

Proceedings of
The Évry Spring School
on

Modelling Complex Biological Systems in the Context of Genomics

May 21st - 25th, 2012

Edited by

Patrick Amar, François Képès, Vic Norris



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

Strohman RC (1977) Nature Biotech 15:199

FOREWORD

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

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

The works of subgroups underpinned the conferences organised in Autrans in 2002, in Dieppe in 2003, in Évry in 2004, in Montpellier in 2005, in Bordeaux in 2006, back to Évry in 2007, in Lille in 2008, in Nice in 2009, in Évry in 2010 and in Sophia-Antipolis in 2011. The conference in Évry in 2012 which as reported here, brought together over a hundred participants, biologists, physical chemists, physicists, statisticians, mathematicians and computer scientists and gave leading specialists the opportunity to address an audience of doctoral and post-doctoral students as well as colleagues from other disciplines.

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

Patrick Amar, Gilles Bernot, Marie Beurton-Aimar, Attila Csikasz-Nagy, Eric Goles, Janine Guespin, Jürgen Jost, Marcelline Kaufman, François Képès, Pascale Le Gall, Ivan Junier, Reinhard Lipowsky, Jean-Pierre Mazat, Victor Norris, William Saurin, El Houssine Snoussi.



ACKNOWLEDGEMENTS

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

Of course the organisation team would like to express gratitude to all the staff of the *IBIS Styles Evry Cathédrale* hotel for the very good conditions we have found during the conference.

Special thanks to the Epigenomics project for their assistance in preparing this book for publication. The cover photography shows a view of University of Évry.

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

They were:

- Centre National de la Recherche Scientifique (CNRS):
<http://www.cnrs.fr>
- Genopole® Évry:
<http://www.genopole.fr>
- GDRE CNRS 513 Biologie Systémique:
http://www.mpi-magdeburg.mpg.de/CNRS_MPG
- Consortium BioIntelligence (OSEO)
- Institut National de Recherche en Informatique et en Automatique (INRIA):
<http://www.inria.fr/>
- GDR CNRS 3003 Bioinformatique Moléculaire:
<http://www.gdr-bim.u-psud.fr>

THE EDITORS



INVITED SPEAKERS

LAURA ADAM,	Virginia Tech, VA (USA)
ZVIA AGUR	Institute for Medical BioMathematics (IL)
GUILLAUME BESLON	INSA Lyon (F)
JORGE CARNEIRO	Gulbenkian Institute, Lisboa (PT)
OLIVIER COLLIN	ENS Paris (F)
DIRK DRASDO	INRIA Rocquencourt (F)
TOM ELLIS	Imperial College London (UK)
JEAN-MARIE FRANÇOIS	LISBP-INSA Toulouse (F)
DIETER W. HEERMANN	Univ. Heidelberg (D)
THOMAS HÖFER	DKFZ Heidelberg (D)
STEFAN KLUMPP	MPI-KG, Potsdam (D)
FRANCIS LÉVI	INSERM, Villejuif (F)
ALBERT LIBCHABER	Rockefeller Univ., New York (USA)
STEPHEN MUGGLETON	Imperial College London (UK)
ROLAND R. NETZ	Technical Univ. Munich (D)
ADRIEN RICHARD	Laboratoire I3S, Sophia Antipolis (F)
PASQUALE STANO	Università degli Studi di Roma Tre, Roma (I)
RODOLPHE THIEBAUT	INSERM U897, Univ. Bordeaux (F)

CONTENTS

PART I INVITED TALKS 13

SYNTHETIC BIOLOGY 1

JEAN-MARIE FRANÇOIS

Alternative to Petrochemistry: Exploiting the biomass for generating chemical synthons by Synthetic Biology 13

TOM ELLIS

Constructing a genome from parts - challenges and opportunities 15

GUILLAUME BESLON

In silico experimental evolution: can we study real evolution with false organisms? 17

SYNTHETIC BIOLOGY 2

PASQUALE STANO

Semi-synthetic minimal cells: relevance in origin of life and synthetic biology 19

STEPHEN MUGGLETON

Developing robust Synthetic Biology designs using automatically chosen microfluidic experiments 21

BIOPHYSICS OF THE CELL

ROLAND R. NETZ

Friction and Hydration Repulsion Between Hydrogen-Bonding Surfaces 23

DIETER W. HEERMANN

The 3D Organization of the Nucleus 25

STEFAN KLUMPP

Economy of molecular machines 27

SYSTEMS IMMUNOLOGY**THOMAS HÖFER**

Dynamic regulatory network govern T-cell proliferation and differentiation 29

JORGE CARNEIRO

To Be Announced 31

RODOLPHE THIEBAUT

Gene expression data for assessing vaccine response: application to HIV vaccine research 31

SYSTEMS ONCOLOGY**FRANCIS LÉVI**

Systems chronopharmacology for the personalization of cancer chronotherapeutics 33

ZVIA AGUR

Systems Oncology for Personalizing Drug Therapy of Cancer Patients: From Theory to Clinical Treatment 35

DIRK DRASDO

Towards predictive quantitative modeling of tissue organization on histological scales: imaging, image analysis and modeling of liver regeneration and growing multi-cellular spheroids 37

PART II 10TH ANNIVERSARY TALKS 39**ALBERT LIBCHABER**

KEYNOTE LECTURE: Possible effects of temperature gradients for the origin of life hypothesis and possible genomic selections 39

ADRIEN RICHARD

Positive and negative feedbacks in discrete models of gene networks 41

PART III	ARTICLES	43
	GILLES BERNOT AND JANINE GUESPIN-MICHEL <i>Modelling Complex Biological Systems in the Context of Genomics: a Decade of History to Savor and Ponder</i>	43
	VIC NORRIS AND PATRICK AMAR <i>Life on the scales: initiation of replication in E. coli</i>	55
	GHISLAIN GANGWE NANA, ARMELLE CABIN, DAVID GIBOUIN, FABRICE LEFEBVRE, ANTHONY DELAUNE, LAU- RENT JANNIERE, CAMILLE RIPOLL AND VIC NORRIS <i>Suitability of Secondary Ion Mass Spectrometry (SIMS) for studies of metabolic heterogeneity</i>	79
	JEAN-PAUL COMET, GILLES BERNOT, APARNA DAS, FRANCINE DIENER, CAMILLE MASSOT AND AMÉLIE CESSIEUX <i>Simplified models for the mammalian circadian clock</i>	85
PART IV	WORKSHOPS	107
	JEAN-PIERRE MAZAT <i>Metabolism Modelling</i>	107
	HÉDI A. SOULA <i>Introduction to Stochastic Simulation Algorithm</i>	121
	FRANÇOIS LE FÈVRE <i>From automatic and expert annotation to the reconstruction of genome-scale metabolic networks and models</i>	123
PART V	POSTERS	143
	ADRIEN BASSO-BLANDIN, FRANCK DELAPLACE <i>GUBS: A rule based behavioral language for synthetic biology</i>	143
	SAYANTANI BASU ROY, FRANCK GAUTHIER, MATTHIEU FALQUE, OLIVIER C. MARTIN <i>Interference among Meiotic Crossovers varies between and within chromosomes in Arabidopsis thaliana</i>	145

EMMANUEL CHAPLAIS, GILLES CHIOCCHIA, MAXIME BREBAN, HENRI-JEAN GARCHON <i>A genetic and metagenomic systems biology approach to elucidate predisposing factors to Spondylarthritis with genes networks</i>	147
APARNA DAS, FRANCINE DIENER, GILLES BERNOT, JEAN-PAUL COMET <i>Simplified mathematical model for the mammalian circadian clock</i>	149
QIAN GAO, DAVID GILBERT, MONIKA HEINER, FEI LIU, DANIELE MACCAGNOLA AND DAVID TREE <i>Multiscale Modelling and Analysis of Planar Cell Polarity in the Drosophila Wing</i>	150
THIBAUT LEPAGE, IVAN JUNIER AND FRANÇOIS KÉPÈS <i>A polymer model for studying long supercoiled DNA molecules</i>	151
ADRIEN HENRY, YULANG LUO, AREEJIT SAMAL AND OLIVIER C. MARTIN <i>Benchmarking gene regulatory networks in silico</i>	153
ABDALLAH ZEMIRLINE <i>Modelling the Cerebral Cortex</i>	154
PART VI LIST OF ATTENDEES	158



Alternative to Petrochemistry: Exploiting the biomass for generating chemical synthons by Synthetic Biology

Jean-Marie François¹

¹ Université de Toulouse, INSA & Toulouse White Biotechnology excellence center
Laboratoire d'Ingénierie des Systèmes Biologiques & procédés,
UMR -CNRS5504 & INRA 792

Abstract

From growing oil prices and the expected shortage of raw materials of fossil origin during the next decades, the fermentation-based processes for (bio)chemical production arises that are based on the exploitation of renewable resources. The product family of amino acids represents a particularly striking example for our growing technological ability to replace petrol-based syntheses of bulk chemicals by production processes. The amino acid methionine and its hydroxyl analog (HMTB), which is key for poultry production, and reaches an annual production volume of ≈ 800.000 tons, is still exclusively produced from petrol. In this context, the development of alternative synthetic pathways for methionine production has become very challenging and of utmost interest. We herein develop a new strategy that combines the capacity of microorganisms to produce highly functionalized carbohydrates with high reaction rates of chemical syntheses by the *ex nihilo* design and implementation of a synthetic pathway to generate a synthon as precursors of many highly relevant biotechnological products in a host microorganism from renewable carbon sources.

This chemical building block (synthon) will then be a platform for the (bio)synthesis of other biotechnological relevant molecules. Scientifically speaking, the presented project derives its innovative character from the fact that the *ex nihilo* implementation of a metabolic pathway represents an unprecedented scientific challenge in the context of bulk chemical production. The commercial output of this project is witnessed by the direct and strong involvement of the industrial partner in order to expand its position in food and animal-nutrition and find biotechnological alternatives to its currently chemical activities. This ongoing ambitious and challenging project gathers complementary experts and skills from different domains, namely enzyme engineering, functional genomics, metabolic engineering, chemical biology, and bioinformatics/biomathematic, creating an interdisciplinary core of expertises which is managed within the new TWB (Toulouse White Biotechnology) R & D Centre of Excellence in Toulouse.



Constructing a genome from parts - challenges and opportunities

Tom Ellis¹

¹ Imperial College, London, United Kingdom

Abstract

The genome is often called the book of life and so to realise DNA-based artificial life we need to start writing our own books. So far synthetic biology has been the lead research field addressing this question. Some researchers have recently succeeded in rewriting a complete copy of their favourite book, adding a few watermarks on the way to prove so (i.e. J.Craig Venter Institute), while most other researchers in this field have instead concentrated on writing extra chapters that add new twists when added to their favourite books (i.e. synthetic biology devices, circuits and pathways). What will it take, therefore, for these researchers to be able to write an entire new book and in doing so produce an artificial cell? This grand challenge of writing a complete genome from parts is one that our group is working towards through a variety of research projects in yeast and *E. coli* model organisms. At the smallest scale, we have begun creating part libraries for predictable gene expression control and are designing new technologies to scale these beyond simple synthetic biology devices. At the largest scale we are engaged in complete synthesis of a chromosome and are investigating the context dependencies associated with genome locus and structure with the goal of determining the hidden rules of writing a book of life.



***In silico* experimental evolution:
can we study real evolution with false organisms?**

Guillaume Beslon¹

¹ IXXI - Institut Rhône-Alpin des Systèmes Complexes, INRIA – LIRIS –
équipe Beagle, INSA-Lyon Computer Science dept., Lyon, France.

Abstract

In the lecture, I will focus on the use of *in silico* models to discover new mechanisms in Darwinian evolution. I will shortly present various *in silico* formalisms and models and show how they can be used to better understand how evolution has shaped the molecular structure of biological organisms.



Semi-synthetic minimal cells: relevance in origin of life and synthetic biology

Pasquale Stano¹

¹ Università degli Studi di Roma Tre, Roma, Italy

Abstract

Semi-synthetic minimal living cell (SSMCs) can be defined as those cell-like compartments - based on lipid vesicles - having the minimal and sufficient number of components to be considered alive [1].

We will discuss the relevance of SSMCs in the origin of life and in synthetic biology. In fact, SSMCs can be considered models of primitive cells, which are characterized by a minimal molecular complexity, yet displaying some living-like properties. On the other hand, the SSMC technology might bring about a novel generation of synthetic biology tools for a variety of applications, for example in drug delivery or novel bio-ICTs approaches.

To date, several research group are actively progressing in designing and constructing simple and complex experimental models - that can be considered as intermediate steps toward the construction of SSMCs. In particular, the current state of the art consists in expressing one or more functional proteins inside liposomes [2]. Here we will discuss the conceptual and the experimental advancements related to the construction of SSMCs in general, and to protein synthesis in particular.

References

- [1] Luisi P. L., Ferri F., Stano P., *Approaches to semi-synthetic minimal cells: a review*. Naturwissenschaften, 2006. **93** 1-13.
- [2] Stano P., Carrara P., Kuruma Y., Souza T., Luisi P. L., *Compartmentalized reactions as a case of soft-matter biotechnology: Synthesis of proteins and nucleic acids inside lipid vesicles*. Journal of Material Chemistry, 2011. **21** 18887-18902



Developing robust Synthetic Biology designs using automatically chosen microfluidic experiments.

Stephen Muggleton¹

¹ Computational Bioinformatics Laboratory Department of Computing Imperial College London, 180 Queen's Gate LONDON SW7 2BZ, United Kingdom

Abstract

Synthetic Biology is an emerging discipline that is providing a conceptual framework for biological engineering based on principles of standardisation, modularity and abstraction. For this approach to achieve the ends of becoming a widely applicable engineering discipline it is critical that the resulting devices are capable of functioning according to a given specification in a robust fashion. This talk describes ongoing work aimed at using experimental validation and revision based on the development of a microfluidic robot scientist to support the empirical testing and automatic revision of robust component and device-level designs. The approach will be based on probabilistic and logical hypotheses generated by active machine learning. A previous paper based on the author's design of a Robot Scientist appeared in Nature and was widely reported in the press. The proposed techniques will extend those in the PI's previous publications in which it was demonstrated that the scientific cycle of hypothesis formation, choice of low-expected cost experiments and the conducting of biological experiments could be implemented in a fully automated closed-loop.



Friction and Hydration Repulsion Between Hydrogen-Bonding Surfaces

Roland R. Netz¹

¹ Fachbereich Physik Freie Universität Berlin, Germany

Abstract

The dynamics and statics of polar surfaces are governed by the hydrogen-bonding network and the interfacial water layer properties. Insight can be gained from all-atomistic simulations with explicit water that reach the experimentally relevant length and time scales. Two connected lines of work will be discussed:

- On surfaces, the friction coefficient of bound peptides is very low on hydrophobic substrates, which is traced back to the presence of a depletion layer between substrate and water that forms a lubrication layer. Conversely, friction forces on hydrophilic substrates are large. A general friction law is presented and describes the dynamics of hydrogen-bonded matter in the viscous limit.
- The so-called hydration repulsion between polar surfaces in water is studied using a novel simulation technique that allows to efficiently determine the interaction pressure at constant water chemical potential. The hydration repulsion is shown to be caused by a mixture of water polarization effects and the desorption of interfacial water.



The 3D Organization of the Nucleus

Dieter W. Heermann¹

¹ Institute for Theoretical Physics University of Heidelberg
Heidelberg, Germany

Abstract

In vivo, chromosomes due to dynamic association with protein factors fold to form three-dimensional structures. Many experiments have paved the way to understand folding and the nuclear architecture of the genome for at least in eucaryotes as well as in procaryotes. Based on these experiments and models a fundamental understanding of the key principles that drive the physical organization has become possible through the concepts of loop and entropy. Using the concept of loop, models are now able to reproduce the results from several of the experiments within the framework of loop and entropy. Linking biological function to structure they moreover are able to make predictions.



Economy of molecular machines

Stefan Klumpp¹

¹ Max Planck Institute of Colloids and Interfaces, Potsdam, Germany

Abstract

Protein-coding genetic sequences are read out in two steps, transcription and translation. Rapid cell growth requires high rates of protein synthesis and therefore fast-growing cells contain larger amounts of the machines that drive these processes, RNA polymerases and ribosomes. The requirement for large concentrations of these machines, which are themselves proteins (or RNA-protein complexes in the case of ribosomes), imposes a number of constraints on the allocation of cellular machinery. These will be discussed for the case of bacteria.



Dynamic regulatory network govern T-cell proliferation and differentiation

Thomas Höfer¹

¹ German Cancer Research Center (DKFZ) and Bioquant Center
Heidelberg, Germany

Abstract

The last decade has seen an enormous advance of our knowledge on the molecules of the immune system and their interactions. By comparison, we still know rather little about the dynamics of the regulatory networks that are formed by the molecular players. We have developed an iterative approach of mathematical modelling and experimentation to dissect the regulatory networks that govern the proliferation and differentiation decisions of T cells. The models address different levels of molecular organization, including the activation of a gene in its chromatin context, the kinetics of gene-regulatory networks and spatio-temporal patterns in intercellular signalling. In this talk, I will report how quantitative and time-resolved analyses of single cells and cell populations have helped to elucidate a core gene network driving T cell differentiation, uncover emergent phenomena that underlie the decision of a cell to proliferate upon antigenic stimulation, and understand the mechanistic basis of stochastic cytokine gene expression



Gene expression data for assessing vaccine response: application to HIV vaccine research

Rodolphe Thiébaud¹

¹ INSERM U897, ISPED, Bordeaux Segalen Univ, Bordeaux, France

Abstract

Gene expression data has already been successfully used for assessing several vaccines such as yellow fever [1] or flu [2]. This has led to the 'system vaccinology' concept [3]. Analysing the change in gene expression after vaccination allows to better understand the biological mechanism leading to a given immune response and may also allow to predict longer term responses. In the case of yellow fever, an unsuspected role of innate immunity has been discovered [3] by this method.

The analysis of gene expression data requires specific methods to take into account specific issues related to the high dimension of the data sets. The inflation of type 1 error due to statistical test multiplicity is well known, making the use of FDR common. Supervised and unsupervised classifications approaches are also usually executed.

Furthermore, in the context of the immunological response to vaccine, the outcome is measured through various different biological markers. Frequency of cell populations is measured by flow cytometry (with up to 12 colours leading to 2^{12} possible populations), functionality is measured by ELISPOT (e.g. Ifn- γ production), intra-cellular cytokine staining (e.g. il2, tnf, ifn), or luminex (e.g. up to 20 cytokines). Hence, data integration techniques for relating several high dimensional matrices are required. This is the same type of tools than those used to relate transcriptomics with proteomics or metabolomics.

In the present paper, we show the application of some of the most recent statistical and bioinformatics tools for the integrated analysis of gene expression data together with immune biomarkers. In particular, we are presenting extensions of gene set analyses and sparse partial least square discrimination analysis (spls - da) / canonical correlation for longitudinal data [4].

These methods have been applied to two different HIV vaccine trials: i) the VRI/ANRS Vac18 trial evaluating three doses of LIPO5 vaccine in healthy volunteers, with a sub study of cytokines and gene expression after in vitro stimulations; and ii) the VRI/ANRS DALIA1 trial evaluating the safety and the

immunogenicity of a dendritic cell-based therapeutic vaccine as a in 19 HIV infected patients who interrupted their antiretroviral treatment. In the latter trial, 17 repeated measurements of gene expression have been performed. Each measurement done by the Illuminate beadship v4, included 47000 probes.

Finally, high throughput biological assays revolutionized the evaluation of new vaccines but need adapted statistical tools.

References

- [1] Querec TD, Akondy RS, Lee EK, Cao W, Nakaya HI, Teuwen D, *et al.*, *Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans*. Nature Immunol., 2009. **10**:116-125.
- [2] Nakaya HI, Wrammert J, Lee EK, Racioppi L, Marie-Kunze S, Haining WN, *et al.*, *Systems biology of vaccination for seasonal influenza in humans*. Nature Immunol. **12**:786-795.
- [3] Pulendran B., *Learning immunology from the yellow fever vaccine: innate immunity to systems vaccinology*. Nat. Rev. Immunol., 2009. **9**:741-747
- [4] Le Cao KA, Martin PG, Robert-Granie C, Besse P., *Sparse canonical methods for biological data integration: application to a cross-platform study*. BMC Bioinformatics, 2009. **10**:34.

Systems chronopharmacology for the personalization of cancer chronotherapeutics

Francis Lévi¹

¹ Laboratoire “Rythmes Biologiques et Cancers”, UMRS776 INSERM et Université Paris Sud, hôpital Paul Brousse, Villejuif, France

Abstract

Chronotherapeutics aims at improving treatment outcomes through the delivery of medicines according to circadian rhythms. Indeed, xenobiotic metabolism and detoxification, as well as cell cycle events, DNA repair, apoptosis and angiogenesis are rhythmically controlled by the Circadian Timing System (CTS). The CTS is a hierarchical network of molecular clocks in each cell. Each molecular clock generates intracellular circadian oscillations in cell functions, as a result of interwoven transcription/translation feedback loops involving 15 clock genes. The cellular circadian clocks are coordinated by the suprachiasmatic nuclei (SCN), the main circadian pacemaker in the hypothalamus, which directly or indirectly generates an array of physiological rhythms. The CTS then make the molecular clocks tick in synchrony in the host tissues that can otherwise be damaged by anticancer agents. As a result, circadian timing can modify 2- to 10-fold the tolerability of anticancer medications in experimental models and in cancer patients [2].

Improved efficacy is also seen when drugs are given near their respective times of best tolerability. Both experimental and clinical data thus support a unique paradigm for cancer therapy: “the lesser the toxicity, the better the efficacy” (*Innominato et al. 2011*). Stochastic and deterministic mathematical models are addressing the threeway interactions between circadian clocks, cell cycle and drug pharmacodynamics at single cell or cell population levels, while integrative models address the issues of therapeutic optimization at whole body level. Several databased chronotherapeutic models not only confirm experimental and clinical chronopharmacology data, but also allow the exploration of many sources of variability, that would otherwise remain hidden (*Altinok et al. ADDR 2007, Eur J Pharm Sci 2009; Clairambault 2010; Bernard et al. Plos Comput Biol 2010; Ballesta et al. Plos Comput Biol 2011*).

Such “systems chronopharmacology” reveal that optimal cancer chronotherapeutics require circadian entrainment to be robust in healthy cells and weak or disrupted in cancer cells. Indeed host clocks are disrupted whenever anticancer drugs are wrongly dosed or timed (*Ahowesso et al. Chronobiology Int 2011*). Circadian disruption is deleterious for cancer control, since it accelerates experimental and clinical cancer progression (*Filipski et al. JNCI*

2002; 2005; Innominato et al *Cancer Res* 2009). Experimental and clinical evidence further show that female patients display more toxicities than males, and may be more prone to treatment-induced circadian disruption (Giacchetti et al. *JCO* 2006, Lévi et al. *ADDR* 2007). Therefore, non-invasive circadian biomarkers (Scully et al. *Interface Focus* 2011) are critical for the modeling of CTS dynamics and the personalization of cancer chronotherapeutics.

Support: C5SYS project, ANR, ERASysBio+ initiative, an EU ERANET in FP7 and ARTBC, Hospital P Brousse, Villejuif (France)

References

- [1] Lévi F., Altinok A., Goldbeter A., *Circadian rhythms and cancer chronotherapeutics*. In “Cancer Systems Biology, Bioinformatics and Medicine: Research and Clinical Applications”, Frederick Marcus & Alfredo Cesario eds, Springer Dordrecht Heidelberg London New York, 2011, Chap. 15, Part 4, 381-407.
- [2] Lévi F., Okyar A., Dulong S., Innominato P. F., Clairambault J., *Circadian Timing in Cancer Treatments*. *Annu. Rev. Pharmacol. Toxicol.*, 2010, **50**:377-421

Systems Oncology for Personalizing Drug Therapy of Cancer Patients: From Theory to Clinical Treatment

Zvia Agur¹

¹ Institute for Medical BioMathematics (IMBM), Israel

Abstract

Theory suggests a universal Resonance Phenomenon, by which differences in the driving periodicities of the susceptible host and pathological tissues can be taken advantage of in optimizing cancer drug therapy. This theory was employed for developing an optimization method for minimizing efficacy/toxicity ratio of cancer drug regimens. Using this method an optimal combination regimen was tailored to a particular mesenchymal chondrosarcoma (MCS) patient. To this end growth curves and gene expression analysis of tumorgrafts, derived from the patient's lung metastasis, served for creating a mathematical model of MCS progression and adapting it to the tumorgraft setting.

The pharmacokinetics and pharmacodynamics of six drugs were modeled, with model variables being adjusted by patient-specific chemosensitivity tests. The xenografted animals were treated by various monotherapy and combination schedules, and the MCS tumorgraft model was computer simulated under the same treatment scenarios. This mathematical model was then up-scaled to retrieve the MCS patient's tumor progression under different treatment schedules. An average accuracy of 87.1% was obtained when comparing model predictions with the observed tumor growth inhibition in the xenografted animals. Simulation results suggested that a regimen containing bevacizumab applied in combination with once-weekly docetaxel would be more efficacious in the MCS patient than all other simulated schedules. Weekly docetaxel in the patient led to stable metastatic disease and relief of pancytopenia. Using this patient's model it was suggested that the relative advantage of "Dose Dense" drug therapy depends on personal cytokinetic and angiogenic parameters.



Towards predictive quantitative modeling of tissue organization on histological scales: imaging, image analysis and modeling of liver regeneration and growing multi-cellular spheroids.

Dirk Drasdo¹

¹ INRIA Rocquencourt, France

Abstract

A major challenge is to extend quantitative modeling towards the tissue scale. This is fundamental to understand how cells coordinately behave to establish functional tissue structure. Research in this field suffers from a lack of techniques that permits quantification of tissue architecture and its development. To bridge this gap we have established a procedure based on confocal laser scans, image processing and three-dimensional tissue reconstruction, as well as on quantitative mathematical modeling [4, 1]. We illustrate our method by two examples: (1) regeneration after toxic liver damage, and (2) growing multi-cellular spheroids. The example of the regenerating liver has been chosen, because liver function depends on the complex micro-architecture formed by hepatocytes (the main type of cells in liver) and micro-vessels (sinusoids) that ensures optimal exchange of metabolites between blood and hepatocytes. Our model of regeneration after toxic damage captures hepatocytes and sinusoids of a liver lobule during the regeneration process. Hepatocytes are modeled as individual agents parameterized by measurable biophysical and cell-biological quantities [2, 3].

Cell migration is mimicked by an equation of motion for each cell subject to cell-cell-, cell-extra-cellular matrix-, and cell-sinusoid-forces, as well as the cell micro-motility. We demonstrate how by iterative application of the above procedure of experiments, image processing and modeling a final model emerged that unambiguously predicted a so far unrecognized order mechanism essential for liver regeneration. A three-dimensional image analysis clearly confirmed the model prediction. In the second example we consider a hybrid model of a multi-cellular tumor spheroid of a non-small-lung cancer cell line demonstrating how by labeling and image analysis the parameters of the model can iteratively be determined. The model is then be used to simulate the growth of the spheroid and shows an excellent agreement with the experimental data. In both studied examples, the iterative comparison of experiments and data permits to exclude potential candidate mechanisms and guide the experiments towards those mechanisms compatible with the experimental observations.

References

- [1] Hoehme, S., Brulport, M., Bauer, A., Bedawy, E., Schormann, W., Gebhardt, R., Zellmer, S., Schwarz, M., Bockamp, E., Timmel, T., G. Hengstler, J.G., and Drasdo, D., *Prediction and validation of cell alignment along microvessels as order principle to restore tissue architecture in liver regeneration*. Proc. Natl. Acad. Sci. (USA), 2010. **107**(23):10371-10376.
- [2] Drasdo, D., Hoehme, S. and Block, M., *On the Role of Physics in the Growth and Pattern Formation of Multi-Cellular Systems: What can we Learn from Individual-Cell Based Models?* Journal of Statistical Physics (2007). **128**(1-2):287-345.
- [3] Block, M., Schoell, E. and Drasdo, D., *Classifying the growth kinetics and surface dynamics in growing cell populations*. Phys. Rev. Lett. (2007). **99**(24):248101-104.
- [4] Hoehme, S., Hengstler J.G., Brulport M., Schäfer M., Bauer A., Gebhardt R. and Drasdo D., *Mathematical modelling of liver regeneration after intoxication with CCl4*. Chemico-Biological Interaction (2007). **168**:74-93.

KEYNOTE LECTURE

Possible effects of temperature gradients for the origin of life hypothesis and possible genomic selections

Albert Libchaber¹

¹ The Rockefeller University, 1230 York Avenue, New York, USA

Abstract

DNA in a temperature gradient: Temperature differences across porous rocks may feed accumulation, replication and size selection of DNA and RNA. Such non-equilibrium conditions near thermal submarine vents are pieces in the fascinating puzzle of the origin of life. Various stages of this scenario will be experimentally presented.



Positive and negative feedbacks in discrete models of gene networks

Adrien Richard¹

¹ Laboratoire I3S, UMR 7271 UNS-CNRS, Université de Nice, 2000, route des Lucioles, B.P. 121, F-06903 Sophia Antipolis cedex, France.

Abstract

René Thomas stated in the 80's that the presence of positive (resp. negative) feedbacks in the interaction graph of a gene network is a necessary condition for the presence of multiple stable states (resp. sustained oscillations) in the dynamics of the network. In this presentation, we show how these two general rules can be turned into formal theorems when gene networks are modeled by discrete dynamical systems. We also present recent generalisations that highlight the influence of connections between feedbacks.



Modélisation de systèmes biologiques complexes dans le contexte de la génomique : dix ans d'une histoire à savourer et à méditer

Modelling Complex Biological Systems in the Context of Genomics : a Decade of History to Savor and Ponder

Gilles Bernot^{1,3} et Janine Guespin-Michel²

¹ University Nice-Sophia Antipolis, Laboratoire I3S - CNRS UMR 7271, 2000, route des Lucioles, B.P. 121, F-06903 Sophia Antipolis CEDEX, France

² professeure émérite, Université de Rouen, laboratoire de microbiologie - Signaux et MicroEnvironnement, EA4312 IUT, F-27000 Evreux, France

³ Epigenomics Project, Genopole[®], F-91000 Evry, France

Résumé

C'est l'histoire d'une entreprise interdisciplinaire suscitée par Genopole[®] autour des systèmes biologiques complexes dans le contexte de la génomique. Elle s'est structurée d'abord autour d'ateliers interdisciplinaires, aussi libres et joyeux qu'actifs et fructueux, dont les thèmes de travail ont émergé des interactions entre les participants. De ces ateliers, et en forte interaction avec eux, sont nées des écoles annuelles « *Modelling Complex Biological Systems in the Context of Genomics* », il y a tout juste 10 ans (celle-ci est donc la onzième !). Une rétrospective s'imposait, sur les acquis de ces écoles, ce qui s'y est forgé, ce qui s'est transformé, ce qu'elles sont devenues.

Summary

This is the story (and the history) of an interdisciplinary endeavor promoted by Genopole[®], concerning the complex biological systems in the context of genomics. It was structured primarily around interdisciplinary workshops, as free and happy as they were active and fruitful, whose work themes emerged from the interactions between participants. 10 years ago, from these workshops, and strongly interacting with them, were born the annual schools on *Modelling Complex Biological Systems in the Context of Genomics* (this one being therefore the eleventh one !). A retrospective had to be done, on the achievements of these schools, what has been created, what changed, and what they became.

1 *Les débuts : les ateliers et le Programme d'Épigénomique*

Lorsque Genopole[®] a été créée à Évry en 1998, son directeur, Pierre Tambourin, était persuadé que l'interdisciplinarité pourrait jouer un rôle clef pour la biologie dans le contexte de la génomique. Il a très vite su imposer un état d'esprit scientifiquement ouvert où chaque discipline se savait respectée, utile, libre de développer ses propres fondements avec une aide très réactive de Genopole[®] pourvu que les collaborations inter-disciplines soient effectives et très actives. Il n'est sans doute pas exagéré de dire que cette confiance mutuelle entre disciplines a été le terreau qui a permis à la France d'acquérir un excellent rayonnement mondial dans nos domaines.

Pour accélérer le processus de découverte, Pierre Tambourin, en collaboration avec l'IHES et l'Université d'Évry, a souhaité la mise en place rapide d'un institut pluridisciplinaire dont la feuille de route était :

- de créer un centre de vie scientifique intense favorisant des rencontres transdisciplinaires de haut niveau « autour » de la modélisation en post-génomique ;
- de créer un réseau, d'abord national puis européen, qui favorise l'émergence d'idées nouvelles en attirant de très bons scientifiques motivés, n'ayant pas nécessairement la moindre connaissance en biologie ou en modélisation, quitte à les former rapidement ;
- d'assurer une animation scientifique locale au bénéfice des laboratoires du site ;
- d'attirer et d'incuber de jeunes équipes à destination de ces laboratoires.

Il a fait appel à François Képès pour monter cet institut et l'on peut *a posteriori* distinguer trois phases de son développement : la création des « ateliers » francophones par François Képès et Paul Bourguine, la création à proprement parler de l'institut appelé *Le Programme d'Épigénomique* co-dirigé par François Képès et Gilles Bernot, et enfin son internationalisation rapide vers *The Epigenomics Project* (une quatrième phase est en cours avec des collaborations industrielles fortes comme on le verra plus loin). Ce sont les ateliers qui ont permis de donner naissance en 2002 aux écoles thématiques annuelles dont cette école, 10 ans après, est donc la onzième édition.

De ce soutien fort de Genopole[®] est donc née l'idée « un peu folle » pour l'époque de rassembler, dans la durée, des chercheurs en biologie, en informatique, en mathématiques, ainsi qu'en physique et chimie théoriques, sans imposer de quelque façon que ce soit qu'il aient une compétence préalable en bio-informatique au sens large. Genopole[®] les a fait travailler ensemble, sans programme préalable imposé, sans appel à projet, et sans autre « livrables » que de se rencontrer et collaborer activement. Une vingtaine de chercheurs venant de plusieurs disciplines et de plusieurs régions de France et

de Belgique ont ainsi pu se rencontrer librement de façon régulière à Évry : ainsi sont nés les ateliers. Le Programme d'Épigénomique a repris les ateliers sous la responsabilité de Victor Norris et cherché à favoriser systématiquement les idées nouvelles et les nouvelles façons de travailler. Nous avons coutume de dire « L'objet d'expérience, c'est nous » car nous expérimentions comment accélérer le processus de découverte grâce aux chocs des cultures.

Sans appels d'offres, mais non sans projet : On était au début de ce que l'on a appelé la post-génomique. Les technologies à haut rendement se développaient rapidement, et les données s'accumulaient à un rythme vertigineux. Données pour faire quoi ? Devant cette accumulation, beaucoup pensaient qu'elles parleraient d'elles mêmes. L'accumulation suffirait, on verrait bien ce qu'il en émergerait. A l'opposé, les ateliers du Programme d'Épigénomique se sont bâtis sur l'idée que, quelles que soient les performances technologiques, la science progresse toujours par questions et réponses ou réfutations, par hypothèses et théories. Ainsi l'accumulation fantastique des données requiert des méthodes de traitement appropriées, des méthodes nouvelles pour générer les hypothèses, et la modélisation et la simulation peuvent servir à progresser scientifiquement dans cette optique.

