

Proceedings of the Lyon Spring School
on

**Advances in
Systems and
Synthetic
Biology**

March 13th - 17th, 2017

IN MEMORIAM
RENÉ THOMAS
(1928 – 2017)

Edited by

Patrick Amar, François Képès, Vic Norris



Hommage à René Thomas

René Thomas est décédé le 9/01/2017 au soir. Nous avons été prévenus à la fin du Colloque de Luminy, par sa fille Isabelle, qu'il avait été mis sous oxygène à son domicile et que ses enfants étaient tous rassemblés dans sa demeure de Rhode St Genèse, ce qui nous avait beaucoup ému, tant il représentait pour beaucoup d'entre nous : selon les cas, un modèle, un inspirateur, un directeur de thèse, un grand frère, un ami (de 40 ans dans mon cas, sans ombrage, que du bonheur...). Nous avons été accueilli dans la pièce où reposait son cercueil par une projection magnifique de la Dibona, son aiguille préférée du massif de la Meije, dans l'Oisans, vue sous son profil le plus majestueux et tentateur. Les hommages se sont succédés, avec des pleurs et des rires, et des sentiments très forts comme il avait su en susciter chez tous ses parents et amis.

Sa belle fille Françoise Nyssen m'a demandé d'écrire un livre sur lui, je vous en livre la première page et un poème écrit à chaud (ou plutôt à froid, à grand froid...) à l'annonce de son décès.

René Thomas, un bon génie du doute et de la création

Y rôde le saint de la genèse... C'est ainsi que m'avait été décrit, de façon alléchante et complotice, l'antre de René. Le saint était un joyeux génie sautillant, émerveillé permanent du monde et de ses découvertes. Généticien astronome, il aurait plu à son idole Christophe, car il avait des facettes du savant Cosinus et des facéties dignes du Sapeur. Sa muse Inga n'avait rien de Madame Fenouillard et elle le stimulait en tant que premier public et première servie, tant il avait besoin d'aimer, d'aider et d'apprendre. Nous allons, au cours de ces quelques pages, découvrir un homme délicieux et un savant d'exception, ces deux qualités étant rares et indépendantes, ce qui augmente la petitesse de leur occurrence chez un même individu.

Au travers de cette destinée individuelle, nous parlerons de son influence sur cinquante ans de bruit et de fureur scientifique, et de sa marque indélébile sur le futur. Comme le disait un grand ami commun, membre du même cercle des savants poètes disparus (mon maître Jacques Besson, frère quasi-jumeau de Schützenberger, qui se présentait comme un fantôme surgi du cinquecento italien par le miracle de la réincarnation et par la grâce de la toile – cf. https://fr.wikipedia.org/wiki/Jacques_Besson – né à Colombière, hameau de la vallée d'Oulx, aujourd'hui en Italie, un peu à l'est de Briançon et voisin du Val des Prés, très cher à René), les hommes sont fils de leurs oeuvres autant que l'inverse, et c'est d'autant plus vrai qu'ils ont un sourire de soie devant la femme qu'ils aiment, des yeux qui pétillent à l'approche d'un ami et une excitation de vieux bouc (disait René...) devant toute nouveauté scientifique – cela augmentant avec l'âge, ce qu'il fait qu'ils ne vieillissent jamais.

C'est donc parti pour quelques aventures gourmandes, sans doute un peu imaginaires, mais reconstruites par un souvenir aimant de fidèle, de disciple complice et de compère compréhensif.

En 1980, à la demande de René, je devais rechercher un Grenoblois, avec un fort accent du midi, qui disait indifféremment à ses ouailles étudiantes (filles et garçons) auxquelles il apprenait à grimper : « A droite ton pied, grros con(gg)! Plus à gauche la main, gror(rr)os con(gg)! ». Je mis rapidement la main (et le pied) sur (et chez) Louis Lliboutry, Grenoblois mythique, fondateur de la glaciologie française, auteur des premiers relevés topo-géologiques de Patagonie et co-équipier de notre autre Grenoblois célèbre, Lionel Terray, dans la conquête du Fitz-Roy, un des sommets emblématiques (avec les pics Mermoz, Saint-Exupéry et Guillaumet) des Andes australes. J'organisai ensuite la rencontre à Grenoble, l'année suivante. A peu près de même taille, Louis (le pionnier de la glaciologie, féru de topographie et d'alpinisme) et René (le pionnier des acides nucléiques, féru d'astronomie et d'alpinisme) se sont étreints comme des frères... Aussi facétieux l'un que l'autre, ils ont évoqué leur première étreinte : « Comment m'as-tu encordé pour me descendre de cette vire de la Dibona où j'étais tombé, j'étais aux trois-quarts inconscient ? » « Facile, vieux frère, j'ai mis tes bras autour de mon

cou, et en avant la descente en rappel ! » « Hèlà, sans corde ? » « Qu'est-ce que tu crois ? C'était plus facile de se débarrasser de toi en cas de problème ! » René blémit, Louis sourit : « Rassure-toi, dans ces cas-là, ils te serrent plus fort que les noyés... et le secours en montagne allait arriver en bas de la Dibona ! », ce secours en montagne, que René, petit maillon d'une longue chaîne avait contribué, sans le savoir, à créer, en prêtant ses cordes à son ami François Henry (Bruxellois de 22 ans à l'époque), qui, avec Jean Vincendon, allait écrire une grande page de l'histoire de l'alpinisme au Mont-Blanc en 1956. René ne les avait pas accompagnés, car il était à l'époque en pleine aventure du phage λ cf. Thomas R. "Effects of chloramphenicol on genetic replication in bacteriophage lambda", *Virology* 9, (1959) 275–283), et il se contentait alors de ses chers rochers de Freÿr...

Petite ode à René de Bruxelles

Adieu l'ami des bruines,
 Adieu, Waterloo triomphant,
 Sous ton rire, O René !
 Evoquant d'une envolée magique,
 La garde, les gènes, le Sapeur,
 Que je voyais entrer souriants
 Dans la même fournaise.
 Naguère, dans les calanques,
 J'ai mis mes pas dans les tiens
 Et tous les vieux complices,
 n'avaient que toi et ton sourire
 dans leurs pensées muettes.
 Le mistral tourmentait les pins
 Et, remontés au col,
 Un grand froid nous saisit,
 Comme si une ombre passait
 Devant notre soleil.
 Revenus au Centre,
 Dont tu aimais la bastide provençale,
 Pleins de ton souvenir,
 Nous avons parlé arabesques,
 ARNs qu'on nature et dénature,
 Depuis que tu l'as dit au monde,
 Dans ta vie pastorienne,
 Et qui se portent comme des charmes
 Dans la nouvelle biologie de synthèse.
 Chaque fois, tous tes concepts,
 Revenaient comme une vieille antienne,
 Eclairés par ton phare
 Indiquant les nœuds, les réseaux, et le port :
 La logique, les systèmes, les itérations de René,
 Renaissant comme des mots neufs
 Dans la bouche des jeunes savants :
 René a dit, René a fait, René a ri,
 Tu me fais pleurer, grand frère.

Jacques Demongeot, 10.01.2017, Boolean Day.

"The system will escape at 1001"

(René Thomas, Richard d'Ari, *Biological Feedback*, CRC Press : Boca Raton, 1989, p. 53)

“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, El Houssine Snoussi, Ines Thiele, Birgit Wiltschi.



ACKNOWLEDGEMENTS

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

Of course the organisation team would like to express gratitude to all the staff of the *Hôtel des Congrès* for the very good conditions we have found during the conference.

Special thanks to the Epigenomics project for their assistance in preparing this book for publication. The cover photography shows a view of the *Quai Saint-Vincent*, on the left bank of the *Saône* river at Lyon.

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

They were:

- Centre National de la Recherche Scientifique (CNRS):
<http://www.cnrs.fr>
- Genopole® Évry:
<http://www.genopole.fr>
- Institut National des Sciences Appliquées (INSA) LYON:
<http://www.insa-lyon.fr>
- Institut Mérieux:
<http://www.institut-merieux.com>
- Institut National de Recherche en Informatique et en Automatique (INRIA):
<http://www.inria.fr/>
- La Région Auvergne-Rhône-Alpes
- Communautés de Recherche Académique Santé Rhône-Alpes:
<http://www.arc1-sante.rhonealpes.fr>
- GDR CNRS 3003 Bioinformatique Moléculaire:
<http://www.gdr-bim.u-psud.fr>

THE EDITORS



INVITED SPEAKERS

GABRIELE BERG	Institute of Environmental Biotechnology (Head), TU Graz, AT
ROSS CARLSON	Center for biofilm engineering, Montana State U., Bozeman, US
WILLIAM E. BENTLEY	Fishell Inst. of biomedical devices, U. Maryland, US
MIKI EBISUYA	Lab. for Reconstitutive Developmental Biology, RIKEN Quantitative Biology Center, Kobe, JP
HERMANN-GEORG HOLZHUTTER	Inst. of Biochemistry, Charité-Universitätsmedizin Berlin, DE
ELKE K. MARKERT	Integrative Cancer Metabolism, Inst. of Cancer Sciences, U. Glasgow, UK
INA KOCH	Molecular Bioinformatics, Goethe U. Frankfurt, DE
JAMES LOCKE	Sainsbury Laboratory Cambridge, University of Cambridge, UK
DANIEL SEGRÉ	Dynamics and evolution of biochemical networks, Boston U., US
ORKAN TELHAN	School of Design, U. of Pennsylvania, Philadelphia, US
SOPHIE VASSEUR	Centre de recherche en Cancérologie de Marseille, Marseille, FR

CONTENTS

PART I INVITED TALKS 13

DYNAMICAL SYSTEMS BIOLOGY & CONSORTIA

JAMES LOCKE

Noisy and dynamic gene regulation at the single cell level . . . 13

DANIEL SEGRÉ

Dynamics and evolution of metabolic interactions in microbial ecosystems 15

SYNBIO & TISSUE MECHANICAL TRANSDUCTION

MIKI EBISUYA

Synthetic cell-cell communications mimicking multicellular development 17

WILLIAM E. BENTLEY

Synthetic Biology and Biofabrication for Conveyance of Molecular Communication 19

MICROBIOME

ROSS CARLSON

In silico and in vitro analysis of resource allocation in biofilm consortia 21

GABRIELE BERG

To Be Announced 23

COMPUTATIONAL APPROACHES

HERMANN-GEORG HOLZHUTTER

Kinetic Modeling of Liver Metabolism 25

INA KOCH

Petri nets for modeling biochemical systems 27

ORKAN TELHAN

New Mediums of Life 29

METABOLISM**SOPHIE VASSEUR***To Be Announced* 31**ELKE K. MARKERT***Classifying tumours by RNA expression patterns:
Hallmarks, signalling pathways and metabolic pathways
across cancers* 33**PART II ARTICLES 35****ANNIKA RÖHL, ALEXANDER BOCKMAYR***Reaction Splitting and Minimum Sets of
Elementary Flux Modes* 37**GIOVANNI CIATTO, ELISABETTA DE MARIA, CINZIA DI GIUSTO***Spiking Neural Networks as Timed Automata* 55**BERTRAND BEAUVOIT, SOPHIE COLOMBIÉ, RAZANNE ISSA,
JEAN-PIERRE MAZAT, CHRISTINE NAZARET, SABINE PERES***HSM, a Reduced Model of Central Carbon Metabolism.
A Dynamical Approach.* 69**RAJEEV KHOODEERAM, GILLES BERNOT,
JEAN-YVES TROSSET***An Ockham Razor model of energy metabolism* 81**OPHELIE GUINAUDEAU, GILLES BERNOT, ALEXANDRE MUZY,
FRANCK GRAMMONT***Abstraction of the structure and dynamics of the biological
neuron for a formal study of the dendritic integration* 103**NGUYEN V. N. TUNG, MARIE BEURTON-AIMAR,
SOPHIE COLOMBIÉ***Studying Metabolic Networks structure: Elementary Flux Modes
vs Minimal Cut Sets Analysis.* 121

**DIANA SCHEPENS, ASHLEY E. BECK, JEFFREY J. HEYS,
TOMAS GEDEON, ROSS P. CARLSON**
*The Benefits of Resource Partitioning and Division of Labor
in Microbial Consortia* 137

KIRSTEN JENKINS, EDINA ROSTA, ATTILA CSIKASZ-NAGY
Coupled versus consecutive positive feedback loops 149

**CLÁUDIO FRAZÃO, DÉBORA TRICHEZ, THOMAS WALTHER,
JEAN-MARIE FRANÇOIS**
*Rational enzyme engineering towards optimization of
a synthetic 1,3-propanediol pathway* 165

PART III LIST OF ATTENDEES 172

Noisy and dynamic gene regulation at the single cell level

James Locke¹

¹ Sainsbury Laboratory Cambridge, University of Cambridge, UK

Abstract

Our lab is focused on developing a quantitative understanding of signal integration and gene circuit dynamics at the single cell level. It is critical to observe cellular behaviour at the single cell level as traditional approaches that take an average from a population can obscure heterogeneous responses and novel dynamics. We combine mathematical modelling and timelapse microscopy to examine gene circuit dynamics in a variety of model organisms including *B. subtilis*, *Cyanobacteria* and *Arabidopsis*. We are attempting to understand how gene regulatory dynamics are generated, how dynamic gene regulatory pathways are coupled together, and what role cellular noise can play in development.



Dynamics and evolution of metabolic interactions in microbial ecosystems

Daniel Segré¹

¹ Dynamics and evolution of biochemical networks,
Boston Univ., US

Abstract

Metabolism, in addition to being the “engine” of every living cell, plays a major role in the cell-cell and cell-environment relations that shape the dynamics and evolution of microbial communities, e.g. by mediating competition and cross-feeding interactions between different species. Despite the increasing availability of metagenomic sequencing data for numerous microbial ecosystems, fundamental aspects of these communities, such as the unculturability of many isolates, and the conditions necessary for taxonomic or functional stability, are still poorly understood. We are developing mechanistic computational approaches for studying the dynamics and evolution of interactions between different organisms based on the knowledge of their entire metabolic networks. In particular, our open source platform for the Computation of Microbial Ecosystems in Time and Space (COMETS) combines metabolic models with convection-diffusion equations to simulate the spatio-temporal dynamics of metabolism in microbial communities. COMETS has been experimentally tested on small artificial communities, and is scalable to hundreds of species in complex environments. I will discuss recent developments and challenges in the study of natural and synthetic microbial communities.



Synthetic cell-cell communications mimicking multicellular development

Miki Ebisuya¹

¹ Lab. for Reconstitutive Developmental Biology, RIKEN Quantitative Biology Center, Kobe, Japan

Abstract

Cell-cell communication plays crucial roles in multicellular development. One way to understand how cell-cell communication works is to recreate or reconstitute the communication mechanisms *in vitro*. In our lab, we create synthetic gene circuits in mammalian cell culture that make neighboring cells communicate with each other, leading to self-organized behaviors of the cells, such as intercellular symmetry breaking and cellular patterning. I will discuss what we have learned from the reconstituted cell-cell communication systems.



Synthetic Biology and Biofabrication for Conveyance of Molecular Communication

William E. Bentley¹

¹ Fishell Institute of biomedical devices, U. Maryland, US

Abstract

The ability to interconvert information between electronic and biologic systems has already transformed our ability to record and actuate biological function (e.g., EEG, EKG, defibrillators). In parallel, we have begun to demand biological connectivity from electronic consumer products (fitbit, cell phones, etc.). There are significant gaps, however, that must be overcome before biological information can be seamlessly conveyed and before biological function can be electronically “programmed”. A communication gap exists whereby the common vectors for information flow in biology are ions and molecules; they are electrons and photons in electronics. Since there are essentially no “free” electrons in biological systems, there is essentially no direct “translator” of electrons to molecules and vice versa. Gaining access to molecular communication is essential as molecules are the primary vector that drives biological function. There is also a fabrication gap to overcome. It is difficult to construct microelectronic devices that include labile biological components. We are developing tools of “biofabrication” that enable facile assembly of biological components within devices that preserve their native biological function. By recognizing that biological redox active molecules are a biological equivalent of an electron-carrying wire, we have developed biological surrogates for electronic devices, including a biological redox capacitor. We have also turned to synthetic biology to provide a means to sample, interpret and report on biological information contained in molecular communications circuitry. Finally, we have developed synthetic genetic circuits that enable electronic actuation of gene expression. This presentation will introduce the concepts of molecular communication that are enabled by integrating relatively simple concepts in synthetic biology with biofabrication. Our presentation will show how engineered cells represent a versatile means for mediating the molecular “signatures” commonly found in complex environments, or in other words, they are conveyors of molecular communication.



***In silico* and *in vitro* analysis of resource allocation in biofilm consortia**

Ross P. Carlson¹

¹ Dept. of Chemical and Biological Engineering, Center for Biofilm Engineering, Thermal Biology Institute, Montana State University, Bozeman, MT 59717, USA

Abstract

Biofilms are ubiquitous in medical, environmental, and engineered microbial systems. The majority of naturally occurring microbes grow as mixed species biofilms. These complicated consortia are comprised of a large number of cell phenotypes with complex interactions and self-organize into three-dimensional structures. While foundational to the vast majority of microbial life on the planet, the basic design principles including resource allocation strategies of consortia biofilms are still poorly understood.

The overarching goal of this research is to develop an experimentally driven, predictive multiscale modeling framework that generates quantitatively accurate predictions of biofilm dynamics, species distributions and responses to perturbations. The biofilm models are formulated by combining genome-scale metabolic reconstructions of individual microorganism species with reaction-transport equations for nutrient and metabolic byproducts. The biofilm models are multiscale with respect to both time and length scales, with the metabolic models linking genes and proteins to cellular dynamics and the consortia models linking individual cells to community dynamics. The *in silico* research is complemented with *in vitro* studies of spatially resolved biofilm measurements to quantify the physiologies of consortia microorganisms.

The research is being developed using two experimental consortia including an Archaeal extremophile consortium from a Yellowstone National Park hot spring and a medically-relevant, three species, bacterial chronic wound system. These distinct, yet tractable, systems are providing insight into basic consortia structure-function relationships and the competitive partitioning of limiting resources between interacting microbial cells. The studied ecological theories and design principles, including the maximum power principle, are believed relevant to many naturally occurring and engineered consortia.



Understanding functions and interactions of plant microbiomes

Gabriele Berg¹

¹ Institute of Environmental Biotechnology, Graz University of Technology,
8010 Graz, Austria

Abstract

The plant microbiome is a crucial factor for plant growth and health [2]. Although this fact is known more than 100 years, multi-omics technologies revolutionized this field of research. Studies revealed deep insights into microbiome functions. Surprisingly, the functions show high cross-kingdom similarities in general but significant differences in detail. In addition, a high proportion of signatures were identified that was assigned to cell-cell communication and interaction. Exemplarily, analyses of moss, lichen and plant microbiomes by multi-omics will be explained [1, 3]. All insights can be transferred in specific applications in biotechnology. For example, the new perspectives influence plant protection approaches such as biocontrol in agriculture [4]. The development of new tools may impact i) the detection of new bio-resources for biocontrol and plant growth promotion, ii) the optimization of fermentation and formulation processes for biologicals, iii) stabilization of the biocontrol effect under field conditions and iv) risk assessment studies for biocontrol agents. Finally, plant microbiomes are an important reservoir for bioactive molecules; therefore omics technologies can be used to discover new bioactive metabolites, e.g. antibiotics, enzymes and volatiles [5]. Moreover, the central role of the plant microbiome and their beneficial inhabitants for human and environmental health will be analyzed.

References

- [1] Aschenbrenner *et al.* 2016. *Frontiers in Microbiology* 7:180.
- [2] Berg *et al.* 2014. *Frontiers in Microbiology* 5:491.
- [3] Berg *et al.* 2015. *Frontiers in Microbiology* 6:1311.
- [4] Berg 2015. *Microbial Biotechnology* 8:5-7.
- [5] Müller *et al.* 2016. *J Biotechnology* 235:171-80.



Kinetic Modeling of Liver Metabolism

Hermann-Georg Holzhütter¹

¹ Inst. of Biochemistry, Charité-Universitätsmedizin, Berlin, DE

Abstract

In my presentation I will explain the design, implementation, numerical simulation and predictive value of physiology-based kinetic models of the hepatocyte metabolism. The first part is devoted to a kinetic models of the hepatic glucose metabolism that allows to assess the relative importance of different modes of enzyme regulation and their mutual interdependencies in the hepatic control of plasma glucose homeostasis. In the second part I will present a tissue-scale model of the smallest functional liver subunit which takes into account metabolic communication among liver cells by concentration gradients of metabolites and hormones in the blood stream and interstitial space and differential expression of metabolic enzymes depending on the relative spatial position of hepatocytes with respect to the arterial and venous termini of the capillary (metabolic zonation). Finally, I will report on our recent attempts to establish a large-scale metabolic model that encompasses all main pathways of the central carbohydrate and lipid metabolism.



Petri nets for modeling biochemical systems

Ina Koch¹

¹ Molecular Bioinformatics, Goethe U. Frankfurt, DE

Abstract

The lecture introduces the foundations of Petri nets with a focus on the invariant analysis. New developments for modeling signal transduction pathways will be presented. Limits, advantages and challenges will be discussed.



New Mediums of Life

Orkan Telhan¹

¹ School of Design, U. of Pennsylvania, Philadelphia, US

Abstract

Orkan Telhan's research focuses on bridging theories of design and computation with artificial life and synthetic biology research. His work investigates the epistemic origins of making and manipulating life with a focus on the artistic, cultural, and ethical implications of living, semi-living, and life-like artifacts.

In this talk, Telhan will discuss a series of design experiments that interrogate life sciences at different levels from the biosynthesis of endangered molecules to microbial prosthetics, and alternative design tools that network algorithms, organisms, and people.



Metabolic flexibility of pancreatic cancer

Sophie Vasseur¹

¹ Centre de recherche en Cancérologie de Marseille, FR

Abstract

Among cancers in critical clinical needs, pancreatic ductal adenocarcinoma (PDAC) is the most intractable with a 5-year survival below 5% and therefore represents one of the most fatal disease among solid cancer (based on the ratio “Cases/Death”). As a silent killer, its symptoms are so insidious that most people are not diagnosed until the disease has advanced beyond the stage where surgical resection is possible. The presence of a prominent non-tumoral cell compartment within the tumor (a main characteristic of PDAC) directly impacts on patient clinical outcomes. In PDAC, cancer cells are “isolated” by a fortress of stromal cells, composed of very few blood vessels, which distorts the normal architecture of pancreas and limits the oxygen and nutrient diffusion within the tissue. This severe hypoxic environment provides a strong metabolic pressure able to regulate tumor cell growth and favors survival of the most aggressive malignant cells. Therefore, during its development this tumor must harbor metabolic pathways which are probably tied in a complex dialogue with highly metabolic host tissues as muscle, liver and adipose tissue, illustrating PDAC as a real metabolic disease. Hence, a comprehensive picture of the metabolic reprogramming of PDAC is needed in order to develop the most appropriate metabolic therapeutic approach to impede pancreatic cancer progression.

Using PDAC genetic engineered mouse models (Pdx1-Cre; LSL-Kras^{G12D}; Ink4a^{fl/fl} mice), we revealed the transcriptomic metabolic signature of PDAC. In combination with un-/targeted metabolomic approaches, we also highlighted the dynamic of the metabolic pathways preferentially activated in PDAC. Based on these data, our group is now establishing a more comprehensive stratification of the metabolic processes occurring at aggressive adenocarcinoma stage in order to fundamentally advance our understanding of the metabolic demand of PDAC and to select the best metabolic targets for therapy.



Classifying tumours by RNA expression patterns: Hallmarks, signalling pathways and metabolic pathways across cancers

Elke K. Markert¹

¹ Integrative Cancer Metabolism, Inst. of Cancer Sciences,
U. Glasgow, UK

Abstract

With the curation of substantial databases containing high-throughput data from human cancer samples, systems approaches in cancer research are gaining traction. There are two main lines of comparative analysis: “integrative” approaches usually refer to combining data from different high-throughput platforms within the same cohort, while “pan-cancer” approaches aim to understand common and distinct features of the disease across tissues. Here we will focus on the latter, studying molecular features of tumours through their mRNA expression profiles. Our aim is to establish a very basic classification system that would reflect the most dominant expression patterns. The resulting classes can then be analysed for finer structures that might be instrumental in creating the overall effect, for instance the activity of distinct signalling and metabolic pathways.



PART II ARTICLES



Reaction Splitting and Minimum Sets of Elementary Flux Modes

Annika Röhl¹ and Alexander Bockmayr¹

¹ Department of Mathematics and Computer Science, Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany

Abstract

Genome-scale metabolic network reconstructions are widely used in systems biology for *in silico* studies of cellular metabolism. A common approach to analyse these models are so-called *elementary flux modes* (EFMs), which correspond to minimal functional units in the network. Already for medium-sized networks, it may be impossible to compute all the EFMs, due to their huge number. From a practical point of view, this is also often not necessary, because a subset of EFMs may already be sufficient to answer the given biological question. Here we present a method based on reaction splitting to compute minimum sets of EFMs that generate the whole flux space and that can be used to obtain additional EFMs in case these are needed.

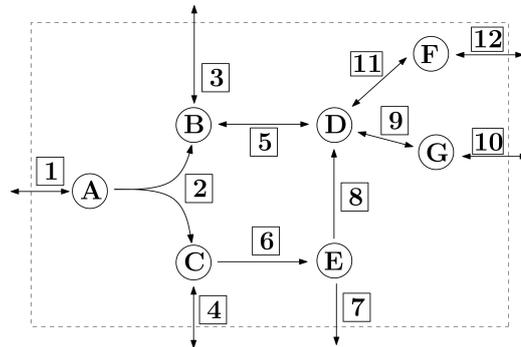
1 Introduction

In systems biology, genome-scale metabolic network reconstructions are widely used to develop *in silico* models of cellular metabolism. Typically the metabolic network is assumed to be in steady-state, i.e., we analyse the so-called *flux cone*

$$\Gamma_S = \left\{ v \in \mathbb{R}^{\mathfrak{R}} \mid Sv = 0, v_{\text{Irr}} \geq 0 \right\}.$$

Here, $S \in \mathbb{R}^{\mathfrak{M} \times \mathfrak{R}}$ denotes the stoichiometric matrix for a set of (internal) metabolites \mathfrak{M} and a set of reactions \mathfrak{R} , see Fig. 1 for a small example. The set $\text{Irr} \subseteq \mathfrak{R}$ contains the irreversible and the set $\text{Rev} = \mathfrak{R} \setminus \text{Irr}$ the reversible reactions. The vectors $v \in \Gamma_S$ are called (feasible) *flux vectors* and can be interpreted as steady-state flux distributions over the network.

To define the elementary flux modes we need the notion of *support* of a flux vector $v \in \Gamma_S$, which is the set $\text{supp}(v) := \{i \in \mathfrak{R} \mid v_i \neq 0\}$ of active reactions in v . An *elementary flux mode* (EFM) [32] is a feasible flux vector $v \in \Gamma_S \setminus \{0\}$ of minimal support (with respect to set inclusion), i.e., there exists no $w \in \Gamma_S \setminus \{0\}$ such that $\text{supp}(w) \subsetneq \text{supp}(v)$. EFMs define minimal sets of reactions that can operate together in steady-state. Minimality means that if any of the reactions is deleted, then the whole flux cannot operate anymore in steady-state.



$$\begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \end{pmatrix}$$

The corresponding stoichiometric matrix.

Figure 1: Metabolites $\mathfrak{M} = \{A, B, \dots, G\}$, reactions $\mathfrak{R} = \{1, 2, \dots, 11, 12\}$, irreversible reactions $\text{Irr} = \{2, 6, 7, 8\}$.

In the example network of Fig. 1 there are 18 EFMs with the following sets of active reactions: $\{1, 2, 3, 4\}$, $\{1, 2, 3, 5, 6, 8\}$, $\{1, 2, 3, 6, 7\}$, $\{1, 2, 3, 6, 8, 9, 10\}$, $\{1, 2, 3, 6, 8, 11, 12\}$, $\{1, 2, 4, 5, 9, 10\}$, $\{1, 2, 4, 5, 11, 12\}$, $\{1, 2, 5, 6, 7, 9, 10\}$, $\{1, 2, 5, 6, 7, 11, 12\}$, $\{1, 2, 5, 6, 8, 9, 10\}$, $\{1, 2, 5, 6, 8, 11, 12\}$, $\{3, 4, 5, 6, 8\}$, $\{3, 5, 9, 10\}$, $\{3, 5, 11, 12\}$, $\{4, 6, 7\}$, $\{4, 6, 8, 9, 10\}$, $\{4, 6, 8, 11, 12\}$, $\{9, 10, 11, 12\}$. The sets $\{3, 5, 9, 10\}$, $\{3, 5, 11, 12\}$, $\{9, 10, 11, 12\}$ are reversible EFMs, which can carry flux in both directions.

EFMs are a popular approach to analyse metabolic networks because every steady-state behaviour of the network can be represented with help of the EFMs [31, 32]. Formally speaking, the EFMs define a *conic basis* of the flux cone Γ_S , i.e., each flux vector $w \in \Gamma_S$ can be represented as a conical combination of the EFMs: $w = \sum_{i=1}^{n_{\text{EFMs}}} \lambda_i v^i$, where v^i are EFMs, $\lambda_i \geq 0$ and n_{EFMs} is the total number of EFMs in the network. While the EFMs are of great theoretical and practical interest, already for medium-sized metabolic networks, it is very difficult to enumerate all of them, due to their exponential number. Thus, over the last 20 years various methods have been developed to compute some or all EFMs in a given metabolic network.

One of the first approaches was the *Double Description Method* [6] and

refinements thereof such as [1, 7, 23] or the nullspace algorithm in [34]. Other approaches are based on graph theory like in [24, 33]. All these methods can compute the whole set of EFMs. In practice, however, they often do not terminate because the number of EFMs gets too large [15] and there is lack of memory. Therefore, a natural idea is to compute a smaller set of EFMs that can still represent the full flux space. In [21, 29], so-called *extreme pathways* were proposed. They correspond to a smaller set of EFMs that generate the whole flux cone, but they also include EFMs that are not needed. Some methods focus on the feasibility of EFMs regarding underlying regulatory rules [12] or thermodynamics [8]. Another possibility is to compute a random subset of EFMs by sampling [20]. The resulting set of EFMs does not generate the whole flux space, but it can give some useful information about it. In [17], a minimum description of the flux cone Γ_S is introduced, based on so-called *minimal metabolic behaviours* (MMBs). Each MMB corresponds to a subset of EFMs with the same set of active irreversible reactions. It can be shown that in order to obtain a minimal conic basis of Γ_S , for each MMB one of the corresponding EFMs has to be chosen [17], see also [11]. More recent approaches based on mixed-integer programming [5, 26] focus on generating a limited number of *shortest* EFMs, i.e., EFMs having only a small number of active reactions. Further extending this work, [4, 22] develop algorithms for enumerating only those EFMs that involve a given set of target reactions. Principal elementary mode analysis [35] is a method to identify a combination of EFMs that best capture the patterns observed in a given set of flux data. For a recent survey on EFMs and their applications, we refer to [37].

Our goal in this paper is to use reaction splitting to identify minimum sets of EFMs sufficient to describe the whole flux space. We start in Sect. 2 by discussing different forms of reaction splitting and then introduce the new concept of *minimum set of elementary modes* (MEMO). In Sect. 3, we describe an algorithm to compute MEMOs and report on some computational results. Sect. 4 concludes and gives an outlook on further work.

2 Methods

2.1 Polyhedral cones

We start by some basic definitions and theorems related to polyhedral cones. For further information we refer to [30]. A subset $\Gamma \subseteq \mathbb{R}^n$ is called a *cone* if any *conic combination* of two elements $x, y \in \Gamma$ belongs to Γ again, i.e., $\lambda x + \mu y \in \Gamma$, for any non-negative $\lambda, \mu \in \mathbb{R}_{\geq 0}$. A cone Γ is called *polyhedral* if there exists a matrix $A \in \mathbb{R}^{m \times n}$ such that

$$\Gamma = \{x \in \mathbb{R}^n \mid Ax \geq 0\}.$$

A cone Γ is *finitely generated* if there exists a finite set of generators $\{g^1, \dots, g^t\} \subseteq \Gamma$ such that every element $x \in \Gamma$ can be written in the form $x = \sum_{k=1}^t \lambda_k g^k$, for some non-negative $\lambda_k \in \mathbb{R}_{\geq 0}$. By a classical theorem of Farkas-Minkowski-Weyl, a cone is polyhedral if and only if it is finitely generated.

For a polyhedral cone $\Gamma = \{x \in \mathbb{R}^n \mid Ax \geq 0\}$, the linear subspace

$$\Lambda = \{x \in \mathbb{R}^n \mid Ax = 0\} = \Gamma \cap (-\Gamma)$$

is called the *lineality space* of Γ . The cone Γ is called *pointed* if $\Lambda = \{0\}$. This means that whenever $x \in \Gamma, x \neq 0$, we have $-x \notin \Gamma$. In other words, Γ does not contain any line $\{\lambda x \mid \lambda \in \mathbb{R}\}$, for $x \neq 0$. By basic linear algebra, the cone Γ is pointed if and only if the matrix A has full column rank, i.e., $\text{rank}(A) = n$.

If Γ is pointed, there is a unique minimum set of generators $\{g_1, \dots, g_t\} \subseteq \Gamma$ corresponding to the extreme rays of Γ . Two rays $x, x' \in \Gamma$ are considered identical if $x' = \lambda x$ for some $\lambda > 0$. A ray $0 \neq x \in \Gamma$ is called an *extreme ray* (ER) of Γ , if there exist no two linearly independent rays $x^1, x^2 \in \Gamma$ such that $x = x^1 + x^2$.

The set of feasible flux vectors in a metabolic network defines a polyhedral cone

$$\Gamma_S = \left\{ v \in \mathbb{R}^{\mathfrak{R}} \mid Av \geq 0, A = \begin{pmatrix} S \\ -S \\ E_{\text{Irr}} \end{pmatrix} \right\},$$

which is called the *flux cone*. Here E_{Irr} denotes an identity matrix where the rows corresponding to the reversible reactions have been deleted. Thus $v_i \geq 0$, for all $i \in \text{Irr}$, and irreversible reactions can carry flux in only one direction.

If all reactions are irreversible, the flux cone Γ_S is pointed and there exists a one-to-one correspondence between the extreme rays of Γ_S and the elementary flux modes [32]. In general, however, the metabolic network may contain reversible reactions together with reversible flux vectors $v \neq 0$ for which $v, -v \in \Gamma_S$, and the flux cone is non-pointed.

2.2 Splitting variables

In many cases it is desirable to have a cone $\Gamma = \{x \in \mathbb{R}^n \mid Ax \geq 0\}$ which is pointed. If the system $Ax \geq 0$ includes constraints $x_j \geq 0$, for all $j \in \{1, \dots, n\}$, then the resulting cone will be pointed. In general, however, a variable $x_j \in \mathbb{R}$ can take negative values, and we cannot simply add the constraint $x_j \geq 0$. To overcome this problem, a well-known method also used in linear programming is to *split* variables. Splitting a variable $x_j \in \mathbb{R}$ means replacing it by two non-negative variables $\bar{x}_j, \bar{x}_{n+j} \geq 0$,

such that $x_j = \bar{x}_j - \bar{x}_{n+j}$. Note that this will change the structure of the cone and increase the dimension of the underlying vector space by 1. To describe this transformation formally, we use a map $\pi_J : \mathbb{R}^n \rightarrow \mathbb{R}^{n+|J|}$, where $J \subseteq \{1, \dots, n\}$ denotes the set of variables $x_j, j \in J$, to be split. For $x \in \mathbb{R}^n$ we get $\pi_J(x) = \bar{x}$ with $\bar{x}_j = x_j$, for all $j \notin J$, and for each $j \in J$:

$$\begin{aligned} \bar{x}_j = x_j & \quad \text{and} \quad \bar{x}_{n+j} = 0 & \quad \text{if } x_j \geq 0, \\ \bar{x}_j = 0 & \quad \text{and} \quad \bar{x}_{n+j} = -x_j & \quad \text{if } x_j < 0. \end{aligned}$$

Applying the map π_J , a polyhedral cone $\Gamma = \{x \in \mathbb{R}^n \mid Ax \geq 0\}$, with $A \in \mathbb{R}^{m \times n}$, is mapped to the polyhedral cone

$$\bar{\Gamma}^J = \{\bar{x} \in \mathbb{R}^{n+|J|} \mid \bar{A}\bar{x} \geq 0\},$$

with the augmented matrix $\bar{A} \in \mathbb{R}^{(m+2|J|) \times (n+|J|)}$, where $\bar{A} = \begin{pmatrix} A & -A_{*,J} \\ E \end{pmatrix}$ and $A_{*,J}$ is the submatrix of A consisting of the columns $A_{*,j}$ with $j \in J$. For all $j \in J$, the matrix E contains as a row the j -th and $(n+j)$ -th unit vector.

The inverse transformation $\pi_J^r : \bar{\Gamma}^J \rightarrow \Gamma$ maps each vector $\bar{x} \in \bar{\Gamma}^J$ to $x = \pi_J^r(\bar{x})$ such that $x_j = \bar{x}_j$, for all $j \notin J$, and $x_j = \bar{x}_j - \bar{x}_{n+j}$, for all $j \in J$. If we apply π_J^r , we say that we *recombine* the variables that were split before. Next we look at the lineality space of $\bar{\Gamma}^J$. Generalising a result from [16], we get:

Lemma 2.1. *Let $\Gamma \subseteq \mathbb{R}^n$ be a polyhedral cone with lineality space Λ . For a set of variables $J \subseteq \{1, \dots, n\}$, the lineality space $\bar{\Lambda}^J$ of the augmented cone $\bar{\Gamma}^J$ is given by:*

$$\bar{\Lambda}^J = \left\{ \begin{pmatrix} x \\ 0 \end{pmatrix} \in \mathbb{R}^{n+|J|} \mid x \in \Lambda \text{ and } x_J = 0 \right\},$$

where $x_J = 0$ means $x_j = 0$, for all $j \in J$.

Proof. For the cone $\Gamma = \{x \in \mathbb{R}^n \mid Ax \geq 0\}$ the lineality space is $\Lambda = \{x \in \mathbb{R}^n \mid Ax = 0\}$. Splitting the variables in J delivers the cone $\bar{\Gamma}^J = \left\{ \begin{pmatrix} x \\ w \end{pmatrix} \in \mathbb{R}^{n+|J|} \mid (A, -A_{*,J}) \begin{pmatrix} x \\ w \end{pmatrix} \geq 0, x_J \geq 0, w \geq 0 \right\}$. For the lineality space we

get:

$$\begin{aligned}
 \bar{\Lambda}^J &= \left\{ \begin{pmatrix} x \\ w \end{pmatrix} \in \mathbb{R}^{n+|J|} \mid (A, -A_{*,J}) \begin{pmatrix} x \\ w \end{pmatrix} = 0, x_J = 0, w = 0 \right\} \\
 &= \left\{ \begin{pmatrix} x \\ 0 \end{pmatrix} \in \mathbb{R}^{n+|J|} \mid (A, -A_{*,J}) \begin{pmatrix} x \\ 0 \end{pmatrix} = 0, x_J = 0 \right\} \\
 &= \left\{ \begin{pmatrix} x \\ 0 \end{pmatrix} \in \mathbb{R}^{n+|J|} \mid Ax = 0, x_J = 0 \right\} \\
 &= \left\{ \begin{pmatrix} x \\ 0 \end{pmatrix} \in \mathbb{R}^{n+|J|} \mid x \in \Lambda \text{ and } x_J = 0 \right\}.
 \end{aligned}$$

□

2.3 Splitting reversible reactions

In the case of metabolic networks, the variables corresponding to the irreversible reactions, by definition, can take only non-negative values. In order to obtain a pointed cone, we can split all reversible reactions into two irreversible ones, see Fig. 2 for an example. This leads to the pointed cone $\bar{\Gamma}_S^{\text{Rev}}$. The uniquely determined extreme rays of this cone are called *extreme currents* [19]. It can be shown that after recombination, they correspond exactly to the EFMs of the metabolic network [7]. In addition, for each split reaction $j \in J$, there exists a 2-cycle $\bar{v} \in \bar{\Gamma}_S^{\text{Rev}}$ with $\bar{v}_j = -\bar{v}_{n+j}$, and $\bar{v}_j = 0$ otherwise. These cycles do not have a biological meaning and can be eliminated (they become zero after applying π_{Rev}^r).

