Modelling biochemical reaction networks in bacteria – From data to models and back

Delphine Ropers

N° d'ordre : 023 2021

HABILITATION A DIRIGER DES RECHERCHES

Délivrée par :

l'Université Claude Bernard Lyon 1

Ecole Doctorale 341

E2M2 : Evolution Ecosystèmes Microbiologie Modélisation

Spécialité : Bioinformatique

Soutenue publiquement le 17/05/2021, par :

Delphine Ropers

Modelling biochemical reaction networks in bacteria

From data to models and back

Devant le jury composé de :

Alexander BOCKMAYR Professeur, Freie Universität Berlin	**Rapporteur**
Matthieu JULES Professeur, AgroParisTech / Chargé de recherche, INRAE	**Rapporteur**
Vassily HATZIMANIKATIS Professeur, Ecole Polytechnique Fédérale de Lausanne	**Rapporteur**
Sandrine CHARLES Professeure, Université de Lyon I	**Présidente**
Muriel COCAIGN-BOUSQUET Directrice de recherche, INRAE	**Examinatrice**
Johannes GEISELMANN Professeur, Université Grenoble - Alpes	**Examinateur**

Université Claude Bernard – LYON 1

Administrateur provisoire de l'Université	M. Frédéric FLEURY
Président du Conseil Académique	M. Hamda BEN HADID
Vice-Président du Conseil d'Administration	M. Didier REVEL
Vice-Président du Conseil des Etudes et de la Vie Universitaire	M. Philippe CHEVALLIER
Vice-Président de la Commission de Recherche	M. Jean-François MORNEX
Directeur Général des Services	M. Pierre ROLLAND

COMPOSANTES SANTE

Département de Formation et Centre de Recherche en Biologie Humaine	Directrice : Mme Anne-Marie SCHOTT
Faculté d'Odontologie	Doyenne : Mme Dominique SEUX
Faculté de Médecine et Maïeutique Lyon Sud - Charles Mérieux	Doyenne : Mme Carole BURILLON
Faculté de Médecine Lyon-Est	Doyen : M. Gilles RODE
Institut des Sciences et Techniques de la Réadaptation (ISTR)	Directeur : M. Xavier PERROT
Institut des Sciences Pharmaceutiques et Biologiques (ISBP)	Directrice : Mme Christine VINCIGUERRA

COMPOSANTES & DEPARTEMENTS DE SCIENCES & TECHNOLOGIE

Département Génie Electrique et des Procédés (GEP)	Directrice : Mme Rosaria FERRIGNO
Département Informatique	Directeur : M. Behzad SHARIAT
Département Mécanique	Directeur M. Marc BUFFAT
Ecole Supérieure de Chimie, Physique, Electronique (CPE Lyon)	Directeur : Gérard PIGNAULT
Institut de Science Financière et d'Assurances (ISFA)	Directeur : M. Nicolas LEBOISNE
Institut National du Professorat et de l'Education	Administrateur Provisoire : M. Pierre CHAREYRON
Institut Universitaire de Technologie de Lyon 1	Directeur : M. Christophe VITON
Observatoire de Lyon	Directrice : Mme Isabelle DANIEL
Polytechnique Lyon	Directeur : Emmanuel PERRIN
UFR Biosciences	Administratrice provisoire : Mme Kathrin GIESELER
UFR des Sciences et Techniques des Activités Physiques et Sportives (STAPS)	Directeur : M. Yannick VANPOULLE
UFR Faculté des Sciences	Directeur : M. Bruno ANDRIOLETTI

Abstract

English abstract

With the advent of new technologies, experimental data in biology has exploded in size and complexity. It is now possible to simultaneously quantify different components of the cell at metabolic, transcriptomic, proteomic, and phenotypic levels. Connecting these different multi-scale and dynamic datasets provides an integrated view of cellular growth and informs us about the underlying molecular networks of genes, RNAs, proteins and metabolites that control the adaptation of the cell to the environment. This is the perspective offered by mathematical modelling and computer simulation, allowing the association of different microscopic and macroscopic scales. This is a difficult problem however, because of the noise and the heterogeneity of the data, and of the size and the nonlinearity of the models. As a consequence, a large number of datasets are only partially analysed and underexploited. This manuscript describes the work I have carried out to improve the utilization of experimental data to gain a better understanding of the adaptation of bacterial growth to a changing environment. This work has been carried out within the Ibis project-team (Inria, Université Grenoble Alpes) with my colleagues, especially the students that I have had the chance to supervise.

After the introductory Chapter 1, I describe in Chapter 2 the modelling of cellular networks using ordinary differential equations as well as simplification and approximation of the models depending on the nature of the available data and the questions addressed. These principles are applied in Chapter 3 to the qualitative analysis of the dynamics of gene networks in the context of the carbon starvation response in *Escherichia coli* bacteria. With the general trend of biology becoming increasingly quantitative, modelling studies require obtaining reliable gene expression and metabolomic data, the analysis of which requires the development of suitable methods described in Chapter 4. Chapter 5 examines the strong link between the activity of the cellular gene expression machinery and bacterial growth rate. This understanding is used to develop a synthetic strain of *E. coli* whose growth control makes it possible to divert the flow of precursors for growth towards the bioproduction of molecules of biotechnological interest. In Chapter 6, large-scale reconstructions of central carbon metabolism are used as platforms to interpret datasets regarding the post-transcriptional regulation of central carbon metabolism in *E. coli*. Chapter 7 is dedicated to the genome-scale analysis of mRNA decay by means of dynamic transcriptomics data. I describe in Chapter 8 ongoing and future projects towards the integrative analysis of microbial growth and resource allocation strategies. The scientific developments of these projects are expected to shape my own research activity in the coming years and that of the future project-team, under creation, that I will lead.

Long French abstract

Avec l'arrivée des nouvelles technologies, les données expérimentales en biologie ont explosé en taille et complexité. Il est désormais possible de quantifier en même temps différents composants de la cellule au niveau métabolique, transcriptomique, protéomique et de caractéristiques phénotypiques comme le taux de croissance. Relier ces différents jeux de données multi-échelles et dynamiques permet d'obtenir une vision intégrée de la croissance cellulaire, en nous renseignant sur la façon dont les réseaux moléculaires sous-jacents de gènes, ARN, protéines et métabolites contrôlent l'adaptation des cellules à leur environnement. C'est le cadre qu'offrent la modélisation mathématique et la simulation informatique, en permettant d'associer les différentes échelles microscopiques et macroscopiques. C'est cependant un problème difficile, du fait du bruit et de l'hétérogénéité des données d'une part, et de la taille et la forme non-linéaire des modèles d'autre part. La conséquence est qu'un grand nombre de jeux de données ne sont que partiellement analysés et sous-exploités.

Ce manuscrit décrit les travaux que j'ai menés pour améliorer l'utilisation de données expérimentales afin d'obtenir une meilleure compréhension de l'adaptation de la croissance bactérienne à un environnement changeant. Ces travaux ont été menés au sein de l'équipe-projet Ibis (Inria, Université Grenoble Alpes) avec mes collègues, en particulier les étudiants que j'ai eu la chance d'encadrer. Après un premier chapitre d'introduction, je décris en chapitre 2 les concepts de base de la modélisation des réseaux biochimiques. Je détaillerai en particulier les reconstructions du métabolisme cellulaire à l'échelle du génome et la modélisation cinétique des réactions enzymatiques, dont les concepts sont utilisés dans plusieurs travaux présentés dans ce manuscrit. La grande dimension et non linéarité des modèles cinétiques complique l'estimation de leurs paramètres et l'analyse de leur dynamique. Je présenterai des travaux sur des simplifications appropriées pour ces modèles selon la nature des données à disposition et les questions abordées, comme la réduction de modèles d'équations différentielles ordinaires (ODE) par séparation des échelles de temps ou l'approximation des modèles ODE par des modèles linéaires par morceaux. Du fait de leur dérivation rigoureuse, les modèles simplifiés retiennent les principales caractéristiques des modèles ODE. Ces approches seront utilisées pour les différents modèles dynamiques présentés dans ce manuscrit.

Dans le chapitre 3, je présente des travaux d'analyse de la dynamique d'un réseau de régulation génique contrôlant la réponse à la privation en carbon de la bactérie *Escherichia coli*. Lors de ces travaux, l'absence de données quantitatives dans la littérature ne permettait pas d'utiliser un modèle ODE pour décrire la dynamique du système. J'ai plutôt analysé la dynamique d'une version linéaire par mocreaux de ce modèle par une approche de modélisation et simulation qualitative. Je décrirai le principe de cette approche avec un exemple simple et son application à l'étude du réseau de la réponse au manque de source de carbone. Cette approche a permis pour la première fois de relier la croissance d'*E. coli* avec les principaux régulateurs transcriptionnels de la bactérie, et de comprendre les cascades de régulations mises en place lors de la réponse à une privation en glucose ou du rédémarrage de croissance sur ce sucre.

L'évolution de la biologie en une science quantitative permet d'obtenir de nombreuses

données d'expression génique et du métabolisme cellulaire. La fiabilité de ces données nécessite le développement de méthodes d'analyse adaptées décrites dans le chapitre 4. Je décrirai des travaux sur l'analyse de données de gènes rapporteurs et l'analyse de données de métabolomique afin de pouvoir reconstruire des profils d'activités de promoteurs et de concentrations de protéines dans le premier cas, et des vitesses d'import et secrétion de métabolites extracellulaires, ainsi que des taux de croissance dans le second cas. Les données quantitatives utilisées dans le reste du manuscrit ont été analysées grâce à ces approches.

Le chapitre 5 s'intéresse au lien étroit entre activité de la machinerie cellulaire d'expression génique et taux de croissance bactérien. A l'aide de modèles simples intégrant des données expérimentales de gènes rapporteurs, nous montrons le rôle clé joué par la machinerie d'expression génique dans l'adaptation globale de l'expression des gènes au cours de la croissance. Ces travaux montrent que le fonctionnement des réseaux biochimiques ne peut être déconnecté de l'état physiologique de la cellule. Cette compréhension est utilisée pour l'ingénierie d'une souche d'*E. coli* synthétique dont le contrôle de la croissance permet de divertir les flux de précurseurs pour la croissance vers la bioproduction de molécules d'intérêt biotechnologique.

Dans le chapitre 6, de grandes reconstructions du métabolisme et différents jeux de données (métabolomique, activités spécifiques) sont utilisées pour étudier la régulation post-transcriptionnelle du métabolisme central carboné chez *E. coli*. Ces travaux ont permis d'expliquer les conséquences physiologiques de l'atténuation du gène de la protéine CsrA et d'identifier des ARNm cibles de cette protéine. Nous avons en outre pu montrer que chez *E. coli* également, le glycogène joue un rôle de stockage de sucre qui sert de source d'énergie pour faciliter la transition de la croissance bactérienne d'une source de carbone à une autre.

Le chapitre 7 s'intéresse à la dégradation de l'ensemble des ARNm d'*E. coli*. Je décrirai le développement d'un modèle simple reposant sur des approches de quasi-équilibre et permettant de prédire la cinétique de dégradation de chacun des ARNm cellulaires de la bactérie *E. coli*. Nous avons pu formuler de nouvelles hypothèses sur le rôle possible de la compétition entre ARNm pour leur fixation au dégradosome lors de l'adaptation de la croissance bactérienne à des changements environnementaux. Nous montrons également que ce mécanisme de compétition joue un rôle physiologique grâce à une approche de modélisation non linéaire à effets mixtes utilisant le modèle mécanistique de la dégradation des ARNm et des jeux de données de transcriptomique dynamique mesurant la cinétique de disparition des ARNm cellulaires.

Le chapitre 8 est dédié à des projets en cours et futurs sur l'analyse intégrative de la croissance microbienne et les stratégies d'allocation de resources des bactéries. Les travaux menés dans le cadre de ces projets vont définir mon activité scientifique dans les années à venir et celle de la future équipe-projet, en cours de création, dont je prendrai la direction.

Acknowledgements

First, I would like to thank Alexander Bockmayr, Matthieu Jules, and Vassily Hatzimanikatis, who kindly accepted to review this manuscript. Alexander guided my first steps in the modelling of biological systems a long time ago. I am delighted to have him on my habilitation committee. I would also like to thank the other members of my committee, Sandrine Charles, Muriel Cocaign-Bousquet, and Hans Geiselmann for their time and availability.

The present habilitation thesis is the result of the research that I have carried out within the Inria project-teams HELIX and IBIS. Many thanks go to my colleagues in these teams for warm and friendly atmosphere, and for the many scientific and non-scientific discussions. Special thanks go to François Rechenmann, Alain Viari, Hidde de Jong, Hans Geiselmann, Michel Page, Eugenio Cinquemani, Aline Marguet, and to the members of the BIOP team at LIPhy, in particular Corinne Pinel, and to Olivier Ali. I would especially like to thank all students and young researchers whom I have had the pleasure to (co-)supervise or otherwise collaborate with: Valentina Baldazzi, Sara Berthoumieux, Ismail Belgacem, Stefano Casagranda, Thibault Etienne, Jérôme Izard, Nils Giordano, Edith Grac, Manon Morin, Stéphane Pinhal, Pedro Tiago Monteiro, and Valentin Zulkower. I am also grateful to Inria for the support to carry out this pluridisciplinary research and for providing a great work environment.

The research presented in this manuscript is the result of a broad collaborative effort. I consider myself lucky to have worked with so many talented and nice people. In addition to the people I mentioned above, many others contributed to this research, directly or indirectly. I would like to thank, in particular, Jean-Luc Gouzé, Aline Métris, Jozsef Baranyi, Grégory Batt, Andreas Kremling, Tomas Gedeon, Laurent Trilling, Yohann Couté, and Myriam Ferro. For succinctness of the manuscript, I have had to omit some of the papers we have worked on together. Be sure, however, that I value our collaborations. Special thanks go to Jean-Luc Gouzé for our many scientific discussions along the years and for his kindness. Thanks to him, model reduction has become less mysterious. I would like also to thank Marie-France Sagot and all members from the Inria project-team ERABLE for our nascent collaboration on the analysis of metabolic graph models.

A significant part of my research is now carried out in tandem with Muriel Cocaign-Bousquet. I would like to warmly thank her for our scientific and non-scientific discussions. I value them so much! I am yearning to go back to Toulouse in the post-COVID times to pursue our brain-storming meetings on mRNA degradation and cellular metabolism, wrapped up at the end of the day by a "crêpe suzette". I am happy that you will be a member of the new project-team that we are creating. Thank you also to all TBI members, in particular to the BLADE team, for the friendly atmosphere and the many scientific exchanges.

And... I would never be where I am without my family and my friends... Warm and heartfelt thanks to you, especially to my parents and my sisters and brother! My deepest gratitude goes to Hidde, and to Quentin, Inès, and Siaka: you are illuminating my life.

Contents

Chapter 1

Introduction

1.1 Context

This manuscript is an overview of my research activities at Inria over the past fifteen years. The common denominator of the work presented is the use of computational systems biology approaches to unravel the complex molecular mechanisms involved in the adaptation of microorganisms to their environment. My work thus heavily relies on experimental data, seeking to maximally exploit and make sense of them. Before I describe my own contributions, I will place them in the context of the constantly evolving field with multiple facets that is systems biology. I will focus in particular on its historical context, because it has shaped the methodology and practice of modern systems biology in general and my own research trajectory in particular.

Much has been said and written about systems biology [Ideker et al., 2001, Kitano, 2001, 2002, Kirschner, 2005, Wolkenhauer, 2001, to name but a few]. Still trying to establish itself as a new field of science, systems biology has about as many definitions as there are systems biologists [Calvert and Fujimura, 2011]. We here summarize the literature by saying that systems biology refers to the body of approaches seeking to understand biological systems at the systems level. However, there is a general agreement on the agenda of systems biology, that it moves away from just cataloguing and characterizing cell components at the molecular level, to look at how they functionally interact and give rise to complex biological processes [Ideker et al., 2001, Kitano, 2001]. Reaching such understanding requires a collection of methods and ideas from various disciplines. It includes the biochemical and molecular biology approaches to analyse network components and their interactions. It shares with physiology the ultimate goal and methods to understand the functioning of living organisms and the underlying mechanisms, and with developmental biology, the analysis of how these mechanisms lead to a succession of physiological states. It makes use of various statistical approaches to analyse the vast amount of molecular data and of computational models of emergent phenomena for predicting the behaviour of biological systems. Eventually, systems biology includes aspects from evolutionary biology and ecology, acknowledging that living organisms are products of selection, which is more difficult to understand on a molecular level [Kirschner, 2005]. The combination of these various disciplines depends on the problem studied and allows to obtain an overall explanation of complex biological phenomena, which would not have been possible otherwise [Brigandt, 2010].

Systems biology is thus an integrative field that has been built on progresses in the disciplines mentioned above and past attempts. As such, it is both an old and a new field in biology [Auffray and Nottale, 2008, Kitano, 2002]. An old field, because system-level understanding in biology was already proposed between the 1930s and the 1970s. Far from being exhaustive, we can cite pioneering works such as the General System Theory introduced by the physiologist L. von Bertalanffy which, applied to biology, aimed at studying living organisms as a whole and considered them as open systems interacting with the environment [Rosen, 1958, von Bertalanffy,

1940, Von Bertalanffy, 1950, 1969]. We owe the notion of homeostasis to the physiologist W.B. Cannon [Cannon, 1929, 1939, Cooper, 2008], while the control engineer N. Wiener introduced the concept of negative feedback [Wiener, 1948], central to understand homeostasis. The name of systems biology itself was coined about fifty years ago [Mesarović, 1968] and first theoretical studies of morphogenesis [Turing, 1952], metabolism [Heinrich and Rapoport, 1974, Kacser, 1973, Mitchell, 1961, Rottenberg et al., 1967, Savageau, 1976], gene expression [Goodwin et al., 1963, Kauffman, 1971], and the nonlinearities of complex biological systems [e.g., Prigogine and Nicolis, 1971] were published over the same period.

All these studies laid the ground for the systems view in biology, but systems biology itself took off only at the beginning of the twenty-first century as a new biological field [Auffray and Nottale, 2008, Kitano, 2002]. Various explanations have been offered. First, early works like those of L. von Bertalanffy were considered too abstract and biological systems too complex to be modelled. As an illustrative example of the latter point, I will cite the French physiologist Claude Bernard. In the nineteenth century, he already advocated the use of mathematics in biology [Bernard, 1865, p238]:

> *"Cette application des mathématiques aux phénomènes naturels est le but de toute science, parce que l'expression de la loi des phénomènes doit toujours être mathématique."*

However, the idea was already there that the complexity of biological phenomena and incomplete knowledge prevent their mathematical study [Bernard, 1865, p238-239]:

> *"Or je pense que les tentatives de ce genre sont prématurées dans la plupart des phénomènes de la vie, précisément parce que ces phénomènes sont tellement complexes, qu'à côté de quelques-unes de leurs conditions que nous connaissons, nous devons non seulement supposer, mais être certain, qu'il en existe une foule d'autres qui nous sont encore absolument inconnues. [...] Ce n'est point que je condamne l'application mathématique dans les phénomènes biologiques, car c'est par elle seule que, dans la suite, la science se constituera ; seulement j'ai la conviction que l'équation générale est impossible pour le moment, l'étude qualitative des phénomènes devant nécessairement précéder leur étude quantitative."*

Other explanations offered for the belated emergence of systems biology are essentially a matter of agenda. Somehow, they have followed the idea of Claude Bernard that an accumulation of detailed information at the physiological, cellular, and molecular levels is needed beforehand. An encyclopedic knowledge of the genes and proteins involved in biological phenomena has thus been developed over the years. On a small scale and through *ad hoc* studies at the beginning, but things have accelerated with the availability of genome sequences and the development of high-throughput approaches [Ideker et al., 2001, Kitano, 2002, Voit, 2017]. They allowed the inventory and quantification of cell components, in a wave of technological developments improving the probing, sensing, and imaging of biological systems. Parallel developments in mathematics, physics, and computer science have made theoretical and simulation tools more powerful and more accessible to a wide audience. Hence, following the development of bioinformatics for the processing and statistical analysis of the vast amount of high-throughput data, systems biology has allowed to make sense of the data, by analysing in

terms of biological networks the relationships between genes, proteins, and their function.

I took the time to survey the history of systems biology to better illustrate why systems biology as we know it nowadays is built on the two pillars of experimental data analysis and mathematical modelling, the so-called top-down and bottom-up approaches. The two approaches are the fruit of the evolution of the field and at the centre of an epistemological debate [Ideker et al., 2001, Kell and Oliver, 2004, Westerhoff and Palsson, 2004]. Top-down systems biology is more related to the modern branch of the domain, with the expansion of molecular biology to genome-wide analyses. A data-driven approach in essence, it is concerned with the analysis of large-scale data sets, whose access has been greatly facilitated by the recent technological, algorithmic and computational advances [Bruggeman and Westerhoff, 2007, Ideker et al., 2001]. The challenge is to gain a system-level understanding by integrating these omics data. The induction task is not trivial due to the fact that data can be multi-variate and multi-layered (e.g. transcriptomics, metabolomics, and proteomics data). The most commonly used models to that aim are often phenomenological, in the sense that they neither describe regulatory mechanisms nor include prior knowledge of the molecular components. They account for correlations between measured concentrations, which allows identifying groups of co-regulated genes and formulating hypotheses on biological mechanisms at work, although correlation does not necessarily mean causality [Bersanelli et al., 2016, Noor et al., 2019]. Top-down systems biology is thus more observational, close to the "naturalist" attitude, and – for its detractors – a "fishing expedition" [Calvert and Fujimura, 2011, Kell and Oliver, 2004].

At the other side of the spectrum, the bottom-up approach is hypothesis driven and more related to the old débuts of the field in the mid-twentieth century. It deduces the functional properties of a subsystem, through the development of mathematical models, which describe mechanistically how components interact and predict how the system behaviour emerges from these interactions [Bruggeman and Westerhoff, 2007]. This approach embraces the aspirations of physics and engineering, with the idea to uncover laws and make the behaviour of biological systems predictable and controllable [Calvert and Fujimura, 2011, Kell and Oliver, 2004, Westerhoff and Palsson, 2004]. This has paved the way for the design of biological systems that function close to specification, at the centre of the synthetic biology agenda [Arkin, 2013, Smolke and Silver, 2011].

While top-down and bottom-up approaches have been often opposed, they are just two sides of the same coin, with the common goal of relating the phenotype to the genotype [Kell and Oliver, 2004, Westerhoff and Palsson, 2004]. The literature in the field reflects a progressive reconciliation of these approaches. Constraint-based modelling of metabolic networks is one such example. A favourite approach to model metabolic networks, it provides a comprehensive representation of the metabolism of an organism, by representing the metabolic network through a series of physico-chemical constraints, including reaction stoichiometry and assumption of steady-state metabolite concentrations [Price et al., 2004, Orth et al., 2010, Volkova et al., 2020]. Even though it requires prior knowledge of the network connectivity, constraint-based modelling pertains more to a top-down approach in my opinion, because it makes extensive use of genomic, transcriptomic, proteomic, and metabolomic data and aids their interpretation [Lewis et al., 2010, Machado and Herrgård, 2014, Ramon et al., 2018, Shlomi et al., 2008, Volkova et al., 2020]. However, constraint-based modelling in its classical implementation is unable to give an

insight into cellular substrate concentrations. That is what kinetic modelling does, but this approach suffers because parametrizing mechanistic models is both costly and time-consuming. Bridging the gap between constraint-based models and kinetic models, and thus between top-down and bottom-up approaches, has been a natural expansion of these approaches. Subsequent improvements to constraint-based methods have allowed us to quantitatively understand metabolism and its regulation, notably by accounting for metabolite concentrations [Henry et al., 2007, Hoppe et al., 2007, Kümmel et al., 2006a] and coupling constraint-based models to mechanistic models [e.g., Cotten and Reed, 2013, Covert et al., 2008, Hanly and Henson, 2011, Lee et al., 2008, Mahadevan et al., 2002, Smallbone et al., 2007, Yizhak et al., 2010]. These models predict dynamic metabolic behaviour without explicitly solving ordinary differential equations. Most recent developments allow the parametrization of constraint-based models, turning them into nearly genome-scale kinetic models [Gopalakrishnan et al., 2020].

The research activities presented in this manuscript reflect these different trends of systems biology, with work relying on the use of mathematical models describing the dynamics of biochemical networks and other works related to the model-based interpretation of high-throughput data to explore and characterize molecular mechanisms involved in bacterial adaptation to environmental changes.

1.2 My journey in systems biology

Trained as a biochemist and molecular biologist at the University of Nancy, my scientific interest has always been in interdisciplinary studies of the functioning of biological systems. Over the course of the years, my research has gradually evolved from experimental biology to computational systems biology, although at times I continue to run wet-lab experiments to answer specific research questions or to acquire new data. Throughout this journey, a major concern has been to obtain good data and/or to make it informative for a tighter integration of models and experiments. The nature of the data, its quality and quantity, were determinant for the methodological choices in all studies. In this sense my work has followed the evolution of systems biology and that of biology becoming more and more quantitative.

My first steps in research were in enzymology, during an internship for my Bachelor's degree. I studied the allosteric regulation of a glycolytic enzyme (GAPDH) of *Bacillus stearothermophilus* using a combination of molecular biology and biochemistry tools. As a master and PhD student, I focused on the regulation of HIV-1 RNA splicing by human proteins. HIV-1 expresses its gene products from a single transcript, which undergoes alternative splicing using a combination of four donor and nine acceptor sites. By means of molecular biology, biochemistry, and bioinformatics tools, I studied the regulatory mechanisms driving the choice of acceptor sites by the host cell splicing machinery [Jacquenet et al., 2001, Ropers et al., 2004, Hallay et al., 2006, Khoury et al., 2009]. In collaboration with Alexander Bockmayer and Damien Eveillard, then at Inria in Nancy, I could analyse by means of mathematical modelling and constraint-programming tools how the competition between activating and inhibitory proteins for their binding in the vicinity of splicing sites determines their usage by the splicing machinery [Eveillard et al., 2003, 2004]. The difficulty consisted in the complexity of RNA splicing and the lack of quantitative information to model the process. This first exposure to mathematical modelling shaped the years to come. I pursued the experience by joining the Helix Bioinformatics group at Inria

Grenoble - Rhône-Alpes as a post-doctoral researcher. In collaboration with Hidde de Jong and Hans Geiselmann (now with the interdisciplinary laboratory of physics, LIPhy), I used qualitative modelling approaches to study the functioning of a genetic regulatory network controlling the carbon starvation response of the bacterium *Escherichia coli*. The mathematical formalism allowed to compensate for the lack of quantitative information, although stress responses were relatively well characterized in *E. coli*. The following years witnessed a rapid growth of dynamical and quantitative data at all cell levels, which opened new avenues for integrative analyses of bacterial adaptation to environmental or genetic cues, looking not only at the gene expression level but also at metabolism and cell physiology. This motivated me to adopt more quantitative modelling approaches to analyse the relation between genotype and phenotype. Now as a research scientist with the systems biology group Ibis, my research activities cover aspects of data analysis, mathematical modelling of biochemical networks, and cell growth, with a tight integration of heterogeneous and multi-layered experimental data.

Analysing complex biochemical systems by means of models and experiments requires a broad expertise. On the biological side, the most important skills are in biochemistry, molecular biology, and microbiology. A good knowledge of the biological systems and of the experimental methods is required for mathematical modelling. On the theoretical side, the major domains of expertise are biostatistics, bioinformatics, and dynamical systems. These are critical for data analysis, model development, analysis, and identification. One of my main objective was to gain know-how in these various domains to achieve my research agenda. I enlarged my spectrum of expertise through various collaborations. Currently, part of my research is nurtured by challenging problems on the regulation of cellular metabolism and mRNA metabolism in particular, in the context of an active collaboration with Muriel Cocaign-Bousquet and her group at the Toulouse Biotechnology Institute. The biochemical models of gene expression and cell growth that I develop quickly grow in complexity and require more advanced tools for their reduction and dynamical analysis. I work on these aspects in collaboration with Jean-Luc Gouzé at Inria Sophia Antipolis - Méditerranée. All these models rely on the integration of data. Their analysis and use in parameter estimation pose challenging statistical problems that I tackle in collaboration with Eugenio Cinquemani within Ibis. The experimental aspects of my research, the synthetic biology application, and the modelling of gene expression and cell growth thrive on many interactions with Hidde de Jong and Hans Geiselmann over the course of the years.

1.3 Overview

In the remainder of this manuscript, I will present a selection of my research activities since I joined Inria. I will focus on bacterial growth adaptation, omitting the works dedicated to HIV-1 RNA splicing published after my PhD thesis [Eveillard et al., 2004, Ropers et al., 2004, Hallay et al., 2006, Khoury et al., 2009]. In Chapter 2, I will discuss mathematical models to investigate the dynamical functioning of gene regulatory networks, first where quantitative data are poor if not absent. Models in this case are qualitative, resulting from simplifications of more complex algebro-differential models that can be used to quantitatively describe better-characterized systems. The latter models typically combine different time scales, which poses interesting questions that we addressed by model reduction and simplification to obtain models that are a good compromise between simplicity and biological realism. Chapter 3 is dedicated

to the application of qualitative models of gene expression to analyse the carbon starvation response of *Escherichia coli*. The development of biology into a quantitative science requires the development of approaches for the analysis of metabolomics data as well as reporter gene data. This will be the topic of Chapter 4. In Chapter 5, I will discuss work related to the analysis and control of cell physiology, and how the study of global mechanisms of control of gene expression led to the development of a synthetic strain of *E. coli* allowing to shift the allocation of resources from growth to production of a metabolite of interest. Chapter 6 is dedicated to the analysis of metabolic networks. Metabolomics and other low to high-throughput data are more and more accessible nowadays, which makes metabolic network models suitable platforms for the integration of experimental data to analyse the regulation of metabolism in various genetic and environmental backgrounds. Chapter 7 presents an analysis of bacterial mRNA decay at the genome-wide level, taking advantage of the availability of dynamical transcriptomic data. The outlook in Chapter 8 is the opportunity to conclude and discuss the future. The annex is concerned with administrative and vitae information, as well as a list of the papers highlighted in the manuscript with their abstract.

Chapter 2

Model approximation and reduction

Mathematical models must be realistic representations of the biological system, from the general topology of the underlying biochemical network to the regulation exerted on the biological processes. However, the modelling task is far from trivial as knowledge of the network topology is not exhaustive and quantitative information is often missing. Different modelling formalisms exist to overcome these limitations and adjust the level of precision to the available data. Here we will restrict the description to deterministic approaches. They are well adapted to study network dynamics in cell populations, which is central to my research activities.