Qui plus est, l'accumulation des données ne devait pas masquer (ou remplacer) la nature complexe des systèmes biologiques qui, à elle seule, par les questionnements dynamiques (dynamiques non linéaires) qu'elle nécessite, justifie le recours à la modélisation et donc la nécessité de développer les techniques mathématique ou informatiques. À la base de tout cela se trouve la biologie, la physiologie des être vivants qui ne se résument pas à leur génome de sorte que la modélisation remet la question biologique au cœur des efforts de recherche en guidant le choix des expériences à mener et des données les plus utiles à accumuler.

L'avant propos des actes de la première école thématique d'Autrans est d'ailleurs explicite :

What are the salient features of the new scientific context within which biological modeling and simulation will evolve from now on ? The global project of high-throughput biology may be summarized as follows. After genome sequencing comes the annotation by "classical" bioinformatics means. It then becomes important to interpret the annotations, to understand the interactions between biological functions, to predict the outcome of perturbations, while incorporating the results from post-genomics studies (of course, sequencing and annotation do not stop when simulation comes into the picture). At that stage, a tight interplay between model, simulation and bench experimentation is crucial.

2 *L'apprentissage de la transdisciplinarité*

La première année, en 2000, il était évident pour Genopole® qu'il fallait faire vite car plusieurs pays se dotaient d'instituts des sciences de la complexité presque dans chaque université, à proximité immédiate des laboratoires. Un premier et unique atelier « généraliste » mélangeant biologistes, informaticiens et mathématiciens fut créé où chacun a raconté aux autres ce qu'il faisait. Il s'agissait de mettre en commun des concepts et du vocabulaire, de bâtir une culture commune. Cela n'a pu se faire que parce que l'ambiance totalement libre a permis à chacun de pouvoir dire « Je ne comprends pas » ou bien « Si je le redis de telle façon avec mes mots, sommes nous toujours d'accord ? ». On prenait le temps qu'il fallait, un exposé prévu pour 30 minutes pouvait durer la matinée, mais on avançait vraiment ensemble.

Nous avons alors constaté que, si nous partagions parfois du vocabulaire entre disciplines, nous n'en partagions pas la sémantique et notre pire ennemi était le contresens. Les schémas de pensée aussi étaient très différents. Pour ne citer qu'un exemple, un informaticien ou un mathématicien veut qu'une affirmation soit *vraie* ou *fausse* alors que pour un biologiste *ça dépend* ; ainsi les uns ont dû admettre que tout raisonnement se fait dans un contexte souvent trop riche pour être entièrement décrit, les autres qu'il faut expliciter de nombreux détails pour modéliser. En 2001-2002, les apports de la modélisation et l'indispensable « retour paillasse » sont devenus des questions très claires dans le contexte de la génomique et finalement, l'atelier a dépassé très vite le cap culturel grâce à la totale liberté des sujets abordés.

Puis est venu le temps de la co-élaboration. Quelques projets ont émergé, et ce premier atelier s'est structuré en 4 ateliers, la plupart des membres participant à 2 ateliers au moins. Ces ateliers différaient par le thème, comme par la manière de travailler. Certains ont perduré, d'autres ont été plus éphémères. Puis avec l'arrivée du Programme d'Épigénomique, d'autres ateliers encore se sont créés ultérieurement, le réseau se renforçait toujours, attirant de nouvelles recrues, des doctorants notamment.

Quelques règles de bonne conduite se sont élaborées pour qu'un atelier existe : véritable interdisciplinarité garantie par des animateurs motivés de plusieurs disciplines, présence de jeunes chercheurs, aucune discipline n'est au service d'une autre, la coopération doit permettre à chacune de faire des recherches originales dans son domaine, un atelier peu actif disparaît de lui-même, *etc.* Il est vite apparu par exemple que la construction collective d'un modèle pouvait devenir une étape extrêmement riche de la recherche de chacun, générant de nombreuses questions inattendues. Cependant aucune contrainte scientifique autre que le contexte de la génomique n'a jamais été édictée et aucun « tableau de bord d'évaluation objectif » n'a jamais été imposé.

Mieux : cela fut clairement à l'origine du succès ; si de nombreuses publications communes ont été, au cours des années, le fruit de ces travaux (environ 70 publications par an faites par des co-auteurs qui ne se sont rencontrés que grâce au Programme d'Épigénomique), elles n'en étaient pas l'objectif prioritaire. Ils s'agissait d'abord de parvenir à travailler ensemble, quel que soit le temps nécessaire, le pari étant que cela ferait émerger des connaissances, et ce pari a été gagné.

La liberté de recherche était le maître mot de ces ateliers, et de leurs succès. Cette optique déjà mentionnée « d'expérience sur nous-même » imposait une structuration des ateliers totalement souple : la fin d'un atelier ne pose aucun problème puisqu'un chercheur participe à plusieurs, de nouveaux domaines peuvent être explorés puis abandonnés, de sorte qu'*in fine* cette « sélection naturelle des ateliers » a permis d'atteindre une couverture assez complète *et une définition* des différents domaines de la modélisation en biologie dans le contexte de la génomique, avec ouverture de nouveaux domaines comme par exemple la biologie de synthèse.

3 L'école d'Autrans en 2002

Fort de ses quatre ateliers, mais néanmoins encore au tout début de la mission que lui avait assignée Genopole[®] le Programme d'Épigénomique se devait d'élargir ce réseau de recherche, attirer des jeunes chercheurs, et convaincre des chercheurs monodisciplinaires expérimentés de l'intérêt de notre domaine scientifique. C'est alors qu'une autre « idée folle » a germé, initier une *école* à partir de ces ateliers. Outre Genopole[®], le CNRS a apporté un soutien sans faille à cette idée (et ce soutien ne s'est jamais essouffé depuis 11 ans). De même les « aînés » plus généralistes du Programme d'Épigénomique, la SFBT et le GDR BIM (alors intitulé IMPG) nous ont soutenus de sorte que la première école a eu lieu il y a 10 ans à Autrans, en mars 2002.

L'école d'Autrans a réuni, autour de la vingtaine des membres des ateliers de l'époque, 40 nouveaux chercheurs et étudiants, dont beaucoup ont par la suite participé à leur tour aux ateliers. L'école fut construite autour d'exposés de une à deux heures le matin et en début d'après midi, suivis de la présentations/discussion du travail des ateliers. Bien qu'il ne soit pas question ici de présenter le contenu des exposés, il paraît important d'insister sur 3 points gages du succès.

- Ces présentations ont couvert un champ très large de disciplines, depuis la biophysique jusqu'à la génomique, depuis la modélisation des réseaux de régulation génétiques, jusqu'aux modélisations multi-échelles, en accord avec l'idée qu'il faut *partir des êtres vivants* et non de leur seul génome, et qu'il faut *aborder les questions sous tous les angles*, et non

pas sous le seul angle de l'analyse calculatoire des données.

- La multidisciplinarité des interventions était donc grande tout en restant orientées vers des sciences de la nature, et à cette époque, c'était très nouveau. Une condition importante à cette compréhension mutuelle a été *l'utilisation de la langue française*, non seulement pour les jeunes chercheurs mais aussi pour tous en raison de la subtilité et de l'étrangeté des idées et des méthodes venant des autres disciplines.
- Enfin, la liberté de recherche et l'enthousiasme des nouveautés explorées par les ateliers ont communiqué à l'école une *atmosphère de travail joyeux*. La volonté de s'entre-découvrir et de comprendre, et surtout *l'enthousiasme suscité par cette démarche nouvelle* a gagné tous les participants, notamment des jeunes. Cette ambiance extrêmement constructive a touché tous les participants, qu'ils aient déjà été convaincus de l'intérêt de la démarche où qu'il l'aient découverte à cette occasion et ceci a été un facteur important de réussite de l'expérience.

À cela il convient aussi d'ajouter les actes, dont la publication chaque année sous forme de livre largement distribué a contribué à donner une stature et une pérennisation à l'entreprise (sans doute d'abord en dépit, puis certainement ensuite grâce, au fait qu'ils ont été en anglais depuis le début). Avec l'aide de Genopole[®] l'école édite de l'ordre de 600 exemplaires de son livre chaque année. Beaucoup d'éditions sont maintenant totalement épuisées mais on peut télécharger une version PDF des années antérieures sur le site web de l'école [12].

Toujours dans l'optique d'attirer des chercheurs confirmés dans notre domaine émergent et prometteur, le Programme d'Épigénomique a mis en place presque en même temps une semaine annuelle de *cours d'introduction avancée à la biologie* : l'objectif, contre toute attente, n'était pas réellement d'enseigner la biologie mais de faire savoir au travers d'exemples ce qu'un biologiste considère comme un résultat de recherche. Au fil du temps, cet objectif a en partie été intégré aux éditions annuelles de l'école thématique.

4 Dix écoles passées, un acquis mais toujours pas la routine

Après Autrans, *la démarche s'est poursuivie*, avec un travail de fond assuré par des ateliers libres et enthousiastes, et l'école thématique annuelle permettant à la fois :

- de divulguer ou enseigner les résultats des ateliers de manière plus formelle, de sorte que les écoles sont des lieux d'apprentissage pour de nouveaux venus et les jeunes chercheurs
- et d'enrichir les travaux de tous les participants par des invités souvent prestigieux et venant d'horizons très différents ; à cet égard, l'école est

un lieu de remise en cause parfois violente (au sens figuré bien sûr) des idées ou des démarches habituelles des participants et certains exposés ont été la source de recherches pluridisciplinaires qui sont devenues de vraies références en biologie des systèmes.

Une salle de réunion en libre accès est d'ailleurs prévue pour faciliter le démarrage de nouvelles collaborations.

Le rapprochement de certaines démarches a parfois pris des allures d'évènement. On peut par exemple citer l'école de Lille organisée par l'Institut de Recherche Interdisciplinaire, où, entre autres, le sujet de l'imagerie pour la biologie, peu ou pas abordé jusqu'alors dans nos écoles, a suscité un enthousiasme fort, catalysé par Bernard Vandebunder.

Grâce à ce mélange systématique des disciplines et des approches, non seulement de nouveaux sujets de recherche émergent, mais également la remise en cause de dogmes devient presque « habituelle » en raison d'inconsistances révélées par des modélisations fondées sur des lois bien établies venant d'autres disciplines.

Pour que cette histoire ne tourne pas au dithyrambe, notons cependant un échec dans l'interdisciplinarité, lors de quelques essais d'introduire des sciences humaines, (histoire ou philosophie des sciences) sous forme de « présentations du soir », tables rondes ou discussions. Ces thèmes ont suscité l'intérêt des participants mais plus en tant que « culture » qu'en tant que coopérations interdisciplinaires plus larges. Ces essais ont d'ailleurs été ensuite abandonnés.

Au cours du temps sont venus se greffer nombre de nouveaux participants, amenant aussi de nouveaux thèmes d'intérêt. Il fallut plus tard jusqu'à trois jours consécutifs tous les deux mois, et pas moins de quatre ou cinq salles de réunion utilisées simultanément, pour permettre à chacun de participer aux ateliers de son choix. Déjà en 2004, lors de la troisième école, près de 250 chercheurs gravitaient autour de ce que l'on peut nommer *un processus*. Processus parce qu'il a su éviter les répétitions à l'identique, innover, se développer et se reconfigurer dynamiquement autour des écoles thématiques.

Dès 2004, il a fallu limiter volontairement le nombre de participants à moins de 100 afin de préserver une ambiance d'école fructueuse et ne pas dériver vers une conférence du domaine. Certaines années cela n'a vraiment pas été facile, tant le succès est grand.

Dans le même temps, le Programme d'Épigénomique accompagnait, cadrait et favorisait le développement de cette communauté européenne naissante. On peut par exemple mentionner les *Think Tanks*, fondés sur l'invitation de chercheurs d'origine diverses, ayant carte blanche pour discuter ensemble pendant plusieurs jours autour d'un thème donné.

Le choix du lieu des écoles successives a dans un premier temps été fondé sur une alternance entre les régions françaises les plus actives du domaine : Autrans en 2002 [1], Dieppe en 2003 [2], Évry en 2004 [3], Montpellier en 2005 [4], Bordeaux en 2006 [5], Évry en 2007 [6], Lille en 2008 [7], Nice en 2009 [8]. Par la suite, avec le rapprochement du consortium BioIntelligence, il a été convenu d'une alternance entre Genopole[®] à Evry et Sophia Antipolis, lieu d'intense activité de ce consortium : Évry en 2010 [9], Sophia Antipolis en 2011 [10], de nouveau Évry en 2012 [11]...

Dans le cadre de l'école, il est important de citer aussi une initiative originale et fructueuse, suggérée originellement par le CNRS et appelée *les tutoriaux*. Il s'agit de comparer des méthodes différentes de modélisation pour un même processus biologique. Le premier exemple fut l'infection de la bactérie *Escherichia coli* par le phage lambda avec ses deux issues possibles, lyse ou lysogénie, concomitantes dans la population. Les tutoriaux représentent un lourd investissement pédagogique et ils se sont mis en place sur plusieurs années sous la houlette de Jean-Pierre Mazat ; après ouverture d'une approche de modélisation une année, ce sont souvent des doctorants qui se chargent des présentations avec beaucoup de succès. C'est vraiment la meilleure manière de montrer dans la pratique, à la fois que plusieurs modèles peuvent rendre compte du même processus, éventuellement sous des facettes différentes, et que chaque modèle permet soit de répondre plus spécifiquement à une question donnée, soit de mieux correspondre à certaines possibilités expérimentales de quantification.

5 Et maintenant...

... cette gageure un peu folle a réussi. Il y a eu une école chaque année et voilà la onzième. Bien sûr, elles ont beaucoup changé au cours du temps, prenant de l'importance, attirant des chercheurs de plus en plus prestigieux, et s'internationalisant... abandonnant depuis longtemps la primeur de la langue française dans les couloirs, langue dont ce chapitre sera sans doute la dernière réminiscence ici (mais nous lui devons bien ça).

Les écoles ont pris une certaine autonomie et beaucoup d'envergure. Les liens très forts entre les ateliers et l'école ont laissé la place depuis 2010 à des liens non moins forts avec le consortium BioIntelligence.

Il faut ici rappeler qu'il y a 10 ans, également à Genopole[®], d'autres pionniers travaillaient sur une autre frontière, elle aussi pluridisciplinaire, pour obtenir et gérer proprement les données à haut débit. On doit citer bien sûr le *Centre National du Séquençage* (le CNS, aussi connu sous le nom de Génoscope) et l'entreprise *Genset*, tous deux programmes de grand séquençage.

Il est remarquable que ces deux frontières se rejoignent aujourd'hui avec des applications industrielles fructueuses. En effet, certains de ces mêmes pionniers scientifiques travaillent activement au transfert vers l'industrie des idées développées durant ces 10 ans d'écoles thématiques. Ils sont à l'origine du consortium *BioIntelligence* qui regroupe six grandes industries (Dassault Systèmes, Ipsen, Pierre Fabre, Sanofi, Servier et Bayer), deux PME (Sobios et Aureus Pharma), ainsi que Genopole[®], l'INRIA et l'INSERM. Ce consortium participe fortement à l'organisation de nos écoles depuis plusieurs années et en a fait l'une de ses manifestations importantes.

D'autres ateliers ou groupes de travail se forment maintenant un peu partout en Europe et peuvent bénéficier des acquis de cette belle aventure. Beaucoup d'entre eux envoient leurs jeunes chercheurs (et aussi moins jeunes...) participer à ces écoles qui sont maintenant une véritable institution.

Cette « installation dans la durée » de l'école thématique ne signifie pas pour autant que la communauté perde de sa capacité d'innovation. De nouveaux challenges arrivent et deviennent importants pour le monde industriel : la médecine personnalisée, les plateformes logicielles offrant de véritables boîtes à outils de prédiction *in silico* des comportements cellulaires, tissulaires, etc, des systèmes d'information capables d'assister ces plateformes logicielles avec des quantités énormes de données fiables et validées, et bien d'autres applications auxquelles nous n'osions même pas imaginer en 2002 qu'elles seraient accessibles seulement 10 ans après !

Cette réussite a permis un véritable transfert d'idées vers l'industrie, objectif permanent de Genopole[®]. Pourtant, l'approche a été presque aux antipodes des méthodes actuellement classiques d'incitation à la recherche appliquée :

- la recherche *fondamentale* a été systématiquement soutenue et le brassage d'idées a été lourdement encouragé à ce niveau ;
- les sujets de recherche ont été non seulement totalement *libres* mais également non planifiés : ce qui a été soigneusement planifié en revanche, c'est la multiplication des occasions de découvertes originales ;
- aucun critère d'évaluation n'a jamais été figé hormis un bouillonnement d'idées obligatoire et une activité forte, pluridisciplinaire, drainant de jeunes chercheurs et conduite par plusieurs chercheurs expérimentés d'horizons scientifiques différents ;
- aucun financement n'a été planifié *a priori* sur un projet déposé avec quelque garantie de réussite que ce soit ;
- aucun « livrable », aucune « time table » n'a jamais envahi le temps des chercheurs impliqués, seulement un objectif simple mais incontournable d'une activité de recherche intense ;

Comment s'opère alors le transfert des idées ? En premier lieu, il faut souligner le dynamisme des partenaires de BioIntelligence et la volonté d'ouverture reconnue de Genopole[®]. Le programme BioIntelligence entend transcrire, pour la R&D biologique et pharmaceutique, les techniques du « Product Lifecycle Management », déjà utilisées dans de nombreux secteurs industriels manufacturiers [13]. Cette volonté nécessite certaines fondations théoriques et technologiques qui, dans ce cadre biologique, sont bien différentes des autres secteurs et ne peut pas être directement dérivée des technologies existantes. Les fondateurs du consortium se sont naturellement tournés vers Genopole[®] pour établir les collaborations de recherche nécessaires à ce projet ambitieux et le Programme d'Épigénomique a été séduit par cette nouvelle aventure.

Il est bien difficile de prédire de quoi sera fait l'avenir et vers quels sujets encore et toujours nouveaux se tournera l'école thématique. On peut peut-être mentionner que, partant de méthodes relevant de la biologie des systèmes, une partie non négligeable des forces de recherche se tourne vers la biologie de synthèse (appelée aussi la biologie synthétique). Encore un domaine complètement nouveau pour lequel on peut faire un parallèle avec la chimie de synthèse qui a fait la « gloire » de quelques uns de nos grand-parents : faire produire au vivant des molécules d'intérêt, par exemple en favorisant artificiellement certaines voies métaboliques, et cette fois ci sans rejets toxiques dans l'environnement. Rendez-vous dans 10 ans, en espérant peut être que l'interdisciplinarité aura réussi d'ici là à inclure les disciplines de sciences humaines, telles que l'éthique, la sociologie et la philosophie des sciences !

Références

- [1] Proc. of the Autrans seminar on Modelling and simulation of biological processes in the context of genomics, March 18th-21st 2002, Amar, Norris, Képès, Norris, Tracqui Ed., Genopole.
- [2] Proc. of the Dieppe spring school on Modelling and simulation of biological processes in the context of genomics, 12-16 may 2003, Amar, Képès, Norris and Tracqui Ed., Platypus Press, ISBN 2-84704-036-6.
- [3] Proc. of the Evry spring school on Modelling and simulation of biological processes in the context of genomics, March 29th - April 2nd 2004, Amar, Comet, Képès and Norris Ed., Platypus Press, ISBN 2-84704-0374.
- [4] Proc. of the Montpellier Spring School on Modelling Complex Biological Systems in the Context of Genomics, Amar, Képès, Molina and Norris Ed., EDP sciences pub., April 4th-8th 2005, ISBN 2-86883-947-9.

- [5] Proc. of the Bordeaux Spring School on Modelling Complex Biological Systems in the Context of Genomics, Amar, Képès, Norris, Beurton-Aimar and Mazat Ed., EDP sciences pub., April 3th-7th 2006, ISBN 978-2-7598-0014-8.
- [6] Proc. of the Evry Spring School on Modelling Complex Biological Systems in the Context of Genomics, Amar, Bernot, Képès and Norris Ed., EDP sciences pub., Evry, April 30th - May 4th, 2007, ISBN 978-2-7598-0019-3.
- [7] Proc. of the Lille Spring School on Modelling Complex Biological Systems in the Context of Genomics, April 7th-11th, 2008, Amar, Képès, Norris and Vandebunder Ed., EDP sciences pub., ISBN 978-2-7598-0075-9.
- [8] Proc. of the Nice Spring school on Modelling Complex Biological Systems in the Context of Genomics, March 30th-April 3rd, 2009, Amar, Képès, Norris and Bernot Ed., EDP Sciences pub., ISBN 978-2-7598-0437-5.
- [9] Proc. of the Evry Spring school on Modelling Complex Biological Systems in the Context of Genomics, March 3rd-7th, 2010, Amar, Képès and Norris Ed., EDP Sciences pub., ISBN 978-2-7598-0545-7.
- [10] Proc. of the Sophia Antipolis Spring school on Modelling Complex Biological Systems in the Context of Genomics, May 23rd-27th 2011, Amar, Képès and Norris Ed., EDP Sciences pub., ISBN 978-2-7598-0644-7.
- [11] Le livre que vous tenez probablement entre les mains en ce moment même...
- [12] <http://epigenomique.free.fr/en/>
- [13] J. Stark *Product Lifecycle Management : 21st century Paradigm for Product Realisation*. 1st Edition, 2004, Springer, ISBN 978-1-85233-810-7.



Life on the scales: initiation of replication in *Escherichia coli*

Vic Norris¹ and Patrick Amar^{2,3}

¹ Theoretical Biology Unit, EA 3829, Department of Biology, University of Rouen
F-76821 Mont Saint Aignan, France

² Laboratoire de Recherche en Informatique, Université Paris-Sud, CNRS UMR 8623
F-91405 Orsay Cedex, France

³ Équipe-projet AMIB, INRIA-Saclay, France

Abstract

At all levels, Life exists on the scales of equilibria. At the level of bacteria, cells either take risks to grow or avoid risks to survive and, in the Dualism hypothesis, it is proposed that these two strategies depend on non-equilibrium and equilibrium hyperstructures, respectively. It is further proposed that the cell cycle itself is the way cells manage to balance the ratios of these types of hyperstructure (corresponding to living on the two scales). Here, we attempt to reinterpret a major event, the initiation of chromosome replication in *E. coli*, in the light of 'scales of equilibria'. This entails thinking in terms of hyperstructures as responsible for intensity sensing and quantity sensing. We outline an automaton approach to the cell cycle that should test and refine the scales concept.

1 Introduction

Living systems are often forced to make a choice: either to interact, take risks and grow or to shut themselves away, hunker down and survive. Living systems are also highly structured. Indeed, in our "ecosystems-first", origins-of-life scenario, the most fundamental of living systems, cells, were highly structured right from the start [1, 2]. We have argued that this structuring takes the form of extensive macromolecular assemblies, termed hyperstructures. Evidence for hyperstructures in both eukaryotes [3] and prokaryotes is steadily accumulating [4, 5, 6, 7]. In bacteria, hyperstructures can be partly described by their position on an equilibrium non-equilibrium axis [4]. Equilibrium (or more exactly, quasi-equilibrium) hyperstructures, which remain stable despite interruptions in the flow of energy or nutrients, allow cells to survive starvation and stresses whilst non-equilibrium hyperstructures, which fall apart as a result of such interruptions, allow growth. The existence of these types of hyperstructures helps, in part, explain how cells overcome one of the major constraints on their evolution, namely that they must be able to

both grow in heaven and survive in hell. This explanation is that a bacterial population comprises cells with different combinations of non-equilibrium and equilibrium hyperstructures which give rise to phenotypes [8]. This gives rise to another question how do cells balance these types of hyperstructures? Put differently and in a scale-free way, this question boils down to 'How do living systems exist on the scales of equilibria?'

At the level of cells, a possible answer to this question involves the cell cycle itself as outlined in the Dualism hypothesis [8]. If the scales of equilibria approach is really fundamental, it should offer a unifying framework for understanding the cell cycle of organisms as different as *Homo sapiens* and *Escherichia coli* and perhaps further still from E.c. to E.T. Here, we re-examine one of the best understood of all systems, the initiation of chromosome replication in *E. coli*, from the standpoint of life on the scales. We make experimental predictions and, given the value of related, automaton approaches to the cell cycle [9], we suggest a simulation approach based on genetic algorithms and a hybrid automaton, HSIM.

2 Problems with current explanations of initiation control

Recent models for initiation have been based on the proportion of the DnaA 'initiator' protein in the ATP-DnaA form, which is active in initiation, or the ADP-DnaA form, which is not. There are several hypotheses to explain how the level of DnaA might increase cyclically to trigger initiation [10]. Lipids have long been known to be involved in initiation [11] and one possibility is that anionic phospholipids might promote conversion from ADP-DnaA to ATP-DnaA *in vivo* as *in vitro* [12]. Another, complementary, possibility is that specific DNA sequences promote this conversion via the formation of homomultimers of the DnaA protein [13]. Following initiation, removal of DnaA may also be achieved by DnaA binding to sequences in the *datA* locus [14]. These proposed mechanisms do not answer the fundamental question of how an increase ATP-DnaA levels would be timed to trigger initiation.

Models for explaining initiation mass must contend with variations in the initiation mass at different growth rates, the complications arising from artificially lengthening the elongation of replication, the absence of effect of the presence of over a score of minichromosomes (that replicate in synchrony with the origin of the chromosome itself), the decrease in the mass of DNA (but unaltered rates of growth and division) when *dnaAts* mutants are grown at semi-permissive temperatures, and the apparently normal timing of initiation of replication of minichromosomes in a strain in which the chromosomal *oriC* was inactive (for references see [8]). Most seriously of all, perhaps, the present of an extra functioning *oriC* in the *E. coli* chromosome does not apparently perturb the cell cycle [15].

3 The Dualism hypothesis

Many, perhaps all, living systems are obliged, at some time in their existence, to balance the requirement to grow with the requirement to survive. Without growth, the proportion of the niche that they fill would decrease (even if they reproduce). Without survival, the game is over (even if conditions permit growth). The fundamental problem is the requirements for growth and survival are opposed. Growth cannot occur without non-equilibrium structures. Survival, particularly in conditions of deprivation, cannot occur without equilibrium (more exactly, quasi-equilibrium) structures. The fundamental question is how do living systems reconcile these contradictory demands. To put it in a picturesque way, how do they manage to live on the scales of equilibria?

The archetypal, terrestrial living system is the cell. Cells contain both non-equilibrium and equilibrium structures or, more precisely, hyperstructures. We propose that the answer to the fundamental question for cells involves them balancing their non-equilibrium and equilibrium constituents. In our hypothesis, this is achieved by the cell cycle which is essentially about sensing the NE:E ratio. The cell cycle does this in two ways (Figure 1):

1. Intensity. The intensity of use of non-equilibrium hyperstructures is sensed and when this is sufficiently high for growth to be limited, a signal is emitted.
2. Quantity. The quantity of equilibrium hyperstructures is sensed and when this is sufficient for daughter cells to be generated, a signal is emitted. It is of likely evolutionary advantage that a quantity sensing mechanism exists to ensure that before a cell embarks on a new cell cycle it has sufficient quantities of membrane, energy and nucleotides to get it through to the end (cell division) even if there is a disaster.

Both intensity and quantity signals are needed for a cell to begin a new cycle and initiate the production of daughter cells. The physical nature of these signals may be different in different living systems.

4 Mechanisms

The two classes of cell cycle signals are generated by two different sets of mechanisms. Here, we give examples of candidate mechanisms.

4.1 Intensity sensing

It can be of evolutionary advantage for a cell to grow (and reproduce) as rapidly as possible in favourable conditions. Such growth involves RNA polymerases transcribing genes to make mRNA, and ribosomes translating mRNA to make proteins. There is a finite minimum distance between two RNA polymerases

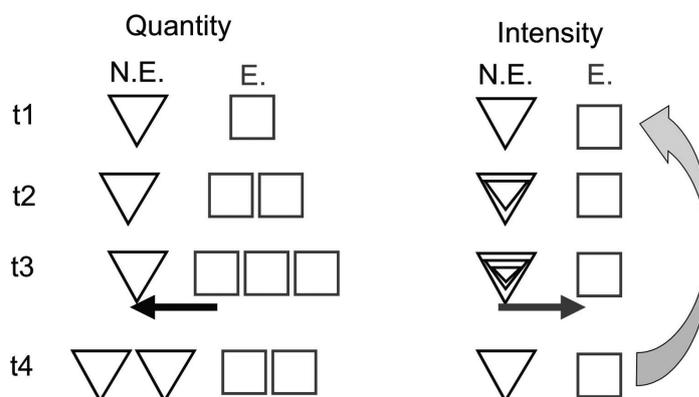


Figure 1: Twin signalling principle. New cycle starts in t3 as a result of two processes changes in quantities and intensities affecting non-equilibrium (triangles) and equilibrium hyperstructures (squares). The quantity of equilibrium hyperstructures, relative to non-equilibrium ones increases until, at time t3 in the cell cycle, a signal is emitted (black arrow on the left) that leads to some of the equilibrium material being transformed into non-equilibrium hyperstructures. In parallel, the non-equilibrium hyperstructures generate equilibrium material with an increasing intensity (t1 to t3) which reaches a threshold in t3 that leads to a complementary signal being emitted (grey arrow on the right). The arrows represent, for example, threshold flows of ions in one case and enzyme locations in the other. Cells in stage t3 are most resistant. t3 is when there is enough structure, material and activity for a cell to commit itself to a new cell cycle.

transcribing a gene [16] (and between two ribosomes translating mRNA [17]). This means that in the absence of the DNA replication part of the cell cycle, the density of RNA polymerases per unit DNA eventually reaches a maximum (i.e., when the polymerases cannot get any closer). From then on, growth can no longer be exponential. A selective pressure therefore exists in favour of cells that replicate their DNA before it becomes limiting. The solution would be *transcriptional sensing* which has been proposed for bacteria [18] but which is equally applicable to eukaryotes. The actual trigger itself of DNA replication could take one or more diverse forms including the creation by coupled transcription-translation (transertion) of a proteolipid hyperstructure [18], a giant elastic or electromagnetic oscillation [19], changes in water structure and ion condensation [8], and changes related to supercoiling [20]. A related solution but which would not involve DNA directly would be via *metabolic sensing* in which the dynamics of cytoskeletal hyperstructures (and hence cell cycle progress) would depend on the activity of metabolic enzymes such that, for example, an enzyme that is catalysing its reaction has a greater affinity for the replication and/or cytoskeletal hyperstructures and helps to stabilise/destabilise them [21].

4.2 *Quantity sensing*

Even in conditions that are generally favourable for growth, there will be times when the non-equilibrium hyperstructures can either only grow linearly or just maintain themselves or shrink, whilst the equilibrium hyperstructures continue to accumulate (red arrow in Figure 1). As the cell grows in the run-up to the key cell cycle event, new material (macromolecules, molecules and inorganic ions) comes from the non-equilibrium hyperstructures feeding both themselves and the equilibrium hyperstructures. In one version of the Dualism hypothesis, this transfer of material is accompanied by other changes such as a redistribution of water structures and changes in the topologies and sizes of both the equilibrium and non-equilibrium hyperstructures (note that resonant frequencies of oscillations are another possibility). This redistribution of material and water structures results in the charge parameters of the equilibrium and non-equilibrium hyperstructures reaching a threshold that leads to ion decondensation and condensation, respectively [8, 22]. This threshold is dependent on the topology and composition of the hyperstructures, such as the alignment of negative charges, rather than, say, the average level of charge. The changes that occur at this threshold time the cell cycle.

5 *The specific case of DnaA in the E. coli cell cycle*

5.1 *DnaA background*

DnaA is a protein found in many bacteria which is commonly believed to be responsible for the initiation of chromosome replication in bacteria [10, 23, 24, 25]. In the Beginning, however, DnaA was not present and a more fundamental mechanism operated. DnaA would therefore have evolved as a veneer on this mechanism. If so, it may prove possible to interpret DnaA action in the framework of Dualism. This we now attempt.

DnaA binds with different affinities to sites in the origin of replication, *oriC*, and elsewhere on the chromosome [26, 27]. The synergistic binding to *oriC* sites results in the unwinding of an AT-rich region and the opening of the double helix that constitutes the first step in initiation [28]. DnaA binds ATP and ADP, can form oligomers, and interacts with lipids. But what controls DnaA? One regulatory mechanism affects the availability of DnaA. A substantial proportion of DnaA binds to the *datA* locus [14]. Another regulatory mechanism depends on the direct activation of DnaA [29] which results from it binding and hydrolysing ATP (only the ATP-DnaA form is active in initiation [30]). Sufficient activation for initiation requires the reactivation of existing ADP-DnaA since *de novo* synthesis of DnaA is not enough [31]. This activation can result from the binding of DnaA to either the membrane or

DnaA-reactivating sequences, DARS1 and DARS2, on the chromosome [13]. Another factor to take into account is that fluorescent DnaA-GFP hybrids bind not only to *oriC* and *datA* to give foci but also elsewhere in the chromosome to give a helix [32, 33, 34].

How can these characteristics of DnaA be accommodated by the Dualism hypothesis?

5.2 Membrane

Firstly, consider the involvement of the membrane. Acidic phospholipids in a fluid phase dissociate ADP or ATP from DnaA so as to convert, in the presence of ATP, inactive ADP-DnaA to the active ATP-DnaA and allow DNA replication *in vitro* [12, 35, 36]; in this, cardiolipin is particularly effective whilst phosphatidylethanolamine and other lipids are either ineffective or inhibitory [37, 38, 39]. Membrane domains have long been proposed to regulate the bacterial cell cycle [40, 41, 42] and there is now considerable evidence for domains enriched in cardiolipin at the poles and between segregating chromosomes [43, 44, 45, 46]. There is also complementary evidence for the existence of domains of high microviscosity, resulting from transertion, around the chromosomes [47, 48] and even for an increase during the cell cycle in the level of cardiolipin and other lipids [47, 48, 49, 50]. In the context of Dualism, cardiolipin-rich, fluid domains might correspond to equilibrium structures such that their accumulation would correspond to *quantity sensing*. More specifically, during the build-up to initiation, transertion hyperstructures around the chromosome become increasingly active (i.e., with more transcription, more translation and more insertion of nascent proteins into membrane) generating and releasing cardiolipin to accumulate elsewhere until a threshold for reactivation of DnaA is attained. This threshold quantity would be the one at which there is sufficient membrane to allow successful completion of a new round of the cell cycle. Note too that this threshold might also depend on domains of cardiolipin acting as a proton trap to sense the energy status of the cell [51].

5.3 DnaA boxes and datA

Secondly, consider DnaA binding to its sites. There are around 350 DnaA proteins per origin (but up to a total of 2500 DnaA in a fast growing cell where there are also many origins) [52] and these proteins can bind with differing affinities to around 300 DnaA boxes in the chromosome [53]. These boxes are found in regions that include *oriC* and *datA*, which are able to bind 45 and 370 DnaA proteins with apparent dissociation constants of 8.6×10^{-9} M and 1.7×10^{-8} M respectively [54]. DnaA binding to *datA* is consistent with a

titration role in initiation because such binding appears to delay initiation or preventing hyper-initiation [55, 56, 57].