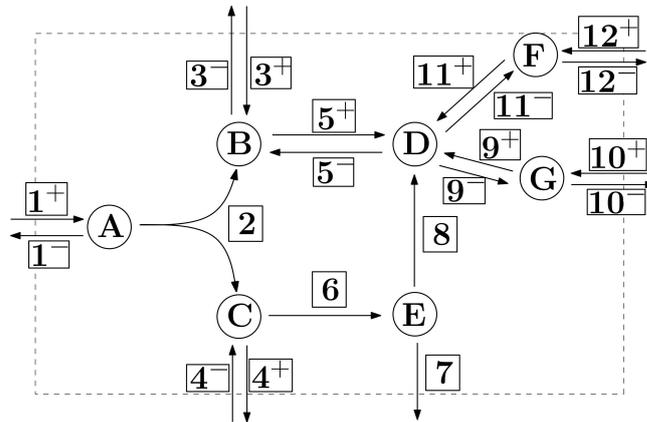


Figure 2: Network where all reversible reactions are split. For example, reaction 3 is split into 3^- and 3^+ .

Splitting all reversible reactions will highly increase the number of variables and thus the dimension of the vector space where the augmented cone lives. Already for medium-sized networks, the number of extreme rays, which corresponds to the number of EFMs (up to the 2-cycles), will be huge and therefore computing the whole set may not be feasible or desirable. There exist methods that split only a subset of the reversible reactions and still allow obtaining a pointed cone. The ERs of the augmented cone define a subset of EFMs in the original cone. The remaining EFMs correspond to rays inside the augmented pointed cone and can be obtained from the ERs by conic combinations.

In the following we introduce a method to determine minimum sets of reversible reactions that have to be split in order to obtain a pointed cone. The resulting sets of ERs (after recombination) define minimum sets of EFMs needed to represent the whole flux space.

2.4 Splitting fully reversible reactions

In [21, 29] it is shown that splitting only the reversible *internal* reactions delivers an augmented cone which is pointed (assuming that there is only one exchange reaction per internal metabolite). This cone is unique and so are its extreme rays. After recombination, these extreme rays are called *extreme pathways (EPs)*. In Fig. 3 an example is shown and the corresponding extreme pathways are given.

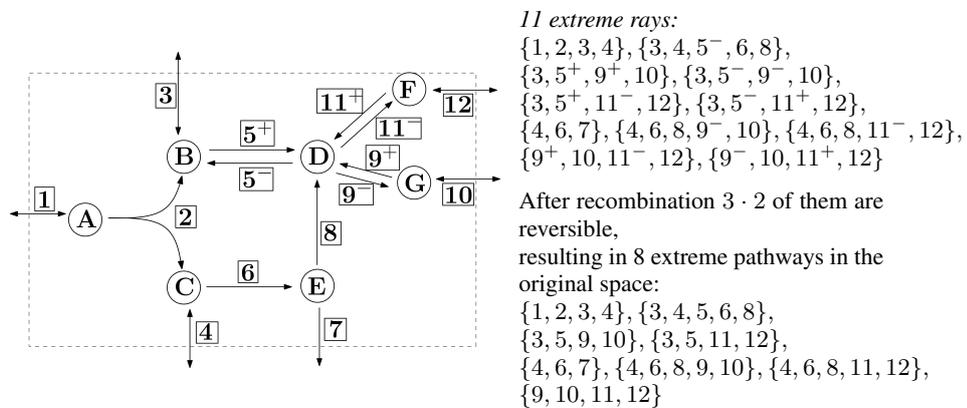


Figure 3: Network where only the reversible internal reactions are split.

To further develop our method we have to distinguish between two types of reversibility [17]. A reversible reaction $j \in \text{Rev}$ is called *fully reversible* if there exists a flux vector $v \in \Gamma_S$ such that $v_j \neq 0$ and $v_i = 0$ for all $i \in \text{Irr}$. Equivalently, $j \in \text{Rev}$ is fully reversible if and only if there is a reversible

flux vector $v \in \Lambda_S$ with $v_j \neq 0$. Otherwise, reaction j is called *pseudo-irreversible*. In our example of Fig. 1 the reactions 3, 5, 9, 10, 11, 12 are fully reversible, whereas reactions 1 and 4 are pseudo-irreversible. Reaction 1 can work in only one direction although it is declared to be reversible in the model. Reaction 4 can work in both directions, but it is not fully reversible. We denote the set of fully reversible reactions with $\text{Frev} \subseteq \text{Rev}$.

Theorem 1. *Let Γ_S be the flux cone of the metabolic network represented by the stoichiometric matrix $S \in \mathbb{R}^{m \times n}$ and the set of irreversible reactions Irr . Splitting all fully reversible reactions delivers a pointed cone $\bar{\Gamma}_S^{\text{Frev}}$.*

Proof. By Lemma 2.1 we have

$$\bar{\Lambda}_S^{\text{Frev}} = \left\{ \begin{pmatrix} v \\ 0 \end{pmatrix} \in \mathbb{R}^{|\mathfrak{A}|+|\text{Frev}|} \mid v \in \Lambda_S \text{ and } v_{\text{Frev}} = 0 \right\}.$$
 According to the definition of fully reversible reactions, these are the only reactions carrying flux in Λ_S . In other words, $v_{\mathfrak{A} \setminus \text{Frev}} = 0$, for all $v \in \Lambda_S$. Thus, if $v \in \Lambda_S$ and $v_{\text{Frev}} = 0$, then $v = 0$. Hence the augmented cone $\bar{\Gamma}_S^{\text{Frev}}$ is pointed. \square

2.5 Splitting a minimum number of fully reversible reactions

We will show in this section that there exist subsets with a minimum number of fully reversible reactions such that after splitting these reactions the augmented cone becomes pointed. These subsets consist of t fully reversible reactions, where t is the dimension of the lineality space of Γ_S . We note that splitting one reaction reduces the dimension of the lineality space by at most 1, see [16]. Therefore at least t fully reversible reactions have to be split in order to obtain a pointed cone.

Before we start, we recall an important concept from matrix theory [9, p.85]. A matrix $B^{\text{rcef}} \in \mathbb{R}^{m \times n}$ is in *reduced column echelon form* if the following properties hold (see Fig. 4):

1. The first non-zero element in column k is a 1 in row j_k , for $k = 1, 2, \dots, r$ (this 1 is called a *pivot*).
2. $1 \leq j_1 < j_2 < \dots < j_r \leq m$ (i.e., for each change in columns from left to right, the pivot appears in a lower row).
3. For $k = 1, \dots, r$, the pivot in column k is the only non-zero element in row j_k .
4. Each of the last $n - r$ columns consists entirely of zeros.

Theorem 2. *Let Γ_S be the flux cone of a metabolic network. If t is the dimension of the lineality space Λ_S , then splitting a certain subset $\text{minFrev} \subseteq \text{Frev}$ of t fully reversible reactions leads to a pointed cone $\bar{\Gamma}_S^{\text{minFrev}}$.*

$$B^{\text{rcef}} = \begin{pmatrix} \textcircled{1} & 0 & 0 & 0 & 0 & 0 \\ * & 0 & 0 & 0 & 0 & 0 \\ 0 & \textcircled{1} & 0 & 0 & 0 & 0 \\ 0 & 0 & \textcircled{1} & 0 & 0 & 0 \\ * & * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 & 0 \\ 0 & 0 & 0 & \textcircled{1} & 0 & 0 \\ * & * & * & * & 0 & 0 \end{pmatrix} \begin{array}{l} \leftarrow j_1 \\ \leftarrow j_2 \\ \leftarrow j_3 \\ \vdots \\ \leftarrow j_r \end{array}$$

Figure 4: B^{rcef} is a matrix in reduced column echelon form, where $*$ are appropriate values in \mathbb{R} .

Proof. We may assume that Γ_S is not pointed, hence $\Lambda_S \neq \{0\}$. Let $B \in \mathbb{R}^{|\mathfrak{R}| \times t}$ be a matrix whose columns $(\tilde{b}^1, \dots, \tilde{b}^t)$ correspond to a basis of the lineality space Λ_S , where $t \geq 1$ is the dimension of Λ_S . By applying elementary column operations, we can obtain the reduced column echelon form B^{rcef} of B , see Fig. 4, which is uniquely determined by B [9, p.85]. The columns (b^1, \dots, b^t) of B^{rcef} define again a basis of Λ_S . The row indices j_1, \dots, j_r for the pivot 1's in B^{rcef} are exactly the indices of the fully reversible reactions we are looking for. To see this, define $\text{minFrev} = \{j_1, \dots, j_r\}$. From $t = \text{rank } B = \text{rank } B^{\text{rcef}} = r$, we get $r = t$. Since $b_{j_k}^k = 1 \neq 0$, for $k = 1, \dots, r = t$, we have $\text{minFrev} \subseteq \text{Frev}$.

After splitting the reactions in minFrev we get the augmented cone $\bar{\Gamma}_S^{\text{minFrev}}$ with lineality space

$$\bar{\Lambda}_S^{\text{minFrev}} = \left\{ \begin{pmatrix} v \\ 0 \end{pmatrix} \in \mathbb{R}^{|\mathfrak{R}| + |\text{minFrev}|} \mid v \in \Lambda_S \text{ and } v_{\text{minFrev}} = 0 \right\}.$$

Since (b^1, \dots, b^t) defines a basis of Λ_S , any $v \in \Lambda_S$ can be written as $v = \sum_{k=1}^t \lambda_k b^k$, for some $\lambda_k \in \mathbb{R}$. The matrix B^{rcef} is in reduced column-echelon form. This means that for each $j \in \text{minFrev}$ there is exactly one b^{k_j} with $b_j^{k_j} \neq 0$, and for all $k \neq k_j$, it holds $b_j^k = 0$. Since for $\begin{pmatrix} v \\ 0 \end{pmatrix} \in \bar{\Lambda}_S^{\text{minFrev}}$ we have $v_j = \sum_{k=1}^t \lambda_k b_j^k = \lambda_{k_j} b_j^{k_j} = \lambda_{k_j} = 0$, for all $j \in \text{minFrev}$, it follows that $\lambda_k = 0$, for all $k \in \{1, \dots, t\}$. This implies $v = 0$ and thus $\bar{\Lambda}_S^{\text{minFrev}} = \{0\}$, which proves that $\bar{\Gamma}_S^{\text{minFrev}}$ is pointed. \square

2.6 Minimum sets of EFMs

After recombination the set of extreme rays of any pointed cone $\bar{\Gamma}_S^{\text{minFrev}}$ obtained by splitting $t = \dim(\Lambda_S)$ fully reversible reactions will be called a

minimum set of elementary modes (MEMO). As we will see in Sect. 3.1, splitting an arbitrary set of t fully reversible reactions does not necessarily lead to a pointed cone. In addition, the minimum set of t fully reversible reactions that have to be split is not unique. Hence the resulting MEMO is also not unique.

We do not further discuss here the minimality of MEMOs and their relationship to minimal metabolic behaviours [17] and the minimal generating set in [11], which will be the topic of a follow-up paper.

3 Results and discussion

3.1 Illustrating example

The MEMOs of a metabolic network correspond to minimum sets of EFMs needed to generate the full flux cone Γ_S . These MEMOs need not be unique. Depending on the reactions that are split, the MEMOs may consist of different flux vectors.

In Fig. 5 we split the fully reversible reactions 3 and 9, which leads to a pointed augmented cone. Thus we can compute a MEMO, which consists of 5 different EFMs. If we split reactions 10 and 12, the augmented cone is again pointed and we obtain another MEMO containing 5 EFMs (see Fig. 6), which is different from the first one.

Splitting arbitrary $t = \dim(\Lambda_S)$ fully reversible reactions does not necessarily deliver a pointed augmented cone. For example, in Fig. 7 the fully reversible reactions 11 and 12 were split. The corresponding augmented cone is not pointed. This can be easily seen because there still exist reversible flux vectors in the transformed network.

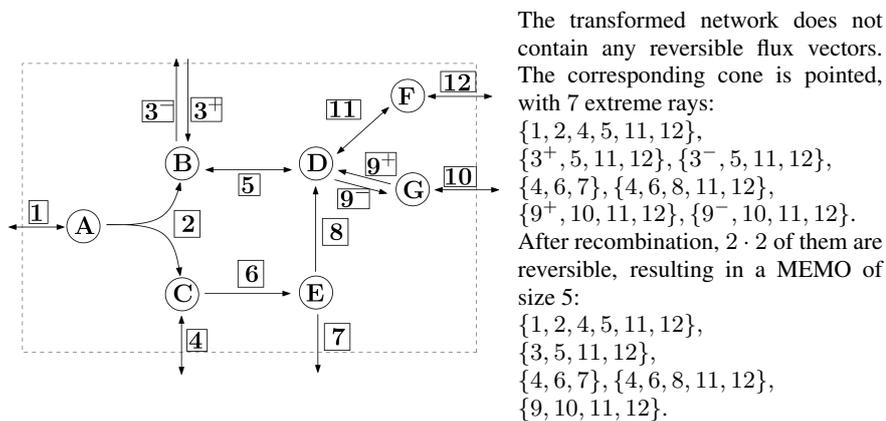


Figure 5: The fully reversible reactions 3 and 9 are split.

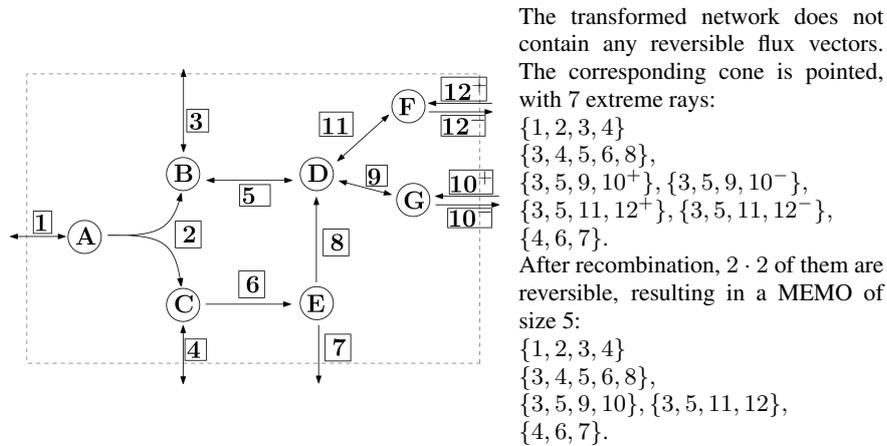


Figure 6: The fully reversible reactions 10 and 12 are split.

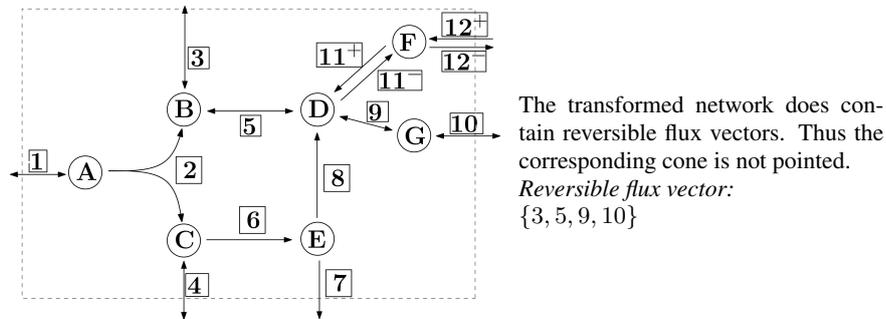


Figure 7: The fully reversible reactions 11 and 12 are split.

3.2 Computational experiments

To compare the size of the MEMOs to the total number of EFMs we implemented our method in MATLAB using POLCO [2] to compute the ERs of a pointed cone. The metabolic network reconstructions were taken from BiGG Models [14], KEGG [13] (together with KEGGtranslator [36]), and BioModels Database [18]. Table 1 summarises the main characteristics of the metabolic networks studied. Since it is usually not possible to enumerate all the EFMs of a genome-scale reconstruction we used the method from [27] to reduce the number of reactions while preserving basic functionalities of the metabolic network. In Table 2, we compare the size of the MEMOs to the number of EPs resp. EFMs, which were computed with POLCO [2] resp. EFMT00L [1]. In the case $t = 0$ the original flux cone is already pointed, i.e., there is no need to split any reaction before computing a MEMO. However, there may still be re-

versible (internal) reactions in the network, resulting in a possible much larger number of EPs or EFMs.

	rxns	mets	rev	frev	t
<i>Escherichia coli</i> carbon metabolism ([3])	34	18	34	34	16
Citrate cycle (TCA) [13]	36	22	32	30	12
Pentose phosphate pathway [13,36]	57	34	38	27	8
Glycolysis / Gluconeogenesis [13,36]	61	32	43	33	13
Pyruvate metabolism [13,36]	81	28	41	24	16
<i>Escherichia coli</i> MG1655 [14]	95	72	46	0	0
<i>Rhizobium etli</i> iOR363 [25]	222	371	71	0	0
<i>Blattabacterium cuenoti</i> iCG238 [10]	323	364	73	5	2
<i>Mus musculus</i> reduced [14]	351	351	177	0	0
<i>M. tuberculosis</i> iNJ661 reduced [14]	427	425	131	0	0
<i>Helicobacter pylori</i> 26695 [28]	452	396	94	31	6
<i>S. Typhimurium</i> LT2 reduced [14]	458	455	142	0	0

Table 1: Characteristics of different metabolic networks.

rxns: number of unblocked reactions of the network

mets: number of metabolites of the network

revs: number of unblocked reversible reactions of the network

frev: number of unblocked fully reversible reactions of the network

t: dimension of the lineality space of the polyhedral cone

4 Conclusion and further work

Based on reaction splitting, we introduced the concept of a minimum set of elementary modes (MEMO) necessary to generate the flux cone of a metabolic network. Next we presented a method to compute these MEMOs. We implemented our algorithm using MATLAB and showed that the size of MEMOs can be by several orders of magnitude smaller than the number of EPs or EFMs.

One drawback is that the MEMOs are not unique because they depend on the set of fully reversible reactions which have to be split. The biological relevance of these different MEMOs has to be further investigated. One can show that the set of irreversible reactions involved in the MEMOs is unique and that they will always be part of the MEMOs independently from the reactions that were split. This set of irreversible reactions is called a minimal metabolic behaviour (MMB) and Larhlimi/Bockmayr already showed in [17] that for each minimal proper face of the flux cone there exists a corresponding MMB. In future work we will further discuss the relationship of MEMOs to MMBs and

	rxns	MEMO	EPs	EFMs
<i>Escherichia coli</i> carbon metabolism [3]	34	14	26	6,421
Citrate cycle (TCA) [13]	36	16	1,306	3,870
Pentose phosphate pathway [13]	57	27	1,582	5,155
Glycolysis / Gluconeogenesis [13]	61	29	1,670	19,464
Pyruvate metabolism [13]	81	45	27,215	47,708
<i>Escherichia coli</i> MG1655 [14]	95	2,572	100,274	N/A
<i>Rhizobium etli</i> iOR363 [25]	222	6,147	N/A	N/A
<i>Blattabacterium cuenoti</i> iCG238 [10]	323	376	N/A	N/A
<i>Mus musculus</i> reduced [14]	351	94,957	N/A	N/A
<i>M. tuberculosis</i> iNJ661 reduced [14]	427	501	22,775	24,233
<i>Helicobacter pylori</i> 26695 [28]	452	150,138	N/A	N/A
<i>S. Typhimurium</i> LT2 reduced [14]	458	97	1,225	1,225

Table 2: Size of MEMOs, number of EPs and EFMs for different networks. For most organisms we used a reduced network such that we were able to compute the MEMOs and EFMs.

rxns: number of unblocked reactions of the network

EPs: number of the extreme pathways

MEMO: size of a minimum set of elementary modes

EFMs: number of the elementary flux modes

N/A: The programs used here (POLCO [2] and EFMTOOL [1]) could not handle the size of the models resp. the number of EPs or EFMs.

the minimal generating set in [11]. We can also show that certain EFMs will never be part of the MEMOs, independently from the reactions which are split.

References

- [1] EFMTOOL. <http://www.csb.ethz.ch/tools/software/efmtool.html>.
- [2] POLCO. <http://www.csb.ethz.ch/tools/software/polco.html>.
- [3] C. CHASSAGNOLE, N. NOISOMMIT-RIZZI, J. W. SCHMID, K. MAUCH, AND M. REUSS, *Dynamic modeling of the central carbon metabolism of escherichia coli*, Biotechnology and bioengineering, 79 (2002), pp. 53–73.

- [4] L. DAVID AND A. BOCKMAYR, *Computing elementary flux modes involving a set of target reactions*, IEEE/ACM Transactions on Computational Biology and Bioinformatics, 11 (2014), pp. 1099–1107.
- [5] L. F. DE FIGUEIREDO, A. PODHORSKI, A. RUBIO, C. KALETA, J. E. BEASLEY, S. SCHUSTER, AND F. J. PLANES, *Computing the shortest elementary flux modes in genome-scale metabolic networks*, Bioinformatics, 25 (2009), pp. 3158–3165.
- [6] K. FUKUDA AND A. PRODON, *Double description method revisited*, in Combinatorics and computer science, Springer, 1996, pp. 91–111.
- [7] J. GAGNEUR AND S. KLAMT, *Computation of elementary modes: a unifying framework and the new binary approach*, BMC Bioinformatics, 5 (2004), p. 1.
- [8] M. P. GERSTL, C. JUNGREUTHMAYER, S. MÜLLER, AND J. ZANGHELLINI, *Which sets of elementary flux modes form thermodynamically feasible flux distributions?*, FEBS Journal, 283 (2016), pp. 1782–1794.
- [9] J. GILBERT AND L. GILBERT, *Linear Algebra and Matrix Theory*, Academic Press, 2014.
- [10] C. M. GONZÁLEZ-DOMENECH, E. BELDA, R. PATIÑO-NAVARRETE, A. MOYA, J. PERETÓ, AND A. LATORRE, *Metabolic stasis in an ancient symbiosis: genome-scale metabolic networks from two blattabacterium cuenoti strains, primary endosymbionts of cockroaches*, BMC Microbiology, 12 (2012), p. 1.
- [11] D. JEVREMOVIC AND D. BOLEY, *Finding minimal generating set for metabolic network with reversible pathways*, BioSystems, 112 (2013), pp. 31–36.
- [12] C. JUNGREUTHMAYER, D. E. RUCKERBAUER, AND J. ZANGHELLINI, *regefntool: Speeding up elementary flux mode calculation using transcriptional regulatory rules in the form of three-state logic*, BioSystems, 113 (2013), pp. 37–39.
- [13] M. KANEHISA AND S. GOTO, *Kegg: kyoto encyclopedia of genes and genomes*, Nucleic acids research, 28 (2000), pp. 27–30.
- [14] Z. A. KING, J. LU, A. DRÄGER, P. MILLER, S. FEDEROWICZ, J. A. LERMAN, A. EBRAHIM, B. O. PALSSON, AND N. E. LEWIS, *Bigg models: A platform for integrating, standardizing and sharing genome-scale models*, Nucleic acids research, 44 (2016), pp. D515–D522.

- [15] S. KLAMT AND J. STELLING, *Combinatorial complexity of pathway analysis in metabolic networks*, Molecular biology reports, 29 (2002), pp. 233–236.
- [16] A. LARHLIMI AND A. BOCKMAYR, *On inner and outer descriptions of the steady-state flux cone of a metabolic network*, in International Conference on Computational Methods in Systems Biology, Springer, 2008, pp. 308–327.
- [17] A. LARHLIMI AND A. V. BOCKMAYR, *A new constraint-based description of the steady-state flux cone of metabolic networks*, Discrete Applied Mathematics, 157 (2009), pp. 2257–2266.
- [18] C. LI, M. DONIZELLI, N. RODRIGUEZ, H. DHARURI, L. ENDLER, V. CHELLIAH, L. LI, E. HE, A. HENRY, M. I. STEFAN, ET AL., *Biomodels database: An enhanced, curated and annotated resource for published quantitative kinetic models*, BMC Systems biology, 4 (2010), p. 92.
- [19] F. LLANERAS AND J. PICÓ, *Which metabolic pathways generate and characterize the flux space? a comparison among elementary modes, extreme pathways and minimal generators*, BioMed Research International, 2010 (2010).
- [20] D. MACHADO, Z. SOONS, K. R. PATIL, E. C. FERREIRA, AND I. ROCHA, *Random sampling of elementary flux modes in large-scale metabolic networks*, Bioinformatics, 28 (2012), pp. i515–i521.
- [21] J. A. PAPIN, J. STELLING, N. D. PRICE, S. KLAMT, S. SCHUSTER, AND B. O. PALSSON, *Comparison of network-based pathway analysis methods*, Trends in biotechnology, 22 (2004), pp. 400–405.
- [22] J. PEY AND F. J. PLANES, *Direct calculation of elementary flux modes satisfying several biological constraints in genome-scale metabolic networks*, Bioinformatics, (2014), p. btu193.
- [23] T. PFEIFFER, J. NU, F. MONTERO, S. SCHUSTER, ET AL., *Metatool: for studying metabolic networks.*, Bioinformatics, 15 (1999), pp. 251–257.
- [24] L.-E. QUEK AND L. K. NIELSEN, *A depth-first search algorithm to compute elementary flux modes by linear programming*, BMC Systems Biology, 8 (2014), p. 1.

- [25] O. RESENDIS-ANTONIO, J. L. REED, S. ENCARNACIÓN, J. COLLADO-VIDES, AND B. Ø. PALSSON, *Metabolic reconstruction and modeling of nitrogen fixation in rhizobium etli*, PLoS Comput Biol, 3 (2007), p. e192.
- [26] A. REZOLA, L. F. DE FIGUEIREDO, M. BROCK, J. PEY, A. PODHORSKI, C. WITTMANN, S. SCHUSTER, A. BOCKMAYR, AND F. J. PLANES, *Exploring metabolic pathways in genome-scale networks via generating flux modes*, Bioinformatics, 27 (2011), pp. 534–540.
- [27] A. RÖHL AND A. BOCKMAYR, *A mixed-integer linear programming approach to the reduction of genome-scale metabolic networks*, BMC Bioinformatics, 18 (2017), p. 2.
- [28] C. H. SCHILLING, M. W. COVERT, I. FAMILI, G. M. CHURCH, J. S. EDWARDS, AND B. O. PALSSON, *Genome-scale metabolic model of helicobacter pylori 26695*, Journal of bacteriology, 184 (2002), pp. 4582–4593.
- [29] C. H. SCHILLING, D. LETSCHER, AND B. Ø. PALSSON, *Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective*, Journal of theoretical biology, 203 (2000), pp. 229–248.
- [30] A. SCHRIJVER, *Theory of linear and integer programming*, Wiley, Chichester; New-York, 1998.
- [31] S. SCHUSTER, D. A. FELL, AND T. DANDEKAR, *A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks*, Nature Biotechnology, 18 (2000), pp. 326–332.
- [32] S. SCHUSTER AND C. HILGETAG, *On elementary flux modes in biochemical reaction systems at steady state*, Journal of Biological Systems, 2 (1994), pp. 165–182.
- [33] E. ULLAH, S. AERON, AND S. HASSOUN, *gef_m: an algorithm for computing elementary flux modes using graph traversal*, IEEE/ACM Transactions on Computational Biology and Bioinformatics, 13 (2016), pp. 122–134.
- [34] R. URBANCZIK AND C. WAGNER, *An improved algorithm for stoichiometric network analysis: theory and applications*, Bioinformatics, 21 (2005), pp. 1203–1210.

- [35] M. VON STOSCH, C. R. DE AZEVEDO, M. LUIS, S. F. DE AZEVEDO, AND R. OLIVEIRA, *A principal components method constrained by elementary flux modes: analysis of flux data sets*, BMC Bioinformatics, 17 (2016), p. 1.
- [36] C. WRZODEK, F. BÜCHEL, M. RUFF, A. DRÄGER, AND A. ZELL, *Precise generation of systems biology models from KEGG pathways*, BMC Systems Biology, 7 (2013), p. 15.
- [37] J. ZANGHELLINI, D. E. RUCKERBAUER, M. HANSCHO, AND C. JUNGREUTHMAYER, *Elementary flux modes in a nutshell: properties, calculation and applications*, Biotechnology journal, 8 (2013), pp. 1009–1016.



Spiking Neural Networks as Timed Automata

Giovanni Ciatto^{1,2}, Elisabetta De Maria² and Cinzia Di Giusto²

¹ Università di Bologna, Italy

² Université Côte d'Azur, CNRS, I3S, France

Abstract

In this paper we show how Spiking Neural Networks can be formalised using Timed Automata Networks. Neurons are modelled as timed automata waiting for inputs on a number of different channels (synapses), for a given amount of time (the accumulation period). When this period is over, the current *potential* value is computed taking into account the current sum of weighted inputs, and the previous decayed potential value. If the current potential overcomes a given *threshold*, the automaton emits a broadcast signal over its output channel, otherwise it restarts another accumulation period. After each emission, the automaton is constrained to remain inactive for a fixed *refractory period* after which the potential is reset. Spiking Neural Networks can be formalised as sets of automata, one for each neuron, running in parallel and sharing channels according to the structure of the network. The inputs needed to feed networks are defined through timed automata as well: we provide a language (and its encoding into timed automata) to model patterns of spikes and pauses and a way of generating unpredictable sequences.

1 Introduction

The brain behaviour has been the object of intensive studies in the last decades: on one side, researchers are interested in the inner functioning of neurons — which are its elementary components — their interactions and how such aspects participate to the ability to move, learn or remember, typical of living beings; on the other side, they emulate nature trying to reproduce such capabilities, e.g., within robot controllers, speech/text/face recognition applications, etc.

In order to achieve a detailed understanding of the brain functioning, both neurons behaviour and their interactions must be studied. Historically, interconnected neurons, “Neural Networks”, have been naturally modelled as directed weighted graphs where vertices are computational units receiving inputs by a number of ingoing edges, called *synapses*, elaborating and possibly propagating them over outgoing edges. Several inner models of the neuron behaviour have been proposed: some of them make neurons behave as binary threshold gates, other ones exploit a sigmoidal transfer function, while, in many cases, differential equations are employed.

According to [15, 17], three different and progressive *generations* of neural networks can be recognised: (i) *first generation* models handle discrete inputs and outputs and their computational units are threshold-based transfer functions; they include McCulloch and Pitt's threshold gate model [16], the perceptron model [8], Hopfield networks [11], and Boltzmann machines [2]; (ii) *second generation* models exploit real valued activation functions, e.g., the sigmoid function, accepting and producing real values: a well known example is the multi-layer perceptron [5, 19]; (iii) *third generation* networks are known as Spiking Neural Networks. They extend second generation models treating time-dependent and real valued signals often composed by *spike trains*. Neurons may fire output spikes according to threshold-based rules which take into account input spikes magnitude and occurrence time [17].

The core of our analysis are Spiking Neural Networks. Because of the introduction of timing aspects (in particular, observe that information is represented not only by spikes magnitudes but also by their frequency), they are considered closer to the actual brain functioning than other generations models. Spiking Neural Networks are weighted directed graphs where edges represent synapses, weights serve as synaptic strengths, and vertices correspond to Spiking Neurons. The latter ones are computational units that may emit (or *fire*) output impulses (*spikes*) taking into account input impulses strength and their occurrence instants. Models of this sort are of great interest not only because they are closer to natural neural networks behaviour, but also because the temporal dimension allows to represent information according to various *coding schemes* [17, 18]: e.g., the amount of spikes occurred within a given time window (*rate coding*), the reception/absence of spikes over different synapses (*binary coding*), the relative order of spikes occurrences (*rate rank coding*) or the precise time difference between any two successive spikes (*timing code*). Several spiking neuron models have been proposed in literature, having different complexities and capabilities. In [13] spiking neuron models are classified according to some *behaviours* (i.e., typical responses to an input pattern) that they should exhibit in order to be considered biologically relevant. The Leaky Integrate & Fire (LI&F) model [14], where past inputs relevance exponentially decays with time, is one of the most studied neuron models because it is basic and easy to use [13, 17]. On the other hand, the Hodgkin-Huxley (H-H) model [10] is one of the most complete and important within the scope of computational neuroscience, being composed by four differential equations comparing neurons to electrical circuits. As one may expect, the more complex the model, the more behaviours it can reproduce, at the price of a greater computational cost for simulation and formal analysis; e.g., in [13], the H-H model can reproduce all behaviours, but the simulation process is really expensive even for just a few neurons being simulated for a small amount of time.

Our aim is to produce a neuron model being meaningful from a biological point of view but also suitable to formal analysis and verification. We intend to exploit model checking algorithms to automatically prove whether our system verifies or not some desired properties. More precisely, this technique could be used to detect non-active portions within some network (i.e., the subset of neurons not contributing to the network outcome), to test whether a particular output sequence can be produced or not, to prove that a network may never be able to emit, to assess if a change to the network structure can alter its behaviour, or to investigate (new) learning algorithms which take time into account.

In this work, we take inspiration from the LI&F model introduced in [6], which relies on the synchronous approach based on the notion of logical time: time is considered as a sequence of logical discrete instants, and an instant is a point in time where external input events can be observed, computations can be done, and outputs can be emitted. The variant we introduce here takes into account some new time-related aspects, such a lapse of time in which the neuron is not active, i.e., it cannot receive and emit. We encode LI&F networks into Timed Automata: we show how to define the behavior of a single neuron and how to build a network of neurons.

Timed Automata [3] are Finite State Automata extended with timed behaviours: constraints are allowed limiting the amount of time an automaton can remain within a particular state, or the time interval during which a particular transition may be enabled. Timed Automata Networks are sets of automata that can synchronise over *channels* communications.

Our modelling of Spiking Neural Networks consists of Timed Automata Networks where each neuron is an automaton alternating between two states: it accumulates the weighted sum of inputs, provided by a number of ingoing weighted synapses, for a given amount of time, and then, if the *potential* accumulated during the last and previous accumulation periods overcomes a given threshold, the neuron fires an output over the outgoing synapse. Synapses are channels shared between the timed automata representing neurons, while *spike* emissions are represented by *synchronisations* occurring over such channels. Timed Automata can be exploited to produce or *recognise* precisely defined spike sequences, too.

The rest of the paper is organised as follows: in Section 2 we describe our reference model, the Leaky Integrate & Fire one, in Section 3 we recall definitions of Timed Automata Networks, and in Section 4 we show how Spiking Neural Networks can be encoded into Timed Automata Networks and how inputs and outputs are handled by automata. Finally, Section 5 summarises our approach and presents some future research directions.

2 Leaky Integrate and Fire Model

Spiking Neural Networks [15] are modelled as directed weighted graphs where vertices are computational units and edges represent *synapses*. The signals propagating over synapses are *trains of impulses: spikes*. Synapses may modulate such signals according to their weight or they could introduce some propagation delay. Synapses are classified according to their weight: *excitatory* if positive, or *inhibitory* if negative.

Computational units represent *neurons*, whose dynamics is governed by two parameters: the *membrane potential* (or, simply, *potential*) and the *threshold*. The former one depends on spikes received by neurons over ingoing synapses. Both current and past spikes are taken into account even if old spikes contribution is lower. In particular, the *leak factor* is a measure of neuron memory about past spikes. The neuron outcome is controlled by the algebraic difference between the membrane potential and the threshold: it is enabled to fire (i.e., emit an output impulse over *all* outgoing synapses) only if such difference is non-negative. Immediately after each emission the neuron membrane potential is reset and the neuron stays in a *refractory period* for a given amount of time. During this period it has no dynamics: it cannot increase its potential as any received spike is lost and therefore it cannot emit any spike.

We focus on the *Leaky Integrate & Fire* model that from an observational point of view is biophysically meaningful [13, 17] but is abstracted enough to be able to apply formal verification techniques such as model-checking. The original definition of Leaky Integrate & Fire traces back to [14]; here we work on an extended version of the discretized formulation proposed in [6].

Definition 1 (Spiking Integrate and Fire Neural Network). *A Spiking Integrate and Fire Neural Network is a tuple (V, A, w) , where:*

- V are Spiking Integrate and Fire Neurons,
- $A \subseteq V \times V$ are the synapses,
- $w : A \rightarrow \mathbb{Q} \cap [-1, 1]$ is the synapse weight function associating to each synapse (u, v) a weight $w_{u,v}$.

We distinguish three disjoint sets V_i of input neurons, V_{int} of intermediary neurons, and V_o of output neurons, with $V = V_i \cup V_{int} \cup V_o$.

A Spiking Integrate and Fire Neuron is a tuple $(\theta_v, p_v, \tau_v, y_v)$, where:

- $\theta_v \in \mathbb{N}$ is the firing threshold,
- $p_v : \mathbb{N} \rightarrow \mathbb{Q}_0^+$ is the [membrane] potential function defined as

$$p_v(t) = \sum_{i=1}^m w_i \cdot x_i(t) + \lambda \cdot p_v(t-1),$$

with $p_v(0) = 0$ and where $x_i(t) \in \{0, 1\}$ is the signal received at the time t by the neuron through its i^{th} input synapses, and $\lambda \in [0, 1]$ is a rational number, representing the leak factor,

- $\tau_v \in \mathbb{N}^+$ is the refractory period,
- $y_v : \mathbb{N} \rightarrow \{0, 1\}$ is the neuron output function, defined as

$$y_v(t) = \begin{cases} 1 & \text{if } p_v(t) \geq \theta_v \\ 0 & \text{otherwise} \end{cases}$$

As shown in the previous definition, the set of neurons of a Spiking Integrate and Fire Neural Network can be classified into input, intermediary, and output ones. The dynamics of each neuron v is defined by means of the set of its firing times $F_v = \{t_1, t_2, \dots\} \subset \mathbb{N}$, also called *spike train*. For each input neuron, the set F_v is assumed to be given as input for the network. For each output neuron, the set F_v is considered an output for the network.

3 Timed Automata

Timed Automata [3] are a powerful theoretical formalism for modelling and verifying real time systems. A timed automaton is an annotated directed (and connected) graph, with an initial node and provided with a finite set of non-negative real variables called *clocks*. Nodes (called *locations*) are annotated with *invariants* (predicates allowing to enter or stay in a location), arcs with *guards*, *communication labels*, and possibly with some variables upgrades and clock *resets*. Guards are conjunctions of elementary predicates of the form $x \text{ op } c$, where $\text{op} \in \{>, \geq, =, <, \leq\}$ where x is a clock and c a (possibly parameterised) positive integer constant. As usual, the empty conjunction is interpreted as true. The set of all guards and invariant predicates will be denoted by G .

Definition 2. A timed automaton TA is a tuple $(L, l^0, X, \Sigma, Arcs, Inv)$, where

- L is a set of locations with $l^0 \in L$ the initial one
- X is the set of clocks,
- Σ is a set of communication labels,
- $Arcs \subseteq L \times (G \cup \Sigma \cup U) \times L$ is a set of arcs between locations with a guard in G , a communication label in $\Sigma \cup \{\epsilon\}$, and a set of variables upgrades (e.g., clock resets);
- $Inv : L \rightarrow G$ assigns invariants to locations.

It is possible to define a synchronised product of a set of timed automata that work and synchronise in parallel. The automata are required to have disjoint sets of locations, but may share clocks and communication labels which are used for synchronisation. We restrict communications to be *broadcast* through labels $b!, b? \in \Sigma$ meaning that a set of automata can synchronise if one is emitting; notice that, a process can always emit (e.g., $b!$) and the receivers ($b?$) must synchronise if they can.

Locations can be normal, urgent or committed. Urgent locations force the time to freeze, committed once not only freeze time but the automaton must leave the location as soon as possible, i.e., they have higher priority.

The synchronous product $TA_1 \parallel \dots \parallel TA_n$ of timed automata, where for each $j \in [1, \dots, n]$, $TA_j = (L_j, l_j^0, X_j, \Sigma_j, Arcs_j, Inv_j)$ and all L_j are pairwise disjoint sets of locations is the timed automaton

$$TA = (L, l^0, X, \Sigma, Arcs, Inv)$$

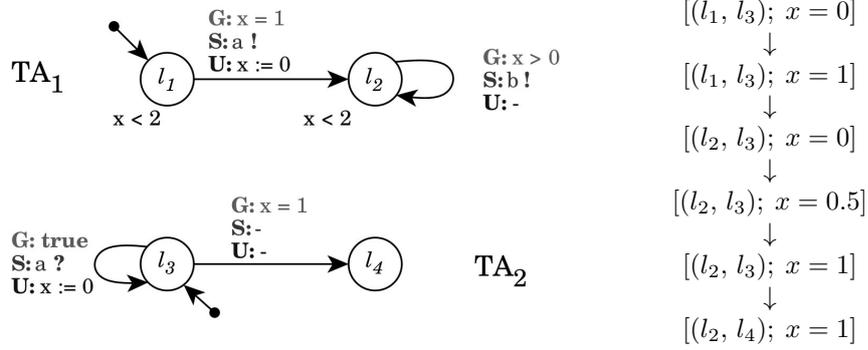
such that:

- $L = L_1 \times \dots \times L_n$ and $l^0 = (l_1^0, \dots, l_n^0)$, $X = \bigcup_{j=1}^n X_j$, $\Sigma = \bigcup_{j=1}^n \Sigma_j$,
- $\forall l = (l_1, \dots, l_n) \in L: Inv(l) = \bigwedge_j Inv_j(l_j)$,
- $Arcs$ is the set of arcs $(l_1, \dots, l_n) \xrightarrow{g, a, r} (l'_1, \dots, l'_n)$ such that for all $1 \leq j \leq n$ then $l'_j = l_j$.