The first section introduces basic concepts of biochemical network modelling. I will notably focus on flux analysis approaches, which will be central in Chapter 6. The description of kinetic modelling approaches will develop approximations used in Michaelis-Menten kinetics, their domain of validity, and issues related to parameter estimation. The analysis of mRNA decay described in Chapter 3 relies on these concepts. In the second section, I will describe some of my work on the approximation and reduction of biochemical models, based on a paper in *IEEE/ACM Transactions on Computational Biology and Bioinformatics* [Ropers et al., 2011]. The approaches described in this article were repeatedly used in the modelling works described elsewhere in the manuscript.

2.1 Deterministic modelling of biochemical network models

2.1.1 General form

Ordinary differential equations are the classical formalism for modelling the dynamics of natural and man-made systems. In the case of biochemical reaction networks, they relate the rate of change of variables to their values by means of mass-balance equations. This gives in matrix notation [Heinrich and Schuster, 1996]:

$$\frac{dx}{dt} = Nv - \mu\, x, \quad x(0) = x_0\,. \tag{2.1}$$

In this system of coupled equations, $x \in \mathbb{R}_+^n$ and $v : \mathbb{R}_+^n \to \mathbb{R}^q$ and denote the vectors of reaction rates and concentrations at time t, respectively, $N \in \mathbb{Z}_+^{n\times q}$ is the stoichiometry matrix, and $\mu \in \mathbb{R}_+$ the growth rate. This type of model is based on the premise that the system is well-stirred and that concentrations can be regarded as continuous quantities. We consider here constant pH and temperature conditions because these two parameters are generally controlled in the experiments. The model is used to describe various biological processes such as transcription, translation, enzymatic reactions, complex formation, and transport reactions between compartments.

2.1.2 Flux analysis

When the biological system under study is at steady state, the balance equation in (2.1) becomes a system of algebraic equations [Heinrich and Schuster, 1996]:

$$N v - \mu x = 0 \,. \tag{2.2}$$

By convention, reactions rates at steady-state are termed fluxes. This mathematical description is typically used for metabolic networks, in which case the flux vector v and the stoichiometry matrix N are restricted to the internal metabolites.

An additional and frequently made assumption is that growth dilution of intracellular metabolites is negligible with respect to the turn-over of metabolite pools by enzymatic reactions:

$$N v = 0 \,. \tag{2.3}$$

The system (2.3) is commonly used to analyse stationary fluxes. There is no longer an explicit dependence of fluxes on concentrations and the fluxes are the new variables of the system.

While we may have measurements for some of them, the vast majority of fluxes in vector v is unknown and makes the system under-determined. It is solved by means of constraint-based modelling approaches. Basically, equation (2.3) constrains fluxes to the null space of the stoichiometry matrix N. There is generally no unique solution to this system but a distribution of feasible stationary solutions. It can be narrowed down by the addition of constraints in the form of inequalities applied to the fluxes of each reaction i [for review, Bordbar et al., 2014]:

$$lb_i \leq v_i \leq ub_i \,. \tag{2.4}$$

For instance, the range of incoming and outgoing fluxes can be fixed on the basis of measured uptake and secretion rates [Mo et al., 2009]. Genetic knock-outs are carried out by setting the bounds of the associated reactions to zero [Edwards et al., 2001]. Thermodynamic constraints allow to eliminate infeasible cycles or to set reaction directions based on measured intracellular metabolite concentrations and the Gibbs free energy of formation [Beard et al., 2004, Henry et al., 2007, Hoppe et al., 2007, Kümmel et al., 2006b, Müller and Bockmayr, 2013, Qian and Beard, 2005]. These simple constraints have evolved to more advanced approaches with the growing availability of omics data. Transcriptomics and proteomics data can be used to fix reaction bounds to zero for reactions corresponding to absent mRNAs or proteins or by linearly adjusting the bounds to the mRNA or protein abundances [Åkesson et al., 2004, Chandrasekaran and Price, 2010, Colijn et al., 2009, Covert et al., 2001, Tian and Reed, 2018, van Berlo et al., 2009, Yizhak et al., 2010]. The use of omics data allows the development of context-specific models [e.g. Agren et al., 2012, Bordbar et al., 2014, Heirendt et al., 2019, Jenior et al., 2020, Jensen and Papin, 2011, Shlomi et al., 2008, Thiele et al., 2020, Wang et al., 2012]. This improves the quality of the predictions and turn metabolic network models into scaffolds for the analysis of high-throughput data. Available information on kinetic parameters can also be used to tighten bounds on reaction fluxes [Cotten and Reed, 2013, Fleming et al., 2010].

The application of constraints to reaction fluxes shrinks the solution space to a biologically relevant region. Possible steady-state flux distributions within the region form a convex

polyhedral cone in a high-dimensional space that can be analysed by different but related approaches. Methods for the analysis of pathways, based on elementary mode analysis [Schuster et al., 1999] or extreme-pathway analysis [Schilling et al., 2000], are used to define the limitations and production capabilities of metabolic systems. Flux balance analysis (FBA) uses linear programming to select an optimal flux distribution within the region, for instance by maximizing an objective function such as the biomass or ATP production [Bonarius et al., 1997, Edwards and Palsson, 1999, Sauer et al., 1998, Varma and Palsson, 1994a,b]. The main difference with metabolic flux analysis is the use in FBA of an objective function [Wiechert, 2001, Zupke and Stephanopoulos, 1994]. Flux variability analysis (FVA) has been developed to identify the possible alternative optima of the FBA solution [Mahadevan and Schilling, 2003]. It determines the maximum and minimum fluxes through each reaction when the flux of the objective function is constrained to its maximum value. The choice of an optimization function in FBA and FVA is not trivial, although optimizing cell growth and energy use has been shown to correctly predict metabolic fluxes in microorganisms [Carlson and Srienc, 2004a,b, Edwards and Palsson, 2000, Edwards et al., 2001, Feist and Palsson, 2010, Ibarra et al., 2002]. Studies have questioned the universality of the assumption of such a metabolic objective [Harcombe et al., 2013, Molenaar et al., 2009, Schuetz et al., 2007]. An unbiased alternative approach is random sampling. It explores the solution space with Monte Carlo approaches without optimizing an objective and returns probability distributions of the fluxes [Haraldsdóttir et al., 2017, Kaufman and Smith, 1998, Keaty and Jensen, 2020, Megchelenbrink et al., 2014, Wiback et al., 2004]. The sampling approach is more and more applied for the analysis of metabolic networks [Bordbar et al., 2014].

Metabolic network modelling is an active field of research with an ever growing range of applications and methodologies. In this section, I only sketched the main constraint-based modelling approaches for the sake of brevity. These approaches make use of genome-scale models of cell metabolism, whose reconstruction is far from trivial and time-consuming [Gu et al., 2019, Gudmundsson et al., 2017]. Fortunately, numerous reconstructions are freely available in databases such as BIGG [King et al., 2016, Norsigian et al., 2020]. The models are encoded and annotated in the standardized SBML format that facilitates their exchange and simulation with the above-mentioned methods on various platforms such as the COBRA toolbox [Becker et al., 2007, Heirendt et al., 2019]. We will use the latter for our analyses of cell metabolism in Chapter 6.

2.1.3 Kinetic models

Beyond metabolic network modelling, stoichiometric models can describe a whole range of different biochemical networks using the form (2.1). Taking into account the interactions between genes, proteins, and RNAs requires one to consider the kinetics of the reactions and their dependence on molecular concentrations and kinetic parameters:

$$\frac{dx(t)}{dt} = Nv(x,p) - \mu\, x, \quad x(0) = x_0, \tag{2.5}$$

with $p \in \mathbb{R}_+^k$ a vector of parameters. The vector of rate laws $v(x,p)$ is a generally nonlinear function of concentrations and parameters.

A variety of mathematical functions exist for kinetic rate laws. The particular choice of a mathematical form depends on the degree of knowledge that we have on the system and the level of precision that we wish. In the absence of precise knowledge, a standard rate law can be chosen. The law of mass action is one example [for a review, see Voit et al., 2015]. It describes the rate v_j of an elementary chemical reaction j (with a single mechanistic step) as being proportional to the product of the concentrations x_i with a given stoichiometry a_i:

$$v_j = k_j \prod_i x_i^{a_i} , \tag{2.6}$$

where k_j is called the rate constant of the reaction. It states that increasing the number of molecules of a given species will increase the probability that they collide. The main advantage of models based on the law of mass action is that they can be determined directly from the elementary reactions and their stoichiometry. Enzymatic reactions are one example. For instance, a free enzyme E catalyses the transformation of a substrate S into a product P by first binding to the free substrate to form an enzyme-substrate complex C. In the example below, the latter can be converted irreversibly into the original enzyme and the product or, if the transformation fails, into the original enzyme and substrate:

$$E + S \underset{k_-}{\overset{k_+}{\rightleftharpoons}} C \overset{k_{cat}}{\rightarrow} E + P , \tag{2.7}$$

where k_+ and k_- are the reaction rate constants and k_{cat} the catalytic constant of the reaction. Applying the mass-action law to the reaction in (2.7), we write the following system of differential equations

$$\begin{array}{rcl} \frac{dx_E}{dt} & = & -k_+ x_E\, x_S + (k_- + k_{cat})\, C , \\ \frac{dx_S}{dt} & = & -k_+ x_E\, x_S + k_-\, C , \\ \frac{dx_C}{dt} & = & k_+ x_E\, x_S - (k_- + k_{cat})\, C , \\ \frac{dx_P}{dt} & = & k_{cat}\, x_C , \end{array} \tag{2.8}$$

and initial conditions: $x_E(0) = x_E^0, x_S(0) = x_S^0, x_C(0) = 0, x_P(0) = 0$. Here it is assumed that no product and complex are present at the start. The total concentration of enzyme and substrate is conserved along the reaction: $x_E^0 = x_E + x_C$ and $x_S^0 = x_S + x_C + x_P$.

A popular model derived from these equations by means of the standard quasi-steady-state approximation (sQSSA) is the Henri-Michaelis-Menten equation [Briggs and Haldane, 1925, Cornish-Bowden, 2015, Michaelis and Menten, 1913]. The sQSSA assumes that the concentration of complex C rapidly equilibrates to its quasi-steady-state value. By solving $dx_C/dt = 0$ in (2.8), and using the mass-conservation relation for the enzyme concentration, we obtain the popular form of the equation describing the rate of accumulation of the product with saturation effects:

$$\frac{dx_P}{dt} = \frac{k_{cat}\, x_E^0\, x_S}{K_m + x_S} , \tag{2.9}$$

where $K_m = \frac{k_- + k_{cat}}{k_+}$ is the Michaelis-Menten constant.

The Michaelis-Menten equation has become a canonical approach to understand enzyme kinetics. It is widely used to estimate k_{cat} and K_m values from product progress curves [Duggleby and Clarke, 1991, Johnson, 2013, Stroberg and Schnell, 2016, Tummler et al., 2014]. The

in-vitro experiments are carried out in conditions of validity of the equation, when the total substrate concentration is much larger than the total enzyme concentration, so that the amount of substrate bound to the enzyme is negligible. This condition, which has been generalized by Segel [1988] and Segel and Slemrod [1989], implies that the quasi-steady-state assumption for the enzyme-substrate complex is valid at a lower enzyme concentration with respect to the sum of the total substrate concentration and K_m value [for review, Schnell and Maini, 2003]:

$$\frac{x_E^0}{K_m + x_S^0} \ll 1\,. \tag{2.10}$$

Validity of the Michaelis-Menten equation is a necessary but not sufficient condition for an accurate estimation of kinetic parameter values [Chen et al., 2010, Choi et al., 2017, Stroberg and Schnell, 2016]. The highly correlated structure and non identifiability of the parameters requires a proper design of dynamic experiments, such as choosing appropriate times for data collection and an initial substrate concentration equal to (as a rule of thumb) two or three times the K_m value [Choi et al., 2017, Duggleby and Clarke, 1991, Stroberg and Schnell, 2016]. The whole difficulty is that a proper assay thus requires a priori knowledge of the K_m value... to determine the K_m value more precisely.

An alternative approximation has been proposed for system (2.8). It overcomes the parameter estimation issues [Choi et al., 2017] and the observation that, *in vivo*, enzyme concentrations are often higher than in enzymatic assays or at least of the same magnitude as their substrate [e.g. Albe et al., 1990]. It is based on the total quasi-steady-state approximation (tQSSA) by replacing the free substrate concentration with its total substrate concentration. The derivation gives a somewhat complex expression [Borghans et al., 1996, Tzafriri, 2003]:

$$\frac{dx_P}{dt} = k_{cat}\,\frac{x_E^0 + K_m + x_S^0 - x_P}{2} - \frac{\sqrt{\left(x_E^0 + K_m + x_S^0 - x_P\right)^2 - 4\,x_E^0\,(x_S^0 - x_P)}}{2}\,. \tag{2.11}$$

From expression (2.11), Tzafriri [2003] developed a simpler approximation called first-order tQSSA. It resembles the sQSSA form of Michaelis-Menten kinetics:

$$\frac{dx_P}{dt} = \frac{k_{cat}\,x_E^0\,x_S}{K_m + x_E^0 + x_S}\,, \tag{2.12}$$

and has the following criteria of validity:

$$x_E^0 + K_m \gg x_S^0 \quad \text{and} \quad K_s \ll K_m\,, \quad \text{or :} \tag{2.13}$$

$$x_E^0 \gg x_S^0 \quad \text{and} \quad x_E^0 \gg K_m \approx Ks\,, \tag{2.14}$$

where K_s is the Van Slyke-Cullen constant defined by $K_s = k_{cat}/k^+$. Compared to sQSSA, the tQSSA approximation extends the parameter domain for which it is reasonable to assume that $dx_C/dt \approx 0$ [Schnell and Maini, 2000, 2003, Tzafriri, 2003]. With this approximation, more accurate and precise estimations can be obtained with a proper experimental design and without prior information on parameter values [Choi et al., 2017]. The tQSSA approximation has been used for modelling different types of biochemical systems [e.g. Ciliberto et al., 2007,

Pedersen et al., 2008b,a, Tzafriri et al., 2002]. We also used it in a study of mRNA degradation kinetics described in Chapter 7.

Various other quasi-steady-state approximations have been proposed, as well as different extensions of those described here [Schnell and Maini, 2003]. In addition, many other expressions allow the modelling of the kinetics of enzymatic reactions, with inhibition, allostery,... or of gene expression with Hill functions for instance. It is beyond the scope of this section to describe these in detail, as there are excellent textbooks on this topic [e.g. Klipp et al., 2016, Segel, 1993, Voit, 2017]. The important message is that the mass-action law, the Michaelis-Menten equation and similar rate laws offer a framework to model a large diversity of biochemical processes, from enzymatic reactions to gene expression. This allows to obtain ODE models for the solution of which numerous numerical methods and software tools are available.

ODE models of biochemical systems are nonlinear, often stiff as they include processes evolving on different time scales, and their size grows quickly as more biochemical processes are considered. As a consequence, they are not amenable to formal mathematical analysis and reliable parameter estimation is a difficult problem due to a generalized lack of data. As we have seen in this section, the problem can be partially solved with a proper experimental design and by using a combination of mass conservation relations and valid quasi-steady-state approximations to reduce the model dimension. In the following section, I will present some personal work related to model approximations of larger biochemical systems and their validity, with two different sets of approaches: model reduction based on time-scale separation and the use of quasi-steady-state approximations, and model approximation by piecewise-linear functions, which is useful when no quantitative data is available.

2.2 Approximation of large kinetic models

Various approximations have been proposed in the literature to reduce the size and complexity of biochemical network models, tailored to typical response functions and time-scale hierarchies found in genetic or metabolic regulation [de Jong and Ropers, 2006b, Heijnen, 2005, Heinrich and Schuster, 1996, Okino and Mavrovouniotis, 1998, Papin et al., 2004, Pecou, 2005, Radulescu et al., 2012, Roussel and Fraser, 2001, Savageau, 2001]. The approximations result in models that are easier to handle mathematically and computationally, while maintaining important properties of the original system. In particular, they reduce the dimension and the number of parameters, and simplify the mathematical form of the equations.

2.2.1 Model reduction based on time-scale separation

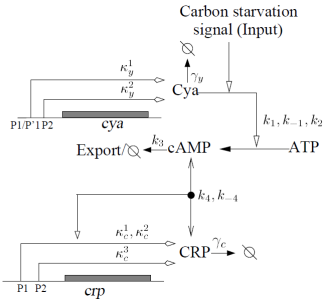
This section is concerned with model-order reduction. The principle is based on the identification of relationships between model variables such that fewer species need to be measured experimentally or simulated. The derivation of the Michaelis-Menten equation is one prototypical example: starting from the mass-action model with four variables, we arrive after reduction at a single ODE describing the time evolution of the product concentration. Product formation can be monitored experimentally, but neither the concentration of the enzyme-substrate complex nor the concentrations of free substrate and free enzyme can be measured. The problem is that many

biochemical systems of interest combine a variety of processes, not only enzymatic reactions, but also gene expression, signalling... The reduction of such large and stiff systems is all but intuitive.

Three general strategies haven been pursued for model-order reduction [Okino and Mavrovouniotis, 1998]: (i) lumping, which transforms the original variables into a vector of lower dimension; (ii) sensitivity analysis neglecting network reactions and species with small impact; and (iii) time-scale analysis identifying the different time scales on which the network species react and considering the fast time-scale reactions and species at steady state. The latter approach is widely used in enzyme kinetics, as we have seen before, and more generally in biochemical network modelling with the usage of quasi-steady-state approximations. The approach has been mathematically formalized by singular perturbation theory [O'Malley, 1991]. We will focus on this approach in what follows. The objective is to obtain models of smaller dimension, with less variables and parameters, while the remaining variables and parameters are easier to measure experimentally. In addition, the parameters must be physiologically meaningful.

Given the ODE model as in (2.5), the first reduction step is to identify the different time scales of the system. This requires *a-priori* knowledge such as the values of the parameters or other information about the biochemical processes. For instance, bacterial pools of metabolites have a turn-over of the order on second, minutes for mRNA pools, and hours for proteins [Shamir et al., 2016]. This is sufficient information in numerous cases. Consider for instance the small biochemical network in Figure 2.1(a). It is part of larger networks that we modelled and studied in [Baldazzi et al., 2010, 2012, Ropers et al., 2006, 2011]. The depletion of a carbon source like glucose is signalled to the cell by the phosphotransferase system. It leads to the activation of the adenylate cyclase Cya and the production of cAMP by the enzyme. This signalling molecule makes a complex with the transcription factor CRP, which binds to promoter regions and activates or inhibits transcription. This small module, which we will later refer to as an activation network, allows the gene expression program to be remodelled so that cells can cope with carbon starvation. In this small network, we identify fast processes, cAMP production and complex formation between CRP and cAMP. The slow processes correspond to the synthesis of CRP and Cya. The corresponding ODE model is shown in Figure 2.1(b). It includes five variables, describing the free concentrations of CRP, Cya, and cAMP, and the concentration of bound Cya and CRP. Rate laws include the Michaelis-Menten equation for cAMP synthesis, the mass-action law for the rate of association and dissociation of CRP·cAMP, while Hill functions are used to represent the regulation of CRP and Cya synthesis by the complex. Its general form is similar to (2.5).

If we are generally able to identify which processes are fast and which ones are slow, difficulties may arise for the separation of variables into fast and slow. We encounter this situation with the ODE model of the activation network shown in Figure 2.1(b). For instance the state equation for the free CRP concentration depends on fast processes, the association and dissociation of the complex CRP·cAMP, and slow processes such as the synthesis of the protein and its degradation. A linear transformation of the variables is needed to uncover the two time scales [Heinrich and Schuster, 1996]. We introduce vectors of slow and fast variables, $x^s \in \mathbb{R}_+^m$ and $x^f \in \mathbb{R}_+^{n-m}$,



(a)

$$
\begin{aligned}
\dot{x}_{y\sim} &= \kappa_y^1 + \kappa_y^2\, h^+(x_{c\sim m}, \theta^2_{c\sim m}, m_{c\sim m}) - \gamma_y\, x_{y\sim} \\
&\quad + (k_{-1} + k_2\, h^+(u_s, \theta_s, m_s))\, x_{y\sim p} - k_1\, x_{y\sim} x_p \\
\dot{x}_{c\sim} &= \kappa_c^1 + \kappa_c^2\, h^+(x_{c\sim m}, \theta^1_{c\sim m}, m_{c\sim m}) - \gamma_c\, x_{c\sim} + k_{-4}\, x_{c\sim m} - k_4\, x_{c\sim}\, x_{m\sim} \\
\dot{x}_{y\sim p} &= k_1\, x_{y\sim}\, x_p - (k_{-1} + k_2\, h^+(u_s, \theta_s, m_s) + \gamma_y)\, x_{y\sim p} \\
\dot{x}_{c\sim m} &= k_4\, x_{c\sim}\, x_{m\sim} - (k_{-4} + \gamma_c)\, x_{c\sim m} \\
\dot{x}_{m\sim} &= k_2\, h^+(u_s, \theta_s, m_s)\, x_{y\sim p} + k_{-4}\, x_{c\sim m} - k_3\, x_{m\sim} - k_4\, x_{c\sim}\, x_{m\sim}
\end{aligned}
$$

(b)

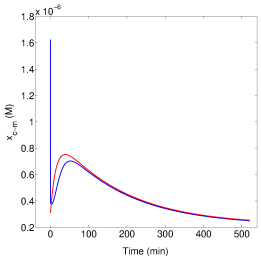
Slow system

$$
\begin{aligned}
\dot{x}_y &= \kappa_y^1 + \kappa_y^2\, h^+(x_{c\sim m}, \theta^2_{c\sim m}, m^2_{c\sim m}) - \gamma_y\, x_y \\
\dot{x}_c &= \kappa_c^1 + \kappa_c^2\, h^-(x_f, \theta_f^2, m_f^2)\; h^+(x_{c\sim m}, \theta^1_{c\sim m}, m^1_{c\sim m}) + \kappa_c^3\, h^-(x_f, \theta_f^1, m_f^1) - \gamma_c\, x_c
\end{aligned}
$$

Fast system

$$
\begin{aligned}
\dot{x}_{y\sim p} &= k_1\,(x_y - x_{y\sim p})\, x_p - (k_{-1} + k_2\, h^+(u_s, \theta_s, m_s) + \gamma_y)\, x_{y\sim p} \\
\dot{x}_{c\sim m} &= k_4\,(x_c - x_{c\sim m})\,(x_m - x_{c\sim m}) - (k_{-4} + \gamma_c)\, x_{c\sim m} \\
\dot{x}_m &= k_2\, h^+(u_s, \theta_s, m_s)\, x_{y\sim p} + (k_3 - \gamma_c)\, x_{c\sim m} - k_3\, x_m
\end{aligned}
$$

(c)



(d)

$$
\begin{aligned}
\dot{x}_y &= \kappa_y^1 + \kappa_y^2 - \gamma_y\, x_y \\
\dot{x}_c &= \kappa_c^1 + \kappa_c^2\, h^+(x_{c\sim m}, \theta^1_{c\sim m}, m_{c\sim m}) - \gamma_c\, x_c \\
x_{c\sim m} &= \frac{h^+(u_s, \theta_s, m_s)\, x_c\, x_y}{K_4\, K_3 + h^+(u_s, \theta_s, m_s)\, x_y} \\
K_3 &= \frac{k_3}{k_2} \\
K_4 &= \frac{k_{-4}}{k_4}
\end{aligned}
$$

(e)

Figure 2.1 – (a) Activation network (from [Ropers et al., 2011]). (b) Detailed ODE model for the activation network. $x_{y\sim}$, $x_{c\sim}$, $x_{y\sim p}$, $x_{c\sim}$, and $x_{m\sim}$ denote the concentrations of free Cya, free CRP, Cya·ATP, CRP·cAMP, and free cAMP, respectively, while u_s denotes the external glucose concentration. The total concentrations of Cya, CRP, and cAMP are referred to as x_y, x_c, and x_m, respectively. h^+ denotes a positive Hill function: $h^+(x, \theta, m) = x^m/(x^m + \theta^m)$. (c) Time-scale separation. (d) Model solutions. The blue curve represents a solution for the concentration variable $x_{c\sim m}$ in the original model. The red curve is the corresponding solution for $x_{c\sim m}$ in the QSS model. After an initial transient the solution of the original model rapidly relaxes to the QSS solution. (e) QSS model for the activation network. The model approximates the original model by coupling the fast variable $x_{c\sim m}$ to the slow variables x_y and x_c. The slow variables are defined as: $x_y = x_{y\sim} + x_{c\sim m}; x_c = x_{c\sim} + x_{c\sim m}; x_m = x_{m\sim} + x_{c\sim m}$.

respectively ($m < n$). These are defined as linear combinations of the original variables x:

$$\begin{bmatrix} x^s \\ x^f \end{bmatrix} = T\,x, \tag{2.15}$$

with $T \in \mathbb{Z}^n \times \mathbb{Z}^n$. This gives for the stoichiometry matrix:

$$\begin{bmatrix} N^s & 0 \\ N^{s'} & N^f \end{bmatrix} = T\,N\,, \tag{2.16}$$

where N^s and N^f are stoichiometry matrices for the slow and fast part, respectively. The matrix T has been chosen to match knowledge on the biochemical reactions involved. The variables $x^s \in \mathbb{R}_+^m$ typically represent total protein concentrations, obtained by summing the concentrations of free and complexed proteins. The variables $x^f \in \mathbb{R}_+^{n-m}$ include protein complexes and metabolites.

The introduction of the fast and slow variables leads to the following reformulation of the system (2.5):

$$\frac{dx^s}{dt} = N^s\,v^s(x^s, x^f), \qquad x^s(0) = x_0^s, \tag{2.17}$$

$$\frac{dx^f}{dt} = N^{s'}\,v^s(x^s, x^f) + N^f\,v^f(x^s, x^f) \approx N^f\,v^f(x^s, x^f), \qquad x^f(0) = x_0^f \tag{2.18}$$

where $v^s(x^s, x^f) \in \mathbb{R}^p$ and $v^f(x^s, x^f) \in \mathbb{R}^{q-p}$ are rate equations for the slow and fast system, respectively. The variable transformation ensures that slow variables change through slow reactions only. In system (2.18), however, we notice that the fast system still includes a slow time scale. This can be seen when we apply the linear transformation to our activation network model, which gives the system in Figure 2.1(c). The consumption rates of the three fast variables by growth dilution and protein degradation (represented by the parameter γ) are on a slow time scale compared to the rate of complex dissociation or export of cAMP. We neglect the slow term, that is, we assume that $N^{s'}\,v^s(x^s, x^f) \ll N^f\,v^f(x^s, x^f)$, as done in system (2.18). The principle was applied to the activation network model in Figure 2.1(b), which gives the slow-fast system in Figure 2.1(c).

Now that we have reformulated our system, we can apply the QSS assumption, based on the hypothesis that the fast variables instantaneously adapt to changes in the slow variables. This amounts to setting $dx^f/dt = 0$ in (2.18). From a mathematical point of view, the QSS approximation restricts the system dynamics to a manifold of lower dimension, an approximation of the so-called slow manifold (see [Heinrich and Schuster, 1996]). After an initial transient, the dynamics of the fast system can be well approximated by an algebraic function of the slow variables: $x^f = g(x^s), g : \mathbb{R}_+^m \to \mathbb{R}_+^{n-m}$. The following QSS model describes the dynamics of the system on the slow manifold:

$$\frac{dx^s}{dt} = N^s\,v^s(x^s, g(x^s)), \qquad x^s(0) = x_0^s, \tag{2.19}$$

The QSS approximation reduces the dimension of the system, but generally also the number of parameters, due to the fact that some of these can be lumped as they no longer independently

occur. The application of the QSS assumption to the model of the example network is shown in Fig. 2.1(d-e). We obtain an algebro-differential model with two state equations and one algebraic equation. The QSS approximation has reduced the total number of parameters from 19 to 13. Numerical simulation using physiologically relevant parameters shows how, after a transient time, the solution of the reduced model converges to the solution of the original one.

I have used this type of QSS approximation in various studies of the carbon starvation response network in *E. coli* [Baldazzi et al., 2010, 2012, Ropers et al., 2006, 2011], as well as transcription-translation [Belgacem et al., 2018], and mRNA degradation [Etienne et al., 2020] (Chapter 7.1). In [Ropers et al., 2011], for instance, we show by numerical studies that a QSS model of a large model of this system is a good approximation of the original model. The reduction allowed to simplify the original model with 14 variables and 63 parameters into a model with 9 variables and 59 parameters. The reduced model no longer includes parameters for which experimental data are generally not available, such as the association and dissociation rate constants, or model variables such as intermediary complexes. In Section 2.2.2, we will discuss how we can further simplify this type of model to qualitatively analyse its dynamics.

2.2.2 Model approximation by means of piecewise-linear functions

How to mathematically model biochemical networks when quantitative data is scarce? Even numerical simulation and analysis of models as simple as the one represented in Figure 2.1(e) can be complicated, due to a generalized lack of physiological values for parameters. Kinetic constants for gene expression like the maximal synthesis rate and the degree of cooperativity in the regulatory mechanisms are typically missing in most cases. Parameter estimation from experimental data could alleviate the problem, but this requires kinetic gene expression data that are also often missing. If no data can be acquired specifically for model estimation, an alternative is to further reduce the model and simplify its mathematical form.