In this titration role, *datA* does not fit readily into the class of equilibrium structures that are subject to quantity sensing (for example, the number of copies of *datA* do not accumulate and progressively bind only ADP-DnaA); unless, of course, *datA* is physically part of an *equilibrium* origin hyperstructure certainly, it does seem close to *oriC* [56, 58]. In this case, the equilibrium material that accumulates (and that is sensed by the origin hyperstructure in its equilibrium state) might include some of the Nucleoid-Associated Proteins or NAPs, which are central to the organisation, replication, segregation, repair and expression of bacterial chromosomes, and which are essential if a new cell cycle is to be completed successfully. The NAPs include HU, H-NS, Fis, Integration Host Factor (IHF), StpA and Dps, each of which binds to several hundred specific sites throughout the chromosomes and which together bind non-specifically to the chromosome at an average of one NAP per hundred base pairs [59, 60, 61, 62, 63]. Several NAPs, including IHF, HU and Dps, associate with DnaA [64, 65]. IHF, for example, binds to many sites on the chromosome [7], and, in bending DNA, often acts together with other NAPs and transcription factors; for example, IHF interacts with DnaA and Fis to initiate replication via IHF binding to a site in *oriC* to stimulate unwinding [66, 67]. This makes IHF a good candidate for playing a role in quantity sensing. It is therefore significant that IHF binds to *datA* and that this binding increases the number of DnaA proteins that bind *datA* [68]. It will be apparent to the reader at this stage that the DNA-binding protein best-placed for the lead role in a quantity sensing hyperstructure perhaps as the first amongst equals is DnaA itself.

Instead of *datA* belonging to the class of quantity sensing, equilibrium hyperstructures, could it belong to the class of intensity sensing, non-equilibrium hyperstructures? Here, one is looking for evidence, for example that the superhelicity of the *datA* region increases during the build-up to initiation and that this releases DnaA. We know of no direct evidence for this (although DnaA binding to DNA is affected by superhelicity, see below). It is, of course, conceivable that a *datA*-IHF-DnaA hyperstructure is involved in *intensity sensing*. Such a role could be related to the proximity to *oriC* of the genes encoding subunits of ATP synthase which themselves are likely to create an ATP synthase transertion hyperstructure with a size and topology that should change with continued growth (and also to the distribution of gyrase sites as discussed in [20]).

5.4 DARSs

What then of DARS1 and DARS2 [13]? These two intergenic regions in the chromosome are directly responsible for part of the conversion of ADP-DnaA into ATP-DnaA needed for initiation. DARS1 (located between *bioD* and *uvrB*) and DARS2 (located between *ygpD* and *mutH*) contain three similarly organised DnaA-binding sites and, in the case of assembly of DnaA proteins into homomultimers on DARS1, the resultant interactions between them reduce the affinity of DnaA for ADP. Since the reactivation of DnaA could be observed *in vitro* with a fragment DARS1, supercoiling is not needed. This excludes one way in which DARSs could form part of an intensity sensing, non-equilibrium hyperstructure. There may, however, be something else related to a quantity sensing, equilibrium hyperstructure.

A PCR-generated fragment containing DARS1 promoted dissociation *in vitro* of ADP-DnaA at 30°C but not at 0°C [13]; the binding of DnaA itself to DARS1 and DARS2 is not affected by these temperatures [13]. Intriguingly, ion condensation onto charged polymers, such as DNA or protein filaments, is related to temperature so that it might occur at 30°C but not at 0°C [69, 70]. Such condensation is related to both the charge of the counterion and the density of charge along the polymer: at the critical charge parameter of the polymer, for example, all the trivalent ions would condense onto the polymer before the divalent ones; moreover, the linear structure of the polymer is important. It may therefore be significant that ATP can recharge DARS-bound DnaA (ATP^{4-} condenses before ADP^{3-}) and that, for this to happen, several copies of DnaA must bind to the DARS [13]. In the Dualism approach, it is speculated that quantity sensing may take the form of positively charged counterions condensing onto equilibrium hyperstructures until the threshold is attained at which counterions decondense from the origin region and the consequent repulsion of the phosphates leads to strand separation and initiation of replication [8, 71, 72]. If there is some substance to this speculation, it would not be surprising to see the same phenomenon recapitulated with the DARS in which they have evolved to become representative of ion condensation on quantity sensing, equilibrium structures. One might take it further and speculate that when the DARS cease to be equilibrium structures they cease to be able to reactivate DnaA. It goes like this: DARS1 and DARS2 are located right next to genes that are expressed in response to DNA damage; if this were to result in a change in the conformation of the DARS (e.g., their becoming single-stranded), the condensation of ATP might cease, ATP-DnaA levels drop and initiation be averted until the damaged DNA had been repaired.

5.5 *DnaA spirals*

To return to intensity sensing, consider again the distribution of DnaA. The protein, is distributed at the membrane either as foci that colocalise with the origin of replication, *oriC* [56], or as a helical spiral [73] or as both (but for *Bacillus subtilis* see [34]). Such hyperstructures may well contain acidic phospholipids such as cardiolipin [74]. The helical form of DnaA may or may not interact with both the numerous sites of DnaA on the chromosome and the membrane [32]. (Note here that the spiral is seen in the nucleoid regions and, even if it can be seen in inter-nucleoid regions, these regions may still contain DNA). If we suppose that the DnaA in this spiral is binding to both its sites on the chromosome and to membrane, there is an interesting possibility that this DnaA hyperstructure is involved in intensity sensing. This is because this spiral hyperstructure might well be responsive to the changes in superhelicity that can accompany growth [20] such that at a threshold superhelicity the spiral would be destabilised and DnaA released from these sites ready to recharged and to participate in initiation; there is indeed evidence that DnaA in some circumstances can be detached from DNA by increased superhelicity [75]. The DnaA spiral would therefore constitute an intensity sensor. In fact, it might be even better than this: the DnaA spiral hyperstructure might act as a quantity sensor by responding to an increase in cardiolipin. This is because acidic phospholipids (such as cardiolipin) in a fluid phase can detach DnaA from DNA [75] and hence an increase in cardiolipin might also be sensed by the DnaA spiral hyperstructure and transduced into the signal for the initiation of replication.

5.6 *H-NS*

In the hyperstructure dynamics approach to the cell, the interactions between hyperstructures determine the phenotype. If, then, the spiral DnaA hyperstructure is indeed acting as a sensor, there should be interactions with other hyperstructures in addition to the origin hyperstructure. These might be expected to include the recently described H-NS hyperstructure, which regulates 5% of all *E. coli* genes and brings together its regulated operons into a two or so clusters per chromosome [7]. Deletions of H-NS halved the time to replicate the chromosome and delayed initiation (insofar as there was a 50% increase in the mass at which initiation occurred) [76]. This would be consistent with the H-NS hyperstructure normally contributing to destabilising the spiral DnaA hyperstructure. It should be noted that IHF, which does not form the same type of hyperstructure [7], has similar effects on the initiation mass and replication times [77].

5.7 Ribonucleotide reductase

In the quantity sensing vein, a sensible cell would ensure that it had enough nucleotides before embarking on a new cell cycle. Deoxyribonucleotides can be generated from the ribonucleotides that constitute RNA (and for ATP and GTP make up the energy currency of the cell) by the enzyme ribonucleotide reductase which is located in the replication hyperstructure [78, 79, 80].

5.8 Central carbon metabolism (CCM)

The central carbon metabolism is the set of biochemical pathways responsible for the transport and oxidation of main carbon sources and, in *E. coli*, comprises the phosphotransferase system, glycolysis, gluconeogenesis, the pentose-monophosphate bypass with the Entner-Dudoroff pathway, Krebs cycle with the glyoxylate bypass and the respiratory chain. A relationship between the enzymes involved in CCM and the elongation step in DNA replication was first observed in *B. subtilis* [81]. This has now been extended to *E. coli* where it has been found that the temperature sensitivity of a strain containing the *dnaB(ts)* mutation could be suppressed by dysfunction of *pgi* or *pta*; *dnaE486(ts)* by dysfunction of *tktB*; *dnaG(ts)* by dysfunction of *gpmA*, *pta* or *ackA*; *dnaN159(ts)* by dysfunction of *pta* or *ackA* [82]. In particular, a full suppression of the thermosensitivity of *dnaA46(ts)* was observed in *pta* (encoding phosphate acetyltransferase) or *ackA* (encoding acetate kinase) double mutants. The acetyl phosphate level, which may be reduced in the Pta- and raised in the AckA⁻ contexts, is thought to act as a global signal in two-component systems linking metabolism with other systems [83]. Could this relationship between DnaA and Pta/AckA reflect intensity sensing? Suppose that free Pta/AckA were to bind DnaA. Then, if the formation of a FDS containing Pta and AckA (and perhaps other CCM enzymes too) were to depend on the intensity of functioning of these enzymes, the availability of ATP-DnaA might increase as the proportion of free Pta/AckA decreased. There is, however, no evidence for direct interaction between DnaA and CCM enzymes and other explanations are possible, as discussed by the authors [81, 82].

5.9 Ribosomes

Interactions between DnaA and ribosomal protein L2 (or truncated L2) have been shown to inhibit initiation *in vitro* and it has been proposed that DnaA may act as a sensor coordinating initiation with cell growth [65]. However, interpreting this result in the context of Dualism is not straightforward. It is not in line with a ribosomal hyperstructure acting as an intensity sensor because, the prediction is that if DNA is becoming limiting, redundant ribosomes may start to accumulate, become inactive and perhaps release L2 (we are trying here

to see past layers of sophisticated feedback controls) which should stimulate initiation. The opposite occurs. It is not in line with quantity sensing either because, again, the prediction is that if ribosomes accumulate as equilibrium structures, release of L2 should again stimulate initiation. Although no hypothesis should be required to explain all the data, Dualism might offer an explanation: the signal that results from the combination of intensity and quantity sensing and that initiates replication also then prevents further initiation (along the lines of the Regulatory Inactivation of DnaA in which, after initiation, the Hda protein plus DNA promote the hydrolysis of active ATP-DnaA to inactive ADP-DnaA [58]). So the prediction from Dualism would then be that in the conversion of equilibrium to non-equilibrium structures - and the consequent redeployment of ribosomes - L2 is released to contribute the inhibition of further initiations. Finally, given that L2 also interacts with a heat shock protein, HtpG, it is conceivable that its inhibition of DnaA is simply a response to stress rather than part of regulation in an unperturbed cycle [84].

6 *Constitutive stable replication*

If the modern *oriC*-DnaA system is a molecular veneer, it should be possible to return to an ancient, more fundamental system. Arguably, such a system has been revealed in *E. coli*. In the absence of *oriC* and DnaA, mutants defective in the RNase, RnhA, are able to replicate the chromosome from five other origins, *oriKs*, in what is known as constitutive stable replication [85]. Like normal initiation, stable replication involves unwinding the DNA helix and making DNA-protein hyperstructures. Initiation is thought to begin at a R-loop (a structure of two strands of DNA and one of RNA); the one strand of RNA is annealed with its complementary strand of DNA, which causes the other strand of DNA to be displaced from the helix; the RNA for the R-loop is transcribed by RNA polymerase and RecA probably then forms or stabilises the R-loop; the RNA in the R-loop is then extended by DNA polymerase I to form a D-loop [85, 86]. The explanation for constitutive stable replication requiring RNase HI is that this enzyme preferentially degrades the RNA primers at sites other than *oriC* because DnaA protects the primer at *oriC* [85].

The challenge for the Dualism hypothesis is to explain stable replication in terms of non-equilibrium and equilibrium hyperstructures responsible for intensity sensing and quantity sensing. The probability of initiation at one of the *oriK* is presumably related to the number and activity of the RNA polymerases at the locus. This would correspond to the class of intensity sensing termed transcriptional sensing. The probability of R-loop formation is increased by the negative superhelicity produced by DNA gyrase which again constitutes a type of intensity sensing (see above) [85].

7 Predictions

Many experimental predictions based on Dualism have already been made [8]. They include:

1. Non-equilibrium and equilibrium hyperstructures. The existence of two types of extended structures with different time scales might be tested by pulse-labelling bacteria with nutrients enriched with isotopes such as ^{13}C , ^{14}C and ^{15}N and imaging using Secondary Ion Mass Spectrometry [87].
2. Ion condensation. If the phenomenon of condensation underlies the recharging of DnaA with ATP by the DARS, this might be demonstrated by combining an *in vitro* system of differentially, isotope-labelled ATP and ADP (or ATP-DnaA and ADP-DnaA) and fragments containing DARS, with the technique of Combing Imaging by Secondary Ion Mass Spectrometry, CIS [88]. This would require testing for the effects of varying temperature, ion concentrations and even water structures.
3. Quantity sensing. Runaway production of equilibrium hyperstructures should lead to initiation. Constructs might be used to induce the production of lipids and NAPs and the effects tested on the numbers of origins (e.g. using isotope-labelling and density gradient centrifugation or runout experiments, or indeed CIS). Such induction is also predicted to have effects on the DnaA spiral hyperstructure.
4. Intensity sensing. Induction of formation of transertion or transembly hyperstructures [8], via systems based on IPTG, arabinose, T7, etc. should lead to initiation and could be tested as in 3. above. Alterations in the DnaA spiral hyperstructure should result from changes in superhelicity and, as in quantity sensing, from changes in the composition of the membrane.
5. Constitutive stable replication. It would be interesting to investigate quantity sensing in this context by, for example, exploring the relationships between constitutive stable replication and the accumulation and distribution of phospholipids such as cardiolipin and of NAPs, where one might expect the genes neighbouring the *oriKs* to be related to lipids and DNA-binding proteins. It has been reported that minichromosomes cannot be replicated in conditions of constitutive stable replication [85, 89], which opens up the possibility that they might be replicated if they were to contain the genes and DNA-binding sites needed for intensity and quantity sensing hyperstructures.

8 *Testing the model with HSIM*

In principle, HSIM could simulate cells containing two sorts of molecule, black and red, to represent the constituents of non-equilibrium and equilibrium hyperstructures respectively. The black molecules would be in the membrane and, on binding an externally supplied substrate, a black molecule would (1) generate x black molecules and y red molecules and (2) undergo a change of state to acquire an affinity for another black molecule leading to the formation of functioning dependent hyperstructures [90]. The red molecules would have a constant affinity for one another and would accumulate in the cytoplasm; a cell cycle signal would turn a proportion of red molecules into blue molecules. A third, blue, molecule or ion is responsible for the dialogue between the two types of hyperstructure and has an affinity for both types; the numbers of this ion are proportional to the volume of the cell; the ion has an affinity for both hyperstructures. In principle, this could allow blue ions to go from the black non-equilibrium hyperstructure to the red equilibrium hyperstructure.

How could this movement be related to the state of the hyperstructures? In the case of the red hyperstructure, it is a simple matter of the size (quantity of constituents) of the hyperstructure: the bigger the red hyperstructure, the more blue ions bind to it. In the case of the black hyperstructure, the intensity has to be sensed by the blue ions. This could be achieved if the state of the black molecules that depends on their activity (and that affects their affinity for one another) also affects their affinity for the blue ion.

Competition for blue molecules is therefore the result of the combination of an increasing quantity of red molecules and a change in the intensity of the black hyperstructure. One question is whether, in this competitive system, there are relationships between the parameters and values of the parameters that can generate a clear, precise signal. Such a signal could come from a positive feedback between, for example, condensation of the blue ion and the stability of the black hyperstructure; one might imagine that the affinity of the blue molecule for black molecules is inversely proportional to the average intensity of the black hyperstructure and, reciprocally, that the affinity of black molecules for one another is decreased in the absence (increased in the presence) of the blue ions. (Note the challenge for the computer scientist here of simulating ion decondensation so as to generate a signal, which may involve a topological change, in an automaton.) Another question is whether such values are accessible (i.e. whether they lie in a Garden of Eden). Emission of a clear signal would result in: (1) conversion of a proportion of red molecules into black ones (corresponding to DNA replication) and (2) cell division in which the spherical cell is cleaved (note that it may prove important to orient the plane of cleavage so as to maximise the segregation of non- equilibrium and equilibrium hyperstructures).

The black molecules, which are responsible for growth, weaken the membrane and, in the event of a stress, may cause the death of the cell. The red molecules strengthen the cell (but do not directly contribute to its growth). In a selective environment in which there can be periods of both growth and stress, variation of the values of the parameters of HSIM is predicted to lead to the selection of populations of cells with particular distributions of hyperstructures and sizes.

9 Discussion

A fundamental characteristic of living systems is, we argue, their ability to balance the conflicting requirements of growth and survival. This entails them balancing the proportion of their mass in the form of two structures that exist on different time scales. In the case of bacteria, in the Dualism hypothesis, these two structures can be equated to non-equilibrium hyperstructures, responsible for growth, and to equilibrium hyperstructures, responsible for survival, and the cell cycle becomes the way to distribute them amongst daughter cells so as to produce a diverse population in which cells are equipped differently for growth and survival [8]. This is what we mean by 'Life on the scales of equilibria'. It has a thermodynamics flavour insofar as the signal to initiate a new cell cycle derives from a combination of intensity sensing and quantity sensing, echoing perhaps intensive and extensive variables [91]. In proposing the Dualism hypothesis, we have also argued that cells also use intensity and quantity sensing to confirm that they have enough resources to complete the cell cycle successfully before initiating it.

The task we have taken on here is to see whether the 'scales of equilibria' approach might serve as a unifying framework for the vast and diverse quantity of molecular information on the cell cycle. We have therefore examined one of the most intensely studied cell cycle events, initiation of chromosome replication in *E. coli*. We show how it may be possible to use the concepts of non-equilibrium, intensity sensing hyperstructures and equilibrium, quantity sensing hyperstructures to class the factors known to be important in the regulation of initiation. These factors include ATP-DnaA, ADP-DnaA, DARS, *datA*, anionic phospholipids, the NAPs and chromosome supercoiling. Other factors which might also have been incorporated include the possibility of metabolic sensing by hyperstructures [21, 81] that include the replication hyperstructure [92] and the EF-Tu hyperstructure [93, 94], and the relationship between calcium concentrations and ATP levels [95]. We consider whether the scales of equilibria approach might be useful in interpreting constitutive stable replication, believed to correspond to the vestiges of an early replication system [85].

Ion condensation, decondensation and water structures might play the lead role in the dialogue between hyperstructures that, in the Dualism hypothesis, results in initiation of replication. In our re-examination of the literature on initiation of chromosome replication in *E. coli*, we find an intriguing clue implicating ion condensation in the activation of DnaA by DARS [13] and even *in vitro* evidence for a role for condensation in the case of another key player in the cell cycle, FtsZ [96].

Finally, we propose a number of experimental tests and, above all, the use of a hybrid automaton, HSIM, to clarify the concepts needed for the scales of equilibria story to become persuasive.

Acknowledgements

We thank Michel Thellier and Camille Ripoll for ideas about thermodynamics and life. For support, we thank the Epigenomics Project, Genopole, DYCOEC (GDR 2984) and the European Network in Systems Biology (GDRE 513).

References

- [1] Hunding, A., et al., *Compositional complementarity and prebiotic ecology in the origin of life*. Bioessays, 2006. **28**(4): p. 399-412.
- [2] Norris, V. and A. Delaune, *QUESTION 1: contingency versus determinism*. Origins of Life and Evolution of Biospheres, 2010. **40**: p. 365-370.
- [3] Narayanaswamy, R., et al., *Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation*. Proc Natl Acad Sci U S A, 2009. **106**(25): p. 10147-52.
- [4] Norris, V., et al., *Toward a Hyperstructure Taxonomy*. Annu Rev Microbiol, 2007. **61**: p. 309-329.
- [5] Llopis, P.M., et al., *Spatial organization of the flow of genetic information in bacteria*. Nature, 2010. **466**(7302): p. 77-81.
- [6] Nevo-Dinur, K., et al., *Translation-independent localization of mRNA in E. coli*. Science, 2011. **331**(6020): p. 1081-4.
- [7] Wang, W., et al., *Chromosome organization by a nucleoid-associated protein in live bacteria*. Science, 2011. **333**(6048): p. 1445-9.
- [8] Norris, V., *Speculations on the initiation of chromosome replication in Escherichia coli: The dualism hypothesis*. Med Hypotheses, 2011. **76**(5): p. 706-16.

- [9] Kamimura, A. and K. Kaneko, *Reproduction of a protocell by replication of a minority molecule in a catalytic reaction network*. Phys Rev Lett, 2010. **105**(26): p. 268103.
- [10] Zakrzewska-Czerwinska, J., et al., *Regulation of the initiation of chromosomal replication in bacteria*. FEMS Microbiol Rev, 2007. **31**(4): p. 378-87.
- [11] Fralick, J.A. and K.G. Lark, *Evidence for the involvement of unsaturated fatty acids in the initiation of chromosome replication in Escherichia coli*. Journal of Molecular Biology, 1973. **80**: p. 459-475.
- [12] Castuma, C.E., E. Crooke, and A. Kornberg, *Fluid membranes with acidic domains activate DnaA, the initiator protein of replication in Escherichia coli*. J Biol Chem, 1993. **268**: p. 24665-24668.
- [13] Fujimitsu, K., T. Senriuchi, and T. Katayama, *Specific genomic sequences of E. coli promote replicational initiation by directly reactivating ADP-DnaA*. Genes Dev, 2009. **23**(10): p. 1221-33.
- [14] Kitagawa, R., et al., *Negative control of replication initiation by a novel chromosomal locus exhibiting exceptional affinity for Escherichia coli DnaA protein*. Genes Dev, 1998. **12**(19): p. 3032-43.
- [15] Wang, X., et al., *Replication and segregation of an Escherichia coli chromosome with two replication origins*. Proc Natl Acad Sci U S A, 2011. **108**(26): p. E243-50.
- [16] Kennell, D. and H. Riezman, *Transcription and translation frequencies of the Escherichia coli lac operon*. Journal of Molecular Biology, 1977. **114**: p. 1-21.
- [17] Brandt, F., et al., *The native 3D organization of bacterial polysomes*. Cell, 2009. **136**(2): p. 261-71.
- [18] Norris, V., *Hypothesis: transcriptional sensing and membrane domain formation initiate chromosome replication in Escherichia coli*. Molecular Microbiology, 1995. **15**: p. 985-987.
- [19] Fröhlich, H., *Long range coherence and energy storage in biological systems*. International Journal of Quantum Chemistry, 1968. **42**: p. 641-649.
- [20] Sobetzko, P., A. Travers, and G. Muskhelishvili, *Gene order and chromosome dynamics coordinate spatiotemporal gene expression during*

- the bacterial growth cycle*. Proc Natl Acad Sci U S A, 2012. **109**(2): p. E42-50.
- [21] Norris, V., et al. *Hypothesis: the cytoskeleton is a metabolic sensor*. in *Modelling complex biological systems in the context of genomics*. 2010. Evry, France: EDP Sciences.
- [22] Ripoll, C., V. Norris, and M. Thellier, *Ion condensation and signal transduction*. BioEssays, 2004. **26**: p. 549-557.
- [23] Kaguni, J.M., *DnaA: controlling the initiation of bacterial DNA replication and more*. Annu Rev Microbiol, 2006. **60**: p. 351-75.
- [24] Mott, M.L. and J.M. Berger, *DNA replication initiation: mechanisms and regulation in bacteria*. Nat Rev Microbiol, 2007. **5**(5): p. 343-54.
- [25] Katayama, T., et al., *Regulation of the replication cycle: conserved and diverse regulatory systems for DnaA and oriC*. Nature Reviews Microbiology, 2010. **8**(3): p. 163-170.
- [26] Hansen, F.G., B.B. Christensen, and T. Atlung, *The initiator titration model: computer simulation of chromosome and minichromosome control*. Research in Microbiology, 1991. **142**: p. 161-167.
- [27] Christensen, B.B., T. Atlung, and F.G. Hansen, *DnaA boxes are important elements in setting the initiation mass of Escherichia coli*. Journal of Bacteriology, 1999. **181**: p. 2683-2688.
- [28] Duderstadt, K.E., K. Chuang, and J.M. Berger, *DNA stretching by bacterial initiators promotes replication origin opening*. Nature, 2011. **478**(7368): p. 209-13.
- [29] Leonard, A.C. and J.E. Grimwade, *Initiating chromosome replication in E. coli: it makes sense to recycle*. Genes Dev, 2009. **23**(10): p. 1145-50.
- [30] Messer, W., *The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication*. FEMS Microbiol Rev, 2002. **26**(4): p. 355-74.
- [31] Kurokawa, K., et al., *Replication cycle-coordinated change of the adenine nucleotide-bound forms of DnaA protein in Escherichia coli*. Embo J, 1999. **18**(23): p. 6642-52.
- [32] Boeneman, K., et al., *Escherichia coli DnaA forms helical structures along the longitudinal cell axis distinct from MreB filaments*. Mol Microbiol, 2009. **72**(3): p. 645-57.

- [33] Nozaki, S., H. Niki, and T. Ogawa, *Replication Initiator DnaA of Escherichia coli Changes its Assembly Form on the Replication Origin During the Cell Cycle*. J Bacteriol, 2009.
- [34] Soufo, C.D., et al., *Cell-cycle-dependent spatial sequestration of the DnaA replication initiator protein in Bacillus subtilis*. Dev Cell, 2008. **15**(6): p. 935-41.
- [35] Yung, B.Y. and A. Kornberg, *Membrane attachment activates dnaA protein, the initiation protein of chromosome replication in Escherichia coli*. Proc Natl Acad Sci U S A, 1988. **85**(19): p. 7202-5.
- [36] Makise, M., et al., *Molecular mechanism for functional interaction between DnaA protein and acidic phospholipids: identification of important amino acids*. J Biol Chem, 2001. **276**(10): p. 7450-6.
- [37] Sekimizu, K. and A. Kornberg, *Cardiolipin activation of dnaA protein, the initiation protein of replication in Escherichia coli*. J Biol Chem, 1988. **263**(15): p. 7131-5.
- [38] Yamamoto, K., et al., *Modulation of Mycobacterium tuberculosis DnaA protein-adenine-nucleotide interactions by acidic phospholipids*. Biochem J, 2002. **363**(Pt 2): p. 305-11.
- [39] Ichihashi, N., et al., *Inhibitory effects of basic or neutral phospholipid on acidic phospholipid-mediated dissociation of adenine nucleotide bound to DnaA protein, the initiator of chromosomal DNA replication*. J Biol Chem, 2003. **278**(31): p. 28778-86.
- [40] Norris, V., *DNA replication in Escherichia coli is initiated by membrane detachment of oriC : a model*. Journal of Molecular Biology, 1990. **215**: p. 67-71.
- [41] Norris, V., *Phospholipid domains determine the spatial organization of the Escherichia coli cell cycle : the membrane tectonics model*. Journal of Theoretical Biology, 1992. **154**: p. 91-107.
- [42] Norris, V., et al., *Hypothesis: hyperstructures regulate initiation in Escherichia coli and other bacteria*. Biochimie, 2002. **84**: p. 341-347.
- [43] Mileykovskaya, E. and W. Dowhan, *Visualization of phospholipid domains in Escherichia coli by using the cardiolipin-specific fluorescent dye 10-N-nonyl acridine orange*. Journal of Bacteriology, 2000. **182**: p. 1172-1175.

- [44] Koppelman, C.-M., et al., *Escherichia coli* minicell membranes are enriched in cardiolipin. *Journal of Bacteriology*, 2001. **183**: p. 6144-6147.
- [45] Kawai, F., et al., *Cardiolipin domains in Bacillus subtilis marburg membranes*. *Journal of Bacteriology*, 2004. **186**: p. 1475-1483.
- [46] Maloney, E., et al., *Localization of acidic phospholipid cardiolipin and DnaA in mycobacteria*. *Tuberculosis (Edinb)*, 2011. **91 Suppl 1**: p. S150-5.
- [47] Fishov, I. and C. Woldringh, *Visualization of membrane domains in Escherichia coli*. *Molecular Microbiology*, 1999. **32**: p. 1166-1172.
- [48] Binenbaum, Z., et al., *Transcription- and translation-dependent changes in membrane dynamics in bacteria: testing the transertion model for domain formation*. *Molecular Microbiology*, 1999. **32**: p. 1173-1182.
- [49] Michel, G.P.F., et al., *Is there a correlation between membrane phospholipid metabolism and cell division?* *Annals of the Institut Pasteur*, 1985. **136A**: p. 111-118.
- [50] Joseleau-Petit, D., et al., *DNA replication initiation, doubling of rate of phospholipid synthesis, and cell division in Escherichia coli*. *Journal of Bacteriology*, 1987. **169**: p. 3701-3706.
- [51] Haines, T.H. and N.A. Dencher, *Cardiolipin: a proton trap for oxidative phosphorylation*. *FEBS Letters*, 2002. **528**: p. 35-39.
- [52] Hansen, F.G., et al., *Initiator (DnaA) protein concentration as a function of growth rate in Escherichia coli and Salmonella typhimurium*. *J Bacteriol*, 1991. **173**(16): p. 5194-9.
- [53] Roth, A. and W. Messer, *High-affinity binding sites for the initiator protein DnaA on the chromosome of Escherichia coli*. *Mol Microbiol*, 1998. **28**(2): p. 395-401.
- [54] Kitagawa, R., et al., *A novel DnaA protein-binding site at 94.7 min on the Escherichia coli chromosome*. *Mol Microbiol*, 1996. **19**(5): p. 1137-47.
- [55] Morigen., A. Lobner-Olesen, and K. Skarstad, *Titration of the Escherichia coli DnaA protein to excess datA sites causes destabilization of replication forks, delayed replication initiation and delayed cell division*. *Molecular Microbiology*, 2003. **46**: p. 245-256.

- [56] Nozaki, S., H. Niki, and T. Ogawa, *Replication initiator DnaA of Escherichia coli changes its assembly form on the replication origin during the cell cycle*. J Bacteriol, 2009. **191**(15): p. 4807-14.
- [57] Felczak, M.M. and J.M. Kaguni, *DnaAcos hyperinitiates by circumventing regulatory pathways that control the frequency of initiation in Escherichia coli*. Mol Microbiol, 2009. **72**(6): p. 1348-63.
- [58] Katayama, T., et al., *Regulation of the replication cycle: conserved and diverse regulatory systems for DnaA and oriC*. Nat Rev Microbiol, 2010. **8**(3): p. 163-70.
- [59] Li, G.-W., O.G. Berg, and J. Elf, *Effects of macromolecular crowding and DNA looping on gene regulation kinetics*. Nature Physics, 2009. **5**: p. 294 - 297.
- [60] Oshima, T., et al., *Escherichia coli histone-like protein H-NS preferentially binds to horizontally acquired DNA in association with RNA polymerase*. DNA Res, 2006. **13**(4): p. 141-53.
- [61] Maurer, S., J. Fritz, and G. Muskhelishvili, *A systematic in vitro study of nucleoprotein complexes formed by bacterial nucleoid-associated proteins revealing novel types of DNA organization*. J Mol Biol, 2009. **387**(5): p. 1261-76.
- [62] Browning, D.F., D.C. Grainger, and S.J. Busby, *Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression*. Curr Opin Microbiol, 2010. **13**(6): p. 773-80.
- [63] Dillon, S.C. and C.J. Dorman, *Bacterial nucleoid-associated proteins, nucleoid structure and gene expression*. Nat Rev Microbiol, 2010. **8**(3): p. 185-95.
- [64] Leonard, A.C. and J.E. Grimwade, *Regulating DnaA complex assembly: it is time to fill the gaps*. Curr Opin Microbiol, 2010. **13**(6): p. 766-72.
- [65] Chodavarapu, S., M.M. Felczak, and J.M. Kaguni, *Two forms of ribosomal protein L2 of Escherichia coli that inhibit DnaA in DNA replication*. Nucleic Acids Res, 2011. **39**(10): p. 4180-91.
- [66] Ryan, V.T., et al., *Escherichia coli prereplication complex assembly is regulated by dynamic interplay among Fis, IHF and DnaA*. Mol Microbiol, 2004. **51**(5): p. 1347-59.
- [67] Swinger, K.K. and P.A. Rice, *IHF and HU: flexible architects of bent DNA*. Curr Opin Struct Biol, 2004. **14**(1): p. 28-35.

- [68] Nozaki, S., Y. Yamada, and T. Ogawa, *Initiator titration complex formed at datA with the aid of IHF regulates replication timing in Escherichia coli*. *Genes Cells*, 2009. **14**(3): p. 329-41.
- [69] Manning, G.S., *Limiting laws and counterion condensation in polyelectrolyte solutions. I. Colligative properties*. *J Chem Phys*, 1969. **51**: p. 924-933.
- [70] Oosawa, F., *Polyelectrolytes*. 1971, New York: Dekker. 1-160.
- [71] Manning, G.S., *Electrostatic free energy of the DNA double helix in counterion condensation theory*. *Biophys Chem*, 2002. **101-102**: p. 461-73.
- [72] von Hippel, P.H., *From 'simple' DNA-protein interactions to the macromolecular machines of gene expression*. *Annu Rev Biophys Biomol Struct*, 2007. **36**: p. 79-105.
- [73] Boeneman, K., et al., *Escherichia coli DnaA forms helical structures along the longitudinal cell axis distinct from MreB filaments*. *Mol Microbiol*, 2009.
- [74] Aranovich, A., et al., *Membrane-catalyzed nucleotide exchange on DnaA. Effect of surface molecular crowding*. *J Biol Chem*, 2006. **281**(18): p. 12526-34.
- [75] Makise, M., et al., *Acidic phospholipids inhibit the DNA-binding activity of DnaA protein, the initiator of chromosomal DNA replication in Escherichia coli*. *Mol Microbiol*, 2002. **46**(1): p. 245-56.
- [76] Atlung, T. and F.G. Hansen, *Effect of different concentrations of H-NS protein on chromosome replication and the cell cycle in Escherichia coli*. *J Bacteriol*, 2002. **184**(7): p. 1843-50.
- [77] Von Freiesleben, U., et al., *Rifampicin-resistant initiation of chromosome replication from oriC in ihf mutants*. *Mol Microbiol*, 2000. **37**(5): p. 1087-93.
- [78] Guarino, E., A. Jimenez-Sanchez, and E.C. Guzman, *Defective Ribonucleoside Diphosphate Reductase Impairs Replication Fork Progression in Escherichia coli*. *J Bacteriol*, 2007. **189**(9): p. 3496-501.
- [79] Odsbu, I., Morigen, and K. Skarstad, *A reduction in ribonucleotide reductase activity slows down the chromosome replication fork but does not change its localization*. *PLoS One*, 2009. **4**(10): p. e7617.