The semantics of a synchronous product $TA_1 \parallel \dots \parallel TA_n$ is that of the underlying timed automaton TA with the following notations. A location is a vector $l = (l_1, \dots, l_n)$. We write $l[l'_j/l_j, j \in S]$ to denote the location l in which the j^{th} element l_j is replaced by l'_j , for all j in some set S . A valuation is a function ν from the set of clocks to the non-negative reals. Let \mathbb{V} be the set of all clock valuations, and $\nu_0(x) = 0$ for all $x \in X$. We shall denote by $\nu \models F$ the fact that the valuation ν satisfies (makes true) the formula F . If r is a clock reset, we shall denote by $\nu[r]$ the valuation obtained after applying clock reset $r \subseteq X$ to ν ; and if $d \in \mathbb{R}_{>0}$ is a delay, $\nu + d$ is the valuation such that, for any clock $x \in X$, $(\nu + d)(x) = \nu(x) + d$.

The semantics of a synchronous product $TA_1 \parallel \dots \parallel TA_n$ is defined as a timed transition system (S, s_0, \rightarrow) , where $S = (L_1 \times \dots \times L_n) \times \mathbb{V}$ is the set of states, $s_0 = (l^0, \nu_0)$ is the initial state, and $\rightarrow \subseteq S \times S$ is the transition relation defined by:

- (silent): $(l, \nu) \rightarrow (l', \nu')$ if there exists $l_i \xrightarrow{g, a, r} l'_i$, for some i , such that $l' = l[l'_i/l_i]$, $\nu \models g$ and $\nu' = \nu[r]$,



(a) The timed automata network $TA_1 \parallel TA_2$ is the synchronous product of the TA_1 and TA_2 automata.

(b) A possible run.

Figure 1: TA_1 and TA_2 start in the l_1 and l_3 locations, respectively, so the initial state is $[(l_1, l_3); x = 0]$. A *timed* transition produces a delay of 1 time unit, making the system move to state $[(l_1, l_3); x = 1]$. A *broadcast* transition is now enabled, making the system move to state $[(l_2, l_3); x = 0]$, broadcasting over channel a and resetting the x clock. Two successive *timed* transitions (0.5 time units) followed by a *broadcast* one will eventually lead the system to state $[(l_2, l_4); x = 1]$.

- (broadcast): $(\bar{l}, \nu) \rightarrow (\bar{l}', \nu')$ if there exist an output arc $l_j \xrightarrow{g_j, b!, r_j} l'_j \in Arcs_j$ and a (possibly empty) set of input arcs of the form $l_k \xrightarrow{g_k, b?, r_k} l'_k \in Arcs_k$ such that for all $k \in K = \{k_1, \dots, k_m\} \subseteq \{l_1, \dots, l_n\} \setminus \{l_j\}$, the size of K is maximal, $\nu \models \bigwedge_{k \in K \cup \{j\}} g_k$, $l' = l[l'_k/l_k, k \in K \cup \{j\}]$ and $\nu' = \nu[r_k, k \in K \cup \{j\}]$;
- (timed): $(l, \nu) \rightarrow (l, \nu + d)$ if $\nu + d \models Inv(l)$.

The valuation function ν is extended to handle a set of shared bounded integer variables: predicates concerning such variables can be part of edges guards or locations invariants, moreover variables can be updated on edges firings but they cannot be assigned to or from clocks.

In Figure 1 we exemplify timed automata usage, we consider the network of timed automata TA_1 and TA_2 with broadcast communications, and we give a possible run.

Throughout our modelling, we have used the specification and analysis tool `Uppaal` [4] which provides the possibility of designing and simulating Timed Automata Networks on top of the ability of testing networks against temporal logic formulae.

4 Spiking Neural Networks modelling

We present here our modelling of the Spiking Integrate and Fire Neural Network via Timed Automata Networks. Let $S = (V_i \cup V_{int} \cup V_{out}, A, w)$ be such a network (as remarked in Section 2 we distinguish between input, intermediary and output neurons). The Timed Automata Network will be obtained as the parallel composition of the encoding of each kind of neuron. More formally:

$$\llbracket S \rrbracket = (\parallel_{N_i \in V_i} \llbracket N_i \rrbracket) \parallel (\parallel_{N_j \in V_{int}} \llbracket N_j \rrbracket) \parallel (\parallel_{N_o \in V_{out}} \llbracket N_o \rrbracket)$$

4.1 Input neurons

The behaviour of input neurons is part of the specification of the network. Here we define two kinds of input behaviours: regular and non-deterministic ones. For each family, we provide an encoding into Timed Automata.

Regular input sequences. Spike trains are “regular” sequences of spikes and pauses: spikes are instantaneous while pauses have a non-null duration. Sequences can be *empty*, *finite* or *infinite*. After each spike there must be a pause except when the spike is the last event of a finite sequence. Infinite sequences are composed by two parts: a finite and arbitrary prologue and an infinite and periodic part composed by a finite sequence of *spike–pause* pairs. More formally the input sequence IS is defined by the grammar production :

$$IS ::= \varepsilon \mid \Phi \mid \Phi \Omega^\omega$$

where Φ is a finite prefix

$$\Phi := P? (sP)^* s$$

and Ω is the part which is repeated infinitely often

$$((sP_1) \cdots (sP_n))^\omega$$

with s representing a spike and $P_i = p[N_i]$ is a pause of a duration N_i .

It is possible to generate an emitter automaton for any regular input sequence. Such an automaton requires a clock t to measure pause durations, a boolean variable s which is *true* every time the automaton is firing and a location for each spike or pause into the sequence. The encoding $\llbracket IS \rrbracket$ is as follows, where α ranges over sub-sequences.:

- $\llbracket \varepsilon \rrbracket = \textcircled{E}$ an *empty* sequence is encoded into an automaton having just one location E without any edge;

- $\llbracket \Phi \rrbracket = \bullet \rightarrow \boxed{\Phi} \longrightarrow \bigcirc{E}$ any *finite* sequence is encoded into a sequence of locations, as described below, where the last one has no outgoing edges and represent the end of the sequence;

- $\llbracket \Phi \Omega^\omega \rrbracket = \bullet \rightarrow \boxed{\Phi} \longrightarrow \boxed{\Omega} \begin{array}{c} \curvearrowright \\ \curvearrowleft \end{array} \bigcirc{R}$ any *infinite* sequence is composed by a finite sub-sequence Φ followed by a finite sub-sequence Ω repeated an infinite amount of times. The two sub-sequences are encoded according to the rules explained below and the resulting automata are connected. Finally, an urgent location \mathbf{R} is added, having an input edge from Ω last location and an output edge to Ω first location.

Any finite sub-sequence is a list of spikes and pauses. They are recursively encoded as follows:

- $\llbracket p[N] \alpha \rrbracket = \bigcirc{P} \xrightarrow[\mathbf{U}: t := 0, s := \text{true}]{\mathbf{G}: t = N} \boxed{\alpha}$ any pause having duration N

and followed by a sub-sequence α is encoded into a location \mathbf{P} with the invariant $t \leq T$ having one outgoing edge connected to the automaton $\llbracket \alpha \rrbracket$; such an edge is enabled if and only if $t = T$ and, if triggered, t is reset and, since pauses are always followed by spikes, s is set to *true*;

- $\llbracket s \alpha \rrbracket = \bigcirc{S} \xrightarrow[\mathbf{U}: s := \text{false}]{\mathbf{S}: y!} \boxed{\alpha}$ any spike followed by a sub-sequence

α is translated to an *urgent* location \mathbf{S} having one output edge connected to the automaton translated from α ; such an edge emits on y if triggered and resets s .

Non-deterministic input sequences. These kinds of input sequences are useful when no assumption is available on neuron inputs. These are random sequences of spikes separated by at least T_{min} time units. Their encoding is shown in Figure 2 and the automaton behaves as follows: it waits in location \mathbf{B} an arbitrary amount of time before moving to location \mathbf{S} , firing its first spike over channel x . Since location \mathbf{S} is *urgent*, the automaton instantaneously moves to location \mathbf{W} , resetting clock t . Finally, from location \mathbf{W} , after an arbitrary amount of time $t \in]T_{min}, \infty [$, it moves to location \mathbf{S} , firing a spike. Notice that an initial delay D may be introduced by adding invariant $t \leq D$ to location \mathbf{B} and guard $t = D$ on edge $(\mathbf{B} \rightarrow \mathbf{S})$.

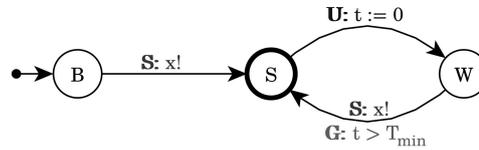


Figure 2: Non-deterministic input sequence automaton

4.2 Intermediary and Output Neurons

The neuron is designed as a synchronous and stateful machine that: i) accumulates potential whenever it receives input spikes within a given *accumulation period*, ii) if the accumulated potential is greater than the *threshold*, the neuron emits an output spike, iii) it waits during a *refractory period*, and restarts from i). We assume that no two input spikes on the same synapse can be received within the same accumulation period (i.e., the accumulation period is shorter than the minimum refractory period of the input neurons of the network). Next, we give the encoding of a neuron into Timed Automata. Notice that this encoding applies to intermediary and output neurons only.

Definition 3. Given an intermediary neuron $N = (\theta, p, \tau, y)$ with m input synapses, its encoding into Timed Automata is $\llbracket N \rrbracket = (L, A, X, \Sigma, Arcs, Inv)$ with:

- $L = \{A, W, D\}$ with D committed,
- $X = \{t\}$
- $\Sigma = \{x_i \mid i \in [1..m]\} \cup \{y\}$,
- $Arcs = \{(A, t \leq T, x_i?, \{a := a + w_i\}, A) \mid i \in [1..m]\} \cup \{(A, t = T, , \{p := a + \lfloor \lambda p \rfloor\}, D), (D, p < \theta, , \{a := 0\}, A), (D, p \geq \theta, y!, , W), (W, t = \tau, , \{a := 0, t := 0, p := 0\}, A)\}$;
- $Inv(A) = t \leq T, Inv(W) = t \leq \tau, Inv(D) = \mathbf{true}$.

The neuron behaviour, described by the Automaton in Figure 3, depends on the following channels, variables and clocks:

- x_i for $i \in [1..m]$ are the m input channels,
- y is the broadcast channel used to emit the output spike,
- $p \in \mathbb{N}$ is the current potential value, initially set to 0,

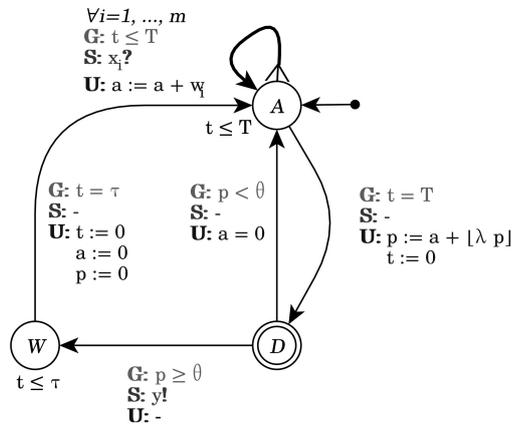


Figure 3: Neuron model.

- $a \in \mathbb{N}$ is the weighted sum of input spikes occurred within the *current* accumulation period; it is 0 at the beginning of each round.

The automaton has three locations: **A**, **D** and **W**. It can move from one location to another according to the following intuitive semantics:

- the neuron keeps waiting in state **A** (for Accumulation) for input spikes while $t \leq T$ and whenever it receives a spike on input x_i , it updates a with: $a := a + w_i$
- when $t = T$, the neuron moves to state **D** (for Decision), resetting t and updating p according to the potential function given in Definition 1 :

$$p := a + \lfloor \lambda \cdot p \rfloor$$

since state **D** is *committed*, it does not allow time to progress, so, from this state, the neuron can move back to state **A** resetting a if the potential has not reached the threshold $p < \theta$, or it can move to state **W**, firing an output spike, otherwise;

- the neuron remains in state **W** (for Wait) for τ time units and then it moves back to state **A** resetting a , p and t .

4.3 Output consumers

In order to have a complete modelling of a Spiking Neural Network, for each output neuron we build an *output consumer* automaton O_y . The automaton,

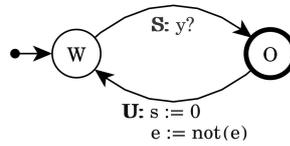


Figure 4: Output consumer automaton.

shown in Figure 4, waits in location **W** for the corresponding output spikes on channel y and as soon as it receives the spike, it moves to location **O**. This location is only needed to simplify model checking queries. Since it is urgent, the automaton instantly moves back to location **W** resetting s the clock measuring the elapsed time since last emission and setting e to its negation, with e being a *boolean variable* which differentiates each emission from its successor.

Thus the encoding of an output neuron N_o is the parallel composition of the encoding of the N_o as if it was an intermediary neuron plus an output consumer on its broadcast channel y :

$$\llbracket N_o \rrbracket = \llbracket N_o \rrbracket \parallel O_y$$

5 Conclusion and Future Works

In this paper we formalised the LI&F model of Spiking Neural Networks via Timed Automata Networks. LI&F neurons are modelled as automata waiting for inputs on a number of different channels, for a fixed amount of time. When such an *accumulation period* is over, the current *potential* value is computed taking into account the current sum of weighted inputs, and the previous decayed potential value. If the current potential overcomes a given *threshold*, the automaton emits a broadcast signal over its output channel, otherwise it restarts its accumulation period. After each emission, the automaton is constrained to remain inactive for a fixed *refractory period* (after which the potential is reset). Spiking Neural Networks composed by more than one neuron can be formalised by a set of automata one for each neuron, running in parallel and sharing channels accordingly. The inputs needed to feed network are defined through Timed Automata as well. We have provided a language and its encoding into Timed Automata to model patterns of spikes and pauses and a way of modelling unpredictable sequences.

We have a complete implementation of the Spiking Neural Network model proposed in the paper via the tool `Uppaal`. It can be found at page [1].

As for future work, we intend to validate our neuron model proving some

characteristic properties expressed in temporal logics via model-checking. Furthermore, we consider this work as the starting point for a number of research directions: we plan to study whether our model cannot reproduce behaviours requiring *bursts* emission capability, as stated in [13] (e.g., tonic or phasic bursting), or some notion of *memory* (e.g., phasic spiking, or bistability). Furthermore, it may be interesting to enrich our formalisations to include modelling of propagation delays or even more complex spiking neuron models like the *theta-neuron* model [7] or Izhikevich's one [12]. Finally it may be interesting to combine learning algorithms with formal-analysis: we would like to exploit reachability properties verification to control weights variations within the scope of existing learning algorithms or strategies, e.g., Hebb's rule [9].

References

- [1] Additional material. https://github.com/gciatto/snn_as_ta.
- [2] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski. Connectionist models and their implications: Readings from cognitive science. chapter A Learning Algorithm for Boltzmann Machines, pages 285–307. Ablex Publishing Corp., Norwood, NJ, USA, 1988.
- [3] R. Alur and D. L. Dill. A theory of timed automata. *Theor. Comput. Sci.*, 126(2):183–235, Apr. 1994.
- [4] J. Bengtsson, K. G. Larsen, F. Larsson, P. Pettersson, and W. Yi. UPPAAL — a Tool Suite for Automatic Verification of Real-Time Systems. In *Proc. of Workshop on Verification and Control of Hybrid Systems III*, number 1066 in Lecture Notes in Computer Science, pages 232–243. Springer-Verlag, Oct. 1995.
- [5] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4):303–314, 1989.
- [6] E. De Maria, A. Muzy, D. Gaffé, A. Ressouche, and F. Grammont. Verification of Temporal Properties of Neuronal Archetypes Using Synchronous Models. In *Fifth International Workshop on Hybrid Systems Biology*, Grenoble, France, Oct. 2016.
- [7] G. B. Ermentrout and N. Kopell. Parabolic bursting in an excitable system coupled with a slow oscillation. *SIAM Journal on Applied Mathematics*, 46(2):233–253, 1986.

- [8] Y. Freund and R. E. Schapire. Large margin classification using the perceptron algorithm. *Machine Learning*, 37(3):277–296, 1999.
- [9] D. O. Hebb. *The Organization of Behavior*. John Wiley, 1949.
- [10] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117(4):500–544, 1952.
- [11] J. J. Hopfield. Neurocomputing: Foundations of research. chapter Neural Networks and Physical Systems with Emergent Collective Computational Abilities, pages 457–464. MIT Press, Cambridge, MA, USA, 1988.
- [12] E. M. Izhikevich. Simple model of spiking neurons. *Trans. Neur. Netw.*, 14(6):1569–1572, Nov. 2003.
- [13] E. M. Izhikevich. Which model to use for cortical spiking neurons? *IEEE Transactions on Neural Networks*, 15(5):1063–1070, Sept 2004.
- [14] L. Lapicque. Recherches quantitatives sur l’excitation électrique des nerfs traitée comme une polarisation. *J Physiol Pathol Gen*, 9:620–635, 1907.
- [15] W. Maass. Networks of spiking neurons: The third generation of neural network models. *Neural Networks*, 10(9):1659 – 1671, 1997.
- [16] W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.
- [17] H. Paugam-Moisy and S. Bohte. *Computing with Spiking Neuron Networks*, pages 335–376. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.
- [18] M. Recce. Pulsed neural networks. chapter Encoding Information in Neuronal Activity, pages 111–131. MIT Press, Cambridge, MA, USA, 1999.
- [19] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Neurocomputing: Foundations of research. chapter Learning Representations by Back-propagating Errors, pages 696–699. MIT Press, Cambridge, MA, USA, 1988.

HSM, a Reduced Model of Central Carbon Metabolism. A Dynamical Approach.

Bertrand Beauvoit¹, Sophie Colombié¹, Razanne Issa², Jean-Pierre Mazat²,
Christine Nazaret³ and Sabine Pérès⁴ (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 & Université 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

We have shown last year [1] the interest of reduced models of metabolism, still representing the main architecture and stoichiometry of the whole metabolism, but with fewer steps which are aggregations of the actual reactions. The reduced model HSM (Human-Scale Model) we have developed (Fig. 1) involves 58 reactions (of which 25 are reversible) and 63 metabolites (of which 35 are internal metabolites) (see appendix 1 for the abbreviations and appendix 2 for the list of metabolites and reactions). The advantage of such models is to be more easily tractable and more understandable. Furthermore, the great interest of reduced models is to permit different theoretical approaches of a given metabolic network.

With this kind of model it was possible to combine the calculation of Elementary Flux Modes (EFM) [2, 3] and FBA analysis [4] to study the production of serine in cancer cells on different substrates [1]. In this short paper we continue the exploration of HSM models in introducing rate equations of the reactions and the metabolites concentrations in order to develop a dynamical study of such a metabolic network.

In this work we were faced with several problems: what is the rate equation of aggregation of successive enzymatic reactions? What is the concentration of the remainder metabolites?

We will show how we deal with these problems with the example of the reduced glycolysis part of our model, which contains 5 reactions instead of 10 in the actual glycolysis.

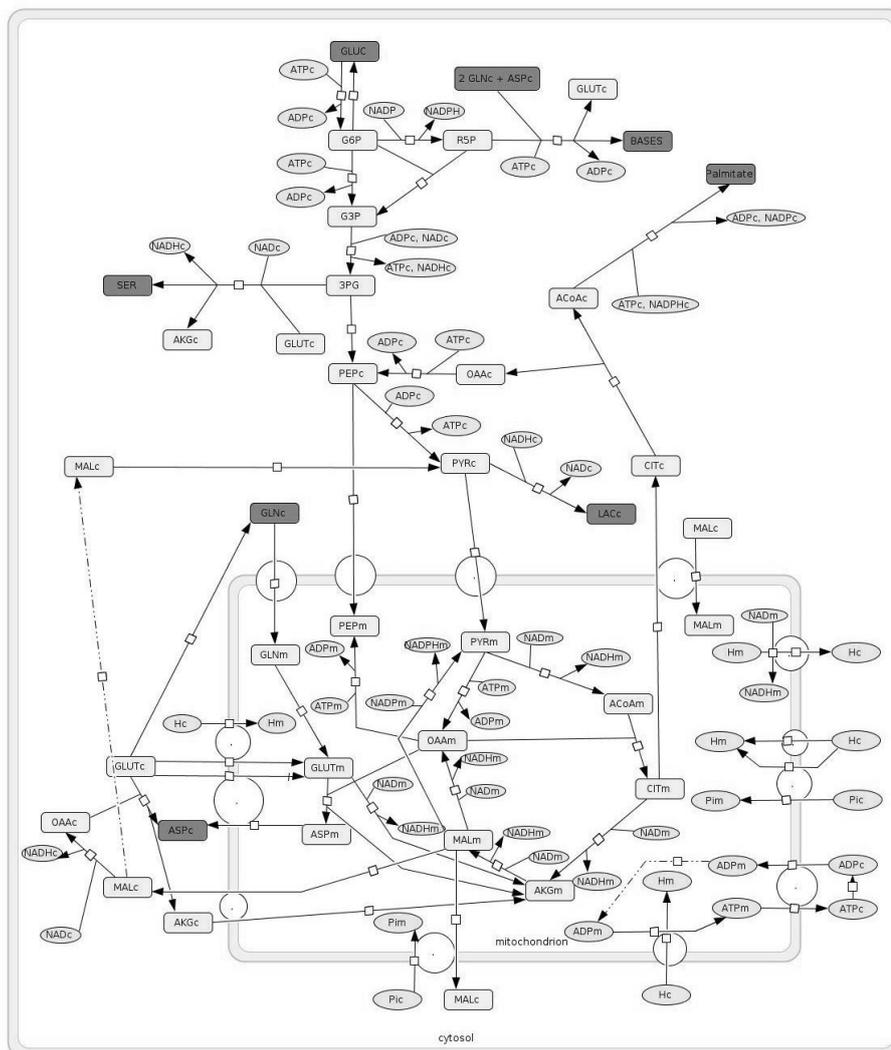


Figure 1: HSM13-3 metabolic network. The abbreviations are given in Appendix 1 and the reactions detailed in Appendix 2.

1 Model of reduced glycolysis

Our reduced model of glycolysis is represented in figure 2. It is essential that the results we obtain when solving the dynamical system representing the reduced network at steady state give fluxes and metabolites concentrations in the range of the physiological values.

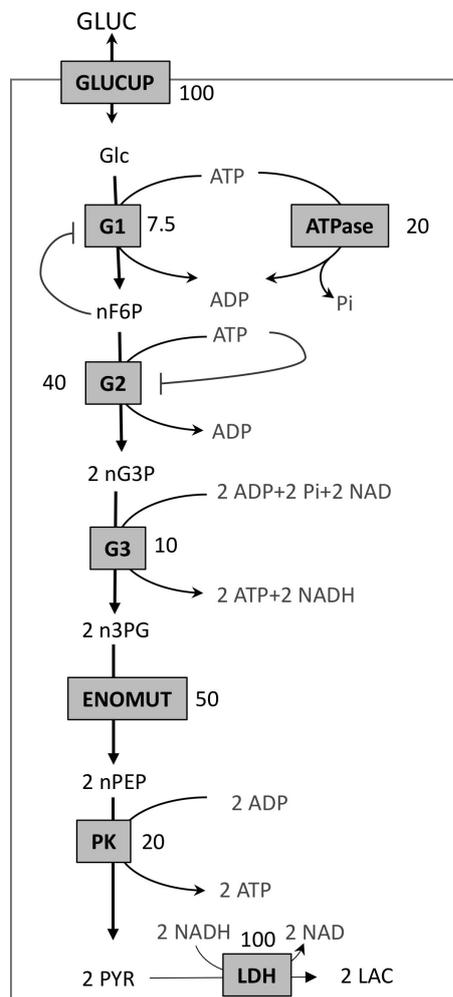


Figure 2: Reduced glycolysis. Abbreviations: GLUCUP: Glucose uptake. G1: hexokinase + phosphoglucose isomerase. G2: phosphofructokinase + aldolase + triose-phosphate isomerase. G3: Glyceraldehyde-3P Dehydrogenase + phosphoglycerate kinase. ENOMUT: Enolase + Phosphoglycerate Mutase. PK: Pyruvate Kinase. LDH: Lactate Dehydrogenase. The abbreviations of metabolites are given in table 2. The numbers in front of the brown squares represented the V_{max} of the (aggregated) steps.

2 The problem of concentrations

In order to have an idea of the concentrations of the different metabolites of glycolysis, we did a survey of the literature which led us to table 1.

In order to preserve the mass amount circulating inside the metabolic network the concentrations of the metabolites at the nodes are taken as the sum of the metabolites concentrations of the species involved in the aggregated reactions. The abbreviation of these new virtual metabolites are the abbreviations of the last metabolite species preceded by “n”. Thus (see the list of abbreviations in Table 1):

$$nF6P = G6P + F6P$$

$$nG3P = F-1,6-P2 + DHAP + G3P$$

$$n3PG = 1,3DPG + 3PG$$

$$nPEP = 2PG + PEP$$

The experimental values of these sums of concentrations are given in table 2.

Metabolite	Abbreviation	Measured Concentrations
Glucose	Glc	5 mM
Glucose-6-Phosphate	G6P	150 μ M
Fructose-6-Phosphate	F6P	40 μ M
Fructose-1,6-bisphosphate	FBP	300 μ M
Glyceraldehyde-3-Phosphate	G3P	50 μ M
Dihydroxyacetone Phosphate	DHAP	350 μ M
1,3-Biphosphoglycerate	13BPG	1 μ M
3-Phosphoglycerate	3PG	125 μ M
2-Phosphoglycerate	2PG	60 μ M
Phosphoenolpyruvate	PEP	25 μ M
Pyruvate	PYR	1 mM
Lactate	LAC	10 mM
Phosphate	Pi	2.5 mM
ATP	ATP	3 mM
ADP	ADP	0.3 mM
NAD	NAD	0.5 mM
NADH	NADH	0.07 mM

Table 1: Mean values of metabolites concentrations in glycolysis.

These values are orders of magnitude of the concentrations extracted from the literature ([8] and [9] for instance).

3 The rate equations

When a reversible step is associated with an irreversible one (case of G1 and G2), the rate equation is essentially the one of the irreversible step. For the association of reversible steps we take mass action law as rate equation with the

New Metabolite	Sum of	Exp. Conc. (mM)	Model at steady-state
Glc	Glc	5	4.94
nF6P	G6P + F6P	0.2	0.5
nG3P	F-1,6-P2+DHAP +G3P	0.7	0.9
n3PG	1,3DPG + 3PG	0.1 – 0.2	0.15
nPEP	2PG + PEP	0.1	0.13
PYR	PYR	1	0.96
LAC	LAC	1	1 fixed
ATP	ATP	3	4.45
ADP	ADP	0.3	0.55
NAD	NAD	0.5	1 fixed
NADH	NADH	0.07	0.05 fixed

Table 2: Experimental and calculated values of the concentrations of the sums of metabolites used in the simplified version of glycolysis depicted in Fig. 2. The column “Exp. Conc. (mM)” is deduced from the values listed in table 1 applying the conventions “nF6P = G6P + F6P etc.” and the column “Model at steady-state” gives the corresponding values derived from the model at steady-state.

product of equilibrium constant as new equilibrium constant of the aggregated steps. Several rate equations are taken from [5] and [6].

Glucose entry: $v_{GLUCUP} = k_{GLUCUP} \cdot GLUC - k_{MGLUCUP} \cdot Glc$.

In the model, the high values of the kinetic constants lead to $Glc = GLUC$ at steady-state.

Step G1: $v_{G1} = NG1/DG1$ with
 $NG1 = VMG1 \cdot Glc / KGlc \cdot ATP / KATP$ and
 $DG1 = (1 + Glc / KGlc + (nF6P^2 / Ki)) \cdot (1 + ATP / KATP)$

This equation involves the inhibition term $(nF6P^2 / Ki)$ [5] essential to insure that the consumption of ATP is not too high at the beginning of glycolysis; this may lead to the consumption of all ATP before its regeneration at the end of glycolysis leading to an arrest of glycolysis when $ATP = 0$. This is the danger of “Turbo effect” well explained in [5].

Step G2: $v_{G2} = VMG2 \cdot nF6P / KnF6P / (1 + nF6P / KnF6P) \cdot ATP / KG2ATP / (1 + ATP / KG2ATP + (ATP / KiATP)^4)$

This equation involves the inhibition term $(ATP / KiATP)^4$ [5], which represents partly the allosteric properties of the phosphofructokinase (PFK).

Step G3:

$$v_{G3} = VMG3 * (n_{G3P} / KG3G3P * NAD / KG3NAD * (5 - ATP) / KG3ADP - 1 / KQG3 * n_{3PG} / KG3n3PG * NADH / KG3NADH * ATP / KG3ATP)$$

Step ENOMUT:

$$v_{ENOMUT} = VMENO * (n_{3PG} / KENOn3PG - 1 / KQENO * PEP / KENOPEP)$$

Steps G3 and ENOMUT are reversible steps represented by mass action law with the composition of equilibrium constants.

Step PK: This step is represented by a generalized Henri-Michaelis-Menten equation $v_{PK} = NPK / DPK$ with:

$$NPK = VMPK * PEP / KPKPEP * (ADP) / KPKADP$$

$$DPK = (1 + PEP / KPKPEP) * (1 + (ADP) / KPKADP)$$

Step LDH: This step is reversible,

$$v_{LDH} = VMLDH * (PYR / KLDHPYR * NADH / KLDHNADH - 1 / KQLDH * LAC / KLDHLAC * NAD / KLDHNAD)$$

Step ATPase: A consumption of ATP is added to simulate the ATP usage of the cell. It permits to reach a steady-state. This step is supposed to be irreversible, modeled by a classical Henri-Michaelis-Menten equation.

$$v_{ATPase} = VMATPase * ATP / (KATP2 + ATP)$$

The parameters values of these equations which fit the experimental values of table 2 are given in the following:

NAD = 1	NADH = 0.05	kGLUCUP=100
kMGLUCUP=100	GLUC = 5	VMG1 = 7.5
KATP = 0.15	KGlc = 1	Ki = 4.422
KnF6P = 0.03	KG2ATP = 0.06	KiATP = 1
VMG2 = 40	VMG3 = 10	KG3G3P = 0.15
KG3NAD=1	KG3NADH =1	KG3ADP =1
KQG3 = 0.157	KG3n3PG = 0.1	KG3ATP = 1
VMPK =25	KPKPEP = 0.1	vMENO = 50
KQENO = 1.056	KENOn3PG =0.1	KENOPEP = 0.1
KPKADP = 0.1	VMLDH = 100	KLDHPYR = 0.4
KLDHNADH = 1	KQLDH = 11000	KLDHLAC = 1
KLDHNAD = 1	LAC = 1	VMATPase = 20
KATP2 = 3		

With these values a steady-state is reached and the concentrations values obtained at steady-state are listed in table 2. One can see that the calculated values are not too different from the average experimental values.

4 Conclusion

This model of glycolysis shows how rate equations can be applied on the reduced metabolic network HSM of central carbon metabolism. We have already developed a simple model of TCA cycle [7] which will be added to this model of glycolysis. The proton motive force (PMF) which is the thermodynamically link between the respiratory chain and the mitochondrial ATP synthesis will be modeled as a special substrate concentration.

More generally, the different steps of HSM reduced network will be modeled with a Henri-Michaelis-Menten equation if they contain an irreversible steps or with a mass action law if all they reactions are reversible.

The interest of a dynamical modeling of a metabolic network is to introduce the concentrations of the metabolites. Furthermore, the fluxes obtained in a dynamical simulations can be analyzed in term of elementary flux modes (EFM) and the yield compared to the flux balance analysis (FBA) of the same network. Finally the sensitivity of the network can be studied, particularly, the Flux Control Coefficients can be calculated.

References

- [1] Bertrand Beauvoit, Sophie Colombié, Jean-Pierre Mazat, Christine Nazaret, Sabine Pérès (2016): *Human Scale Metabolic Network of Central Carbon Metabolism. Application to serine metabolism from glutamine in Cancer Cells*. Proceedings of the Evry Spring School on advances in Systems and Synthetic Biology. P. Amar, F. Képès and V. Norris eds.
- [2] Schuster, S., & Hilgetag, C. (1994). *On elementary flux modes in biochemical reaction systems at steady-state*. Journal of Biological Systems, 2, 165-182.
- [3] Schuster, S., Fell, D. A., & Dandekar, T. (2000). *A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks*. Nature Biotechnology, 18, 326-332
- [4] Jeffrey D. Orth, Ines Thiele & Bernhard Ø Palsson *What is flux balance analysis?* Nature Biotechnology (2010) 28: 245-248.
- [5] Bas Teusink, Michael C. Walsh, Karel van Dam and Hans Westerhoff (1998). *The danger of metabolic pathways with turbo design*. TIBS 23:162- 169.

- [6] Johan H. van Heerden, Meike T. Wortel, Frank J. Bruggeman, Joseph J. Heijnen, Yves J. M. Bollen, Robert Planqué, Josephus Hulshof, Tom G. O'Toole, S. Aljoscha Wahl, Bas Teusink (2014). *Lost in transition: Startup of glycolysis yields subpopulations of nongrowing cells*. *Science* 343, 1245114 (2014). DOI: 10.1126/science.1245114
- [7] Christine Nazaret, Margit Heiske, Kevin Thurley, Jean-Pierre Mazat: *Mitochondrial energetic metabolism : A simplified model of TCA cycle with ATP production*. *J Theor Biol.* 2009 Jun 7;258(3):455-64.
- [8] J. O. Park *et al.*, *Metabolite concentrations, fluxes and free energies imply efficient enzyme usage*. *Nat. Chem. Biol.*, vol. 12, no. 7, pp. 482–489, Jul. 2016.
- [9] W. Chen, E. Freinkman, T. Wang, K. Birsoy, and D. M. Sabatini *Absolute Quantification of Matrix Metabolites Reveals the Dynamics of Mitochondrial Metabolism*. *Cell*, vol. 166, no. 5, p. 1324–1337.

Appendix 1: Abbreviations of reactions

ANT: ADP/ATP exchanger.
ASPUP: Aspartate uptake.
ASYNT: ATP Synthase.
ATPASE: ATP usage.
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.
GOT1: cytosolic Glutamate Oxaloacetate Transaminases.
GOT2: mitochondrial Glutamate Oxaloacetate Transaminases.
GS1: Glutamine synthase .
IDH1: cytosolic Aconitase + Isocitrate dehydrogenase (NADP).
IDH2: mitochondrial Aconitase + Isocitrate dehydrogenase (NADP).
IDH3: mitochondrial Aconitase + Isocitrate dehydrogenase (NAD).

K567: 2-oxoglutarate dehydrogenase + succinate thiokinase + succinate dehydrogenase + fumarase.

L: Leak of the membrane to protons.

LACIO : Input/Output of lactate.

LDH: Lactate Dehydrogenase.

MDH1: cytosolic Malate Dehydrogenase.

MDH2: mitochondrial Malate Dehydrogenase.

ME1: cytosolic Malic enzyme.

ME2: mitochondrial Malic enzyme.

NN: transhydrogenase.

PDH: Pyruvate Dehydrogenase.

PEPCK1: cytosolic PhosphoEnolPyruvate Carboxy Kinase.

PK: Pyruvate Kinase.

PL1: Synthesis of PhosphoLipids.

PL2: beta oxidation of fatty acids.

PL3: Formation of PalmitylCoA.

PP1: Oxidative part of PPP.

PP2: non-oxidative part of PPP.

PUR: Nucleotide Synthesis.

PYC: Pyruvate Carboxylase.

RC1: Complex I of Respiratory Chain.

RC2: Complex (III+IV) of Respiratory Chain.

SEROUT: Output of serine.

SERSYNT: Serine Synthesis = Dehydrogenase + Transaminase and Phosphatase.

T1(CIC): Citrate/Malate exchanger.

T2(OGC): Glutamate/Aspartate exchanger.

T3(DIC): Dicarboxylate carrier.

T4 (GLAST): Malate/2-oxoglutarate exchanger.

T5(PIT): Pi carrier.

T6: Pyruvate transporter in mitochondria .

T7: Phosphoenolpyruvate exchanger cytosol/mitochondria.

T8: Glutamine exchanger cytosol/mitochondria.

T9: Glutamate/H⁺ transporter in mitochondria.

T10: AcetoAcetate /H⁺ transporter in mitochondria .

T11: Palmitoyl-CoA transporter in mitochondria.

T12: cytosolic Pi transport .

T13: Cytosolic Palmitate transport.

Appendix 2: METATOOL entry file of HSM X

-ENZREV

ASYNT ANT ENOMUT G3 GLUD1 GOT1 GOT2 IDH1 IDH2 LACIO LDH
MDH1 MDH2 ME1 ME2 PP2 T1 T2 T3 T5 T7 T8 T9 T12 T13

-ENZIRREV

ASPUP ATPASE CL CS G1 G2 GG3 GG4 GLNUP GLS1 GLUCUP GS1
IDH3 K567 L NN PDH PEPCK1 PK PL1 PL2 PL3 PP1 PUR PYC RC1 RC2
SEROUT SERSYNT T4 T6 T10 T11

-METINT

3PG AcetoAcetate_c AcetoAcetate_m ACoAc ACoAm AKGc AKGm ASPc
ASPm CITc CITm PMF G3P G6P GLNc GLNm GLUCc GLUTc GLUTm
LACc MALc MALm OAAc OAAm PalCoAc PalCoAm Palmitate_c PEPc
PEPm Pic Pim PYRc PYRm R5P SERc

-METEXT

ADPc ADPm ASP ATPc ATPm BASES CO2 CoAc CoAm GLN GLUC HCO3
LAC NADc NADHc NADm NADHm NADPc NADPm NADPHc NADPHm
NH3 O2 Palmitate Pi Q QH2 SER

-CAT

ANT : $ATPm + ADPc + 0.88 PMF = ATPc + ADPm$.

ASPUP : $ASP = ASPc$.

ASYNT : $3 ADPm + 3 Pim + 8 PMF = 3 ATPm$.

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 : $GLUTc + NH3 + ATPc = GLNc + ADPc + Pic$.

IDH1 : $CITc + NADPc = AKGc + NADPHc + CO2$.

IDH2 : $CITm + NADPm = AKGm + NADPHm + CO2$.

IDH3 : CIT_m + NAD_m = AKG_m + NADH_m + CO₂ .
 K567 : AKG_m + NAD_m + Pim + ADP_m + Q = MAL_m + NADH_m + CO₂ + ATP_m + QH₂ .
 L : PMF = .
 LACIO : LAC_c = LAC .
 LDH : PYR_c + NADH_c = LAC_c + NAD_c .
 MDH1 : MAL_c + NAD_c = OAA_c + NADH_c .
 MDH2 : MAL_m + NAD_m = OAA_m + NADH_m .
 ME1 : MAL_c + NADP_c = PYR_c + NADPH_c + CO₂ .
 ME2 : MAL_m + NAD_m = PYR_m + NADH_m + CO₂ .
 NN : NADH_m + NADP_m + PMF = NAD_m + NADPH_m .
 PDH : PYR_m + NAD_m = ACoA_m + NADH_m + CO₂ .
 PEPCK1 : OAA_c + ATP_c = PEP_c + ADP_c + CO₂ .
 PK : PEP_c + ADP_c = PYR_c + ATP_c .
 PL1 : 8 ACoA_c + 7 ATP_c + 14 NADPH_c + 7 HCO₃ = Palmitate_c + 7 ADP_c + 7 Pic + 14 NADP_c + 8 CoA_c + 7 CO₂ .
 PL2 : PalCoA_m + 7 NAD_m + 7 CoA_m + 7 Q = 7 NADH_m + 8 ACoA_m + 7 QH₂ .
 PL3 : Palmitate_c + CoA_c = PalCoA_c .
 PP1 : G6P + 2 NADP_c = R5P + 2 NADPH_c + CO₂ .
 PP2 : 3 R5P = 2 G6P + G3P .
 PUR : R5P + 2 GLN_c + ASP_c + 3 ATP_c = BASES + 2 GLUT_c + 3 ADP_c + 3 Pic .
 PYC : PYR_m + HCO₃ + ATP_m = OAA_m + Pim + ADP_m .
 RC1 : NADH_m + Q = NAD_m + QH₂ + 4 PMF .
 RC2 : 6 QH₂ + 3 O₂ = 6 Q + 36 PMF .
 SEROUT : SER_c = SER .
 SERSYNT : 3PG + GLUT_c + NAD_c = SER_c + AKG_c + NADH_c + Pic .
 T1 : CIT_m + MAL_c = CIT_c + MAL_m .
 T2 : AKG_c + MAL_m = AKG_m + MAL_c .
 T3 : MAL_m + Pic = MAL_c + Pim .
 T4 : GLUT_c + ASP_m + PMF = GLUT_m + ASP_c .
 T5 : Pic + 0.12 PMF = Pim .
 T6 : PYR_c + 0.12 PMF = PYR_m .
 T7 : PEP_c = PEP_m .
 T8 : GLN_c = GLN_m .
 T9 : GLUT_c + 0.12 PMF = GLUT_m .
 T10 : AcetoAcetate_c + 0.12 PMF = AcetoAcetate_m .
 T11 : PalCoA_c = PalCoA_m .
 T12 : Pi = Pic .
 T13 : Palmitate = Palmitate_c .