During my postdoctoral research and in later collaborative projects, I focused on one specific type of simplification, the approximation of sigmoidal Hill functions by means of step functions [Glass and Kauffman, 1973, Mestl et al., 1995] (for review, see de Jong and Ropers 2006a,b). From a biological point of view, the use of step functions corresponds to the assumption that gene activity is switched on or off abruptly instead of progressively, when the concentration of the regulatory protein crosses a threshold. An example is shown in Figure 2.2 for two genes, *rrn* and *fis*, the transcription of which is regulated by protein Fis in a cooperative manner [see Ropers et al., 2006, and references therein]. We have modelled the rate of expression of stable RNAs in response to variations of the Fis concentration by a positive Hill function:

$$f_{rrn}(x_{fis}) = \kappa_{rrn}\, h^+(x_{fis}, \theta^1_{fis}, n_1), \text{ with } h^+(x_{fis}, \theta^1_{fis}, n_1) = \frac{x_{fis}^{n_1}}{x_{fis}^{n_1} + (\theta^1_{fis})^{n_1}} \tag{2.20}$$

where κ_{rrn}, n_1, and θ^1_{fis} are constants denoting the synthesis rate of stable RNAs, the Hill number and the dissociation constant of Fis, respectively. The function $f_{rrn}(x_{fis})$ implies that stable RNAs are maximally expressed at a rate κ_{rrn} if $x_{fis} > \theta^1_{fis}$, whereas they are almost not expressed if $x_{fis} < \theta^1_{fis}$. The expression of $f_{rrn}(x_{fis})$ can be simplified by a step function, thus

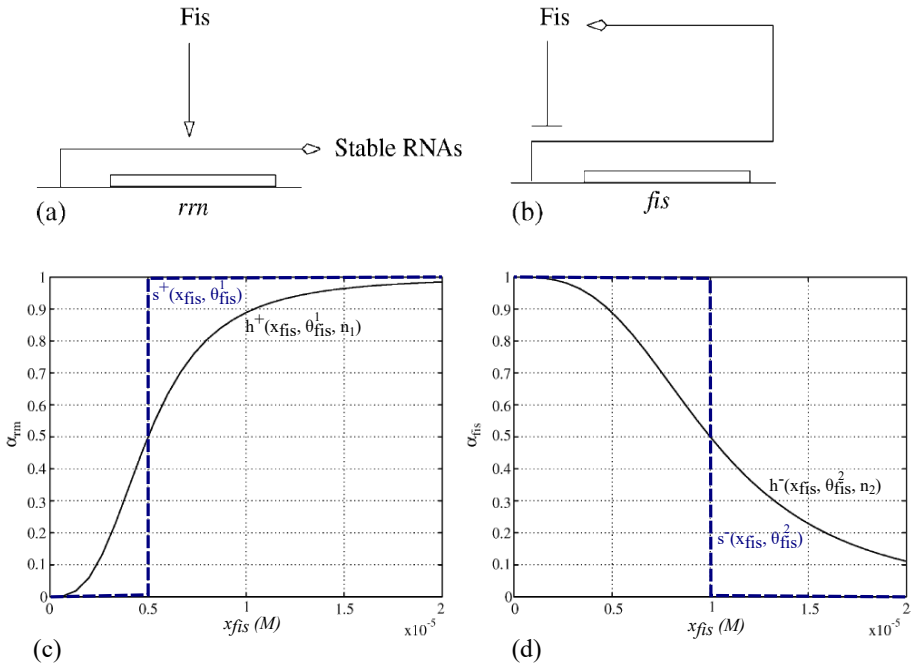


Figure 2.2 – Representation of the regulation of gene expression by means of step functions. Regulation of the expression of genes (a) *rrn* and (b) *fis* by the protein Fis. (c) Activity of the promoter *rrn*P1 (α_{rrn}), normalized on the scale 0 to 1, as a function of the concentration of transcriptional regulator Fis. (d) Idem for the activity of promoter *fis*P α_{fis}. The dotted blue line represents the positive (c) or negative (d) step function used to approximate promoter activities modelled by Hill functions. Adapted from [Ropers et al., 2006].

eliminating the often unknown cooperativity coefficient n_1 (Figure 2.2(c)):

$$f_{rrn}(x_{fis}) \approx \kappa_{rrn}\, s^+(x_{fis}, \theta^1_{fis}), \text{ with } s^+(x,\theta) \begin{cases} 0, if x < \theta\,, \\ 1, if x > \theta\,. \end{cases} \tag{2.21}$$

In the same manner, the auto-inhibition of Fis synthesis can be modelled by a negative Hill function, which we subsequently approximate by a negative step function (Figure 2.2(d)):

$$f_{fis}(x_{fis}) = \kappa_{fis}\, h^-(x_{fis}, \theta^2_{fis}, n_2), \text{ with } h^-(x_{fis}, \theta^2_{fis}, n_2) = \frac{(\theta^2_{fis})^{n_2}}{x^{n_2}_{fis} + (\theta^2_{fis})^{n_2}}\,, \tag{2.22}$$

and

$$f_{fis}(x_{fis}) \approx \kappa_{fis}\, s^-(x_{fis}, \theta^2_{fis}), \text{ with } s^-(x,\theta) \begin{cases} 1, if x < \theta\,, \\ 0, if x > \theta\,. \end{cases} \tag{2.23}$$

The approximation can be extended to more complex systems, as in (2.19), where the dynamics of the fast system at steady state is determined by changes in the slow variables. The fast processes such as complex formation, metabolism, and signalling pathways introduce a coupling between slow variables, represented by a combinatorial expression of sigmoidal functions. For instance, following the application of the quasi-steady-state approximation in Figure 2.1(e), the term for the synthesis rate of protein CRP is a function of the CRP·cAMP concentration $f_{crp} = \kappa^1_c + \kappa^2_c\, h^+(x_{c\sim m}, \theta^1_{c\sim m}, m_{c\sim m})$, which couples the CRP synthesis rate to the CRP and Cya concentrations because the steady-state concentration of the complex

$x_{c\sim m} = h^+(u_s, \theta_s, m_s)\, x_c\, x_y)/(K_4\, K_3 + h^+(u_s, \theta_s, m_s)\, x_y)$ depends on these two variables. The resulting dependency of the CRP synthesis rate is shown normalized on the scale 0 to 1 in Figure 2.3(a).

Two steps are needed to approximate multivariate Hill functions like $h^+(x_{c\sim m}, \theta^1_{c\sim m}, m_{c\sim m})$. In [Ropers et al., 2011], we showed that we can first approximate the multivariate function by a product of Hill functions:

$$h^+(g(x^s), \theta, n) \approx h^+(x^s_1, \theta_1, n_1) \times h^+(x^s_2, \theta_2, n_2) \times ...\text{and:} \quad (2.24)$$

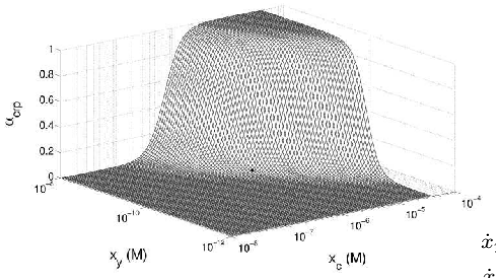
$$h^-(g(x^s), \theta, n) = 1 - h^+(x^s_1, \theta_1, n_1) \times h^+(x^s_2, \theta_2, n_2) \times ... \quad (2.25)$$

We subsequently approximate the Hill functions by step functions:

$$h^+(g(x^s), \theta, n) \approx s^+(x^s_1, \theta_1) \times s^+(x^s_2, \theta_2, n_2) \times ...\text{and:} \quad (2.26)$$

$$h^-(g(x^s), \theta, n) = 1 - s^+(x^s_1, \theta_1, n_1) \times s^+(x^s_2, \theta_2, n_2) \times ... \quad (2.27)$$

Applied to the reduced model of the activation network in Figure 2.1(e), we obtain an approximation for the multivariate function $h^+(x_{c\sim m}, \theta^1_{c\sim m}, m_{c\sim m}) \approx h^+(x_y, \theta_y, m_y)\, h^+(x_c, \theta_c, m_c)\, h^+(u_s, \theta_s, m_s)$, which we further approximate by a product of step functions $h^+(x_{c\sim m}, \theta^1_{c\sim m}, m_{c\sim m}) \approx s^+(x_c, \theta^1_c)\ s^+(x_y, \theta^1_y)\ s^+(u_s, \theta_s)$. This eventually gives the so-called piecewise-linear (PL) model in Figure 2.3(a). The model describes the regulatory logic of CRP expression: the protein is expressed if glucose is depleted ($u_s > \theta_s$), and if there is enough enzyme Cya to produce cAMP ($x_y > \theta_y$) and transcription factor CRP ($x_c > \theta_c$). We showed by numerical studies in [Ropers et al., 2011] that PL models are good approximations of reduced and original models. In the case of the carbon starvation network studied in this paper, the PL approximation reduces the QSS model of 59 parameters to a PL model of 46 parameters [Ropers et al., 2011].



$$\dot{x}_y = \kappa^1_y + \kappa^2_y - \gamma_y\ x_y$$

$$\dot{x}_c = \kappa^1_c + \kappa^2_c\ s^+(x_c, \theta^1_c)\ s^+(x_y, \theta^1_y)\ s^+(u_s, \theta_s) - \gamma_c\ x_c$$

(a) (b)

Figure 2.3 – (a) Approximation of complex sigmoidal functions by a product of Hill functions. The surface represents the plot of $h^+(x_{c\sim m}, \theta_{c\sim m}, m_{c\sim m})$ as a function of x_y and x_c. It is fitted with the product of Hill functions $h^+(x_y, \theta_y, m_y)\, h^+(x_c, \theta_c, m_c)\, h^+(u_s, \theta_s, m_s)$. θ_c and θ_y denote the threshold values determined by curve fitting. (b) PL model for the activation network. From [Ropers et al., 2011].

2.2.3 Discussion and perspectives

In this chapter, I presented work related to the approximation of nonlinear ODE systems, allowing to develop a large variety of biochemical network models, from more detailed to more

abstract models. The various reduced models are rigorously derived, which results in good approximations that preserve the main features of the global system dynamics [Ropers et al., 2011]. Quasi-steady-state and piecewise-linear approximations will be repeatedly used in the various biological problems addressed in the following chapters.

A key step of the method presented in Section 2.2.1 is the ability to solve a fast system at steady state. This is no longer possible when biochemical networks are much larger and include slow variables that are indirectly coupled through numerous intermediate metabolic and signalling processes. This point was addressed in follow-up studies in the context of the post-doctorate of Valentina Baldazzi, with the development and application of a method allowing the inference of PL models from more complex QSS models by means of time-scale arguments and sensitivity criteria from metabolic control analysis [Baldazzi et al., 2010, 2012]. I also worked on time-scale separation in the context of transcription-translation models in collaboration with the team of Jean-Luc Gouzé (Inria Sophia Antipolis - Méditerranée) [Belgacem et al., 2018].

Other related work was carried out in the context of the PhD thesis of Stefano Casagranda, whom I co-supervised with Jean-Luc Gouzé. We developed an approach, Principal Process Analysis, to determine the contribution of each biological process to the output of a dynamical system. Due to the general form of biochemical network models in (2.5), the rates of the processes appear in a linear additive manner in the right-hand side of the ODEs. In [Casagranda et al., 2015, 2018], we introduced new quantities to weigh the influence of each process on the dynamical change of each model variable. Processes that are inactive in some time windows, because they do not influence the dynamics of the system, are removed from the model. This allows the creation of submodels for each time window that only contain the active processes. This procedure leads to the simplification of the system to its core mechanisms. The simplified system can be further studied, to understand the role of each active process in the system dynamics. Formally, the method is not a model reduction approach because it does not preserve mass balances. However, it allows to dissect the complex dynamics of original models through the analysis of simplified versions of these models in given time windows.

There is no firm response to the question as to which model formalism is the most appropriate. A trade-off has to be found between the quantitative and mechanistic information available and the complexity of the problem studied. ODE models of gene expression and metabolism and their reduced QSS models offer a powerful framework if molecular concentrations and kinetic constants are known or can be precisely inferred from experimental data. When these models are too large for a dynamical analysis, or when information on gene expression dynamics, signalling or metabolic pathways is incomplete or missing, PL models are a good alternative. This is the situation faced in the following chapter with the study of the carbon starvation response of the bacterium *E. coli*. The PL model of the gene regulatory network controlling the stress response is analysed by means of qualitative approaches.

Chapter 3

Qualitative analysis of the dynamics of gene regulatory networks

The work described in this chapter represents the main achievement of my post-doctorate in the Helix Bioinformatics group [Ropers et al., 2006]. The piecewise-linear (PL) model developed in this study is the first attempt to connect *E. coli* growth with a network of transcription regulators involved in the adaptation of *E. coli* to its nutritional environment. At the time, there was a lot of information in the literature about these regulators, but expression data was sparse, often obtained in diverse steady-state environmental conditions, and for different strains. Because of the lack of quantitative information, I used a qualitative modelling and simulation approach to analyse the functioning of the network of transcription regulators.

In the following section, I will illustrate this approach with the PL model of a cross-inhibition network. This simple network is part of the carbon starvation network, whose dynamics will be analysed in Section 3.2.

3.1 Qualitative modelling and simulation of piecewise-linear models

An advantage of the use of step functions is that they facilitate the analysis of the qualitative dynamics of the PL models (de Jong et al. 2004; for review, de Jong and Ropers 2006a,b). The threshold values of step function variables partition the phase space into hyper-rectangular regions, in each of which the system behaviour is qualitatively homogeneous. The continuous phase-space dynamics of the system can be discretised into a state transition graph. This graph is composed of states corresponding to regions of the phase space as well as transitions between these states. The state transition graph describes the possible qualitative behaviours of the system and allows to determine attractors and their attainability. Such analysis is difficult to carry out with the corresponding quantitative QSS models [Ropers et al., 2007].

An example of qualitative analysis is shown with a small cross-inhibition network in Figure 3.1(a). The gene *crp*, which we have encountered in previous examples, is involved in a positive feedback loop with gene *fis*, while *fis* expression is also limited by autoinhibition. Panel (b) of the figure shows the corresponding PL differential equation model, in the case of carbon starvation (when CRP has already been activated by the binding of cAMP upon glucose depletion). In the absence of quantitative information, we qualitatively order the concentrations and threshold parameters (panel (c)). For instance, the concentration of protein Fis varies between a minimal and a maximal value denoted 0 and max_{fis}, respectively. When its concentration reaches the threshold θ^1_{fis}, the protein binds to *crp* promoter and inhibits CRP expression. At

higher intracellular levels, when it reaches the threshold concentration θ^2_{fis}, the protein represses its own expression by binding to its promoter region. The dynamics of the system in the phase plane is shown in panel (d). The system possesses three equilibrium points, two stable points – one characterized by a high concentration of Fis and low concentration of CRP, and the other by a low concentration of Fis and high concentration of CRP – and one unstable point. In the upper left region characterized by $0 < x_{fis} < \theta^1_{fis}$ and $\theta_{crp} < x_{crp} < max_{crp}$, the step function $s^-(x_{crp}, \theta_{crp})$ evaluates to 0, while $s^-(x_{fis}, \theta^1_{fis})$ and $s^+(x_{fis}, \theta^2_{fis})$ are equal to 0. The PL model in this specific region simplifies to the following system:

$$\frac{\dot{x}_{fis}}{dt} = -\gamma_{fis}\, x_{fis}\,, \tag{3.1}$$

$$\frac{\dot{x}_{crp}}{dt} = \kappa_{crp} - \gamma_{crp}\, x_{crp}\,. \tag{3.2}$$

Trajectories in this region asymptotically converge to the point $x_{fis} = 0$ and $x_{crp} = \frac{\kappa_{crp}}{\gamma_{crp}}$. To each region of the phase space corresponds a simplified system of linear equations with an associated dynamical behaviour. After discretization of the phase space, we obtain the state transition graph in panel (d), which includes three qualitative equilibrium states, two stable and one unstable one. As can be seen with this simple example, the qualitative modelling approach allows a quick scan of the qualitative dynamics of the system, without numerical information on parameter values [de Jong et al., 2004]. It has been implemented in the computer tool Genetic Network Analyzer [de Jong et al., 2003, Batt et al., 2012], allowing the analysis of large state transition graphs of complicated gene expression network models.

3.2 Qualitative analysis of the carbon starvation response in *E. coli*

A variety of processes are involved in the adaptation of *E. coli* bacteria to a carbon source, from the remodelling of gene expression, metabolism, and DNA topology to the adaptation of the growth rate, and the lack of mechanistic information. Based on literature data, we reconstructed a first network of six genes that are believed to play a key role in the response of the cell to carbon source availability (Figure 3.2(a)). The network includes genes involved in the transduction of the carbon starvation signal (the global regulator of transcription *crp* and the adenylate cyclase *cya*; panels (a) and (b)), metabolism (the global regulator *fis*), cellular growth (the *rrn* genes coding for stable RNAs, needed in high numbers in exponential phase of growth for ribosome biogenesis and protein synthesis), and DNA supercoiling, an important modulator of gene expression (the topoisomerase *topA* and the gyrase *gyrAB*; panels (a) and (c)). The resulting network is of course far from exhaustive and leaves aside other known regulators. However, it allows to analyse how the key genes included function together. The outcome of their interactions is hard to predict because they are involved in many feedback loops, such as the mutual inhibition of *fis* and *crp*, the auto-inhibition of *fis* mediated by Fis itself and by the DNA supercoiling or the auto-inhibition of CRP and Cya. Do these network components and shared interactions allow to reproduce and explain biological observations?

The network was reconstructed through a bottom-up approach, by gathering literature and database knowledge on the different genes. As described in Chapter 2, writing the correspond-

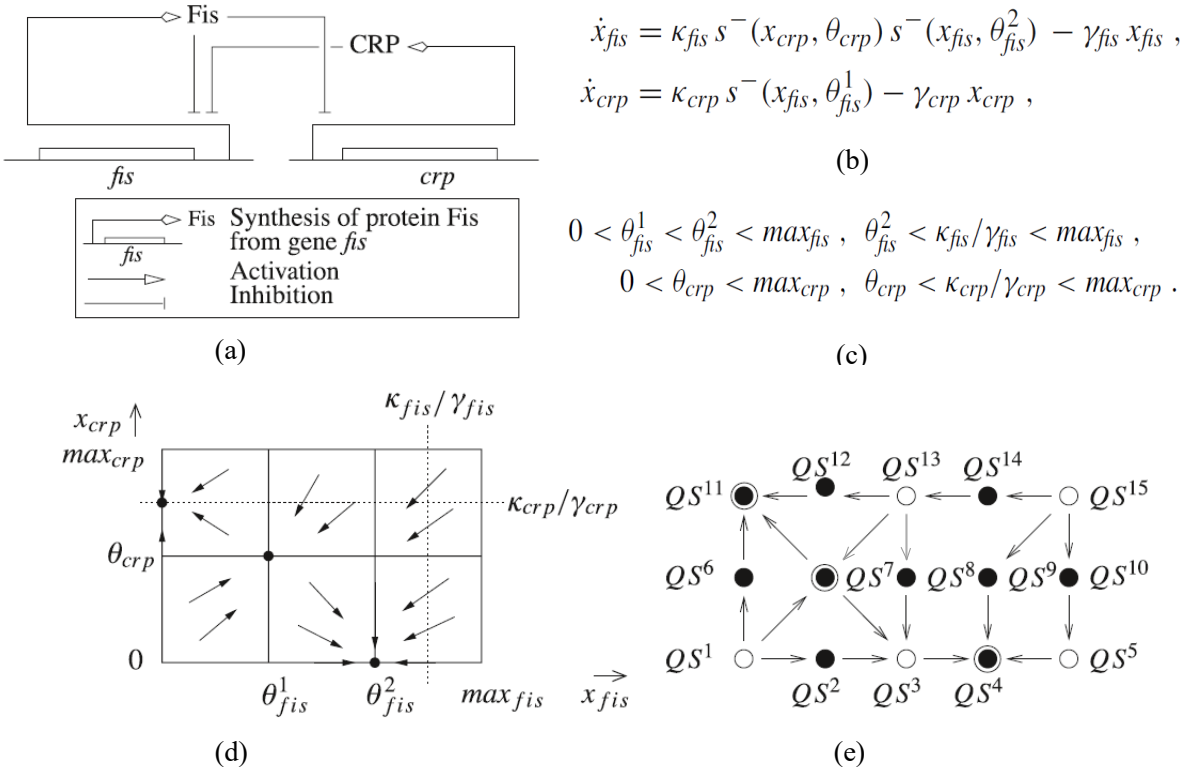
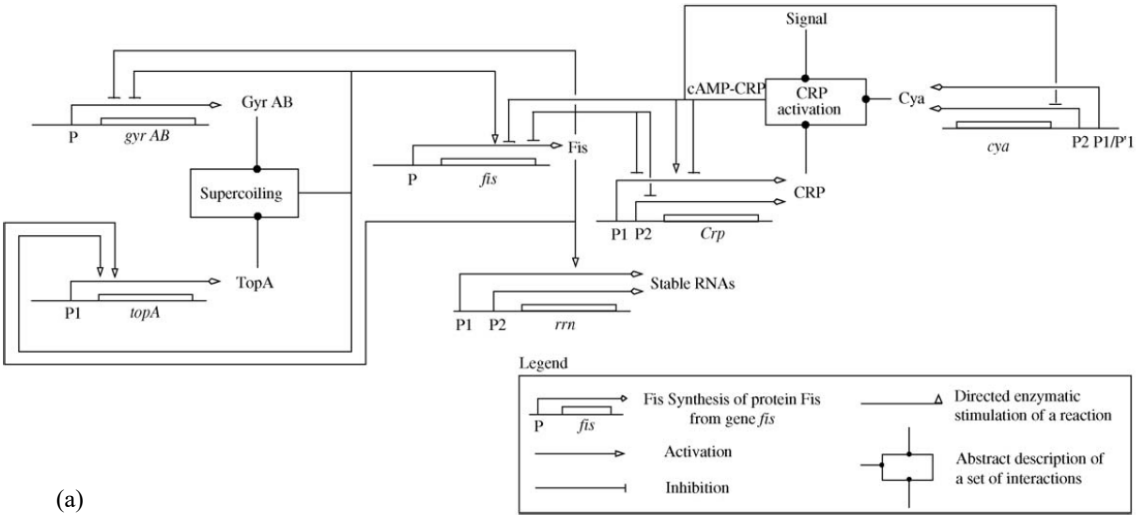
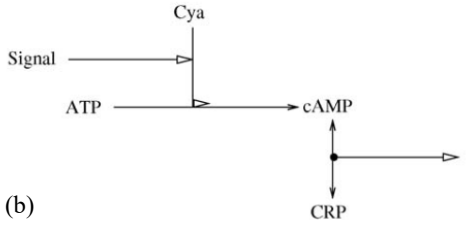
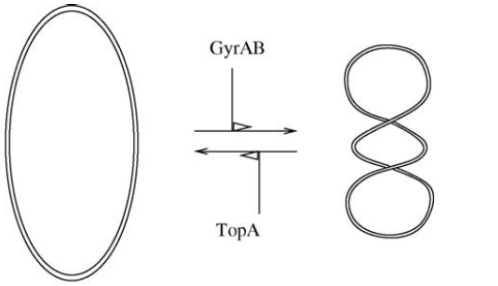


Figure 3.1 – (a) Simple genetic regulatory network composed of the genes *fis* and *crp*. In conditions of carbon starvation, when CRP is activated by cAMP, Fis and CRP inhibit each other's expression. (b) PL differential equation model of the small network. (c) Inequality constraints on threshold and rate parameters associated to PL models. (d) Sketch of the dynamics in each domain of the phase space for the two-gene network. Dots represent the equilibrium points of the system. (e) State transition graph for the two-gene network. QS denotes a qualitative state. The qualitative equilibrium states are circled. Adapted from [Ropers et al., 2007].







(d) $$\dot{x}_{topA} = \kappa^1_{topA} + \kappa^2_{topA}\, s^+(x_{gyrAB}, \theta^3_{gyrAB})\, s^-(x_{topA}, \theta^1_{topA})\, s^+(x_{fis}, \theta^4_{fis}) - \gamma_{topA}\, x_{topA}$$

$$0 < \kappa^1_{topA}/\gamma_{topA} < \theta^1_{topA} < \theta^2_{topA} < \theta^3_{topA} < (\kappa^1_{topA} + \kappa^2_{topA})/\gamma_{topA} < max_{topA}$$

Figure 3.2 – (a) Network of key genes, proteins, and regulatory interactions involved in the carbon starvation response network in *Escherichia coli*. The boxes 'CRP activation' and 'supercoiling' are detailed in Panels (b) and (c). The graphical conventions [Kohn, 2001] are explained in the legend. (d) PL differential equation and parameter inequality constraints for the topoisomerase TopA. From [Ropers et al., 2006].

ing PL model required the prior development of a kinetic model and its reduction based on time-scale separation to obtain a QSS model describing the coupling of slow variables – Fis, CRP, GyrAB, TopA, stable RNAs – with the fast processes – DNA supercoiling and activation of CRP by the phosphotransferase system. The coupling is schematically represented in Figure 3.2(b,c). We eventually obtained a PL model of seven variables, one concentration variable for each of the six proteins and one input variable representing the presence or absence of the carbon starvation signal. Seven differential equations, one for each variable, and forty inequality constraints describe the dynamics of the system. As an illustration, the differential equation and the parameter inequality constraints for the state variable x_{topA} are given in Figure 3.2(d). For instance, the constraint $0 < \kappa^1_{topA}/\gamma_{topA} < \theta^1_{topA}$ expresses that without stimulation of the *topA* promoter, the TopA concentration decreases towards a background level, below the threshold θ^1_{topA}.

The PL model is able to predict growth arrest and the entry into stationary phase following carbon depletion, as well as growth resumption to exponential growth when glucose is available again [Ropers et al., 2006]. Two state transition graphs are obtained for each of these simulation scenarios. One include all possible qualitative behaviours from the initial state representative of the exponential growth conditions to the equilibrium state corresponding to stationary phase conditions and the opposite for the other graph. As an illustration, a representative path along the graph for carbon starvation response is shown in Figure 3.3. It reproduces experimental observations from the literature, such as the decrease of protein Fis concentration in response to carbon depletion and growth arrest due to the reduced production of stable RNAs. The chain of molecular events can be deduced from the prediction and the network structure. The lack of glucose activating CRP, the protein represses Fis synthesis. The decrease of the protein concentration alleviates the inhibition of CRP synthesis. The signal of carbon starvation is thus amplified by the mutual inhibition module involving *fis* and *crp*. The reduction of Fis levels affects the DNA supercoiling module as well and cell growth is halted by the arrest of stable RNA production. Predictions of growth resumption when glucose is added were more surprising, as damped oscillations of Fis, stable RNA, and GyrAB concentrations were predicted to occur [Ropers et al., 2006]. This led to launching experiments in the group to verify the reality of this behaviour. I will come back to this in Chapter 5.

3.3 Discussion and perspectives

The PL model developed in this study is the first attempt to connect *E. coli* growth with a network of transcriptional regulators based on the literature information. The qualitative analysis of the PL model allowed to bypass the lack of quantitative information and understand the cascades of molecular processes involved in the stress response [Ropers et al., 2006]. This work also allowed to point at inconsistencies in the literature concerning the regulation of transcription factors. We will come back to this point in Chapter 5, where the global control of gene expression will be shown to be an important regulatory mechanism of the environmental adaptation of *E. coli*.

Although PL models are reduced and approximated versions of larger nonlinear ODE models, they can be complex and generate large state-transition graphs including many

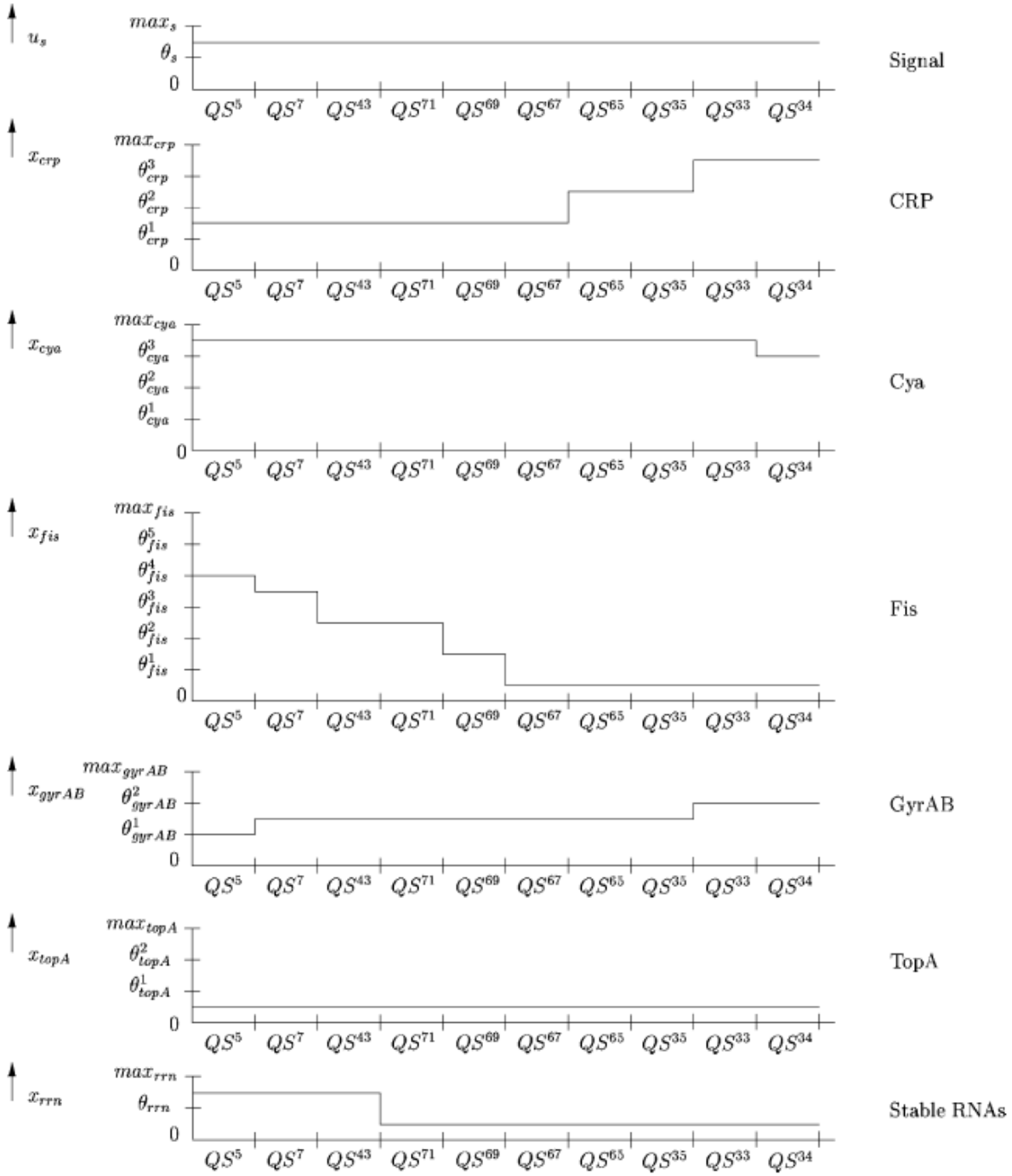


Figure 3.3 – Temporal evolution of the protein and stable RNA concentrations in a typical qualitative behaviour in the state transition graph, when Signal is present ($\theta_s < u_s \leq max_s$). The behaviour represents the molecular events accompanying the transition from exponential to stationary phase following carbon starvation. From [Ropers et al., 2006].

alternative qualitative behaviours. This makes it impossible to analyse the system dynamics by manual inspection and to compare model predictions to available experimental data. Ideally, one would like to query if the graph is consistent with our understanding or with the partial and heterogeneous information from the literature. For instance, is there a path in the graph reproducing the experimental observation that Fis concentration decreases when there is a lack of glucose? And is there a concomitant increase of the CRP concentration? This type of problem was addressed in subsequent work by Grégory Batt and Pedro Monteiro during their PhD theses in our group. The qualitative modelling approach was reformulated to allow the application of formal verification tools based on model checking, together with the development of an approach for the automatic translation of frequent biological questions into the temporal logic formulas used in model checking. This facilitates the application of model-checking approaches for the analysis of state transition graphs by non-expert users. I participated in the development of these queries and the application of model-checking to analyse the PL model of the network in Figure 3.2(a) [Batt et al., 2005] and an extension of this network that I derived [Monteiro et al., 2008].