- [80] Sanchez-Romero, M.A., F. Molina, and A. Jimenez-Sanchez, *Organization of ribonucleoside diphosphate reductase during multifork chromosome replication in Escherichia coli*. *Microbiology*, 2011. **157**(Pt 8): p. 2220-5.
- [81] Janniere, L., et al., *Genetic evidence for a link between glycolysis and DNA replication*. *PLoS ONE*, 2007. **2**: p. e447.
- [82] Maciag, M., et al., *Genetic response to metabolic fluctuations: correlation between central carbon metabolism and DNA replication in Escherichia coli*. *Microb Cell Fact*, 2011. **10**: p. 19.
- [83] Wolfe, A.J., *Physiologically relevant small phosphodonors link metabolism to signal transduction*. *Curr Opin Microbiol*, 2010. **13**(2): p. 204-9.
- [84] Motojima-Miyazaki, Y., M. Yoshida, and F. Motojima, *Ribosomal protein L2 associates with E. coli HtpG and activates its ATPase activity*. *Biochem Biophys Res Commun*, 2010. **400**(2): p. 241-5.
- [85] Kogoma, T., *Stable DNA replication: interplay between DNA replication, homologous recombination and transcription*. *Microbiology and Molecular Biology Reviews*, 1997. **61**: p. 212-238.
- [86] Sandler, S.J., *Requirements for replication restart proteins during constitutive stable DNA replication in Escherichia coli K-12*. *Genetics*, 2005. **169**(4): p. 1799-806.
- [87] Byrne, M.E., et al., *Desulfovibrio magneticus RS-1 contains an iron- and phosphorus-rich organelle distinct from its bullet-shaped magnetosomes*. *Proc Natl Acad Sci U S A*, 2010. **107**(27): p. 12263-8.
- [88] Cabin-Flaman, A., et al., *Combed Single DNA Molecules Imaged by Secondary Ion Mass Spectrometry*. *Anal Chem*, 2011. **83**(18): p. 6940-6947.
- [89] Hong, X. and T. Kogoma, *Absence of a direct role for RNase HI in initiation of DNA replication at the oriC site on the Escherichia coli chromosome*. *J Bacteriol*, 1993. **175**(20): p. 6731-4.
- [90] Amar, P., et al., *A stochastic automaton shows how enzyme assemblies may contribute to metabolic efficiency*. *BMC Syst Biol*, 2008. **2**: p. 27.
- [91] Raine, D.J., et al., *Networks as constrained thermodynamic systems*. *Comptes Rendus de l'Academie des Sciences*, 2003. **326**: p. 65-74.

- [92] Batto, A.F., et al. *From metabolic hyperstructures to DNA replication complexes and back again.* in *Modelling complex biological systems in the context of genomics*. 2008. Evry, France: EDP Sciences.
- [93] Mayer, F., *Cytoskeletons in prokaryotes.* Cell Biology International, 2003. **27**: p. 429-438.
- [94] Defeu Soufo, H.J., et al., *Bacterial translation elongation factor EF-Tu interacts and colocalizes with actin-like MreB protein.* Proc Natl Acad Sci U S A, 2010. **107**(7): p. 3163-8.
- [95] Naseem, R., et al., *ATP regulates calcium efflux and growth in E. coli.* J Mol Biol, 2009. **391**(1): p. 42-56.
- [96] Popp, D., et al., *Suprastructures and dynamic properties of Mycobacterium tuberculosis FtsZ.* J Biol Chem, 2010. **285**(15): p. 11281-9.



Suitability of Secondary Ion Mass Spectrometry (SIMS) for studies of metabolic heterogeneity

Ghislain Gangwe Nana¹, Armelle Cabin¹, David Gibouin¹, Fabrice Lefebvre¹, Anthony Delaune¹, Laurent Janniere², Camille Ripoll¹ and Vic Norris¹

¹ EA 3829, Department of Biology, University of Rouen
F-76821 Mont Saint Aignan, France

² MEGA Laboratory, Institute of System and Synthetic Biology, Génopole Campus I,
rue Henri Desbruères, 91000 Evry, France

Abstract

Failure to appreciate heterogeneity when growing bacteria may prove a powerful source of artefacts. Unfortunately, the nature and extent of heterogeneity, both within and between bacteria, is far from clear. Here, we investigate the suitability for metabolic studies of combining the labelling of *Bacillus subtilis* with the stable isotopes ¹³C and ¹⁵N with analysis of the distribution of these isotopes with Secondary Ion Mass Spectrometry.

1 Introduction

A bacterial population exhibits balanced growth when every extensive property of the growing system (e.g. mass) increases by the same factor over a time interval [1]; it has, however, been argued that this definition does not include cell number and that the term may be used to describe an exponentially growing population in which cells are filamenting [2]. A bacterial population, whether growing or not, is in steady state when, for example, its intensive properties (e.g. DNA/mass or origin/mass) are independent of time [3]. Ole Maaloe, the founder of the Copenhagen school of bacterial growth physiology, once advised that 'every effort must be made to achieve [true steady-state growth] before commencing serious measurements' in other words, exponential growth alone is not sufficient [4]. Despite such advice, it is often assumed pragmatically by experimentalists that after a couple of hours exponential growth, the bacteria in a population will be relatively similar. There are, however, good reasons to question this assumption. Steady-state growth can be attained only during the exponential phase of a culture and only when the composition of the growth medium is constant and this may require considerable effort [2]. This all boils down to heterogeneity. How and why does heterogeneity vary? What is the full extent of intracellular and intercellular heterogeneity in bacteria?

A comprehensive evaluation of bacterial heterogeneity even during conditions of steady state growth may not be straightforward because the generation of heterogeneity is fundamental to bacteria which must develop strategies to survive and flourish in a fluctuating environment that presents both opportunities and dangers. No single phenotype in a population will suffice. Phenotypic heterogeneity of bacterial populations is widely evidenced [5, 6, 7] but just how far phenotypic heterogeneity is related to metabolic and structural heterogeneity is less clear. We have argued that phenotypic heterogeneity results from differences in the cell's complement of hyperstructures alias the extended assemblies of molecules and macromolecules that serve a function [8]. Such differences are speculated to arise because bacteria use the cell cycle to vary the ratio of (1) the static, redundant and robust quasi-equilibrium structures needed to survive in a difficult environment with (2) the dynamic, efficient and fragile non-equilibrium structures needed to grow in a favourable environment [9].

Studies of heterogeneity can benefit from using Secondary Ion Mass Spectrometry (SIMS) to determine the distribution of isotopes. This technique allows the isotopic composition of sections of cells and tissues to be determined on scales down to 50 nm and has been used to study the incorporation of ^{13}C and ^{15}N into bacteria and other cells [10, 11, 12]. Here, we describe a method that will allow SIMS to be used to investigate intracellular and intercellular heterogeneity in *Bacillus subtilis*, a model organism that readily differentiates, following labelling during exponential growth as commonly done by microbiologists.

2 Materials and Methods

B. subtilis was grown at 40°C with shaking in Spizizen medium containing per litre 50mM CaCl_2 , 14g K_2HPO_4 , 6g KH_2PO_4 , 2g $(\text{NH}_4)_2\text{SO}_4$, 1g $\text{C}_6\text{H}_5\text{Na}_3\text{O}_7 \cdot 7.2\text{H}_2\text{O}$, 2mmol/L MgSO_4 , 11 $\mu\text{g}/\text{mL}$ Fe III citrate, 10 $\mu\text{mol}/\text{L}$ MnCl_2 , 1 μM FeSO_4 , 4 $\mu\text{g}/\text{mL}$ FeCl_3 , 0.2% D-glucose, 0.01% tryptophan and 0.1% casein hydrolysate.

After dilution of an overnight culture (1/50), bacteria were grown to exponential phase where the doubling time was 42 minutes. At the start of the exponential phase, three hours after this dilution, ^{13}C -D-glucose (99% ^{13}C , ISOTECH, USA) was added at 11mM and/or $^{15}\text{NH}_4\text{Cl}$ (98% ^{15}N , ISOTECH, USA) at 0.150 mM to give final ratios of $^{13}\text{C}:^{12}\text{C}$ of 1 and $^{15}\text{N}:^{14}\text{N}$ of 1. 10 mL samples were then taken at 90 minutes. Growth was stopped by adding 10 mL of Spizizen medium at 4°C. All subsequent manipulations were then performed at or below 4°C. Samples were centrifuged in a Sigma 3K18C at 6000 rpm (8700g) for 10 minutes and the pellets were resuspended in 0.1mol/L cacodylate plus 0.04% MgCl_2 (to help avoid autolysis), then centrifuged again at 6000 rpm for 15 min.

In the following dehydration step, performed at 4°C, the pellet was resuspended in 25% ethanol for 20 minutes, then 50% ethanol for 15 minutes and centrifuged at 6000 rpm (8700 g) for 10 minutes. The pellet was then suspended in 10L of 50% ethanol and 2L were deposited on a silicon chip (which had previously been cleaned by sonication during successive immersions in distilled water, 96% ethanol, acetone, 3:1 H₂SO₄:H₂O₂, water, and finally 96% ethanol). The dehydrated bacteria on the chip were then baked in a vacuum at 50°C. They were then analysed using a Zeiss electron microscope and a CAMECA NanoSIMS 50. The latter entailed bombardment of the chip with a primary beam of Cs⁺ ions, a mass resolution of 5000, and the counting of five of the following secondary ions at the same time from the same place on the surface of the sample: ¹²C, ¹³C, ¹²C¹⁴N, ¹²C¹⁵N, ¹³C¹⁵N, and ³²S. The counts were then analysed using ImageJ. The naturally occurring ratio of ¹⁵N:¹⁴N is 0.0036 and ¹³C:¹²C is 0.011. Isotope ratios (such as ¹²C¹⁵N:¹²C¹⁴N) were used to avoid artefacts due to matrix effects as previously described [10, 13].

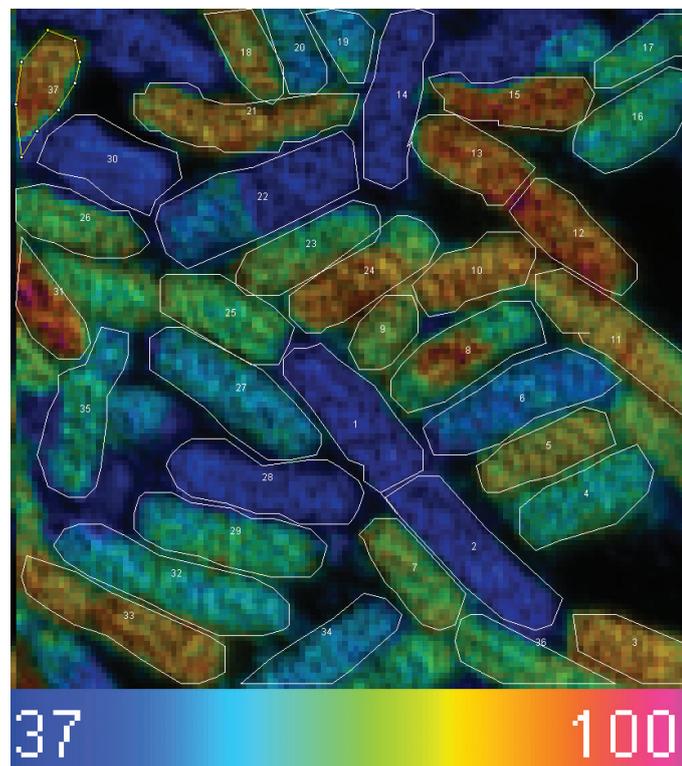


Figure 1: Isotope distribution within a population of *B. subtilis*. Cells were labelled with ¹⁵N for 90 minutes at the start of exponential growth and visualised using the HSI representation in which blue (e.g. 37 left) and magenta (e.g. 100 right) corresponds to the ratio of (¹⁵N/¹⁴N) × 100.

To visualise the distribution of isotope ratios, a Hue Saturation Intensity (HSI) transformation of the ratio image was used with a hue code spanning the spectrum of colours from blue (low isotope ratio threshold) to magenta (high isotope ratio threshold) (NRIMS, <http://www.nrims.hms.harvard.edu>). The significance of results was also determined using the ratio of, for example, $^{15}\text{N}:(^{14}\text{N}+^{15}\text{N})$ with the SIMSAD program (Delaune, in preparation).

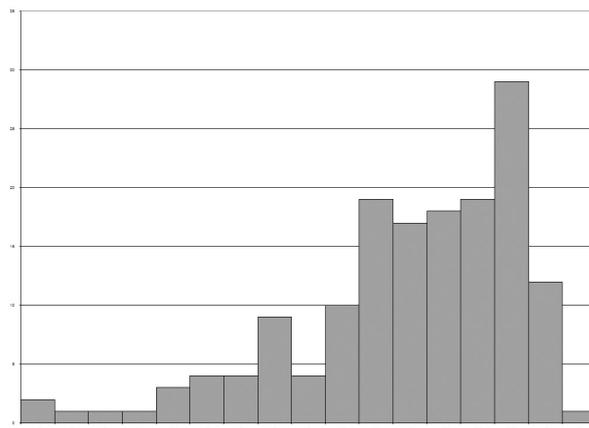


Figure 2: Distribution of the isotopic fraction $^{15}\text{N}/(^{15}\text{N}+^{14}\text{N})$. Cells of *B. subtilis* were labelled with ^{15}N for 90 minutes at the start of exponential growth. 154 bacteria were analysed and put into 17 bins of equal width (isotopic fraction 0.02). Bin 1 is 0.15-0.17 and bin 17 is 0.47-0.49. The ordinate gives the number of bacteria in each bin.

3 Results and Discussion

In the preliminary experiments reported here, we labelled *B. subtilis* over 90 minutes, which corresponds to roughly two mass doublings on average (the doubling time of the population was 42 minutes). Different levels of incorporation of ^{15}N are clearly discernible in Figure 1: there is heterogeneity in label at both the intracellular and cellular levels. The histogram in Figure 2 shows that this heterogeneity is quite extensive. Other data (unpublished) reveal that the bacteria most heavily labelled with ^{13}C are also most heavily labelled with ^{15}N , indeed, our data show that the incorporation of ^{13}C is directly proportional to incorporation of ^{15}N . Moreover, the pixels corresponding to the maximum level of ^{13}C and ^{15}N are located within the same region of the bacterium (see Table 1 for the case of 120 minute labelling). This is in spite of the fact that differences between ^{13}C and ^{15}N might be expected if there are significant differences on the 100 nm scale in the distributions of C and N due to differences in the composition of the different regions. We also took serial

sputter 'slices'; the considerable similarity between these slices shows that the considerable difference within slices is not due to artefacts linked to position and topology. Note that artefacts can also arise when taking ratios due to low absolute numbers of counts and a tailor-made program, SADSIM, can be used to increase the significance of results (Delaune, in preparation).

Position of pixels corresponding to maximum labelling	% of cells showing pixels in this position
^{13}C at a pole	64
^{15}N at a pole	83
Colocation of ^{13}C and ^{15}N at a pole	48
Colocation of ^{13}C and ^{15}N throughout the cell	64

Table 1: Location of the pixel corresponding to the highest level of ^{13}C and ^{15}N within *B. subtilis*. 31 cells were analysed labelling after 120 minute.

It is clear that SIMS can be used to generate quantitative data useful in the study of heterogeneity. The questions that may be addressed include (1) whether an exponentially growing population of cells always comprises cells that are growing slower than others, (2) what proportion of a stationary phase population of cells grows immediately on addition of fresh medium and (3) what are the dynamics and locations of equilibrium and non- equilibrium structures within cells.

Acknowledgements

For support, we thank the Epigenomics Project, Genopole, DYCOEC (GDR 2984) and the European Network in Systems Biology (GDRE 513).

References

- [1] Campbell, A., *Synchronization of cell division*. Bacteriol Rev, 1957. **21**(4): p. 263-72.
- [2] Fishov, I., A. Zaritsky, and N.B. Grover, *On microbial states of growth*. Molecular Microbiology, 1995. **15**: p. 789-794.
- [3] Painter, P.R. and A.G. Marr, *Mathematics of microbial populations*. Annu Rev Microbiol, 1968. **22**: p. 519-48.
- [4] Ingraham, J.L., O. Maaloe, and F.C. Neidhardt, *Growth of the Bacterial Cell*. 1983: Sunderland: Sinauer Associates. 435.

- [5] Balaban, N.Q., et al., *Bacterial persistence as a phenotypic switch*. Science, 2004. **305**: p. 1622-1625.
- [6] Smits, W.K., O.P. Kuipers, and J.W. Veening, *Phenotypic variation in bacteria: the role of feedback regulation*. Nat Rev Microbiol, 2006. **4**(4): p. 259-71.
- [7] Avery, S.V., *Microbial cell individuality and the underlying sources of heterogeneity*. Nat Rev Microbiol, 2006. **4**(8): p. 577-87.
- [8] Norris, V., et al., *Toward a Hyperstructure Taxonomy*. Annu Rev Microbiol, 2007. **61**: p. 309-329.
- [9] Norris, V., *Speculations on the initiation of chromosome replication in Escherichia coli: The dualism hypothesis*. Med Hypotheses, 2011. **76**(5): p. 706-16.
- [10] Lhuissier, F., et al., *Secondary ion mass spectrometry imaging of the fixation of ^{15}N -labelled NO in pollen grains*. J Microsc, 2000. **198**(Pt 2): p. 108-15.
- [11] Guerquin-Kern, J.L., et al., *Progress in analytical imaging of the cell by dynamic secondary ion mass spectrometry (SIMS microscopy)*. Biochimica Biophysica acta, 2005. **1724**(3): p. 228-38.
- [12] Steinhäuser, M.L., et al., *Multi-isotope imaging mass spectrometry quantifies stem cell division and metabolism*. Nature, 2012. **481**(7382): p. 516-9.
- [13] Peteranderl, R. and C. Lechene, *Measure of carbon and nitrogen stable isotope ratios in cultured cells*. J Am Soc Mass Spectrom, 2004. **15**(4): p. 478-85.

Simplified models for the mammalian circadian clock

Jean-Paul Comet¹, Gilles Bernot¹, Aparna Das²,
Francine Diener², Camille Massot¹ and Amélie Cessieux¹

¹ Laboratoire I3S, UMR 7271 UNS-CNRS, Université de Nice,
2000, route des Lucioles, B.P. 121, 06903 Sophia Antipolis CEDEX, France.
e-mail : {bernot, comet}@unice.fr

² Laboratoire J.A. Dieudonné, UMR CNRS 7351, Université de Nice,
Parc Valrose, 06108 Nice Cedex 02, France.
e-mail : {aparna, diener}@unice.fr

Abstract

Numerous biological mechanisms are synchronized by the circadian rhythm in species so diverse as mushrooms, drosophiles or mammals. Because of its ubiquity and of its implication in numerous cellular functions, we believe important to be able to extract the main “coarse-grain” mechanisms common to a majority of organisms. In this chapter we consider both differential equation models and discrete models and we deliberately choose to push the simplicity of the model as far as possible, focusing only on a few biological behaviours of interest. The hope is to get the essential abstract causalities that govern these behaviours.

1 Introduction

The growing interest for the circadian cycle in biology is motivated by several facts, including an important proportion of the population with staggered working timetable, the effect of jet lag, and, also importantly, the drug chronotherapy. Many of these studies involve careful modelling where the circadian cycle interacts with other subsystems such as the metabolic pathway influenced by a particular drug [1]. A good design of such complex models implies, among others, a proper global understanding of the main features of the circadian cycle behaviours. In this perspective the goal of this paper is to design models of the circadian cycle that focus on these main features whilst being sufficiently simple to offer a human comprehension of the global circadian system.

Mathematical models of biological systems are often classified into three main frameworks.

- Continuous models using differential equations constitute probably the most frequently used framework, see for example [2].

- With the growing interest for logical approaches, discrete models constitute a class of frameworks which focus on the causality relationships between events, this kind of models is well suited to take into account qualitative information [3, 4].
- Lastly the stochastic approaches form a sort of intermediate class of frameworks: the global dynamics of the model is divided into elementary steps to which are attributed some probabilities [5].

Among others, Leloup and Goldbeter [6] proposed a carefully detailed differential model of the circadian clock; Forger and Peskin [7] proposed a stochastic model. There does not yet exist any qualitative model of the mammalian circadian clock, which we incidentally propose in this article because discrete approaches push further simplicity. As a first step towards simplification we propose two successive differential models reflecting a two-step simplification of the Leloup-Goldbeter's model. We then jump directly to a discrete model in order to take the day-night alternation into account, this additional simplification being sufficiently intuitive to avoid addressing a stochastic modelling.

In section 2 we describe the main molecular mechanisms that constitute circadian cycle in the suprachiasmatic nucleus. Although partly governed by the light/dark alternation, it is known that this cycle is actually an endogenous cycle, governed by an internal mechanism in each cell and the goal of section 3 is to introduce two simplified EDP models and section 4 establishes on the smaller model a mathematical proof of the existence of sensible parameters that lead to a limit cycle in the dynamics. Section 5 introduces a supplementary simplification using the discrete Thomas' framework whereas introducing the light influence on the system. Lastly section 6 takes into consideration some delays and gives an obvious explanation of the main reasons for the robustness against the day length variations.

2 Molecular processes underlying the mammalian circadian clock

Mammalian circadian clocks regulate numerous biological functions (sleeping, locomotive activity, internal temperature, hormonal secretions and so on). In the absence of temporal reference point, the circadian rhythm shows a sustained oscillation both at the cellular and the organism levels, indicating that the body possesses an endogenous daily clock. Although each cell is equipped with an internal clock, there are external synchronizers, called Zeitgebers ("donors of time"), such as light, social rhythms, food or physical exercise that maintain a 24-hour oscillation.

In mammals there is an internal supervisor made of two neuronal groups situated at the base of the hypothalamus, that is called the *suprachiasmatic*

nucleus (SCN) [8]. All the cells that constitute the SCN exhibit a rhythmic circadian clock and they are strongly synchronized (apparently using some local signals). The cells in the peripheral tissues (liver, muscles...) also exhibit a rhythmic behaviour [9] but it seems that the SCN plays the role of main oscillator synchronising the peripheral oscillators. Injuries of the SCN entail a visible disappearance of the global rhythm, and a transplant restores it [10, 11, 12]. In the suprachiasmatic nucleus, the main Zeitgeber is light. This “pacemaker” receives signals from the retina *via* the retinohypothalamic tract which is independent from the vision mechanism [13]. According to [14] these signals are passed on by the SCN into the pineal gland which, through the night-secretion of melatonin, informs the whole body of the arrival of night.

At the molecular level in the SCN cells, the regulation of the circadian cycle is driven by the expression of the clock genes, generating a feedback loop detailed in Figure 1:

- During the day, the protein complex BMAL1/CLOCK induces the expression of clock genes by fixing their promoters. It stimulates the production of CRY1 and CRY2 (cryptochromes), as well as of PER1, PER2 and PER3, RORA and REV-ERB α . Proteins RORA activate *bmal1* and *cry1* while REV-ERB α inhibits them [15, 16]. Moreover, in the suprachiasmatic nucleus, the light activates the transcription of genes *per* by inducing the acetylation of their promoters, that allows the slackening of the chromatin and their transcription [17]. Proteins PER are then phosphorylated by CKI ϵ (Casein Kinase I ϵ), see [18]. Then these proteins can follow three different pathways: they can be destroyed *via* ubiquitination, or they can form complexes PER1-PER2-CKI ϵ which accumulate in the cytoplasm, or they can also form complexes PER1-CRY1-CKI ϵ and PER2-CRY2-CKI ϵ which also accumulate in the cytoplasm¹.
- During the night, complexes PER1-PER2-CKI ϵ and PER-CRY-CKI ϵ are hyperphosphorylated by CKI ϵ before entering the nucleus where they inhibit BMAL1/CLOCK [19]. This degradation probably originates from CKI ϵ since it can lead to the degradation of complexes (PER or CRY)-PER-CKI ϵ -BMAL1/CLOCK by phosphorylating BMAL1 [18]. Thus during the night there is no transcription anymore of the circadian genes, leading to a negative retro-control of PER and CRY on themselves.
In the cytoplasm, PER-CRY-CKI ϵ captures REV-ERB α . RORA can then activate the transcription of BMAL1 which will form a complex with CLOCK. Nevertheless a period of time is mandatory to neutralize

¹Let us remark that the function of PER3 is not yet known.

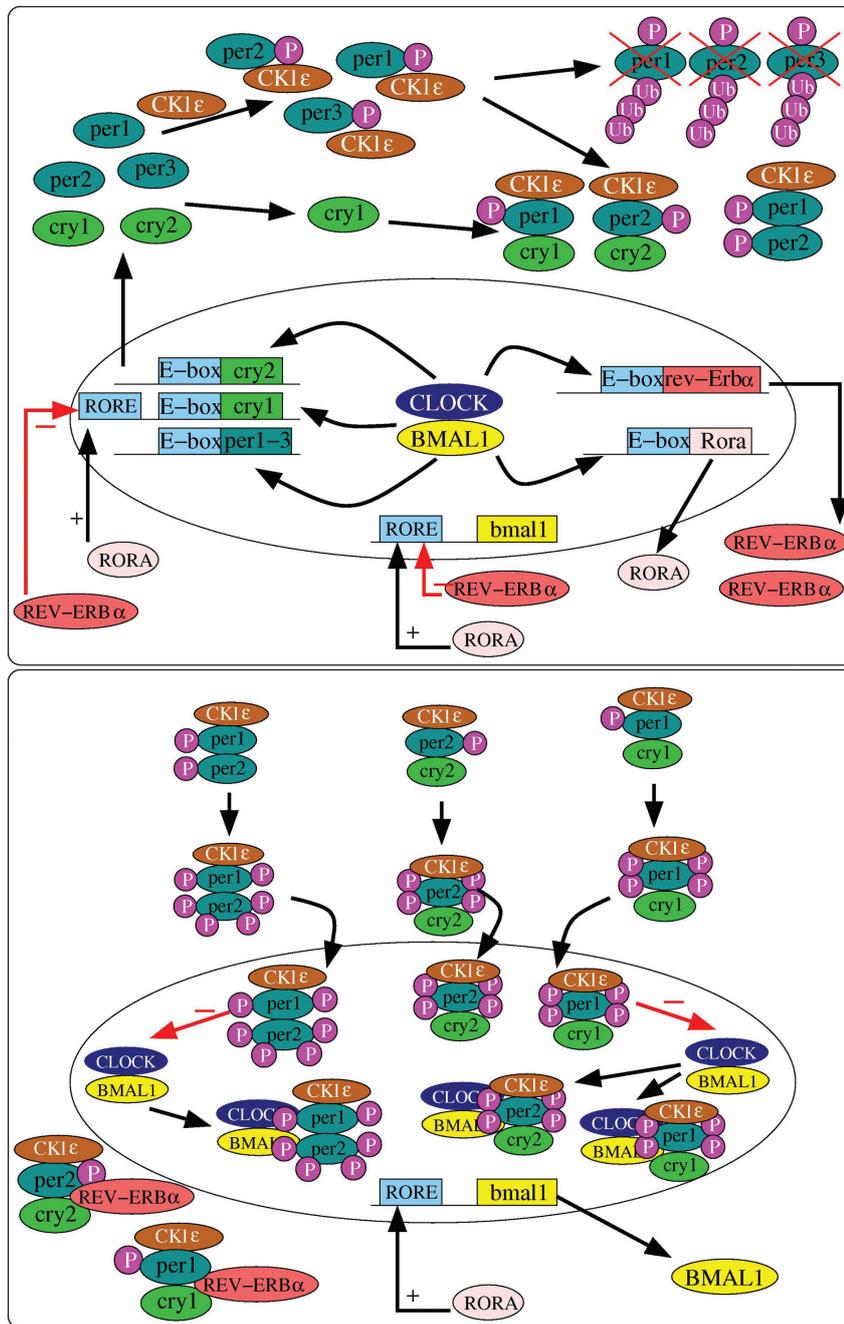


Figure 1: Molecular regulations of the mouse circadian clock during the day (top) and during the night (bottom).

all molecules REV-ERB α , this explains why the peak of transcription of BMAL1 takes place several hours before dawn. At dawn, proteins CRY and PER have all been destroyed by the degradation of complexes. REV-ERB α is consequently again active and inhibits *bmal1* via RORE. The cycle can then restart [20, 21, 22].

3 Simplified EDP models for the mammalian circadian clock

In this section we propose a mathematical model of the circadian cycle obtained as a simplification of the model of Leloup and Goldbeter [6]. In the next section we will establish the existence of a limit cycle with a default period of 24 hours.

We will focus in this section on the behavior of the complex PER-CRY. During the day, the proteins PER and CRY are phosphorylated by CKI ϵ and form complexes PER-CRY-CKI ϵ which then accumulate in the cytoplasm. During the night, PER-CRY-CKI ϵ captures REV-ERB α in the cytoplasm and will be hyperphosphorylated by CKI ϵ before entering the nucleus where it inhibits its own transcription.

The continuous time model of Leloup and Goldbeter [6] contains 16 differential equations. The considered genes are *per*, *cry*, *bmal1*, *clock* and *rev-erb α* . They take into account the phosphorylation cycles, membrane exchanges between nucleus and cytosol, the translations of mRNA into protein, as well as the association of PER and CRY proteins to form the complex PER-CRY. The trajectories resulting from the equations can be cyclic with a sensible circadian period. Goldbeter and Leloup [6] also focus on the behavior of the concentrations over time under some light-dark alternation patterns: they consider the constant darkness case as well as the case “12 hours light – 12 hours dark.” Their model is quite close to the biological reality, despite some simplifications. This model also correctly predicts several mutant behaviours of the circadian rhythm.

3.1 Model with 8 variables

In this subsection, we reduce the original model of Leloup and Goldbeter without losing too much information. This approach allows us to get a simpler model with 8 variables and 8 equations, leading to a better focus on the main involved mechanisms and on their importance within the overall schema.

We first decided to remove CLOCK protein because, although important to form a complex with BMAL1, its amount is known to be constant and is never a critical resource. Secondly, following Leloup and Goldbeter, we removed REV-ERB α from our EDP model. In fact the action of REV-ERB α on *bmal1* is hidden in the negative feedback from the PER-CRY complex in the nucleus

to the PER and CRY in the cytosol, see Figure 2. Lastly, we have represented as a unique step both translation and traduction, leading to the removal of all mRNA variables.

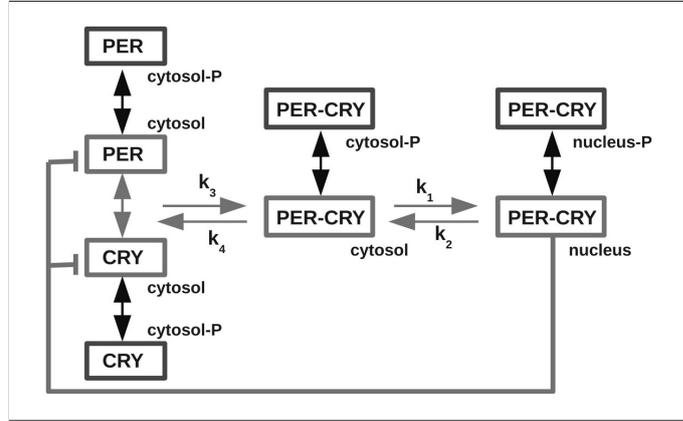


Figure 2: The 8 variable model contains a negative feedback loop (grey arrows) where (1) the PER-CRY complex of the nucleus inhibits the proteins PER and CRY, (2) PER and CRY form a complex in the cytosol and (3) this complex enters the nucleus. Black arrows represent reversible phosphorylations.

So, the model of Figure 2 is described by a set of 8 kinetic equations with 28 parameter values. We have chosen similar parameter values to those of Leloup and Goldbeter, with very few modifications.

$$\frac{dP_C}{dt} = \frac{K^n}{K^n + PC_N^n} v_1 - v_{1p} \frac{P_C}{k_p + P_C} + v_{2p} \frac{P_{CP}}{k_{dp} + P_{CP}} - k_3 P_C C_C + k_4 P_C C_C - kd_1 P_C \quad (1)$$

$$\frac{dC_C}{dt} = \frac{K^n}{K^n + PC_N^n} v_2 - v_{1c} \frac{C_C}{k_p + C_C} + v_{2c} \frac{C_{CP}}{k_{dp} + C_{CP}} - k_3 P_C C_C + k_4 P_C C_C - kd_2 C_C \quad (2)$$

$$\frac{dPC_C}{dt} = -v_{1pc} \frac{PC_C}{k_p + PC_C} + v_{2pc} \frac{PC_{CP}}{k_{dp} + PC_{CP}} + k_3 P_C C_C - k_4 P_C C_C - k_1 PC_C + k_2 PC_N - kd_3 PC_C \quad (3)$$

$$\frac{dPC_N}{dt} = -v_{3pc} \frac{PC_N}{k_p + PC_N} + v_{4pc} \frac{PC_{NP}}{k_{dp} + PC_{NP}} + k_1 PC_C - k_2 PC_N - kd_4 PC_N \quad (4)$$

$$\frac{dP_{CP}}{dt} = v_{1p} \frac{P_C}{k_p + P_C} - v_{2p} \frac{P_{CP}}{k_{dp} + P_{CP}} - v_{dpc} \frac{P_{CP}}{k_d + P_{CP}} - k_{dn} P_{CP} \quad (5)$$

$$\frac{dC_{CP}}{dt} = v_{1c} \frac{C_C}{k_p + C_C} - v_{2c} \frac{C_{CP}}{k_{dp} + C_{CP}} - v_{dcc} \frac{C_{CP}}{k_d + C_{CP}} - k_{dn} C_{CP} \quad (6)$$

$$\frac{dPC_{CP}}{dt} = v_{1pc} \frac{PC_C}{k_p + PC_C} - v_{2pc} \frac{PC_{CP}}{k_{dp} + PC_{CP}} - v_{dpc} \frac{PC_{CP}}{k_d + PC_{CP}} - k_{dn} PC_{CP} \quad (7)$$

$$\frac{dPC_{NP}}{dt} = v_{3pc} \frac{PC_N}{k_p + PC_N} - v_{4pc} \frac{PC_{NP}}{k_{dp} + PC_{NP}} - v_{dpcn} \frac{PC_{NP}}{k_d + PC_{NP}} - k_{dn} PC_{NP} \quad (8)$$

Variables P_C and C_C correspond to the amount of PER and CRY proteins respectively in the cytosol, and PC_C and PC_N are the amount of PER-CRY complex in the cytosol and in the nucleus respectively. The complexation of PER and CRY follows the rate k_3 , with corresponding dissociation rate k_4 . The use of Hill functions in equations (1) and (2) simulate the inhibition of PER-CRY (PC_N) on PER (P_C) and CRY (C_C) with n as degree of cooperativity.

3.2 Model with 4 variables

In order to mathematically prove the existence of a limit cycle, we again simplify the model of Figure 2: the model of figure 3 does not incorporate the phosphorylation of the proteins and, consequently, contains four variables instead of eight variables. The dynamics of cytosolic PER protein (P_C) and

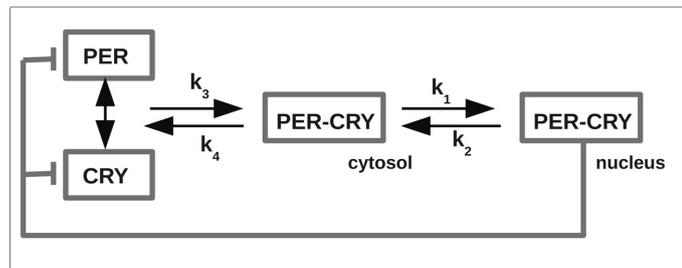


Figure 3: The 4 variable model in which the phosphorylation steps are omitted.