An Ockham Razor model of energy metabolism

Rajeev Khoodeeram^{1,2}, Gilles Bernot¹ and Jean-Yves Trosset³

¹I3S, CNRS - Univ. of Nice-Sophia Antipolis, F-06903 Sophia Antipolis, France

²Université Des Mascareignes, Mauritius

³Bio-Information Research Laboratory, Sup'Biotech, 94800 Villejuif, France

*This article is dedicated to René Thomas who unfortunately left
this world last January*

Abstract

We present an abstract model of energy metabolism that aims at understanding how activity level of biological functions and combination of nutrients influence metabolic shifts. One of the most frequently observed transition is between respiration and fermentation which is induced by high intake of glucose even in the presence of oxygen (Crabtree and Warburg effects). This glycolytic phenotype is observed in many micro-organisms including parasites and is also shared by all cancer cell lines which makes the Warburg glycolytic phenotype one of the most efficient target in oncology. Nutrients influence production yield of high added value compounds and the study of metabolic shifts is also of concern in bioproduction and fermentation processes. In order to help understanding how major metabolic actors influence these transitions, we developed an abstract and qualitative model of energy metabolism. To facilitate the interpretation of our results with respect to biological knowledge we restrict our variables to key metabolic or cellular components such as pathways, cellular functions, nutrients and important cofactors that play the role of regulators in this cellular system. Primary results on global dynamic phenotypes such as metabolic oscillations and Warburg/Crabtree effects are presented in this chapter. Our simulations have been done using a new software called DyMBioNet.

1 Introduction

Highly proliferative cells such as micro-organisms play a major role in biotechnology as their high turnover provides interesting yields for the industrial bio-

synthesis of high added value molecules such as food complement or bio-fuel. To adapt between cellular maintenance or cell growth (i.e. production of biomass) or between primary and secondary metabolism cells modify their metabolism with respect to environmental conditions. Nutrients, presence/absence of oxygen, carbon/nitrogen ratio are external regulators of cellular economy. They may induce metabolic shift such as the short term Crabtree effect that shows immediate shift from respiration to fermentation upon addition of excess sugar or the long term Crabtree effect that arise under steady-state conditions at high growth rates [1]. This effect which persists even in the presence of oxygen has also been observed in cancer cells by Otto Warburg in the early 20th century. The Warburg phenotype evolved towards a seemingly irreversible status due to the accumulation of mutations whereas the Crabtree effect is clearly reversible [2]. In the rest of this chapter, we do not make further distinctions between short and long term Crabtree effect or between reversible Crabtree and irreversible Warburg effect as this is not the scope of this chapter.

The Warburg glycolytic phenotype occurs in all tumor cell lines and therefore appears as a common anti-cancer target [3, 4, 5]. Other therapeutic areas such as infectious diseases (e.g. parasitic diseases), anti-aging or obesity also strongly depend on metabolism energetics. Controlling cellular fate is crucial not only to develop therapeutic strategy against infectious diseases or cancer but is also central to optimize yield in industrial bio-processes.

To study how metabolism can control such global cellular phenotypes, we developed an abstract model of energy metabolism. We made use of the Ockham's razor principle which asserts that between two equivalent models, the simpler, the better.

The next five sections describe: i) our qualitative modelling approach of energy metabolism, ii) a global view of energy metabolism and especially, the trade-off between efficient versus inefficient metabolism (respiration vs fermentation), iii) the Thomas modelling framework, iv) our qualitative model and the associated kinetic parameters, and finally v) four dynamics of energy metabolism under respiration, fermentation and Crabtree/Warburg initial conditions using DyMBioNet software.

2 Energy trade off in Cell Metabolism

Metabolism can be summarized as an oscillation between catabolism and anabolism. Catabolism degrades nutrients to extract electrons to be used during anabolism (synthesis of biological macro-molecules) or store in cofactors as a reservoir of energy. More specifically, electrons are stored in: i) cytoplasmic NAD(P)H which plays the role of electron container to synthesize

biomolecules or redox potential to release fermentation products, ii) mitochondrial NADH to create proton gradients and ATP through ATP-synthase in oxydative phosphorylation chain thanks to electron acceptor role of oxygen, iii) the third main reservoir concerns electron-pair bonds with primary metabolites (nucleotides, aromatic amino acids), plants alkaloids or flavor compounds issued from fermentation process. Most of secondary metabolites of industrial interest (i.e. food, biofuel) are produced during fermentation. The production yield of these metabolites are therefore sensitive to the metabolic modes of the cell especially to respiration and fermentation.

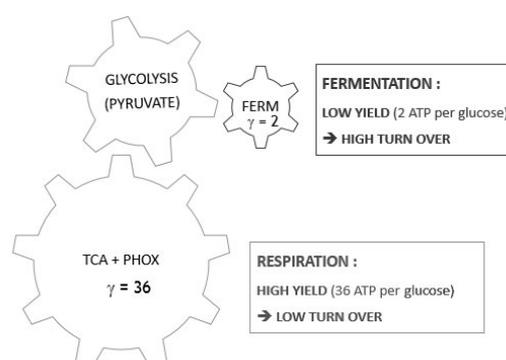


Figure 1: Energy metabolism: trade-off between fermentation and respiration. The Glycolysis pathway (upper wheel) is connected through pyruvate to two other pathways: fermentation with a fast turn over and Krebs cycle which has a high efficiency but lower turn over.

The shift from a highly efficient metabolism (respiration) to an inefficient metabolism (fermentation) at high glucose intake is one of the most important effects characterizing the Crabtree or Warburg effect mentioned above. The energy yield is decreasing from 36 to 2 molecules of ATP per molecule of glucose but the turnover of glucose (number of glucose molecule degraded per unit of time) is higher (Figure 1). These metabolic modes impact cell growth rate [6] and bio-production of secondary metabolites [7].

3 A coarse-grained qualitative approach of energy metabolism

In order to study the impact of external conditions (nutrient, drugs) on the global phenotypes of energy metabolism, especially the Crabtree or Warburg effect, we focus on the biological actors that are directly related to these phenotypes. The variables of our model correspond therefore to coarse-grained

metabolic descriptors such as metabolic pathways, biological functions, key cofactors or cellular nutrients (Figure 2).

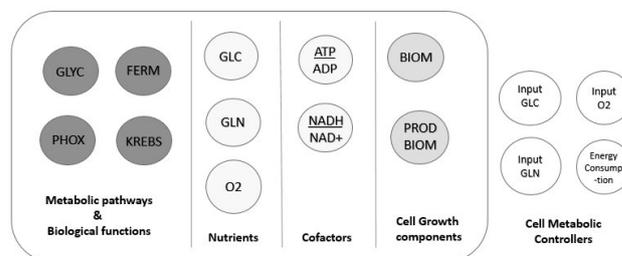


Figure 2: Four main classes of actors for the energy metabolism: biological functions or pathways (left), nutrients and cofactor (middle), cell growth components (right) and controls of cellular inputs (white).

The advantage of this qualitative modelling is to get closer connection with experimental facts as global effect of cellular functions or pathways are usually well known. The parametrization of the model depends only on biological knowledge and not on molecular interactions for which kinetics constants are more difficult to estimate *in vivo*. The model includes four classes of cellular variables and one class of external cursors controlling nutrients and consumption of energy. The four internal cell variables are: i) biological functions or metabolic pathways such as Glycolysis, Fermentation, Krebs cycle and Oxidative Phosphorylation (red circles in figure 2), ii) nutrients, which concern carbon source symbolized by glucose (GLC), nitrogen source symbolized by glutamine (GLN) and oxygen (O₂), iii) energy and redox cellular level symbolized by the ratio of cofactors ATP/ADP and NADH/NAD⁺ and finally iv) the biomass-related components that correspond to anabolism.

The ATP/ADP ratio captures the energetic balance of the cell. It encapsulates all the other related molecules such as GTP, AMP and inorganic phosphate (Pi). Other important nutrients necessary for cell growth such as vitamins are supposed to be present implicitly. This abstraction level is not as high for the NADH/NAD⁺ ratio which corresponds explicitly to the redox molecular components directly involved in the Glycolysis pathway and Krebs cycle. The NADPH/NADP⁺ cofactors are the redox energy cofactors for anabolism which are not explicitly involved in the Crabtree or Warburg transitions. For that reason there were only implicitly included into the "Biomass Production" variables (blue cycle in Figure 2).

4 The Thomas Modelling Framework

René Thomas has developed in the 70's a discrete modelling approach for gene networks [8, 9] where a gene never consumes its activators or inhibitors contrarily to metabolism where a product consumes its substrate. In our qualitative model of energy metabolism, variables corresponding to biological functions or metabolic pathways do not consume their resources. For example, Glycolysis “activates” the Krebs cycle but it is nevertheless not “consumed” by Krebs. We consequently decided to use the Thomas formalism instead of a formalism dedicated to metabolic reactions such as BIOCHAM [10, 11]. Exceptionally, for cofactors and nutrients, there are consumptions which are then modelled by negative retro-actions (see full model in Figure 8).

4.1 Interaction network with multiplexes

The static representation of a biological network can be drawn using a directed graph, in which directed arrows (activation/inhibition) represent the action of one variable on its target variable. An example is given for the interaction between Glycolysis and Krebs in Figure 3.

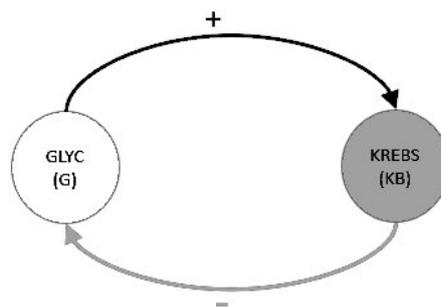


Figure 3: An oversimplified metabolic relation between Glycolysis and Krebs (only for pedagogical purposes). The end product of Glycolysis is a nutrient for Krebs (black positive arrow). Moreover a high level of activity of Krebs can produce citrate, which may inhibit Glycolysis (grey negative arrow). Note that Glycolysis is not *consumed* by Krebs.

Moreover, in formal modelling, actors are not always independent variables. It is often necessary to group together several concomitant molecular conditions into a single statement for an activation or an inhibition to be present. This is the role of a *multiplex*, that encodes the basic important

logical conditions for the action to be present [12]. If the multiplex is FALSE, then its action (activation or inhibition) is ignored (see Figure 4). More than one multiplex can act on a given variable. Each multiplex represents only one resource for its target node, i.e. one well-defined but complex biological regulation event. The naive representation of Figure 3 can be replaced by 2 multiplexes (Figure 4) that encapsulate more detailed information than just + and - signs.

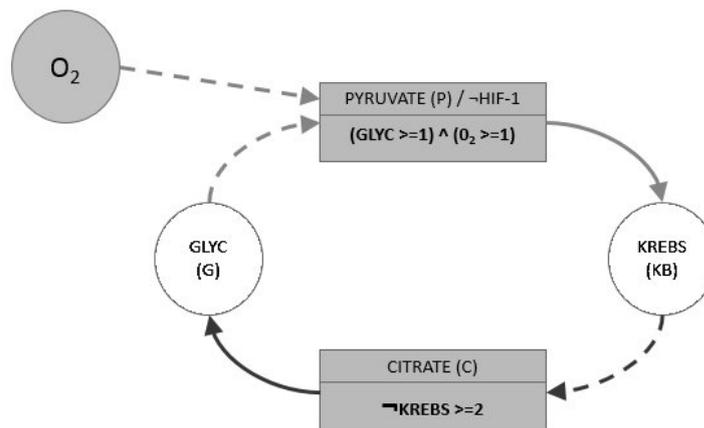


Figure 4: Simplified interactions between Glycolysis and Krebs with multiplexes (green boxes). Top multiplex is called PYRUVATE to mention the metabolite connecting the two pathways without making it an explicit variable of the model, note that oxygen is also needed to act on Krebs (\wedge stands for the conjunction). Similarly, CITRATE refers to the inhibitory effect of high level of citrate on phosphofruktokinase. More precisely, it reduces glycolysis by enhancing the inhibitory effect of ATP. So, a high level of Krebs may finally inhibit Glycolysis (\neg stands for the negation of being a resource for Glycolysis).

The top multiplex of Figure 4 informs that pyruvate, the end-product of GLYCOLYSIS will aliment KREBS. The information that pyruvate is a resource of KREBS is encoded into a multiplex called “Pyruvate” (denoted by P) which says symbolically that GLYCOLYSIS is functional (“ $GLYC \geq 1$ ”) and can activate KREBS (Figure 4). Note that there is no retroactive loop on this top part to say that KREBS consumes pyruvate and therefore GLYCOLYSIS. Note also that there is an additional condition in the Pyruvate multiplex which concerns oxygen. This is a rough illustration of our specific question, i.e. the shift between aerobic or anaerobic metabolism in cancer cells or microorganisms. In this particular case, the assertion that oxygen is present stipulates

(in cancer cell) the absence of Hypoxia-Induced Factor 1 (HIF1) which inhibits mitochondria in anoxic condition. The statement on the presence of oxygen in the “P” multiplex ensures therefore that HIF-1 is “off”.

The second multiplex at the bottom of Figure 4 mentions that citrate (produced by KREBS) can inhibit GLYCOLYSIS (through PhosphoFructoKinase) if accumulated in the cytoplasm. The minus inhibition sign is then logically encoded by a negative logical statement (the negation \neg) in the second multiplex concerning KREBS.

4.2 Activity levels and thresholds

To each variable is assigned a number of qualitative activity states. Certain variables like O₂ are boolean describing only the absence or presence of this nutrient. Other variables are multivalued in order to capture finer level of biological activities. These qualitative states are defined according to a given question of interest.

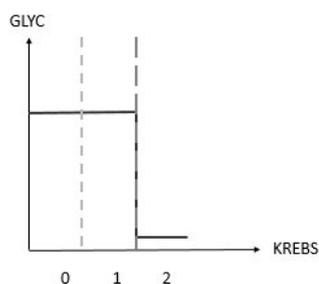


Figure 5: Three activity states for KREBS. The inhibition of GLYCOLYSIS by citrate multiplex is effective only if CITRATE is in excess, which occurs when KREBS is above the threshold of 2 (over-expressed)

For sake of simplicity, GLYCOLYSIS and KREBS variables as well as PYRUVATE and CITRATE multiplexes are represented by G, KB, P and C respectively. In addition, the assignment of a name to a multiplex is important as it denotes a particular molecular mechanism in the energy metabolism. The name is given to help biologists to define the kinetic parameters of the model (see Table 1). The variable KREBS is multivalued (0,1,2) in multiplex C. In this multiplex, Krebs is prefixed by a \neg symbol, which means that it inhibits glycolysis (represented also by a red line) only when it is over-expressed (0=low, 1=high, 2=over-expressed) as depicted in Figure 5.

4.3 Kinetic parameters

A particular combination of resources acting on a variable determines its fate if we leave the variable evolve without limit in time, the rest of the system being supposedly frozen. The long time activity level of the variable of interest corresponds to what is often referred to as kinetic parameters, which can be represented using $K_{\nu,\omega}$, where ν is the variable and ω represents the set of resources (identified by solid arrows in Figure 4).

A variable that possesses n multiplexes as potential resources, has 2^n different potential sets of resources and consequently 2^n kinetic parameters. In Figure 4, KREBS (respectively GLYC) has multiplex P (respectively multiplex C) as the sole input and therefore 2 possible sets of resources each. Table 1 lists randomly chosen values for the parameters K.

Kinetic Parameters	Values
K_{KB}	0
$K_{KB,P}$	1
K_G	0
$K_{G,C}$	2

Table 1: An arbitrary set of kinetic parameters. The first and third parameters indicate that in absence of resources, respectively the Krebs cycle (KB) and the glycolysis (G) are attracted toward a state (0) that is too low to allow them having an action on the rest of the system. If the multiplex PYRUVATE becomes a resource for Krebs (second parameter) then Krebs is attracted toward a “normal” activity level (1). If the multiplex CITRATE becomes a resource of glycolysis (that is an *absence* of citrate, fourth parameter) then, according to this toy pedagogical example, glycolysis is attracted to a its highest level (2).

These kinetic parameters are not “dynamical” parameters as they remain fixed during the simulations but they give all the information that is necessary to deduce the complete dynamics of the system.

Determining the values for the set of K parameters is based purely on the biological knowledge. Using formulas from the multiplex, we can quickly deduce the set of multiplexes which are resources. In Figure 4, multiplex P is a resource of Krebs only when glycolysis is equal or above 1. In a similar vein, multiplex C is a resource of G when Krebs is less than 2. This method gives us a total of 9 parameters, as shown in Table 2.

Current state		Resources ($K_{\nu,\omega}$)		Attracting State	
G	KB	K_G	K_{KB}		
0	0	C	-	2	0
0	1	C	-	2	0
0	2	-	-	0	0
1	0	C	P	2	1
1	1	C	P	2	1
1	2	-	P	0	1
2	0	C	P	2	1
2	1	C	P	2	1
2	2	-	P	0	1

Table 2: States, applicable parameters and their values. These are only toy values according to Table 1, chosen to obtain Figure 6.

4.4 Transition graphs

In a dynamical system, a transition from a current state (denoted by η) to a next state (denoted by η'), can be represented by $\eta \rightarrow \eta'$, as seen in Figure 6. As demonstrated in [15], the probability that all variables pass through their respective thresholds at the same time is negligible *in vivo*. This means that the system is asynchronous and therefore only one variable is likely to evolve (increase or decrease) over a unit of time while other variables remain unchanged. As demonstrated by Houssine Snoussi, this way of discretizing the state space is compatible with continuous approaches such as stepwise linear differential equations [13]. From Table 2, assuming normoxic condition in which case Oxygen is always present ($O_2=1$) in the cells, we would obtain the synchronous state graph of Figure 6a.

Focusing on state (2,2) in Figure 6b, we obviously see that imaginary continuous trajectories (green) would quit the state (2,2) by going either in state (1,2) or in state (2,1). According to [13], the René Thomas approach retains the transitions (blue) : (2,2) \rightarrow (1,2) and (2,2) \rightarrow (2,1). By doing so in each state, we get the transition graph of Figure 6c.

More formally,

- ν can change from state $\eta(\nu)$ to state $\eta'(\nu) = \eta(\nu) + 1$ only if $K_{\nu,\omega} > \eta$ and it is then required that $\eta'(x) = \eta(x)$ for all $x \neq \nu$.
- ν can change from state $\eta(\nu)$ to state $\eta'(\nu) = \eta(\nu) - 1$ only if $K_{\nu,\omega} < \eta$ and it is then required that $\eta'(x) = \eta(x)$ for all $x \neq \nu$.
- Accordingly, a state η is *stable* when for all variables x , $K_{x,\omega} = \eta(x)$ (i.e. there is no variable ν that gives rise to a possible transition).

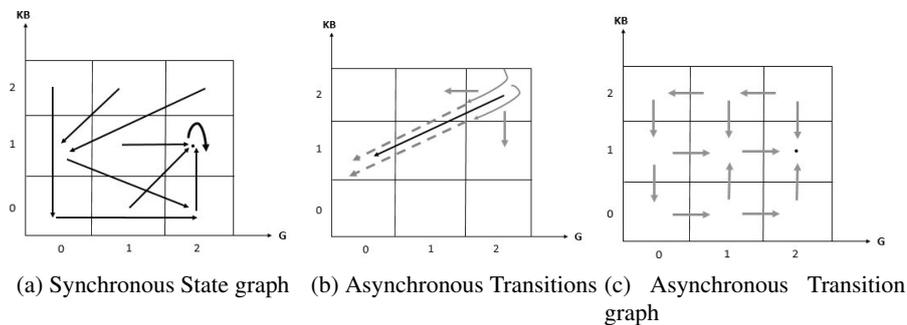


Figure 6: State transitions. The asynchronous dynamic (a) directly obtained from Table 2 would be biologically incorrect because it contains “jumps” from 0 to 2 that do not reflect a continuous increase of activity and because it contains simultaneous variable changes (diagonal arrows). Asynchronous variable changes are required for compatibility with models based on differential equations (b) where the probability to cross two frontiers at the same time is null. So, Table 2 results finally to the asynchronous transition graph (c).

5 Our qualitative model

5.1 Introduction

Based on the Thomas framework, we applied a method [14] in 5 steps as follows:

1. What are the actors? (variables of the system, see section 3)
2. Which variables or combination of variables act on a given variable? (interaction graph)
3. How many qualitative levels for each variable and in which order are the targets influenced by a given variable? (link between targets and qualitative levels that determines the elementary comparisons in the multiplexes)
4. Identify the Kinetic parameters
5. Validate the global behavior of the model

5.2 Threshold order

For each main actor of section 3, we have firstly looked at the actors on which it acts. For example, Krebs acts on Glycolysis, on the production of biomass

(PROD-BIOM), as well as GLN and NADH/NAD⁺. After a first round of discussions based on the literature, it appears that Krebs acts on Glyc and Prod. BIOM mainly via the excess production of Citrate and at the same level of Citrate. So, this gives rise to three targets for Krebs as in Figure 7 (for the moment ignoring the thresholds):

We then assume that for a reason or another, Krebs goes progressively from a state where it is completely off to a state where it is running at its maximum (over-expressed).

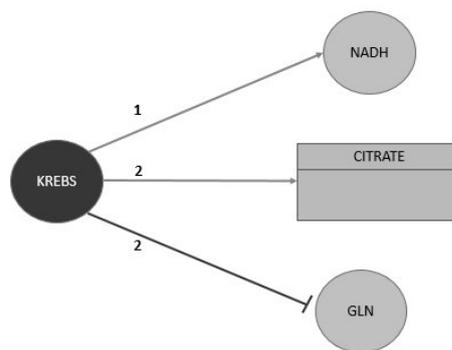


Figure 7: Threshold order for Krebs. Krebs acts on three actors including one multiplex. NADH is the first actor to be activated as Krebs's level rises up and, according to the purpose of our model, one sees no reason in the literature to distinguish the abstract levels of Krebs where CITRATE and GLN are produced by Krebs (else Krebs would get an additional level 3, which seems useless according to the questions under consideration).

As soon as Krebs is active, it produces NADH. So, this is the first actor activated by Krebs (threshold of 1). Krebs may increase its activity by consuming alphaKeto Glutarate (α KG), one end-product of glutaminolysis (cancer phenotype) which has the effect to produce citrate that exits mitochondria to be transformed into malate. This citrate-malate shuttle has also the effect of replenishing NAPDH for the biomass production. Both actions (consumption of alpha-ketoglutarate, end-product of glutaminolysis and over-production of citrate is a result of high level of krebs (level 2) without further distinction.

This reasoning allowed us to put the thresholds 1 and 2 in Figure 7. We did the same work for all the variables and finally we obtain the interaction graph of Figure 8.

Our abstract model of energy metabolism is made of 10 variables, 11 multiplexes (i.e. regulation mechanisms) and 4 metabolic controllers to setup the external nutrient conditions.

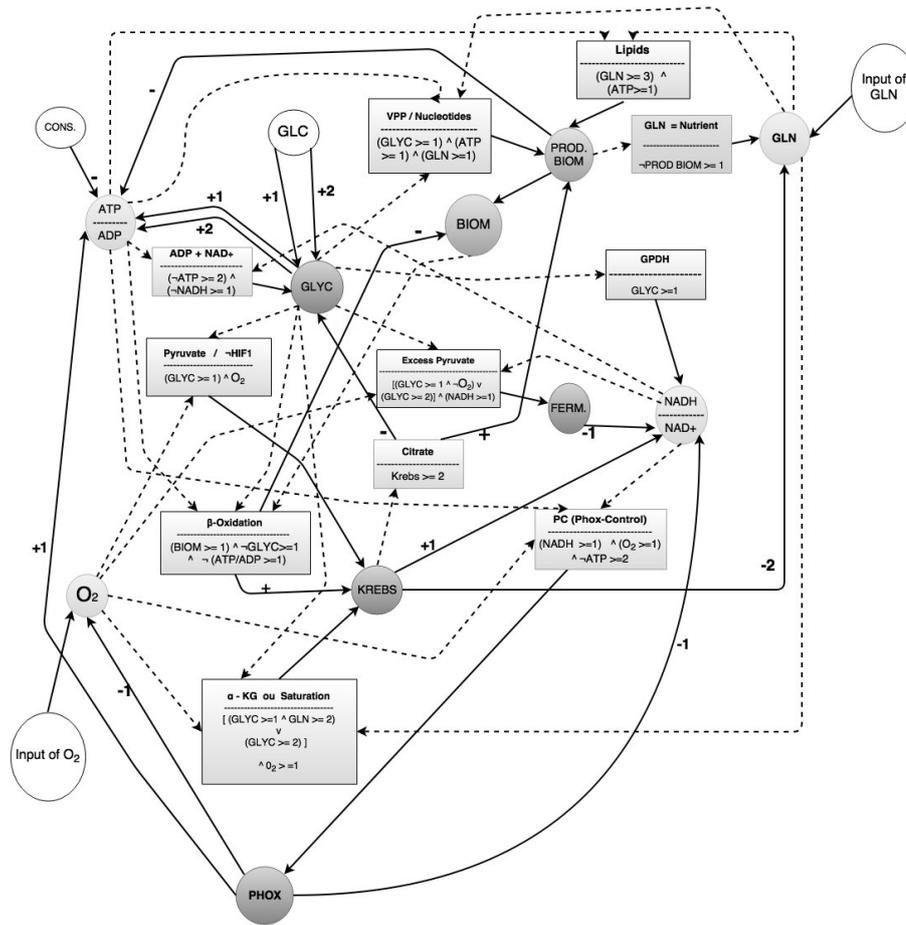


Figure 8: Model for energy metabolism resulting from answering three questions: i) what are the actors? ii) what interacts with what? and iii) which priority? The last component (kinetic parameters) is given in Table 4.

5.3 Identification of the K parameters

For each actor ν of Figure 8, we have considered the 2^n possible sets of resources and for each of them, we have determined the level towards which ν is attracted. To do so, we consider a virtual experiment where the considered resources of ω of ν are “frozen” in the system and we inventory which ones of the targets of ν would be directly affected by ν .

If the previous stage of the method (threshold order) has been properly done, there is a level l of ν such that all targets with a threshold $t \leq l$ are

affected, then $K_{\nu,\omega} = l$. For example, consider the 3 possible resources of Krebs(Kb) in Figure 8. There are 8 configurations to treat. For example, when $\omega = \{P\}$ (only Pyruvate), there is normal pyruvate (from glycolysis) and Krebs is attracted towards a level where it is strong enough to produce NADH, but not enough to inhibit GLN or produce excess of Citrate. So, $K_{Kb,P} = 1$. In table 3, a – (respectively x) means absence (respectively presence) of a resource. The column *values* corresponds to the threshold levels of the variable (here the variable Krebs can take only 0,1 and 2 as possible values). If a *C* appears as value, this means this condition cannot occur.

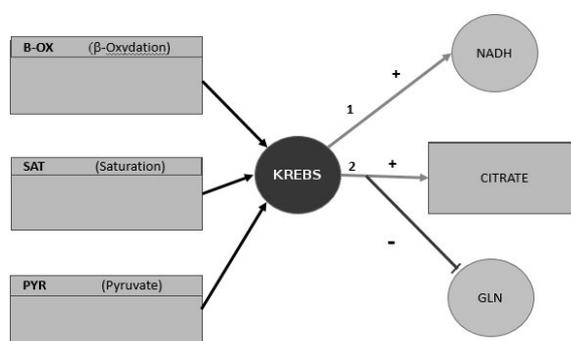


Figure 9: Identification method for the kinetic parameters of Krebs: we consider successively each of the 8 possible subsets of $\{\beta\text{-OX}, \text{SAT}, \text{PYR}\}$, assume that Krebs benefits from that subset (say ω) of resources for an infinite period of time, and finally deduce from the literature on which targets Krebs will finally act. Then $K_{KREBS,\omega}$ equals 0 if no target activated, 1 if only NADH is activated, and else 2, see Table 3.

We did the same work for the 7 other K parameters. Some sets of resources are inconsistent. For example Pyruvate and β -Oxidation are contradictory. Inconsistent sets of resources are identified by C in Table 3, where only 4 sets of resources are consistent. The same work is applied for all variables and we got the whole set of 100 parameters of the model, owing to the large panel of convergent literature (see Table 4).

6 Results

6.1 The DyMBioNet Software

We have developed an extension of SMBioNet [15], DyMBioNet (short for Dynamic Modelling of Biological Networks), which we use to simulate the

Resources for Krebs, $K_{KREBS,\omega}$			
<i>PYR</i>	β -OX	α -KG	Values
-	-	-	0
-	-	x	C
-	x	-	1
-	x	x	C
x	-	-	1
x	-	x	2
x	x	-	C
x	x	x	C

Table 3: Set of resources for Krebs. “-” and “x” respectively stand for the absence or the presence of a resource. A qualitative level for the KREBS variable is assigned (under the column “values”) for each of the $2^3 = 8$ possible combinations of resources for this central variable. Some combinations are contradictory (C).

dynamics of the energy metabolism. DyMBioNet also includes proof checking techniques with Temporal Logic (CTL) to confirm the existence of certain metabolic states under specific nutrient or drug conditions. It has a built-in user-friendly interface and a suitable chart for demonstrating how the system evolves over time. This formal logic framework will also help in the future to propose most pertinent experiments to validate or refute certain hypotheses concerning the Warburg effect.

In the following subsections, we illustrate four of the key behaviours that participate to validate our model.

6.1.1 Respiration under normal condition

The nutrients were set up for normal respiration conditions, i.e. presence of oxygen, normal level of glucose intake and low level of nitrogen source (GLN). Simulation shows that under these nutrient conditions, the system is maintained in respiration mode or return to respiration if it starts in fermentation mode (FERM = 1 at the initialisation of the simulation).

In this respiration condition, glutamine is not an important ingredient to fuel this machinery. The default value of 0 for the variable *Input_GLN* corresponds to a basic level of input nitrogen source. The GLN nutrient variable was used to model the excess of glutamine or the glytaminolysis phenotype which is often activated in cancer cells [16]. A representative simulation under normal respiration mode is shown in Figure 10.

Metabolic oscillations can be observed from all the simulations we performed.

Variable	K Parameters & values
ATP/ADP	$K_{ATP/ADP,\{cons.,phox\}} = 1,$ $K_{ATP/ADP,\{cons.,phox,glyc2\}} = 1,$ $K_{ATP/ADP,\{cons.,phox,pbm\}} = 2,$ $K_{ATP/ADP,\{cons.,phox,pbm,glyc2\}} = 2,$ $K_{ATP/ADP,\{glyc1,pbm,glyc2\}} = 1,$ $K_{ATP/ADP,\{glyc1,phox,pbm,glyc2\}} = 1,$ $K_{ATP/ADP,\{glyc1,cons.,glyc2\}} = 1,$ $K_{ATP/ADP,\{glyc1,cons.,pbm\}} = 2,$ $K_{ATP/ADP,\{glyc1,cons.,pbm,glyc2\}} = 2,$ $K_{ATP/ADP,\{glyc1,cons.,glyc2\}} = 1,$ $K_{ATP/ADP,\{glyc1,cons.,phox\}} = 1,$ $K_{ATP/ADP,\{glyc1,cons.,pbm\}} = 2,$ $K_{ATP/ADP,\{glyc2,cons.,pbm,glyc2\}} = 2$
BIOM (BM)	$K_{BM,\{\beta-ox\}} = 1, K_{BM,\{\beta-ox,pbm\}} = 1$
FERM	$K_{FERM,\{ex-pyr\}} = 1$
GLYC	$K_{GLYC,\{NAD+,GLU1\}} = 1,$ $K_{GLYC,\{NAD+,GLU1, GLU2\}} = 1,$ $K_{GLYC,\{NAD+,CIT,GLU1\}} = 1,$ $K_{GLYC,\{NAD+,CIT,GLU1,GLU2\}} = 2$
GLN	$K_{GLN,\{in_gln\}} = 3, K_{GLN,\{in_gln,krebs\}} = 3,$ $K_{GLN,\{in_gln,pbm\}} = 3, K_{GLN,\{in_gln,krebs,pbm\}} = 3$
KREBS	$K_{KREBS,\{\beta-OX\}} = 1, K_{KREBS,\{PYR\}} = 1,$ $K_{KREBS,\{PYR,\alpha-KG\}} = 2$
NADH	$K_{NADH,\{Krebs,Phox\}} = 1,$ $K_{NADH,\{Krebs,glyc,phox\}} = 1,$ $K_{NADH,\{glyc,krebs,ferm\}} = 1,$ $K_{NADH,\{glyc,phox,ferm\}} = 1,$ $K_{NADH,\{krebs,phox,ferm\}} = 1,$ $K_{NADH,\{glyc,krebs,phox,ferm\}} = 1$
OXYG	$K_{O_2,\{in_O_2\}} = 1, K_{O_2,\{in_O_2,phox\}} = 1$
PHOX	$K_{PHOX,\{PC\}} = 1$
PROB.BIOM (PBM)	All K parameters are equal to 1, except $K_{PBM,\{\}}$
Control Variables for c in GLC,CONS., GLN or IN_O_2)	K_c =value assigned to c

Table 4: List of K parameters for the whole model (K parameters that do not appear in the table have values 0)

The three basic metabolites which are often used to observed metabolic oscillations O₂, NADH and ATP are oscillating between low and medium state. Glycolysis (GLYC) is also oscillating, which is due to the NADH/NAD⁺ and

ATP/ADP oscillators. The biomarkers of respiration and metabolic oscillations (NADH via Krebs) and PHOX oscillate as well. The model therefore reproduces the basic metabolic oscillating behaviours.

Let us point out that discrete abstract frameworks are, by construction, not quantitative. Within quantitative frameworks one could try to identify parameters with a sufficient precision in order to determine whether the oscillations are damped. Here the price to pay for abstraction is to be unable to address this question.

6.1.2 Biomass production conditions

Glutamine, which circulates with the highest concentration among amino acids, serves as a major bioenergy substrate and nitrogen donor for proliferating cells [16]. The amount of ATP in highly proliferative cell is not dramatically different from a quiescent cell in respiratory mode. The addition of glutamine triggers the accumulation of biomass. In this context, glucose and glutamine are considered essential nutrients providing ATP and carbon skeletons for building blocks of macromolecules respectively. This justifies the following set of initial values [$GLU = 1$, $Input_OXYG = 1$ and $Input_GLN = 1$] for this simulation that are critical nature of proliferative phenotypes (see Figure 11).

All the biomarkers for oscillating metabolism are clearly visible during this simulation in which other nodes in the graph keep their same initial state. The availability of glutamine (by turning [$GLU = 1$, $Input_OXYG = 1$]) has an effect on the production of biomass and later on the biomass. The nitrogen source for building blocks like nucleotides as well as DNA coincides with the presence of glutamine in the cell. We can also observe that glutamine at a given moment becomes an active nutrient for Krebs cycle through alpha Ketoglutarate. This is marked by the shift of Krebs from level 1 to 2 in the simulated chart.

6.2 Fermentation Condition

In normal fermentation, the minimal required conditions are firstly $O_2=0$ (a very low level or no more oxygen available) and the presence of either a low or high glucose concentration in cells. We investigate the effects of consumption of ATP in the metabolic system. It means that PHOX will stay at level 0 after a certain amount of time. Whether we start with an initial value of 0 or 1 for fermentation, there will be a tendency for fermentation to go towards 1 and stay at 1. In all fermentation processes, the principal purpose is to regenerate NAD^+ so that glycolysis can continue.

Most of the biomarkers of fermentation (GLYC, NADH, FERM and ATP/ADP)

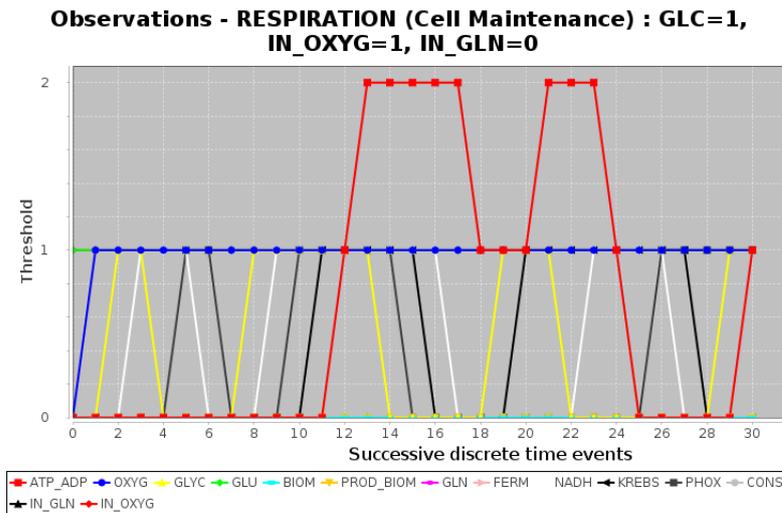


Figure 10: Respiration (Cell Maintenance). The initial state expresses normal cell maintenance conditions with medium level for glucose ($GLC=1$), presence of oxygen input and no excess of glutaminolysis or anaplerotic reactions. Time goes from left to right, with a unique variable change at each time step. So, one sees when a variable increases due to a line of its color that rises, and a decreasing line indicates a decrease. Increases and decreases are always from one unit at each time step.

have the tendency to oscillate: Glycolysis is high when NAD^+ and ATP are high, and Glycolysis is low when $NADH$ and ADP is high! We can also witness that Krebs and Phox as well, stay at level=0 during fermentation which is in accordance with biological observations.

6.3 Crabtree/Warburg effect

The capability to ferment sugars into ethanol is a key metabolic trait of yeasts. Crabtree-positive yeasts use fermentation even in the presence of oxygen, where they could, in principle, rely on the respiration pathway. This biologically observed phenomenon is surprising because fermentation has a much lower ATP yield than respiration (2 ATP vs. approximately 18 ATP per glucose) [2]. This normally occurs at high glucose level ($GLU=2$) and $Input_Oxygen=1$.

In accordance with these biological observations: Initially, the model state is in the Krebs phase and shortly afterwards it ends in the fermentation phase. Consequently, our model is able to reproduce the Crabtree effect.

Observations - Respiration (Croissance) with GLUC=1, IN_OXYG=1 and IN_GLN=1

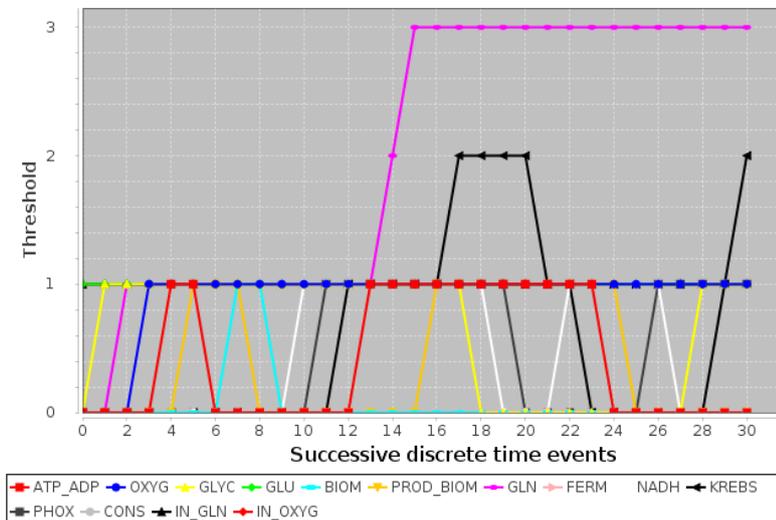


Figure 11: Respiration during cell division. The initial state expresses normal cell growth conditions with medium level for glucose, presence of oxygen input and presence of anaplerotic reactions.

7 Conclusion

Our coarse-grained modelling of energy metabolism allows us to study how main actors of metabolism (including nutrients) can influence or affect global metabolic phenotypes such as metabolic oscillations or metabolic transitions between fermentation and respiration. The high level of abstraction has the advantage to directly relate the variables of the model to the biological knowledge or readouts from cellular phenotypic screens.