The models that we developed were reconstructed and manually parametrised from the literature information. I collaborated on several other projects in this manner, for the piecewise-linear modelling of the transcriptional regulatory network controlling the response of the model eukaryote *Saccharomyces cerevisiae* to the agricultural fungicide mancozeb [Monteiro et al., 2011], or the network involved in the adaptation of food pathogens to osmotic stress [Ropers and Métris, 2016, Métris et al., 2017]. Alternative approaches exist though, such as model inference from reporter gene data as done in our group in the framework of the PhD thesis of Diana Stefan [Stefan et al., 2015]. Note that the PL models can be further approximated by logical models [Abou-Jaoudé et al., 2016, de Jong, 2002, Kauffman, 1969, Thomas and d'Ari, 1990]. This opens the door to other advanced tools and methods for the analysis of discrete models. In [Corblin et al., 2009], for instance, I collaborated with the group of Laurent Trilling (Univ. Grenoble Alpes) on the reconstruction of discrete versions of the model in [Ropers et al., 2006] that are consistent with biological data expressed as constraints.

One may argue that not everyone masters model reduction into QSS models and their approximation into PL models whose dynamics can be qualitatively analysed. Based on my experience, I advised the (former) company Genostar in a follow-up project. The company was involved in the commercialization and distribution of a bioinformatics platform including Genetic Network Analyzer. The project led to the development of an approach facilitating the rigorous building of PL models able to address the problem of coupling slow variables through fast processes, which was implemented as a graphical tool in subsequent releases of Genetic Network Analyzer [Batt et al., 2012]. This approach facilitates PL model building by less informed modellers.

The work described in this chapter was motivated by the lack of quantitative information to parametrize the PL model and verify its predictions. This led Hans Geiselmann and his group to mount an experimental program, in which I was involved as well, for the development of strains and plasmids expressing reporter genes that could be used for monitoring gene expression in

various strains and conditions. This work was in line with the general trend of biology becoming increasingly quantitative. As a consequence of these developments, there was an interest in trying alternative approaches based on the analysis of quantitative ODE models or their reduced versions. I will describe this work in Chapters 6 to 7. The availability of new experimental approaches with new types of data opens new methodological questions as well, such as the robust inference of biological quantities from the primary data than can be used for the analysis and validation of quantitative ODE models. These methodological questions will be the subject of the following chapter.

Chapter 4

Analysis of dynamical gene expression and metabolomics data

In the course of the years, biology has moved from experimental approaches based on qualitative or semi-quantitative experiments (e.g., western-blots) to more quantitative ones. The most recent ones include transcriptomics and proteomics data for instance, measurements of gene expression along growth by means of reporter genes, or dynamical metabolomics measurements. The two later approaches prove essential in a number of modelling studies described in the following chapters. Most of the time, the time-series data need to be processed in order to estimate the relevant biological quantities. For purposes such as model validation against experimental data, parameter estimation, or model-based analysis of the data, it is important that these estimations be unbiased and robust to experimental noise, while remaining able to capture precisely rapid variations of the signals, which commonly occur during growth-phase transitions.

In this chapter, I will present work on the use of regularized linear inversion methods for the reconstruction of interesting quantities from metabolomics (Section 4.1) or gene expression (Section 4.2) data. I chose to highlight these papers because of their relevance for many laboratories. Indeed time-series data accumulate in these places because they are becoming easier to acquire, but their information content is far from exploited due to the difficulties to analyse them. The method for gene expression analysis was developed by Valentin Zulkower as part of his PhD thesis, which I co-advised with Hidde de Jong and Hans Geiselmann [Zulkower et al., 2015]. The method for metabolomics data analysis was developed by Eugenio Cinquemani. For this study, I brought the problem, analysed the data, and coordinated the project [Cinquemani et al., 2017]. The method and results were presented at the joint 2017 ISMB-ECCB conference, and published in the corresponding special issue of *Bioinformatics* [Cinquemani et al., 2017].

4.1 Estimation of time-varying growth, uptake and secretion rates from dynamic metabolomics data

The accumulation of extracellular metabolites or their disappearance from the growth medium provides interesting information about intracellular physiology [Kell et al., 2005]. The time profiles are used to compute uptake and secretion rates that can be related to intracellular fluxes in flux balance and metabolic flux analyses, as will be shown in Chapter 6. Despite its apparent simplicity, the problem of estimating time-varying uptake and secretion rates from measurements of extracellular metabolites is challenging. First the available data are noisy and sparse, despite continuous progress in metabolomics methods. Second, the time-course

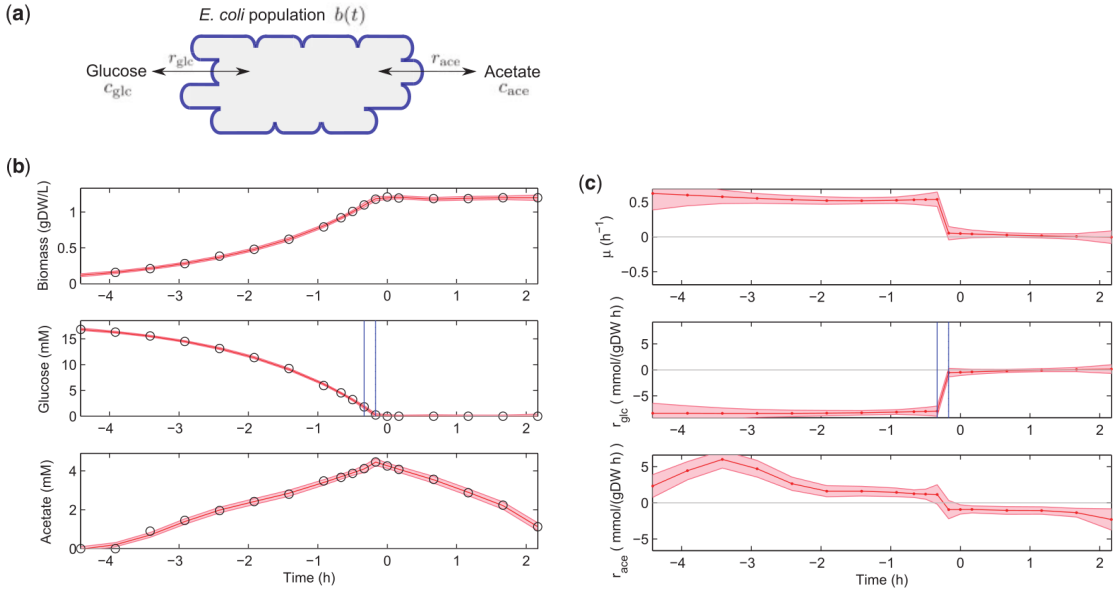


Figure 4.1 – Reconstruction of growth and metabolite exchange rates from *E.coli* diauxic growth experiments with the method in [Cinquemani et al., 2017]. (a) Model scheme defining variables and rates. (b) Data (circles) from [Morin et al., 2016] and EKS estimates with 95% credibility intervals (solid curve and shaded band) of biomass and concentration profiles at all times. (c) EKS rate estimates and credibility intervals (red curve and shaded band). Vertical blue lines show the periods of fast transitions detected in data-preprocessing).

profiles of different extracellular metabolites are strongly correlated. Uptake and secretion rates are indeed proportional to the size of the growing population of cells consuming or producing the metabolites [Stephanopoulos et al., 1998]. Third, the profiles are subject to discontinuities, resulting from sudden changes in the functioning of metabolism. Transition from growth on glucose to acetate is one example, see Figure 4.1(b) for an example (data from Chapter 6; [Morin et al., 2016]). The acetate suddenly stops accumulating in the growth medium upon glucose depletion and is right after consumed by cells as they resume their growth. Such sharp transition of metabolic regime is difficult to account for. The consequence is that exometabolomics data are often under-exploited, in the sense that in many studies the time-varying rates of uptake and secretion are not computed to only focus on the steady-state regime for which data analysis is easier. More advanced approaches are needed to analyse the data.

ODE models of microbial growth are traditionally used to estimate exchange/growth rates from metabolomics and growth data. The general form of the model describes the dynamics of biomass concentration $b(t)$ and concentration of metabolite i at time t $c_i(t)$ [de Jong et al., 2017a]:

$$\begin{aligned} \dot{b}(t) &= \mu(t)\, b(t), \\ \dot{c}_i(t) &= r_i(t)\, b(t), \quad i = 1, \ldots, n. \end{aligned} \tag{4.1}$$

The model is based on the assumption that changes in concentrations $c_i(t)$ are due to the growth, uptake and secretion rates, omitting the negligible degradation rate of extracellular metabolites, inflow and outflow of the medium in the bioreactor. Growth rate $\mu(t)$ and exchange

rates $r_i(t)$ must be reconstructed over a whole experimental period $t \in \mathscr{T}$ from noisy measurements of b and of c_i taken at possibly different time instants. Traditionally, this is achieved by differentiating a fit of the corresponding measurements and computing the estimated rates by means of Equation 4.2:

$$\begin{aligned} \mu(t) &= \frac{\dot{b}(t)}{b(t)}, \\ r_i(t) &= \frac{\dot{c}_i(t)}{b(t)}. \end{aligned} \tag{4.2}$$

Such approach is said indirect, as the data are first smoothed before the quantities of interest are reconstructed via the model in Equation 4.2. This results in the propagation of estimation errors that cannot be controlled. While spline fitting of the data is often the preferred smoothing method [Wahba, 1990], the optimal placement of spline knots is a difficult problem, in particular when there are sudden changes of regime, and the quality of the estimate needs to be assessed in further processing, typically through bootstrap approaches. Finally, the approach does not account for the coupling of the different model equations through the biomass, while the exploitation of this important information could improve the rate estimation procedure. Instead each rate is estimated independently from the others.

We addressed the above issues in [Cinquemani et al., 2017]. Our (direct) approach is based on the use of the explicit model in Equation 4.1, together with an appropriate statistical approach for the estimation of the exchange rates. We considered the reconstruction of rates from measured concentrations and biomass as an input estimation problem, which can be solved using linear inversion methods. That is, the growth, uptake and secretion rates are the input ($u(t) = [\mu(t)\ r_1(t)\ \cdots\ r_n(t)]^T$), and the biomass and metabolite concentrations the output ($x(t) = [b(t)\ c_1(t)\ \cdots\ c_n(t)]^T$) of the linear system $\frac{dx(t)}{dt} = u(t)\,x(t)$, so that rates can be estimated by linear inversion from the noisy biomass and metabolite concentration measurements $y(t)$ modelled as: $y(t) = x(t) + e(t), t \in \mathscr{T}$, with $e(t)$ a Gaussian measurement noise. The input profile $u(t)$ is assumed to be piecewise continuous, so that the solution of the ODE system is well determined, but not necessarily smooth.

As it stands, the problem of estimating uptake/secretion and growth rates is ill posed. There are an infinity of solutions that explain the data for a given set of initial conditions, some of which are not realistic from a biological point of view. For instance, irregular ("wiggly") profiles may fit slowly changing measurements of the biomass or the metabolite concentrations. The problem must be regularized to obtain a unique, acceptable solution [Wahba, 1990]. While it avoids overfitting the data, regularization is challenging because the data are characterized by a combination of slow and fast variations of metabolite concentrations and biomass. In [Cinquemani et al., 2017], we solved this by formulating a Bayesian regularized estimation problem. Each unknown rate profile $u_i(t)$ with $i = 1, ..., n+1$ is modelled as the outcome of a random Gaussian process $\frac{dv_i(t)}{dt} = \gamma(t)\,\omega_i(t)$ and $\frac{du_i(t)}{dt} = v_i(t)$, where w_i is the standard white Gaussian noise and $\gamma_i(t)$ the regularization factor. $u_i(t)$ is hence modelled as a double-integral of white noise, which implies it being continuously differentiable with variability (the probability distribution of its derivative) determined by the magnitude of $\gamma_i(t) > 0$. $\gamma_i(t)$ is a function of time, where larger values of the factor around a time point allow for rapid changes of $u_i(t)$ around that time. By this modelling, the problem of estimating u_i at any time t becomes that of computing the best estimates in a Bayesian perspective, provided the measurement model for the set of all data $\mathscr{Y} = \mathscr{Y}_1, ...\mathscr{Y}_{n+1}$

with $\mathscr{Y}_i = y_i(t) : t \in \mathscr{T}_i$, to calculate the a posteriori expectation

$$\mathbb{E}[\big(r_1(t), \ldots, r_n(t), \mu(t)\big)|\mathscr{Y}] \tag{4.3}$$

at all times of interest $t \in \mathscr{T}$. This approach favours smooth solutions. Two data preprocessing steps are carried out separately to obtain Bayesian priors for the smoothing factors for the slow and fast dynamics, and the detection of regions where fast dynamics take place. The first one is based on a cubic spline interpolation of the data and generalized cross validation, separately for every i, while the second detects times at which concentrations drop to zero, a fact generally associated with a change of metabolic regime. The regularization estimation problem is solved by a dynamical smoothing approach based on an Extended Kalman Smoother (EKS) [Kailath et al., 2000]. It allows to obtain a smoothed estimate for both the rates and states (concentrations and biomass) at time t, along with credible intervals, given past and future measurements. In practice, the Kalman filter - in its extended version adapted for non-linear systems - allows in a first step to estimate states and rates at time t from the knowledge of past measurements from 0 to t and updates the solution given the measurement at $t + 1$. In a second step, the smoothing allows to estimate states and rates on the interval $[0, T]$, given measurements $y(0 : T)$. Smoothed estimates are more accurate than the filtered ones because more data are used [Cinquemani et al., 2017].

We tested the approach on simulated data and real data: those obtained from the batch cultures of wild-type *E. coli* obtained during diauxic growth on glucose and acetate (Chapter 6; [Morin et al., 2016]), as well as fed-batch culture data of *L. lactis* obtained at TBI [Cinquemani et al., 2017]. Results obtained with *E. coli* are shown in Figure 4.1. The method is able to capture the abrupt regime changes, while providing us with stable growth and exchange rates during periods of slow dynamics (either during growth on glucose before time 0 of glucose exhaustion or during growth on acetate, after time 0). We validated our approach by comparing metabolic flux analysis results obtained either with exchange/growth rates obtained from smoothing splines (those used in the data preprocessing step) or the EKS approach. The smoothing spline estimates yield distributions of intracellular fluxes that are much less precise and non-intuitive. On the contrary, the EKS approach allows to make accurate predictions of intracellular fluxes when compared to literature data. This shows the importance of precise rate estimation for metabolic flux analysis using the EKS approach. I used the approach for this precise reason in further studies of *E. coli* metabolism, such as the one on the role of glycogen in metabolic adaptation of *E. coli* described in Section 6.2.

4.2 Estimation of promoter activities and protein concentration profiles from reporter gene data

Reporter gene experiments, which enable to monitor gene expression at high time resolution in a non-intrusive way, are commonly used for the study of gene activities in response to environmental changes, such as in Chapter 5. In these experiments, a gene coding for a fluorescent (or luminescent protein) is designed to have the same promoter as a natural gene of the bacteria, and therefore be driven by the same regulations. A bacterial strain carrying this synthetic gene, either on the chromosome or on plasmid, is grown in a microplate or a

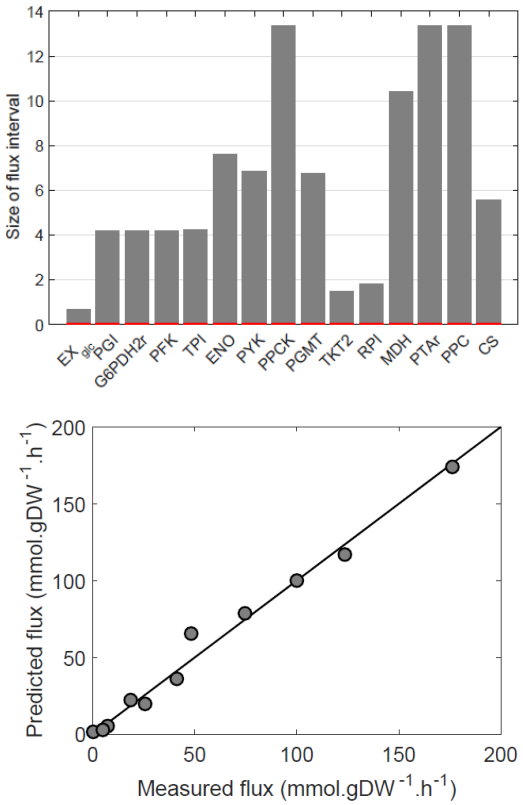


Figure 4.2 – Comparison of metabolic flux analysis results using rate estimates obtained with the EKS method in [Cinquemani et al., 2017] or smoothing splines. (a) Intervals of fluxes in the optimal solutions for 15 selected reactions in the carbon central metabolism obtained with the smoothing spline (grey) and EKS (red) estimates. (b) Comparison of predicted and measured intracellular fluxes for eleven reactions in the carbon central metabolism (EX_{glc}, PGI, G6PDH2r, PGMT, GAPD, PYK, PDH, PPC, MDH, TKT1, and TKT2). Fluxes are given relatively to the specific glucose consumption rate.

bioreactor with frequent measurements of fluorescence (or luminescence) and absorbance, which are then processed using models of gene expression to estimate the temporal profiles of the population's growth rate, promoter activity (transcription rate) and intracellular concentration of the protein encoded by the gene of interest. The relation between promoter activity and the observed fluorescence and absorbance signals is indirect. Models of gene expression are thus needed to infer the quantities of interest. In past studies, we addressed the problem of estimating promoter activities through an indirect approach, where we first smoothed the primary data using cubic splines and then injected the resulting approximating functions into a gene expression model (see [de Jong et al., 2010] for the method and [Boyer et al., 2010] for its software implementation). As pointed out in Section 4.1, however, this leads to estimation errors that are difficult to control. In addition, and in the same way that sharp metabolic shifts are observed during growth transitions, gene expression abruptly changes as well, which is difficult to capture with smoothing splines. Here also, the problem of reconstructing interesting quantities from reporter gene signals requires advanced methods. In a subsequent study [Zulkower et al., 2015], we showed how the analysis of reporter gene data could be also formulated as an input estimation problem and be solved by means of regularized linear inversion.

I briefly illustrate the principle of the method with a simple ODE model describing reporter gene expression in one step:

$$\frac{dR(t)}{dt} = a(t)\,V(t) - \gamma_r\,R(t), \tag{4.4}$$

where $R(t)$ [mmol] is the time-varying quantity of the reporter protein in the growing bacterial population and $a(t)$ [mmol min^{-1} L^{-1}] the synthesis rate of the reporter protein per unit of population volume $V(t)$ [L]. The latter is also called gene activity or promoter activity in the literature [Ronen et al., 2002], assuming a proportionality between transcription and translation rates. The fact that we consider total amount of molecules instead of concentrations as done usually allows to drop the term of growth dilution from the equation. γ_r is the degradation constant of the reporter [min^{-1}] and can be easily measured [de Jong et al., 2010]. More detailed models can be used, for instance to describe the synthesis of the reporter mRNA, as well as the immature and mature forms of the reporter protein [Zulkower et al., 2015]. This is appropriate in situations where the fluorescent protein has a relatively long maturation time or the reporter mRNA is stable, for instance.

Absorbance and fluorescence measurements carried out in microplate readers are generally assumed to be proportional to the volume of the growing bacterial population V and the total amount R of reporter protein in the population, respectively. The following measurement model thus relates the absorbance $\tilde{V}$ and fluorescence $\tilde{R}$ measurements at time t_i to the volume and reporter protein quantities:

$$\begin{aligned} \tilde{V}(t_i) &= \alpha\,V(t_i) + \nu_i, \\ \tilde{R}(t_i) &= \beta\,R(t_i) + \nu'_i, \end{aligned} \tag{4.5}$$

where ν_i, ν'_i represent the measurement noise and α, β are unknown proportionality coefficients. The concentration of the reporter protein can be simply obtained from the ratio $\tilde{R}(t_i)/\tilde{V}(t_i)$.

We solve a first estimation problem, by inferring the growth rate of the population from the

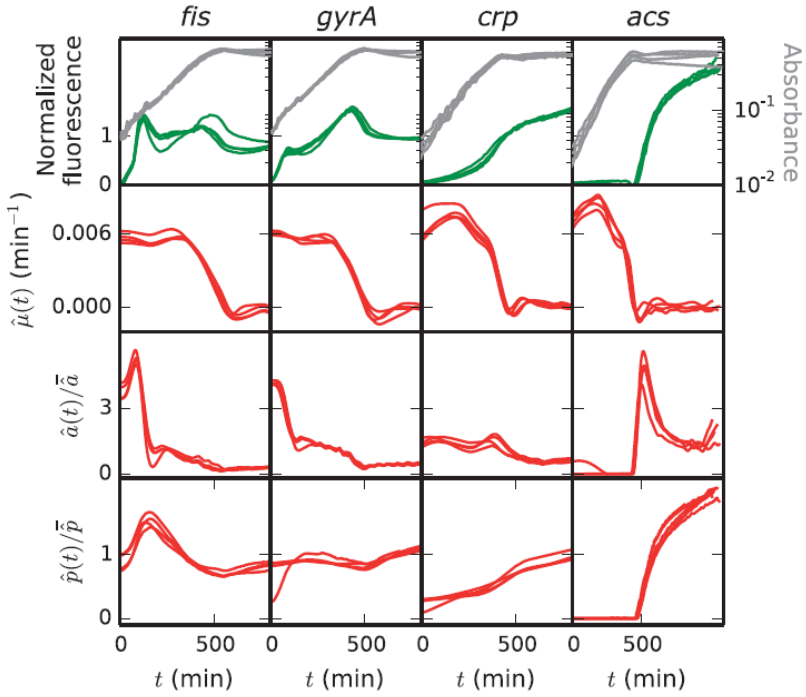


Figure 4.3 – Fluorescence and absorbance data obtained from reporter gene experiments in *E. coli* and estimations of growth rate, promoter activity and protein concentration from these data. The measured fluorescence and absorbance signals are shown in the top row. The estimations of growth rate, promoter activity and protein concentration are denoted by $\hat{\mu}(t)$, $\hat{a}(t)$, and $\hat{p}(t)$, respectively. The fluorescence signal,$\hat{a}(t)$, and $\hat{p}(t)$ have been divided by their mean as they have different orders of magnitude for *E. coli* genes *fis*, *gyrA*, *crp* and *acs*. For each signal, four replicates are shown, corresponding to different wells of the microplate.

absorbance measurements via the following growth equation:

$$\frac{d(\alpha V)(t)}{dt} = \alpha\, V(t)\, \mu(t) \simeq \bar{V}(t)\, \mu(t), \tag{4.6}$$

where $\mu(t)$ [min^{-1}] represents the growth rate and $\bar{V}(t)$ an interpolated version of the measurements $\tilde{V}(t)$. As it is posed, the estimation of growth rate $\mu(t)$ and initial conditions $(\alpha V)(t_0)$ is a linear estimation problem, where $\mu(t)$ is the input (assumed to be piecewise constant) and $\tilde{V}(t)$ the observed output. Similarly to the analysis of metabolomics data, the problem is ill-posed and regularization is needed to solve the linear inversion problem. In the present case, we imposed a Tikhonov regularization on the first derivative of the growth rate [Zulkower et al., 2015], penalizing rapid successive variations, and we set the regularization parameter by generalized cross validation. We illustrate the results obtained in Figure 4.3 with data obtained from *E. coli* cells growing on glucose. Cells were transformed with plasmids carrying a transcriptional fusion of *gfp* with promoters of genes involved in bacterial adaptation to carbon source availability. The estimated growth rates are shown on the second row of Figure 4.3 for four replicates. As can be seen, the growth rate drops from its maximal value to zero upon glucose exhaustion. The estimation is thus able to capture the sudden arrest of bacterial growth.

Along with an estimated growth rate, the linear inversion gives us the estimated volume of the growing cell population $\widehat{\alpha V(t)}$, which we use in replacement of the cell volume in Equation 4.4

to solve the second estimation problem:

$$\frac{dR(t)}{dt} = \frac{a(t)}{\alpha}\,\widehat{\alpha\,V(t)} - \gamma_r\,R(t). \tag{4.7}$$

This is again a linear, time-varying system with input $a(t)$ and output R(t), which can be solved by means of the same linear inversion approach as above. The reconstructed promoter activities for the four reporter genes in *E. coli* are shown in the third row of Figure 4.3. The method correctly infers the known fast changes in gene expression from the data, such as induction of *fis* following a nutrient upshift [Ali Azam et al., 1999] and *acs* upon glucose exhaustion [Berthoumieux et al., 2013, Wolfe, 2005], while avoiding over-fitting outside the transition region.

The host protein and its reporter counterpart share the same promoter and control sequences, but do not necessarily have the same degradation rate. Based on the knowledge of the degradation rate constant, it is also possible to reconstruct the concentration of the protein of interest expressed from the natural gene on the chromosome. To that aim, we develop the following model for the expression of the host protein:

$$\begin{aligned} \frac{dP(t)}{dt} &= \frac{a(t)}{\alpha}\,\widehat{\alpha\,V}(t) - \gamma_p\,P(t) \\ p(t) &= \alpha\,\frac{\alpha\,P(t)}{\widehat{\alpha\,V}(t)} \end{aligned} \tag{4.8}$$

where $P(t)$ [mmol] is the time-varying quantity of the host protein, $p(t)$ [mmol L^{-1}] its concentration, and γ_p [h^{-1}] its degradation constant. We assume γ_p to be approximately known, bearing in mind that most bacterial proteins are stable, with half-lives well over 10 h [Larrabee et al., 1980]. If the reporter protein is stable as well, the default choice of $\gamma_p = \gamma_r$ usually leads to good results, in the sense of returning a protein concentration that is a smoothed version of the reporter concentration. We have again a linear system in Equations 4.7-4.8, where the equations for the reporter and host protein quantity are coupled by a shared input $a(t)$. The linear inversion procedure was adapted for linear equations with a shared input in [Zulkower et al., 2015], resulting in an estimate of $p(t)$ from absorbance and fluorescence measurements. The estimated protein concentrations for the *E. coli* reporter genes are shown in the bottom row of Figure 4.3. The degradation constant of Fis was measured in a former study ($d_p = 0.0065$ min^{-1}; [de Jong et al., 2010]), whereas the other proteins were assumed long-lived ($d_p = 0.001$ min^{-1}) like most *E. coli* proteins [Larrabee et al., 1980]. For instance, the estimated profile of Fis reproduces correctly the known transient accumulation of the protein following a nutrient upshift [Ali Azam et al., 1999, de Jong et al., 2010], which is consistent with its key role in activating the synthesis of stable RNAs needed for growth [Dennis et al., 2004].

More generally we tested the overall approach on simulated data. The results show that it allows a robust reconstruction of promoter activities, growth rates and protein concentrations from fluorescence and absorbance signals [Zulkower et al., 2015].

4.3 Discussion and perspectives

In this chapter, we have seen through two case studies how regularized linear inversion methods allow the reconstruction of interesting quantities from metabolomics or gene expression data. The two methods are able to capture the abrupt changes of metabolism and gene expression

during growth transition, and to avoid over-fitting the data when the system dynamics is slow. They are also able to deal with low signal-to-noise ratios such as the weakness of absorbance signal at the beginning of the experiment. Two different regularization methods haven been chosen: a stochastic approach based on a Bayesian framework for the metabolomics data and a deterministic approach using Tikhonov regularization for the reporter gene data. The approaches can be extended to retrieve time-varying profiles in a much wider range of problems, such as single-cell measurements of fluorescence signals. A necessary condition for their application is that the data should depend linearly on the signal to be estimated.

An important aspect of these approaches is that they come with a software implementation [Cinquemani et al., 2017, Martin et al., 2019, Zulkower et al., 2015], which makes them accessible to a larger audience. Eugenio Cinquemani and I are considering the development of new versions of the software for the analysis of metabolomics data. We envision to enhance the robustness of the method to noise in data at the beginning of the growth kinetics, and to include a graphical user interface for use by non-experts. This will be the object of an Inria technological development project proposal to be submitted in the near future.

Data analysed with these approaches were crucial for the work described in the rest of the manuscript. The following chapter provides one such example. Reporter gene data were used to understand cell physiology and based on the insights this provided, we proposed a strategy to control the bacterial growth.

Chapter 5

Analysing and controlling cell physiology

The work presented in this chapter is the unexpected development of the analysis of the carbon starvation response described in Chapter 3. Follow-up studies made us realise the major role played by the global control of gene expression in the adaptation of bacterial growth, resulting from adjustments in the activity of the gene expression machinery. This work on global control was performed as part of the PhD thesis of Sara Berthoumieux. I co-supervised the early work of Sara on the global control of gene expression during her Master thesis and was closely involved in this part of her PhD project for the analysis and interpretation of the data. I also designed and performed some of the genetic constructions used in the study. This work and other work published at the same time, drew attention to the fact that the functioning of biochemical networks cannot be disconnected from the physiological state of the cell [Berthoumieux et al., 2013].

The observation motivated the follow-up study and the bet that adjusting the intracellular level of either ribosomes or RNA polymerase could be used to control the growth rate. This work was performed by Jérôme Izard as part of his PhD thesis, whom I co-supervised with Hans Geiselmann and Stéphan Lacour, and by the post-doctoral researcher Cindy Gomez Balderaz, co-supervised by Hans Geiselmann and Hidde de Jong. While controlling the synthesis of ribosomal proteins was not met with success, the strategy to control expression of the two limiting $\beta\beta'$ subunits proved more promising [Izard et al., 2015]. In addition to the co-advisorship and research design, I performed some experiments myself and contributed to the data analysis, notably the microfluidics data. This study was my first encounter with synthetic biology approaches. It was also a long and rich scientific adventure, which continues nowadays through a new line of research on our team agenda, dedicated to the analysis of resource allocation problems (see Chapter 8).