CRY protein (C_C) and cytosolic PER-CRY (PC_C) and nuclear PER-CRY (PC_N) is governed by the following system of four kinetic equations:

$$\frac{dP_C}{dt} = \frac{K^n}{K^n + PC_N^n} v_1 - k_3 P_C C_C + k_4 P C_C - kd_1 P_C \quad (9)$$

$$\frac{dC_C}{dt} = \frac{K^n}{K^n + PC_N^n} v_2 - k_3 P_C C_C + k_4 P C_C - kd_2 C_C \quad (10)$$

$$\frac{dP C_C}{dt} = k_3 P_C C_C - k_4 P C_C - k_1 P C_C + k_2 P C_N - kd_3 P C_C \quad (11)$$

$$\frac{dP C_N}{dt} = k_1 P C_C - k_2 P C_N - kd_4 P C_N \quad (12)$$

The next section is devoted to the investigation of the dynamics of the 4 variable model *via* numerical simulations and bifurcation diagram analysis. Most results were obtained using MATHEMATICA.

4 Analysing the 4-variable EDP model

Not surprisingly, the existence of the cycle depends on the degradation rate of the complex PER-CRY in the nucleus. Thus kd_4 is the control parameter. When all parameters are fixed except kd_4 , we observe that a Hopf bifurcation takes place with disappearance of the limit cycle for kd_4 crossing some value. This implies that for a sufficiently small value of kd_4 there exists one limit cycle.

To study the stability of the equilibrium points we apply the Lyapunov's first (indirect) method [23]. If $kd_4 = 0.41$, the system (9-12) has one fixed point. The Jacobian matrix of the system (9-12) at the equilibrium point is

$$\begin{pmatrix} -0.277892 & -0.162354 & 0.06 & -11.1584 \\ -0.227892 & -0.212354 & 0.06 & -12.2743 \\ 0.227892 & 0.162354 & -0.19 & 0.06 \\ 0 & 0 & 0.08 & -0.16 - kd_4 \end{pmatrix}$$

This matrix has 4 eigenvalues.

- Two of them are real eigenvalues with negative real part.
- The two others are complex conjugates and their common real parts:
 - are positive when the parameter kd_4 is less than the bifurcation value, which is close to 0.42 in our case (see the figure 4(a), 4(b) and 4(c))
 - and become negative when the parameter crosses this bifurcation value.

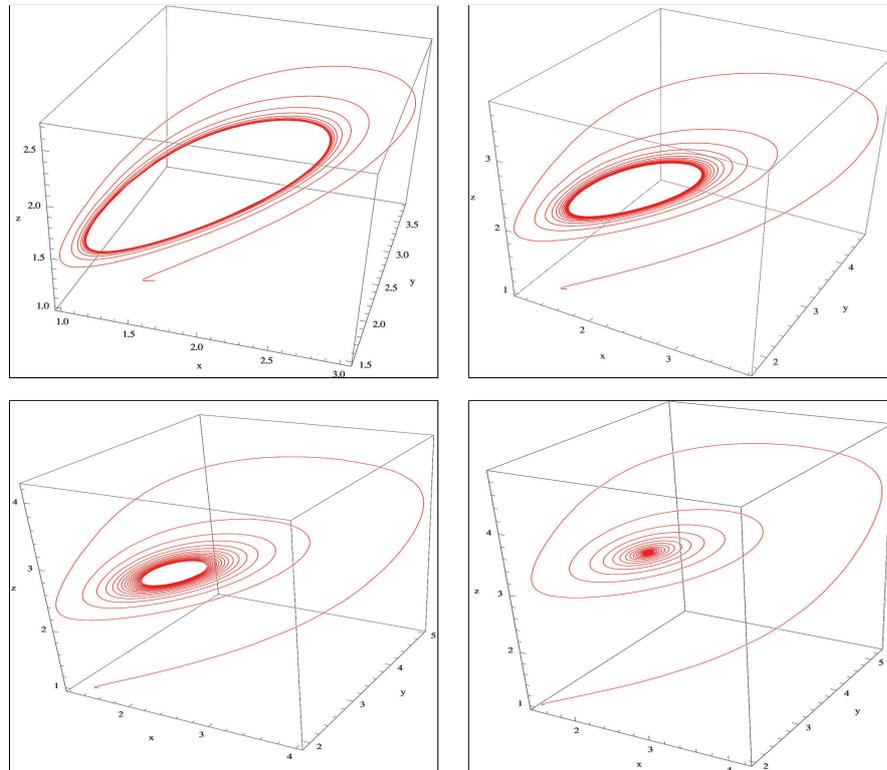


Figure 4: 3D projections of some phase trajectories. Common parameters are $K = 0.4$, $n = 15$, $v_1 = 2$, $v_2 = 2.2$, $k_1 = 0.08$, $k_2 = 0.06$, $k_3 = 0.08$, $k_4 = 0.06$, $kd_1 = 0.05$, $kd_2 = 0.05$ and $kd_3 = 0.05$, with $(1.5, 2, 1, 0.5)$ as initial condition. In (a), $kd_4 = 0.25$. In (b), $kd_4 = 0.35$. In (c), $kd_4 = 0.42$. In (d), $kd_4 = 0.55$. We observe a stable limit cycle for Figures (a), (b) and (c) only.

This computation can be made easily using the Routh-Hurwitz criterion [24] applied to the characteristic polynomial of the Jacobian matrix.

It is also possible to get the same bifurcation schema with kd_3 around 0.39, where kd_3 is the degradation rate of the complex PER-CRY in the cytosol.

Following Gérard, Gonze and Goldbeter in [25], we suspect that 3 variables are needed to observe the same behaviour (this will be the object of a future research).

5 Simplified purely discrete model

The discrete modelling frameworks aim at providing a logical explanation of the observed behaviours from a qualitative point of view. For biological

networks, René Thomas [26] established a framework where the continuous phase space is partitioned into qualitative regions in such a way that, within a region, the action of a species on other species is qualitatively constant. Trajectories are then abstracted by transitions from a region to another one. René Thomas has defined this abstraction in such a way that it is consistent with the continuous framework [27].

5.1 Thomas' discrete modeling framework

To define a Thomas' model one has first to design an *interaction graph* whose nodes are most of the time the molecular species of the biological network under consideration, and whose edges are labelled with signs: "+" for activations and "-" for inhibitions. For example, similarly to model of Figure 3, the genes *per* and *cry* have a positive action on the complex PER-CRY in the cytosol, which in turn has a positive action on the complex PER-CRY in the nucleus, and lastly the former inhibits the genes *per* and *cry*, see Figure 5.

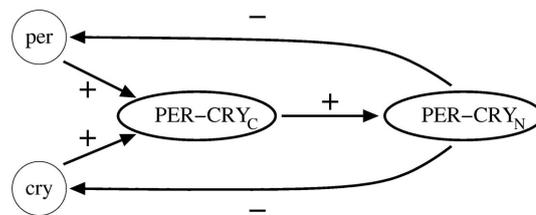


Figure 5: Thomas' model with 4 variables.

René Thomas has remarked that, when a gene x acts on a gene y , there is a threshold in the concentration space of x such that :

- if the concentration of x is below the threshold, x is unable to regulate y , and the regulation is inactive,
- if its concentration passes the threshold, x is able to regulate y and the regulation is active.

More generally, when x has an action on several targets $y_1, y_2 \dots y_k$, there are k thresholds (one threshold associated each target). Then from a qualitative point of view, x can have $k + 1$ different behaviours: either it does not regulate any target, either it regulates a unique target, either it regulates 2 targets, and so on. Thus, the concentration space of the variable x can be divided into $k + 1$ intervals where k is the number of targets. These intervals are called *abstract levels* and are denoted by integers in $[0, k]$.

The qualitative dynamics of the network is then defined by the successive *states* of each variable, where a state is an interval number. Let us consider a variable y regulated by x_1, x_2, \dots, x_n . A qualitative state being given, there is a subset ω of $\{x_1, x_2, \dots, x_n\}$ such that $x_i \in \omega$ if and only if the qualitative state of x_i denotes an interval greater than the $x_i \rightarrow y$ threshold. According to René Thomas, the abstract level toward which y is attracted only depends on ω . Consequently, if y has n regulators there are 2^n different parameters for variable y , which we denote by $K_{y,\omega}$, one for each possible subset ω of regulators of y .

A model of a gene network is consequently defined by

- an interaction graph $G = (V, E)$ where edges are labelled with a sign and an integer (which is the number of the interval immediately after the threshold),
- the family of parameters $\{K_{y,\omega}\}$, y being any member of V , and ω being any subset of $G^-(y)$, the set of predecessors of y in the graph G .

A model of a gene network being given (in particular all parameter values are known), we associate a state graph:

- a state is an assignment s that associates to each variable of the graph an integer value representing a possible interval for this variable (see above),
- a transition $s \rightarrow s'$ from one state to another state is obtained by modifying only *one* variable assignment as follows:
 - the modified variable x must satisfy $s(x) \neq K_{x,\omega}$ where ω is the set of active regulators of x according to the state s ,
 - if $s(x) < K_{x,\omega}$, then $s'(x) = s(x) + 1$ and $s'(y) = s(y)$ for all $y \neq x$,
 - if $s(x) > K_{x,\omega}$, then $s'(x) = s(x) - 1$ and $s'(y) = s(y)$ for all $y \neq x$.

Thus the dynamics of the model is *nondeterministic* and *asynchronous* because at a particular state, even if several variables are attracted towards a different value, only one variable can change simultaneously (because there is no reason that several variables reach their threshold exactly at the same time). Moreover variables never jump several thresholds as the discrete model is an abstraction of a continuous phenomenon.

5.2 Oversimplified circadian clock and introducing light influence

In first approximation the two negative cycles of Figure 5 are known to act roughly in the same time sustaining the circadian oscillation. Moreover the main role of *per* and *cry* is to produce the PER-CRY complex. It suggests that if we wish to simplify again, the next step is to amalgamate *per* and *cry* into an abstract “set of genes” that we denote G . Then we get a unique cycle with three nodes in which the only role of PER-CRY_C is to “produce” PER-CRY_N. So we remove the node PER-CRY_C, leading to the model of Figure 6 (left) where PER-CRY_N is abbreviated as PC .

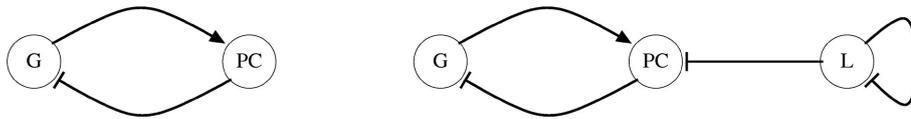


Figure 6: (left) A simplified model of the interactions in the mammalian circadian clock. (right) Adding the light variable in order to take into account the driving of the system by light/dark alternation.

Since light trains the rhythmicity of the mammalian circadian clock by inhibiting the transport of PER-CRY_C into the nucleus, it has an inhibitor effect on complex PC in Figure 6 (right).

Moreover because of the natural alternance of day and night, the model has to be able to produce an oscillation of light. This is done by adding a fictitious auto-inhibition: light can be “on” or “off”, when it is “on”, it has an inhibitor effect on itself and light tends to go “off” and conversely.

5.3 Identifying the discrete parameters

The interaction graph being given, it remains many possible parameter values leading to a large number of possible models with different state graphs. Obviously, a parameterization is valid if the corresponding state graph is coherent with the biological knowledge about the dynamics of the system.

Let us first treat the parameters for the variable L (light). The fictitious feedback loop must generate oscillations, thus $K_{L,\{L\}} = 1$ and $K_{L,\{G\}} = 0$ (otherwise no oscillation on L could be observed).

We also know that the circadian clock oscillates in constant darkness ($L = 0$). So, the feedback loop $G \leftrightarrow PC$ generates oscillation and we must have $K_{G,\{G\}} = 1$ and $K_{G,\{PC\}} = 0$ (otherwise no sustained oscillation on L could be observed).

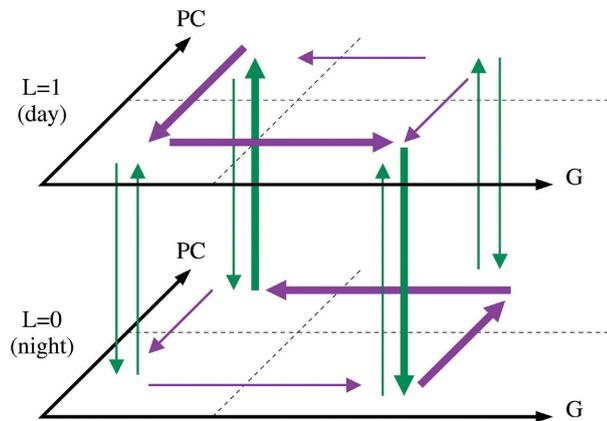


Figure 7: Qualitative dynamics of a simplified model.

It remains to identify the parameters associated with PC . Four situations have to be studied :

1. When G is off and light is on, there is no transcription of PC and existing complexes PER-CRY are kept outside the nucleus by light. Then PC (in the nucleus) cannot increase nor stay at 1. This property exactly means $K_{PC,\{L\}} = 0$.
2. When G is on and light is on, there is transcription of PER-CRY in the cytosol but light prevents these complexes to enter the nucleus. In other words $K_{PC,\{G,L\}} = 0$.
3. When G is off and light is off, there is no transcription of PER-CRY complexes in the cytosol and *a fortiori* they do not enter the nucleus, leading to $K_{PC,\{\}} = 0$.
4. When G is on and light is off, then PER-CRY complexes accumulate in the cytosol and in the absence of light, they can enter the nucleus. So, $K_{PC,\{G\}} = 1$.

The parameters being identified and following section 5.1, we build the state graph of Figure 7. Green transitions stand for alternation of days and nights whereas blue ones represent the evolution of genes G and complexes PC . Thick transitions sketch the common path for a regular day-night alternation; let us start from the state $(L, G, PC) = (1, 0, 1)$:

- During the day, complexes PC (in the nucleus) are first degraded (transition $(1, 0, 1) \rightarrow (1, 0, 0)$) and genes are transcribed (transition $(1, 0, 0) \rightarrow (1, 1, 0)$).

- Then night falls (transition $(1, 1, 0) \rightarrow (0, 1, 0)$).
- During the night, the concentration of complexes PC grows and becomes sufficient to inhibit G (transitions $(0, 1, 0) \rightarrow (0, 1, 1)$ and $(0, 1, 1) \rightarrow (0, 0, 1)$).
- At dawn, the transition $(0, 0, 1) \rightarrow (1, 0, 1)$ closes the circadian cycle.

Notice that, whatever the state, one can switch light *ad libitum*. This results from the negative auto-inhibition of L that, within the standard logical approach of René Thomas, cannot be temporized (all vertical transitions of Fig. 7).

6 Discrete model with delays

Clearly, a valuable model of the circadian clock needs to integrate chronometric information to predict behaviours which could usefully be confronted with experimental data. The idea developed in this section is to manually associate with each transition of the qualitative state graph, a delay which represents the time spent to go from a state to a neighbour state along a transition. This simple solution is made possible by the small number of transitions in our state graph. For bigger models more elaborated *hybrid* frameworks are required: Siebert and Bockmayr proposed an automaton product which allows to handle time [28], Batt and co-workers used verification tools on timed automata to learn about the possible qualitative behaviors of the network under a whole range of uncertain delay parameters [29]. We also proposed such sophisticated frameworks [30, 31, 32] where one of our motivations was to take into consideration small successive accumulations.

6.1 Labelling the state graph with delays

In Figure 8 we label each transition of the state graph with a delay. For example, if we consider that days and nights last the same duration, each of vertical transitions are labelled with 12 hours (others transitions require some more elaborated reasoning).

Let us first remark that in the setting of René Thomas, at any state and consequently at any time, it is impossible to have two opposite transitions. If a variable can evolve, then it can *either* increase or decrease. We consequently introduce a set of clocks, one clock per variable. When active, the clock of a given variable measures the time elapsed since that variable can increase (resp. decrease). Within one state, if there are several outgoing transitions, the next triggered transition will be the one which is associated with the variable whose clock reaches its delay.

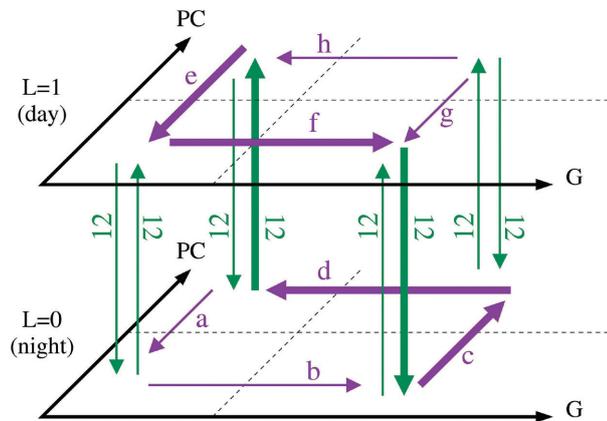


Figure 8: A qualitative model with delays remaining unknown

When a transition is fired, the clock of the involved variable, say x , is of course reset to 0. But some other clocks have also to be reset: When a variable is directly influenced by x , its possibility of variation may change, in which case the corresponding clock has also to be reset to 0.

6.2 Identification of delays for the mammalian circadian clock

The goal of this subsection is to identify the parameters a to g of Figure 8 using standard knowledge about the circadian cycle:

- During the night, the time necessary for firing transitions associated with delays c and d is equal to 12:

$$c + d = 12 \quad (13)$$

A shorter sum would lead to trigger the transition labelled by a before the up light transition of the figure. Conversely a longer sum would lead to trigger the up light transition before the end of the transition labelled by d .

- The trajectory from $(L = 1, G = 0, PC = 1)$ to $(L = 1, G = 1, PC = 0)$ is performed during the day. Thus,

$$e + f \leq 12 \quad (14)$$

A longer sum would lead to trigger the down light transition before the end of the transition labelled by f . On the contrary, a sum shorter than 12 hours will lead to the state $(l = 1, G = 1, PC = 0)$, which is stable in

the plane $L = 1$ and consequently the trajectory will stay there, waiting for the down light transition without modifying the discrete trajectory in the state graph.

- In constant night, the mammalian circadian cycle stabilizes around 24 hours, see [33, 34]. Thus,

$$a + b + c + d = 24 \quad (15)$$

- The transcription of genes G does not depend directly on the light L . Then

$$b = f \quad (16)$$

$$d = h \quad (17)$$

- When light is on, PC is inhibited. Thus a is notably smaller than e :

$$a > e \quad (18)$$

The resulting constraints (13-18) admit an infinity of solutions. The following delays satisfy these constraints and seem sensible with respect to current knowledge:

$$\begin{array}{cccc} a = 7 & b = 5 & c = 7 & d = 5 \\ e = 1 & f = 5 & g = ? & h = 5 \end{array}$$

The delay g is not valuated. Indeed, for the purpose of this chapter, the interest of the transition labelled by g is only that it induces a stable state in the plane $L = 1$, whatever the actual value g .

6.3 Robustness to day length

The parameters values proposed in the previous subsection are remarkably robust to day length variations. The previous model imposes a constant and equal durations of nights and days. Indeed, it reflects the durations of nights and days during the periods of equinoxes. In the northern hemisphere at the latitude 50 degrees north, the duration of night increases up to 16 hours in the winter, see Figure 9-left, and it decreases down to 8 hours in the summer, see Figure 9-right.

To illustrate the clock manipulation in our dynamics, let us consider the winter model (left part of the figure).

1. Let us start from the beginning of the night with a state where $(L, G, PC) = (0, 1, 0)$ and all clocks are set to 0. Because $c = 7$ is smaller than 16, the next state is $(L, G, PC) = (0, 1, 1)$. The clock of PC is of course reset to 0 and the clock of G is also reset to 0 because PC now inhibits G . And it remains 9 hours for light to go up.

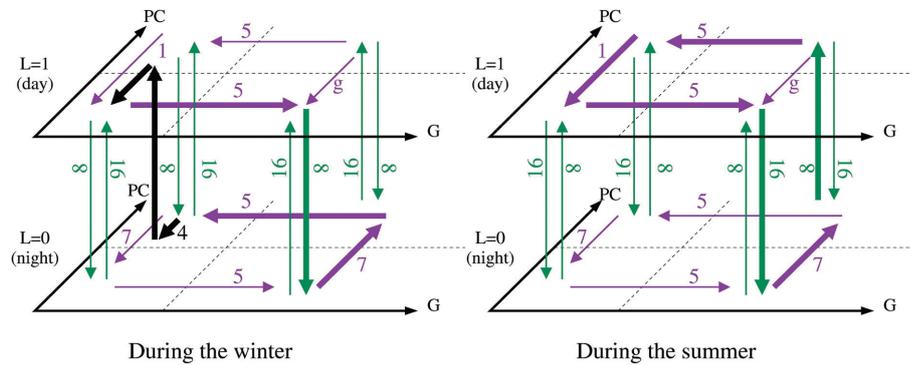


Figure 9: Influence of seasons on the qualitative model of the mammalian circadian cycle. **(left)** during the winter. **(right)** during the summer.

2. From $(0, 1, 1)$, because $d = 5$ is smaller than 9, the next state is $(0, 0, 1)$. The clock of G is reset to 0 and the clock of PC is also reset to 0 because G does not activate PC anymore. And it remains 4 hours for light to go up.
3. From $(0, 0, 1)$, because $a = 7$ is greater than 4, the next state is $(1, 0, 1)$. The clock of L is reset to 0 and the clock of PC is not reset to 0 because it continues to decrease. However, $e = 1$ and $a = 7$ and in first approximation we can consider that $\frac{4}{7}$ of PC have been degraded. Consequently, the clock for PC becomes $\frac{4 \times e}{a} = \frac{4}{7}$. It remains $\frac{3}{7}$ of an hour for PC to become equal to 0.
4. From $(1, 0, 1)$, because $\frac{3}{7}$ is smaller than 8, the next state is $(1, 0, 0)$. The clock of PC is reset to 0 and the clock of G is also reset to 0 because PC now activates G . And it remains $8 - \frac{3}{7} = 7 + \frac{4}{7}$ hours for light to go down.
5. From $(1, 0, 0)$, because $f = 5$ is smaller than $7 + \frac{4}{7}$, the next state is $(1, 1, 0)$. The clock of G is reset to 0 and the clock of PC is also reset to 0 because $(G, PC) = (1, 0)$ is a stable state in the plane $L = 1$ and consequently, PC does not decrease anymore. And it remains $2 + \frac{4}{7}$ hours for light to go down.
6. After $2 + \frac{4}{7}$ hours, we leave the state $(1, 1, 0)$ to go back to $(0, 1, 0)$. The clock of L is reset to 0, the clock of PC is also reset to 0, and the clock of G is equal to $2 + \frac{4}{7}$.
7. From $(0, 1, 0)$, because $c = 7$ is smaller than 16, the next state is $(0, 1, 1)$. The clock of PC is reset to 0 and the clock of G is also reset to 0; it remains 9 hours for light to go up and the circadian cycle is phased again.

The summer model behaves similarly (right part of the figure) leading to the observation that our model is robust to the alternance of seasons.

7 Discussion

The designer of a model in the context of genomics is often torn between two options: to elaborate a rich model reflecting as much biological knowledge as possible in a consistent way, or to design a simplified model dedicated to a given family of questions in order to study the main causalities at a coarse-grained scale. This chapter was inspired by the second philosophy.

This approach may induce drastic simplifications, however it provides chains of causalities in the broad outlines, which make it possible to apprehend a problem in its entirety. For example, it is frequently observed that a jet lag from East to West induces less disorders than travels from West to East. Indeed:

- When one travels from East to West, one increases the awakening time, thus one increases the time of exposure to light. In the simplified model of Figure 7, we observe that a longer stay in the upper states (where $L = 1$) does not change the trajectories because they reach the locally stable state $(L, G, PC) = (1, 1, 0)$, and when night comes, the cycle is immediately synchronised again.
- When one travels from West to East, one does not feel like sleeping, the model continuing its trajectory until the locally stable state $(L, G, PC) = (1, 1, 0)$. Since the full day duration is shorter, the night length ($L = 0$) is reduced. At night ($(L, G, PC) = (0, 1, 0)$), for a sufficiently significant jet lag, the system does not reach the state $(L, G, PC) = (0, 1, 1)$ before light switches on again. And when light switches on, the system is “prisoner” of the transition $(1, 1, 1) \rightarrow (1, 1, 0)$, unable to cross the states $(1, 0, 1)$ and $(1, 0, 0)$ and the circadian cycle is consequently strongly disturbed.

Of course, this explanation is crude. It nevertheless gives a good intuitive idea of the main causality mechanisms involved in jet lag.

More generally, reducing the number of variables may help establishing mathematical results, as illustrated in Section 4 where we have established the existence of a limit cycle. Many other global properties can be established, and, afterwards, studying how far the properties established on a simple model can be propagated on a more detailed model, belongs to the scope of multilevel modelling.

References

- [1] De Maria, E., Fages, F., Rizk, A., Soliman, S.: Design, optimization and predictions of a coupled model of the cell cycle, circadian clock,

- DNA repair system, irinotecan metabolism and exposure control under temporal logic constraints. *Theor. Comput. Sci.* **412**(21) (2011) 2108–2127
- [2] Murray, J.: *Mathematical Biology, I. An introduction*. Springer (2002 (third edition))
- [3] Thomas, R., d'Ari, R.: *Biological Feedback*. CRC Press (1990)
- [4] Chabrier, N., Fages, F.: Symbolic model checking of biochemical networks. In: *CMSB'03: Proceedings of the first workshop on Computational Methods in Systems Biology*. Volume 2602 of *Lecture Notes in Computer Science.*, Springer-Verlag (2003) 149–162
- [5] Gillespie, D.: A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Physics* **22**(4) (1976) 403–434
- [6] Leloup, J.C., Goldbeter, A.: Modeling the mammalian circadian clock: Sensitivity analysis and multiplicity of oscillatory mechanisms. *J.T.B.* **230** (2004) 541–562
- [7] Forger, D., Peskin, C.: Stochastic simulation of the mammalian circadian clock. *PNAS* **102**(2) (2005) 321–324
- [8] Reghunandanan, V., Reghunandanan, R.: Neurotransmitters of the suprachiasmatic nuclei. *J. of Circadian Rhythms* **4** (2006) 2
- [9] Bell-Pedersen, D., Cassone, V., Earnest, D., Golden, S., Hardin, P., Thomas, T., Zoran, M.: Circadian rhythms for multiple oscillators: lessons from diverse organisms. *Nat. Rev. Genet.* **6**(7) (2005) 544–556
- [10] Klein, D., Moore, R., Reppert, S.: *Suprachiasmatic nucleus, the mind's clock*. Oxford University Press (1991)
- [11] Lesauter, J., Silver, R.: Output signals of the SCN. *Chronobiol. Int.* **15** (1998) 535–550
- [12] Weaver, D.: The suprachiasmatic nucleus: a 25-year retrospective. *J.Biol.Rhythms* **13** (1998) 100–112
- [13] Johnson, R., Moore, R., Morin, L.: Loss of entrainment and anatomical plasticity after lesions of the hamster retinohypothalamic tract. *Brain Res.* **460** (1988) 297–313

- [14] Maywood, E., O'Neill, J., Reddy, A., Chesham, J., Prosser, H., Kyriacou, C., Godinho, S., Nolan, P., Hastings, M.: Genetic and molecular analysis of the central and peripheral circadian clockwork of mice. In: *Cold Spring Harb Symp Quant Biol*. Volume 72. (2007) 85–94
- [15] Gallego, M., Virshup, D.: Post-transcriptional modifications regulate the ticking of the circadian clock. *Nature Reviews Molecular Cell Biology* **8** (2007) 139–148
- [16] Sato, T., Panda, S., Miraglia, L., Reyes, T., Rudic, R., McNamara, P., Naik, K., Fitzgerald, G., Kay, S., Hogenesch, J.: A functional genomics strategy reveals Rora as a component of the mammalian circadian clock. *Neuron*. **43**(4) (2004) 527–37
- [17] Belden, W., Loros, J., Dunlap, J.: CLOCK leaves its mark on histones. *Trends Biochem Sci*. **31**(11) (2006) 610–3
- [18] Eide, E., Vielhaber, E., Hinz, W., Virshup, D.: The circadian regulatory proteins BMAL1 and cryptochromes are substrates of casein kinase I ϵ . *J. Biol. Chem*. **277**(19) (2002) 17248–54
- [19] Froy, O., Miskin, E.: Effect of feeding regimens on circadian rhythms: implications for aging and longevity. *Aging* **2**(1) (2010) 7–27
- [20] Albrecht, U. and Bordon, A., Schmutz, I., Ripperger, J.: The multiple facets of Per2. *Cold Spring Harb. Symp. Quant. Biol.* **72** (2007) 95–104
- [21] Maury, E., Moynihan Ramsey, K., Bass, J.: Circadian rhythms and metabolic syndrome: from experimental genetics to human disease. *Circ. Res.* **106** (2010) 447–462
- [22] Vanselow, K., Kramer, A.: Role of phosphorylation in the mammalian circadian clock. *Cold Spring Harb. Symp. Quant. Biol.* **72** (2007) 167–76
- [23] Lyapunov, A.: *The general problem of the stability of motion*. Taylor & Francis, London, UK (1992)
- [24] Gradshteyn, I., Ryzhik, I.: *Table of Integrals, Series and Products*. Translation edited and with a preface by A. Jeffrey and D. Zwillinger, Academic Press, San Diego, Calif, USA, 6th edition (2000)
- [25] Gérard, C., Gonze, D., Goldbeter, A.: Dependence of the period on the rate of protein degradation in minimal models for circadian oscillations. *Phil. Trans. R. Soc. A* **367** (2009) 4665–4683

- [26] Thomas, R.: Logical analysis of systems comprising feedback loops. *J. Theor. Biol.* **73**(4) (1978) 631–56
- [27] Snoussi, E.: Qualitative dynamics of a piecewise-linear differential equations : a discrete mapping approach. *Dynamics and stability of Systems* **4** (1989) 189–207
- [28] Siebert, H., Bockmayr, A.: Incorporating time delays into the logical analysis of gene regulatory networks. In: CMSB. Volume 4210 of LNCS. (2006) 169–183
- [29] Batt, G., Ben Salah, R., Maler, O.: On timed models of gene networks. In: FORMATS. Volume 4763 of LNCS., Springer (2007) 38–52
- [30] Ahmad, J., Bernot, G., Comet, J.P., Lime, D., Roux, O.: Hybrid modelling and dynamical analysis of gene regulatory networks with delays. *ComplexUs* **3**(4) (2007) 231–251
- [31] Comet, J.P., Bernot, G.: Introducing continuous time in discrete models of gene regulatory networks. In: Proc. of the Nice Spring school on Modelling and simulation of biological processes in the context of genomics. EDP Sciences, ISBN : 978-2-7598-0545-7 (2010) 61–94
- [32] Comet, J.P., Fromentin, J., Bernot, G., Roux, O.: A formal model for gene regulatory networks with time delays. In: CSBio'2010. Volume 115 of CCIS., Springer (2010) 1–13
- [33] Oster, H., Yasui, A., van der Horst, G., Albrecht, U.: Disruption of mCry2 restores circadian rhythmicity in mPer2 mutant mice. *Genes Dev.* **16** (2002) 2633–2638
- [34] Van der Horst, G., Muijtjens, M., Kobayashi, K., Takano, R., Kanno, S., Takao, M., de Wit, J., Verkerk, A., Eker, A., van Leenen, D., Buijs, R., Bootsma, D., Hoeijmakers, J., Yasui, A.: Mammalian Cry1 and Cry2 are essential for maintenance of circadian rhythms. *Nature* **398** (1999) 627–630
- [35] Ahmad, J., Roux, O., Bernot, G., Comet, J.P., Richard, A.: Analysing formal models of genetic regulatory networks with delays: Applications to lambda phage and T-cell activation systems. *Int. J. Bioinformatics Research and Applications* **4**(3) (2008) 240–262
- [36] Ahmad, J., Bernot, G., Comet, J.P., Lime, D., Roux, O.: Hybrid modelling and dynamical analysis of gene regulatory networks with delays. *ComplexUs* **3**(4) (2006, Cover Date: November 2007) 231–251

Appendix

Theorem: Let $x = x^*$ be an equilibrium point of a nonlinear system:

$$\dot{x} = f(x),$$

where $f : D \rightarrow \mathcal{R}^n$ is continuously differentiable and $D \subset \mathcal{R}^n$ is the neighbourhood of the equilibrium point x^* . Let λ_i denote the eigenvalues of the Jacobian matrix $A = \frac{\delta f}{\delta x} |_{x^*}$ then the following are considered.

- If $Re\lambda_i < 0$ for all i , then $x = x^*$ is asymptotically stable.
- If $Re\lambda_i > 0$ for one or more i , then $x = x^*$ is unstable.
- If $Re\lambda_i \leq 0$ for all i and at least one $Re\lambda_j = 0$, then $x = x^*$ may be either stable, asymptotically stable, or unstable.

Since A is only defined at x^* , stability determined by the indirect method is restricted to infinitesimal neighborhoods of x^* .

To study the signs of the real parts of eigenvalues, we have the following criterion [24].

Routh-Hurwitz Criterion:

Given the polynomial $P(\lambda) = \lambda^n + a_1\lambda^{n-1} + \dots + a_{n-1}\lambda + a_n$, where the coefficients $a_i, i = 1, 2, \dots, n$ are real constants, define the n Hurwitz matrices

$$\begin{aligned} H_1 &= [a_1], \\ H_2 &= \begin{bmatrix} a_1 & 1 \\ a_3 & a_2 \end{bmatrix} \\ &\vdots \\ H_n &= \begin{bmatrix} a_1 & 1 & 0 & 0 & \cdots & 0 \\ a_3 & a_2 & a_1 & 1 & \cdots & 0 \\ a_5 & a_4 & a_3 & a_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & a_n \end{bmatrix}, \end{aligned}$$

where $a_i = 0$ if $i > n$.

All of the roots of the polynomial have negative real part if and only if the determinants of all Hurwitz matrices are positive: $\det H_i > 0, i = 1, 2, \dots, n$. Routh-Hurwitz criteria for $n = 3$ are $a_1 > 0, a_3 > 0$ and $a_1 a_2 - a_3 > 0$ and for $n = 4$ are $a_1 > 0, a_2 > 0, a_3 > 0, a_4 > 0$ and $a_1 a_2 a_3 a_4 - a_1^2 a_4^2 - a_3^2 a_4 > 0$.

Metabolism Modelling

Jean-Pierre Mazat¹

¹ IBGC CNRS UMR 5095 et Université Bordeaux 2

Glossary

- Metabolism: the metabolic network at work; often synonymous with metabolic network.
- Enzyme: protein with catalytic activity. Synthesized in the cell by other proteins and protein- RNA complexes.
- Enzymatic reaction: biochemical reaction catalysed by an enzyme.
- Metabolic network: The metabolic network or metabolism (which is more the metabolic network at work) is the set of all the reactions taking place in a cell.
- Metabolite: molecule synthesized, degraded or/and transformed in the cellular metabolism.

1 Introduction

1.1 What is a metabolic network ?

Cells house a great number of chemical reactions, which split the nutriment we eat in smaller molecules and produce energy (ATP molecules for instance) from the oxygen we breathe in.