This abstract model of energy metabolism reproduces the basic aspects of energy metabolism dynamics such as metabolic oscillations and the Warburg/Crabtree effect. This metabolic transition from respiration to fermentation is confirmed with the DyMBioNet simulation when the glucose intake variable is modified from medium (GLC=1) to high (GLC=2).

A long term goal of this study is to propose pertinent experiments using formal logic to confirm or refute certain hypotheses concerning energy metabolism or to study the consistency of anti-Warburg strategies. For other problematics such as optimizing trade off between biomass and storage for bioproduction, new variables need to be incorporated to make the model more suitable for these new problems.

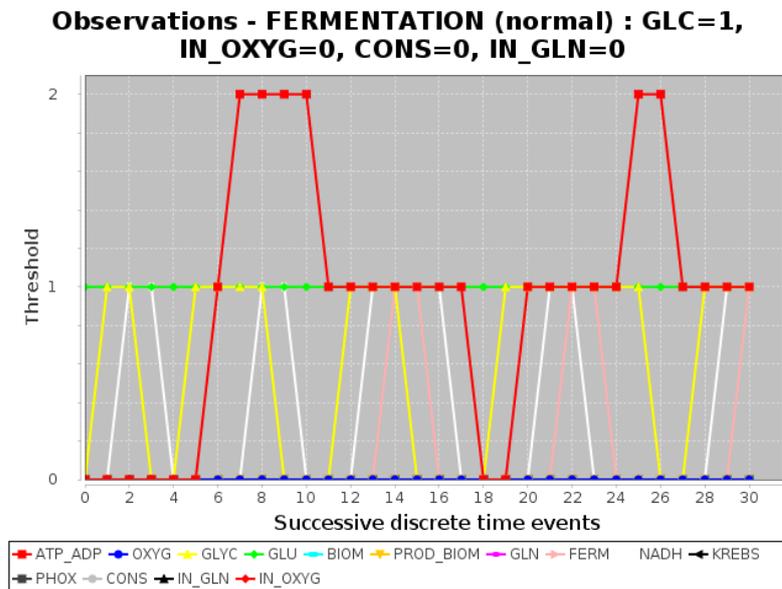


Figure 12: Initial state leading to fermentation, characterized by an absence of oxygen input and normal nutrient level (GLC=1).

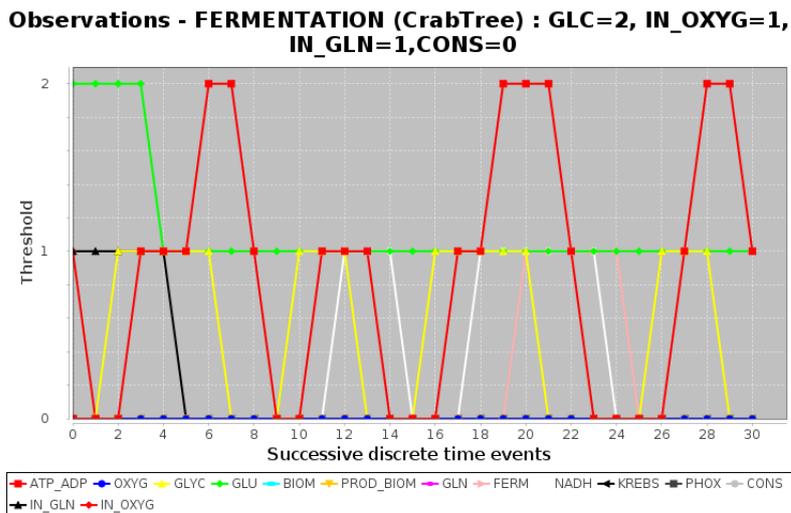


Figure 13: Fermentation showing the Crabtree effect under high glucose input, in the presence of oxygen

References

- [1] Hagman A, Säll T, Piskur J. Analysis of the yeast short-term Crabtree effect and its origin. *FEBS Letter* 2014, 281(1), 4805–4814.
- [2] Diaz-Ruiz R, Rigoulet M, Devin A. The Warburg and Crabtree effects: On the origin of cancer cell energy metabolism and of yeast glucose repression. *Biochimica et Biophysica Acta*, 2011, 1807, 568–576.
- [3] Levine AJ, Puzio-Kuter AM. The Control of the Metabolic Switch in Cancers by Oncogenes and Tumor Suppressor Genes. *Science* 2010, 330,1340–1344.
- [4] Tran Q, Lee H, Park J, Kim SH, Park J. Targeting Cancer Metabolism - Revisiting the Warburg Effects. *Toxicology Research*, 2016, 32(3),177–193.
- [5] Schwartz L, Buhler L, Icard P, Lincet H, Steyaert JM. Metabolic Treatment of Cancer: Intermediate Results of a Prospective Case Series. *Anticancer Research* 2014, 34(2),973–980.
- [6] Molenaar D, van Berlo R, de Ridder D, Teusink B. Shifts in growth strategies reflect tradeoffs in cellular economics. *Molecular Systems Biology* 2009, Epub 2009 Nov 3.
- [7] Gehrman E, Glässer C, Jin Y, Sendhoff B, Drossel B, Hamacher K. Robustness of glycolysis in yeast to internal and external noise. *Physical Review E* Epub 2011 Aug 8.
- [8] Thomas R., 1978. Logical analysis of systems comprising feedback loops. *J. Theoretical Biology* 1978, 73(4),631–656.
- [9] Thomas, R., Thieffry D, Kaufman M. Dynamical behavior of biological regulatory networks-I. *Bulletin Mathematical Biology* 1995, 57(2),247–276.
- [10] Fages F, Soliman S, Chabrier-Rivier N. Modelling and querying interaction networks in the biochemical abstract machine BIOCHAM. *Journal of Biological Physics and Chemistry* 2004, 4,64–73.
- [11] Calzone L, Fages F, Soliman S. BIOCHAM: an environment for modeling biological systems and formalizing experimental knowledge. *Bioinformatics* 2006, 22(14),1805–1807.

- [12] Bernot G, Comet J-P, Khalis Z. Gene regulatory networks with multiplexes . *Proceedings of European Simulation and Modelling Conference*, 2008, Le Havre, France, ISBN 978-90-77381-44-1,423–432.
- [13] Snoussi E. Qualitative dynamics of a piecewise-linear differential equations: a discrete mapping approach. *Dynamics Stability Systems* 1989, 4,189–207.
- [14] Bernot G, Comet J-P, Snoussi E. Formal methods applied to gene network modelling . *Logical Modeling of Biological Systems* , L. Fari del Cerro and K. Inoue Eds., Bioengineering and health science series. ISBN 978-1-84821-680-8, 2014, 245–289.
- [15] Bernot G, Comet J-P, Richard A, Guespin J. Application of formal methods to biological regulatory networks: extending asynchronous logical approach with temporal logic; *Journal of Theoretical Biology* 2004, 229,339–347.
- [16] Jin L; Alesi GN,Kang S. Glutaminolysis as a target for cancer therapy *Oncogene* 2016, 35(28): 3619–3625.
- [17] Snoussi E, Thomas, R, Logical identification of all steady states: the concept of feedback loop characteristic states.*Bulletin Mathematical Biology* 1993, 55(5),973–991.



Abstraction of the structure and dynamics of the biological neuron for a formal study of the dendritic integration

Ophelie Guinaudeau¹, Gilles Bernot¹, Alexandre Muzy¹
and Franck Grammont²

¹ Universite Cote d Azur, CNRS UMR 7271, Laboratoire I3S, France

² Universite Cote d Azur, CNRS UMR 7351, Laboratoire JAD, France

Abstract

Understanding how neurons integrate the thousands of inputs they receive is a fundamental issue of neuroscience research. For this purpose, we define a new model for studying the impact of the dendritic morphology on the neuronal function. Following the Cable Theory application to neuron modelling, we propose relevant abstractions to reduce the number of parameters while keeping biophysical accuracy. This allows us to demonstrate a theorem characterizing structural equivalence classes of neurons sharing the same input/output (I/O) function. The theorem implies that the dendritic morphology is, surprisingly, not as critical as expected with respect to the I/O function of the neuron.

1 Introduction

Understanding brain organization and the way it processes neuronal information is an interdisciplinary worldwide challenge [6]. Here we focus on the I/O function at the single neuron scale with a particular emphasis on neuron morphology. It is known since decades that dendritic arborization is the part of the neuron where most of the neuronal computation is performed. However, it has been largely neglected up to now in computational neuroscience, faced with the difficulty to reduce its complexity. For this purpose, we decided to systematically use the remarkable abstraction capabilities of theoretical computer science. We propose the first neuron model integrating dendritic morphology, based on formal methods. It permits to prove rigorous properties about the role of the dendrites morphology in the I/O function of the neuron.

An input signal received on the dendritic tree far from the soma can easily undergo a 40-fold attenuation [19]. It follows that strong distal excitatory signals may be annihilated by a weak inhibitory signal received closer to the soma. In accordance, inhibitory synapses seem to be mostly located on proximal dendrites in some cell types [12, 1].

Section 2 introduces the basics about the neuron biology. Section 3 quickly describes the existing neuron models and introduces our framework. Formal methods commonly separate static descriptions from dynamics ones. Section 4

thus defines the static description of our framework, mainly allowing to describe the structure of a neuron. Section 5 defines the dynamic description, allowing to rigorously link any input signal to its output signal for any given neuron. Our formal approach allows establishing a necessary and sufficient condition for two different neurons to have the same I/O function. Section 6 describes this result. We finally discuss the impact of neuron morphology on its function in light of our result.

2 Archetypical biological neuron

2.1 Structure-function relationship

An archetypical neuron consists of a cell body called the *soma* and of two kinds of extensions: *dendrites* on the one hand and an *axon* on the other hand (cf. Figure 1). Nervous signals travel from dendrites to the axon passing through the soma. Dendrites are tree structures which can be highly branched. Nervous impulses are received at specific points called *synapses*, mostly located all over the dendrites. These input signals then propagate along the dendrites and accumulate in the soma which behaves as a bath with tap turned on. If soma potential exceeds a given threshold, a nervous impulse is generated and transmitted to adjacent neurons through the axon, partially emptying the bath. It should be noted that there are different types of neurons whose structure may widely vary from the archetypical one [17].

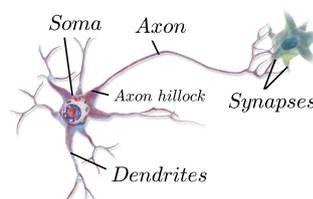


Figure 1: Structure of archetypical biological neuron.

2.2 Resting potential

The neuron, as any cell, is delimited by a membrane. It is a lipid bilayer pierced by channels conferring a selective permeability to specific ions. At rest, the neuronal membrane has a polarity: the interior of the cell is negatively charged compared to the extracellular medium. This difference of potential of about -70mV is called the *resting potential*, due to an unequal distribution of ions on both sides of the membrane. When the permeability is disrupted, ionic flows are generated leading the membrane potential to deviate from the resting

value. A depolarization is an increase in the potential beyond its resting value and a hyperpolarization is a decrease.

2.3 Action potential

An *action potential* (AP) is the physiological support of what we call neuronal information. It is a sudden and transient reversal of the membrane potential. It is generated at the axon hillock where the membrane is rich in Na^+ voltage-dependent channels (cf. Figure 1). As the extracellular media is highly concentrated in Na^+ , the opening of these channels causes a massive inflow of Na^+ resulting in a strong depolarization. These channels are rapidly inactivated and closed. At the same time, voltage-dependent K^+ channels open, leading to a K^+ outflow, returning the potential to its resting value. This repolarization is usually followed by a slight hyperpolarization. The amplitude and duration of APs are constant parameters for a given neuron type. Therefore both duration and amplitude cannot encode the nervous signal. Therefore, the intensity depends only on APs frequency.

2.4 Synaptic transmission

Although there are electrical synapses, most of them are chemical. Incoming APs trigger the release of neurotransmitters in the synaptic cleft. These small molecules bind to receptors on the postsynaptic neuron membrane causing ionic channels to open. This results in a membrane potential change, called a *postsynaptic potential* (PSP), whose voltage is proportional to the intensity of the stimulation. The PSP can be either a depolarization or a hyperpolarization being thus respectively excitatory (EPSP) or inhibitory (IPSP). If the APs frequency is sufficiently high, their individual effects (PSP) are added: it is called *temporal summation*.

2.5 Propagation through the dendritic tree

PSPs generated at synapses propagate along the dendrites towards the soma. As thousands of synapses are distributed over the dendritic tree, the signals are combined all along: it is called *spatial summation*. The spreading of the electrical signal in dendrites is decremental: the potential tends to return to its resting value due to leak channels allowing ions to cross the membrane following the electrochemical gradient.

2.6 Integration by the soma and axon transmission

The soma accumulates all the signals having undergone both temporal and spatial integrations. There is a threshold below which the potential change at the soma has no consequence. However, if the depolarization is strong enough, a new AP is generated to be transmitted to other neurons via the axon. The reaching of this threshold actually triggers the opening of voltage-dependent Na^+ channels at the origin of the APs (Section 2.3). The maximal theoretical APs frequency is imposed by the *refractory period* in which no AP can be generated despite a suprathreshold potential. It is due to a transient inactivation of the Na^+ voltage-dependent channels just after each AP beginning. This absolute refractory period is followed by a relative one during which the stimulation must be stronger than usual to trigger the opening.

The newly generated APs propagate along the axon towards their output synapses in a regenerative way. It means that, contrary to the signal conduction in dendrites, the APs do not undergo any alteration.

3 Single neuron models

Neuron models described in the literature can be categorized according to their goal of modelling, separating computational from biophysical models.

3.1 Computational neuron models

Computational models are considered as computational units thus highly idealized. The first example is the formal neuron proposed by McCulloch and Pitts in 1943 [11]. It is a binary neuron which performs a weighted sum of its inputs. A Heavyside function is applied to calculate its output: 1 if the sum exceeds the threshold and 0 otherwise. It allows elementary logic calculations and it is mostly used as part of networks for artificial intelligence purposes. Basically, it is a bio-inspired calculation and it should not be used for a deep understanding of the neuronal functioning.

3.2 Biophysical neuron models

Biophysical models are mostly used to understand the neuronal behaviors. They allow to focus on specific mechanisms and usually consider the membrane potential as the key variable. The potential depends on the electrical membrane properties. One of the most famous biophysical model was presented by Hodgkin and Huxley in 1952 [7]. It describes the dynamics of ion channels governing the initiation of the AP by a set of non linear differential

equations. This model is accurate and compatible with experimental observations. However, it is very complicated and difficult to validate because of its large number of parameters [13]. This model thus led to simplified models such as the FitzHugh-Nagumo model [5].

3.2.1 The Integrate-and-Fire model

The *Integrate-and-Fire (I&F)* model, proposed by Lapicque in 1907, does not focus on the molecular mechanisms governing the AP. It focuses on the I/O function of the neuron [10, 3]. The membrane is described as an electrical circuit constituted of a capacitor in parallel with a resistor. In the *leaky-I&F* model, an additional term is added to take into account the leak of ions through the membrane: $C \frac{dV}{dt} = -g_l(V - V_0) + I(t)$, where V is the membrane potential, C the capacitance, V_0 the resting potential, g_l the leak conductance and $I(t)$ the injected current (external or synaptic). This differential equation describes the potential dynamics below the threshold. The AP initiation is not explicitly represented: in addition, when V reaches the threshold, an AP is generated. At the same time, the potential is reset and the threshold value is updated to take into account the refractory period [18]. There are other extensions of the I&F model such as quadratic or exponential models [2].

3.2.2 Cable Theory applied to dendrites

Most of the existing neuron models are punctual, meaning that the neuron is equivalent to a point, ignoring its morphology. However, there are biophysical models focusing on the influence of structural characteristics on the nervous signal propagation. One of them is the Cable Theory applied to dendrites, proposed by Rall, a pioneer of dendrites modelling, in the 1960's [14, 15]. *Cable Theory* is based on a second order partial differential equation developed by Lord Kelvin in 1850 to describe the attenuation of the electrical signal spreading along a submarine cable. This concept was later applied to dendrites [15]. The dendrites are considered as cylindrical cables along which an electrical signal passively propagates. This phenomenon is described as follows: $\lambda \frac{\partial^2 V}{\partial x^2} = \tau \frac{\partial V}{\partial t} + V$, where V is the membrane potential variation from the resting value, x is the traveled distance over the cylinder, t is the time, $\lambda = \sqrt{\frac{r_m}{r_i}}$ is the space constant and $\tau = r_m \times c_m$ is the time constant where r_i is the intracellular resistance, r_m the membrane resistance and c_m the membrane capacitance.

Given an initial condition, the equation is solved analytically for different boundary conditions, assuming an infinite or a finite cable. The equation describes the attenuation undergone by the nervous signal during its conduction,

taking into account the length of dendrites, their diameter and the composition of the membrane. It is the basis of essentially all simulators taking the neuronal morphology explicitly into account such as Neuron or GENESIS [1]. However, the main drawback of this approach is the number of parameters involved [13].

3.3 A framework dedicated to the study of dendritic integration

In the remainder of this chapter, we study the impact of the morphology of the dendritic trees on the neuron I/O function. For this purpose, we have developed the first formal neuron model integrating dendrites morphology. We are at the interface between computational and biophysical models as we are not directly interested in the cellular mechanisms involved in neuron properties but rather in the “computational properties” of the neuron. Notwithstanding, our parameters can always be correlated with observable biophysical entities. Our major contribution is to model this process by mixing discrete and continuous modelling. Since the AP is known for its speed and its stereotypical properties, it is reasonable to consider it as an instantaneous event identified by its time of occurrence (conventionally called *spike*). More precisely, in our model a neuron receives spikes sequences at synapses. Those discrete inputs are immediately converted into continuous signals which then conduct through the dendritic tree towards the soma. The soma integrates all the signals and when a threshold is reached, a spike is emitted on the output. Our soma modelling is inspired by the leaky-I&F model. To investigate the dendritic integration, we focus on dendrites modeling, choosing Cable Theory as a basis. Although based on noticeable hypotheses [4], Cable Theory is credible and it allows very efficient abstractions. We have reduced the number of parameters while keeping the biophysical accuracy. Thanks to this approach, we are able to study equivalent classes of dendritic integration.

It is worth noting that our framework will include an *abstract* modelling of the ionic charge flows in the excited neuron. These charges are directly related to the membrane potential difference compared to the resting value, depending on the local membrane properties. Charges travel through the dendritic tree and they are summed at branching points. Finally, they control the soma potential and consequently the spike emissions. From our abstract point of view, we will make no difference between “charges” and “potentials.”.

4 Static description of a neuron

According to computer science, a *tree* is recursively defined as a root node to which is attached a *forest* of children trees. Our neuron model consists of a forest of dendritic trees connected to a root soma and we ignore the axon as it

transmits the signal from the soma without any loss (cf. Figure 2) . Synapses on the dendritic forest are the input ports of the system. They receive sequences of spikes and each of them triggers a local change of the electrical potential. The potential reaches a maximal absolute value after a short period of time and then returns progressively to its resting value. The kinetic is proper to each synapse and characterized by the parameters given in Definition 1.

Definition 1. [Synapse] A synapse is a triplet $s = (\nu_s, \hat{\tau}_s, \check{\tau}_s)$ where:

- ν_s is a non-zero real number called maximal potential of a spike for s . If $\nu_s > 0$ then s is said excitatory, otherwise it is said inhibitory;
- $\hat{\tau}_s$ and $\check{\tau}_s$ are strictly positive real numbers respectively called rise time and descent delay of the potential.

Continuous signals generated at synapses then propagate in the dendritic trees towards the soma. We choose to divide dendrites into homogeneous elementary compartments delimited by branching points and synapses (Definition 4). In agreement with Cable Theory, we hypothesize a passive signal propagation with leakage. Whatever the boundary conditions, the analytical solution of the linear cable equation is of the form $V(x) = V_0 \times \alpha$ where α , is a variable depending on x . Therefore, we decided to describe the potential at the output of a compartment as equal to the potential at its input attenuated by a coefficient α after a delay δ (Definition 2). Grouping the parameters of the Cable Theory into those two abstract parameters greatly simplifies our model.

Definition 2. [Compartment] A compartment is a couple $c = (\delta_c, \alpha_c)$ where:

- δ_c is a real number greater or equal to zero called the crossing delay for c ;
- α_c is a real number such that $\alpha_c \in]0, 1]$, called the attenuation at the end of c . If $\delta_c = 0$, then $\alpha_c = 1$.

The soma accumulates input signals coming from the dendrites making its potential changing gradually. At the same time, there is a leak of charges making the potential slowly returning to its resting value. The soma is also characterized by its activation threshold from which it can emit a spike, and by the duration of the refractory periods (Definition 3). Note that the relative refractory period can be technically expressed by an augmented threshold.

Definition 3. [Soma] A soma is a tuple $\nabla = (\theta, \hat{\theta}, \rho, \hat{\rho}, \gamma)$ where all the parameters are strictly positive real numbers: θ is called the activation threshold, $\hat{\theta}$ the threshold augmentation, ρ the absolute refractory period, $\hat{\rho}$ the relative refractory period and γ is called the leak.

Definition 4. [Neuron] A neuron is a labelled rooted tree satisfying the following conditions: any non-root node having 0 or 1 child is labelled by a synapse, any branch is labelled by a compartment and any non-root node having at least 2 children is called a branching (and is not labelled). The root of the tree is labelled by a soma. Given a neuron N , we note $Sy(N)$ the set of its synapses and $Co(N)$ the set of its compartments. Moreover, the direct children of the soma are called dendritic trees. Finally, we note \mathcal{N} the set of all neurons.

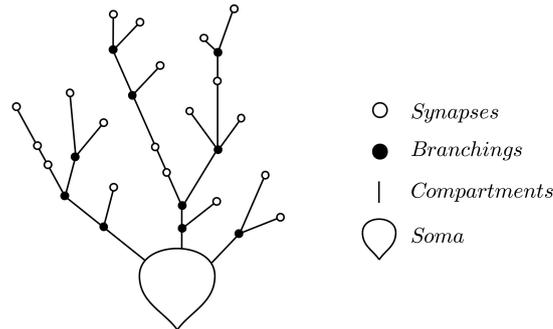


Figure 2: Schematic representation of our neuron model structure.

In computer science, the parent of a node is its only neighbour on the way to the root. The root is the only parentless node. Children of a node are its neighbours except its parent node. In the neuron model, the information goes from the leaves (synapses) to the root (soma). To avoid any confusion, we use the terms input/output and predecessor/successor to replace the couple child/parent. Given a neuron N and a compartment $y \xrightarrow{c} z$: y is called the *input node* of c , z the *output node* of c , c the *output compartment* of y and c the *input compartment* of z . We note $In(z)$ the set of the input compartments of z . A *predecessor* of $y \xrightarrow{c} z$ is a compartment in N of the form $x \xrightarrow{c'} y$ and we note $Pred(c)$ the set of the predecessors of c . Also, we call *contributor compartments* of c ($CC(c)$), the compartments at delay 0 from y . We define the *synaptic contributors* ($CS(c)$) in a similar way. More formally:

Definition 5. [Contributors] Given a neuron N and a compartment $y \xrightarrow{c} z$, $\forall c' \in Pred(c)$, the set of the contributor compartments of c noted $CC(c)$ is defined inductively by:

- If $\delta_{c'} \neq 0$ then $c' \in CC(c)$;
- If $\delta_{c'} = 0$ then $CC(c') \subset CC(c)$.

Moreover, the set of the synaptic contributors of $y \xrightarrow{c} z$ noted $CS(c)$ is the subset of $Sy(N)$ defined inductively by:

- If $y \in Sy(N)$ then $y \in CS(c)$;
- $\forall c' \in Pred(c)$, if $\delta_{c'} = 0$ then $CS(c') \subset CS(c)$.

5 The state of a neuron and its dynamics

To describe the state of a neuron and its dynamics, we need to introduce the notion of “segment”:

Notation 1. Given a set E , a segment with values in E is an application $\omega : [0, t] \rightarrow E$ where $t \in \mathbb{R}^+ \cup \{+\infty\}$ with the convention that $[0, +\infty] = \mathbb{R}^+$. We note Sgt the set of all the segments and we endow $Sgt(E)$ with a partial internal law of concatenation as follows:

if $\omega_1 : [0, t_1] \rightarrow E$ and $\omega_2 : [0, t_2] \rightarrow E$ are two segments such that $t_1 \in \mathbb{R}^+$ and $\omega_1(t_1) = \omega_2(0)$, then $\omega_1 \cdot \omega_2$ is the concatenated segment ¹ $\omega_1 \cdot \omega_2 : [0, t_1 + t_2] \in E$ such that:

- $(\omega_1 \cdot \omega_2)(t) = \omega_1(t)$ if $t \leq t_1$
- $(\omega_1 \cdot \omega_2)(t) = \omega_2(t - t_1)$ if $t \geq t_1$

We note $\prod_{i=1}^n \omega_i$ for $\omega_1 \cdot \omega_2 \cdot \dots \cdot \omega_n$, the concatenation of the n segments.

Moreover, if \square is a binary operation on E , it can be extended as follows: if ω_1 and ω_2 are segments of the same length t_0 , we note $\omega_1 \square \omega_2 : [0, t_0] \rightarrow E$ the segment such that $(\omega_1 \square \omega_2)(t) = \omega_1(t) \square \omega_2(t)$ for all $t \in [0, t_0]$. When $\square = +$, we accept the notation $\sum_{i=1}^n \omega_i$ for $\omega_1 + \omega_2 + \dots + \omega_n$.

The input signals are received at the synapses in the form of infinite segments taking value 1 at the times of the spikes and 0 otherwise (Definition 6). The output signal will be of the same type so that our modelling opens a way of building a network where the input of a neuron would come from the output of other ones.

Definition 6. [Signal] A signal is a segment $\omega : \mathbb{R}^+ \rightarrow \{0, 1\}$ such that:

$$\exists r \in \mathbb{R}^{*+}, \forall t \in \mathbb{R}^+, (\omega(t) = 1 \Rightarrow (\forall t' \in]t, t+r[, \omega(t') = 0))$$

The carrier of ω is defined by: $Car(\omega) = \{t \in \mathbb{R}^+ | \omega(t) = 1\}$. A signal such that $Car(\omega)$ is a singleton $\{u\}$ is called a spike at the time u , noted ω^u .

We call “trace”, the potential change triggered by a spike. It directly depends on the synapse parameters (Definition 1). This continuous variation

¹by convention $(t_1 + \infty) = +\infty$

is exponential from a biophysical point of view [15, 16]. For the sake of simplicity, we approximate it as linear (Definition 7). It seems reasonable as the experiments do not always follow the theoretical model and show a significant variability [8]. The temporal summation observed in biology is reproduced by making the sum of the respective traces of the successive spikes. It gives what we call the trace of the signal (cf. Figure 3).

Definition 7. [Trace of a signal] *The trace of a spike ω^u on a synapse s is the segment v_{s,ω^u} defined by:*

- If $t \leq u$ then $v_{s,\omega^u}(t) = 0$;
- If $u \leq t \leq u + \hat{\tau}_s$ then $v_{s,\omega^u}(t) = \frac{\nu_s}{\hat{\tau}_s}(t - u)$;
- If $u + \hat{\tau}_s \leq t \leq u + \hat{\tau}_s + \check{\tau}_s$ then $v_{s,\omega^u}(t) = \frac{\nu_s}{\check{\tau}_s}(u + \hat{\tau}_s + \check{\tau}_s - t)$;
- If $u + \hat{\tau}_s + \check{\tau}_s \leq t$ then $v_{s,\omega^u}(t) = 0$

Moreover, given an input signal ω_s on a synapse s , the trace of ω_s is defined by the real segment $v_{s,\omega_s} = \sum_{u \in \text{Car}(\omega_s)} v_{s,\omega^u}$.

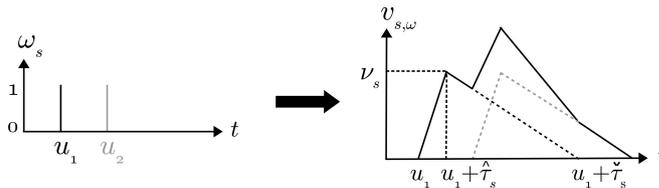


Figure 3: *Trace of a signal.* The trace of a signal is the sum of the traces of its spikes. Each spike received at a synapse s causes a maximum variation ν_s of the potential after a time $\hat{\tau}_s$ followed by a return to the resting potential with a delay $\check{\tau}_s$.

The state of a neuron at a given time is the value of the potential at every point of it. It includes the state of the soma and the state of all the compartments (Definition 9).

The value of the potential at the soma is not sufficient to characterize its state. One must also know the time elapsed since the last emitted spike to manage refractory periods. We thus define the state of the soma as a couple (e, p) where e is the elapsed time since the last spike and p is the current soma potential. Due to the biological properties of the neuron, there is a constraint on this couple which has to be nominal (i.e. normal): when p is above the threshold, e is necessarily less than the refractory period duration (Definition 8). Indeed,

when p exceeds the threshold, there are only two possibilities: either e is in the refractory period, or a spike is emitted and thus e is reset to 0.

Definition 8. [Nominal] Given a soma $\nabla = (\theta, \hat{\theta}, \rho, \hat{\rho}, \gamma)$, a couple (e, p) where $e \in \mathbb{R}^+$ and $p \in \mathbb{R}$, is nominal if $(e < \rho)$ or $(p < \theta)$ or $(p < \theta + \hat{\theta} + \frac{\hat{\theta}(\rho - e)}{\hat{\rho}})$. We note $Nominal(\nabla)$ the set of all the nominal couples.

Based on the compartment definition (Definition 2) and knowing the potential at its input node, it is easy to calculate the potential at its output after its crossing delay. The potential at each point of a compartment at a given time is deduced from this relationship. We thus define the state of a compartment c at a time h as a segment $v_c^h(t)$ which describes the evolution of the potential at its output between h and $h + \delta_c$ (cf. Figure 4).

Definition 9. [State of a neuron] The state of a neuron N is a triplet $\eta = (V, e, p)$ where:

- V is a family of segments, indexed by $Co(N)$, the set of the compartments of N ; each segment is of the form $v_c : [0, \delta_c] \rightarrow \mathbb{R}$ where δ_c is the crossing delay of the compartment c . For each compartment c :

$$v_c(\delta_c) \sim \left(\sum_{c' \in CC(c)} v_{c'}(0) \right) \alpha_c$$

where by convention the comparator " \sim " is: " $=$ " if its input node is a branching, " \geq " if its input node is an excitatory synapse or " \leq " if its input node is an inhibitory synapse;

- $e \in \mathbb{R}^+$ represents the elapsed time since the last spike and $p \in \mathbb{R}$ is called the soma potential such that the couple (e, p) is nominal for the soma of N .

We note ζ_N the set of all the states of the neuron N .

The compartments dynamics is done by "segments sliding" by an arbitrary timestep Δ (cf. Figure 5). The potential at the input of c is calculated from the potential at the output of its contributor compartments while taking into account the spikes received at synaptic contributors. More formally:

Theorem 1. [Dynamics of the compartments]

Let a neuron N , an initial state $\eta^I = (V^I, e^I, p^I)$ and an input signal $\mathcal{S} = \{\omega_s\}_{s \in Sy(N)}$. There exists a unique application $\mathcal{V} : \mathbb{R}^+ \times Co(N) \rightarrow Sg(\mathbb{R})$, which associates a segment $v_c^h(t) : [0, \delta_c] \rightarrow \mathbb{R}$ for each couple $(h, c) \in \mathbb{R}^+ \times Co(N)$ and such that, for any $\Delta \leq \inf(\{\delta_c | c \in N \wedge \delta_c > 0\})$ and for each couple (h, c) :

1. $v_c^0 = v_c^I$
2. for any $t + \Delta \leq \delta_c$, $v_c^{h+\Delta}(t) = v_c^h(t + \Delta)$,
3. for any t such that $\delta_c - \Delta \leq t \leq \delta_c$, $v_c^{h+\Delta}(t) = \left(\sum_{c' \in CC(c)} v_{c'}^h(t - \delta_c + \Delta) + \sum_{s \in CS(c)} v_{s, \omega_s}^h(t - \delta_c + \Delta) \right) \alpha_c$.

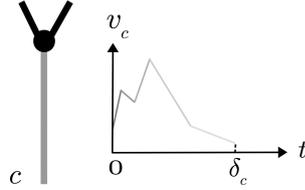


Figure 4: *The state of a compartment.* The state of a compartment c is the potential at its output between $t = 0$ ($v_c(0)$) and the crossing delay of c ($v_c(\delta_c)$).

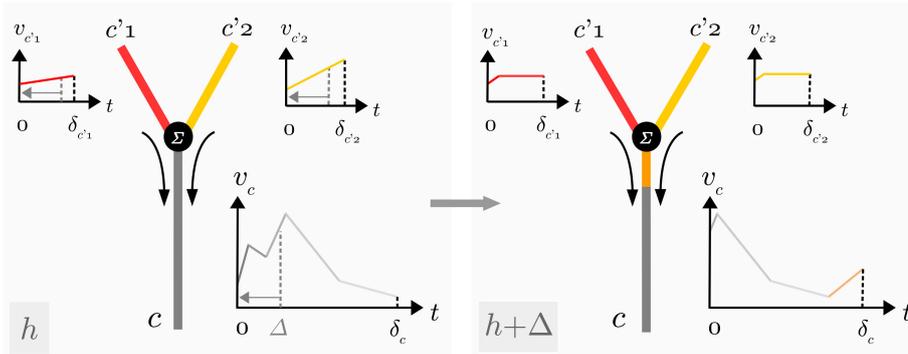


Figure 5: *Dynamics of the compartments.* The state of the compartments at time $h + \Delta$ can be calculated from the state of the compartments at time h by “sliding.” The potential at the input of a compartment is the sum of the potentials at the output of its contributors.

Lastly, the soma dynamics is inspired by the leaky-I&F model. At a time t , the soma potential depends on inputs coming from the dendritic trees (denoted $F(t)$) and on the leak γ applied to the current value of the potential. From an initial condition (e_0, p_0) , we can define the P_F function which describes the evolution of the soma potential over time (Lemma 1). To account for the refractory period, the couple (e, p) is forced to remain nominal (Definition 8). So when p reaches the threshold, its value is reduced (by θ) and e is reset to 0. P_F is hence discontinuous at this set of times that defines the carrier of the output signal (Definition 6).

Lemma 1. [Technical lemma] *Given a soma $\nabla = (\theta, \hat{\theta}, \rho, \hat{\rho}, \gamma)$, there exists a unique family of functions $P_F : \text{Nominal}(\nabla) \times \mathbb{R}^+ \rightarrow \mathbb{R}$ indexed by the set of continuous and lipschitzian functions $F : \mathbb{R}^+ \rightarrow \mathbb{R}$, such that for any couple $(e_0, p_0) \in \text{Nominal}(\nabla)$, P_F satisfies:*

- $P_F(e_0, p_0, 0) = p_0$
- $\forall t \in \mathbb{R}^+$ the right derivative $\frac{dP_F(e_0, p_0, t)}{dt}$ exists and is equal to $F(t) - \gamma \cdot P_F(e_0, p_0, t)$
- $\forall t \in \mathbb{R}^+$, $\ell = \lim_{u \rightarrow t^-} (P_F(e_0, p_0, u))$ exists and:
 - if $(t + e_0, \ell) \in \text{Nominal}(\nabla)$ then $P_F(e_0, p_0, t)$ is continuous and is derivable if $t > 0$ therefore $P_F(e_0, p_0, t) = \ell$,
 - otherwise, for any $u \geq t$, $P_F(e_0, p_0, u) = P_G(0, \ell - \theta, u - t)$ where for any $x \in \mathbb{R}^+$, $G(x) = F(x + t)$.

The lemma hereinabove allows us to define the dynamics of a neuron which associates a state to each time.

Definition 10. [Dynamics of a neuron] *Given a neuron N , an initial state $\eta^I = (V^I, e^I, p^I)$ and an input signal $\mathcal{S} = \{\omega_s\}_{s \in \text{Sy}(N)}$, the dynamics of N is the infinite segment $d : \mathbb{R}^+ \rightarrow \zeta_N$ defined by:*

- $d(0) = \eta^I$
- $\forall h \in \mathbb{R}^+$, $d(h) = \eta = (V, e, p)$ where:
 - $V = \{\mathcal{V}(h, c)\}_{c \in \text{Co}(N)} = \{v_c^h\}_{c \in \text{Co}(N)}$ where the application \mathcal{V} is the one from Theorem 1;
 - Consider beforehand $F(\bar{h}) = \sum_{c \in \text{In}(\nabla)} v_c^{\bar{h}}(0)$.
 F is lipschitzian at \bar{h} as in any point its derivative is between $\left(\sum_{s \in \text{Sy}(N)} \frac{-\nu_s}{\tau_s} \right)$ and $\left(\sum_{s \in \text{Sy}(N)} \frac{\nu_s}{\tau_s} \right)$. According to Technical lemma 1, there exists a unique function P_F such that $P_F(e^I, p^I, 0) = p^I$ and $\forall \bar{h}$, $\frac{dP_F(e^I, p^I, \bar{h})}{d\bar{h}} = F(\bar{h}) - \gamma \cdot P_F(e^I, p^I, \bar{h})$. Therefore, if $P_F(e^I, p^I, \bar{h})$ is continuous on the $]0, \bar{h}]$ interval, then $e = e^I + \bar{h}$. Otherwise, let \bar{h}' be the greatest \bar{h} such that $P_F(e^I, p^I, \bar{h})$ is discontinuous, then $e = h - \bar{h}'$.
 - Considering the previous P_F function, $p = P_F(e^I, p^I, h)$.

6 Remarkable properties

We found that any neuron can be reduced to a *pin-holder*. A “pin-holder” is a neuron where each dendritic tree is simply one synapse linked to the soma by a single compartment. The pin-holder corresponding to a neuron can be obtained by applying the decomposition function defined below:

Definition 11. [Decomposition function] We note \mathcal{P} , the set of all the pin-holders. The decomposition function $fd : \mathcal{N} \rightarrow \mathcal{P}$ is the application which associates to any neuron $N \in \mathcal{N}$ the neuron from \mathcal{P} built as follows:

- $fd(N)$ has the same soma than N : ∇ ,
- $fd(N)$ has the same set of synapses than N : $Sy(fd(N)) = Sy(N)$,
- For each synapse $s \in Sy(N)$, there exists a unique path made of compartments c_1, \dots, c_n (where for all i , $c_i = (\delta_{c_i}, \alpha_{c_i})$) from ∇ to s in N . Therefore, the compartment linking ∇ to s in $fd(N)$ is the couple (δ_s, α_s) such that $\delta_s = \sum_{i=1}^n \delta_{c_i}$ and $\alpha_s = \prod_{i=1}^n \alpha_{c_i}$.

The definition of the state of a pin-holder differs from the state of its corresponding neuron only at the compartments level, as the soma remains the same.

Definition 12. [State of a pin-holder] Given the state of a neuron N , we note $fd_N : \zeta_N \rightarrow \zeta_{fd(N)}$ the application which associates to each state $\eta = (V, e, p)$ of N , the state $fd_N(\eta) = (\bar{V}, e, p)$ of $fd(N)$ such that for each synapse s of $fd(N)$ and c its output compartment:

$$\bar{v}_c = v_{c_1} \cdot (v_{c_2} \times \alpha_1) \cdot \dots \cdot (v_{c_n} \times \prod_{i=1}^{n-1} \alpha_i) = \prod_{i=1}^n \left(v_{c_i} \times \prod_{j=1}^{i-1} \alpha_j \right)$$

such that c_1, \dots, c_n is the path of compartments from ∇ to s in N where for each i , $c_i = (\delta_{c_i}, \alpha_{c_i})$.

Thanks to the pin-holder concept we have introduced, we demonstrated the following theorem:

Theorem 2. [The pin-holder theorem] Let N_1 and N_2 be two neurons. If $fd(N_1) = fd(N_2)$ then, for any input signal \mathcal{S} , N_1 and N_2 have the same soma dynamics meaning that the evolution of the (e, p) couples with time is the same after a certain delay. Therefore, they have the same output signal (ω_{∇}) after a certain delay.

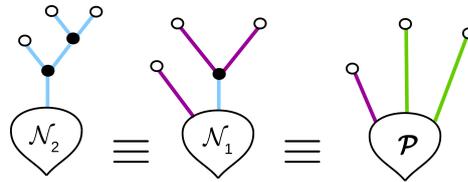


Figure 6: *The pin-holder theorem.* The theorem defines equivalent classes of structure and demonstrates that the exact shape of dendritic trees is not critical in the I/O function of the neuron. The neurons \mathcal{N}_1 , \mathcal{N}_2 and \mathcal{P} represented here belong to the same class. The parameters are $(\alpha = 0.1, \delta = 1)$ for the blue compartments, $(\alpha = 0.25, \delta = 2)$ for the purple compartments and $(\alpha = 0.125, \delta = 3)$ for the green compartments.