5.1 Contribution of cell physiology to the global control of gene expression

In the work discussed in Chapter 3, we modelled the various genetic regulatory interactions involved in the response of *E. coli* to carbon starvation. The model summarized the view that gene expression changes accompanying adaptation of bacterial growth to the stress result from combined effects of positive and negative transcription regulators. In the absence of quantitative and dynamical data in the literature that could be confronted to the model, J. Geiselmann and his group launched in the wake of this study a new experimental program, with my help. It aimed at quantifying the dynamics of gene expression in *E. coli* by means of reporter genes. Using the approach described in Chapter 4, we reconstructed promoter

activities from the fluorescence and absorbance measurements. An example is provided in Figure 5.1 for genes of a subpart of the carbon starvation response network. They include the global regulators of transcription Fis and CRP, as well as the metabolic gene *acs* coding for the acetylCoA synthetase. This enzyme allows the use of acetate as a carbon source. It is inhibited by Fis and activated by the complex CRPcAMP. Our former study suggested that Fis and CRP are part of a regulatory switch controlling growth adaptation, with each gene inhibiting each other's expression, allowing the expression of Fis in exponential phase and that of CRP in stationary phase. The promoter activities in Figure 5.1 tell a different story. The profiles for CRP and Fis are similar, with high levels in exponential phase that drop as cell growth rate slows down. The observation of an induction of Acs at the entrance in stationary phase is consistent with other reports [Wolfe, 2005]. We also obtained a similar gene expression profile for Fis in a former study with a different experimental approach [de Jong et al., 2010].

The similarity of gene expression profiles for *fis* and *crp* suggests that a common control mechanism might be at work. The two genes share at least the same gene expression machinery, whose activity is known to vary with the physiological state of the cell, notably the growth rate in steady-state conditions [Dennis and Bremer, 2008]. The concentration of available RNA polymerase is one example, but this quantity is difficult to monitor directly [Klumpp and Hwa, 2008]. As an indirect read-out of the global physiological state we therefore decided to use a constitutive promoter, the pRM promoter of phage λ, whose activity is controlled by the transcription and translation machinery and the pools of precursor metabolites, but not by any particular transcription factor in *E. coli* [Berthoumieux et al., 2013]. The activity of the promoter is shown in Figure 5.1(e). It is stationary in exponential phase and stabilizes to a lower level when cells cease growing. Considering this profile as representative of the physiological state of the cell, we developed an approach to dissect the contribution of the transcription factors and the global physiological state to gene expression. The promoter activity was then formulated as follows:

$$p(t) = k\, p_1(t)\, p_2(t), \tag{5.1}$$

with k [M min^{-1}] representing the maximum promoter activity. The dimensionless term $p_1(t)$, for convenience assumed to vary between 0 and 1, quantifies the modulation of the promoter activity by the global physiological state, for instance through the availability of free RNA polymerase. The dimensionless term $p_2(t)$, also varying between 0 and 1, accounts for the effect of transcription factors and other specific regulators, and may take the form of sigmoidal regulation functions as seen in Chapter 3. Normalization of the promoter activities with a reference state t^0 - chosen here at growth transition - and a subsequent logarithmic transformation gave the following expression:

$$\log \frac{p(t)}{p^0} = \log \frac{p_1(t)}{p_1^0} + \log \frac{p_2(t)}{p_2^0}\,, \tag{5.2}$$

with $p_1^0 = p_1(t^0)$ and $p_2^0 = p_2(t^0)$.

This simple model can be used to test different hypotheses on the contribution of global and specific effects on gene expression. For instance, dominance of the global physiological effect makes the effect of transcription factors negligible. This is translated by $p_2(t) \approx p_2^0$ and thus

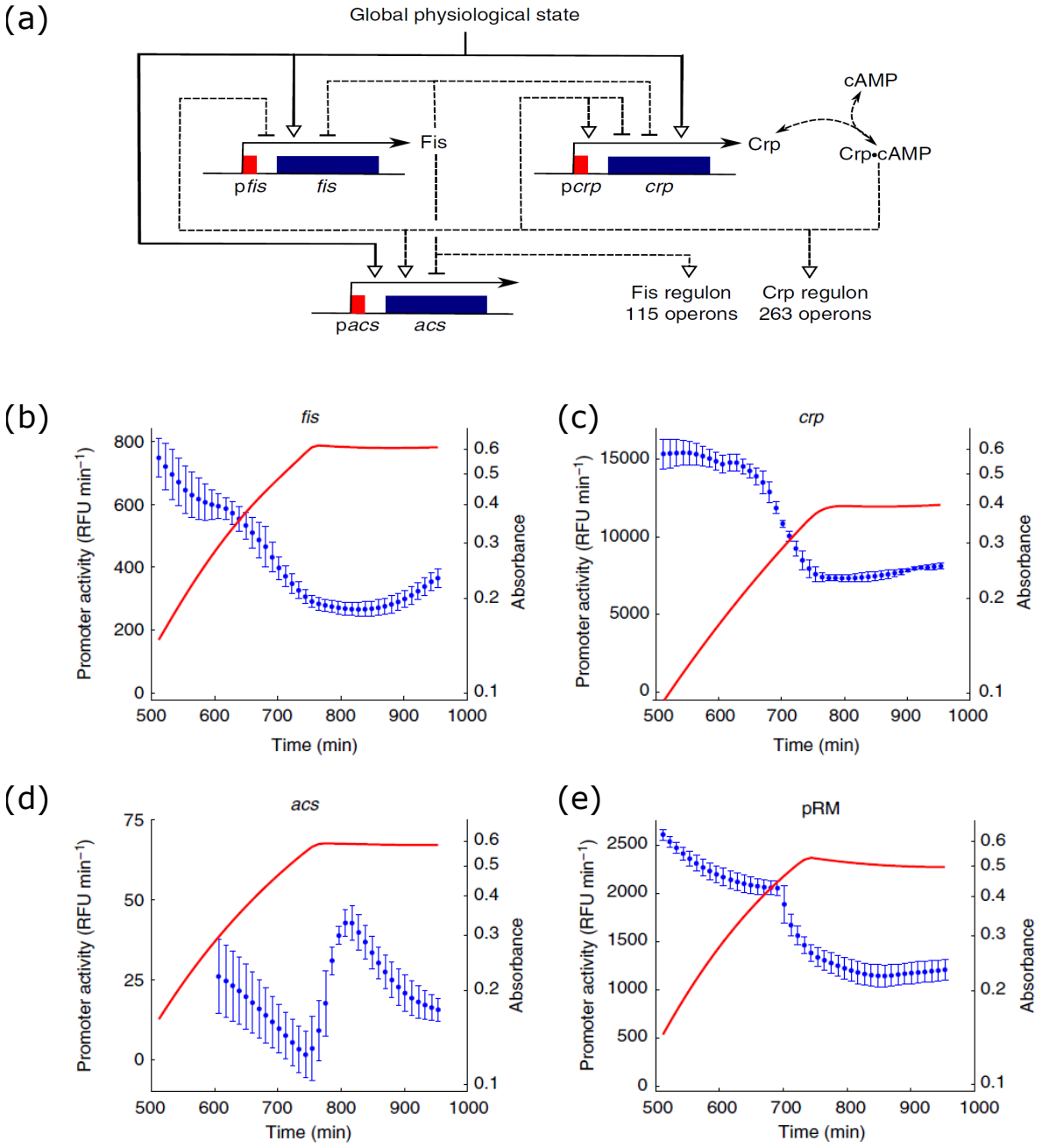


Figure 5.1 – (a) Central regulatory circuit involved in the control of *E. coli* carbon metabolism, consisting of the two pleiotropic transcription factors Crp and Fis and their regulatory interactions. The global physiological state affects the expression of all genes in the network. Genes are shown in blue and promoters in red. Specific regulatory interactions are indicated by dashed lines and the effect of the physiological state by solid lines. (b-e) Experimental monitoring of transcriptional response of network. Time-varying activity of *fis* promoter (b, blue), derived from GFP reporter data, and absorbance (solid line, red). (c–e) Idem for the activities of the *crp* and *acs* promoters, as well as the activity of the pRM promoter of phage λ. The latter promoter is constitutive in the conditions studied and reflects the global physiological state of the cell. From [Berthoumieux et al., 2013].

$\log(p_2(t)/p_2^0) \simeq 0$ in the right-hand side of Equation (5.2). In Figure 5.2(a-c), for instance, we test the contribution of global physiological effects (represented by pRM activity) to the promoter activity of *fis*, *crp*, and *acs* from Figure 5.1. Variation of the log normalized activity of pRM (representing the contribution of global physiological effects) explains the variation of the log normalized activity of promoter *fis* and *crp*, while it does not for *acs*. Taking into account both global and specific effects are necessary for this gene. It is induced at the onset of stationary phase, when its transcription factor CRP is itself activated by cAMP. Using dynamic data on intracellular cAMP concentration obtained from measurements of external cAMP (Figure 5.2(e)), we were able to show that the log normalized variation of intracellular cAMP concentration explains the variation of the remaining log normalized activity of *acs* promoter, once we have subtracted the contribution of the global physiological effect. These results and others obtained with variant strains in [Berthoumieux et al., 2013] show the importance of taking into account global physiological effects in gene expression. The specific regulatory effects by transcription factors appear to fine-tune the gene expression level set by global effects. This new vision on gene expression was confirmed in parallel for different promoters in different environmental conditions and micro-organisms (*E. coli* and yeast) [Gerosa et al., 2013, Keren et al., 2013]. These various works have revived past work in the 1950s-70s on the dependency of the macromolecular composition of the cell on growth rate, in particular the RNA and protein content [for review Dennis and Bremer, 2008]. They also paved the way to a synthetic biology approach in a follow-up study, aiming at controlling bacterial growth.

5.2 A synthetic biology approach to control bacterial growth

In the previous section, we analysed the natural strategies evolved by bacteria to adjust gene expression to growth changes through a modification of the activity of the gene expression machinery. This allows reallocating the available resources to the production of RNAs and proteins in the proportions needed at the new growth rate. Changing these resource allocation strategies has many applications in both fundamental microbiology and biotechnology. As we reviewed in [de Jong et al., 2017b], some of them re-engineer the transcription and translation machinery to re-allocate resources towards the production of proteins or compounds of interest. In [Izard et al., 2015], we proposed such an approach, by modifying the natural control of the synthesis of the two limiting RNA polymerase $\beta\beta'$ subunits in *E. coli*.

We engineered an *E. coli* strain, in which we replaced the natural promoter of genes *rpoBC* by a synthetic *lac* promoter inducible by IPTG and decoupled the synthesis of ribosomal proteins encoded by genes *rplKAJL* from that of the $\beta\beta'$ subunits (Figure 5.3(a); see [Izard et al., 2015] for more details). We also added two chromosomal copies of the *lacI* gene to confer mutational robustness to the strain. Inhibition of the synthetic *lac* promoter exerted by the *lac* repressor is relieved by IPTG addition. This leads to the expression of the two limiting $\beta\beta'$ subunits allowing RNA polymerase formation and cell growth. We characterized the physiology of the strain in [Izard et al., 2015] and the main results are shown in Figure 5.3. At high IPTG concentrations, the growth curves of the engineered "R" strain and the reference wild-type "W" strain are comparable (Panel (b)), as well as their growth rates at steady state in exponential phase (Panel (c)). The reference strain "W" is a wild-type strain of *E. coli* including the two extra copies of *lacI* gene. At low IPTG levels, cells stop growing after some time and their

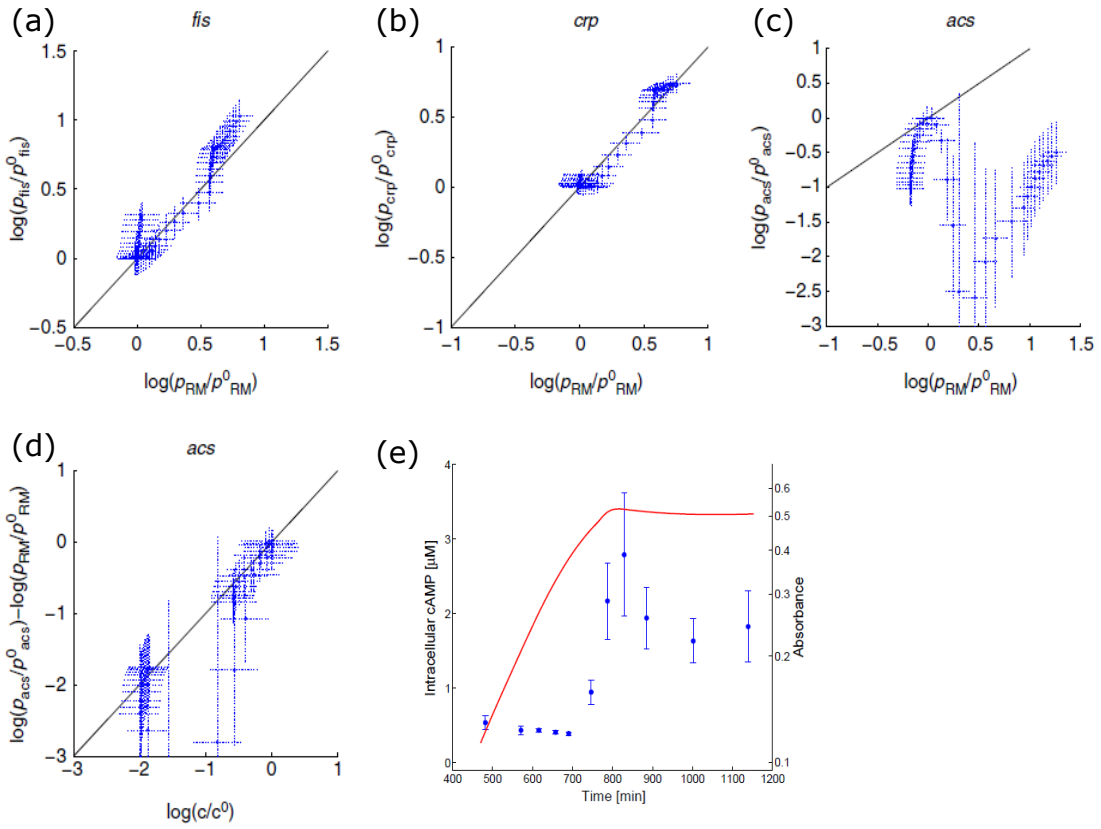


Figure 5.2 – Predicted and observed control of *fis*, *crp*, and *acs* activity by CRPcAMP and the physiological state of the cell, in various experimental conditions and genetic backgrounds. (a) Predicted (–, black) and measured (•, blue) relative activity of the *fis* promoter ($\log(p_{fis}(t)/p^0_{fis})$) as a function of the relative activity of the pRM promoter ($\log(p_{RM}(t)/p^0_{RM})$). (b-c) Idem for *crp* and *acs*. (d) Predicted (–, black) and measured (•, blue) remaining relative activity of the *acs* promoter after subtraction of the effect of global physiological parameters ($\log(p_{acs}(t)/p^0_{acs}) - \log(p_{RM}(t)/p^0_{RM})$) and as a function of the relative intracellular cAMP concentration ($\log(c(t)/c^0)$). (e) Absorbance (red) and derived concentration of intracellular cAMP from measurements of the external cAMP concentration. The confidence intervals in the plots have been computed from experimental replicas. From [Berthoumieux et al., 2013].

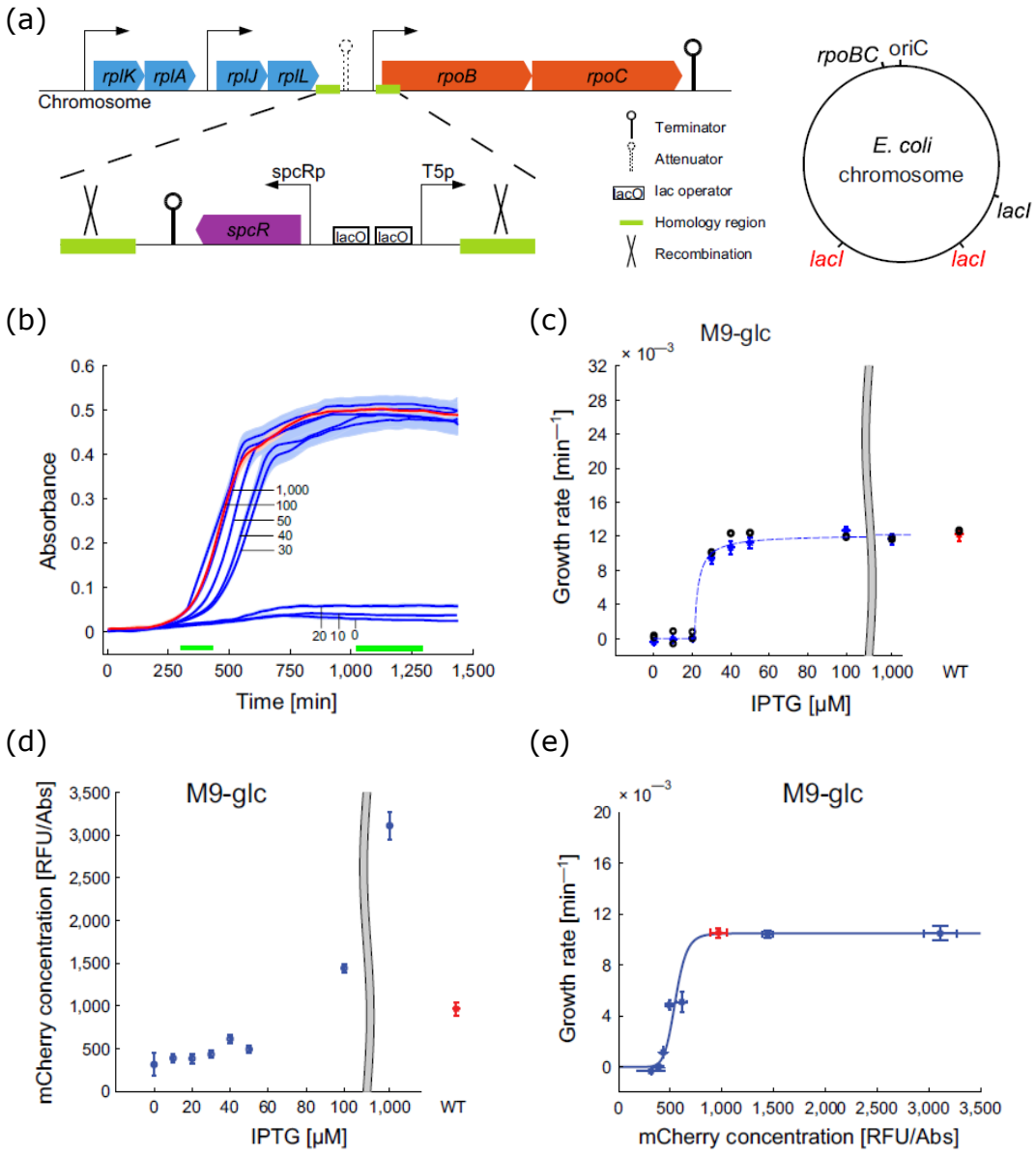


Figure 5.3 – (a) Construction of an *E. coli* strain with inducible expression of the *rpoBC* genes encoding the $\beta\beta'$ subunits of RNA polymerase. (b) Growth kinetics in a microplate of the engineered "R" strain (blue) and the wild-type strain "W" (red) at different concentrations of IPTG added to M9 minimal medium supplemented with 0.2% glucose. The growth rates are typically computed in the time intervals indicated by the green bars. (c) Growth rate estimated from the absorbance data for the R and W strains. The black circles represent the growth rates obtained with the R and W strains in shake flask experiments. (d) Quantification of the $\beta\beta'$ subunits of RNA polymerase using a chromosomal fusion of the *rpoC* gene with the gene encoding the mCherry fluorescent reporter protein. (e) Quantitative dependence of the growth rate on $\beta\beta'$ concentrations, computed from the data in (d) and the measured growth rates at the different IPTG concentrations. The blue curve is a Hill function with a Hill coefficient of 10. From [Izard et al., 2015].

growth rate is null. These results obtained in M9 minimal medium supplemented with glucose were also observed in other growth media and genetic backgrounds. Labelling protein β' in the R and W strains with the fluorescent protein mCherry allows quantifying the expression of the RNA polymerase subunits (Panel (d)). A high IPTG concentration results in the overexpression of the subunits compared to the wild-type situation. At low IPTG, the subunit levels is too low to support growth (Panel (e)). We have thus a growth switch, with growth that can be turned on or off by the simple addition or removal of inducer.

Time-lapse microscopy in a microfluidics experiment followed by the quantification of individual cell growth rates showed that the switch is reversible and that almost all growth-arrested cells resume growth when IPTG is added back to their medium (Figure 5.4(a)). Using the engineered strain allows to reallocate resources from biomass formation to the production of a metabolite of interest, glycerol in this case. We transformed the R and W strains with a plasmid carrying the genes that code for the glycerol pathway in yeast. Following the growth arrest and the depletion of RNA polymerase, the production yield of glycerol in the R strain is twice higher than in the W strain (Figure 5.4(b)-(d)), and close to the maximal theoretical yield [Izard et al., 2015]. This example shows that growth-arrested cells remain metabolically active and represent an interesting platform for biotechnological applications. In a follow-up study, we showed in collaboration with the company Metabolic Explorer that the synthetic growth switch still works efficiently in liter-scale bioreactors (Ropers *et al.*, submitted). We also characterized the physiology of the growth-controlled strain, by assessing the reorganization of the metabolome, transcriptome, and proteome induced by the growth switch. The data confirms the idea that the growth-controlled strain is functioning like a bag of active metabolic enzymes.

5.3 Discussion and perspectives

The global control mechanism evidenced in our study in [Berthoumieux et al., 2013] is the consequence of changes in the pools of ribosomes and RNA polymerases, which affects transcription and translation rates directly, modifies gene expression on a global scale, and ultimately affects cell macromolecular composition (reviewed in [Jun et al., 2018]). Our work shows that these observations, originally made for *E. coli* cells in steady-state growth, remain valid in dynamic environments: when growth slows down following glucose exhaustion in the growth medium, changes in the expression of global regulators of transcription can be attributed to these global effects rather than to specific regulations [Berthoumieux et al., 2013]. This shows that the functioning of biochemical networks cannot be disconnected from the physiological state of the cell [Berthoumieux et al., 2013]. Following work took this phenomenon into consideration to analyse the cellular metabolism and mRNA decay (see Chapters 6 and 7).

The exploitation of global control of gene expression for biotechnological purposes allowed us to engineer an *E. coli* strain and turn it into a growth switch [Izard et al., 2015]. I am still full of wonder at the results obtained with this strain. While flexibility characterizes the metabolism of bacterial strains and often oppose modifications aiming at hijacking it for bio-production purposes for instance, the gene expression machinery resembles an Achille's heel. It has such a profound impact on cell processes, that it becomes possible to intervene directly in the resource

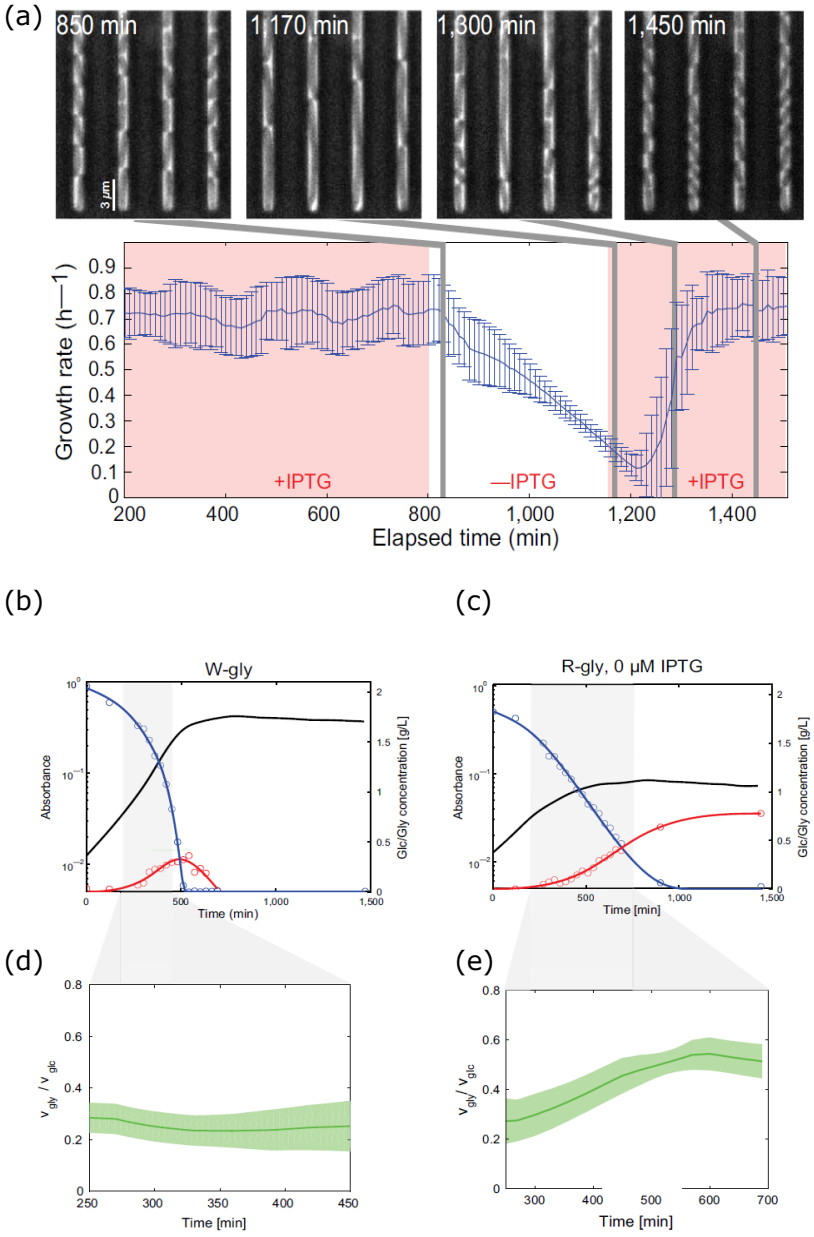


Figure 5.4 – (a) Growth arrest by external control of *rpoBC* is reversible. The R strain was grown in a microfluidics device and phase-contrast images were acquired every 10 min. Cells were grown in M9 minimal medium with glucose, initially in the presence of 1M IPTG. 6 h after removing IPTG from the medium, the cells are elongated. About 100 min after adding back lM IPTG into the medium, the elongated cells divide and resume normal growth. The growth rates in the plot are the weighted mean of the growth rates of 100 individual cells. The glycerol-producing W strain (W-gly) (b) and R-gly strain (c) were grown in shake flasks in M9 minimal medium supplemented with 2 g/L glucose. The optical density (OD_{600}, black), the glycerol (red), and glucose (blue) concentrations were measured in samples taken at intervals of about 30 min using coupled enzyme assays. (d) The instantaneous yield of glycerol production in the W-gly is computed by dividing the glucose consumption rate by the glycerol production rate, in a time interval in which the derivatives of the glucose and glycerol concentration curves are well defined (between 250 and 450 min). (e) Idem for the R-gly strain (in the time interval between 250 and 700 min).

allocation scheme by targeting the machinery as we did in our study. The engineered strain is a great tool, which has utility both in fundamental research and biotechnology. On the one hand, it allows the analysis of the natural resource allocation strategies that have evolved and, on the other hand, it constitutes a new method for producing metabolites, peptides and recombinant proteins, which we patented in Europe and the USA [Geiselmann et al., 2015].

The work described so far focused on small subnetworks controlling the adaptation of *E. coli* to a carbon source. The main reason is that we restricted our analysis to measurable parts of the *E. coli* network, for which data was hence available. However, the increased availability of omics data makes it possible to analyse biological processes involved in bacterial adaptation at a larger scale. This will be the subject of the two following chapters dedicated to the analysis of the regulation of metabolism and mRNA decay at the genome-wide level in *E. coli*.

Chapter 6

Metabolic network models as platforms for integrating omics data

This chapter is dedicated to the analysis of the regulation of metabolism in *E. coli* using constraint-based models, which augment the information content of the data. The work presented in this chapter was performed as part of the PhD thesis of Manon Morin funded by an INRA-Inria PhD grant. Manon was co-supervised by Muriel Cocaign-Bousquet, Brice Enjalbert, and myself. Muriel was the main supervisor and I lead and performed the modelling work and data integration shown in the two studies. With this work, I made my first steps in constraint-based modelling and the exploitation of metabolomics data. This is the reason why I chose to highlight the two corresponding papers [Morin et al., 2016, 2017]. I remain impressed to see how much information can be extracted from the data, based on the sole information on the measured uptake and secretion rates, as well as the growth rate, plus additional intracellular metabolite concentrations. I used a similar approach in the framework of the PhD thesis of Stéphane Pinhal, whom I co-supervised with Hidde de Jong and Johannes Geiselmann, dedicated to the analysis of growth inhibition by acetate [Pinhal et al., 2019].

6.1 Post-transcriptional regulation of central carbon metabolism in *E. coli*

The possible influence of mRNA stability on metabolic activity has long been ignored. With the identification of global regulators such as the carbon storage regulator system (CSR) and of small RNAs controlling the stability and/or translation of metabolic genes, it has become clear that post-transcriptional regulations add a new layer to the already complex network controlling cell metabolism in *E. coli* [Kotte et al., 2010, Kochanowski et al., 2013]. To what extent this added complexity is determinant in the adaptation of bacterial growth to the environment was unclear when we started this study. We focused on the CSR system, whose functioning is paradoxical. CSR consists of the dimeric mRNA binding protein CsrA and small regulatory RNAs CsrB/C, which inhibit CsrA activity. Both of these noncoding RNAs are targeted by the protein CsrD, which triggers their RNase E-dependent degradation (see Pourciau et al. 2020 for a recent review). The system is essential for growth on glycolytic media, where pleiotropic regulators like ppGpp, CRP, and RpoS have larger regulons but are not essential. Past studies conducted with a multitude of different strains and conditions showed that attenuating the *csrA* gene through deletion of the last 10 amino acids leaves cells viable, albeit with a strongly reduced growth and perturbed biofilm formation, motility, and the accumulation of the storage sugar glycogen [Esquerré et al., 2016, Romeo et al., 1993]. Our objective was to obtain a more global view of the CSR regulon and understand the physiology of the attenuated *csrA* strain and other CSR deletion mutants. Is really the accumulation of glycogen responsible for the

essentiality of CsrA, as proposed in the literature [Timmermans and Van Melderen, 2009]? Or is there an alternative explanation to this phenomenon? Do post-transcriptional regulation play an important role in shaping the central carbon metabolism? We characterized CSR targets among the central carbon metabolism genes to obtain responses to these questions, through a multi-scale analysis of growth properties, mRNA levels, enzyme activities, fluxes and metabolite concentrations in CSR variant and wild-type strains, as summarized in Figure 6.1.