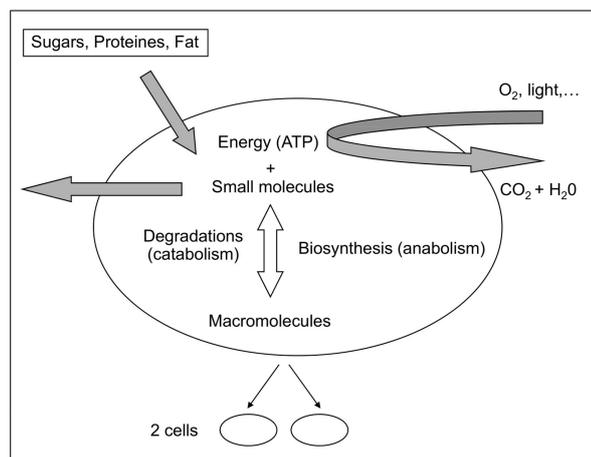


Figure 1: Cell metabolism, or the cell as an open system.

From these molecules other reactions take place to synthesize the basic molecules of the cell, such as amino-acids, nucleotides and nucleosides, lipids, etc. These molecules have a molecular weight (MW) between about twenty (18 for water) and about a thousand (NADH, lipids, etc.; see table 1). With these elementary molecules, the cell can synthesize several macromolecules (protein and nucleic acid biosynthesis for instance). As a matter of fact, organisms are built almost entirely from water and about thirty small precursor molecules (amino acids, aromatic bases of nucleic acids, sugars, palmitate, glycerol and choline).

All these synthesis and degradations constitute the *cell metabolism* or the *metabolic network* or the *metabolism*. In addition these reactions are regulated and the regulation network is presumably as or more complex than the cell metabolism itself.

In this chapter, for the sake of simplicity, we will limit ourselves to the part of metabolism which concerns the synthesis (anabolism) and the degradation (catabolism) of the small basic molecules of the cell (amino acids, aromatic bases of nucleic acids, sugars, palmitate, glycerol and choline etc.). These molecules will be called *metabolites* in the following. We will not deal with the regulation of the metabolic network

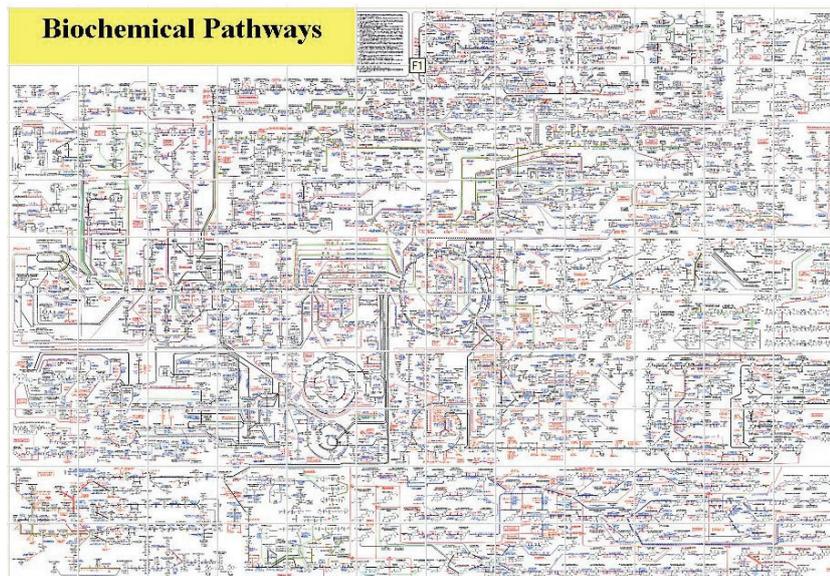


Figure 2: The metabolic network.

Even with all these restrictions, there remains a complex system (Fig. 2) with about 750 reaction and about 500 metabolites, slightly variable according to the cell type.

Computer scientists tend to derive a graph representation of the metabolic maps (Fig. 3). One has however to be aware that most of the metabolic reactions are bi-molecular, i.e. involve two substrates and two products, which complicate the graph representation because the two products of a reaction are usually not the two substrates of the following reaction.

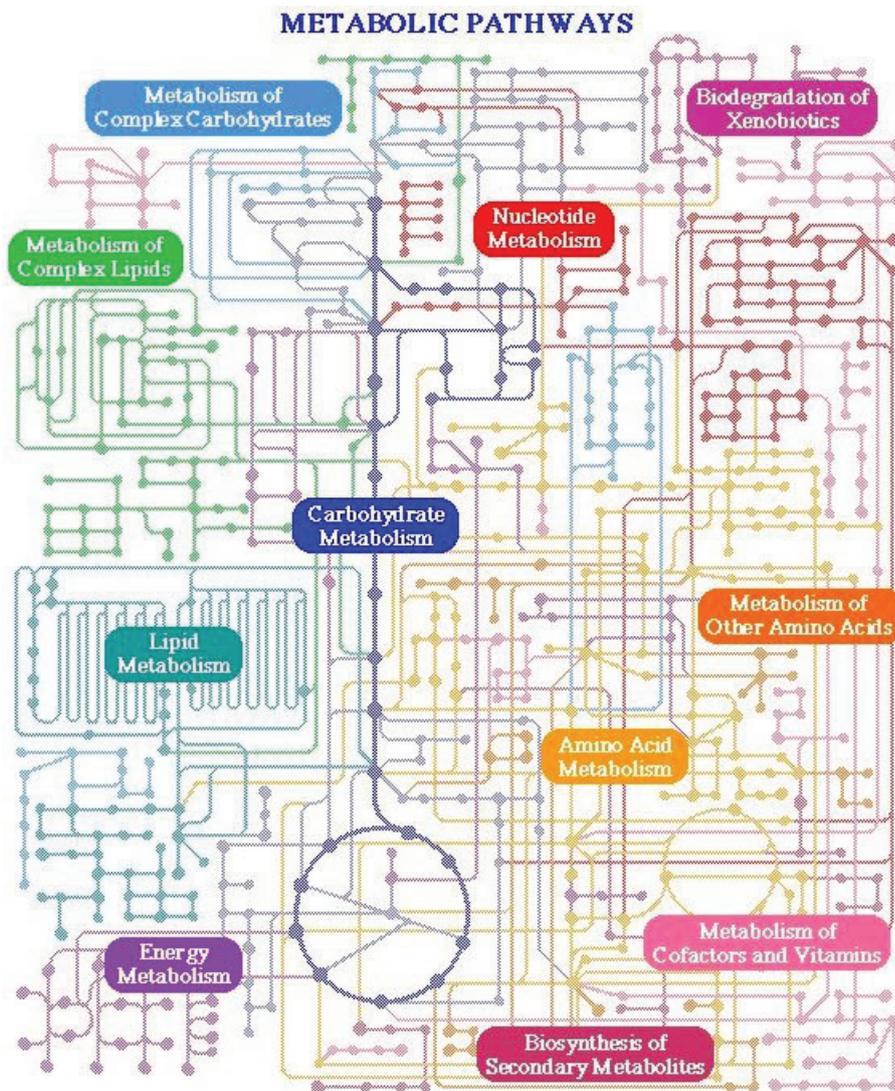


Figure 3: The structure of the metabolic network (from Kyoto Encyclopedia of Genes and Genomes). In this map, each dot represents an intermediate; each line represents an enzyme that acts on an intermediate

1.2 Why do cells use enzymes ?

Most of the cell reactions are either impossible in the living conditions or very slow. Thus they need to be catalyzed. The enzymes are these catalyzers. Enzymes are proteins, i.e. linear chain of amino acids which fold, adopting a special conformation able to favour a particular reaction.

The molecular weight of an enzyme is usually between 10000 and 1 million; it means the the MW of an enzyme is at least two order of magnitude higher than those of substrates and products (however, an enzyme can also act on other enzymes or more generally on macromolecules; we wil not deal with these cases here); the interactions of the substrates at the surface or in pockets of the enzyme put them in special conformations (analog to transition complex) and in this way favour the reaction.

It must be pointed out that the specificity of an enzyme is twofold. First there is a specificity of reaction : the enzyme E1 transforms the substrate S in P1 and another enzyme E2 transforms the substrate S in another product P2 etc. For instance pyruvate dehydrogenase (1.2.1.51) transforms pyruvate in acetyl-CoA, while lactate dehydrogenase (1.1.1.27) gives lactate from the same metabolite pyruvate. The second specificity is the one regarding substrates binding; for instance pyruvate dehydrogenase will bind only pyruvate; when metabolism needs to perform the same reaction on another substrate, α -ketoglutarate for instance, another enzyme has to be designed the α -ketoglutarate-dehydrogenase, specific of α -ketoglutarate (and not of pyruvate) and able to perform exactly the same chemical reaction.

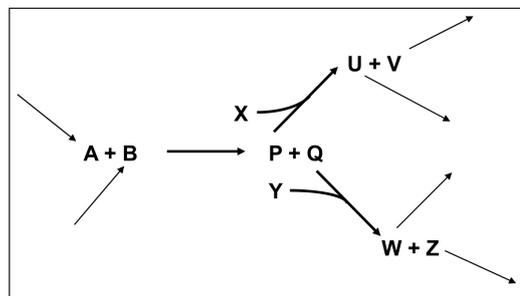


Figure 4: Entanglement of metabolic reactions

Let us nevertheless emphasise that specificities are not usually completely strict. Other metabolites, more or less chemically related to the “official” substrate are able to bind the same binding site on a given enzyme. This specificity notion (scale) is closely related to the problem of enzymes (genes) annotation : what is (are) the “official” substrate(s) of an enzyme? In other words which name will be given to a given enzyme.

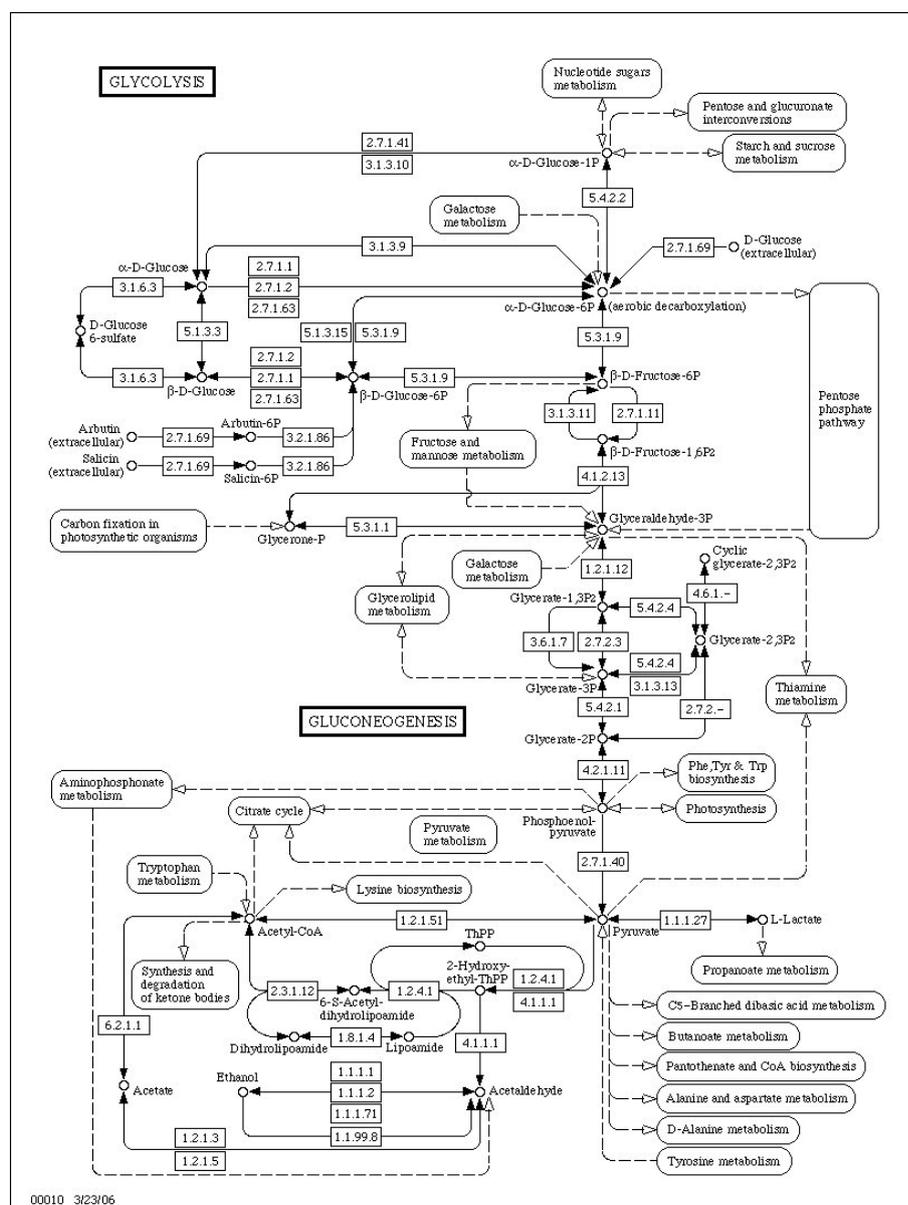


Figure 5: The pyruvate cross-roads (from KEGG)

Usually the substrate with the better affinity for an enzyme is the "official" one, but it is not always the case due to historical reasons. Furthermore, some enzyme can have a wide specificity accommodating several related substrates (for instance hexokinase with hexoses, which are the family of sugars with six carbon atoms).

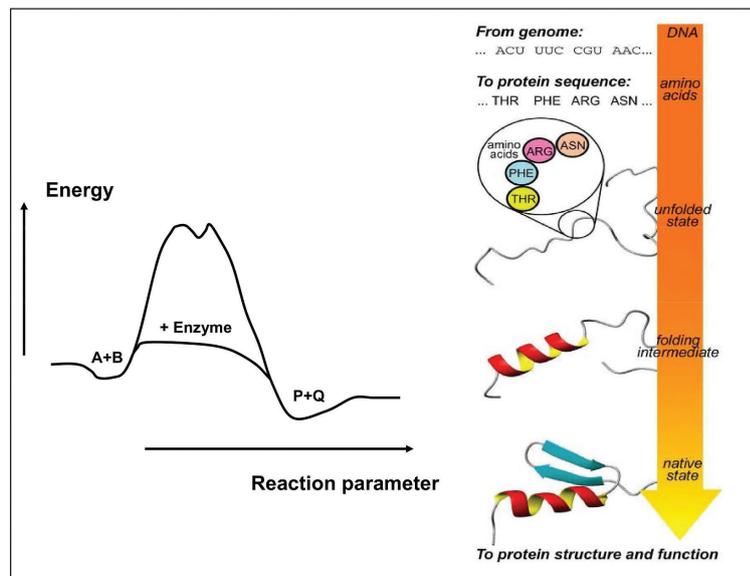


Figure 6: Enzyme folding

The last remark concerns the dimensions. On Fig. 7 are drawn, at the same scale, an average enzyme (diameter 5nm) and an average metabolite (0.5 nm). It is better understood how an enzyme can play its catalysing role in bending the substrate in an adequate, reacting, conformation. Table 1 gives also some dimension of biological objects.

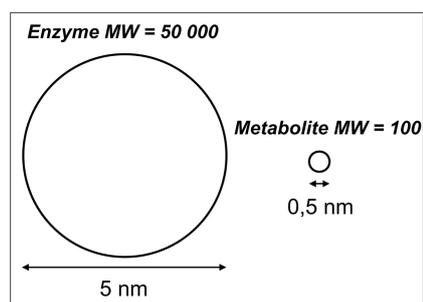


Figure 7: Average enzyme and average metabolite at the same scale.

Type	Size (nm)	MW	Observations
Water (H_2O)	0.26	18	
Alanine (Amino acid)	0.5	89	
Glucose	0.7	180	
ATH	1.6	507	MW of acid form
NADH		741	dipotassium salt
Phospholipid	3.5	750	
Myoglobin	3.6	16900	small protein
Hemoglobin	6.8	65000	medium protein
Cytochrom b		42592	part of complex of respiratory chain in mitochondria
Complex III	$7 \times 7 \times 11$	240000	
Ribosome (E. coli)	18	$2.8 \cdot 10^6$	
Membrane	7-9		thickness
Lysosome	250-500		
Peroxisome	500		
Mitochondrion	1.5μ		
<i>E. coli</i>	2μ	2 pg	
Nucleus	$4 \mu - 6 \mu$		
Chloroplast	8μ	60 pg	spinach leaves
Hepatic cell	20μ	$8 \mu g$	

Table 1: Sizes of molecules and cells.

2 Enzyme kinetics

2.1 Essential concepts

In metabolism modelling it is essential to know the *rates* with which enzymes catalyse the reactions. Biochemists often intermingle the reaction R_i catalyzed by the enzyme E_i with the rate V_i ; it will be sometimes necessary in modelling to precise the variable actually under study. From analogy with the chemical catalysis, V. Henri (Henri, 1902 and 1903) proposed that the substrate bind firstly to the enzyme and then reacts to give the product; this scheme will be extended later by Briggs and Haldane in a fundamental paper of one and a half page long (Briggs and Haldane, 1925).

Modelling of this scheme in condition for which $P_0 = 0$ (the usual conditions of measurement which are different of the *in vivo* conditions) is written in the following dynamical system:

$$\begin{cases} \frac{d[ES]}{dt} = k_1[E][S] - (k_{-1} + k_2)[ES] & (1) \\ v = \frac{d[P]}{dt} = k_2[ES] & (2) \end{cases}$$

with the conservation equations:

$$\begin{cases} [E] + [ES] = [E]_{total} & (3) \\ [S] + [ES] = [S]_{total} = [S]_0 & (4) \end{cases}$$

List of variables (4): [E], [ES], [S] and [P] with 2 conservation relationships
 \rightarrow 2 independent variables [ES] and [P] for instance.

List of parameters (3): k_1 , k_{-1} and k_2 .

Initial conditions: $[E]_0 = [E]_{total}$, $[ES]_0 = 0$, $[S]_0 = [S]_{total}$, $[P] = 0$

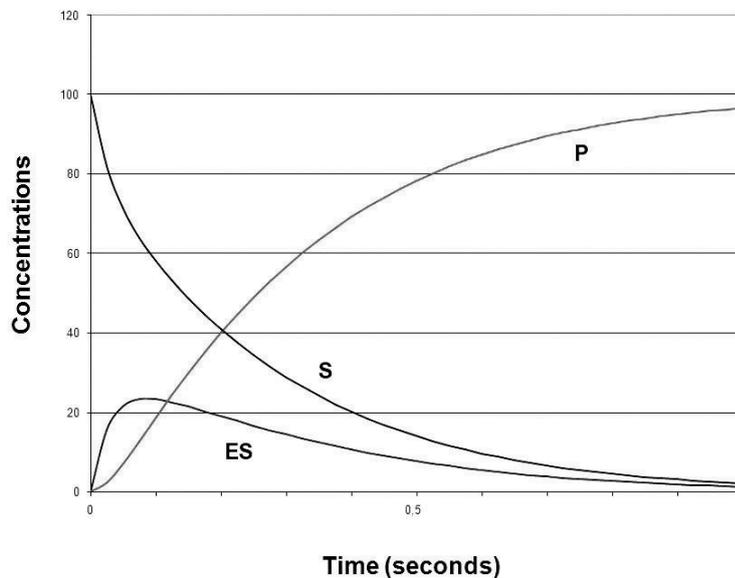


Figure 8: Time course of the different species of metabolites and enzymes in the simple enzymatic reaction: $E + S \leftrightarrow ES \rightarrow E + P$, where $E_{total} = 100$, $S_{total} = 100$, $k_1 = 1$, $k_{-1} = k_2 = 100$

At the beginning (Fig. 8), S binds E to form the complex [ES] which dissociates partially to form P. If the abscissa is time in seconds, one can see that the phenomenon is too fast to be easily observed. An experimental solution to this problem is to decrease the quantity of enzyme; because an enzyme is a

catalyst, its quantity can be in minute amount. One get the behaviour described in Fig. 9).

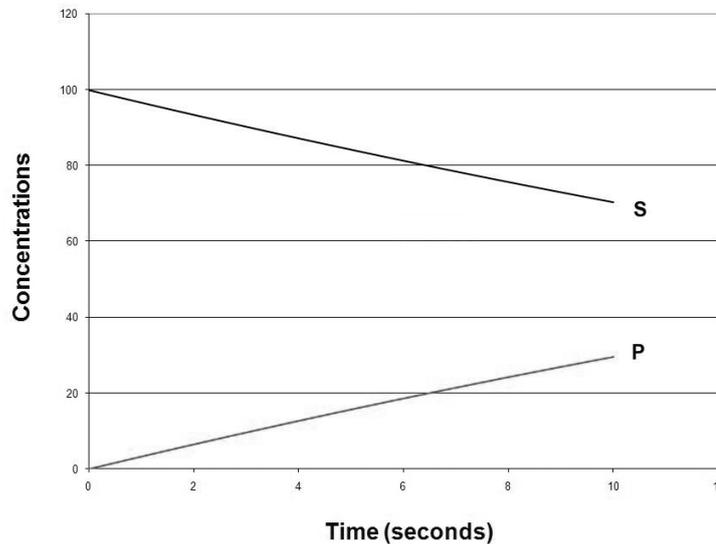


Figure 9: Time course of the different species of metabolites and enzymes in the simple enzymatic reaction: $E + S \leftrightarrow ES \rightarrow E + P$ with low enzyme concentration, where $E_{total} = 0.1$, $S_{total} = 100$, $k_1 = 1$, $k_{-1} = k_2 = 100$

The time course of P production is much longer and can be observed during several seconds. Furthermore the P production is quasi linear, i.e. is performed at a constant rate Fig. 9. This is due to the fact that the variation of [ES] is weak (between 0.033 and 0.026) and cannot be seen on the graph. In these conditions one can write $d[ES]/dt = 0$ which allows to derive the Henri-Michaelis-Menten equation proposed below in paragraph 1.

It should be pointed out also that in these conditions, the first reaction ($E + S \leftrightarrow ES$) is displaced from its thermodynamic equilibrium which gives $[ES]_{eq} = [E]_{eq} = 0.05$ ($K_{eq} = [E] \times [S]/[ES] = k_{-1}/k_1 = 100$). This is due to the high value of k_2 which is of the same order than k_{-1} .

In conclusion one can see that modelling with differential equations is easy. One can model metabolic networks with hundreds of kinetic equations and follow the time course of the metabolites concentrations towards a possible steady-state. However such a modelling assume that the cellular medium is homogeneous and that one can speak of concentrations, i.e. that the number of metabolites and enzyme molecules per volume unit is high enough. This is not always the case.

2.2 Henri-Michaelis equation

In the conditions of Fig. 9, i.e. $d[ES]/dt = 0$, one can solve the system of equations (1) to (4) shown in section 2.1, which becomes a system of algebraic equations to obtain:

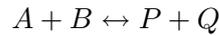
$$V = \frac{d[P]}{dt} = \frac{V_M[S]_0}{K_m + [S]_0} \quad (5)$$

with:

$$K_m = \frac{k_{-1} + k_2}{k_1} \text{ and } V_M = k_2[E]_0$$

This rate equation due to V. Henri (Henri, 1902 and 1903) on the one hand and to Briggs and Haldane (Briggs and Haldane, 1925) on the other hand is usually known as the Michaelis-Menten equation.

The approximations which are underlying equation (5) are $[E]_0 \ll [S]_0$ and necessitate to be in a time range for which $d[ES]/dt$ is close to zero (quasi steady-state hypothesis). But Cornisch-Bowden & Hofmeyr (2005) properly show that this equation does not generally apply *in vivo*, because, in this case the product concentration is not zero (the product of a reaction is the substrate of the following reaction(s)). Furthermore, in general, a reaction deals with two substrates and two products:



One is thus conducted to write a differential system of the form:

$$\begin{cases} \frac{d[EA]}{dt} = k_A[E][A] - (k_{-A} + k_{AB})[ES] \\ \frac{d[EAB]}{dt} = k_B[EA][B] - (k_{-AB} + k_2)[EAB] \\ \text{etc.} \end{cases}$$

One can, according to the same sort of approximations as above, i.e. quasi steady-state of the enzymatic species E, EA, EB, EAB etc., propose a rather general and suitable phenomenological equation (Chassagnole et al. 2001, Cornisch-Bowden & Hofmeyr, 2005):

$$v = \frac{V_{AB} \frac{[A][B]}{K_A K_B} - V_{BA} \frac{[P][Q]}{K_P K_Q}}{\left[1 + \frac{[A]}{K_A} + \frac{[P]}{K_P}\right] \left[1 + \frac{[B]}{K_B} + \frac{[Q]}{K_Q}\right]} \quad (6)$$

or

$$v = \frac{\left[1 - \frac{[P][Q]}{K_{eq}[A][B]}\right] \left[\frac{V_{AB}}{K_A K_B}\right]}{\left[1 + \frac{[A]}{K_A} + \frac{[P]}{K_P}\right] \left[1 + \frac{[B]}{K_B} + \frac{[Q]}{K_Q}\right]} \quad (7)$$

K_A, K_B, K_P and K_Q , are the so called Michaelis constants for the metabolites A, B, P and Q, analog to the parameter K_m for the substrate S in equation (5). V_{AB} corresponds to the maximal rate of the reaction in the direction from A and B towards P and Q (P corresponds to A and Q to B in the reaction) and V_{BA} corresponds to the maximal rate in the reverse reaction. These parameters are linked through the Haldane relationship:

$$K_{eq} = \frac{[P]_{eq}[Q]_{eq}}{[A]_{eq}[B]_{eq}} = \frac{V_{AB}K_PK_Q}{V_{BA}K_AK_B}$$

which is introduced in eq (7).

In equation (7) the first term in brackets in the numerator expresses the distance of the concentrations [A], [B], [P] et [Q] to the equilibrium. The second term in brackets in the numerator expresses the efficacy of the enzyme due to its concentration and to the rate constants of products production (analog to the term $k_2[E]_0$ in the Michaelis-Menten equation).

3 Metabolic networks

3.1 Why is a metabolic network more than the sum of its reactions?

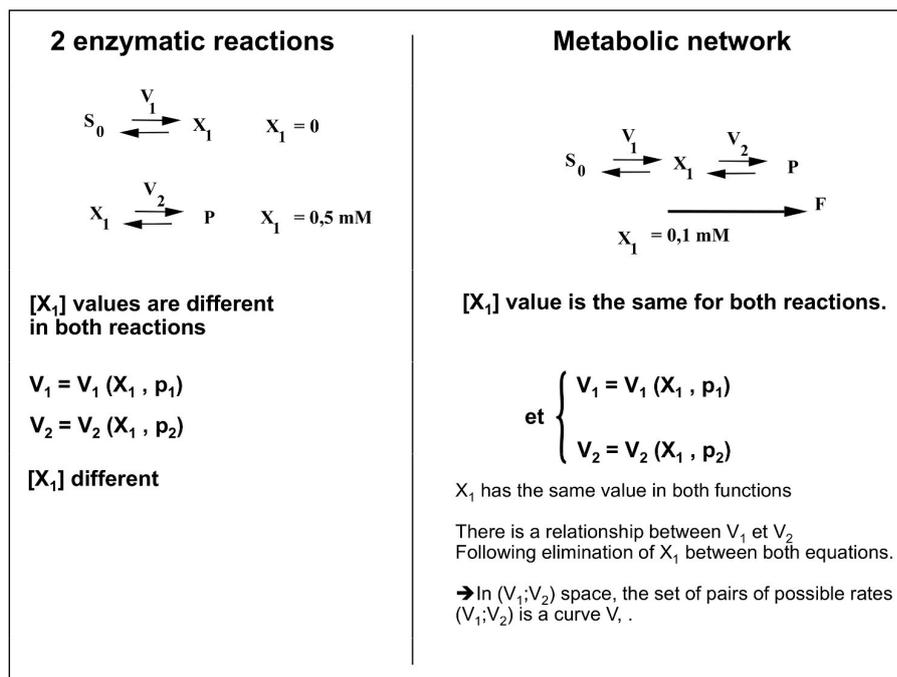


Figure 10: What is a metabolic network?

The figure 10 gives the answer. It is also illustrated by the laboratory practice. When an enzyme is studied, one use substrate and product concentrations (product concentration equal to zero) which are not the one encountered *in vivo*. Typically in the case of Fig. 10, the first reaction will be studied at zero concentration of the product X_1 , while the second reaction will be studied with variable non-zero concentrations of X_1 . On the contrary in the cell, both reactions are linked by X_1 concentration, which, in the cell, is always the same (if the intracellular medium is supposed homogeneous; see the discussion later). In a metabolic network the metabolites concentrations are determined by the network structure and the rate equations. This point is shown in the right part of Fig. 10. In a metabolic network, the intermediate metabolites are the links between the reactions. From the mathematical point of view one can derive relationships between the rates by elimination of the intermediate metabolites concentrations, so that the rate of possible rates is a one-dimensional space in the case of example of Fig. 10 (see Fig. 11).

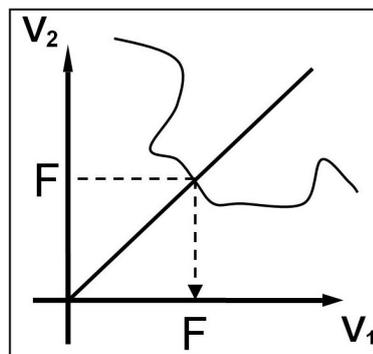


Figure 11: Phase space

3.2 Is cellular medium homogeneous?

The answer is most of the time probably not. Any view of a cell image obtained with an electronic or a photonic microscope indicates clearly an heterogeneous arrangement. Furthermore there are several examples of compartmentation and of channelling (Agius and Sherratt, 1997).

However, most of the modelling of metabolism assume (without saying it explicitly most of the time) that the cell medium is homogeneous. The reason is that the same study in an heterogeneous medium is much more difficult; a first study in an homogeneous medium can afford a first description of the main behaviour of the metabolic network. Differential equations are well adapted to concentrations supposed to be the same in all directions.

3.3 Cellular metabolism as an open system

There are two ways to approach a metabolic network as shown in Fig. 12 below:

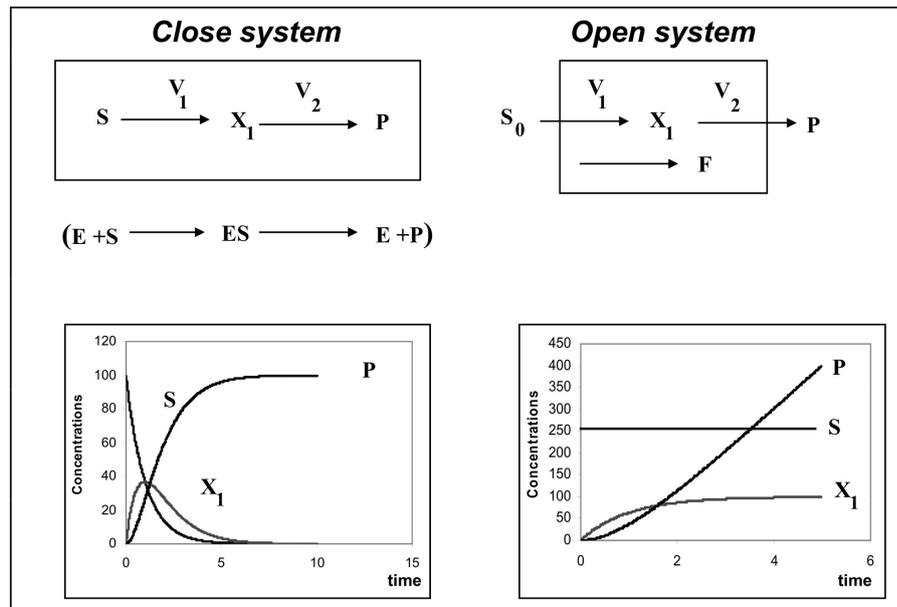


Figure 12: Two ways of considering a metabolic network

Either as a close system, which tends towards an equilibrium with usually several concentrations equal to zero due to (quasi) irreversible reactions. This situation does not correspond to the cellular situation for which the external metabolites are continuously imported and products exported. Even if one considers only part of metabolism one uses to assume that the concentration of the metabolites outside the considered part of the network (S_0 and P in Fig. 12) are constants or produced at constant rate. In these conditions the metabolite concentrations inside the network reach a steady-state with non zero constant values.

3.4 The steady-state

For the calculation of the flux in a metabolic network, one uses to consider that the intermediate metabolites have a constant value, i.e. that the production and the consumption of a given metabolite balances. This hypothesis is fairly well experimentally verified. On periods of time lasting several minutes and sometimes more, one can assume that most of the metabolite concentrations

are constant. One has however to be aware that this hypothesis does not hold when the physiological conditions are changed : rest to work transition for a muscle cell for instance. Some pathways are oscillatory : glycolysis, production of insulin by the β -cells in pancreas.

In the following we will accept the hypothesis of steady-state, knowing that it is not always satisfied.

Introduction to Stochastic Simulation Algorithm

Hédi A. Soula¹

¹ Unité 870 INSERM / INRA 1235, INSA, La Doua, 69100 Villeurbanne, France

Abstract

Most chemical systems are described using the Ordinary Differential Equations (ODE) formalism where chemical species variations are determined by the overall reactions schemes and concentrations. This formalism has been translated to biology as well for biochemical reactions - transcription, translation, binding kinetics as well.

The ODE formalism describes implicitly the deterministic limit of such a system: the final result is the average of all possible realizations of the chemical reactions. In many theoretical and computational problems, we do not wish to ignore the intrinsic fluctuations of a particular solution. Indeed, in many cases, the variability of one trajectory is needed for realism's sake. This question has led to several algorithms that describe and compute 'particular' and realistic trajectories of any chemical system.

The topic of this lecture will be focused on one of them: the Gillespie algorithm to describe chemical reactions also called the Stochastic Simulation Algorithm (SSA).

We will present a simple version of the mathematical background behind the SSA and provide some hints on the proof. We will propose several practical study cases in form of classical biochemical/physiological examples: protein production and degradation, transcription/translation, binding and enzymatic kinetics. In all cases, we will compare with the classical frame. Finally, we will discuss several refinements of the algorithms - such as tau-leap for stiff system.



From automatic and expert annotation to the reconstruction of genome-scale metabolic networks and models

François Le Fèvre¹, Damien Mornico¹, Eugeni Belda¹, David Vallenet¹,
Claudine Médigue¹

¹ Laboratoire d'Analyses Bioinformatiques pour la Génomique et le Métabolisme, UMR 8030 CEA/Genoscope - CNRS - Université d'Evry, F-91057 Evry, France

1 Abstract

Metabolism is the set of enzyme-catalyzed biochemical reactions that occur in all cells of all living organisms. These reactions, often numbering in the order of thousands, operate together as a network to progressively transform chemicals from the surrounding environment into the energy and chemical species needed by the cells. While knowledge on metabolic reactions has been slowly accumulated for a few model organisms over the past century, the advent of genome sequencing and annotation now enables to transfer part of this knowledge to new organisms and quickly draft their own metabolic networks.

In this course, we will present current resources and tools for automatic and expert annotation of genomic object involved in metabolism and for reconstruction of the metabolic network of a newly sequenced organism, with a bias toward microbial species. The focus will be put on the comprehensiveness of the reconstruction, in the aim of identifying as exhaustively as possible the metabolic capabilities of the studied organism with a focus on resources and tools available in Microme. Microme, a 15-partner project funded by the Framework 7 programme of the European Commission, is a new bioinformatics infrastructure for the curation and integration of bacterial reactions and pathways, for genome annotation and for the reconstruction of metabolic networks.