The interesting result is that a pin-holder is the canonical representative of a large set of neurons (cf. Figure 6). According to Theorem 2, these neurons have the same soma dynamics and therefore the same output for a given input. This result shows that the dendrites morphology is, surprisingly, not critical when only taking into account the I/O function of the neuron. Instead of the precise morphology, attenuation and delay are the key parameters in this function.

Moreover, the normalization into a pin-holder is an important reduction of the neuron complexity and this is very efficient for computational purposes.

Notice that the pin-holder is a sort of extension of the classical formal neuron (Section 3). Indeed, the α parameter of the compartments in our model is comparable to the weights applied to the inputs in the classical version. The main difference is the delay brought by our δ parameter.

7 Conclusion

The complex neuronal information processing emerges from an appropriate arrangement of a large number of neurons, each behaving as a device with potentially rich computational capabilities. Focusing on the I/O function at the scale of an individual neuron is thus a fundamental step in understanding the whole brain function. Cable Theory provides a good approximation to link the structure of the neuron to its function. As we are interested in the impact of dendrites morphology, we decided to base our work on this founding model. We proposed relevant abstractions to reduce the number of parameters while keeping the biophysics relevance. This enabled us to demonstrate the *pin-holder theorem* showing that a very large number of different dendritic tree structures share the same I/O function. Consequently, and unexpectedly,

under the assumptions of Cable Theory, it implies that the precise morphology does not have a critical impact on the neuron I/O function, only delays and attenuations matter. It then comes that the dendritic morphology is probably essentially driven by the fact that neuronal structure is the result of a progressive development during neuroembryogenesis.

Among the basic Cable Theory assumptions, the passive conduction of the signals through the dendrites is a strong one: Some neurons exhibit active mechanisms at the level of dendrites [8]. Another assumption is that synapses parameters do not noticeably vary in time although inhibitory synapses seem to be more effective when an excitatory signal passes close to them [9, 17]. Taking into account these properties to model more complex neurons would probably modify our basic results, but we provide a first track on how the inputs distributed over the dendritic tree interact in time and space to determine the I/O function of the neuron.

A future direction is to take benefit from the *pin-holder theorem* in order to study interconnected neurons. Our framework opens up perspectives for precise network studies for two principal reasons. First, thanks to our hybrid approach, we would only need to work on discrete signals as the continuous part of the system is restricted to the “internal” part of the neuron. Also, the *pin-holder theorem* allows to drastically reduce the neuron model complexity and this is a strong advantage for simulating and formally reasoning on networks.

References

- [1] Bower, J.M., Beeman, D. (2003) The book of GENESIS: exploring realistic neural models with the General Neural Simulation System. *Internet Edition*.
- [2] Brette, R., Gerstner, W. (2005) Adaptive Exponential Integrate-and-Fire Model as an Effective Description of Neuronal Activity. *American Physiological Society*, **94(5)**: 3637-3642.
- [3] Brunel, N., Van Rossum, M.C. (2007). Lapicque s 1907 paper: from frogs to integrate-and-fire. *Biological cybernetics*, **97(5-6)**: 337-339.
- [4] Byrne, J.H., Roberts, J.L. (2004) From molecules to networks: an introduction to cellular and molecular neuroscience. *Elsevier Inc*.
- [5] FitzHugh, R. (1961) Impulse and physiological states in models of nerve membrane. *Biophysical journal*, **1**: 445-466.

- [6] Grillner, S., Ip, N., Koch, C., Koroshetz, W., Okano, H., Polachek, M., Poo, M., Sejnowski, T.J. (2016) Worldwide initiatives to advance brain research. *Nature neuroscience*, **19(9)**: 1118-1122.
- [7] Hodgkin, A., Huxley, A.F. (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, **117(4)**: 500.
- [8] Koch, C. (2004) Biophysics of computation: information processing in single neurons. *Oxford university press*.
- [9] Koch, C., Poggio, T., Torres, V. (1982) Retinal ganglion cells: a functional interpretation of dendritic morphology. *Science*, **298(1090)**: 227-263.
- [10] Lapique, L. (1907) Recherches quantitatives sur l'excitation électrique des nerfs traitée comme une polarisation. *J. Physiol. Pathol. Gen.*, **9(1)**: 620-635.
- [11] McCulloch, W.S., Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, **5(4)**: 115-133.
- [12] Megias, M., Emri, Z.S., Freund, T.F., Gulyas, A.I. (2001) Total number and distribution of inhibitory and excitatory synapses on hippocampal CA1 pyramidal cells. *Neuroscience*, **102(3)**: 527-540.
- [13] Popper, K. (2014) Conjectures and refutations: The growth of scientific knowledge. *Routledge*.
- [14] Rall, W. (1959). Branching dendritic trees and motoneuron membrane resistivity. *Experimental neurology*, **1(5)**: 491-527.
- [15] Rall, W. (1962) Theory of physiological properties of dendrites. *Annals of the New York Academy of Sciences*, **96(4)**: 1071-1092.
- [16] Rall, W. (2011). Core conductor theory and cable properties of neurons. *Comprehensive physiology*.
- [17] Stuart, G., Spruston, N., Hausser, M. (2016) Dendrites. *Oxford*.
- [18] Tuckwell, H.C. (1988) Introduction to Theoretical Neurobiology Volume 1. Linear Cable Theory and Dendritic Structure. *Cambridge University Press*.
- [19] Williams, S.R., Stuart, G.J. (2002) Dependence of EPSP efficacy on synapse location in neocortical pyramidal neurons. *Science*, **295(5561)**: 1907-1910.



Studying Metabolic Networks structure: Elementary Flux Modes vs Minimal Cut Sets Analysis.

Nguyen V. N. Tung¹, Marie Beurton-Aimar², Sophie Colombié³

¹ EMBL-EBI, Wellcome Genome Campus, Cambridge, UK

² LaBRI, UMR 5800, Univ. Bordeaux, France.

³ INRA Bordeaux, UMR 1332, Fruit Biology and Pathology, Bordeaux, France.

Abstract

In recent year, amount of biological data has rapidly increased due to high-throughput experiments on living organisms. This fact leads to the emerging trend towards studying and analyzing of large metabolic networks. The computation of network structural properties could help to explore inherent metabolic behaviors. Elementary Flux Modes (EFMs) analysis is one of the tools able to give a useful insight into such metabolic networks. However, the results produced by this tool could be so big that their analysis is not feasible easily. Recently, a dual approach to EFMs calculus has been developed with calculation of Minimal Cut Sets (MCSs), i.e. the identification of sets of reactions which prevent the execution of an objective reaction. EFMs are minimal sets of enzymes that operate at steady state with all irreversible reactions used in the appropriate direction, whereas MCSs are minimal sets of reactions which can block the yield of a certain product from given substrates. The results of MCSs are expected smaller than those of EFMs in a given metabolic network. In order to verify this hypothesis, we have tested this assumption with 6 networks each having different size and structure: one example of a single pathway, the TriCarboxylic Acid (TCA) cycle, three descriptions of the mitochondria energy metabolism (in 3 different tissues: *liver*, *yeast* and *muscle*) and two models of the central carbon metabolism of heterotrophic plant cells. We have found that the number of MCSs is higher than EFMs's for TCA cycle and mitochondrial networks. Conversely, with biggest networks, the central metabolism plant cells, the number of MCSs is smaller than those of EFMs but remains very large. However this analysis has revealed that the size of each cut set does not grow with the size of the network and the study of small MCSs, i.e. the capacity to stop some production from stopping a small number of reactions, can lead to identify very interesting reactions patterns which are more or less necessary for the network activity. Future works will focus on this point.

1 Introduction

Metabolic networks are well-known to be one of the most complex cellular systems [36, 33] regarding on the number of shared metabolites by different reactions. In this context, specific modeling and analysis methods have to be designed. A number of studies are concentrated on the characterization of metabolic networks by means of graph theory to gain an insight into the global network structure. In that way, finding topological properties can help us to analyze biological structures [2, 6, 37]. Depending on the goals of the analysis, these works appear in different research themes: graph-theoretical analysis [15, 5], Petri nets models [32, 10], flux balance analysis [17, 23], metabolic pathway analysis [30, 29, 22, 24]...

Concerning metabolic pathway analysis, computing feasible routes through the network is one of the methods addressing large networks. Several modeling methods can be used to do it such as Petri net network, Extreme Pathways or Elementary Flux Modes (EFMs) computation. In this work we have focused on EFMs which use the stoichiometric matrix to find set of reactions which can interact together respecting steady state assumption. One of the well-known problem about EFMs computation is the size of the results. Recently, an additional work has been derived from EFMs to give a dual view of them and to provide a smaller solution. It is the computation of Minimal Cut Sets (MCSs) [19]. We have tested these two methods and compared their results for 6 different networks which exhibit different size and level of complexity modeled. In this article after a rapid presentation of these two methods and a description of the networks, we will show that when the network size increases some differences can be observed between EFMs and MCSs results.

2 Analyze the reliability of the metabolites production

Elementary Flux Modes (EFMs) analysis has been introduced by Schuster [31] to analyze metabolic pathways. It is a constraint-based approach which can be used to calculate all biologically meaningful pathways through a network [30]. This method is useful to gain an insight into metabolism of living organisms and to identify all genetically independent pathways that are inherent in a metabolic network (see [35] for more complete explanations). An EFM is a unique and non-decomposable set of reactions. From the stoichiometric matrix, EFMs are computed by selecting groups of reactions which interact together and respecting the well-known steady-state mass balancing equation as follows:

$$\frac{dS}{dt} = Nv \quad (1)$$

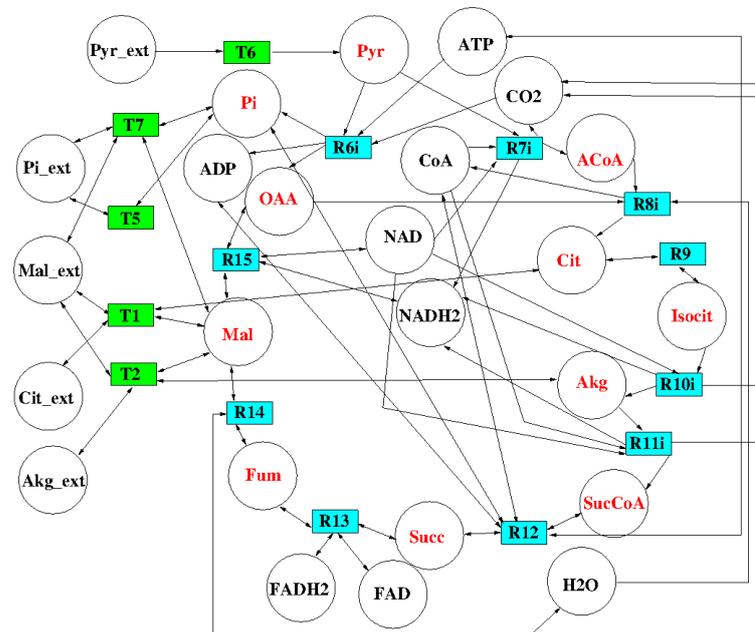


Figure 1: TCA cycle network. The metabolites are the circles: the red labels represent the internal metabolites, the other ones (black colors) represent the external metabolites. The rectangles are the reactions: the shapes filled by cyan are the transporters. The arrows denote the direction of reactions, a double arrow means that the reaction is reversible.

S is a vector of concentration values, N is the $m \times r$ stoichiometric matrix made up of m metabolites and r reactions, and v is the r -dimensional (flux) vector of the reaction rates. At the steady state, Nv has to be equal 0. Grafahrend-Beleau et al. [13] have shown that computing the set of EFMs of a given network is equivalent to compute the set of t -invariants of the network modeled through a Petri net.

In the small example network of TCA cycle given in Figure 1 (for details see [26]), we can see 15 reactions and 25 metabolites (11 internal and 14 external ones). Applying EFMs computation, 16 EFMs have been found. To analyze this result, for example, we can consider the case of production of external citrate obtained by firing the transporter reaction $T1$. Figure 2 shows the 7 EFMs/available routes to fire $T1$ and if any of them cannot be operated (by the inhibition of one reaction belonging to the EFMs) there is no more way to activate $T1$.

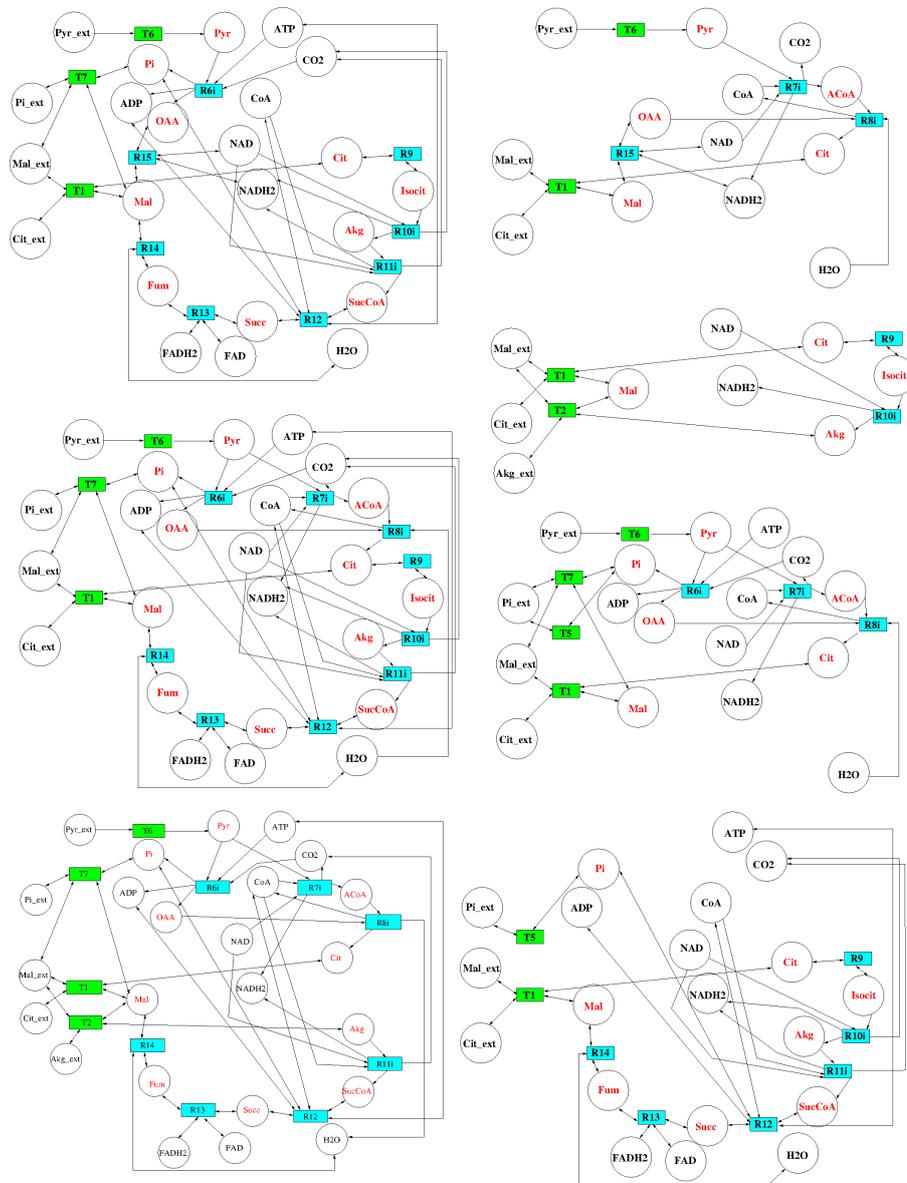


Figure 2: 7 EFMs containing T1 to produce external citrate.

A number of tools were specifically designed to compute all EFMs of a given metabolic network: METATOOL [28], CellNetAnalyzer [20] (also known as a successor to METATOOL), efmtool [34] and regEfmtool [11] (re-using efm-tool and adding logical rules). These tools implement an algorithm based

on linear algebra and its complexity is exponential, especially for metabolic networks including many connected pathways [21]. As the number of obtained EFMs can be huge, enumerating all possible pathways that contain a given reaction is a difficult task [1]. The question of “*how to suit biological reasoning to such large results*” stays open. Classification of EFMs can be done by clustering methods. Due to their specificity: each EFMs is unique and minimal, classical hierarchical clustering does not offer satisfying results. Overlapping clustering seems to be more promising for this task. A classification method for EFMs, ACoM [27], has been proposed, based on motif findings with overlapping clustering tools. In [13] a classification of t-invariants is also studied using another agglomerative clustering algorithms. But most often the size of the results to classify still remains a major difficulty. Going back to classical graph theory methods, another way to extract knowledge from networks has been explored: computing graph diameter, average degree of nodes, average path length... and authors such as Barabási [6], Fell [12] or Jeong [15] have confirmed that metabolic networks can exhibit behaviors similar to small world networks and can be explored that way to find organization, links or hubs through metabolic networks [9]. More recent works have suggested using a dual view on the problem of finding feasible routes and of searching ways to cut access to a specific reaction and then to inhibit it. The next sections will present this method.

3 Fragility of metabolic networks analysis

A new algorithm to find Minimal Cut Sets (MCS) through metabolic networks has been proposed by Klamt and Gilles [19]. Whereas EFMs can identify all possible and feasible metabolic routes for a given network at steady state, MCSs computing allows finding a set of reactions able to eliminate a given objective functioning. An MCS is a unique and the smallest set of reactions whose removal from the network would stop a given metabolic function. As the Klamt algorithm allows to compute all MCS of a given metabolic network, EFMs and MCSs complement each other in a duality based relationship [18, 4].

MCSs concept is closed to the problem of finding the minimum cuts in a graph, that is a well-known approach to characterize the structure of networks in graph theory [3, 7]. The main idea can be expressed like that:

- if a few of links are cut or otherwise fail, the network may still be able to transmit messages between any pair of its nodes,
- if enough links fail, there will be at least one pair of nodes that cannot communicate with each other.

Thus an important measure of the reliability of a network is the minimum number of links that must fail to disconnect the graph. In that way, MCSs can be used for studying the fragility of a network structure and identifying suitable targets to stop metabolic functionality. If we go back to the TCA cycle network, MCSs computation gives 54 MCSs and 27 of them contain $T1$. One can be disappointed by this result because the number of MCSs is not always smaller than EFMs and does not seem to be a good solution for the problems which their size of results are huge. In a first and rapid conclusion, the goal claimed by [14] to use MCSs results to minimize the needed effort to understand the structure of complex networks seems to be far away. But it is noted that the TCA cycle contains only one function/pathway and perhaps is not a good example for the purpose of testing automatic analysis of complex networks. Automatic tools address large networks composed of several pathways and it is why we now present results obtained with 5 other networks describing cell energy metabolism or plant central carbon metabolism.

4 Description of the metabolic networks

To compare the results provided by the computation of EFMs and MCSs, we have used 3 networks modeling energetic metabolism of mitochondria in 3 different tissues: muscle, liver, yeast, and 2 networks modeling different versions of the central metabolism of plant cells. One, called network plant cell A (PCA), is the smallest network consisting of 70 metabolites and 78 reactions described in more details in [8] and depicted in Figure 3. The second one, called PCC, has more reactions and metabolites. The additional reactions describe the amino acids synthesis towards external metabolites.

Energetic metabolism of mitochondria is often described as a set of five main pathways: oxidation of phosphorylants, TCA cycle (also called Krebs cycle), ketone bodies, beta-oxidation, and a part of ornithine cycle. Depending on the tissues, some variations can be observed. The 3 retained models concern muscle and liver (*Homo sapiens*) and yeast (*S. cerevisiae*). Both mitochondrion of muscle and yeast do not contain an urea cycle. Mitochondria of yeast does not include beta-oxidation as well as production/consumption inside ketone bodies and three specific reactions have been added for TCA cycle and five transport reactions linked to them [25].

Next, we look at the metabolic network of heterotrophic plant cells (details can be found in [8]). Figure 3 shows the network including the main pathways of the central carbon metabolism in plant cells: glycolysis (black color), the TCA cycle (blue color), the pentose phosphate pathway (pink color), the starch

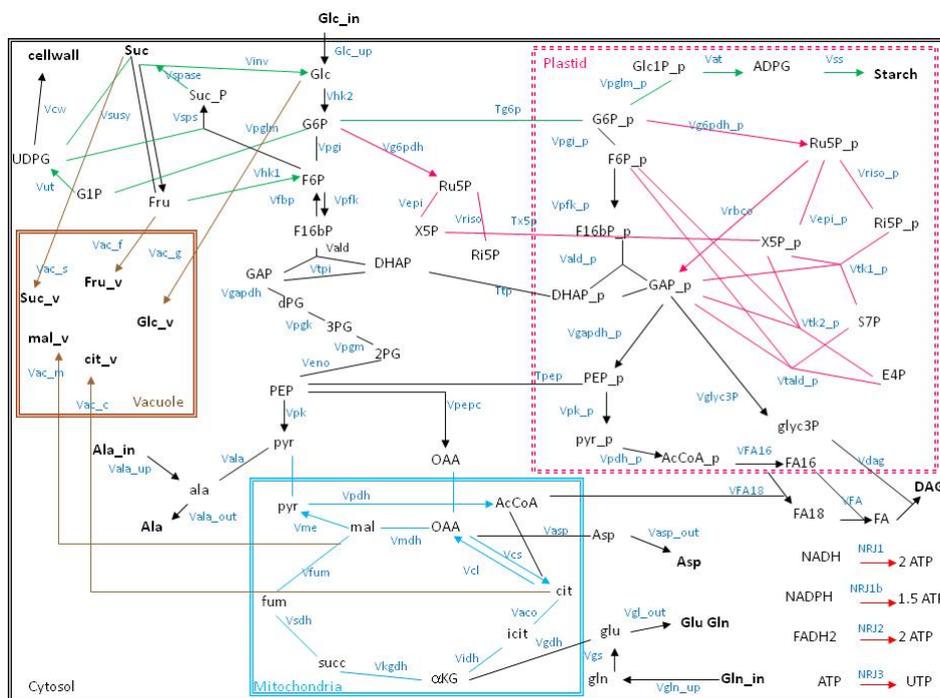


Figure 3: Enlarged metabolic network of a heterotrophic plant cells. Each color indicates one pathway: blue for the TCA cycle, black for glycolysis and also for the fluxes towards output metabolites, pink for the PPP, green for the sucrose and starch synthesis, red for respiration and brown for storage in vacuole. External metabolites are in bold. Irreversible reactions are indicated by unidirectional arrows.

and sucrose pathways (green color) and the storage reactions towards the vacuole (brown color). Due to its autotrophic nature, the plant synthesizes its own respiratory substrates (mainly carbohydrates) which then serve as substrates for the TCA cycle. The TCA cycle provides precursors for several biosynthetic processes, such as nitrogen fixation and biosynthesis of amino acids. The pentose phosphate pathway includes the irreversible oxidative branch, whereas the non-oxidative branch is reversible (recycling of pentose-phosphates from fructose phosphate and triose-phosphate). In the starch and sucrose pathways, sucrose is metabolized in cytosol, whereas starch is metabolized in plastids from imported hexose phosphates (G1P or G6P). Several effluxes are illustrated: protein synthesis from several amino acids (glutamate and glutamine, aspartate and alanine), lipid synthesis (diacyl glycerol) from plastidial pyru-

vate and trioses, synthesis of cell wall polysaccharides from UDP-glucose, sugars (glucose, fructose and sucrose) and storage of organic acids (malate and citrate) in vacuoles. Subcellular compartments, such as mitochondria and plastids, can lead to potentially reversible transport of metabolites such as G6P, X5P, PEP and DHAP.

In the first version (PCA), there are 70 different metabolites. Among them, 15 metabolites are external, meaning that they are carbon sources or carbon sinks (nutrients, waste products, stored and excreted products, and precursors for further transformation). These are exogenous glucose and amino acids (glutamine and alanine), CO₂, sugars (sucrose, glucose and fructose) and organic acids (citrate and malate) stored in vacuoles, amino acids for protein synthesis (aspartate, alanine, glutamate and glutamine), cell wall polysaccharides, starch and lipids. The metabolites named cofactors (ATP, NADH, NADPH, and FADH₂) can be internal which means that they are expected to be balanceable at steady state. The PCC version has 89 reactions and 74 metabolites, 24 metabolites are external. The direction (reversible or irreversible) of the reactions is derived from thermodynamic properties. The complete list of reactions in METATOOL format is supplied at:

<http://dept-info.labri.u-bordeaux.fr/~beurton/MetabolicNetworks>

5 *Elementary Flux Modes computation*

To compute EFMs, the original METATOOL software (module 5.0.3 beta in CellNetAnalyzer¹) has been chosen, a program with a graphical user interface for the analysis of metabolic networks running under Matlab (Mathworks, Inc.). Table 1 shows the different number of EFMs obtained for six networks. For these six networks, the number of EFMs could be considered to be very large and it grows exponentially with the number of reactions as it is expected. We can see that for the first simple network (TCA cycle) the number of EFMs is in the same range than the number of reactions. The number of reactions in the mitochondria networks is double than those of TCA cycle and the number of EFMs is multiplied by hundreds. At least, the number of reactions in the plant cell networks is doubled again and the number of EFMs is multiplied by thousands. After this global remark we can take a look more precisely different cases. Into the muscle and liver mitochondria networks, one can see that the smaller number of reactions leads to the greater number of EFMs. This suggests that the structures of these two networks are different; even if they share a large part of the reactions. Concerning the two plant cell networks, one can note that the number of EFMs becomes very big. The difference of 11

¹<http://www.mpi-magdeburg.mpg.de/projects/cna/cna.html>

reactions between PCA and PCC leads to a number of EFMs 10^2 times bigger for PCC. It means that these differences create many new routes through the network.

Networks	Nb. React	Nb. Int. Meta	Nb. EFMs
TCA cycle	16	13	16
Muscle	37	31	3,253
Liver	44	36	2,307
Yeast	40	34	4,627
PCA	78	55	114,614
PCC	89	50	9,319,997

Table 1: Global description of six networks. Here, the abbreviations are explained: Nb. React = Number of Reactions; Nb.Int Meta = Number of Internal Metabolites; Nb. EFMs = Number of Elementary Flux Modes

6 Finding all Minimal Cut Sets

The next step is to compute MCSs in order to verify whether the hypothesis of Klamt about the size of MCSs result could work in our cases. We have computed all MCSs for all possible objective functions in the networks so that we can compare these results to these ones of EFMs which computes all possible routes through the networks.

Table 2 summarizes the obtained results of six networks. In the case of the TCA cycle, the number of MCSs is three times bigger than the number of EFMs. As a consequence, computing the number of feasible pathways could be less complex than computing the different ways used to cut the network. The three mitochondria networks exhibit the same kinds of results. The number of MCSs is at least ten times bigger the number of EFMs. However, note that there is a difference between liver and yeast results: the number of MCSs for yeast is much bigger than liver. Therefore, EFMs and MCSs probably do not capture the same information about how the networks are connected. For two plant cell networks, we have obtained less MCSs than EFMs. Beside, surprisingly, while PCA has double number of reactions comparing to yeast and the EFMs number in PCA is thirty times bigger than the EFMs number in yeast, the MCSs number in both PCA and yeast are not very different. Thus, sometimes a big network could be less difficult to analyze than another smaller one.

In addition to comparing the number of EFMs and MCSs, we have taken a look on their length. The length is an important parameter to run procedures

Networks	Nb. EFMs	Nb. MCSs	EFMS Length	MCSs Length
TCA cycle	16	54	8.3 (4/12)	3.8 (3/4)
Muscle	3,253	42,534	17.7 (2/23)	10.2 (6/12)
Liver	2,307	47,203	16.7 (2/24)	11.4 (6/14)
Yeast	4,637	90,318	15.3 (2/22)	11.6 (7/13)
PCA	114,614	93,009	37.7 (2/53)	11.2 (4/18)
PCC	9,319,997	2,815,375	33.1 (1/48)	11.8 (4/18)

Table 2: Global comparison between the number of MCSs and EFMs.

The fourth and fifth columns denote the EFMs and MCSs average lengths, respectively. The minimum and maximum length of EFMs and MCSs respectively are put in the parentheses.

used to post-treat the results, for example: finding common motifs and classification. The length of an MCS is the minimal number of reactions whose removal obstructs a given objective reaction. On other words, the length of an MCS is defined by the number of reactions constituting that MCS. The length of an EFM is the number of reactions belonging to the route. The last two columns in Table 2 show the obtained results for six networks. One can note that the distribution of lengths is totally different. Whereas the length of EFMs grows with the size of networks, the MCSs length remains stable (the TCA network could be considered independently because of its very small size).

Figure 4 shows the distribution of MCSs lengths for 5 networks. The TCA network has not been included in this comparison because its number of reactions and the number of MCSs are too small to be considered. Even five distributions seem to be heterogeneous, it is shown that a big number of MCSs does not imply a large range of values. Intuitively, most MCSs are not so long. Remarkably, the yeast network exhibits a different profile with a lot of MCSs (more than 50%) close to the maximum length.

Due to software simulation limitations (e.g., segmentation fault or more than ten days to obtain the results), we have not been able to compute EFMs and MCSs of PCA and PCC with CNA. The new application, regEFMtool written by [16] based on efmtol developed by [34], provided us an efficient tool for making a computation of EFMs. The computation of PCA was obtained after less than one hour with regEFMtool. Moreover, this program allows us to consider a specific objective reaction to inhibit and so to compute only the set of EFMs concerning this reaction. The available algorithms for MCSs computation through CNA or efmtol require the EFMs to be computed first and reuse the EFMs matrix to compute MCSs. Future works will be focused on the study of small MCSs because they defined small, i.e easy to analyze,

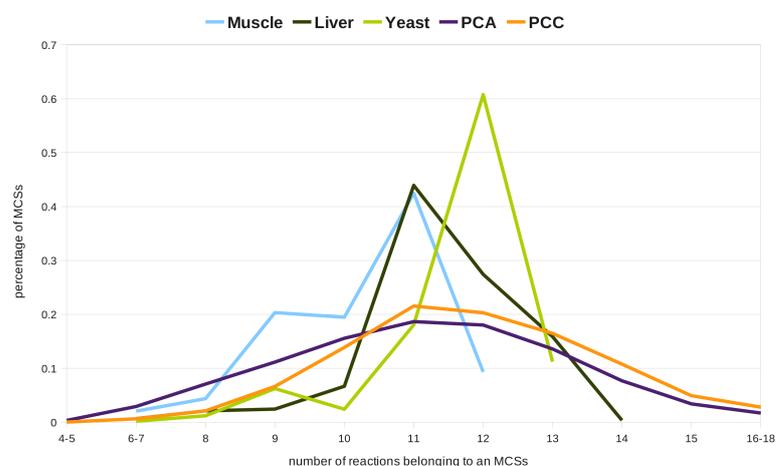


Figure 4: Histogram of the MCSs length.

sets of reactions which are able to stop productions of metabolites. If we consider these sets, we can defined patterns of reactions which are critical for the network. By definition of MCSs if all reactions belonging to one MCS are unavailable, it is not possible to find feasible routes through the network. Moreover, if two distant reactions, i.e. not directly connected by the same shared metabolite, belong to the same MCSs that signifies that they define a kind of dependency between them, because at least one of them is required.

7 Conclusion

Nowadays, a large amount of data is available through big international databases to describe metabolic networks at the cellular level. However, these metabolic networks containing several connected pathways are complex to analyze. From the last decade, a lot of methods have been developed to support in finding automatically characteristics of such networks, for instance, finding set of reactions interacting together to guarantee a certain biological function. The EFMs analysis is one of these methods and it is actually implemented in some softwares. Unfortunately, the size of the results grows exponentially with the network size. Recently, a dual view of EFMs, called MCSs, used to compute a set of reactions which is able to disconnect/cut a given metabolic functionality of the network, has been studied. The results of this new method are expected to be smaller than those of EFMs. We have tested this hypothesis with 6 different networks from a very simple one containing only one pathway

(i.e., the TCA cycle) to more complex ones like central metabolism of plant cells. Our experiments showed that in the cases of the TCA cycle and the 3 networks of mitochondria, the hypothesis is not confirmed because the number of the obtained MCSs is greater than the number of EFMs. But with the larger networks, which have more than 70 reactions, the number of MCSs seems not to grow exponentially. It is worth to note that both EFMs and MCSs results are impossible to analyze manually and they require automatic procedures to do this task. Beside, classification and finding common motifs are common way to achieve this analysis but these methods are influenced by a few of parameters such as the size of the classified elements difficult to fix. One can note that if the average length of the EFMs (i.e. the number of reactions belonging to one EFM) depends on the network size, the MCSs average size does not. Regarding on this information we have take a look at small size MCSs and in future works we will study how these small cut sets can be used to identify reaction patterns through the networks.

Acknowledgements

This work is supported by a grant from Vietnamese Government. The authors thank Le Xuan Quynh and Preston for their useful comments of the manuscript.

References

- [1] V. Acuna, F. Chierichetti, V. Lacroix, A. Marchetti-Spaccamela, M.-F. Sagot, and L. Stougi. Modes and cuts in metabolic networks: Complexity and algorithms. *Biosystems*, 95(1):51 – 60, 2009.
- [2] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:47–97, June 2001.
- [3] H. Ariyoshi. Cut-set graph and systematic generation of separating sets. *IEEE Transactions on Circuit Theory*, 19(3):233 – 240, may 1972.
- [4] K. Ballerstein, A. von Kamp, S. Klamt, and U.-U. Haus. Minimal cut sets in a metabolic network are elementary modes in a dual network. *Bioinformatics*, 28(3):381–387, Feb. 2012.
- [5] A.-L. Barabási and E. Bonabeau. Scale-free networks. *Scientific American*, 288:60–69, May 2003.
- [6] A.-L. Barabási and Z. N. Oltvai. Network biology: understanding the cell’s functional organization. *Nature Reviews Genetics*, 5(2):101–113, Feb. 2004.

- [7] C. Berge. *Hypergraphs: combinatorics of finite sets*. Elsevier Science, 1987.
- [8] M. Beurton-Aimar, B. Beauvoit, A. Monier, F. Vallée, M. Dieuaide-Noubhani, and S. Colombié. Comparison between elementary flux modes analysis and ¹³C-metabolic fluxes measured in bacterial and plant cells. *BMC Systems Biology*, 5(95), 2011.
- [9] M. Beurton-Aimar, N. Parisey, F. Vallée, and S. Colombié. Identification of functional hubs through metabolic networks. In *ALifeXII, 12th International Conference on the Synthesis and Simulation of Living Systems*. Odense, 19-23 august.
- [10] C. C. Petri net modelling of biological networks. *Brief Bioinform*, Jul 2007.
- [11] D. E. R. C. Jungreuthmayer and J. Zanghellini. regefmtree: Speeding up elementary flux mode calculation using transcriptional regulatory rules in the form of three-state logic. *Biosystems*, 113:37–39, 2013.
- [12] D. A. Fell and A. Wagner. The small world of metabolism. *Nature Biotechnology*, 18:1121–1122, November 2000.
- [13] E. Grafahrend-Belau, F. Schreiber, M. Heiner, A. Sackmann, B. Junker, S. Grunwald, A. Speer, K. Winder, and I. Koch. Modularization of biochemical networks based on classification of petri net t-invariants. *BMC Bioinformatics*, 9(90):doi:10.1186/1471-2105-9-90, 2008.
- [14] O. Hädicke and S. Klamt. Computing complex metabolic intervention strategies using constrained minimal cut sets. *Metabolic Engineering*, 13:204–213, 2011.
- [15] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabasi. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, October 2000.
- [16] C. Jungreuthmayer, D. E. Ruckerbauer, and J. Zanghellini. Utilizing gene regulatory information to speed up the calculation of elementary flux modes. *ArXiv e-prints*, Aug. 2012.
- [17] K. Kauffman, P. Prakash, and J. Edwards. Advances in flux balance analysis. *Curr Opin Biotechnol*, 14(5):491–496, 2003.
- [18] S. Klamt. Generalized concept of minimal cut sets in biochemical networks. *Bio Systems*, 83(2-3):233–247, Apr 2005.

- [19] S. Klamt and E. D. Gilles. Minimal cut sets in biochemical reaction networks. *Bioinformatics*, 20(2):226–234, 2004.
- [20] S. Klamt, J. Saez-Rodriguez, and E. Gilles. Structural and functional analysis of cellular networks with cellnetanalyzer. *BMC Systems Biology*, 1(1):2, 2007.
- [21] S. Klamt and J. Stelling. Combinatorial complexity of pathway analysis in metabolic networks. *Molecular biology reports*, 29(1-2):233–236, 2002.
- [22] S. Klamt and J. Stelling. Two approaches for metabolic pathway analysis? *Trends in Biotechnology*, 21(2):64 – 69, 2003.
- [23] J. D. Orth, I. Thiele, and B. O. Palsson. What is flux balance analysis? *Nat. Biotechnol*, 28(3):245–248, Mar 2010.
- [24] J. A. Papin, J. Stelling, N. D. Price, S. Klamt, S. Schuster, and B. O. Palsson. Comparison of network-based pathway analysis methods. *Trends in Biotechnology*, 22(8):400 – 405, 2004.
- [25] S. Pérès, M. Beurton-Aimar, and J. Mazat. Analysis of large set of elementary modes: application to energetic mitochondrial metabolism. In *European Conference on Complex Systems*, 2005.
- [26] S. Pérès, M. Beurton-Aimar, and J. Mazat. Pathway classification of tca cycle. *Systems Biology, IEE Proceedings*, 153(5):369–371, september 2006.
- [27] S. Pérès, F. Vallée, M. Beurton-Aimar, and J. Mazat. Acom: a classification method for elementary flux modes based on motif finding. *Biosystems*, 103(3):410–419, 2011.
- [28] T. Pfeiffer, I. S. Valdenbro, J. C. Nuño, F. Montero, and S. Schuster. Metatool: for studying metabolic networks. *Bioinformatics*, 15(3):251–257, 1999.
- [29] N. D. Price, J. A. Papin, and B. O. Palsson. Determination of redundancy and systems properties of the metabolic network of helicobacter pylori using genome-scale extreme pathway analysis. *Genome Research*, 12(5):760–769, 2002.
- [30] S. Schuster, D. A. Fell, and T. Dandekar. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnology*, 18(3):326–332, 2000.

- [31] S. Schuster and C. Hilgetag. On elementary flux modes in biochemical reaction systems at steady state. *Journal of Biological Systems*, 2(2):165–182, 1994.
- [32] R. Srivastava, M. S. Peterson, and W. E. Bentley. Stochastic kinetic analysis of the Escherichia coli stress circuit using sigma(32)-targeted antisense. *Biotechnol Bioeng*, 75(1):120–129, Oct 2001.
- [33] J. Stelling, S. Klamt, K. Bettenbrock, S. Schuster, and E. D. Gilles. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420(6912):190–193, 2002.
- [34] M. Terzer and J. Stelling. Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics*, 24(19):2229–2235, 2008.
- [35] C. Trinh, A. Wlaschin, and F. Sreenc. Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. *Applied Microbiology and Biotechnology*, 81(5):813–826, January 2009.
- [36] A. Wagner and D. A. Fell. The small world inside large metabolic networks. In *Proceedings of the Conference of The Royal Society in London B*, volume 268, pages 1803–1810, April 2001.
- [37] X. Zhu, M. Gerstein, and M. Snyder. Getting connected: analysis and principles of biological networks. *Genes Dev*, 21(9):1010–24, 2007.



The Benefits of Resource Partitioning and Division of Labor in Microbial Consortia

Diana Schepens¹, Ashley E. Beck^{2,4}, Jeffrey J. Heys³, Tomas Gedeon¹ and Ross P. Carlson^{3,4}

¹ Department of Mathematical Sciences, Montana State University-Bozeman, Montana, 59717 USA

² Department of Microbiology and Immunology, Montana State University-Bozeman, Montana, 59717 USA

³ Department of Chemical and Biological Engineering, Montana State University-Bozeman, Montana, 59717 USA

⁴ Center for Biofilm Engineering, Montana State University-Bozeman, Montana, 59717 USA

Abstract

Microbial communities composed of different species are frequently observed in nature. There is growing interest in understanding the advantages of living in a mixed species community as compared to a monoculture. Understanding the effects of microbial interactions will lead to better control of environmental microbial systems responsible for nutrient cycling, medical systems like the human GI tract or chronic wounds, and applied systems like biofuel synthesis. We examine two different advantages associated with interacting microbial communities operating as a food chain. The first advantage is the minimization of anabolic resource investment required to drive a constant flux through a series of enzyme-catalyzed reactions. The second advantage is enhanced productivity of a biofilm community when inhibitory byproducts are consumed by a scavenger population.