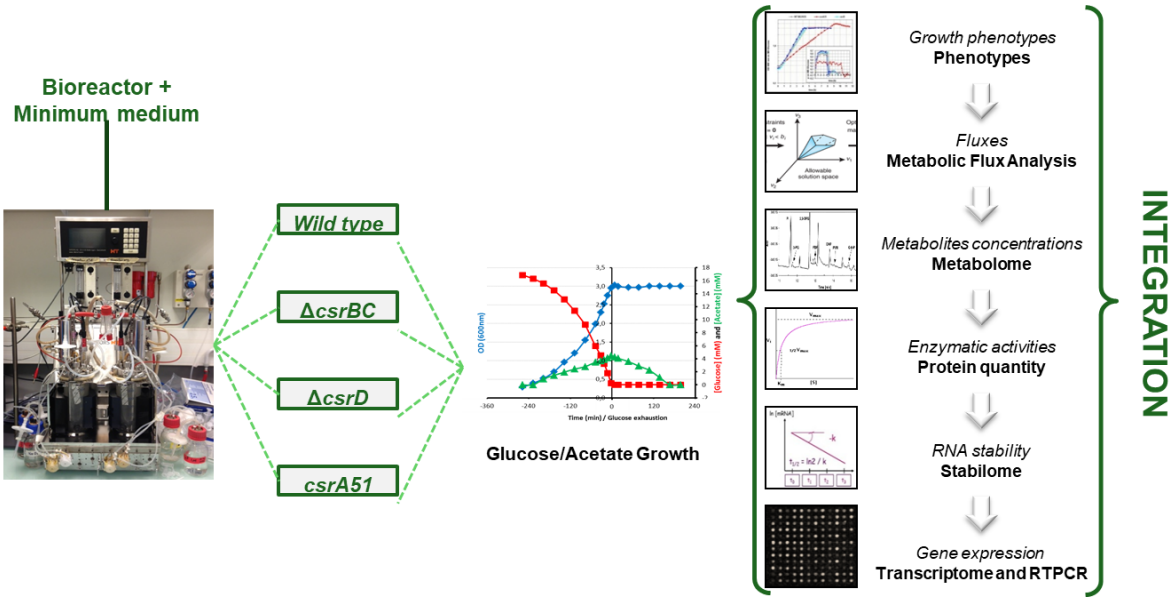


Figure 6.1 – Systematic measurement of different molecular levels in wild-type and mutant/deletion strains of the CSR system during a glucose-acetate diauxie in *E. coli*.

In a first study, we probed the activity of central carbon metabolism in the attenuated *csrA* and wild-type strains growing exponentially in minimal M9 medium supplemented with glucose [Morin et al., 2016]. Figure 6.2 shows in bold the names of metabolites and reactions in glycolysis/gluconeogenesis, glycogenesis, pentose-phosphate pathway, and the Krebs cycle for which experimental data was obtained in the study - a measured concentration, enzyme specific activity and/or mRNA level obtained by qRT-PCR. The corresponding data set is shown in Figure 6.3. The flux through each reaction was determined by metabolic flux analysis. In the absence of transcriptomics data at the time (they were obtained later on, [Morin et al., 2020]), we could not develop a condition-specific model from a generic metabolic reconstruction [Bordbar et al., 2014]. Instead we tried to exploit as much as we could the available metabolomics data to develop models as specific to the strains and conditions studied as possible. We adapted the iAF1260 genome-scale reconstruction of *E. coli* metabolism [Orth et al., 2010] for the metabolism and storage of glycogen, without distinguishing between the different forms of the molecule, because only the total pool of glycogen was quantified. Maintenance fluxes were determined from previous measurements of steady-state growth rates and glucose uptake rates [Esquerré et al., 2014].

We then formulated a limited number of uptake and secretion constraints that were directly motivated by the composition of the growth medium and the utilization of glucose as the sole

carbon source and the secretion of acetate and CO_2. These were determined by means of the approach described in Chapter 4. Intracellular metabolite concentrations were used to specify thermodynamic constraints that enforce reaction directionality, based on the determination of the Gibbs energy of the reactions at room temperature, and physiological pH and ionic strength [Fleming et al., 2009, Henry et al., 2006, 2007]. The resulting constraint model was then used to predict the intracellular metabolic fluxes. Instead of using a principle of optimality to solve the model, we performed a metabolic flux analysis, where we sought to minimize the difference between the measured and predicted fluxes, including the biomass production rate. This was formulated as a linear programming problem, similarly to what was done in [Lee et al., 2012]:

$$\begin{aligned} \min \sum_{j=1}^{p}(u_j^+ + u_j^-) \text{ subject to:} & \\ N\,v = 0, & \\ v^l \leq v \leq v^u, & \\ v_j - u_j^+ + u_j^- = u_j^M, \text{ for all } j = 1, ..., p, u_j^+, u_j^- \geq 0\,. & \quad (6.1) \end{aligned}$$

with v the vector of steady-state fluxes with lower bounds v_l and upper bounds v_u, N the stoichiometry matrix, u_j^+ and u_j^- non-negative dummy fluxes, and u_M the vector of p measurements of exchange fluxes and growth rate. We assumed that the first p elements of v correspond to the measured fluxes. We further analysed the optimal solutions by flux variability analysis [Mahadevan and Schilling, 2003] to determine the minimum and maximum flux values satisfying the constraints and consistent with the measurements. This allowed restricting the possible values of intracellular fluxes to very tight intervals, as shown in Figure 6.3.

A genetic perturbation such as CsrA attenuation may affect metabolism in different manners: directly, through the modification of the expression of CsrA target genes, or indirectly through modifications of the growth rate or of upstream or downstream reactions, which are propagated through the metabolic pathway (Figure 6.3(e)). The effect of growth rate on metabolism could be assessed by using data from a previous study on continuous cultures growing at rates comparable to the attenuated and wild-type strains (0.3 and 0.6 h^{-1}) [Esquerré et al., 2014] (see Figure 6.3). We separated the direct and indirect effect of CsrA attenuation on metabolic fluxes using a hierarchical regulation analysis [ter Kuile and Westerhoff, 2001, van Eunen et al., 2011]. The approach quantifies the contribution of changes in gene expression or metabolite pool to the flux change. We write the rate of an enzyme-catalysed reaction at steady-state $J = v(e, x, K) = f(e) \times g(x, K)$, in which v is the rate, e the enzyme concentration, x the vector of metabolite concentrations (substrates, products, and effectors), and K, the vector of dissociation constants. A logarithmic transformation dissects the flux into two terms, one depending on the enzyme concentration and the other, on the metabolite concentrations: $log\ J = log\ f(e) + log\ g(x, K)$. The change of flux between the wild-type and *csrA51* strains is written: $\Delta log\ J = \Delta log\ f(e) + \Delta log\ g(x, K)$. The contribution of gene expression and metabolic control to the flux change is quantified by dividing the latter expression as follows:

$$\frac{\Delta log\ J}{\Delta log\ J} = \frac{\Delta log\ f(e)}{\Delta log\ J} + \frac{\Delta log\ g(x, K)}{\Delta log\ J} = \rho_h + \rho_m = 1\,, \quad (6.2)$$

in which ρ_h is the hierarchical regulation coefficient and ρ_m, the metabolic regulation coefficient. In our conditions, $f(e)$ corresponds to the measured specific activity SA of the enzyme. We can therefore determine ρ_h directly from the experimental data and the fluxes: $\rho_h = \Delta logSA/\Delta logJ$, and deduce $\rho_m = 1 - \rho_h$. For each reaction, sets of hierarchical regulation coefficients were calculated from the specific activities of the different replicates of each strain, and the lower and upper bounds for the fluxes, which gives the boxplots in Figure 6.3(f).

These experimental and model analyses allowed us to demonstrate the strong control of the upper part of glycolysis by the CSR post-transcriptional regulatory system [Morin et al., 2016]. Attenuation of CsrA activity results in a decrease of most glycolytic activities, especially the phosphofructokinase. This causes an accumulation of metabolites in the upper part of glycolysis before the phosphofructokinase step and results in a glucose–phosphate stress controlling negatively the sugar uptake. This strongly affects the bacterial growth rate and could explain the essentiality of *csrA* gene for growth on glycolytic substrates. The glucose-phosphate stress can be relieved by restoring PfkA activity in the *csrA* mutant strain.

6.2 Post-transcriptional regulation of metabolic adaptation

Our previous study demonstrated the major role played by the CSR system in the control of upper glycolysis, but its putative role during metabolic adaptation remained to be established. In addition we showed that metabolic reprogramming rather than glycogen accumulation causes a growth defect when CsrA is attenuated: does glycogen have a physiological role in *E. coli*? This polysaccharide is the main storage form of glucose, from bacteria such as *Escherichia coli* to yeasts and mammals. Although its function as a sugar reserve in mammals is well documented, the role of glycogen in bacteria is not that clear. In a follow-up study using the same data set obtained with our multi-scale analysis of central carbon metabolism (Figure 6.1), we analysed gene expression and metabolic pools in CSR variants and wild-type strain transitioning from growth on glucose to growth on acetate [Morin et al., 2017].

The main results of this study are summarized in Figure 6.4. The figure illustrates the evolution of biomass and pools of glucose, acetate and glycogen, as cells grow first on glucose and then transition to acetate when the preferred carbon source is depleted (Panels (a)-(d)). Glycogen is also consumed during the second growth phase. The timing of acetate consumption varies between strains (Panel (c)). We could not attribute it to differences in gene expression levels, but to differences in the energetic status. Analyses of the adenylate energy charge AEC = [ATP + 1/2 ADP]/[ATP + ADP + AMP] before glucose depletion showed a high energetic status, typical of exponential growth, while it drops after glucose depletion, to low levels in the *csrBC* strain reminiscent of dying cells (Panel (e)). Using the metabolic model previously developed in [Morin et al., 2016], we predicted by flux balance analysis the maximal ATP fluxes that the various wild-type and CSR variant strains are able to produce following glucose depletion, in the range of measured glycogen and acetate rates (Panel (f)). Acetate and glycogen appear to cover the needs for maintenance energy. Preventing glycogen synthesis through deletion of the *glgC* gene does not allow cells to maintain their energetic status after glucose depletion (Panel (g)). This study brings a new vision to the physiological role of glycogen. The stored polysaccharide

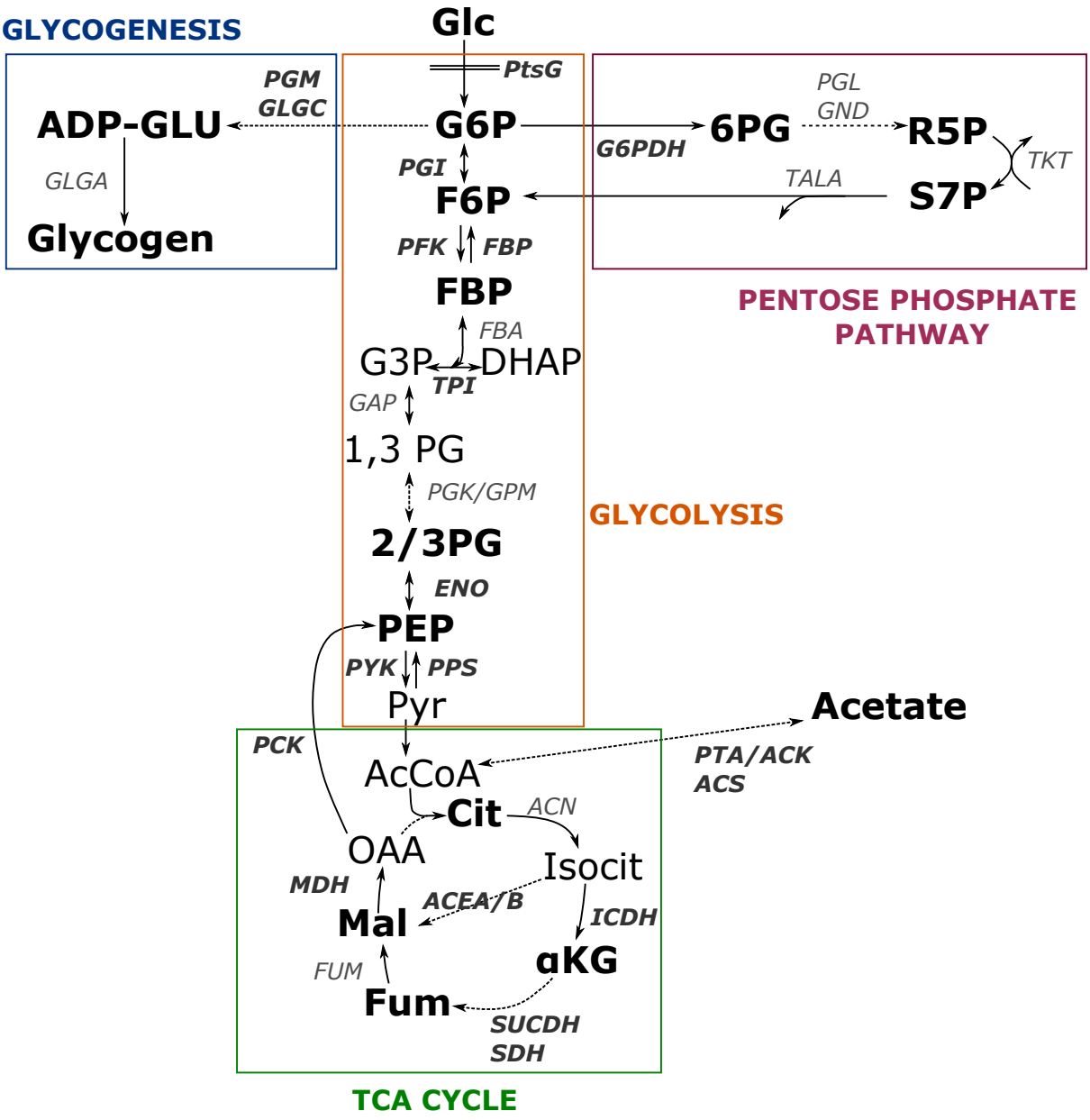


Figure 6.2 – Schematic representation of the main reactions in central carbon metabolism including the glycolysis, pentose-phosphate pathway, glycogenesis, and tricarboxylic acid (TCA) cycle. Metabolites in bold were quantified by metabolomics. Reactions shown in bold were monitored, by qRT-PCR measurement of the mRNA levels and/or determination of the enzymatic activities. Metabolic network analysis allowed to predict the flux value through each metabolic reaction. Dashed lines represent aggregated reactions. 2/3PG corresponds to the pool of 2PG and 3PG that cannot be distinguished in metabolomics experiments. The corresponding data is shown in Figure 6.2.

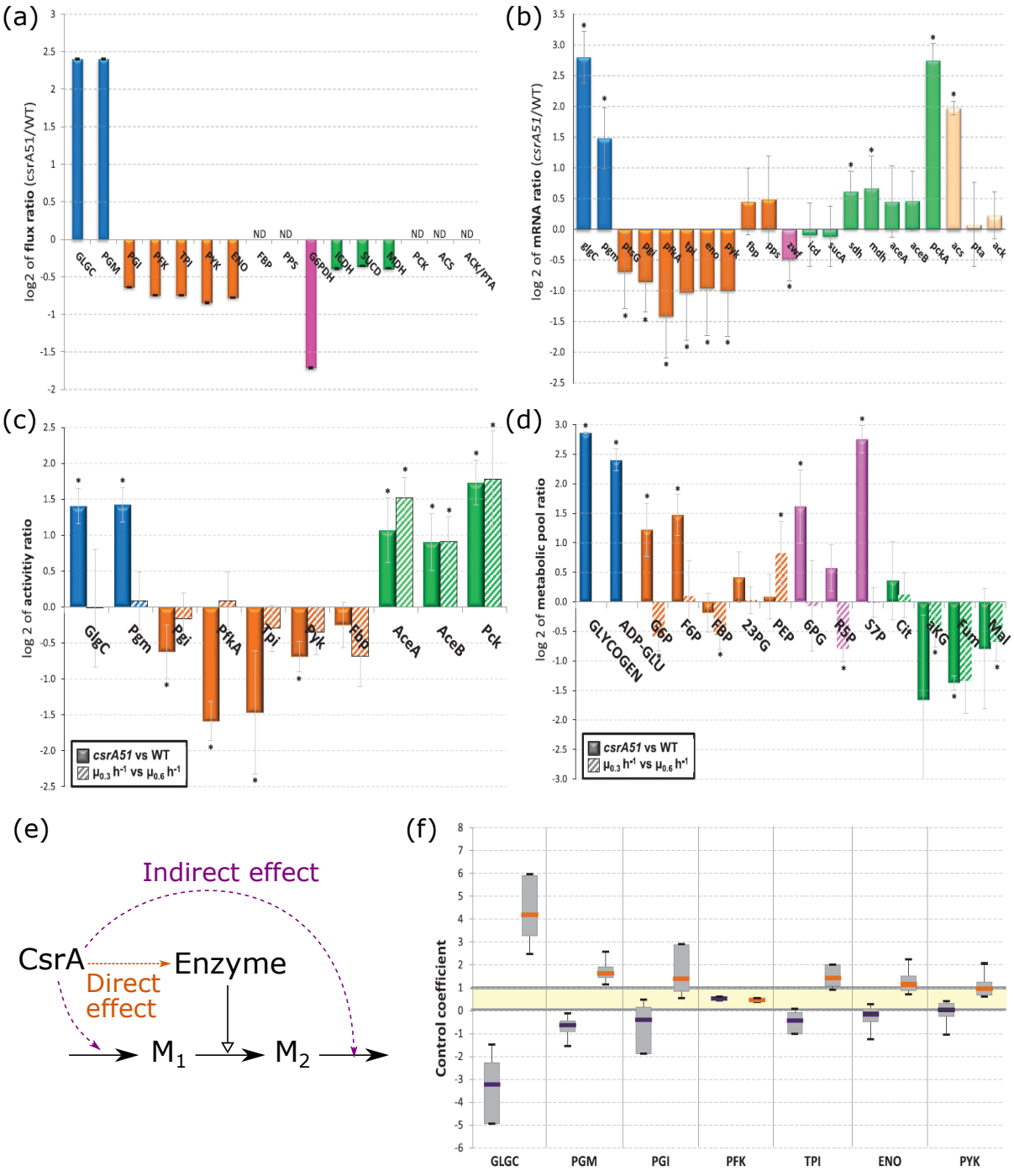


Figure 6.3 – Comparison of the (a) fluxes, (b) mRNA levels, (c) enzyme activities, and (d) metabolite concentrations in the central carbon metabolism between the wild-type and the *csrA51* strain or at different growth rates. The displayed values correspond to the log2 of the ratio of the *csrA51* strain to its isogenic wild type. Hatched columns represent log2 of ratio between values in the wild-type growing at μ =0.3h^{-1} (chemostatic cultures) to the same strain at μ =0.6h^{-1} (batch cultures) (data from Esquerré et al. 2014). (e) Schematic representation of direct and indirect effects of CsrA on metabolic fluxes. (f) Hierarchical and metabolic regulation coefficients are shown as boxplots. Boxes represent the interquartile range (IQR) between the first and third quartiles. Whiskers denote the lowest and highest values within 1.5 3 IQR from the first and third quartiles. Purple, median metabolic coefficient; orange, median hierarchical coefficient.

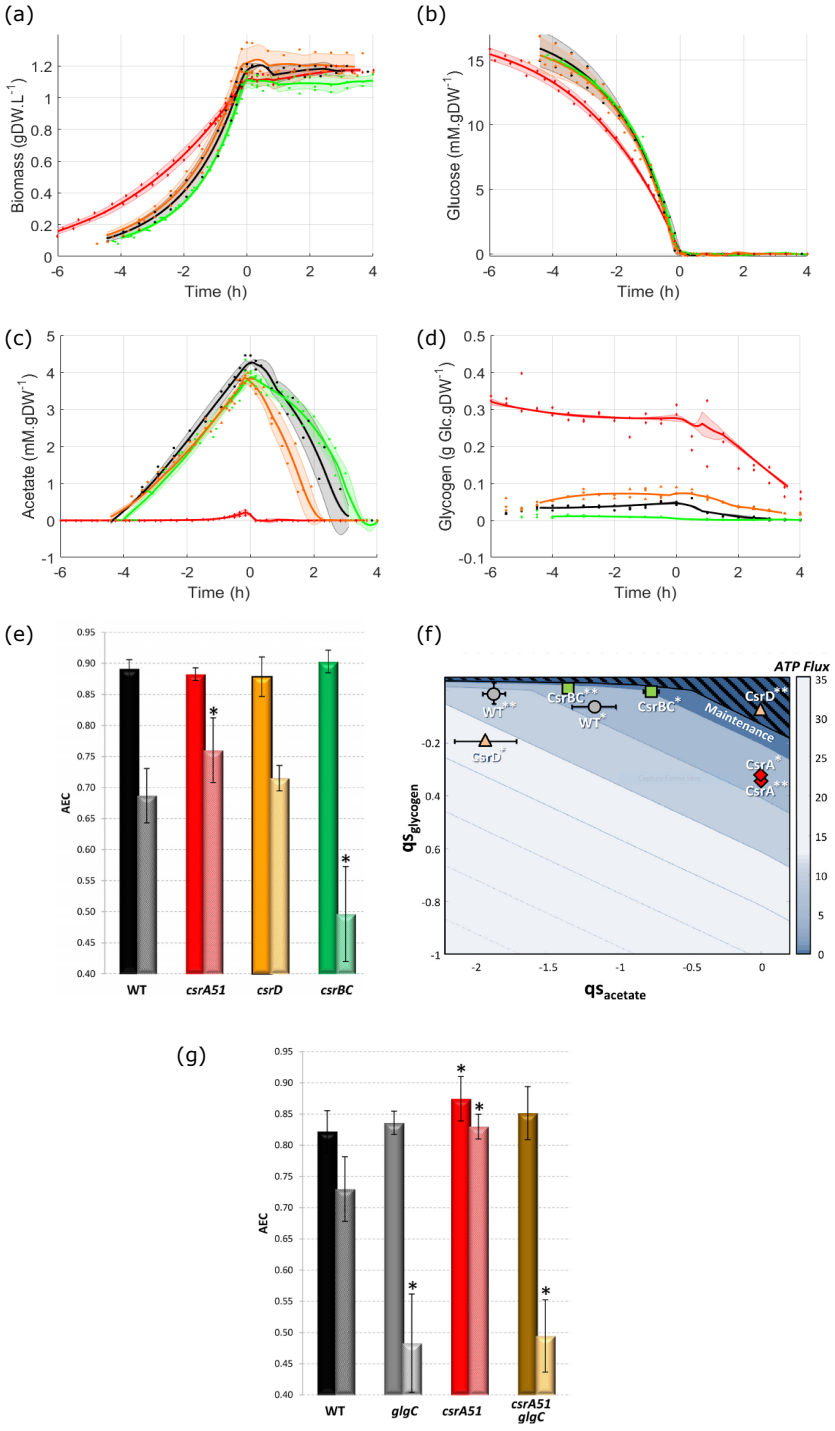


Figure 6.4 – Behaviour of CSR system mutants during the glucose-acetate transition. Changes in (a) the biomass concentration, (b) extracellular glucose concentration, (c) extracellular acetate concentration and (d) glycogen concentration. All data concerning the replicates are displayed as dots, and the fitted average value of each strain is displayed as a line. Shaded areas represent ±1 standard deviation. WT, black circles; *csrBC* mutant, green squares; *csrD* mutant, orange triangles; *csrA51* mutant, red diamonds. (e) Adenylate energy charges (AEC) of the four strains, during glucose consumption (plain bars) and 1.5 h after glucose exhaustion (striped bars). A significant difference between a mutant and the WT is represented by an asterisk ($P < 0.05$ [t-test]). (f) Maximum flux of ATP predicted by the constraint-based model in [Morin et al., 2016]. The bar to the right of the plot indicates the ATP flux values. The four strains were assessed at two different times 90 min (*) and 150 min (**). (g) Effect of the deletion of *glgC* gene on the AEC of the wild-type and *csrA51* strains.

is utilized during transition to new carbon source to provide energy for growth resumption. As major regulator of glycogen storage, the CSR system appears to play a key role for the fitness of *E. coli* cells transitioning from growth on glycolytic to gluconeogenetic substrates.

6.3 Discussion and perspectives

Integration of data with models augments the information content of the data. The two studies presented here provide one such example: the constraint-based model integrating metabolomics measurements allowed us to determine intracellular fluxes, while small models in hierarchical regulation analyses integrating the fluxes and enzyme specific activities enabled us to disentangle the direct and indirect regulatory effects of CsrA.

It is, however, difficult to causally relate the predicted changes of fluxes to changes in gene expression, metabolism, and other interesting parameters such as growth rate. While the approach in [Morin et al., 2016] enabled us to perform this analysis for a subset of reactions, the next step will be to develop an approach allowing to identify at the genome-scale level the specific parts of metabolism directly affected by a perturbation. I will come back to this point in Chapter 8, where I will discuss some work started on this subject.

The field of constraint-based modelling is evolving fast, notably with the development of approaches allowing the integration of omics data [Bordbar et al., 2014]. The recent obtention of transcriptomics data such as those obtained in the same experimental conditions [Morin et al., 2020] will help us improve our analysis of the post-transcriptional regulations of cell metabolism. Work in this direction has recently started with the internship of Amélie Caddéo (Univ. Grenoble Alpes), whom I am supervising. I will discuss this specific point in Chapter 8, for a new project in which we seek to understand the cross-talk between RNA metabolism and central carbon metabolism.

Chapter 7

Analysis of bacterial mRNA decay

This chapter is devoted to the study of mRNA degradation kinetics. Transcriptomics experiments by means of microarrays or RNAseq enable the monitoring of the decay of cellular mRNAs at the genome scale following a transcription arrest. Simple models assuming exponential decay are generally used to determine mRNA half-lives from the degradation profiles [Laguerre et al., 2018]. However, the assumption that mRNA decay follows first-order kinetics does not permit a detailed investigation of the regulatory mechanisms responsible for the diversity of degradation profiles observed experimentally. Can we develop a model of mRNA degradation that would be a good compromise between a mechanistic description of the degradation process and simplicity for the confrontation of the model with experimental data?

We addressed the question in the framework of the PhD thesis of Thibault Etienne, funded by an INRA-Inria PhD grant, which I co-advised with Muriel Cocaign-Bousquet. I was the main supervisor of Thibault. We made the bet that more information on the regulatory mechanisms could be extracted from dynamic omics data by confronting them to mechanistic models of mRNA degradation instead of exponential models. Our final objective was to be able to obtain more information than just the determination and analysis of mRNA half lives from a large data set obtained in a former study monitoring the degradation of 4254 cell mRNAs in steady-state cultures of *E. coli* growing at four different rates [Esquerré et al., 2014]. The corresponding work is described below. In a first section, I describe the mechanistic modelling of mRNA degradation and how we used the model to show that competition between mRNAs could affect the degradation kinetics [Etienne et al., 2020]. In a follow-up study described in Section 7.2, work that will be submitted soon, we show the physiological relevance of this phenomenon using dynamic omics data [Etienne et al., In preparation]. To the best of my knowledge, this work on mRNA decay is the first example of the interpretation of transcriptomics data by means of a mechanistic model that allows to identify regulatory mechanisms and analyse their contribution to the cell physiology.

7.1 Competitive effects in bacterial mRNA decay

One of the most common approaches for the experimental determination of mRNA half-life is the monitoring of residual mRNA concentrations following transcriptional arrest. An antibiotics like rifampicin is often used, which blocks the elongation of transcription by RNA polymerase. After a delay during which RNA polymerase completes the mRNAs it had started to transcribe before antibiotics addition, mRNAs no longer accumulate and are progressively degraded by the cell machinery called degradosome (Figure 7.1, yellow panel). Sequence and structure characteristics of the mRNAs are factors known to result in variations of the degradation rate, as well as regulations by small RNAs and/or RNA-binding proteins such as Hfq. Since many of the degradation profiles resemble an exponential decay, the curves are fitted by an exponential

function and the estimated rate constant is used for the calculation of mRNA half-lives. This method and some variants are the reference for the determination of mRNA half-life [Laguerre et al., 2018]. Alternative modelling approaches have analysed the coupling of degradation with other cellular processes, but they are too detailed to allow a thorough confrontation of model predictions to data (for review, Roux *et al.*, submitted).