2 Introduction, overall reconstruction process

Many applications motivate the reconstruction of global metabolic networks. From a descriptive point of view, they provide an overview of all metabolic reactions that can proceed in the cell. For instance, one can then study the structure of the metabolic network, locate each reaction within the whole network, or determine which enzymes are associated with a given metabolic pathway. Studying networks across several organisms introduce new evolutionary issues: how do metabolic pathways evolve? Various types of experimental data also benefit from being integrated on a reconstructed metabolic network [17]. Useful insights often only emerge from experimental data when they are interpreted in the context of the whole metabolism. At a larger scale,

studying the global metabolic network can help in relating metabolic functions found in the genome with observed cell phenotypes, such as the ability for the cell to grow on given chemical environments, or its tendency to excrete particular by-products. Comprehensiveness of the reconstruction is a crucial requirement in this type of application, as global phenotypes result from the operation of the whole set of reactions. The size and complexity of metabolic networks, however, often hinder direct deductions of their system-level properties, e.g. their dynamic behaviors or their global phenotypes. Such deductions are therefore mostly performed using mathematical or computational models of metabolism [7, 12]. An accurate knowledge of metabolic networks is here again key in building faithful models.

In this brief article, we will cover the most useful and fundamental parts of metabolic network reconstruction. Many more details can be found in other comprehensive reviews, which focus for instance on practical aspects [33], conversion to metabolic models [9], or reconstruction of biological networks of broader types [11].

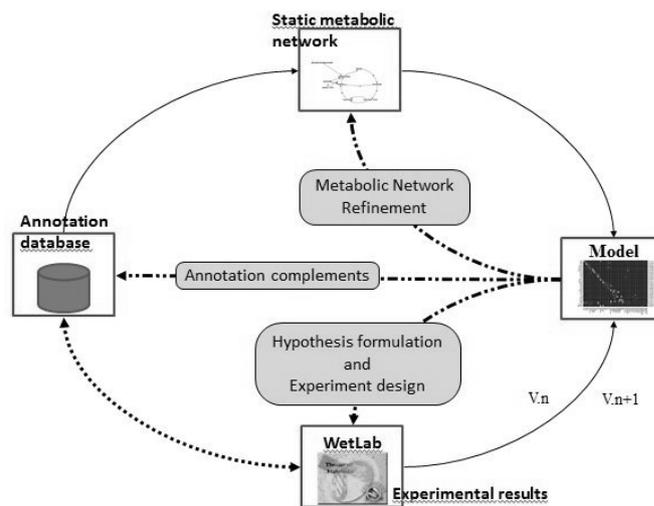


Figure 1: Iterative reconstruction of metabolic networks and models. Data extracted from the annotation database is used to build a static metabolic network, which is then converted to a metabolic model. Analysis of the model in regards of experimental results allows refining the metabolic network, completing genome annotation and formulating hypotheses that could be tested at the workbench.

Network reconstruction generally proceeds in three steps which gradually build and refine the metabolic network. First, information about metabolic

reactions is extracted from the genome annotation. Depending on the amount of biochemical knowledge available for the organism, this step can account for a large fraction of the reconstructed network, especially for barely studied organisms. Then, the draft network is completed with organism-specific reactions and metabolic pathways which could not be extracted from the genome annotations. A significant amount of metabolic information is indeed present in the literature and needs to be incorporated in the network. Finally, the completeness of the network is evaluated and missing reactions inferred. These three steps are described in the following sections.

From this metabolic network, scientists can design a mathematical model that can be iteratively refined by comparing its predictions to experimental data, thereby increasing biological knowledge on the organism.

Several platforms allow confronting experimental results against *in silico* experiments. For instance, CycSim [13] supports the design of knockout experiments: simulation of growth phenotypes of single or multiple gene deletion mutants on specified media and comparison of these predictions with experimental phenotypes.

The screenshot shows the CycSim web application interface. The main content area displays the results of a Flux balance Analysis (FBA) for a specific phenotype (FBAP). The analysis ID is Sec13666-93fd-4814-994e-161ec342d63, and the date is Thu Mar 17 11:33:56 GMT+100 2011. The results are presented as a comparison of *in silico* predictions with experimental phenotypes.

		in vivo			Accuracy score	Consistent predictions
		growth	no growth	no experiment		
in silico	growth	55	0	8		
	no growth	0	3	24		

Below the table, there is a section for 'Details results' which includes a legend and a table of growth phenotypes for different strains across various media. The table has columns for Strain and media: abaylyl succinate, abaylyl acetate, abaylyl butanedioyl, abaylyl asparagine, abaylyl d. glucarate, abaylyl glucose, abaylyl lactate, abaylyl quinolate, and abaylyl succinate ura. The rows list various strains including Wild type, ACIAD0019, ACIAD0015, ACIAD0022, ACIAD0023, ACIAD0024, ACIAD0028, ACIAD0031, ACIAD0032, and ACIAD0034. Each cell in the table contains a small icon representing the growth phenotype (e.g., 'in silico', 'in vivo', 'no growth').

Figure 2: CycSim, a web application dedicated to *in silico* experiments with genome-scale metabolic models coupled to the exploration of knowledge from BioCyc and KEGG. <http://www.genoscope.cns.fr/cycsim>

3 From genome annotation to metabolic reactions

The first step of metabolic network reconstruction is mostly gene-centric. Leveraging on the ability to predict functions for a large fraction of genes, one can derive metabolic reactions from genes with enzymatic functions.

Thanks to dramatic improvements in genome sequencing technology, the number of available genome sequences is quickly increasing (see <http://genomeonline.org> for a review of past and ongoing sequencing projects). Since genome and protein databases GenBank, CMR, IMG, MicroScope, SEED, or UniProt for instance (see Table 1) collect genome information for a wide fraction of sequenced organisms, genome annotation is readily available for many organisms. If not, the raw genome sequence needs to be processed and annotated [23] using integrated annotation platforms such as Microscope [34], IMG [21] or RAST [3].

Focus on Microscope

The MicroScope platform, which provides support for the (re-)annotation of microbial genomes and their comparative analysis (<http://www.genoscope.cns.fr/agc/microscope>), currently hosts about 100 different projects covering roughly 1020 bacterial genomes.

Following a fully automated annotation and analysis process, data is made available through advanced graphical Web interfaces for manual refinement and for the exploration of comparative genomics analysis results. To these ends, various tools are made available to the platform user, such as dynamic procedures allowing one to analyse genomic islands, synteny conservation, "pan" and "core" genomes within microbial species, gene and metabolic phyloprofiles, etc [34]. The expert annotations gathered in MicroScope (an average of 50 000 a year, created by more than 1000 curators having a personal account on the platform) continue to improve the quality of bacterial genome annotations.

Following the main research activities at the Genoscope (i.e, in-depth re-examination of the metabolic functions of a bacterial organism and discovery of novel enzymatic activities), and in the context of an European collaborative project named Microme, recent evolutions of the MicroScope platform have focused on the prediction and curation of prokaryote metabolic pathways. New strategies for functional annotation of prokaryote genomes have been developed, especially a genomic and metabolic context-based inference method that can propose candidate genes for sequence-orphan enzymatic activities [30]. Results of these strategies are now made available through new MicroScope web interfaces. Procedures and graphical interfaces dedicated to the curation of enzymatic activities, of Gene- Protein-Reaction associations and of predicted metabolic pathways have also been developed. Expert annotations of this metabolic data are directly imported into the Microme repository to form

a core of high-confidence data used in the projection strategy (i.e, to predict new GPRs associations and metabolic pathways in microbial genomes).

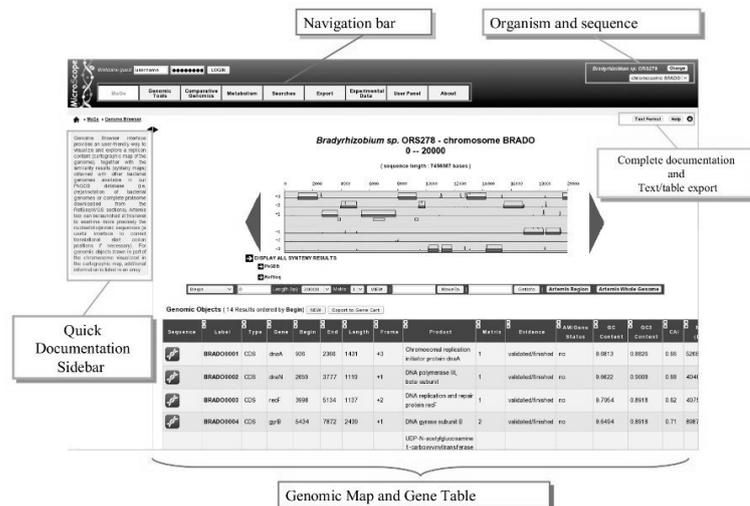


Figure 3: Overview of the Microscope platform:
<https://www.genoscope.cns.fr/agc/microscope>.

The easiest method to translate gene functional annotations into metabolic reactions relies on Enzyme Commission (EC) numbers which classifies most enzyme activities [4]. Virtually all genome annotation databases associate EC numbers with gene functions and all metabolic databases relate existing EC numbers with the reactions they store (see Table 1). In this process, EC numbers act as direct links between genes and reactions.

Yet, EC numbers present several drawbacks which require using further methods. First, a small set of recently identified enzymatic activities are lacking EC numbers, a consequence of the slow curation process of the EC classification. Second, full EC numbers are not always assigned to gene functions, bringing uncertainty about which activities are really predicted. And third, the specificity of enzyme activities is not explicitly defined for a few EC numbers. For these reasons, additional tools have been developed. They either directly translate textual annotations into metabolic reactions [18] or perform metabolic annotations de novo from genome sequence [32, 25, 2]. These latter tools do not benefit, however, from all improvements made to "classical" annotation platforms, a limitation which may restrict their efficiency in the future.

The BioCyc collection of metabolic databases rely on a such a software, Pathway Tools (see Table 2 and ref. [6, 18]), to build metabolic networks. Pathway Tools actually contains a module called PathoLogic which combines

Name	Link	Comment
MicroScope	http://www.genoscope.cns.fr/agc/mage	Microbial genome annotation database
IMG	http://img.jgi.doe.gov	Microbial genome annotation database
CMR	http://cmr.jcvi.org	Microbial genome annotation database
SEED	http://seed-viewer.theseed.org	Genome annotation database
UniProt	http://www.uniprot.org	Comprehensive database of curated and predicted protein functions
BRENDA	http://www.brenda-enzymes.info	Database of curated enzyme information
TransportDB	http://www.membrane-transport.org	Database of predicted membrane transport proteins
Enzyme	http://expasy.org/enzyme	Database of enzymatic activities

Table 1: Resources and software tools for microbial metabolic network reconstruction: **Genome annotation/enzyme databases**

EC numbers with textual annotations and knowledge of metabolic pathways to infer metabolic reactions from genome annotation.

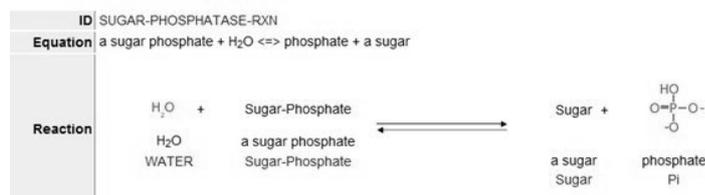


Figure 4: For instance, the reaction with EC#3.1.3.23 is not fully defined in MetaCyc: the main substract/product is labeled as Sugar-phosphate. Integrating this reaction needs to resolve which sugar-phosphate is the enzyme substrate.

In order to solve those problems, reactions are now resolved not necessarily through their EC numbers but using of reactions database such as Rhea [1]. Rhea is a manually annotated database of chemical reactions; each reaction is defined by its chemical equation: list of compounds and their associated stoichiometry (and eventually its compartment origin in/out for transport re-

Name	Link	Comment
ChEBI	http://www.ebi.ac.uk/chebi	Database of chemicals of biological interest
Rhea	http://www.ebi.ac.uk/rhea	Manually annotated database of chemical reactions
BioCyc	http://www.biocyc.org	Collection of organism-specific metabolic databases
MetaCyc	http://www.metacyc.org	Multiorganism, curated metabolic pathway database based on BioCyc
MicroCyc	http://www.genoscope.cns.fr/agc/microcyc	Collection of organism-specific metabolic databases based on BioCyc, for all organisms included in MicroScope. (more than 500 organisms available)
Reactome	http://www.reactome.org	Curated database of metabolic pathways.
Microme	http://www.microme.eu	A specific instance of Reactome dedicated to microbial world and a web portal to tools and resources of all Microme partners.
KEGG	http://www.genome.jp/kegg	Comprehensive database of compounds, reactions, metabolic pathways, and genes/proteins
UniPathway	http://www.grenoble.prabi.fr/obiwarehouse/unipathway	Database of curated metabolic pathways, linked to UniProt proteins
UM-BBD	http://umbbd.msi.umn.edu	Database of microbial biodegradation pathways

Table 2: Resources and software tools for microbial metabolic network reconstruction: **Metabolic databases**

actions) In Rhea, each compound is mapped to a compound of ChEBI [22], a database of chemicals of biological interest, ensuring that the reaction is balanced, unique and mapped to the real catalytic activity. Furthermore, each gene or protein functional annotation must be directly linked to their corresponding reaction(s). Several initiatives now focus on directly specifying metabolic reactions during the genome annotation step. The SEED system already associates genes to their metabolic roles [8], and UniProt is formalizing the description of metabolic functions with the UniPathway resource [24] (see Table 2).

Name	Link	Comment
BIGG	http://bigg.ucsd.edu	Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions
The Model SEED	http://www.theseed.org/models	A resource for the generation, optimization, curation, and analysis of genome-scale metabolic models.
BioModels	http://www.ebi.ac.uk/biomodels-main	A repository of peer-reviewed, published, computational models
CycSim	http://www.genoscope.cns.fr/cycsim	An online tool for exploring and experimenting with genome-scale metabolic models.

Table 3: Resources and software tools for microbial metabolic network reconstruction: **Models databases**

Thanks to such initiatives, the process of translating gene functions into metabolic reactions is likely to become much easier in the future. A new module in Microscope platform has been developed in order to directly link the MetaCyc reaction to gene without using the EC number as a key. This feature allows flagging the relationship between gene and reactions with the following evidence tags:

- "validated" : reaction has been manually linked to this gene by users,
- "annotated" : reaction has been linked to homologous gene and transferred here from a close genome,
- "predicted" : reaction has been linked to this gene by the pathway-tools algorithm.

A critical point in metabolic model reconstruction is the ability to distinguish isozymes from enzymatic complexes because the impact of a gene deletion will not have the same consequences. If the deleted gene belongs to a complex, the enzymatic activity will be lost, whereas in the case of isozymes, the enzymatic activity will remain present in the cell due to the existence of a "redundant" enzyme. Protein complexes could be built manually from literature or inferred by homology. One strategy is to use a pivot organism: for each protein complex experimentally demonstrated in the pivot organism, a similar complex is inferred in the studied organism if, for each subunit of the pivot complex, a candidate polypeptide can be detected by homology (e.g. Blast alignments with */bidirectional/* best hits) [36].

Name	Link	Comment
Pathway Tools /Pathway Hole Filler	http://bioinformatics.ai.sri.com/ptools/	Automated reconstruction of metabolic pathways from genome annotation and MetaCyc pathway database. Pathway Hole Filler: identification of candidate genes for orphan metabolic activities
OptFlux	http://www.optflux.org	An open-source software platform for in silico metabolic engineering
YanaSquare	http://yana.bioapps.biozentrum.uni-wuerzburg.de	Software helping in metabolic model reconstruction, linked to KEGG data
GEMSiR	http://sb.nhri.org.tw/GEMSiR	A software platform for genome-scale Metabolic models Simulation, Reconstruction and Visualization.
Cobra tool box	http://opencobra.sourceforge.net	Quantitative prediction of cellular metabolism with constraint-based models
SurreyFBA	http://sysbio3.fhms.surrey.ac.uk/SurreyFBA.zip	A command line tool and graphics user interface for constraint-based modeling of genome-scale metabolic reaction networks

Table 4: Resources and software tools for microbial metabolic network reconstruction: **Software tools for metabolic network reconstruction**

Finally, transport reactions are key reactions to be integrated to the metabolic network. They allow defining the interface between the cell and the environment. TransportDB provides tools to annotate a genome by finding the transporters.

4 Adding organism-specific metabolic pathways

Using an organism-specific metabolic pathways resource is a key point in delivering a well annotated network. Some databases enable to annotate the existence of each pathway: this is the first step to define an organism-specific

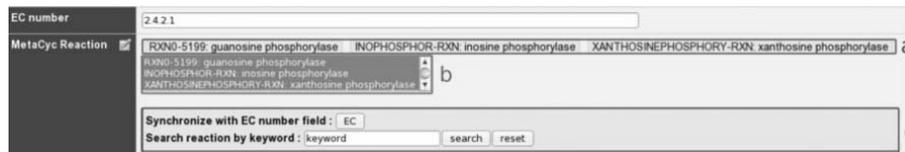


Figure 5: This field allows user to link one or more metabolic reactions from MetaCyc (BioCyc) to the current edited gene. A - Reactions presented at the top of the field have been manually curated by an annotator. B - A multiple selection list gives quick access to all predicted (unselected) or curated (selected) reactions linked to this gene. C - A search box allows one to quickly access MetaCyc reactions corresponding to either EC numbers from previous EC number field or a given keyword.

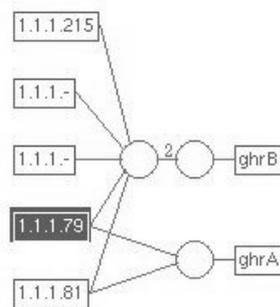


Figure 6: Protein complexes and isozymes representation in EcoCyc. At the right end, the genes (purple boxes) are translated into polypeptide (first orange circle). These polypeptides are able to make protein complexes that can aggregate to form the final catalyst.

metabolic pathways resource. Microscope platform provides the Pathway Curation tool which presents a list of predicted MicroCyc pathways in a given organism, coming from Pathway-tools software results, for which status can be curated by the annotator. Each pathway could have the following evidence status:

- predicted: Predicted by the BioCyc pathologic algorithm (default one),
- validated: Curated as a functional pathway (all the reactions of the pathway are supposed to exist in the organism),
- variant needed: The predicted pathway is not completely correct for the organism (i.e. some reactions may not be present in the organism but no better pathway definition exists in MetaCyc). Thus, a new pathway variant definition is needed,
- unknown: Not enough evidence to declare the pathway as functional (i.e. validated status),

- non-functional: The pathway has been lost in the organism and is no more functional (i.e. due to gene loss or pseudogenisation events),
- deleted: Curated as a false positive prediction.

Pathway Curation
Acinetobacter baylyi ADP1

① Current pathway curation status: (248 predicted)

- ✓ 4 validated
- ⊗ 0 variant needed
- ? 2 unknown
- ⊗ 0 non functional
- ✗ 2 deleted

② Search new pathway by keyword:

Display pathway hierarchy: ON OFF

Showing 11 to 256 of 256 results Show All Results Search: Copy CSV Print

?	⊗	✗	Curation [256]	Pathway	Completion	Reactions nb
?	⊗	✗	[validated]	1CMET2-PWY : formylTHF biosynthesis I	0.91	11
?	⊗	✗	[validated]	ACETOACETATE-DEG-PWY : acetoacetate degradation (to acetyl CoA)	1	2
?	⊗	✗	[validated]	ALADEG-PWY : alanine degradation I	1	2
?	⊗	✗	[validated]	ALANINE-SYN2-PWY : alanine biosynthesis II	1	1
⊗	?	✗	[unknown]	ALANINE-VALINESYN-PWY : alanine biosynthesis I	0.67	3
✓	?	⊗	[predicted]	ALKANEMONOX-PWY : two-component alkanesulfonate monooxygenase	1	2
✓	?	⊗	[predicted]	ANARESP1-PWY : respiration (anaerobic)	0.69	13
✓	?	⊗	[predicted]	ARGSYN-PWY : arginine biosynthesis I	1	9

③

Figure 7: Main overview for Pathway Curation tool in Microscope.

Nevertheless, not all information on an organism's metabolic network can be found in its genome annotation. Previous experimental studies of its metabolism may well have identified metabolic pathways for which the corresponding genes remain unknown (e.g. an alanine biosynthesis pathway in *E. coli*, see <http://biocyc.org/ECOLI/NEW-IMAGE?object=PWY0-41>) or whose functions were not correctly propagated into their annotations. Also, not all metabolic pathways are represented in global metabolic databases, which are being built by reviewing the past literature, a still ongoing process.

For these reasons, the draft metabolic networks generated from genome annotation need to be curated and completed with the maximum amount of additional organism-specific biochemical information. Such information is mainly found in the literature, but also in some specialized biochemical databases. For instance, BRENDA extracts information on enzymes directly from the literature [19] and UM-BBD stores useful information on numerous microbial biocatalytic activities and biodegradation pathways [10] (see Table 2).

When including new reactions into the draft network, one must remain consistent with the reaction metabolite identifiers. Given the size of the meta-

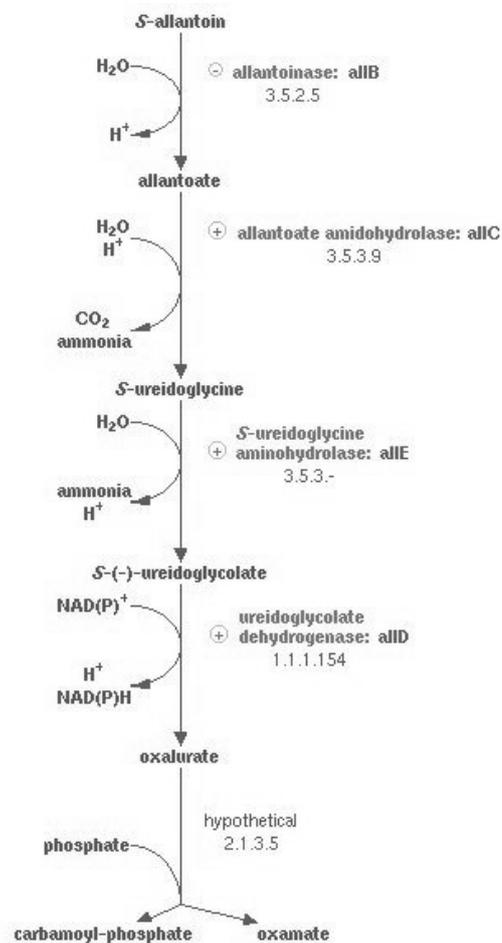


Figure 8: In EcoCyc, the allantoin degradation IV (anaerobic) presents a reaction (EC# 2.1.3.5) that do not have any associated enzyme protein.

biological networks, this technical issue is critical as designating a given metabolite with distinct identifiers would prevent from finding connections between reactions that share the metabolite, and therefore would hinder the identification of pathways. Similarly, mathematical models of metabolism crucially depend on a consistent naming of metabolites. This issue is better solved by systematically referring to a unique database of metabolites, be it KEGG, MetaCyc, ChEBI or PubChem (see Table 2). These databases usually cross-reference their entries.

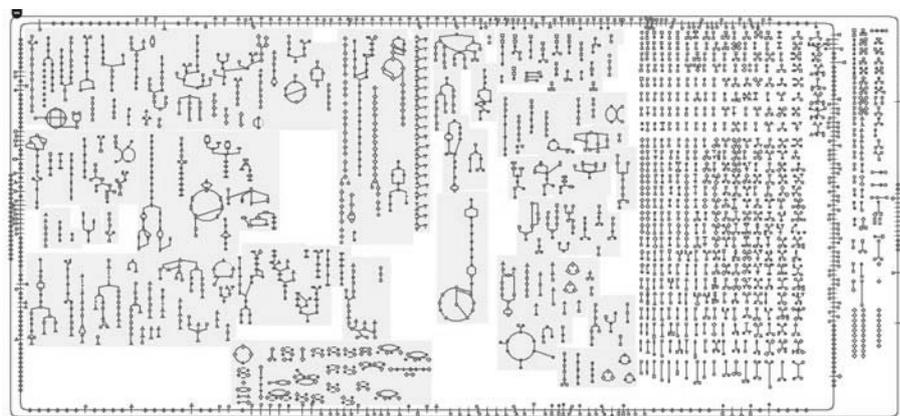


Figure 9: Interactive overview of Escherichia coli K12 metabolic network. Screenshots from EcoCyc [19] (<http://ecocyc.org/overviewsWeb/celOv.shtml>)

As regards software tools enabling curation of genome-scale metabolic networks, very few of them are currently available. Again, Pathway Tools provides an interface to curate, extend and visualize a locally reconstructed metabolic network [12] (see Figure 1). YanaSquare imports metabolic networks automatically generated in KEGG and allows curating it using a spreadsheet-like interface [29]. OptFlux [26] allows importing metabolic models in System Biology Markup Language format [16]. General metabolic databases such as KEGG or SEED provide tools to visualize their reconstructed network, but do not allow curation.

Initiatives are being taken to develop further network curation tools (see for instance the EU-funded project Microme, <http://www.microme.eu>) and more convenient solutions will be hopefully available in a near future.

5 Identifying and filling gaps in metabolic networks

Once a global metabolic network is obtained, its completeness can be assessed by examining it globally or at the pathway level. Incomplete metabolic pathways, reactions disconnected from the rest of the network, or metabolites that are only produced and never consumed (so-called dead-end metabolites) are actually as many clues that the reconstructed network include gaps that need to be filled for the network to represent a consistent functioning set of metabolic pathways.

A few methods have been designed to identify and fill such metabolic gaps. Pathway Tools makes use of predefined metabolic pathways: each pathway

that is almost complete in the metabolic network is considered to be entirely present. The few missing reactions are therefore added to the network, thereby filling the gaps in the pathway [18]. Other more elaborate methods make use of graph- based or constraint-based mathematical representation of the network (see ref. [9] for a more thorough review). The GapFill method for instance attempts to fill metabolic gaps by adding minimal sets of reactions from a reference database that remove dead-end metabolites [27].

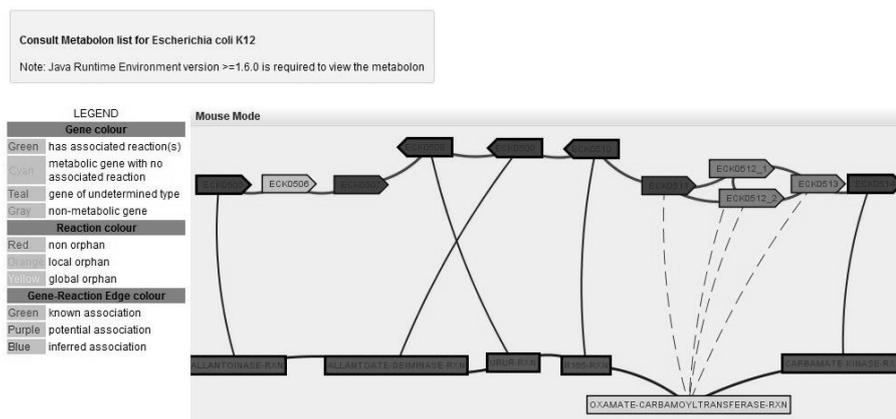


Figure 10: Screenshot of the CanOE Web interface. This metabolon is composed of 6 reactions (red and yellow) covering the complete pathway for the anaerobic degradation of allantoin, in which one reaction is orphans in *E. coli* (yellow): oxamate carbamoyltransferase (OXTCase). The missing activity in *E. coli* K-12 (the OXTCase) has yet to be associated to any genes in any organism and is thus a global orphan activity, despite its presence having been biochemically demonstrated in *Streptococcus allantoicus*, and even reported in *E. coli*. The CanOE metabolon bearing this reaction contained 5 gap genes (ECK0506-507 and ECK0511 to 0513 gray, blue and green) that could serve as candidate genes. The link between a reaction and a gene could be experimentally demonstrated (green) or proposed by CanOE in purple.

Reactions that are inferred by pathway projection process constitute orphan activities, for which no related gene is known [20]. Additional methods have therefore been introduced to identify genes for orphan activities. While bioinformatic tools may help in identifying suitable candidate genes (e.g. Pathway Hole Filler [15], implemented in Pathway Tools), promising experimental setups are being developed to systematically screen gene enzymatic functions [9, 5]. CanOE (Candidate genes for Orphan Enzymes) is a four-step bioinformatics strategy that proposes ranked candidate genes for sequence-orphan enzymatic activities (or orphan enzymes for short) [30]. The first

step locates "genomic metabolons", i.e. groups of co-localized genes coding proteins catalyzing reactions linked by shared metabolites, in one genome at a time. These metabolons can be particularly helpful for aiding bioanalysts visualize relevant metabolic data. In the second step, they are used to generate candidate associations between un-annotated genes and gene-less reactions. The third step integrates these gene-reaction associations over several genomes using gene families, and summarizes the strength of family-reaction associations by several scores. In the final step, these scores are used to rank members of gene families which are proposed for metabolic reactions. These associations are of particular interest when the metabolic reaction is a sequence-orphan enzymatic activity. The CanOE strategy has been integrated in the MicroScope platform and is available at this URL:

<http://www.genoscope.cns.fr/agc/microscope/metabolism/canoe.php>.

6 *Toward mathematic models of genome-scale metabolism*

The end result of the network reconstruction process should be a comprehensive, consistent, and organized listing of metabolic reactions. It should provide an accurate overview of the organism's metabolism, especially when linked to convenient visualization software.

Reasoning on the metabolic network, however, requires going a step further as its size and complexity prevent from performing simple predictions from it. Several mathematical modeling frameworks have therefore been proposed to formalize metabolic networks, perform systems-level predictions and integrate various types of experimental data. Reviewing these frameworks would be out of the scope of this short article, and has already been extensively covered elsewhere [17, 31].

Deriving a metabolic model from a reconstructed network usually necessitates collecting additional data on reactions and metabolites and further checking their consistency. For instance, constraint-based models, which are widely used for genome-scale modeling, critically rely on stoichiometric coefficients and reaction reversibilities. Since subsequent model predictions directly depend on these data as well as on the correctness of the metabolic network, specific care should be taken when creating models from metabolic networks. Here again, the model reconstruction issue has been largely covered in a few reviews [9, 11].

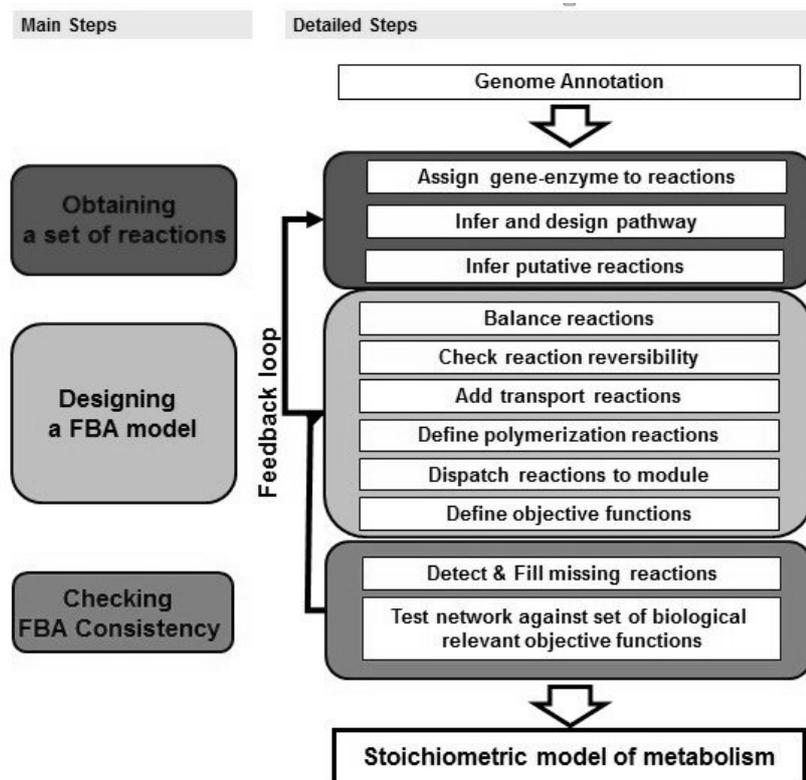


Figure 11: Standard protocol to design network and model from genome annotation.

References

- [1] Alcántara, Rafael, Kristian B Axelsen, Anne Morgat, Eugeni Belda, Elisabeth Coudert, Alan Bridge, Hong Cao, et al. 2012. "Rhea—a manually curated resource of biochemical reactions". *Nucleic Acids Research* 40 (Database issue): D754-760. doi:10.1093/nar/gkr1126.
- [2] Arakawa, Kazuharu, Yohei Yamada, Kosaku Shinoda, Yoichi Nakayama, et Masaru Tomita. 2006. "GEM System: automatic prototyping of cell-wide metabolic pathway models from genomes". *BMC Bioinformatics* 7: 168. doi:10.1186/1471-2105-7-168.
- [3] Aziz, Ramy K, Daniela Bartels, Aaron A Best, Matthew DeJongh, Terrence Disz, Robert A Edwards, Kevin Formsma, et al. 2008. "The RAST Server: rapid annotations using subsystems technology". *BMC Genomics* 9: 75. doi:10.1186/1471-2164-9-75.

- [4] Bairoch, Amos. 2000. "The ENZYME database in 2000". *Nucleic Acids Research* 28 (1): 304-305.
- [5] Beloqui, Ana, María-Eugenia Guazzaroni, Florencio Pazos, José M Vieites, Marta Godoy, Olga V Golyshina, Tatyana N Chernikova, et al. 2009. "Reactome array: forging a link between metabolome and genome". *Science (New York, N.Y.)* 326 (5950): 252-257. doi:10.1126/science.1174094.
- [6] Caspi, Ron, Tomer Altman, Kate Dreher, Carol A Fulcher, Pallavi Subhraveti, Ingrid M Keseler, Anamika Kothari, et al. 2012. "The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases". *Nucleic Acids Research* 40 (Database issue): D742-753. doi:10.1093/nar/gkr1014.
- [7] Curran, Kathleen A, Nathan C Crook, et Hal S Alper. 2012. "Using flux balance analysis to guide microbial metabolic engineering". *Methods in Molecular Biology (Clifton, N.J.)* 834: 197-216. doi:10.1007/978-1-61779-483-4_13.
- [8] Disz, Terry, Sajia Akhter, Daniel Cuevas, Robert Olson, Ross Overbeek, Veronika Vonstein, Rick Stevens, et Robert A Edwards. 2010. "Accessing the SEED genome databases via Web services API: tools for programmers". *BMC Bioinformatics* 11: 319. doi:10.1186/1471-2105-11-319.
- [9] Durot, Maxime, Pierre-Yves Bourguignon, et Vincent Schachter. 2009. "Genome-scale models of bacterial metabolism: reconstruction and applications". *FEMS Microbiology Reviews* 33 (1): 164-190. doi:10.1111/j.1574-6976.2008.00146.x.
- [10] Ellis, Lynda B M, Dave Roe, et Lawrence P Wackett. 2006. "The University of Minnesota Biocatalysis/Biodegradation Database: the first decade". *Nucleic Acids Research* 34 (Database issue): D517-521. doi:10.1093/nar/gkj076.
- [11] Feist, Adam M, Markus J Herrgård, Ines Thiele, Jennie L Reed, et Bernhard Ø Palsson. 2009. "Reconstruction of biochemical networks in microorganisms". *Nature Reviews. Microbiology* 7 (2): 129-143. doi:10.1038/nrmicro1949.
- [12] Feist, Adam M, et Bernhard Ø Palsson. 2008. "The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*". *Nature Biotechnology* 26 (6): 659-667. doi:10.1038/nbt1401.