1 Introduction

Microorganisms acquire resources from their environment, which are processed via series of enzymes into cellular energy and biomass, and metabolic byproducts are secreted. Species that are efficient at deriving cellular energy and biomass from available resources typically grow and reproduce faster than less efficient species [11]. This simplified view suggests communities should consist of a monoculture of the most 'efficient' species, sometimes termed a 'superorganism'. Superorganisms are not observed in nature. Instead, complex communities of interacting microorganisms are observed [6]. Potential advantages of mixed communities relative to monocultures include efficiency gains

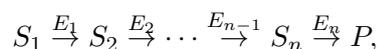
from specialization, the development of advanced functions, the ability to better tolerate environmental fluctuations, the utilization of inhibitory byproducts by other community members, and lower susceptibility to a single predatory attack, among many possible advantages.

The focus here is on food chains within microbial communities, specifically, systems where one population consumes metabolites excreted by others. These types of interactions are frequently referred to as syntrophic interactions [13]. Food chains are commonly observed in natural and applied systems including the cycling of carbon during the degradation of lignocellulosic material and the cycling of nitrogen in waste water treatment facilities [10]. Additionally, food chains have been observed to evolve repeatedly in the laboratory environment. For example, chemostats inoculated with a single *Escherichia coli* strain and operated under carbon-limited conditions will evolve spontaneously into a community of cross-feeding sub-strains. The evolved communities consist of one strain specializing in glucose catabolism and other strains specializing in catabolism of secreted inhibitory byproducts like acetate. The overall system demonstrates a 15% improvement in biomass production from the limiting resource glucose [9].

While there are many potential benefits for microorganisms existing in a mixed community, the focus here is on a theoretical analysis of two possible benefits associated with resource exchange: enhanced return on limiting anabolic resource and lowered byproduct inhibition. The first benefit is quantified by solving a constrained optimization model of a series of reactions controlled by resource investment into different substrate and enzyme pools, and the second benefit of reduced byproduct accumulation is quantified using individual-based models of a biofilm community.

2 First Benefit: Enhanced Return on Anabolic Resource Investment

Cellular flux is driven by combinations of resource investments into both substrate and enzyme pools. Consider an n reaction enzymatic pathway



where the substrate S_1 is transformed into product P with the help of n enzymes, E_1, \dots, E_n . In a limited resource environment, microorganisms that maximize the functional return on resource investment will be competitive [2]. To examine the trade-off between investment into combinations of substrate and enzyme pools, we assume that the cell can partition an essential resource (e.g., carbon) into optimal combinations of substrate pools utilized in the enzymatic pathway and into building the enzymes E_1, \dots, E_n . Both substrate

pools and enzymes represent resource investments but with different resource liquidities. If substrate S_i consists of b_i carbon atoms and enzyme E_i consists of a_i carbon atoms, then the total carbon investment for the enzymatic pathway is:

$$C := \sum_{i=1}^n ([E_i]a_i + [S_i]b_i).$$

Our model of the enzymatic pathway takes the form of a set of differential equations:

$$\dot{S}_1 = k_{a_1}S_0 - \frac{V_1[S_1]}{\kappa_1 + [S_1]}, \quad \dot{S}_2 = \frac{V_1[S_1]}{\kappa_1 + [S_1]} - \frac{V_2[S_2]}{\kappa_2 + [S_2]}, \quad \dot{P} = \frac{V_n[S_n]}{\kappa_n + [S_n]} - k_{a_2}[P],$$

where $k_{a_1}S_0$ is a constant source term and $[P]$ is the concentration of the final product. Here $V_i = [E_i]k_i$ and we assume a one-to-one stoichiometric relationship between S_i and S_{i+1} as catalyzed by E_i . Assuming the pathway is at steady state, we obtain

$$k_{a_1}S_0 = \frac{V_1[S_1]}{\kappa_1 + [S_1]} = \frac{V_2[S_2]}{\kappa_2 + [S_2]} = \dots = \frac{V_n[S_n]}{\kappa_n + [S_n]} = k_{a_2}[P]$$

and thus $k_{a_1}S_0$ is the steady state flux through the pathway. The concentration of the final product $[P]$ is proportional to $k_{a_1}S_0$.

We then formulate the pathway investment minimization problem as the strategy of minimizing the amount of anabolic substrate for a given flux:

$$\begin{aligned} \min_{\{[E_i], [S_i], i=1, \dots, n\}} C([E_i], [S_i]) \quad & \text{subject to } n \text{ constraints} \quad (1) \\ \frac{k_i[E_i][S_i]}{\kappa_i + [S_i]} &= k_{a_1}S_0 \text{ for } i = 1, \dots, n. \end{aligned}$$

The optimization is over substrate S_i and enzyme E_i concentrations. Solving (1) using Lagrange multipliers results in the following system of equations:

$$a_i = \lambda_i \frac{k_i[S_i]}{\kappa_i + [S_i]}, \quad b_i = \lambda_i \frac{k_i \kappa_i [E_i]}{(\kappa_i + [S_i])^2}, \quad k_{a_1}S_0 = \frac{k_i[E_i][S_i]}{\kappa_i + [S_i]}.$$

Solving the system of equations gives:

$$\begin{aligned} [S_i] &= \sqrt{k_{a_1}S_0 \frac{\kappa_i a_i}{k_i b_i}}, \\ [E_i] &= \frac{b_i}{a_i \kappa_i} [S_i] (\kappa_i + [S_i]) = \sqrt{k_{a_1}S_0 \frac{b_i \kappa_i}{k_i a_i}} + \frac{k_{a_1}S_0}{k_i}. \end{aligned}$$

Therefore the minimum anabolic investment to achieve a flux $k_{a_1}S_0$ through the enzymatic pathway is:

$$C = \sum_{i=1}^n ([E_i]a_i + [S_i]b_i) = \sum_{i=1}^n \left(\frac{a_i}{k_i} k_{a_1} S_0 + 2\sqrt{\frac{\kappa_i a_i b_i}{k_i} k_{a_1} S_0} \right)$$

The impact of different flux levels on the relative total investment requirement is shown in Figure 1. This figure shows that reducing the flux from a baseline value of 1.0 causes an increase in the relative investment cost, but increasing the flux leads to a decrease in the relative investment cost. The key observation we make is that $C(k_{a_1}S_0)$ is an increasing, concave down function of $k_{a_1}S_0$:

$$\frac{\partial C}{\partial(k_{a_1}S_0)} = \sum_{i=1}^n \left[\frac{a_i}{k_i} + \frac{1}{\sqrt{k_{a_1}S_0}} \sqrt{\frac{\kappa_i a_i b_i}{k_i}} \right] > 0,$$

$$\frac{\partial^2 C}{(\partial(k_{a_1}S_0))^2} = -\frac{1}{2} \sum_{i=1}^n \left[\frac{1}{\sqrt{(k_{a_1}S_0)^3}} \sqrt{\frac{\kappa_i a_i b_i}{k_i}} \right] < 0.$$

As a consequence, if one compares the carbon investment, C , needed to achieve a flux $k_{a_1}S_0$ to the investment needed to achieve twice the flux, $2k_{a_1}S_0$, this investment is *less* than $2C$, representing an enhanced functional return on a given anabolic resource investment. From an ecological perspective, the complete oxidation of a substrate at a fixed flux in a single population will require more anabolic resource investment (approximately 10% more for the parameter choices in Figure 1) than in a food chain of two interacting microbial populations that split the same substrate oxidation pathway. This observation assumes efficient metabolite exchange between the two interacting populations.

3 Second Benefit: Inhibitory Byproduct Consumption

All metabolisms have byproducts which can negatively influence chemical thermodynamics, as well as stress microorganisms by creating inhibitory local environments. Common metabolic byproducts that can be inhibitory include organic acids and alcohols such as formic acid, acetic acid, ethanol, pyruvic acid, lactic acid, or glycerol. An existing individual-based model of microbial growth in a biofilm was modified to compare the growth of a monoculture generalist population to a community of cross-feeding populations [14]. Biofilms are aggregates of cells encapsulated within a polymeric matrix of microbial origin. Biofilm communities were studied because product inhibition is especially significant in an environment where metabolite transport is limited to

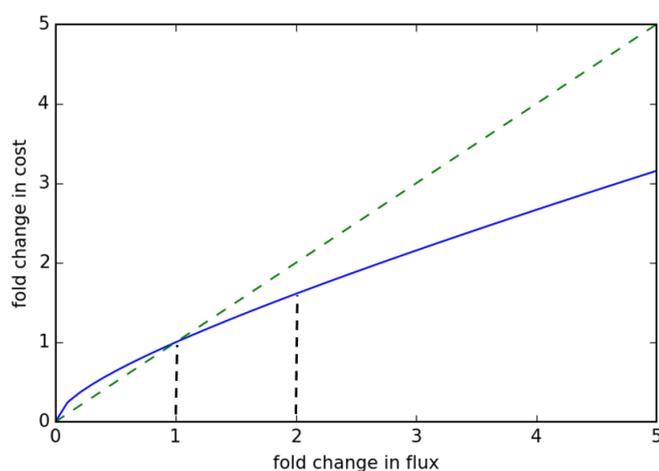


Figure 1: For a fold change in the flux, the fold change in the total anabolic resource investment can be higher for a reduced flux or lower for an increased flux (solid line). The diagonal dashed line (green) shows an equal fold change in resource investment for a desired fold change in flux. The vertical dashed lines (black) highlight the 1- and 2-fold changes in flux.

diffusion; convective transport is negligible in typical biofilms [12]. Communities of different microbial populations are compared in the simulations. The communities are based on three unique populations:

1. a generalist population that consumes sugar and produces an inhibitory byproduct (e.g., acetate) as well as CO_2 ,
2. a producer population that consumes sugar and produces only an inhibitory byproduct (e.g., acetate), and
3. a scavenger population that consumes the inhibitory byproduct as its only substrate, producing CO_2 .

These representative microorganisms and their interactions, which are summarized in Figure 2, are relevant to many natural and engineered systems, including the syntrophic consortium studied by Bernstein et al. [3, 4].

Individual-based models, in which individual cells are explicitly modeled and substrate concentrations are modeled using a continuum approximation, are a natural choice for biofilm community simulations. Numerous individual-based models have been developed, and they often focus on different physical aspects of the biofilm environment, from growth to chemical transport

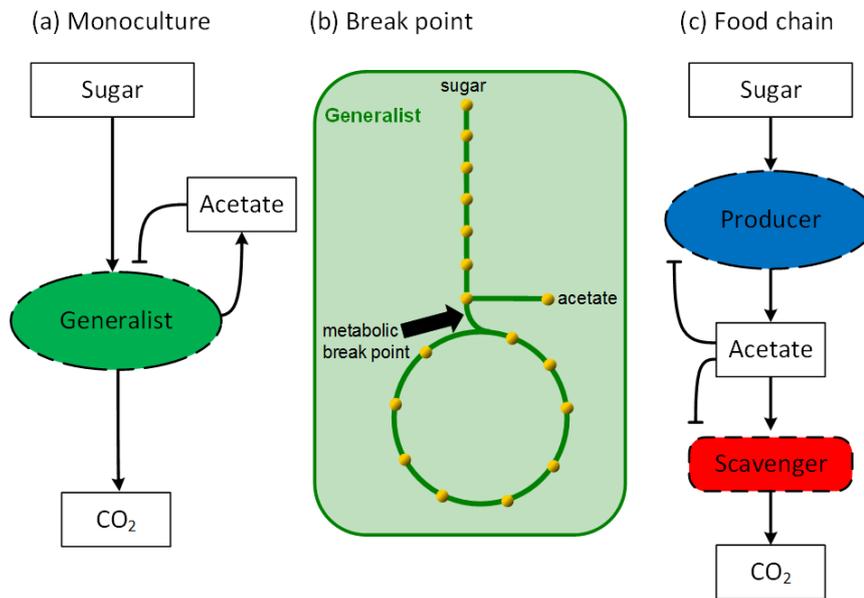


Figure 2: (a) The monoculture contains only the generalist strain, which consumes sugar and produces both an inhibitory byproduct and CO_2 . (b) The break point in the metabolic pathway, which can represent a number of commonly observed byproducts, is shown for division of the pathway between the two specialists. (c) The food chain consists of a producer that consumes sugar and produces the inhibitory byproduct, and a scavenger that consumes (and is inhibited by) the byproduct from the producer and secretes CO_2 .

to detachment [1, 7, 8]. The individual-based biofilm model analyzed here is described in two sections: (1) the diffusible substrate model, and (2) the microbial cell model.

3.1 Diffusible Substrate Model

Substrates considered in the biofilm model include a sugar, which serves as the reduced carbon source for the generalist and producer populations, and a reduced metabolic byproduct such as acetate or lactate, which is the carbon and energy source for the scavenger population. Local substrate, S , concentrations within the biofilm are modeled using the reaction-diffusion equation:

$$\frac{dS}{dt} = \mathcal{D}_S \nabla^2 S + \sum_i \left(\frac{1}{Y_{s,i}} \frac{dX_i}{dt} \right) \quad (2)$$

where \mathcal{D}_S is the diffusivity of substrate S within the biofilm, $Y_{s,i}$ is the yield of microbial cell i consuming or producing substrate S , and $\left(\frac{dX_i}{dt}\right)$ is the specific growth rate of the individual microbial cell i . The diffusivity of sugar is $0.85 \text{ mm}^2/h$ assuming a six-carbon sugar, and the inhibitory byproduct diffusivity is $0.99 \text{ mm}^2/h$ assuming a short chain organic acid. The other parameters in equation 2 are dependent on the specific population phenotype and are defined in the next section. The domain is assumed to have walls (i.e., no-flux boundary conditions) on the bottom and sides, while the top is assumed to have a fixed substrate concentration (1.0 g/L for sugar and 0.0 g/L for inhibitory byproduct). The initial condition is a concentration of zero for all substrates, set in part to capture the initial slow growth phase when the biofilm is initially being established.

3.2 Microbial Cell Model

While substrates are modeled using a continuum approximation, microorganisms are modeled as discrete cells, which allows each cell to potentially have a unique metabolism. For the presented study, three microbial populations are modeled. The growth rates of the generalist population and the producer population are described by:

$$\frac{dX_i}{dt} = \left(\frac{\mu_i G}{K_{G,i} + G} \right) \left(\frac{K_{I,i}}{K_{I,i} + A} \right) X_i$$

where G is the local sugar concentration (g/L), μ_i is the maximum specific growth rate of population i , $K_{I,i}$ is the byproduct inhibition constant for population i , $K_{G,i}$ is the half-saturation constant for population i , and A is the local byproduct concentration (g/L). The effect of byproduct concentration on specific growth rate of the producer and generalist populations is shown in Figure 3a where the byproduct acts as an inhibitor to growth. The specific growth rate of the scavenger population is calculated using:

$$\frac{dX_s}{dt} = \left(\frac{\mu_s A}{K_A + A + \frac{A^2}{K_{IA}}} \right) X_s$$

where K_A is the half-saturation constant and K_{IA} is the byproduct inhibition constant. The effect of byproduct concentration on the growth rate of the scavenger population is shown in Figure 3b where the byproduct is both the sole energy source and inhibitor. The parameter values used in the cell growth equations are summarized in Table 1.

The individual biofilm model algorithm is implemented using the Python programming language and is publicly available on github [15]. The algorithm begins with a setup phase where the initial cell distributions and initial

Param.	Value	Param.	Value	Param.	Value
μ_g	0.4 h^{-1}	μ_p	0.36 h^{-1}	μ_s	0.4 h^{-1}
$K_{G,g}$	0.05 g/L	$K_{G,p}$	0.05 g/L	K_A	0.005 g/L
$K_{I,g}$	0.04 g/L	$K_{I,p}$	0.15 g/L	K_{IA}	0.5 g/L
$Y_{G,g}$	0.15 gX/gG	$Y_{G,p}$	0.14 gX/gG		
$Y_{A,g}$	0.25 gX/gA	$Y_{A,p}$	0.22 gX/gA	$Y_{A,s}$	0.3 gX/gA

Table 1: Parameters used in the individual-based biofilm model where g is the generalist population, p is the producer population, and s is the scavenger population. See [5] and references therein for source information.

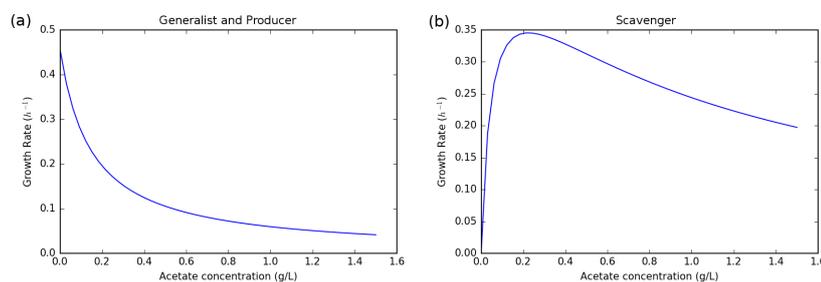


Figure 3: (a) The effects of byproduct inhibition on the specific growth rate of the generalist and producer populations at a fixed sugar concentration of 1 g/L , and (b) the effects of byproduct concentration on the scavenger population where the byproduct of the producer population is both the sole energy source and an inhibitor.

substrate concentrations are set. Then, a time loop is initiated and each time step begins with a calculation of the growth of each cell based on the current substrate concentration at the individual cell location. The mass of each cell is increased based on the growth rate and time step. Next, the reaction-diffusion equation is solved to determine updated substrate concentrations. Finally, any cells that have increased in mass beyond a threshold (twice their initial mass for the results shown here) are divided into two cells with one occupying the original location and the other being randomly placed within a cell diameter of the parent microorganism. If any cells are overlapping, i.e., occupying the same space, an iterative smoothing process is used to spread out the cells until no overlap exists.

3.3 Individual-Based Biofilm Model Results

The impact of resource partitioning is illustrated via comparison of three different microbial communities. Initial conditions consisted of 36 cells for all

System	<i>g</i> count	<i>p</i> count	<i>s</i> count	total count
1	389	0	0	389
2	0	288	147	435
3	165	192	109	466

Table 2: Cell counts after simulating 10 hours of growth for an initial community of 36 cells (evenly divided among the relevant three populations). *g* is the generalist population, *p* is the producer population, and *s* is the scavenger population. The values represent the average from 8 independent simulations.

communities; when interacting communities were analyzed, the 36 cells were divided equally between population types, e.g., for the two-population communities there were initially 18 cells from each population. The first, and simplest, community consisted of only the generalist population. The second community consisted of the producer and scavenger populations, and the third community combined generalist, producer, and scavenger populations.

Table 2 summarizes the final population cell counts after simulating 10 hours of growth for all three communities. In all cases, the initial community was comprised of 36 total cells that were randomly placed within two cell diameters of the bottom of the domain. The presented population cell counts are averages of 8 different simulations. The standard deviation of the 8 simulations based on different random starting locations was 5-10 cell counts. It is also important to emphasize that in the mathematical model, starting location is the only source of randomness.

Figure 4a shows a representative cell distribution for a monoculture community of generalists (green). The cells are drawn larger than their actual size to facilitate visualization. Cell clustering was a result of random initial cell seeding. Figure 4a also shows contours of the byproduct concentration, an inhibitor for this population. The highest byproduct concentrations are found in the center of large cell clusters and inhibit cell growth. This is the least productive community, having the smallest final cell count.

The second community consisted of producer and scavenger populations. The initial community consisted of 18 cells from each population, and an example of a final cell distribution is shown in Figure 4b. The producer population (shown as blue) generated more inhibitory byproduct than the generalist population; however, the highest byproduct concentrations were comparable to the generalist monoculture due to the function of the scavenger population (red). The scavenger population grew fastest in regions of high byproduct concentration. The average byproduct concentration in this community was

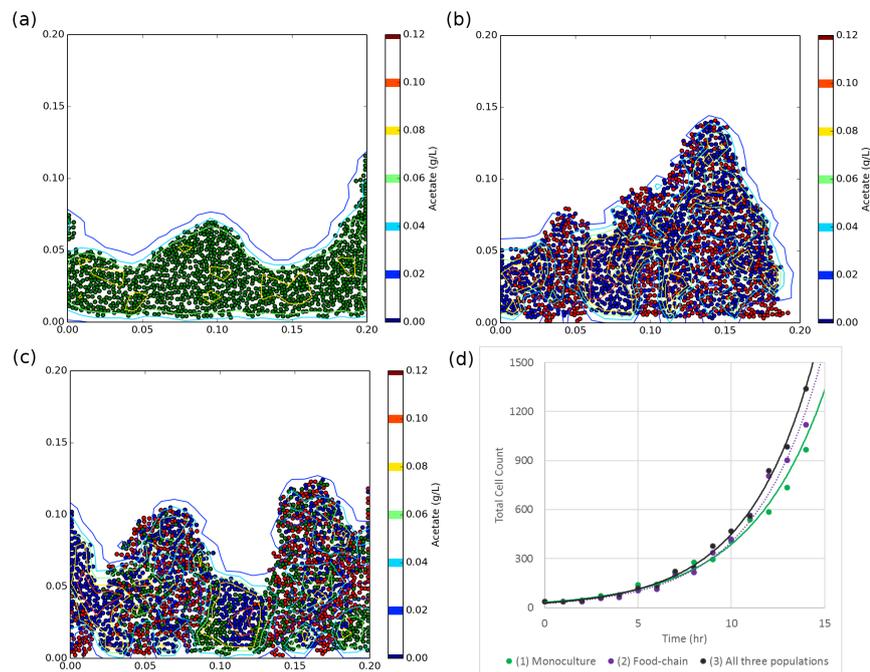


Figure 4: Examples of final microorganism populations after 15 hours of growth for: (a) a monoculture community comprised of the generalist population (green circles), (b) a food-chain community consisting of producer (blue) and scavenger (red) populations, and (c) a diverse community consisting of all three populations: generalist (green), producer (blue), and scavenger (red). (d) Total cell count as a function of time for the monoculture (green), producer and scavenger combined (purple), and all three populations combined (black). The trendlines are best-fit exponential curves.

lower than the generalist monoculture, and the overall biomass productivity (i.e., final total cell count) of the community was significantly ($\sim 10\%$) higher.

The third community was a combination of the first two communities and consisted of all three populations considered here: (1) generalist, (2) producer, and (3) scavenger. The initial community was 12 cells of each population for a total of 36 cells. As shown in Table 2, this combined community had the largest biomass productivity. An example of a final cell distribution is shown in Figure 4c. In some cases, this community had the highest byproduct concentration due to the larger populations of generalist and producer. However, the average concentration never rose significantly higher than the other two communities due to byproduct consumption by the scavenger population.

The total population of the three different scenarios as a function of time is shown in Figure 4d. For the first 5 hours, the monoculture population (shown in green) is slightly higher than either the producer-scavenger community (purple) or the three populations combined (shown in black). After 6 hours, the mixed communities increase in population more rapidly with the community with all three populations showing the most rapid increase in total cell numbers.

4 Conclusions

Microbial food chains can represent an ecologically competitive partitioning of environmental resources. Resource partitioning via cross-feeding can potentially benefit a community in numerous ways, and two specific advantages are examined here. First, it was shown through analysis of resource partitioning into substrate and enzyme pools that division of labor via interacting populations can maximize cell function (e.g., enzymatic flux) for a limiting anabolic resource investment. Second, dividing metabolic processes between cell populations allows the community to avoid a large accumulation of inhibitory byproducts, enhancing overall community productivity. Since different metabolic byproducts present varying degrees of inhibition, resource investment, and energy demand, there is a wide range of benefits from resource sharing in microbial communities. As research and commercial interest in microbial consortia increases, a better quantitative understanding of the benefits of resource sharing will be important for rational engineering and control.

Acknowledgments

This work was supported by National Science Foundation award DMS-1361240 and National Institutes of Health award U01EB019416.

References

- [1] Alpkvist, E., C. Picioreanu, M.C.M. van Loosdrecht, and A. Heyden, (2006) Three-dimensional biofilm model with individual cells and continuum EPS matrix. *Biotech. Bioeng.*, **95**(5):961-79.
- [2] Beck, A.E., K.A. Hunt, H.C. Bernstein, and R.P. Carlson, (2016) Interpreting and designing microbial communities for bioprocess applications, from components to interactions to emergent properties. *Biotechnologies for Biofuel Production and Optimization*. **1**: 407-432.

- [3] Bernstein, H.C., S.D. Paulson, and R.P. Carlson, (2012) Synthetic *Escherichia coli* consortia engineered for syntrophy demonstrate enhanced biomass productivity. *J. Biotechnology*, **157**(1):159-66.
- [4] Bernstein, H.C. and R.P. Carlson, (2014) Design, construction, and characterization methodologies for synthetic microbial consortia. *Methods Mol. Biol.*, **1151**:49-68.
- [5] Harvey, E., J. Heys, and T. Gedeon, (2014) Quantifying the effects of the division of labor in metabolic pathways. *J. Theor. Biol.*, **360**:222-42.
- [6] Holland, J.N. and D.L. DeAngelis, (2009) Consumer-resource theory predicts dynamic transitions between outcomes of interspecific interactions. *Ecol. Lett.*, **12**: 1357-66
- [7] Kreft, J.U., C. Picioreanu, J.W.T. Wimpenny, and M.C.M. van Loosdrecht, (2001) Individual-based modelling of biofilms. *Microbiology-Sgm.*, **147**:2897-912.
- [8] Picioreanu, C., J.U. Kreft, and M.C.M. van Loosdrecht, (2004) Particle-based multidimensional multispecies biofilm model. *Appl. Env. Microbiology*, **70**:3024-40.
- [9] Rosenzweig, R.F., R.R. Sharp, D.S. Treves, and J. Adams, (1994) Microbial evolution in a simple unstructured environment: Genetic differentiation in *Escherichia coli*. *Genetics*, **137**(4):903-17.
- [10] Schink, B., (2002) Synergistic interactions in the microbial world. *Anton Leeuw Int. J.*, **81**(1-4):257-61.
- [11] Smith, H.L. and P. Waltman, (1995) The theory of the chemostat. *Cambridge University Press*, Cambridge.
- [12] Stewart, P.S. and J.B. Raquepas, (1995) Implications of reaction-diffusion theory for the disinfection of microbial biofilms by reactive antimicrobial agents. *Chem. Eng. Sci.*, **50**(19): 3099-104.
- [13] Stump, S. M. and C. A. Klausmeier, (2016) Competition and coexistence between a syntrophic consortium and a metabolic generalist, and its effect on productivity. *J Theo Biology* **404**: 348-360.
- [14] Venters, M., R.P. Carlson, T. Gedeon, and J.J. Heys, (2017) Effects of spatial localization on microbial consortia growth, *PLoS ONE*, *accepted*.
- [15] <https://github.com/jeffheys/biofilmSegregation>

Coupled versus consecutive positive feedback loops

Kirsten Jenkins¹, Edina Rosta², Attila Csikasz-Nagy^{1,3}

¹ Randall Division of Cell and Molecular Biophysics, Faculty of Life Sciences and Medicine, King's College London, London, United Kingdom

² Department of Chemistry, Faculty of Natural and Mathematical Sciences, King's College London, London, United Kingdom

³ Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Budapest, Hungary

Abstract

Biological systems use various network motifs in order to create a desired emergent behaviour. One of the prominent motifs is the positive feedback loop. Positive feedback loops have the ability to create a sharp switch-like transition between two steady states. Positive feedback loops that create switch-like behaviour can also be coupled to each other in various ways: they could be directly coupled through a shared component, or flipping one switch can drive a second downstream switch without sharing a common component, which we term as consecutive switches. Here we investigate the kinetic differences these coupling variants can bring, and the effect of these upon a biological system using as an example two eukaryotic cell cycle checkpoints, G1/S and metaphase to anaphase. We find that consecutive positive feedback loops create different global behaviours that differ from those created by directly coupling the positive feedback loops. Directly coupled switches always flip all species in the network at the same moment; consecutive switches can also do this under some conditions, but they also have the ability to switch some species first and other species later, introducing an intermediary state.

1 Introduction

Dynamical systems modelling builds networks, which explain how different species interact with each other, with the aim of understanding the global behaviour that results from the network. Examples of these global behaviours that are particularly relevant to biology, is switch-like transitions and oscillations, due to the regularity with which they are experimentally observed [1].

From a dynamical perspective, a switch is a rapid and large alteration in the activity of a species over a small change in input. Switch-like transitions are created through the presence of one or more positive feedback loops (PFBLs) within the system, and are one of multiple possible behaviours that can be produced using PFBLs [2-6]. A PFBL is a loop of either purely activations,

or activations and an even number of inhibitions within a network. This motif often causes each species to activate its own activators/inactivate its own inhibitors; creating an amplification of the signal that drives the rapid alteration in concentration of the molecule.

Some switches have the further property of being bistable [6–8]. Bistability, is the phenomena that is observed when an input that creates a switch will not reverse the effect; but remains in the new state, causing hysteresis. A different value of the input is required to revert the switch to the original state (hysteresis). Bistability is caused by two stable points existing for a given parameter set such that the final state of the system will then depend upon the initial conditions (e.g. initial concentrations of molecules) as well the parameters of the system (e.g. inputs, thresholds and rate constants). Bistability is of particular importance in biology, as it provides the system with resistance to external noise and perturbation, as well as a sudden, fast and significant change in state over a small change in system input [8].

The cell cycle is of particular interest when attempting to understand biological switch-like transitions, as it is a system that all living organisms undergo in order to grow and divide, figure 1. In eukaryotes, the cell cycle can be viewed as a series of abrupt transitions between each phase that converge upon a closed loop [9]. As growth and division need to be tightly regulated, there are three checkpoints within the cell cycle, where the cell is able to confirm if previous processes are complete and no errors have occurred. These checkpoints are at the G1/S transition, the G2/M phase transition, and the metaphase to anaphase transition [10]. At each of these checkpoints the cell has a gene regulatory network that inhibits progress to the next stage of the cycle until the cell is ready. The change of phase then creates a fast change in many molecular concentrations. For this reason, the regulatory system should act like a switch, and once the phase change is completed, bistability should ensure that noise cannot flip the switch back [11–14].

Biology often uses more than one PFBL to create a switch-like transition through coupling them together. This can be done either directly through a shared component, or flipping one switch, which can drive a second downstream switch. This system we have named a consecutive switch. One example of a directly coupled PFBL network is controlling the metaphase to anaphase transition [15–17]. It has been previously found that directly coupling PFBLs increases the stability and resistance to noise of the switches, and changes the speed of switches [8, 18, 19]. The cell cycle can also be described as a series of switch-like transitions that drive each other in turn. Not only does the whole cell cycle appear as a series of consecutive PFBLs that create sudden, large concentration changes in key molecules, but also within the

G1/S transition (the first cell cycle checkpoint) in budding yeast (*S. cerevisiae*) multiple sequential PFBLs sequential drive one other [20, 21].

We therefore propose to investigate the differences between coupled and consecutive PFBLs upon the properties of a switch. We shall utilise the cell cycle checkpoints, in particular the metaphase to anaphase and G1/S transitions, as an example to examine the effects of the differences upon coupled bistable biological systems.

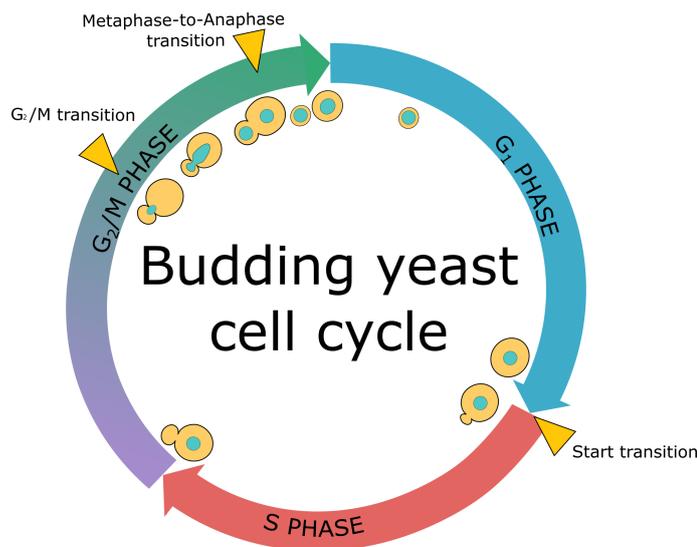


Figure 1: Budding yeast cell cycle. The cell cycle begins in G1 phase and the cell grows, the start transition ensures that sufficient size has been reached prior to S phase where DNA is replicated. G2 phase allows further growth, the G2/M transition ensures DNA has been replicated without error, and M phase is where the cell divides with the metaphase to anaphase transition ensuring proper alignment of the spindles prior to DNA separating into mother and daughter cells.

2 Model

We have created simplified networks of metaphase to anaphase transition and the G1/S transition, in order to compare the effects of directly coupled versus consecutive PFBL networks upon the resulting switch-like behaviour, shown in figure 2.

The metaphase to anaphase transition shows a directly coupled PFBL network in *S. cerevisiae*. This transition inhibits the cyclin dependent kinase activity (Cdc28) so that the cell divides [22]. The first positive feedback loop

involves Sic1 (an inhibitor of Cdc28) and Cdc28 [15, 16]. During metaphase Cdc28 activity is high which inhibits Sic1. During the transition, Cdc14 concentration increases which activates Sic1 and inhibits Cdc28. The second positive feedback loop is formed by Cdh1 (also a Cdc28 inhibitor) and Cdc28 [17]. Cdc14 also activates Cdh1 which likewise inhibits Cdc28 (figure 2, left panel).

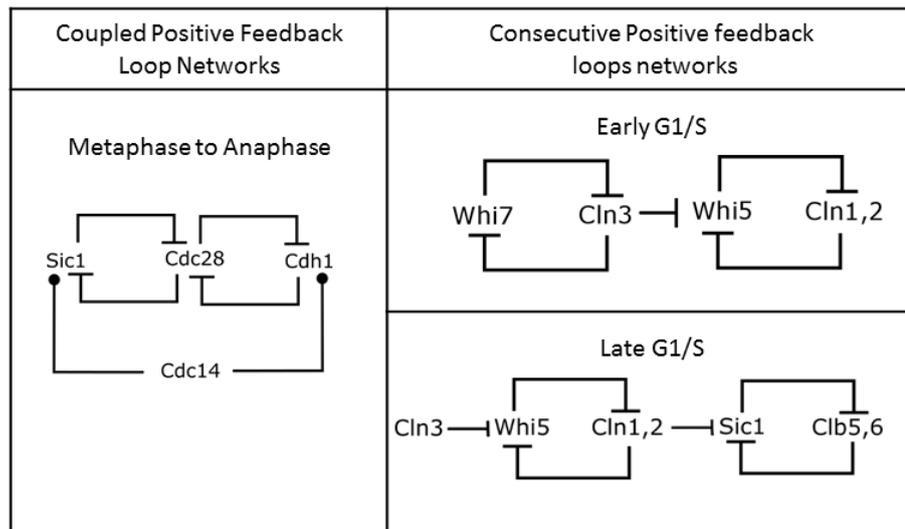


Figure 2: Simplified cell cycle checkpoint transition networks for *S. cerevisiae* that demonstrate coupled and consecutive PFBL networks. An example of a classic, well understood coupled PFBL network is the metaphase to anaphase transition which inactivates Cdc28, through Sic1 and Cdh1 (left). Two consecutive PFBL networks are created using the G1/S transition by splitting the G1/S transition into early and late processes. This enable each network to contain two PFBLs that are sequential (right). The two networks differ in that in early G1/S (right, top), Cln3 synthesis is increased so that the input is inside the first PFBL; whereas for the late G1/S network (right, bottom) Cln3 is still increased as the input but is external to the PFBL it is affecting.

To examine the effect of consecutive PFBLs upon a network, we use the G1/S transition in *S. cerevisiae*. In late G1, at the beginning of the G1/S transition (early G1/S in the model), cyclin Cln3 is produced in the ER but inhibited from being released to its destination in the nucleus by Whi7 [23–25]. Once Cln3 accumulates to a sufficiently high concentration, it is able to induce the phosphorylation and inactivation of Whi7. This allows Cln3 to bind the cyclin dependent kinase (Cdk), Cdc28, further inhibiting Whi7. Cln3/Cdc28 then enters the nucleus, creating the first PFBL. Nuclear accumulation of Cln3/Cdc28 drives a second positive feedback loop that involves Cln1,2 and Whi5 [12,26–

29]. Cln3/Cdc28 phosphorylates and inactivates Whi5, which is inhibiting the transcription factor required for S phase (SBF). The release of SBF starts to transcribe Cln1,2 which binds to Cdc28, and further inhibits Whi5, allowing the activation of SBF. Cdc28/Cln1,2 also targets the Cdk inhibitor Sic1. Upon the presence of Cdc28/Cln1,2, the inhibition of Cdk/Clb5,6 by Sic1 is surpassed by Cdk/Clb5,6 inducing the degradation of Sic1, [30] generating a third positive (double negative or antagonistic) feedback loop [31].

As this transition contains three as opposed to two PFBLs, we are able to split this network into early G1/S using the first and second PFBLs (figure 2, right top); and late G1/S containing the second and third PFBLs (figure 2, right bottom). This allows us to explore if the position of the input, Cln3, has an effect upon the resulting behaviour of the network. In early G1/S, the rate at which Cln3 is synthesized is increased manually, affecting Cln3 concentration, so the input to the system is within the first PFBL. Whereas in the late G1/S model, Cln3 concentration is increased manually to create the input for the network and this is external input to the first PFBL.

We turned each wiring diagram in figure 1 into a mathematical model. The models are deterministic and created using ordinary differential equations (ODEs), (table 1). Each of the species is able to be in an ON or an OFF state and all the transitions between the ON and OFF states are governed by Goldbeter-Koshland kinetics [32], such that a reaction rate, r , is defined as:

$$r_{off \rightarrow on} = \frac{k_1 [Catalyst_{on}] [X_{off}]}{K_{M1} + [X_{off}]} \text{ and } r_{on \rightarrow off} = \frac{k_2 [Catalyst_{off}] [X_{on}]}{K_{M2} + [X_{on}]}$$

where k is the kinetic parameter and K_M is the Michaelis-Menten constant for the reaction. The concentrations of catalysts in each reaction are the concentrations of inhibitors or activators such that: inhibitions will act as catalysts to $r_{on \rightarrow off}$ and activations will act as catalysts to $r_{off \rightarrow on}$. A leakage of 0.1 is added to all the models for all transition reactions (table 1). Static thresholds are also added when a reaction has no catalyst. These conditions ensure that all variables remain positive, which is important since they represent molecular concentrations.

	Equations
Metaphase / Anaphase	$dsic1on/dt = k1*(totsic1-sic1on)*(0.1+cdc14) / (KM6+(totsic1-sic1on))$ $-k2*sic1on*(0.1+cdc28on)/(KM1+sic1on)$ $dcdc28on/dt = k3*(totcdc28-cdc28on)*(0.1+Threshold) / (KM2+(totcdc28-cdc28on))$ $-k4*cdc28on*(sic1on+cdh1on+0.1) / (KM3+cdc28on)$ $cdh1on/dt = k5*(totcdh1-cdh1on)*(0.1+cdc14) / (KM4+(totcdh1-cdh1on))$ $-k6*cdh1on*(0.1+cdc28on)/(KM5+cdh1on)$
Early G1/S	$dcln3off/dt = s*cln3synthrate-k1*cln3off*(Threshold1+0.1) / (KM6+cln3off)+k2*cln3on*(whi7on+0.1) / (KM1+cln3on) -k3*cln3off$ $dcln3on/dt = (k1*cln3off*(Threshold1+0.1)/(KM6+cln3off)) -k2*cln3on*(whi7on+0.1)/(KM1+cln3on) -k10*cln3on$ $dwhi7on/dt = k4*(totwhi7-whi7on)*(Threshold2+0.1) / (KM2+(totwhi7-whi7on))-k5*whi7on*(cln3on+0.1)/(whi7on+KM3)$ $dcln1on/dt = (-k6*cln1on*(whi5on+0.1)/(KM4+cln1on)) +k7*(totcln1-cln1on)*(Threshold4+0.1) / (KM5+(totcln1-cln1on))$ $dwhi5on/dt = (-k8*whi5on*(cln1on+cln3on+0.1) / (KM6+whi5on))+k9*(totwhi5-whi5on)*(Threshold3+0.1)/(KM7+(totwhi5-whi5on))$
Late G1/S	$dwhi5on/dt = k1*(totwhi5-whi5on)*(Threshold1+0.1) / (KM6+(totwhi5-whi5on))$ $-k2*whi5on*(cln1on+cln3+0.1)/(KM1+whi5on)$ $dcln1on/dt = k3*(totcln1-cln1on)*(Threshold2+0.1) / (KM2+(totcln1-cln1on))$ $-k4*cln1on*(whi5on+0.1)/(KM3+cln1on)$ $dsic1on/dt = k5*(totsic1-sic1on)*(Threshold3+0.1) / (KM4+(totsic1-sic1on))$ $-k6*sic1on*(cln1on+clb5on+0.1)/(KM5+sic1on)$ $dclb5on/dt = k7*(totclb5-clb5on)*(Threshold4+0.1) / (KM7+(totclb5-clb5on))$ $-k8*clb5on*(sic1on+0.1)/(KM8+clb5on)$

Table 1: Model equations for each network using ordinary differential equations. OFF states are equal not expressly written can be calculated as OFF=total – ON, the exception is Cln3off in early G1/S, which is written above.