The development of models of mRNA degradation and their analysis is described in [Etienne et al., 2020]. Based on the experimental literature, we showed that we can assimilate mRNA decay by the degradosome to a macro-reaction catalysed by the major enzyme RNase E. While degradation of each mRNA is usually studied independently from other mRNAs, we realised that cellular mRNAs share the same machinery of degradation and should compete for binding to RNase E because they outnumber the enzymes within cells. Modelling of the degradation process with mass-action law led to two models, including the mechanism of competition or not. At the moment of reducing the models, we noticed that the standard QSS approximation was not applicable in our conditions. First, because the total and not the free mRNA concentration is most likely measured in the transcriptomics assays (mRNAs are deproteinized during cell extraction). Second, because available literature data indicate an excess of RNase E with respect to individual mRNAs, while the total concentration of all cell mRNAs is way larger than the enzyme concentration. These conditions motivated the use of the total QSSA described in Section 2.1.3, since it considers the total substrate concentration and has a larger domain of validity than sQSSA, which should facilitate the future step of parameter estimation. Application of the first-order tQSSA gives for the model describing single mRNA degradation, in the absence of competition:

$$\frac{d}{dt}m_i(t) = -\frac{kcat\ E_0\ m_i(t)}{Km_i + E_0 + m_i(t)} \tag{7.1}$$

with, as seen before, the following domain of validity (Section 2.1.3):

$$E_0 + Km_i \gg m_{i0} \quad \text{and} \quad K_i \ll Km_i \quad \text{or :} \tag{7.2}$$

$$E_0 \gg m_{i0} \quad \text{and} \quad E_0 \gg Km_i \approx K_i\,. \tag{7.3}$$

Building upon previous works by Pedersen et al. [2007] and Tang and Riley [2013], we developed an approximate version of the tQSSA form which includes competition between all cellular mRNAs:

$$\frac{d}{dt}m_i(t) = -\frac{k_{cat}\ E_0\ m_i(t)}{Km_i\left(1 + \sum\limits_{j\neq i}\frac{m_j(t)}{Km_j}\right) + E_0 + m_i(t)} \tag{7.4}$$

with $j = 1\cdots n, n = 4254$ mRNAs, and any of the following conditions implying the validity of the approximation:

$$\begin{aligned}
E_0 \ll Km_i^{app}(0) + m_{i0} \quad &\text{and} \quad K \lessapprox Km_i^{app}(0) + m_{i0}\,,\\
Km_i \gg \sum_{i=1}^{n} m_{i0} \quad &\text{and} \quad K \ll Km_i^{app}(0)\,,\\
Km_i \gg \sum_{i=1}^{n} m_{i0} \quad &\text{and} \quad E_0 \gg K \gtrapprox Km_i^{app}(0)\,,\\
E_0 \gg Km_i^{app}(0) + m_{i0} \quad &\text{and} \quad E_0 \gg K\,,
\end{aligned}$$

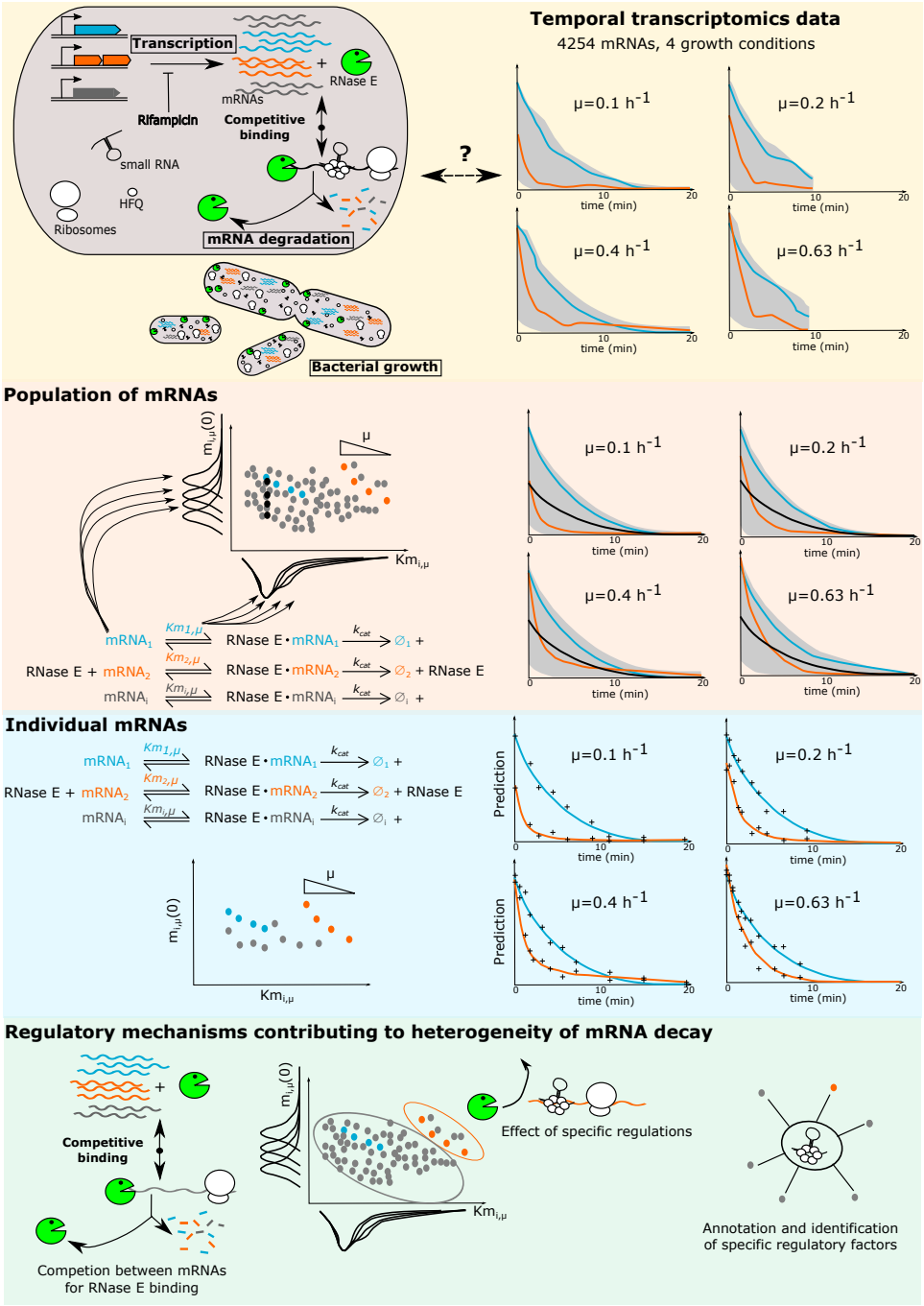


Figure 7.1 – Analysis of the physiological control of bacterial mRNA decay. Yellow panel: the objective of the study is to formulate hypotheses on the regulatory cellular mechanisms that could explain the observed adjustments of mRNA degradation profiles in *E. coli* cells growing at four different rates, like competition between mRNAs, protection of mRNAs by elongating ribosomes, or specific regulations involving small RNAs and RNA-binding proteins (CSR, Hfq,...). Orange panel: Nonlinear-mixed effect (NLME) modelling of bacterial mRNA decay using the degradation model in (7.4). We consider that the parameter values of individual mRNAs (Km values and initial concentrations) are drawn from a distribution common to the population of mRNAs. The black line represents the degradation profile of an average mRNA characterized by mean Km and initial concentration values. Blue panel: the parameters of individual mRNAs allow to generate degradation profiles that fit well the data. Green panel: parameters with values different from the population mean indicate the possibility of underlying regulatory mechanisms. Further analyses (e.g. functional annotation, enrichment analysis) are carried out to propose hypotheses on the possible regulatory mechanisms at work. From [Etienne et al., In preparation].

where $K = \frac{k_{cat}}{min\{k_{1,i}\}}$. The degradation introduces a coupling between mRNAs. For the experimental conditions described in Figure 7.1 (yellow panel) for instance, this brings the model with competition to a large system of 4254 ordinary differential equations for each of the four environmental conditions. The model differs from the one in (7.1) by the multiplication of enzyme affinity with a competition term depending on the concentration of the other cell mRNAs. This shows that competition between mRNAs decreases enzyme affinity globally for all mRNAs.

To test if competition between mRNAs has any impact in the kinetics of degradation, we numerically simulated the models by means of parameters and initial conditions from the literature satisfying the above conditions [Etienne et al., 2020]. The profiles generated with the competitive model vary from seemingly exponential to more linear profiles (Figure 7.2(a-d)). Such diversity is difficult to reproduce in the absence of competition and is observed experimentally [Etienne et al., 2020]. The explanation lies in the titration of RNase E at the beginning of the kinetics and as long as the total mRNA concentration remains sufficiently high, which delays the onset of degradation (Figure 7.2(e,f)). During this period of time, mRNA competition slows down the degradation rate. By means of rate response coefficients evaluating the sensitivity of the degradation rate to changes in mRNA concentrations, we further showed that 1) competition differentially affects the fate of mRNAs – it stabilizes mRNAs with low affinities and destabilizes those with high affinities – and 2) it explains the observed negative correlations between mRNA concentrations and half lives indicating that more abundant mRNAs are degraded more rapidly [Esquerré et al., 2014, Esquerré et al., 2015, Nouaille et al., 2017].

The lag effect caused by mRNA competition was a striking result. Residual transcription is generally associated with delayed degradation, in particular for long genes or genes towards the 3' end of operons [Chen et al., 2015]. There is no direct evidence of this phenomenon in the literature, through measurements of the free or bound concentration of RNase E, for instance. To test this possibility, we analysed dynamical transcriptomics data in *E. coli*. We estimated the delay before degradation for 3140 mRNAs of this data set [Esquerré et al., 2014], as well as the maximal time needed to transcribe each of them using an elongation rate constant experimentally determined in [Chen et al., 2015]. Among the 2454 mRNAs that are not immediately degraded after rifampicin addition, 51% of them have a delay before degradation larger than the time needed for transcription. The delay is even twice for 21% of them (Figure 7.3). While this quick analysis clearly underestimates the number of mRNAs for which transcription elongation takes a shorter time than the delay before degradation, it indicates that residual transcription is not the sole determinant for the delay in the data set studied. Competition between mRNAs could well be another one. In a follow-up study, we verified the reality of the phenomenon of competition, by using data obtained in [Esquerré et al., 2014].

7.2 Integrative analysis of mRNA degradation

Esquerré *et al.* [Esquerré et al., 2014] showed that mRNA stability is one mechanism used by *E. coli* bacteria to adjust gene expression to their growth rate. Based on the data and our models, can we make hypotheses on the regulatory mechanisms at work? Is competition one of them? The principle of this study, which we are finalizing, is sketched in Fig. 7.1. We have shown in our study that nonlinear mixed effects (NLME) modelling [Lavielle, 2014] can be used to infer the parameters of the degradation model in (7.4) from dynamical transcriptomics data obtained by microarrays in *E. coli* cells growing at four different growth rates [Esquerré et al., 2014]. This framework generally yields good estimation results [Gonzalez et al., 2013].

In the NLME framework, we consider that the parameter values of individual mRNAs are drawn from a distribution common to the population of mRNAs. Concretely, we describe the time-series data

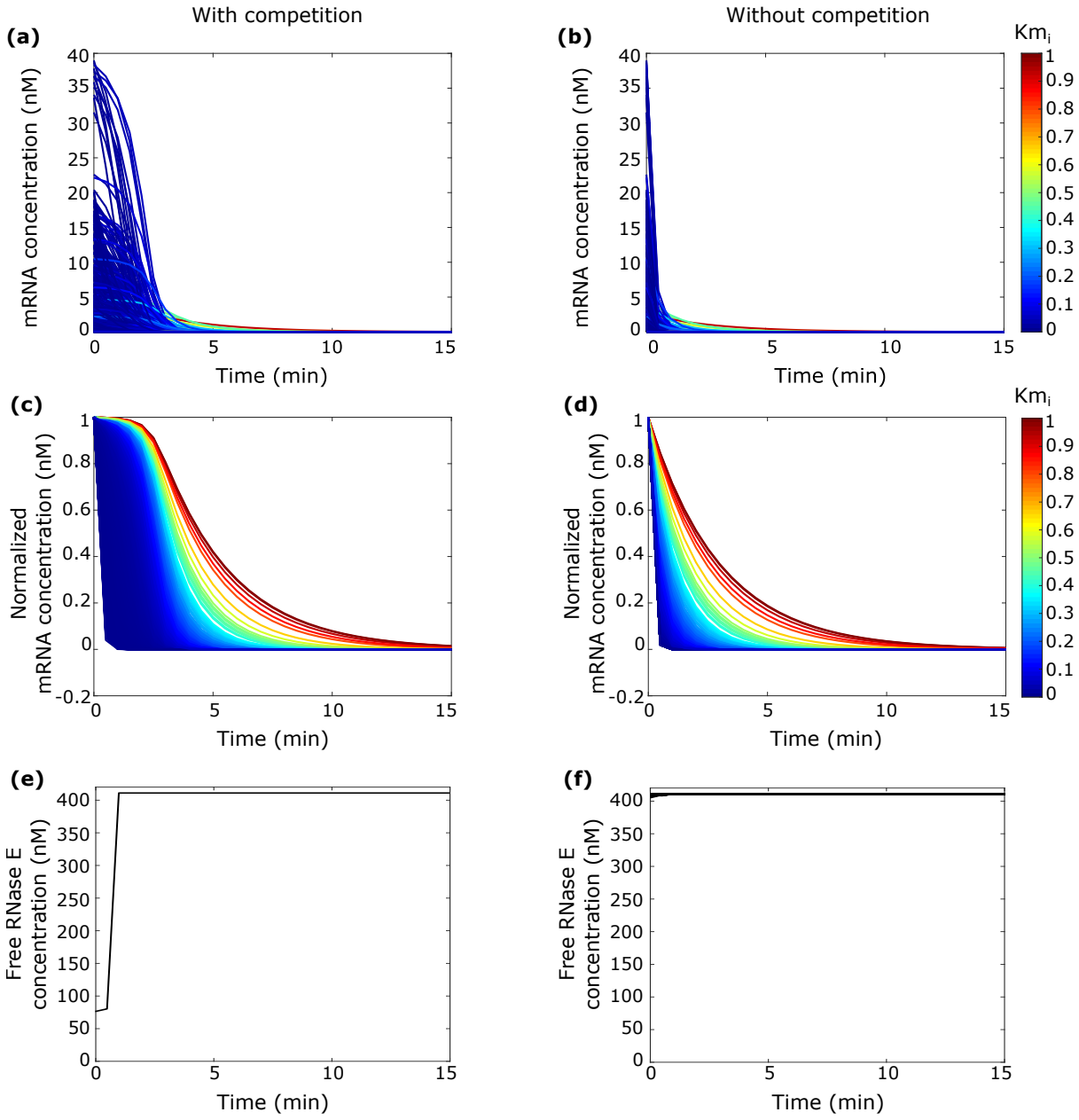


Figure 7.2 – Numerical simulation of mRNA degradation kinetics in isolated and competitive systems. Predicted profiles for the (a) competitive and (b) isolated systems. The profiles are normalized to their respective initial concentrations for the competitive (c) and isolated (d) systems. Predicted free RNase E concentrations for (e) the competitive system ($E_{free} = E_0 - \sum_{i=1}^{n} c_i(t)$) and (f) the isolated one ($E_{i,free} = E_0 - c_i(t)$). 4312 curves are displayed in this case, due to the lack of coupling between mRNAs. The colour bars on the right side represent the normalized gradient of Km values, on a scale from zero (the minimal Km value) to one (maximal value). From [Etienne et al., 2020].

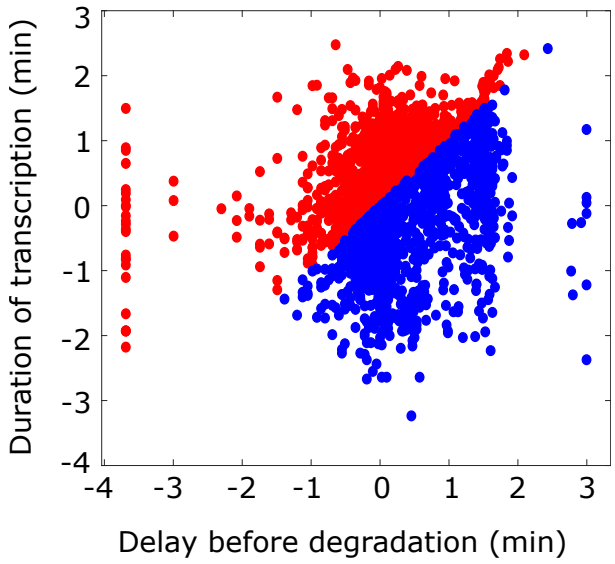


Figure 7.3 – Role of residual transcription in the retardation of degradation. On log scales and for each mRNA with a non negligible delay, the duration of transcription is plotted versus the delay before degradation. mRNAs with a delay before degradation higher than the time needed for transcription are displayed in blue and red, otherwise. From [Etienne et al., 2020].

with the following measurement model:

$$Y_{i,t_{j,\mu},\mu} = f(m_{i,\mu}(t_{j,\mu}), \Phi_{i,\mu}) + g(f(m_{i,\mu}(t_{i,j,\mu}), \Phi_{i,\mu}), \theta) \times \varepsilon_{i,j,\mu}. \tag{7.5}$$

Here, $Y_{i,t_{j,\mu},\mu}$ is the observed concentration of mRNA i at time $t_{j,\mu}$ in a given growth condition μ, $\varepsilon_{i,j,\mu} \sim \mathcal{N}(0,1)$ is the residual error, and the function g the residual error model with a vector of noise parameters θ. Function f is the solution to the ODE system (7.4). The vector of parameters $\Phi_{i,\mu}$ is a function of a vector of population parameters Φ_{pop}, fixed effects β_μ, and two vectors of random effects describing, respectively, the individual variability of mRNAs, $\eta_i \sim \mathcal{N}(0,\Sigma)$, resulting from differences in mRNA characteristics and regulations, and additional individual variability between growth conditions, $\eta_{i,\mu} \sim \mathcal{N}(0,\Omega)$:

$$\Phi_{i,\mu} = \Phi_{pop} + \beta_\mu \times \Phi_{pop} + \eta_i + \eta_{i,\mu}\,. \tag{7.6}$$

The estimation problem amounts to inferring the fixed effects (β_μ) and parameter distributions describing the population (via parameters Φ_{pop}, Σ, and Ω), from which, together with the data, are then derived the specific parameters for individual mRNAs. This approach allowed us to estimate the individual kinetic parameters of the 4254 *E. coli* mRNAs in the four growth conditions. The physiological interpretation of the parameters is under way. Current results indicate that competition between mRNAs for binding to RNase E is indeed a global regulatory mechanism adjusting mRNA stability to growth rate. Additional specific regulatory mechanisms mediated by small RNAs or RNA-binding proteins (HFQ, CSR system...) appear to fine tune the stability of no more than a fifth of the mRNAs [Etienne et al., In preparation].

7.3 Discussion and perspectives

The modelling of mRNA decay described in this chapter has allowed the identification of a new regulatory mechanism of mRNA degradation. It relies on the competition between mRNAs for their binding to RNase E. In the first minutes after the onset of degradation, the enzyme is titrated by the myriad of cellular mRNAs. Sole those with higher concentrations and/or higher affinities have more chances to be degraded. Competition on the one hand tends to stabilize mRNAs by increasing the

competitive effect at the whole-cell level. On the other hand, it affects mRNAs individually through the modification of their affinity: competition stabilizes mRNAs with low affinities and destabilizes those with high affinities. This global mechanism allows to adjust degradation rates to the intracellular mRNA concentrations. The latter are in large part regulated at the transcriptional level [Esquerré et al., 2014, Morin et al., 2020]. This means that degradation is directly coupled to transcription and that mRNA competition is the cornerstone of this regulatory mechanism. Surprisingly, less than 20% of cellular mRNAs are the target of additional specific regulations.

The observation that the competition between mRNAs is a major global control mechanism, while specific regulatory mechanisms concern a small fraction of cellular mRNAs, is an unexpected result. This is nevertheless reminiscent of the global changes of the gene expression machinery studied in Chapter 5 that contribute more to transcription than specific regulatory mechanisms. It provides another example that biological processes cannot be studied without considering the physiological state of the cell. This work on mRNA decay was recently highlighted in a press release for the general public by INRAE[1].

The NLME framework was critical in this study to allow the estimation of model parameters from microarray data and to cope with the noise in the data. While sub-sets of time-series data are generally analysed in studies of mRNA degradation, our approach is exhaustive since all cell mRNAs are included and their degradation profiles fitted. This allows to draw conclusions valid at the entire cell level. In addition, our approach is modular and scalable. It can be extended to other types of models and to other single-cell or population data, as well as additional heterogeneous data. For instance, in the context of the internship of Olivier Feudjio (Univ. de Paris), co-advised by Thibault Etienne and myself, we are currently applying our approach to new models and data sets, in order to study the role played by the localization of RNase E in mRNA stability.

One limitation of our approach, however, is the use of model simplification to remove the coupling of the ODEs introduced by the competition term. The price to pay is that the simplified model does no longer allow the analysis of the impact of extreme initial concentrations or Km values on bacterial mRNA decay. Although our study in [Etienne et al., 2020] suggests that this is not a key factor for the decay of most mRNAs, it would be interesting to study this specific point using the transcriptomic data. This requires additional developments of our approach that will be discussed in Chapter 8.

[1] https://www.inrae.fr/actualites/comprendre-comment-bacteries-sadaptent-leur-environnement

Chapter 8

Outlook

In this manuscript I have provided an overview of my past and current research activities at the interface of systems biology, bioinformatics, and microbiology. The material has been organized into five chapters around representative publications that I co-authored. As discussed in each chapter, the work has been carried out in the context of several research projects, in collaboration with a number of colleagues in France and abroad, and the co-supervision of PhD students and post-doctoral researchers. Additional research activities that have not been discussed in this manuscript relate to past and current collaborations with Andreas Kremling (TU München, Germany), Laurent Trilling (Univ. Grenoble Alpes), Tomas Gedeon (Montana State University, USA), Aline Métris and Jozsef Baranyi (formerly with Institute of Food Research, Norwich, UK), and Jean-Luc Gouzé (Inria Sophia Antipolis - Méditerranée). The evolution of the work described, from the qualitative to the quantitative modelling of biological systems, reflects the increased availability of quantitative data characterizing bacterial growth at multiple levels of cell organization. Most of my current and future research activities will be dedicated to the challenge of interpreting multiple heterogeneous data sets with mathematical models of bacterial growth and obtaining a more integrated view of cellular physiology. I describe these future directions of research below.

8.1 Genome-scale analysis of microbial physiology

8.1.1 Genome-scale analysis of cell metabolism

Foundations for this line of research have been laid by the on-going ANR project RIBECO (2018-2022), in which I am principal investigator for Inria Grenoble - Rhône-Alpes. In this project, we seek to learn more about the energetic burden imposed by the mRNA life cycle. Indeed, the continuing cycle of mRNA synthesis and degradation raises energetic constraints detrimental to cellular growth, in particular when the substrate consists of poor carbon sources. Our aim is to elucidate the connection between central carbon metabolism and RNA metabolism. While most of the work described in Chapter 6 relied on metabolomics data only, more data will be obtained in this project, such as transcriptomic and mRNA half-live data at the genome-wide level. This opens new opportunities to develop specific genome-scale models including both gene expression and metabolic pathways [O'Brien et al., 2013, Salvy and Hatzimanikatis, 2020]. Integration of these data sets is not an easy task as each comes with its own noise and bias, as seen in Chapter 4 for reporter gene and metabolomics data. I will therefore continue working on the problem of estimating biological quantities from primary data in collaboration with Eugenio Cinquemani and Hidde de Jong, as well as their implementation in user-friendly software.

Understanding the reprogramming of cell metabolism following a genetic or environmental perturbation is complicated. The main difficulty is to causally relate the predicted changes of fluxes to changes in gene expression and growth rate. A flux can vary directly with, for instance, the concentration of the enzyme catalysing the reaction and indirectly in response to the perturbation of another metabolic flux propagating through the network. I will therefore seek to develop approaches allowing one to identify at the genome-scale level the specific parts of metabolism directly affected by a perturbation. I have started investigating this question in collaboration with Marie-France Sagot and her team at Inria Lyon in the framework of the project MuSE, funded by the Complex

Systems Institute in Rhône-Alpes (IXXI), which I coordinated (2018-2020). We adapted a recently developed mixed-integer linear programming approach [Pusa et al., 2020], which we plan to generalize to the integration of multiple data sets (microarray and proteomics data, growth rates) and to apply to the problems studied in the context of RIBECO. Knowledge on RNA and energy metabolism obtained in this way should help us to design strategies that alleviate the energetic burden of the mRNA life cycle for biotechnological purposes. In particular, we will reengineer strains to help the efficient degradation of carbon sources derived from the pretreatment of agricultural and forestry residues.

The know-how developed for the analysis of metabolism and bacterial growth of *E. coli* can be applied to other organisms as well. I will work on two related projects in the coming months. For instance, Hidde de Jong and I have started a collaboration with Luiz de Carvalho from the Francis Crick Institute in London to analyse metabolic networks for the genus *Mycobacterium*, which we will reconstruct, and use these models to propose novel hypotheses on metabolic bottlenecks in the growth of mycobacteria. Another nascent project concerns the analysis of metabolic alterations during the development of the Parkinson disease, in collaboration with Florence Fauvelle from the Grenoble Institute of Neurosciences.

8.1.2 Genome-scale analysis of mRNA decay

The NLME framework generally yields good estimation results [Gonzalez et al., 2013] and it was critical in our integrative study of mRNA degradation. However, we had to overcome a number of issues relative to the noise of the data and identifiability problems, which may impede the extension of this approach to other high-throughput data sets taking into account additional biological processes, which is our objective in the RIBECO project. For instance in Equation 7.4, each reaction rate v_i for a given mRNA i is a saturating function of the concentration and kinetic parameters of mRNA i, but also of the other mRNAs. This introduces a coupling between the ODEs resulting in correlations between model parameters, which could be intensified with the inclusion of other biological mechanisms. To circumvent the problem in our study, we simplified the model to decouple the equations. In collaboration with Aline Marguet, recently recruited in our group, and Eugenio Cinquemani, I will consider alternative NLME approaches by explicitly modelling the source of correlations introduced by the competition between mRNAs, based on an approach described in [Marguet et al., 2019]. Alternative approximations of the coupled models are also currently investigated in the context of the post-doctorate of Thibault Etienne.

Extending our approach to other data sets is not trivial, as each comes with its own noise. Another line of research will thus concern the development of appropriate preprocessing steps and error models in order to use heterogeneous data sets. Altogether this work should allow the exploitation of various time-series data sets of *E. coli* growth, metabolism, and gene expression obtained at TBI in the framework of the ANR RIBECO. A deeper understanding of the observed connections between mRNA degradation and the central carbon metabolism, and the regulatory mechanisms involved, would help us reach our final goal in the RIBECO project. That is, to propose changes in the life cycle of specific mRNAs and thus improve the degradation of vegetal biomass by microorganisms.

8.2 Resource allocation strategies in natural and engineered microorganisms

This direction of research is a development of our work in Chapter 5, in which we described a growth switch enabling the reallocation of nutrient resources from bacterial growth to the production of compounds. In order to understand the functioning of the growth switch on the molecular level, I am currently developing a mechanistic model of the gene expression machinery in *E. coli*, in collaboration with Hidde de Jong, Hans Geiselmann, and Jean-Luc Gouzé at Inria Sophia Antipolis - Méditerranée. In line with previous efforts in this direction [Dourado and Lercher, 2020, Weiße et al., 2015], the idea

is to provide a coarse-grained picture of the different macromolecular components, completed with the addition of RNA polymerase and its external control.

While many experimental studies monitored the macromolecular composition of *E. coli* cells growing at various steady-state growth rates, dynamical data for the gene expression machinery are scarce, if not absent. I will conduct experiments in the laboratory of Hans Geiselmann at LIPhy to provide us with original data on the dynamic adaptation of macromolecular cell composition to improve model estimation and prediction. A more detailed understanding of the growth switch will leverage the development of optimal strategies for producing metabolites or (heterologous) proteins, but on the fundamental level it may also provide novel insights into the dynamics of the adaptation of gene expression to environmental perturbations. Moreover, in the context of the RIBECO project, we expect the model to be useful for the design of strategies that tune the life cycle of specific mRNAs.

Another related line of research concerns the analysis of resource allocation strategies in bacteria. The foundations for this research axis have been laid by the ANR project Maximic (2017-2022), which involves modellers and experimentalists from our project-team, as well as other members of LIPhy, and specialists in control theory from Inria Sophia Antipolis - Méditerrannée (BIOCORE and McTAO project-teams), as well as Tomas Gedeon from Montana State University (USA). In this study, cells are viewed as self-replicators that try to grow optimally. The models in this case do not piece together all known biochemical reactions, but provide a coarse-grained picture of key cellular functions that captures the major fluxes of material and energy passing through the cell and fuelling growth. Such models may be instrumental for explaining a fundamental trade-off between rate and yield in the growth of microorganisms, that is, the fact that in microorganisms rapid growth generally comes at the cost of less efficient growth [Lipson, 2015].

8.3 From project-team IBIS to MICROCOSME

Inria project-teams have a maximal duration of twelve years. Our current project-team IBIS is hence terminating its life. With former permanent Ibis members (Hidde de Jong, Eugenio Cinquemani, Aline Marguet, Hans Geiselmann) and Muriel Cocaign-Bousquet, we are creating a new Inria project-team, MICROCOSME. I will take the lead of this new team, whose creation is currently under instruction. The on-going and future directions of research described are part of two of the four research axes of MICROCOSME. The other two axes are concerned with the analysis of the variability of bacterial growth, and heterogeneity within communities consisting of different microbial species and control of communities for biotechnological applications.
The start of a new scientific adventure...

Appendix

Curriculum Vitae

Delphine ROPERS

Born 29/11/1975 in Essey-lès-Nancy (54)
French nationality
Married, two children

Current position

Position: Research scientist at Inria
Professional address: Centre de recherche Grenoble - Rhône-Alpes, Inovallée, 655 Avenue de l'Europe - CS 90051, 38334 Montbonnot Cedex, France
Phone: +33 4 76 61 53 72
E-mail: delphine.ropers@inria.fr
Web: https://team.inria.fr/ibis/delphine-ropers/

Professional Experience

2017 -	Research scientist (Chargée de recherche classe normale) at Inria, Grenoble - Rhône-Alpes research centre (project-team Ibis)
2008 - 2017	Research scientist - first grade (Chargée de recherche 1re classe) at Inria, Grenoble - Rhône-Alpes research centre (project-team Ibis)
2006 - 2008	Research scientist - second grade (Chargée de recherche 2e classe) at Inria, Grenoble - Rhône-Alpes research centre (project-team Helix)
2003 - 2006	Post-doctoral researcher at Inria, Grenoble - Rhône-Alpes research centre (project-team Helix)
2002 - 2003	Doctoral researcher at CNRS Nancy
2001 - 2002	Temporary lecturer (ATER) at Université de Nancy I (96h)
1998 - 2001	Doctoral researcher at Université Nancy I
1998 - 2001	Instructor (monitorat) at Université Nancy I

Academic Education

1998 - 2003	Doctoral thesis in Molecular and Cellular Biology at Laboratoire de Maturation des ARN et Enzymologie Moléculaire at CNRS/Université Nancy I. Title of the thesis: Experimental study of the role of SR proteins in the regulation of the HIV-1 virus RNA splicing, responsible of the human immunodeficiency, and mathematical modelling of these regulations.
1996 - 1998	Master of Science in Biochemistry at University of Nancy I. Title of the thesis: "Etude des éléments activant ou inhibant en *cis* l'épissage de l'ARN du virus VIH-1".
1993 - 1996	Bachelor of Science in Biochemistry at University of Nancy I.

Scientific Production

Articles in preparation

1. T.A. Etienne, L. Girbal, E. Cinquemani, M. Cocaign-Bousquet, **D. Ropers**. Integrative analysis of the physiological control of bacterial mRNA decay.

Submitted articles

1. **D. Ropers**, Y. Couté, L. Faure, S. Ferré, D. Labourdette, A. Shabani, L. Trouilh, P. Vasseur, G. Corre, M. Ferro, M.-A. Teste, J. Geiselmann, H. de Jong. A multi-omics study of bacterial growth arrest in a synthetic biology application.

2. C. Roux, T.A. Etienne, E. Hajnsdorf, **D. Ropers**, A.J. Carpousis, M. Cocaign-Bousquet, L. Girbal. The essential role of mRNA degradation in understanding and engineering *E. coli* metabolism.

Peer-reviewed articles

1. T.A. Etienne, M. Cocaign-Bousquet, **D. Ropers** (2020). Competitive effects in bacterial mRNA decay. *Journal of Theoretical Biology*, 504:110333.