- [13] Le Fèvre, F, S Smidtas, C Combe, M Durot, Florence d'Alché-Buc, et V Schachter. 2009. "CycSim—an online tool for exploring and experimenting with genome-scale metabolic models". *Bioinformatics (Oxford, England)* 25 (15): 1987-1988. doi:10.1093/bioinformatics/btp268.
- [14] Gevorgyan, Albert, Michael E Bushell, Claudio Avignone-Rossa, et Andrzej M Kierzek. 2011. "SurreyFBA: A Command Line Tool and Graphics User Interface for Constraint-Based Modeling of Genome-Scale Metabolic Reaction Networks". *Bioinformatics* 27 (3): 433-434. doi:10.1093/bioinformatics/btq679.
- [15] Green, Michelle L, et Peter D Karp. 2004. "A Bayesian method for identifying missing enzymes in predicted metabolic pathway databases". *BMC Bioinformatics* 5: 76. doi:10.1186/1471-2105-5-76.
- [16] Hucka, M, A Finney, H M Sauro, H Bolouri, J C Doyle, H Kitano, A P Arkin, et al. 2003. "The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models". *Bioinformatics (Oxford, England)* 19 (4): 524-531.
- [17] Joyce, Andrew R, et Bernhard Ø Palsson. 2006a. "The model organism as a system: integrating "omics" data sets." *Nat Rev Mol Cell Biol* 7 (3): 198-210. doi:10.1038/nrm1857.
- [18] Karp, Peter D, Suzanne M Paley, Markus Krummenacker, Mario Landresse, Joseph M Dale, Thomas J Lee, Pallavi Kaipa, et al. 2010. "Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology". *Briefings in Bioinformatics* 11 (1): 40-79. doi:10.1093/bib/bbp043.
- [19] Lang, Maren, Michael Stelzer, et Dietmar Schomburg. 2011. "BKM-react, an integrated biochemical reaction database". *BMC Biochemistry* 12: 42. doi:10.1186/1471-2091-12-42.
- [20] Lespinet, Olivier, et Bernard Labedan. 2005. "Orphan enzymes?" *Science (New York, N.Y.)* 307 (5706): 42. doi:10.1126/science.307.5706.42a.
- [21] Markowitz, Victor M, Konstantinos Mavromatis, Natalia N Ivanova, I-Min A Chen, Ken Chu, et Nikos C Kyrpides. 2009. "IMG ER: a system for microbial genome annotation expert review and curation". *Bioinformatics (Oxford, England)* 25 (17): 2271-2278. doi:10.1093/bioinformatics/btp393.
- [22] de Matos, Paula, Nico Adams, Janna Hastings, Pablo Moreno, et Christoph Steinbeck. 2012. "A database for chemical proteomics:

- ChEBI". *Methods in Molecular Biology (Clifton, N.J.)* 803: 273-296. doi:10.1007/978-1-61779-364-6_19.
- [23] Médigue, Claudine, et Ivan Moszer. 2007. "Annotation, comparison and databases for hundreds of bacterial genomes". *Research in Microbiology* 158 (10): 724-736. doi:10.1016/j.resmic.2007.09.009.
- [24] Morgat, Anne, Eric Coissac, Elisabeth Coudert, Kristian B. Axelsen, Guillaume Keller, Amos Bairoch, Alan Bridge, Lydie Bougueleret, Ioannis Xenarios, et Alain Viari. 2012. "UniPathway: a resource for the exploration and annotation of metabolic pathways". *Nucleic Acids Research* 40 (D1): D761-D769. doi:10.1093/nar/gkr1023.
- [25] Notebaart, Richard A, Frank H J van Enckevort, Christof Francke, Roland J Siezen, et Bas Teusink. 2006. "Accelerating the reconstruction of genome-scale metabolic networks". *BMC Bioinformatics* 7: 296. doi:10.1186/1471-2105-7-296.
- [26] Rocha, Isabel, Paulo Maia, Pedro Evangelista, Paulo Vilaça, Simão Soares, José P Pinto, Jens Nielsen, Kiran R Patil, Eugénio C Ferreira, et Miguel Rocha. 2010. "OptFlux: an open-source software platform for in silico metabolic engineering". *BMC Systems Biology* 4: 45. doi:10.1186/1752-0509-4-45.
- [27] Satish Kumar, Vinay, Madhukar S Dasika, et Costas D Maranas. 2007. "Optimization based automated curation of metabolic reconstructions". *BMC Bioinformatics* 8: 212. doi:10.1186/1471-2105-8-212.
- [28] Schellenberger, Jan, Richard Que, Ronan M T Fleming, Ines Thiele, Jeffrey D Orth, Adam M Feist, Daniel C Zielinski, et al. 2011. "Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0". *Nature Protocols* 6 (9):1290-1307. doi:10.1038/nprot.2011.308.
- [29] Schwarz, Roland, Chunguang Liang, Christoph Kaleta, Mark Kühnel, Eik Hoffmann, Sergei Kuznetsov, Michael Hecker, Gareth Griffiths, Stefan Schuster, et Thomas Dandekar. 2007. "Integrated network reconstruction, visualization and analysis using YANASquare". *BMC Bioinformatics* 8: 313. doi:10.1186/1471-2105-8-313.
- [30] Smith, Alexander, Claudine Médigue, Eugeni Belda, Alain Viari, et David Vallenet. 2012. "The CanOE strategy: integrating genomic and metabolic contexts across multiple prokaryote genomes to find candidate genes for orphan enzymes". *PLoS Computational Biology*.

- [31] Stelling, Jörg. 2004. “Mathematical models in microbial systems biology”. *Current Opinion in Microbiology* 7 (5): 513-518. doi:10.1016/j.mib.2004.08.004.
- [32] Sun, Jibin, et An-Ping Zeng. 2004. “IdentiCS—identification of coding sequence and in silico reconstruction of the metabolic network directly from unannotated low-coverage bacterial genome sequence”. *BMC Bioinformatics* 5: 112. doi:10.1186/1471-2105-5-112.
- [33] Thiele, Ines, et Bernhard Ø Palsson. 2010. “A protocol for generating a high-quality genome-scale metabolic reconstruction”. *Nature Protocols* 5 (1): 93-121. doi:10.1038/nprot.2009.203.
- [34] Vallenet, D, S Engelen, D Mornico, S Cruveiller, L Fleury, A Lajus, Z Rouy, et al. 2009. “MicroScope: a platform for microbial genome annotation and comparative genomics”. *Database: The Journal of Biological Databases and Curation* 2009: bap021. doi:10.1093/database/bap021.
- [35] Vallenet, David, Stefan Engelen, Damien Mornico, Stéphane Cruveiller, Ludovic Fleury, Aurélie Lajus, Zoé Rouy, et al. 2009. “MicroScope: a platform for microbial genome annotation and comparative genomics”. *Database: The Journal of Biological Databases and Curation* 2009: bap021. doi:10.1093/database/bap021.
- [36] Vieira, Gilles, Victor Sabarly, Pierre-Yves Bourguignon, Maxime Durot, François Le Fèvre, Damien Mornico, David Vallenet, et al. 2011. “Core and panmetabolism in *Escherichia coli*”. *Journal of Bacteriology* 193 (6): 1461-1472. doi:10.1128/JB.01192-10.

GUBS : A rule based behavioral language for synthetic biology

Adrien Basso-Blandin¹, Franck Delaplace¹

¹Laboratoire IBISC - Université d'Evry, F-91034 Evry, France

Abstract

GUBS (Genomic Unified Behavior Specification Language) is a domain specific language dedicated to synthetic biology. The objective of the GUBS project is to define a framework to automatically synthesize biological functions from a specification written in GUBS [1].

The field of synthetic biology is currently looking forward principles and tools to ease the design of biological synthetic functions. In this endeavor, computer-aided-design (CAD) environments play a central role by providing the required features to engineer systems: specification, analysis and tuning [2, 6, 8, 4, 5]. Languages and CAD environments for synthetic biology are announced as one of the key milestone for the second wave of synthetic biology to overcome the complexity of large synthetic system design [7]. Nonetheless, this domain is still in its infancy and transforming this vision into a practical reality remains a daunting challenge.

Currently, programming language for synthetic biology mainly address the design of the structure, namely the component/biobrick assembly. Hence, a program relates to a DNA sequence description (e.g., GENOCAD [4], GEC [6]). The structural description is an indispensable step in the synthesis chain and provides the exact description of devices. However, the level of abstraction seems to be too restrictive for tackling *in-silico* genome engineering [3] because the required size of program is subject to error and likely un-computable. Hence, we propose to add a new layout where functions are described behaviorally which will be transformed during the compilation process into a structural description.

GUBS is a behavioral specification language describing causal relation with logical rules. In brief, the language contains four concepts : stimulus reaction as input, membrane compartmenting, regular behavior (occurring all the time) and singular behavior (occurring sometimes). GUBS's semantic is based on multi-modal Hybrid logic. The goal of the compilation is to select components in a database (e.g., registry of parts) whose assembly behaves similarly to the specified program.

References

- [1] A. Basso Blandin. développement d'un langage pour la biologie synthétique. Mémoire de Master 2 MOPS, September 2012.
- [2] Lesia Bilitchenko, Adam Liu, Sherine Cheung, Emma Weeding, Bing Xia, Mariana Leguia, J Christopher Anderson, and Douglas Densmore. Eugene—a domain specific language for specifying and constraining synthetic biological parts, devices, and systems. *PloS one*, 6(4):e18882, January 2011.
- [3] Peter a Carr and George M Church. Genome engineering. *Nature biotechnology*, 27(12):1151–62, December 2009.
- [4] Michael J Czar, Yizhi Cai, and Jean Peccoud. Writing DNA with GenoCAD. *Nucleic acids research*, 37(Web Server issue):W40–7, July 2009.
- [5] Franck Delaplace, Hanna Kludel, and Amandine Cartier-Michaud. Discrete causal model view of biological networks. In *Proceedings of the 8th International Conference on Computational Methods in Systems Biology - CMSB '10*, pages 4–13, New York, New York, USA, September 2010. ACM Press.
- [6] M Pedersen. Towards programming languages for genetic engineering of living cells. *Journal of the Royal Society, Interface*, 6 Suppl 4:S437–50, 2009.
- [7] Priscilla E M Purnick and Ron Weiss. The second wave of synthetic biology: from modules to systems. *Nature reviews. Molecular cell biology*, 10(6):410–22, 2009.
- [8] P. Umesh, F. Naveen, C.U.M. Rao, and S.A. Nair. Programming languages for synthetic biology. *and Synthetic Biology*, 4(4):265–269, 2010.

Interference among Meiotic Crossovers varies between and within chromosomes in *Arabidopsis thaliana*

Sayantani Basu Roy¹, Franck Gauthier¹, Matthieu Falque¹,
Olivier C. Martin¹

¹ UMR de Génétique Végétale, INRA Univ Paris-Sud CNRS, Ferme du Moulon,
F-91190 Gif-sur-Yvette, France

Abstract

Crossovers form during meiosis and are responsible for shuffling alleles, a major driver of adaptation and evolution. Each pair of homologous chromosomes will form at least one crossover (in the great majority of species), and when multiple crossovers arise, they typically exhibit “interference”, occurring rarely in close proximity. To model this phenomenon, statistical and mechanistic approaches have been used based mainly on the “Gamma” stationary renewal process and the “beam-film” model. We found significant differences between male and female meiosis in *Arabidopsis thaliana* using both of these modeling approaches. Significant trends were also observed in interference heterogeneity between distinct regions of the same chromosome. This takes us towards a new class of models that incorporate intra-chromosomal variation of interference strength.



A genetic and metagenomic systems biology approach to elucidate predisposing factors to Spondylarthritis with genes networks

Emmanuel Chaplais¹, Gilles Chiocchia², Maxime Breban^{1,2},
Henri-Jean Garchon¹

¹ Immunology-Hematology department, Cochin Institute, INSERM U1016,
CNRS UMR8104, 27 rue du Faubourg Saint-Jacques, F-75014 Paris, France

² Rheumatology clinics, Ambroise-Paré Hospital, and University of Versailles
Saint-Quentin-en-Yvelines, France

Abstract

Spondyloarthritis (SpA) is the second most frequent chronic inflammatory rheumatism (CIR) with prevalence of 0.3% in adults in France. It has a multifactorial etiology with a strong genetic component dominated by the well-established association with HLA-B27. While B27 is present in more than 80% of the patients, only 4% of B27+ subjects from the general population, however, develop SpA. Genome-wide association studies (GWAS) have detected a number of single-nucleotide polymorphisms (SNPs) but, as is now commonplace in multifactorial diseases, they explain poorly the missing heredity. To tackle this issue, our laboratory initiated a variety of specific approaches. First, we took advantage of the availability of large families with multiple cases of SpA and we conducted linkage studies using high-density genotyping microarrays followed by family-based association studies, in 154 families representing 1068 subjects (448 patients). We identified three new linked regions with nonparametric lod scores > 3 (on 6p12, 13q and Xqter), in addition to the MHC and the SPA2 locus we previously identified on chr.9. In addition, using this same dataset, we detected 2000 associated SNPs at the 1% FDR. In parallel, we characterized the transcriptome of dendritic cells, which play a central role in pathogenesis of SpA, and identified differentially expressed genes in patients compared with either familial or unrelated controls. Finally, we set out to characterize the diversity of gut microbiome in 20 SpA patients and in an equivalent number of controls, based on recent discoveries in inflammatory bowel diseases, a group of diseases that also occur frequently in SpA patients.

A major but promising challenge is now to integrate these different layers and kinds of genetic and functional information to discern relevant pathophysiological mechanisms, accounting for their likely heterogeneity, and hopefully to identify key genes as potential therapeutic targets. We describe, as a thesis

project, a systems biology modeling approach for SpA that aims at combining genetic, transcriptomic and metagenomic data, in the context of available information in the B27-transgenic rat model for SpA and in public databases. First, we draw lists of associated SNPs and differentially expressed genes by classical statistical analyses of experimental data. We associate genes with SNPs, depending on their genomic location relative to genes and/or to DNA-binding motifs. Using GO term and metabolic pathways enrichment analysis, we then build an associative network, with genes as nodes, disease impact as node size and gene "interrelationships" as edges. Interrelationships between genes will be inferred by classical data mining but also from our specific experimental data (epistasis between SNPs, eQTLs in trans etc). Finally, we will integrate metagenomic data, using schemes yet to be defined, and also the genetic linkage data, including with the MHC. The latter, perhaps, are the most specific aspect of our data set and should endow the projected network with unique heuristic properties.

Simplified mathematical model for the mammalian circadian clock

Aparna Das¹, Francine Diener¹, Gilles Bernot², Jean-Paul Comet²

¹ Lab. J. A. Dieudonné, Department of Mathematics, University of Nice
Sophia-Antipolis, 28, Avenue Valrose, 06108 Nice Cedex-2, France

² Lab. I3S, UMR 6070 UNS and CNRS, Algorithmes-Euclide-B,
2000 route des Lucioles, B.P. 121, F-06903, Sophia-Antipolis, France

Abstract

In this poster we present a mathematical model as a system of four ordinary differential equations to better understand the intrinsic mechanism of circadian oscillations inside the cell. This model is a simplified version of a model proposed by Leloup and Goldbeter [1, 2]. The four variables are the concentrations of two proteins PER and CRY and the concentration of the complex PER-CRY, first in the cytosol and then in the nucleus. The dynamic of these four variables shows a limit cycle, the period of which may be chosen equal to 24 hours for specific (and realistic) values of the parameters. The main purpose of this poster is to study the Hopf bifurcation which take place with disappearance of the limit cycle for the degradation parameter of the complex PER-CRY in the nucleus crossing some value.

References

- [1] Leloup J.-C. and Goldbeter A., (2003) Toward a detailed computational model for the mammalian circadian clock. *Proc. Natl. Acad. Sci. USA* **100**: 7051-7056.
- [2] Leloup J.-C. and Goldbeter A., (2004) Modeling the mammalian circadian clock: Sensitivity analysis and multiplicity of oscillatory mechanisms. *J. Theor. Biol.* **230**: 541-562.

Multiscale Modelling and Analysis of Planar Cell Polarity in the *Drosophila* Wing

Qian Gao^{1a}, David Gilbert^{1a}, Monika Heiner², Fei Liu²,
Daniele Maccagnola³ and David Tree^{1b}

¹ a: School of Information Systems, Computing and Mathematics;

b: School of Health Sciences and Social Care, Brunel University, UK

² Computer Science Department., Brandenburg Univ. of Tech. Cottbus, Germany

³ Dipartimento di Informatica Sistemistica e Comunicazione, Università degli Studi di Milano-Bicocca, Italy

Abstract

Multiscale modelling in Systems Biology goes beyond the traditional approach of modelling at just one spatial scale. Until now most models have been at the intracellular level, and indeed have largely ignored locality within the cell. However with the growth of data available at different scales, a need has emerged to increase the spatial scope of models of biological systems to enable descriptions at the intercellular (cell-cell), tissue, organ and even whole organism scales.

Multiscale modelling attempts to model across several spatial scales, and introduces a series of challenges, e.g. repetition, variation, organisation, replication or deletion of components; hierarchical organisation; communication between components and mobility. We have chosen the biological example of Planar Cell Polarity (PCP) signalling in the *Drosophila* wing, which illustrates several of these issues. PCP is required for many developmental events in both vertebrates and non-vertebrates, and defects in PCP in vertebrates lead to developmental abnormalities in multiple tissues. We apply an approach using the new concept of hierarchically coloured Petri Nets (HCPN) to model a tissue comprising multiple cells hexagonally packed in a honeycomb formation in order to describe the phenomenon of PCP. This exemplar illustrates the power of our approach for describing spatial multiscale problems in biological systems, which requires computational experiments over very large underlying models represented by systems of Ordinary Differential Equations (ODEs). We have constructed a family of related models, permitting different hypothesis to be explored regarding the mechanism underlying PCP. These models each comprise about 150,000 ODEs. In addition, our models include the description of the effect of genetic mutations by inserting a patch of mutated cells (clone) lacking one key signalling protein into a wild-type field of cells. We have applied a set of analytical techniques including clustering and model checking over time series of primary and secondary data. Our models support the interpretation of biological observations reported in the literature.

A polymer model for studying long supercoiled DNA molecules

Thibaut Lepage¹, Ivan Junier^{2,1} and François Képès¹

¹Epigenomics Project / institute of Systems and Synthetic Biology, Genopole®, CNRS, University of Évry, France

²Centre for Genomic Regulation, Barcelona, Spain

Abstract

Most bacterial chromosomes are circular and negatively supercoiled. This superhelicity strongly affects the 3D structure of DNA. In particular, when supercoiled, circular DNA folds into braided conformations (plectonemes). These superstructures can bring distant parts of the chromosome close to one another in space. Thus, they can have a crucial effect on the juxtaposition of genes and hence on the properties of transcriptional co-regulation. Superhelicity is also known to play an important role in transcriptional regulation by affecting the structure of promoter regions [1]. It is suspected to regulate transcription at a global scale in organisms with a very small genome, where no or few transcription factors are known [2].

Several models have been proposed to study supercoiled DNA. Quasi-atomic models [3, 4] are restricted to small molecule lengths (a few helices). Brownian dynamics allows the simulation of larger molecules but, still, is not convenient to investigate large molecule sizes [5]. Monte Carlo simulations of a several kilobasepair-long coarse-grained polymer model of DNA have been carried out [6]. However, the topological constraints imposed by the double-helix structure of DNA, namely, the conservation of the linking number, were not ensured along the simulations. Here, we present a model respecting the conservation of the linking number all the way through the Monte Carlo procedure and enabling the simulation of long DNA molecules.

References

- [1] A. Travers and G. Muskhelishvili (2005), DNA supercoiling - a global transcriptional regulator for enterobacterial growth? *Nat. Rev. Microbiol.*, **3**: 157-169.
- [2] C. J. Dorman (2011), Regulation of transcription by DNA supercoiling in *Mycoplasma genitalium*: global control in the smallest known self-replicating genome. *Molec. Microbiol.*, **81**: 302-304.

- [3] M. Sayar, B. Avşaroğlu and A. Kabakçioğlu (2010), Twist-writhe partitioning in a coarse-grained DNA minicircle model. *Phys. Rev. E*, **81**: 041916.
- [4] T. B. Liverpool, S. A. Harris and C. A. Laughton (2008), Supercoiling and Denaturation of DNA Loops. *Phys. Rev. Lett.*, **100**: 238103.
- [5] K. Klenin, H. Merlitz, and J. Langowski (1998), A Brownian Dynamics Program for the Simulation of Linear and Circular DNA and Other Wormlike Chain Polyelectrolytes. *Biophys. J.*, **74**: 780-788.
- [6] Z. Yang, Z. Haijun, and O.-Y. Zhong-can (2000), Monte Carlo Implementation of Supercoiled Double-Stranded DNA. *Biophys. J.*, **78**: 1979-1987.

Benchmarking gene regulatory networks in silico

Adrien Henry¹, Yulang Luo¹, Areejit Samal^{1,3} and Olivier C. Martin^{1,2}

¹ LPTMS, CNRS UMR 8626, Bât 100, Univ Paris-Sud, France

² INRA, Univ Paris-Sud, CNRS, INA PG, UMR0320 / UMR8120
Génétique Végétale, F-91190 Gif-sur-Yvette, France

³ Institute for Systems Biology, Seattle, Washington, United States of America

Abstract

Biological networks have numerous structural features that distinguish them from random networks: degree distribution, assortativity, presence of motifs – that is over-representation of small sub-graphs, etc. The standard benchmark against which one detects these kinds of features is the ensemble of networks obtained by edge randomization of the biological network under consideration.

In this work, we ask whether such "unusual" features might follow from the functional capabilities of biological networks. We tackle this question in silico for the case of gene regulatory networks. We model transcriptional dynamics via a Boolean update rule that may be different for each gene. Our benchmark ensemble of "functional" in silico regulatory networks is then the set of networks or "genotypes" – defined by an interaction graph and for each gene a Boolean function – that reproduce biologically-specified expression patterns. The space of all such functional genotypes is huge and so we sample it by Markov Chain Monte Carlo. Using this benchmark ensemble, we then redefine the notion of "unusual" features, and find that instead biological networks are no longer atypical. Thus it is plausible that the unusual features found in gene regulatory networks are in a sense "necessary" since they arise from the functional constraints associated with enforcing given expression patterns, at least within our modeling framework.

References

- [1] Z. Burda, A. Krzywicki, O.C. Martin and M. Zagorski, (2011) Motifs emerge from function in model gene regulatory networks. *PNAS* 108 (42) 17263-17268.
- [2] Kauffman, S., (1993) The origins of order. *Oxford University Press*, Oxford.

Modelling the Cerebral Cortex

Abdallah Zemirline¹

¹ Lab-STICC, UMR 3192, Université de Brest, BP 809, F-29238 Brest Cedex 3, France

Abstract

We propose to model the cerebral cortex by a BF-hypergraph; where a node represents the cell body of a neuron and its axons (resp. dendrites) are represented by a F-harc (B-harc).

1 Introduction

Living beings evolve by adapting to the world complexity and its unsteadiness. In particular, the human brains analyse the confrontations with difficult situations, extract information from salient events, and learn from that knowledge how in the future to behave efficiently in the presence of similar situations.

When J.Z. Tsien and his colleagues [1] attempted to unravel how the brain makes memories, by means of a set of relevant experiments and some powerful mathematical analyses, as well as by the recording of the activity of more 200 neurons in awake mice, they discovered the basic mechanism used by the brain to draw vital information from experiences and turn that information into memories. Their results consolidate the fact that a large population of interacting neurons is necessary to observe in vivo for explaining how the brain achieves perceptions and memories.

2 BF-Hypergraphs for Modelling the Cerebral Cortex

The cerebral cortex is the outer layer of the cerebrum. It consists of neurons, i.e. the neuronal cell bodies, the unmyelinated fibers, and the myelinated axons. The neocortex is the area of the cortex which consists only of the neuronal cell bodies and unmyelinated fibers, called the grey matter, and surrounding the deeper white matter constituted of the myelinated axons. In humans and large mammals, it presents deep grooves and wrinkles which allow to develop functions responsible for enhanced cognitive skills such as speech, language, working memory, consciousness, reflexion, abstract thought, and imagination.

A neuron [2] is a nerve cell having a cell body or soma (with a nucleus) that processes information, and several cytoplasm extensions (axons and dendrites) that allow it, by means of electrical and chemical signaling, to dispatch (axons) and receive (dendrites) informations. Roughly, signal exchanges between different neurons are made at the level of the synapses. Although the synapses are often constituted between axons of cells and dendrites of other cells, there are other types of synaptic junctions, between axons, between dendrites, and between axons and cellular bodies. A neuron may have several thousands of axons and much more of dendrites.

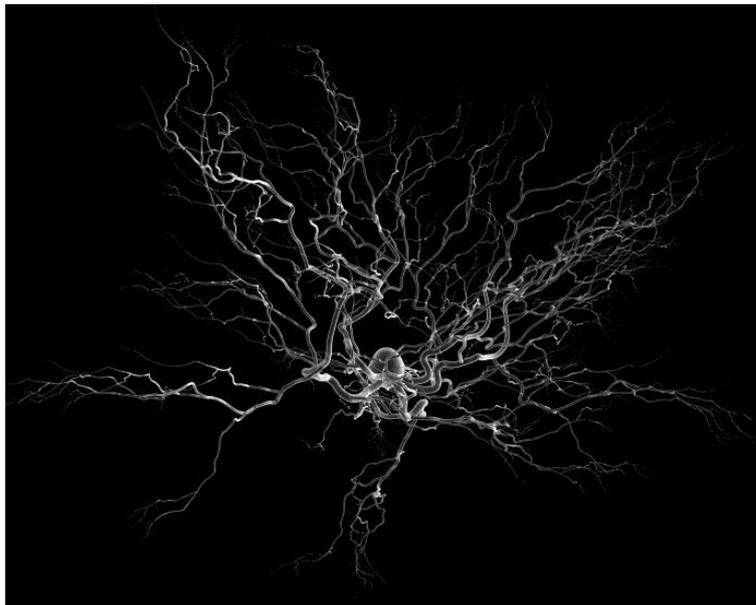


Figure 1: Hereafter, an image of a neuron (GNU General Public Licence), it comes from the website: <http://fr.wikipedia.org/wiki/Fichier:Neuron-SEM-2.png>

Classical neural networks [3], like for example Hopfield neural network [4], are rather representations of the neocortex. They have a structure of some particular graphs, cliques, bipartite graphs, etc. In these models, only nodes are used to represent cell bodies, but the representations of axons and dendrites are missing. Would there an efficient model for the cerebral cortex, I do not know.

To explore especially that internal representation of external events in the brain is achieved not by recording exact details of those events, but rather by recreating its own selective pictures based on cognitive importance, as it

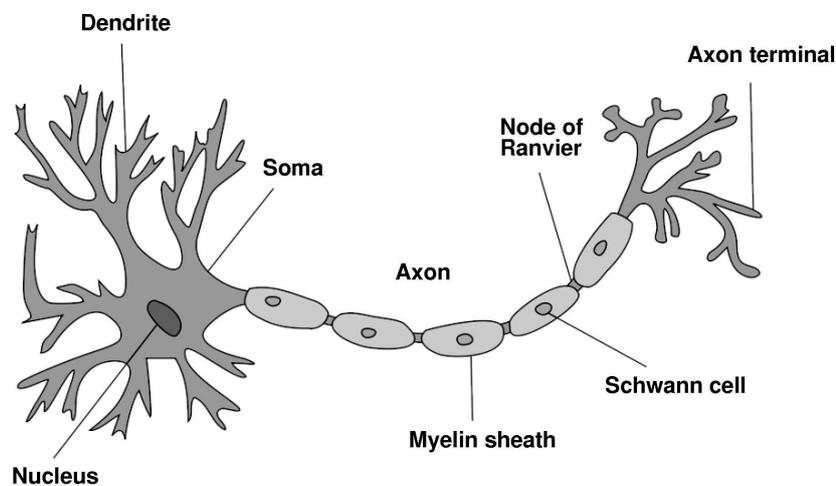


Figure 2: This is a schema of a neuron from wikipedia, the free encyclopedia. A more detailed picture (placed in the public domain by its author LadyofHats) representing a complete neuron cell diagram can be found in the website : http://wikipedia.org/wiki/Myelin_sheath_gap

is suggested in [1], and in order to get universal categorizations of internal brain representations across individuals and species, a relevant modelling of the cerebral cortex as well as efficient simulations of the internal brain representations are required.

This is why we describe hereafter a modelling of the cerebral cortex, based on directed hypergraphs.

A directed hypergraph [5] is a pair $H = (V, A)$ where V is the set of vertices, and E is the set of hyperarcs. It is a generalization of a directed graph. A hyperarc is an arc which may have several heads and several tails. A hyperarc a is then a pair $(tail(a), head(a))$, where $tail(a)$ (resp. $head(a)$) is the subset of V of all tails (resp. heads) of a . The subsets $head(a)$ and $tail(a)$ are supposed disjoint.

A hyperarc a is a F-harc (resp. B-harc) if $head(a)$ (resp. $tail(a)$) has precisely one element.

A F-hgraph (resp. B-hgraph) is a hypergraph whose hyperarcs are F-harcs (resp. B-harcs). A BF-hgraph is a hypergraph which admits both F-harcs and B-harcs. Let us observe that every hypergraph can be transformed into a BF-hgraph by replacing every hyperarc a by the two hyperarcs $(v_a, head(a))$ and $(tail(a), v_a)$, where v_a is a new node.

We propose to model the cerebral cortex by a BF-hypergraph; where a node represents the cell body of a neuron and its axons (resp. dendrites) are represented by a F-harc (resp. B-harc).

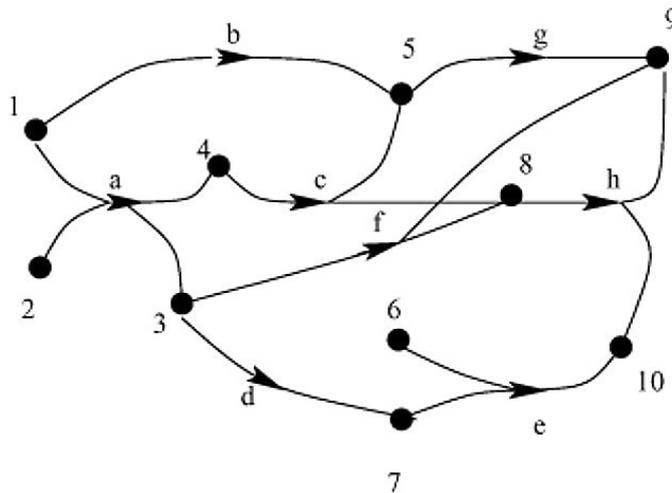


Figure 3: A directed hypergraph

Besides, a r -uniform directed hypergraph $H = (V, A)$ is a directed hypergraph whose all hyperarcs are of size r .

References

- [1] L. Lin, R. Osan, and J. Z. Tsien, *Organizing principles of real-time memory encoding : Neural clique assemblies and universal neural codes*. Trends Neurosci., 2006. **29**(1): 48-57.
- [2] Kenneth Saladin, *Anatomy and Physiology: The Unity of Form and Function*. McGraw-Hill Companies Inc, 2010.
- [3] A. Zemirline *et al.*, *High Level Course*, Proceedings of the Autrans seminar on Modelling and simulation of biological processes in the context of Genomics, 2002. pp. 281-293.
- [4] John J. Hopfield, *Neural networks and physical systems with emergent collective computational abilities*. Proc. Nat. Acad. Sci., Biophys., 1982. **79**(8): 2554-2558.
- [5] Claude Berge, *Hypergraphs*. North-Holland Mathematical Library, 1989.

LIST OF ATTENDEES

(April 5th, 2012)

ADAM Laura	(ladam@vbi.vt.edu)
AMAR Patrick	(pa@lri.fr)
BASSO BLANDIN Adrien	(abasso@ibisc.univ-evry.fr)
BASU ROY Sayantani	(sbasu@moulon.inra.fr)
BERNOT Gilles	(bernot@unice.fr)
BEURTON-AIMAR Marie	(beurton@labri.fr)
BOUFFARD Marc	(marc.bouffard@u-psud.fr)
CHAPLAIS Emmanuel	(emmanuel.chaplais@inserm.fr)
COLLIN Olivier	(ocollin@biologie.ens.fr)
DAS APARNA	(aparna.das@unice.fr)
DELAPLACE Franck	(franck.delaplace@ibisc.univ-evry.fr)
DOULAZMI Mohamed	(mohamed.doulazmi@upmc.fr)
ELATI Mohamed	(mohamed.elati@issb.genopole.fr)
ELLIS Tom	(t.ellis@imperial.ac.uk)
GAO Qian	(Qian.Gao@brunel.ac.uk)
HENRY Adrien	(adrien.henry@u-psud.fr)
JARAMILLO Alfonso	(Alfonso.Jaramillo@issb.genopole.fr)
JESTER brian	(jester_brian@hotmail.com)
JUNIER Ivan	(i.junier@gmail.com)
KÉPÈS François	(francois.kepes@issb.genopole.fr)
KLUMPP Stefan	(klumpp@mpikg.mpg.de)

LE FÈVRE François	(flefevre@genoscope.cns.fr)
LEPAGE Thibaut	(thibaut.lepage@issb.genopole.fr)
LE GALL Pascale	(pascale.legall@issb.genopole.fr)
LEVI Francis	(francis.levi@inserm.fr)
MAZAT Jean-Pierre	(Jean-Pierre.Mazat@phys-mito.u-bordeaux2.fr)
MEDIGUE Claudine	(cmedigue@genoscope.cns.fr)
MOLINA Franck	(franck.molina@sysdiag.cnrs.fr)
MUGGLETON Stephen	(s.muggleton@imperial.ac.uk)
NDIAYE Ndeye Fatou	(iseg.senegal@gmail.com)
NORRIS Victor	(victor.norris@univ-rouen.fr)
PERES Sabine	(sabine.peres@lri.fr)
RAFIEI Nafiseh	(rafiei@issb.genopole.fr)
ROSU Bianca	(redtwo_ank@yahoo.com)
SENÉ Sylvain	(sylvain.sene@ibisc.univ-evry.fr)
STANO Pasquale	(stano@uniroma3.it)
THIERRY Alain	(alain.thierry@sysdiag.cnrs.fr)
VIROLLE Marie-Joelle	(marie-joelle.virolle@igmors.u-psud.fr)
ZELISZEWSKI Dominique	(dominique.zeliszewski@issb.genopole.fr)
ZEMIRLINE Abdallah	(zemirlin@univ-brest.fr)
ZINOVYEV Andrei	(andrei.zinovyev@curie.fr)