To analyse the system and create the plots provided in the results, the input parameter for each network is varied and the steady state(s) of the system for that parameter set are plotted to create a bifurcation diagram. In order to be able to vary Cln3 synthesis in the early G1/S model we include terms in the ODE model for Cln3off to be synthesized and both Cln3off and Cln3on to be degraded, which means that the total concentration of Cln3 varies parameter set. Experimental results show that Cln3 concentration varies with volume, and has been used in many models [20, 21, 33] The parameter sets and initial conditions for all plots provided in the results are presented in the Appendix, table 2.

3 Results

Bifurcation analysis allows us to examine where a system will alter its qualitative behaviour when an input is changed due to external signals, through the evolution of fixed points. In particular, this will enable us to identify switch-like transitions in the system. Switch behaviour is created when the steady state curve, or attractor, that the system is currently attracted to is destroyed, forcing the system to rapidly jump from the initial steady state curve to a new steady state curve over a very small change in input parameter [34].

The bifurcation diagram for the coupled PFBL network (metaphase to anaphase transition) found that one switch-like transition is created that turns off Cdc28 (figure 3). This can be observed by following the thick curve for [Cdc28] from when [Cdc14]=0 ([Cdc28]=10) until the bifurcation point. If [Cdc14] is increased any higher, then the steady state of [Cdc28] is close to zero.

The bifurcation diagrams for the consecutive PFBLs exhibit new behaviour that is not possible for coupled PFBLs (figure 3, bifurcation diagrams for early and late G1/S transition). These diagrams show that, as the input parameter is increased, it is possible for the switch to occur in one of two ways. On the left hand diagrams for the early and late G1/S transitions, Cln3 and Cln1,2/Cln1,2 and Clb 5,6 respectively, turn on at the same value of input parameter (see figure 3 and table 3). We call this “compound” switch behaviour, as 2 switch-like transitions occur together making the appearance of only one abrupt transition between steady states. This means that one of the switches is effectively hidden and is not observed as a separate switch. In the right hand panels of the G1/S models, the parameters within the system have been changed (see appendix for complete set of parameters for each plot in figure 3). It can be seen that the first and second PFBLs do not flip at the same parameter values (see figure 3 and table 3).

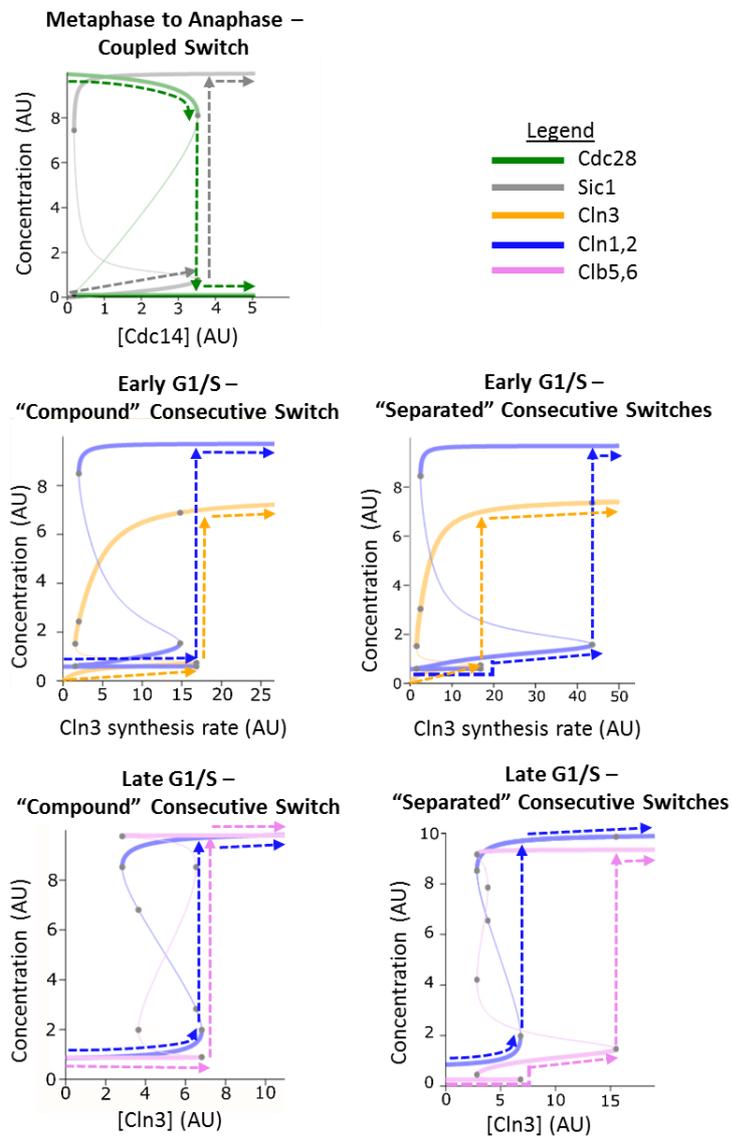


Figure 3: Bifurcation Diagrams of coupled and consecutive networks. Solid lines represent steady states; black circles are bifurcation points; and coloured arrows are the concentrations for each species as the transition is completed. The directly coupled PFBL creates one switch-like transition, whereas the consecutive PFBLs have either “compound” or “separated” switch-like transitions. Equations of these models are in table 1 and see table 2 in the appendix for parameter alterations used to create the “compound” and “separated” switch behaviours. Figures created using Oscill8 [37].

We term this a “separated” consecutive switch behaviour. The Cln3 for early G1/S and Cln1,2 in late G1/S turn on at a lower input parameter than the Cln1,2 and Clb5,6 respectively. This occurs because the first switch-like transition occurs but this does not sufficiently suppress the inhibitor in the second PFBL to induce the second switch-like transition, creating an intermediary stable state. Further suppression of the inhibitor or activation of the activator eventually causes the second PFBL to create a switch-like transition.

It was observed that both the “compound” and “separated” switch behaviour scenarios could occur for the proposed consecutive FBFL wirings. This suggests that the position of the input, whether internal or external, does not affect the emergent behaviour of the system.

4 Discussion

It is observed that directly coupling PFBLs creates one switch-like transition that has previously been shown to increase bistability and decrease perturbations caused by noise [19]. The cell often couples PFBLs but it is also becoming increasingly apparent that the cell ‘stacks up’ consecutive PFBLs to create one phase change. This is most well known in the G1/S transition [14, 20].

It is therefore important to attempt to understand the differences between these two network architectures upon the switch-like transitions. It was observed that by using consecutive PFBLs driving each other, multiple switch-like transitions are created as opposed to a single switch behaviour. The two different networks tested also showed that these switch-like transitions may either be seen at different times or that one may be hidden and effectively unobservable.

The resulting appearance and number of switch behaviours in a system is dependent upon the parameters. This could allow models to be constrained in parameter space but also implies that experiments need to be designed to identify if there are hidden PFBLs within a network. In biology, mutants can be created to knockout molecules. We propose this as a method to enable experimentalists to remove particular PFBLs and identify of these extra switches and increase the accuracy of systems level networks.

We hypothesize that the cell may have either switch behaviour occur at different times or overlap PFBLs to create a “compound” switch behaviour depending upon the requirements of the network. For example, in the G1/S transition Cln1,2 triggers budding in *S. cerevisiae* and Clb5,6 initiates DNA replication [30, 35, 36]. The cell creates the bud prior to DNA replication, in normal conditions, but once budding has occurred DNA replication should

begin. Therefore, we would expect one switch-like transition to drive another switch-like transition, with each occurring at different times with an intermediary stage between them. These intermediary states may also be present in the cell cycle which is a large series on consecutive switches. The intermediary states may create hidden phases of the cell cycle that need to be completed in order to trigger the next state.

In contrast to this, the cell can use multiple PFBLs to create a “compound” switch behaviour that has a greater change in concentrations between the ON and OFF state of the final molecule in the network. The final molecule is observed to have almost background concentration change as the input parameter is increased until the bifurcation point, where it completely turns on. This may be especially important in a situation where a small concentration change in a molecule may have a negative impact upon cell health or wastes cellular resources, for instance a transcription factor that once turned on starts transcribing genes related to a different phase. This may also be the case for Clb5,6 as S phase processes should not begin prior to completion of G1.

In conclusion, we observe that consecutive PFBLs can create different properties to those of coupled or single positive feedback loops and therefore may be particularly useful in some biological systems [8, 18, 19].

Acknowledgments

This work is supported by the EPSRC Centre for Doctoral Training in Cross-Disciplinary Approaches to Non-Equilibrium Systems (CANES, EP/L015854/1).

References

- [1] Tyson JJ, Albert R, Goldbeter A, Ruoff P, Sible J. Biological switches and clocks. *J R Soc Interface* 2008;5 Suppl 1:S1–8. doi:10.1098/rsif.2008.0179.focus.
- [2] Novák B, Tyson JJ. Design principles of biochemical oscillators. *Nat Rev Mol Cell Biol* 2008;9:981–91.
- [3] Tyson JJ, Chen KC, Novak B. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr Opin Cell Biol* 2003;15:221–31.
- [4] Thomas R. Laws for the dynamics of regulatory networks. *Int J Dev Biol* 2002;42:479–85.

- [5] Tyson JJ, Novák B. Functional motifs in biochemical reaction networks. *Annu Rev Phys Chem* 2010;61:219.
- [6] Griffith JS. Mathematics of cellular control processes II. Positive feedback to one gene. *J Theor Biol* 1968;20:209–16.
- [7] Thomas R, D’Ari R. *Biological Feedback*. CRC Press; 1990.
- [8] Ferrell JE. Feedback regulation of opposing enzymes generates robust, all-or-none bistable responses. *Curr Biol* 2008;18:R244–5.
- [9] Tyson JJ. Models of cell cycle control in eukaryotes. *J Biotechnol* 1999;71:239–44.
- [10] Morgan D. *The Cell Cycle: Principles of Control (Primers in Biology)*. New Sci Press London 2006.
- [11] Charvin G, Oikonomou C, Siggia ED, Cross FR. Origin of irreversibility of cell cycle start in budding yeast. *PLoS Biol* 2010;8:e1000284. doi:10.1371/journal.pbio.1000284.
- [12] Skotheim JM, Di Talia S, Siggia ED, Cross FR. Positive feedback of G1 cyclins ensures coherent cell cycle entry. *Nature* 2008;454:291–6. doi:10.1038/nature07118.
- [13] Doncic A, Falleur-Fettig M, Skotheim JM. Distinct interactions select and maintain a specific cell fate. *Mol Cell* 2011;43:528–39. doi:10.1016/j.molcel.2011.06.025.
- [14] Zhang T, Schmierer B, Novák B. Cell cycle commitment in budding yeast emerges from the cooperation of multiple bistable switches. *Open Biol* 2011;1:110009. doi:10.1098/rsob.110009.
- [15] López-Avilés S, Kapuy O, Novák B, Uhlmann F. Irreversibility of mitotic exit is the consequence of systems-level feedback. *Nature* 2009;459:592–5.
- [16] Simonetta M, Manzoni R, Mosca R, Mapelli M, Massimiliano L, Vink M, et al. The influence of catalysis on mad2 activation dynamics. *PLoS Biol* 2009;7:e1000010.
- [17] Holt LJ, Krutchinsky AN, Morgan DO. Positive feedback sharpens the anaphase switch. *Nature* 2008;454:353–7.
- [18] Brandman O, Ferrell JE, Li R, Meyer T. Interlinked fast and slow positive feedback loops drive reliable cell decisions. *Science* (80-) 2005;310:496–8.

- [19] Cardelli L, Csikász-Nagy A, Dalchau N, Tribastone M, Tschaikowski M. Noise reduction in complex biological switches. *Sci Rep* 2016;6.
- [20] Adames NR, Schuck PL, Chen KC, Murali TM, Tyson JJ. Experimental testing of a new integrated model of the budding yeast START transition n.d.;24061.
- [21] Chen KC, Csikasz-Nagy A, Gyorffy B, Val J, Novak B, Tyson JJ. Kinetic Analysis of a Molecular Model of the Budding Yeast Cell Cycle. *Mol Biol Cell* 2000;11:369–91. doi:10.1091/mbc.11.1.369.
- [22] Farr KA, Cohen-Fix O. The metaphase to anaphase transition. A case of productive destruction. *Eur J Biochem* 1999;263:14–9.
- [23] Tyers M, Tokiwa G, Futcher B. Comparison of the *Saccharomyces cerevisiae* G1 cyclins: Cln3 may be an upstream activator of Cln1, Cln2 and other cyclins. *EMBO J* 1993;12:1955–68.
- [24] Dirick L, Böhm T, Nasmyth K. Roles and regulation of Cln-Cdc28 kinases at the start of the cell cycle of *Saccharomyces cerevisiae*. *EMBO J* 1995.
- [25] Yahya G, Parisi E, Flores A, Gallego C, Aldea M. A Whi7-Anchored Loop Controls the G1 Cdk-Cyclin Complex at Start. *Mol Cell* 2014;53:115–26. doi:10.1016/j.molcel.2013.11.015.
- [26] Costanzo M, Nishikawa JL, Tang X, Millman JS, Schub O, Breitkreuz K, et al. CDK Activity Antagonizes Whi5, an Inhibitor of G1/S Transcription in Yeast. *Cell* 2004;117:899–913. doi:10.1016/j.cell.2004.05.024.
- [27] de Bruin RAM, McDonald WH, Kalashnikova TI, Yates J, Wittenberg C. Cln3 activates G1-specific transcription via phosphorylation of the SBF bound repressor Whi5. *Cell* 2004;117:887–98. doi:10.1016/j.cell.2004.05.025.
- [28] Schmoller KM, Turner JJ, Kõivomägi M, Skotheim JM. Dilution of the cell cycle inhibitor Whi5 controls budding-yeast cell size. *Nature* 2015;526:268–72. doi:10.1038/nature14908.
- [29] Wagner M V, Smolka MB, de Bruin RAM, Zhou H, Wittenberg C, Dowdy SF. Whi5 regulation by site specific CDK-phosphorylation in *Saccharomyces cerevisiae*. *PLoS One* 2009;4:e4300. doi:10.1371/journal.pone.0004300.

- [30] Nash P, Tang X, Orlicky S, Chen Q, Gertler FB, Mendenhall MD, et al. Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. *Nature* 2001;414:514–21.
- [31] Ferrell JE. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr Opin Cell Biol* 2002;14:140–8.
- [32] Goldbeter A, Koshland DE. An amplified sensitivity arising from covalent modification in biological systems. *Proc Natl Acad Sci U S A* 1981;78:6840–4.
- [33] Csikász-Nagy A, Palmisano A, Zámboorszky J. Molecular network dynamics of cell cycle control: transitions to start and finish. *Cell Cycle Synchronization Methods Protoc* 2011:277–91.
- [34] Kuznetsov Y. *Elements of applied bifurcation theory*. Vol. 112. Springer Science and Business Media; 2013.
- [35] Baroni MD, Monti P, Alberghina L. Repression of growth-regulated G1 cyclin expression by cyclic AMP in budding yeast. *Nature* 1994;371:339–42.
- [36] Mendenhall MD, Hodge AE. Regulation of Cdc28 cyclin-dependent protein kinase activity during the cell cycle of the yeast *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* 1998;62:1191–243.
- [37] Conrad E. <http://oscill8.sourceforge.net/> 2004.

Appendix

Cell cycle transition	Parameters and initial conditions that will create bifurcation diagrams
Metaphase to Anaphase	<p>Initial conditions – Sic1on = 0 Cdh1on = 0 Cdc28on = 10</p> <p>Parameters – $k_1 = 1.1$ $K_{M1} = 1$ $k_2 = 1$ $K_{M2} = 1$ $k_3 = 1$ $K_{M3} = 1$ $k_4 = 1$ $K_{M4} = 1$ $k_5 = 1$ $K_{M5} = 1$ $k_6 = 1$ $K_{M6} = 1$ Totsic1 = 10 Threshold = 2 Totcdh1 = 10 Totcdc28 = 10</p>
Early G1/S Compound	<p>Initial conditions – cln3off = 0 cln3on = 0 whi7on = 10 cln1on = 0 whi5on = 10</p> <p>Parameters – $s = 10$ $k_1 = 1$ $k_2 = 1$ $k_3 = 1$ $k_4 = 1$ $k_5 = 1$ $k_6 = 1$ $k_7 = 1$ $k_8 = 1$ $k_9 = 1$ $k_{10} = 0.5$ totcln1 = 10 totwhi7 = 10 totwhi5 = 10 $K_{M1} = 1$ $K_{M2} = 1$ $K_{M3} = 1$ $K_{M4} = 1$ $K_{M5} = 1$ $K_{M6} = 1$ $K_{M7} = 1$ Threshold1 = 3.9 Threshold2 = 1 Threshold3 = 9.5 Threshold4 = 4</p>
Early G1/S Separated	<p>Initial conditions – cln3off = 0 cln3on = 0 whi7on = 10 cln1on = 0 whi5on = 10</p> <p>Parameters – $s = 10$ $k_1 = 1$ $k_2 = 1$ $k_3 = 1$ $k_4 = 1$ $k_5 = 1$ $k_6 = 1$ $k_7 = 1$ $k_8 = 1$ $k_9 = 1$ $k_{10} = 0.5$ totcln1 = 10 totwhi7 = 10 totwhi5 = 10 $K_{M1} = 1$ $K_{M2} = 1$ $K_{M3} = 1$ $K_{M4} = 1$ $K_{M5} = 1$ $K_{M6} = 1$ $K_{M7} = 1$ Threshold1 = 3.9 Threshold2 = 1 Threshold3 = 9.4 Threshold4 = 5</p>
Late G1/S Compound	<p>Initial conditions – cln1on = 0 whi5on = 10 clb5on = 0 sic1on = 10</p> <p>Parameters – $k_1 = 1$ $k_2 = 1$ $k_3 = 1$ $k_4 = 1$ $k_5 = 1$ $k_6 = 1$ $k_7 = 1$ $k_8 = 1$ totcln1 = 10 totwhi5 = 10 totclb5 = 10 totsic1 = 10 $K_{M1} = 1$ $K_{M2} = 1$ $K_{M3} = 1$ $K_{M4} = 1$ $K_{M5} = 1$ $K_{M6} = 1$ $K_{M7} = 1$ $K_{M8} = 1$ Threshold1 = 10 Threshold2 = 5 Threshold3 = 10 Threshold4 = 5</p>
Late G1/S Separated	<p>Initial conditions – cln1on = 0 whi5on = 10 clb5on = 0 sic1on = 10</p> <p>Parameters – $k_1 = 2$ $k_2 = 2$ $k_3 = 1$ $k_4 = 1$ $k_5 = 1$ $k_6 = 1$ $k_7 = 0.1$ $k_8 = 1$ totcln1 = 10 totwhi5 = 10 totclb5 = 10 totsic1 = 10 $K_{M1} = 1$ $K_{M2} = 1$ $K_{M3} = 1$ $K_{M4} = 1$ $K_{M5} = 1$ $K_{M6} = 1$ $K_{M7} = 1$ $K_{M8} = 1$ Threshold1 = 10 Threshold2 = 5 Threshold3 = 10 Threshold4 = 22.6</p>

Table 2: Parameters and initial conditions used to create bifurcation diagrams found in results

	Concentrations for compound switch transition	Concentrations of Cln1/Sic1 for separated switch transition
Early G1/S transition	[Cln3] switch behaviour observed at cln3 synthesis rate=16.8 from 0.75 to 7.0 [Cln1/2] switch behaviour observed at cln3 synthesis rate=16.8 from 0.6 to 9.7	[Cln3] switch behaviour observed at cln3 synthesis rate=16.8 from 0.75 to 7.0 [Cln1/2] switch behaviour observed at cln3 synthesis rate=43.6 from 1.6 to 9.7
Late G1/S transition	[Cln1/2] switch behaviour observed at [cln3] = 6.8 from 2.0 to 9.7 [Sic1] switch behaviour observed at [cln3] = 6.8 from 0.9 to 9.8	[Cln1/2] switch behaviour observed at [cln3] = 6.8 from 2.0 to 9.7 [Sic1] switch behaviour observed at [cln3] = 15.5 from 1.5 to 9.3

Table 3: Concentrations of components that are activated in a switch-like manner and the input values at which this behaviour occurs.



Rational enzyme engineering towards optimization of a synthetic 1,3-propanediol pathway

Cláudio Frazão¹, Débora Trichez¹, Thomas Walther^{1,2} and Jean-Marie François^{1,2}

¹ LISBP, CNRS, INRA, INSA, Université de Toulouse, 135 Avenue de Rangueil, 31077 Toulouse, France

² TWB, 3 rue Ariane, 31520 Ramonville-St. Agne, France

Abstract

We have previously constructed a 1,3-propanediol de novo synthesis pathway starting from malate, composed of six consecutive enzymatic reaction steps. An important step in this pathway consists in the conversion of the monocarboxylic acid (L)-dihydroxybutyrate (DHB) into the keto-acid 2-oxo-4-hydroxybutyrate (OHB). In this study, we utilized a rational approach to engineer a DHB dehydrogenase by using bacterial malate dehydrogenase (eMDH) as the template enzyme. Structural analysis of (L)-malate dehydrogenase and (L)-lactate dehydrogenase enzymes that act on sterically cognate substrates revealed key residues in the substrate and co-substrate binding sites which are responsible for substrate discrimination. Accordingly, mutations were introduced into three different protein regions (substrate binding site, exterior loop and the vicinity of the coenzyme nicotinamide ring) yielding a total of twenty-four eMDH variants that were characterized on DHB, malate and lactate. The highest DHB dehydrogenase activity of $20.42 \pm 0.19 \text{ U}\cdot\text{mg}^{-1}$ was found for the quintuple mutant R81A:M85Q:I12V:G179D:D86S which corresponded to an increase of 2042-fold compared to the wild-type enzyme.

1 Introduction

A central goal of the bioeconomy consists in reducing our dependence on fossil resources through next-generation biomanufacturing by focusing on the industrial production of bio-based specialty chemicals [1]. To accomplish this goal, non-natural biochemical pathways are needed, in which synthetic biology – including redesign and protein engineering – emerges as a critical tool.

1,3-Propanediol (PDO) is an important chemical that can be utilized as monomer for the synthesis of high-value polymers. PDO can naturally be synthesized from glycerol by two enzymatic steps: glycerol dehydration to 3-hydroxypropionaldehyde (HPA) by a glycerol dehydratase, and HPA reduction to PDO by 1,3-propanediol oxidoreductase. No natural microorganisms have however been found to directly use sugars to produce PDO [2].

We recently constructed a non-natural biosynthetic pathway towards PDO production from malate using *Escherichia coli* as a microbial chassis [3,4]. Since malate is a TCA cycle intermediate, the proposed route can be engineered in selected hosts with the ability to utilize different and cheap sugars. The proposed PDO synthetic pathway is composed of six enzymatic steps: (1) phosphorylation of malate to 4-phospho-malate, (2) reduction of 4-phospho-malate to malate semialdehyde, (3) reduction of malate semialdehyde to (L)-2-dihydroxybutyrate (L-DHB), (4) oxidation of L-DHB to 2-keto-4-hydroxybutyrate (OHB), (5) decarboxylation of OHB to HPA, (6) reduction of HPA to PDO. Enzyme engineering towards substrate specificity remains as a critical task towards improving pathway efficiency.

In this work, we report on the construction of an efficient DHB dehydrogenase (step no. 4) by engineering the malate dehydrogenase from *Escherichia coli*.

2 Materials and methods

2.1 Cloning, mutagenesis, expression and purification

E. coli DH5 α (New England Biolabs) was routinely used for the construction of plasmids. The *mdh* gene encoding the wildtype (L)-malate dehydrogenase of *E. coli* (eMDH) was amplified by PCR and cloned into vector pET-28a+ (Novagen). Point mutations were introduced by PCR based on site-directed mutagenesis. The resulting PCR products were digested by DpnI at 37 °C to remove template DNA and transformed into competent cells. Mutated plasmids were verified to carry the desired mutations by sequencing.

Enzymes were expressed in *E. coli* BL21(DE3) cells (New England Biolabs) in 200 mL Luria-Bertani (LB) medium supplemented with 50 μ g/mL kanamycin (37 °C, 200 rpm) that were inoculated from an overnight culture at OD600 of 0.05 and grown to OD600 of 0.6 before protein expression was induced for 3h by addition of 1 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) to the culture medium. Cells were harvested by centrifugation and pellets were stored at -20 °C until further analysis. Protein purification starting from frozen cell pellets was performed according to [3].

2.2 Enzymatic assays

Protein concentrations were determined by the method of Bradford (Bio-Rad) and enzymatic activities on malate, DHB, and lactate were measured at 37 °C by monitoring the absorbance at 340 nm which is characteristic for NADH using an Epoch 2 microplate spectrophotometer (BioTek). Initial rate measurements were carried out for reaction mixtures (250 μ L) that contained variable

concentrations of substrate, 10 mM NAD⁺, 179 mM glycine buffer (pH 9.0), 5 mM MgCl₂, 50 mM KCl and appropriate amount of enzyme. Three substrates, including (L)-malate, (L)-lactate and a racemic mixture of (D/L)-DHB, were tested at in a concentration range between 0 and 50 mM.

3 Results

3.1 Strategy for DHB dehydrogenase design

Malate dehydrogenase catalyzes the reversible NAD⁺-dependent oxidation of malate to oxaloacetate. Because of the structural similarity between the C4 carboxylic acids DHB and malate, the dimeric eMDH enzyme was selected as a design template to engineer an enzyme which catalyzes the oxidation of DHB to OHB.

Malate and lactate dehydrogenases are both 2-hydroxy-acid dehydrogenases. They share a similar tertiary structure in which the nucleotide binding domain and catalytic residues are conserved but yet maintain a respective selectivity for di- and monocarboxylic acids. Key differences between the two enzymes are found in the sequences of a loop region that covers the active-site. In this regard, loop closure in LDH is similar to that in MDH, as implied by a common kinetic mechanism. Crucially, contrarily to MDH, LDH loop closure occurs without the favorable energy contributions of a stabilizing salt bridge formed between a second substrate carboxylate group absent in (L)-lactate and a second arginine guanidinium group in the enzyme. Since (L)-lactate and (L)-2,4-dihydroxybutyrate (DHB) are both monocarboxylic acids, identification of compensatory stabilizing loop interactions in lactate dehydrogenase afforded a design strategy for the engineering of DHB dehydrogenase.

3.2 Identification of the target mutation sites for engineering MDH

Sequence and structure comparison between MDH and LDH were used to identify putative target residue positions for mutation. LDH acts on a substrate which has only one carboxy group, contrarily to MDH that acts on a substrate presenting a second negatively charged carboxy group. Since DHB has also only one carboxy group, we have step-wise replaced eMDH residues by corresponding amino acids in LDH.

Comparison of MDH and LDH active-site loop sequences revealed the presence of three internal loop stabilizing interactions specific to LDH. The corresponding eMDH target mutation sites were step-wise replaced by corresponding amino acids in LDH.

Analysis of lactate dehydrogenase multiple sequence alignments shows the notable presence of smaller residues at (equivalent) positions in eMDH close

to the coenzyme nicotinamide ring. Yin and co-workers [5] demonstrated the favorable impact of an I12V mutation in the engineering of lactate dehydrogenase activity in eMDH.

3.3 Enzyme activity and specificity

The specific activities of wild-type eMDH and the twenty-four generated mutant enzymes were measured on (L)-malate, (D/L)-DHB and (L)-lactate at initial substrate concentrations of 50 mM. (L)-lactate was included in the analysis because of its structural similarity to DHB and malate. Enzymatic assays at substrate concentrations of up to 50 mM revealed that the activity of the mutant enzymes did not saturate on all substrates. Therefore, substrate affinity could not be determined. It is of note that we could only use a racemic mixture of (D/L)-DHB. Since eMDH is only active in (L)-stereoisomers it is very likely that the activities and substrate affinities of the MDH mutants on (L)-DHB are actually higher.

The wild-type enzyme eMDH had a very low DHB dehydrogenase activity ($0.01 \pm 0.00 \text{ U mg}^{-1}$) which was more than 6000-fold smaller than on its natural substrate malate ($62.25 \pm 1.33 \text{ U mg}^{-1}$) (Fig. 1a). No (L)-lactate dehydrogenase activity was found for this enzyme.

Point mutations were first introduced into the substrate-binding region of eMDH either individually or as groups (Fig. 1b). The substitution of arginine at position 81 by glutamine, cysteine and alanine residues (R81Q, R81C and R81A, respectively) resulted in a sharp decrease of enzyme activity on (L)-malate up to approximately 98% (Fig. 1a, 1b) confirming the pivotal role of R81 in determining substrate specificity of MDH enzymes [6]. Conversely, mutants in position R81 showed improved (L)-lactate and DHB dehydrogenase activity when compared to the wild-type enzyme. While the variant R81Q showed the highest activity on (L)-lactate among this group ($0.28 \pm 0.01 \text{ U.mg}^{-1}$), the mutant R81A was the one with highest DHB dehydrogenase activity ($0.33 \pm 0.01 \text{ U.mg}^{-1}$).

We therefore continued to combine the mutation R81A with mutations on the mobile loop which closes the active site (at residues A80, M85, D86, G179, T211) and mutations in the coenzyme nicotinamide ring vicinity NAD-cofactor binding site (at residue I12).

In Fig. 1c, we can observe the effect of grouping mutations at the substrate binding site and exterior loop. While no improved activity on (L)-lactate was found, most of eMDH variants exhibited higher activity on DHB, and notably the mutant R81A:M85Q ($2.02 \pm 0.11 \text{ U.mg}^{-1}$), suggesting a synergistic effect when introducing residue alterations at these two different regions of the protein. Interestingly, the increase in DHB dehydrogenase activity was also followed by an increase in malate dehydrogenase activity. Upon

simultaneous introduction of mutations in substrate binding site and in the coenzyme nicotinamide ring vicinity (Fig. 1d), the same trend was observed. The mutant R81A:I12V exhibited 1.7-fold higher DHB dehydrogenase activity if compared to best eMDH* with mutations at only the substrate binding site (Fig. 1b, 1d).

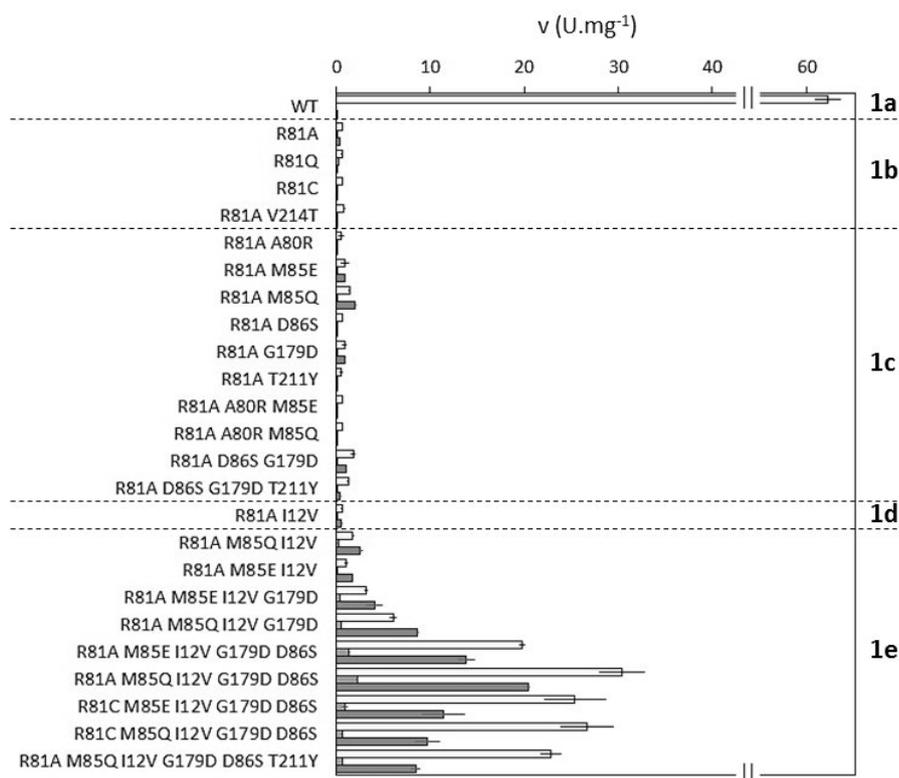


Figure 1: Specific activities of wild-type eMDH (1a) and mutant enzymes (1b-e) on (L)-malate (grey bars), (L)-lactate (red bars) and (D/L)-DHB (green bars). Mutants were divided into four groups according to the cumulation of mutations in up to three functional regions: 1b, substrate binding site; 1c, substrate binding site and exterior loop; 1d, substrate binding site and coenzyme nicotinamide ring vicinity; 1e, substrate binding site, exterior loop and coenzyme nicotinamide ring vicinity. Initial substrate concentration was 50 mM.

To investigate whether DHB dehydrogenase activity could still be improved, we decided to go further by combining mutations at the substrate binding site, exterior loop and vicinity of the coenzyme nicotinamide ring (Fig. 1e). Variants from this group were found to be the most interesting ones with

improvement in DHB dehydrogenase activity being noteworthy (up to 10-fold when comparing to R81A:M85Q), whilst activity on (L)-lactate was improved up to 7.8-fold when comparing to the single mutant R81Q. Again, mutant variants exhibiting improved activity on DHB showed increased activity on (L)-malate (up to 48% of eMDH activity), suggesting that a compromise between DHB and (L)-malate dehydrogenase activities is required.

Overall, the quintuple mutant R81A:M85Q:I12V:G179D:D86S (eMDH*₅) showed the highest activity on DHB (20.42 ± 0.19 U.mg⁻¹) and also on (L)-lactate (2.20 ± 0.03 U.mg⁻¹), while the triple mutant R81A:M85E:I12V (eMDH*₃) preferred DHB as a substrate over the original substrate (Table 1). Both variants target substitutions at three functional regions.

Clone	Specific activity (U.mg ⁻¹)			Specificity ^c log ₂ (V _D /V _M)
	(L)-Malate	(D/L)-DHB	(L)-Lactate	
eMDH	62.25 ± 1.33	0.01 ± 0.00	0.00 ± 0.00	-13.60 ± 0.01
eMDH* ₅ ^a	30.37 ± 2.35	20.42 ± 0.19	2.20 ± 0.03	-0.57 ± 0.03
eMDH* ₃ ^b	1.01 ± 0.17	1.77 ± 0.00	0.15 ± 0.01	0.81 ± 0.07

(a) eMDH* R81A:M85Q:I12V:G179D:D86S

(b) eMDH* R81A:M85E:I12V

(c) Ratio of the specific activities of one enzyme on (D/L)-DHB and (L)-malate. Activities were measured at 50 mM.

Table 1: Specific activity of best eMDH* on tested substrates at 50 mM

4 Conclusions

Creating enzymes with novel properties by protein engineering is a key enabling technology for synthetic biology, which aims at both improving productivities of natural products and enlarging the panel of chemicals that can be produced by microorganisms.

We are currently implementing a 6-step synthetic pathway which produces PDO from malate by employing 5 hitherto unknown enzymatic activities. In the present report we describe the engineering of a DHB dehydrogenase by using eMDH as the template enzyme. Wild-type eMDH had only very residual and hardly measurable activity on the DHB substrate (0.01 ± 0.00 U mg⁻¹). Upon introduction of identified point mutations at distinct protein functional regions, we were able to obtain improved DHB dehydrogenase activities in all the twenty-four generated eMDH variants. The highest DHB dehydroge-

nase activity of $20.42 \pm 0.19 \text{ U.mg}^{-1}$ was found for the quintuple mutant R81A:M85Q:I12V:G179D:D86S which corresponded to an increase of 2042-fold compared to the wild-type enzyme. On the other hand, the triple mutant R81A:M85E:I12V showed the highest specificity by preferring DHB over its original substrate (L)-malate.

The characteristics of these two engineered DHB dehydrogenase are highly promising and we shall test these new enzymes upon in vivo implementation of the synthetic PDO pathway.

Acknowledgements

The study was supported by a grant aid from Toulouse Transfer Technology and ANR (Synpathic project, ANR14-CE06-0024) to JMF, which also provided financial support for CF doctoral thesis. DT was supported by a post-doctoral grant (Science without borders program) provided by the CAPES foundation (Ministry of Education, Brazil).

References

- [1] Mak WS, Tran S, Marcheschi R, Bertolani S, Tompson J, Baker D, Liao JC. Integrative genomic mining for enzyme function to enable engineering of a non-natural biosynthetic pathway. *Nature Communications*, 2015. **6**.
- [2] Chen Z, Geng F, Zeng AP. Protein design and engineering of a de novo pathway for microbial production of 1, 3-propanediol from glucose. *Biotechnology Journal*, 2015. **10**(2): p. 284-289.
- [3] Walther T, François JM. Microorganism modified for the production of 1, 3-propanediol. U.S. Patent Application No. 14/410,732, 2013.
- [4] Walther T, Dressaire C, Cordier H, François JM. Method of production of 2, 4-dihydroxybutyric acid. U.S. Patent no. 9,238,829, 2016.
- [5] Yin Y, Kirsch JF. Identification of functional paralog shift mutations: conversion of *Escherichia coli* malate dehydrogenase to a lactate dehydrogenase. *Proceedings of the National Academy of Sciences*, 2007. **104**(44): p. 17353-17357.
- [6] Bell JK, Yennawar HP, Wright SK, Thompson JR, Viola RE, Banaszak LJ. Structural analyses of a malate dehydrogenase with a variable active site. *Journal of Biological Chemistry*, 2001. **276**(33): p. 31156-31162.

LIST OF ATTENDEES

(February 3rd, 2017)

AMAR Patrick	(patrick.amar@sys2diag.cnrs.fr)
BENTLEY William E.	(bentley@umd.edu)
BERG Gabriele	(gabriele.berg@tugraz.at)
BEURTON-AIMAR Marie	(beurton@labri.fr)
BOUYIOUKOS Costas	(costas.bouyioukos@univ-paris-diderot.fr)
CARBONELL Pablo	(pablo.carbonell@manchester.ac.uk)
CARLSON Ross	(rossc@montana.edu)
CHEVALIER Philippe	(philippe.chevalier@chu-lyon.fr)
CSIKASZ-NAGY Attila	(attila.csikasz-nagy@fmach.it)
DI FILIPPO Marzia	(m.difilippo4@campus.unimib.it)
EBISUYA Miki	(miki.ebisuya@riken.jp)
GUINAUDEAU Ophélie	(guinaude@i3s.unice.fr)
HEUX Stéphanie	(heux@insa-toulouse.fr)
HOLZHUTTER Hermann-Georg	(hermann-georg.holzhuetter@charite.de)
HOREMANS Steff	(steff.horemans@issb.genopole.fr)
ISSA Razanne	(razanne.issa@ibgc.cnrs.fr)
JENKINS Kirsten	(kirsten.jenkins@kcl.ac.uk)
KAUFMAN Marcelle	(mkaufman@ulb.ac.be)
KÉPÈS François	(francois.kepes@issb.genopole.fr)
KOCH Ina	(ina.koch@bioinformatik.uni-frankfurt.de)

LE GALL Pascale	(pascale.legall@issb.genopole.fr)
LE GOUELLEC Audrey	(alegouellec@chu-grenoble.fr)
LOCKE James	(James.Locke@slcu.cam.ac.uk)
MADEC Morgan	(morgan.madec@unistra.fr)
MARKERT Elke	(Elke.Markert@glasgow.ac.uk)
MAZAT Jean-Pierre	(jean-pierre.mazat@phys-mito.u-bordeaux2.fr)
MOULIN Cécile	(moulin@lri.fr)
NASSER William	(william.nasser@insa-lyon.fr)
NORRIS Victor	(victor.norris@univ-rouen.fr)
PALAMA Tony	(palama@insa-toulouse.fr)
PERES Sabine	(sabine.peres@lri.fr)
REVERCHON Sylvie	(Sylvie.reverchon-pescheux@insa-lyon.fr)
RÖHL Annika	(annika.roehl@fu-berlin.de)
ROSATI Elise	(elise.rosati@gmail.com)
SEGRÉ Daniel	(dsegre@bu.edu)
TELHAN Orkan	(otelhan@design.upenn.edu)
TOPARLAK Omer Duhan	(omerduhan.toparlak@unitn.it)
TROSSET Jean-Yves	(jytrosset@gmail.com)
TZONEVA Gannie	(gannie.tzoneva@regeneron.com)
ZELISZEWSKI Dominique	(dominique.zeliszewski@issb.genopole.fr)