2. M. Morin, B. Enjalbert, **D. Ropers**, L. Girbal, and M. Cocaign-Bousquet (2020). Genome-wide stabilization of mRNA during a 'feast-to-famine' growth transition in *Escherichia coli*. *mSphere*, 5:e00276-20.

3. S. Pinhal, **D. Ropers**, J. Geiselmann, H. de Jong. Acetate metabolism and the inhibition of bacterial growth by acetate (2019). *Journal of Bacteriology*, 201:e00147-19.

4. I. Belgacem, S. Casagranda, E. Grac, **D. Ropers**, J.-L. Gouzé (2018). Reduction and stability analysis of a transcription-translation model of RNA polymerase. *Bulletin of Mathematical Biology*, 80(2), 294-318.

5. A. Kremling, J. Geiselmann, **D. Ropers**, H. de Jong (2018). An ensemble of mathematical models showing diauxic growth behaviour. *BMC Systems Biology*, 12:82.

6. S. Casagranda, S. Touzeau, **D. Ropers**, J.-L. Gouzé (2018). Principal process analysis of biological models. *BMC Systems Biology*, 12(1):68.

7. H. de Jong, S. Casagranda, N. Giordano, E. Cinquemani, **D. Ropers**, J. Geiselmann, J.-L. Gouzé (2017). Mathematical modeling of microbes: Metabolism, gene expression, and growth. *Journal of the Royal Society Interface*, 14:20170502.

8. M. Morin, **D. Ropers**, E. Cinquemani, J.C. Portais, B. Enjalbert, M. Cocaign-Bousquet (2017). The Csr system regulates *Escherichia coli* fitness by controlling glycogen accumulation and energy levels. *mBio*, 8(5): e01628-17.

9. H. de Jong, **D. Ropers**, J. Geiselmann (2017). Resource reallocation in bacteria by reengineering the gene expression machinery. *Trends in Microbiology*, 25(6):480-493.

10. E. Cinquemani, V. Laroute, M. Cocaign-Bousquet, H. de Jong, D. Ropers (2017). Estimation of time-varying growth, uptake and excretion rates from dynamic metabolomics data. *Bioinformatics*, 33(14):i301-i310.

11. A. Métris, S.M. George, **D. Ropers** (2017). Piecewise linear approximations to model the dynamics of adapation to osmotic stress by food-borne pathogens. *International Journal of Food Microbiology*, 240:63-74.

12. **D. Ropers** and A. Métris. Osmotic stress response to NaCl in *Escherichia coli*: qualitative modeling and simulation data (2016), *Data in Brief*, 9: 606-612.

13. M. Morin, **D. Ropers**, F. Létisse, S. Laguerre, J.C. Portais, M. Cocaign-Bousquet, B. Enjalbert (2016). The post-transcriptional regulatory system CSR controls the balance of metabolic pools in upper glycolysis of *Escherichia coli*. *Molecular Microbiology*, 100(4):686-700.

14. J. Izard, CDC Gomez Balderas, **D. Ropers**, S. Lacour, X. Song, Y. Yang, AB Lindner, J. Geiselmann, H. de Jong (2015). A synthetic growth switch based on controlled expression of RNA polymerase *Molecular Systems Biology*, 11(840).

15. A. Kremling, J. Geiselmann, **D. Ropers**, H. de Jong, Understanding carbon catabolite repression in *Escherichia coli* using quantitative models (2015). *Trends in Microbiology*, 23(2):99-109.

16. V. Zulkower, M. Page, **D. Ropers**, J. Geiselmann, H. de Jong (2015), Robust reconstruction of gene expression profiles from reporter gene data using linear inversion. *Bioinformatics*, 15;31(12):i71-9

17. M. Trauchessec, M. Jaquinod, A. Bonvalot, V. Brun, C. Bruley, **D. Ropers**, H. de Jong, J. Garin, G. Bestel-Corre, M. Ferro (2014), Mass spectrometry-based workflow for accurate quantification of *E. coli* enzymes : how proteomics can play a key role in metabolic engineering, *Molecular and Cellular Proteomics*. 13(4):954-968.

18. G. Baptist, C. Pinel, C. Ranquet, J. Izard, **D. Ropers**, H. de Jong, J. Geiselmann (2013), A genome-wide screen for identifying all regulators of a target gene, *Nucleic Acids Research*, 41(17) :e164.

19. S. Berthoumieux, H. de Jong, G. Baptist, C. Pinel, C. Ranquet, **D. Ropers**, J. Geiselmann (2013), Shared control of gene expression in bacteria by transcription factors and global physiology of the cell, *Molecular Systems Biology*, 9:634. Editors' choice in Science

20. V. Baldazzi, **D. Ropers**, J. Geiselmann, D. Kahn, H. de Jong (2012), Importance of metabolic coupling for the dynamics of gene expression following a diauxic shift in *Escherichia coli*, *Journal of Theoretical Biology*, 295: 100-115.

21. G. Batt, B. Besson, P.-E. Ciron, H. de Jong, E. Dumas, J. Geiselmann, R. Monte, P.T. Monteiro, M. Page, F. Rechenmann, **D. Ropers** (2012), Genetic Network Analyzer: A tool for the qualitative modeling and simulation of bacterial regulatory networks, J. van Helden, A. Toussaint, D. Thieffry (eds), Bacterial Molecular Networks: Methods and Protocols, *Methods in Molecular Biology*, Humana Press, Springer, New York, 439-462.

22. P.T. Monteiro, P.J. Dias, **D. Ropers**, A.L. Oliveira, I. Sá-Correia, M.C. Teixeira and A.T. Freitas (2011), Qualitative modeling and formal verification of the FLR1 gene mancozeb response in *Saccharomyces cerevisiae*, *IET Systems Biology*, 5(5): 308-316.

23. **D. Ropers**, V. Baldazzi, H. de Jong (2011), Model reduction using piecewise-linear approximations preserves dynamic properties of the carbon starvation response in *Escherichia coli*, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 8(1):166-181.

24. V. Baldazzi, **D. Ropers**, Y. Markowicz, D. Kahn, J. Geiselmann, H. de Jong (2010), The carbon assimilation network in *Escherichia coli* is densely connected and largely sign-determined by directions of metabolic fluxes, *PLoS Computational Biology*, 6(6):e1000812.

25. H. de Jong, C. Ranquet, **D. Ropers**, C. Pinel, J. Geiselmann (2010), Experimental and computational validation of models of fluorescent and luminescent reporter genes in bacteria, *BMC Systems Biology*, 4:55.

26. F. Boyer, B. Besson, G. Baptist, J. Izard, C. Pinel, **D. Ropers**, J. Geiselmann, H. de Jong (2010), A MATLAB program for the analysis of fluorescence and luminescence reporter gene data, *Bioinformatics*, 26(9):1262-1263.

27. F. Corblin, S. Tripodi, E. Fanchon, **D. Ropers**, L. Trilling (2009), A declarative constraint-based method for analyzing discrete genetic regulatory networks, *BioSystems*, 98(2):91-104.

28. J.-M. Saliou, C.F. Bourgeois, L. Ayadi-Ben Mena, **D. Ropers**, S. Jacquenet, V. Marchand, J. Stévenin and C. Branlant (2009), Role of RNA structure and protein factors in the control of HIV-1 splicing, *Frontiers in Biosciences*, 14:2714-2729.

29. P.T. Monteiro, **D. Ropers**, R. Mateescu, A.T. Freitas, H. de Jong (2008), Temporal logic patterns for querying dynamic models of cellular interaction networks, *Bioinformatics*, 24(16): i227-i233. Special issue ECCB-2008.

30. H. de Jong, **D. Ropers** (2006), Strategies for dealing with incomplete information in the modeling of molecular interaction networks, *Briefings in Bioinformatics*, 7(4):354-363.

31. H. Hallay, N. Locker, L. Ayadi, **D. Ropers**, E. Guittet, C. Branlant (2006), Biochemical and NMR study on the competition between proteins SC35, SRp40, and heterogeneous nuclear ribonucleoprotein A1 at the HIV-1 Tat exon 2 splicing site, *Journal of Biological Chemistry*, 281(48):37159-74.

32. **D. Ropers**, H. de Jong, M. Page, D. Schneider, J. Geiselmann (2006), Qualitative simulation of the carbon starvation response in *Escherichia coli*. *BioSystems*, 84(2):124-152.

33. G. Batt, **D. Ropers**, H. de Jong, J. Geiselmann, R. Mateescu, M. Page and D. Schneider (2005), Validation of qualitative models of genetic regulatory networks by model checking: Analysis of the nutritional stress response in *Escherichia coli*. *Bioinformatics*, 21(Suppl 1):i19-i28, special issue ISMB-2005.

34. D. Eveillard, **D. Ropers**, H. de Jong, C. Branlant, A. Bockmayr (2004), A multi-scale constraint programming model of alternative splicing regulation, *Theoretical Computer Science*, 135(1): 3-24.

35. **D. Ropers**, L. Ayadi, R. Gattoni, S. Jacquenet, L. Damier, C. Branlant, J. Stévenin (2004), Differential effects of the SR proteins 9G8, SC35, ASF/SF2 and SRp40 on the utilization of the A1 to A5 splicing sites of HIV-1 RNA, *Journal of Biological Chemistry*, 279(29): 29963-29973.

36. S. Jacquenet, **D. Ropers**, P. Bilodeau, L. Damier, A. Mougin, M. Stoltzfus, C. Branlant (2001), Conserved stem-loop structures in the HIV-1 RNA region containing the A3 3' splice site and its cis-regulatory element: possible involvement in RNA splicing, Nucleic Acids Research, 29(2): 464-478.

Conference papers

1. S. Casagranda, **D. Ropers**, J.-L. Gouzé (2015), Model reduction and process analysis of biological models. *Proceedings of 23rd Mediterranean Conference on Control and Automation* (MED 2015)

2. I. Belgacem, E. Grac, **D. Ropers**, J.-L. Gouzé (2014), *Proceedings of 21st International Symposium on Mathematical Theory of Networks and Systems*, Jul 2014, pp 1383-1386, Groningen, Netherlands.

3. **D. Ropers**, V. Baldazzi, H. de Jong (2009), Reduction of a kinetic model of the carbon starvation response in Escherichia coli, in *Proceedings of the 15th IFAC Symposium on System Identification*, SYSID 2009.

4. P.T. Monteiro, **D. Ropers**, R. Mateescu, A.T. Freitas, H. de Jong (2008). Temporal logic patterns for querying qualitative models of genetic regulatory networks. In: M. Ghallab, C.D. Spyropoulos, N. Fakotakis, N. Avouris (eds.), *Proceedings of 18th European Conference on Artificial Intelligence (ECAI 2008)*, IOS Press, Amsterdam, 229-233

5. **D. Ropers**, H. de Jong, J.-L. Gouzé, M. Page, D. Schneider, J. Geiselmann (2005), Piecewise-linear models of genetic regulatory networks : Analysis of the carbon starvation response in *Escherichia coli*. *Proceedings of ECMTB 2005, Mathematical Modeling of Biological Systems*, Volume I. A. Deutsch, L. Brusch, H. Byrne, G. de Vries and H.-P. Herzel (eds), Birkhäuser, Boston, 83-96.

6. G. Batt, H. de Jong, J. Geiselmann, M. Page, **D. Ropers**, D. Schneider (2005). Qualitative analysis and verification of hybrid models of genetic regulatory networks: Nutritional stress response in *Escherichia coli*. M. Morari, L. Thiele (eds), *Hybrid Systems: Computation and Control (HSCC 2005)*, *Lecture Notes in Computer Science*, 3414, Springer-Verlag, Berlin, 134-150

7. G. Batt, H. de Jong, J. Geiselmann, M. Page, **D. Ropers**, D. Schneider (2005). Qualitative analysis and verification of hybrid models of genetic regulatory networks: Nutritional stress response in Escherichia coli. M. Morari, L. Thiele (eds), Hybrid Systems: Computation and Control (HSCC 2005), *Lecture Notes in Computer Science* 3414, Springer-Verlag, Berlin, 134-150

8. D. Eveillard, **D. Ropers**, H. de Jong, C. Branlant, A. Bockmayr (2003), Multiscale modeling of alternative splicing regulation, C. Priami (ed.), *Computational Methods in Systems Biology (CMSB03)*, *Lecture Notes in Computer Science* 2602, Springer-Verlag, Berlin, 75-87.

Book and book chapters (in English and in French)

1. S. Bourgoin-Voillard, C. Durmort, J. Geiselmann, A. Le Gouëllec, **D. Ropers**, M. Sève (2015). La cellule procaryote en biotechnologies in *Les biotechnologies en santé – Tome 1 : Introduction aux biotechnologies en santé*. Coordonnateurs : S. Bourgoin-Voillard, W. Rachidi, M. Sève. Ed. Lavoisier.

2. V. Baldazzi, P.T. Monteiro, M. Page, **D. Ropers**, J. Geiselmann, H. de Jong (2011), Qualitative analysis of genetic regulatory networks in bacteria, W. Dubitzky, J. Southgate, H. Fuss (eds.), *Understanding the Dynamics of Biological Systems: Lessons Learned from Integrative Systems Biology*, Springer-Verlag, Berlin, pp 111-130.

3. **D. Ropers** (2011), *De la complexité génomique à la diversité protéique - Analyse par modélisation et expériences de la régulation de l'épissage alternatif de l'ARN du virus VIH-1*, Editions Universitaires Européennes, Saarbrücken (Germany), ISBN 978-613-1-56368-3.

4. G. Batt, R. Casey, H. de Jong, J. Geiselmann, J.-L. Gouzé, M. Page, **D. Ropers**, T. Sari, D. Schneider (2006). Qualitative analysis of the dynamics of genetic regulatory networks using piecewise-linear models. In: E. Pecou, S. Martinez, A. Maass (eds.), *Mathematical and Computational Methods in Biology*, Editions Hermann, Paris, 206-239

5. H. de Jong, **D. Ropers** (2006), Qualitative approaches towards the analysis of genetic regulatory networks, Z. Szallasi, V. Periwal, J. Stelling (eds), *System Modeling in Cellular Biology: From Concepts to Nuts and Bolts*, MIT Press, Cambridge, MA, 125-148.

Proceedings and journal papers in French, science popularization articles

1. D. Ropers, with T. Vieville (2019), Mais comment éduquer les garçons à l'équité des genres au niveau informatique et numérique ? Colloque "Un rêve pour les filles et les garçons : la SCIENCE", Colloque des associations Parité Science, Femmes & Sciences, de l'Union des professeurs de physique et chimie de l'académie de Grenoble et de la section Alpes de la Société Française de Physique, Grenoble, 9 novembre 2019, p55-70.

2. G. Batt, H. de Jong, J. Geiselmann, J.-L. Gouzé, M. Page, **D. Ropers**, T. Sari, D. Schneider (2007), Analyse qualitative de la dynamique de réseaux de régulation génique par des modèles linéaires par morceaux, *Technique et Science Informatique*, 26(1-2):11-45.

3. **D. Ropers**, H. de Jong, J. Geiselmann. Modélisation de la réponse au stress nutritionnel de la bactérie *Escherichia coli*, *Biofutur*, 275:36-39, 2007

4. H. de Jong, **D. Ropers**, C. Chaouiya, D. Thieffry. Modélisation, analyse et simulation de réseaux de régulation génique. *Biofutur*, 252:36-40, 2005

Patent

Johannes Geiselmann, Hidde de Jong, Delphine Ropers, Jérôme Izard, Method for producing metabolites, peptides and recombinant proteins, EP3047031A1, US9816123B2, WO2015036622A1
https://patents.google.com/patent/US9816123B2/en

Participation in projects

2018 - 2022	ANR project (from the French National Research Agency) RIBECO "Engineering RNA life cycle to optimize economy of microbial energy: application to the bioconversion of biomass-derived carbon sources". Principal Investigator (PI), work-package leader.
2018 - 2022	ANR project Maximic "Optimal control of microbial cells by natural and synthetic strategies". Participant.
2018 - 2020	Funding from the Institut of Complex Systems IXXI and the Federation of Systems Biology BioSyl "Multi-Omics and Metabolic models iNtegration to study growth Transition in *Escherichia coli*". Coordinator.
2016 - 2020	PhD grant INRA-Inria and Inria/ANR funding. Main PhD advisor (co-adviser with Muriel Cocaign-Bousquet).
2016 - 2020	ANR project MEMIP "Mixed-Effects Models of Intracellular Processes - Methods, Tools and Applications". Participant.
2012 - 2017	Projet Investissement d'Avenir RESET (Investments for the Future programme of ANR) "Arrest and restart of the gene expression machinery » (2012-2017 ; https://project.inria.fr/reset/fr/). Co-PI, work-package leader.
2012 - 2017	Contrat Jeune Scientifique INRA-Inria (PhD grant and two year-postdoctoral research fellowship for Manon Morin; 2012 – Octobre 2017). Co-advisor with Muriel Cocaign-Bousquet.
2010 - 2013	ANR project GemCo "Model reduction, experimental validation, and control for the gene expression machinery in *E. coli*". WP leader.
2008 - 2013	Action d'Envergure Inria ColAge "Natural and engineering solutions to the control of bacterial growth and aging: A systems and synthetic biology approach". Participant.
2006 - 2009	FP6 european project EC-MOAN "Scalable Modeling and Analysis Techniques to Study Emergent Cell Behavior - understanding the *E. coli* stress response". Participant.
2006 - 2009	ANR project MetaGenoReg "Towards an understanding of the interrelations between metabolic and gene regulation: *E. coli* carbon metabolism as a test case". Participant.

Teaching

2020 -	Teaching on data integration in course "Artificial Intelligence for Omics", Master 1 AI for Health, Univ Grenoble-Alpes, 16h/year.
2013 -	Course "Modelling in Systems Biology", 2nd year student at Phelma school - Grenoble INP/Master 1 Nanobiotechnology - Univ Grenoble-Alpes, 16h/year. Responsible of course.
2012 - 2018	Lecture on mathematical modelling of biological systems, 5th year biotechnology students at INSA Toulouse, 4h/year.
2007 -	Teaching on cell systems modelling in course "Molecular tools for health", Master 1 Health engineering, Univ Grenoble-Alpes, 14h/year.
2006 - 2008	Course Modelling of Gene Regulatory Networks, PhD Program in Computational Biology, Instituto Gulbenkian de Ciencia, Lisbon (Portugal), 1 week/year
2004 - 2008	Teaching in Bioinformatics course, Master 1 Computer Science, Univ Joseph Fourier (Grenoble), 13h/year.
2004 - 2007	Teaching in course "Modelling of metabolism", Master 2 aMIV, Univ Claude Bernard (Lyon), 5h/year.

Supervision

PhD students

Thibault Etienne	Doctoral school Evolution Ecosystems Microbiology Modelling (E2M2), Univ Lyon. Co-advisor with Muriel Cocaign-Bousquet (principal advisor). Defence: November 2020.
Stefano Casagranda	Doctoral school Information and Communication sciences and technologies (STIC), Univ Nice. Co-advisor with Jean-Luc Gouzé. Defence: June 2017.
Manon Morin	Doctoral school Eco-Agro-Bio-Sciences (SEVAB), Univ Paul Sabatier, Toulouse. Co-advisor with Muriel Cocaign-Bousquet. Defence: November 2015.
Stéphane Pinhal	Doctoral school of chemistry and life sciences (CSV), Univ Joseph Fourier, Grenoble. Co-advisor with Johannes Geiselmann and Hidde de Jong. Defence: March 2015.
Valentin Zulkower	Doctoral school of Mathematics, Information Sciences and Technologies, and Computer Science (MSTII), Univ Joseph Fourier, Grenoble. Co-advisor with Johannes Geiselmann and Hidde de Jong. Defence: March 2015.
Jérôme Izard	Doctoral school of chemistry and life sciences (CSV), Univ Joseph Fourier, Grenoble. Co-advisor with Johannes Geiselmann and Stéfan Lacour. Defence: December 2012.

BSc and MSc students

Amélie Caddéo	M1, Univ Grenoble - Alpes. Defence: June 2021.
Olivier Feudjio	M1, Univ de Paris. Co-supervision with T. Etienne. Defence: June 2021.
Tommy Burnoud	M1, Univ Grenoble - Alpes. Defence: June 2020.
Arieta Shabani	M1, Univ Grenoble - Alpes. Co-supervision with H. de Jong. Defence: June 2019.
Eric Cumunel	M2, Univ Lyon 1. Co-supervision with M.-F. Sagot. Defence: July 2019.
Naina Goel	M1, Univ . Paris-Dauphine. Co-supervision with H. de Jong. Defence: June 2018.
Alex Uchenna Anyaegbunam	M2, Univ Paris-Dauphine. Co-supervision with E. Cinquemani. Defence: June 2016.
Keerthi Kurma	M1, Phelma - Grenoble INP. Co-supervision with E. Cinquemani. Defence: June 2015.
Julien Sauvage	2nd year, Phelma - Grenoble INP. Defence: October 2014.
iGEM Grenoble-EMSE-LSU team	MSc students, Univ Joseph Fourier, Grenoble INP, ENSM Saint-Etienne & Louisiana State University. iGEM 2013 competition.
Nils Giordano	M2, ENS Paris & Univ Pierre et Marie Curie. Defence: June 2012.
iGEM Grenoble team	MSc students, Univ Joseph Fourier, Grenoble INP & Polytech Grenoble. iGEM 2012 competition.
Stéphane Pinhal	M2, Univ Pierre et Marie Curie. Defence: June 2011.
iGEM Grenoble team	MSc students, Univ Joseph Fourier & Grenoble INP. iGEM 2011 competition.
Dishank Gupta	BSc, Institute of Technology, Banaras Hindu University, India. Defence: August 2010.
Ying Song	M1, Univ Joseph Fourier, Grenoble. Defence: July 2010.
Vaibhhav Sinha	MSc, IIT Kharagpur, India. Defence: August 2009.
Mohammed El Amine Youcef	M1, Univ Joseph Fourier Grenoble. Defence: July 2009.
Yan Cao	MSc, Zhejiang University, China. Defence: March 2009.

Scientific and Administrative Duties

2019 -	Co-coordinator of mentoring program
2014 -	"Référent-chercheur" at Inria Grenoble - Rhône-Alpes
2009 -	First-aid rescue worker
2017 - 2018	Member of Inria strategic plan working group
2015 - 2019	Nominated member of the Inria Evaluation Committee
2014 - 2018	Member of the Comité d'Etudes Doctorales at Inria Grenoble - Rhône-Alpes
2010 - 2015	Member of the Commission de Formation Permanente at Inria Grenoble - Rhône-Alpes
2007 - 2015	Representant of Inria in the scientific board of the Complex Systems Institute of Lyon (IXXI)
2007 - 2019	Member of the steering committee of the Rhône-Alpes Seminar on Modeling in the Life Sciences SEMOVI

Committees

PhD progress committee

Charlotte Roux	Univ Paul Sabatier, Toulouse, 2018-2021
Manon Barthe	Univ Paul Sabatier, Toulouse, 2017-2021
Irene Ziska	Univ Claude Bernard, Lyon, 2017-2020
Martin Wannagat	Univ Claude Bernard, Lyon, 2012-2015
Alice Julien-Laferrière	Univ Claude Bernard, Lyon, 2012-2015
Claire Villiers	Univ Joseph Fourier, 2010-2013
Sirichai Sunya	Univ Paul Sabatier, 2007-2010

PhD committee

Thibault Etienne	PhD co-advisor. Univ Claude Bernard, Lyon. December 2020
Joël Espel	Invited member. Univ Grenoble-Alpes. October 2020.
Ronan Duchesne	Examiner. ENS Lyon. December 2019.
Marianyela Petrizzelli	Examiner. Univ Paris-Sud. July 2019.
Stefano Casagranda	PhD co-advisor. Univ Nice. June 2017.
Manon Morin	PhD co-advisor. Univ Paul Sabatier, Toulouse. November 2015
Ismail Belgacem	Examiner. Univ Joseph Fourier, Grenoble. March 2015
Stéphane Pinhal	PhD co-advisor. Univ Joseph Fourier, Grenoble. March 2015
Valentin Zulkower	PhD co-advisor. Univ Joseph Fourier, Grenoble. March 2015
Anna Zukhova	Examiner. Univ Bordeaux. December 2014
Claire Villiers	Examiner. Univ Joseph Fourier, Grenoble. October 2013
Jérôme Izard	PhD co-advisor. Univ Joseph Fourier, Grenoble. December 2012

Selection committee

2021	Inria Starting and Advanced Research positions
2020	CR Inria admission panel
	CR Inria Grenoble - Rhône-Alpes
2019	INRA-Inria PhD grants
	CR Inria Bordeaux Sud-Ouest

	CR Inria (national)
	Assistant Professor, Univ Rennes
2018	Assistant Professor, INSA Lyon
	CR Inria Grenoble - Rhône-Alpes
	CR Inria (national)
	INRA-Inria PhD grant
2017	CR2 Inria Sophia-Antipolis - Méditerranée
	CR1 Inria
	Assistant Professor, INSA Lyon
	PhD grant INSERM-Inria
2016	CR Inria admission panel
	CR Inria Lille - Nord-Europe
2015	Assistant Professor, INSA Lyon
	CR Inria admission panel
	Assistant Professor, INSA Lyon
	Assistant Professor, INSA Lyon
2008	IR INRA

Reviewing activities and program committees

Funding agencies	ANR, BMBF (Germany), DAAD (Germany), BBSRC (UK), NWO (The Netherlands), NSC (Poland), Univ Grenoble-Alpes, PEPS
Journal articles	Nature Communications, Cell Systems, Bioinformatics, Biosystems, Current Biotechnology, BMC Systems Biology, PLoS One, Journal of Theoretical Biology, Biophysical Journal...
Books and book chapters	CRC Press, Wiley, Springer
Program committees	JOBIM (2007, 2008, 2010, 2011, 2017, 2021), CSBio (2019, 2020), ECCB 2020

Scientific animation

2018 –	Participation to the creation of the community "Microbial bioinformatics" for the European infrastructure ELIXIR
2008 – 2014	Creation and co-organization of the Grenoble seminar on Complex Systems
2011	Co-organization with Eugenio Cinquemani of workshop on Identification and Control of Biological Interaction Networks
2007 – 2017, 2019	Steering committee member of Modelling in Life Sciences seminar in Rhône-Alpes

Selected articles

The articles that I highlighted in the manuscript are listed below, together with their abstract.

[Ropers et al., 2011] **Model reduction using piecewise-linear approximations preserves dynamic properties of the carbon starvation response in *Escherichia coli***. Delphine Ropers, Valentina Baldazzi, Hidde de Jong. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 8(1), 166-181, 2011.
The adaptation of the bacterium *Escherichia coli* to carbon starvation is controlled by a large network of biochemical reactions involving genes, mRNAs, proteins, and signalling molecules. The dynamics of these networks is difficult to analyze, notably due to a lack of quantitative information on parameter values. To overcome these limitations, model reduction approaches based on quasi-steady-state (QSS) and piecewise-linear (PL) approximations have been proposed, resulting in models that are easier to handle mathematically and computationally. These approximations are not supposed to affect the capability of the model to account for essential dynamical properties of the system, but the validity of this assumption has not been systematically tested. In this paper we carry out such a study by evaluating a large and complex PL model of the carbon starvation response in *E. coli* using an ensemble approach. The results show that, in comparison with conventional nonlinear models, the PL approximations generally preserve the dynamics of the carbon starvation response network, although with some deviations concerning notably the quantitative precision of the model predictions. This encourages the application of PL models to the qualitative analysis of bacterial regulatory networks, in situations where the reference time-scale is that of protein synthesis and degradation.

[Ropers et al., 2006] **Qualitative simulation of the carbon starvation response in *Escherichia coli***. Delphine Ropers, Hidde de Jong, Michel Page, Dominique Schneider and Johannes Geiselmann. *Biosystems* 84 (2), 124-152, 2006.
In case of nutritional stress, like carbon starvation, *Escherichia coli* cells abandon their exponential-growth state to enter a more resistant, non-growth state called stationary phase. This growth-phase transition is controlled by a genetic regulatory network integrating various environmental signals. Although *E. coli* is a paradigm of the bacterial world, it is little understood how its response to carbon starvation conditions emerges from the interactions between the different components of the regulatory network. Using a qualitative method that is able to overcome the current lack of quantitative data on kinetic parameters and molecular concentrations, we model the carbon starvation response network and simulate the response of *E. coli* cells to carbon deprivation. This allows us to identify essential features of the transition between exponential and stationary phase and to make new predictions on the qualitative system behaviour following a carbon upshift.

[Cinquemani et al., 2017] **Estimation of time-varying growth, uptake and excretion rates from dynamic metabolomics data**. Eugenio Cinquemani, Valérie Laroute, Muriel Cocaign-Bousquet, Hidde de Jong, Delphine Ropers, *Bioinformatics* (Proceedings of the 25th ISMB/16th ECCB), 33(14):i301–i310.
Motivation. Technological advances in metabolomics have made it possible to monitor the concentration of extracellular metabolites over time. From these data, it is possible to compute the rates of uptake and excretion of the metabolites by a growing cell population, providing precious information on the functioning of intracellular metabolism. The computation of the rate of these exchange reactions, however, is difficult to achieve in practice for a number of reasons, notably noisy measurements, correlations between the concentration profiles of the different extracellular metabolites, and discontinuties in the profiles due to sudden changes in metabolic regime. *Results.* We present a method for precisely estimating time-varying uptake and excretion rates from time-series measurements of extracellular

metabolite concentrations, specifically addressing all of the above issues. The estimation problem is formulated in a regularized Bayesian framework and solved by a combination of extended Kalman filtering and smoothing. The method is shown to improve upon methods based on spline smoothing of the data. Moreover, when applied to two actual datasets, the method recovers known features of overflow metabolism in *Escherichia coli* and *Lactococcus lactis*, and provides evidence for acetate uptake by *L. lactis* after glucose exhaustion. The results raise interesting perspectives for further work on rate estimation from measurements of intracellular metabolites.

[Zulkower et al., 2015] **Robust reconstruction of gene expression profiles from reporter gene data using linear inversion. Valentin Zulkower, Michel Page, Delphine Ropers, Johannes Geiselmann & Hidde de Jong. *Bioinformatics*, 31(12), i71-i79, 2015.**
Motivation: Time-series observations from reporter gene experiments are commonly used for inferring and analyzing dynamical models of regulatory networks. The robust estimation of promoter activities and protein concentrations from primary data is a difficult problem due to measurement noise and the indirect relation between the measurements and quantities of biological interest.
Results: We propose a general approach based on regularized linear inversion to solve a range of estimation problems in the analysis of reporter gene data, notably the inference of growth rate, promoter activity, and protein concentration profiles. We evaluate the validity of the approach using in silico simulation studies, and observe that the methods are more robust and less biased than indirect approaches usually encountered in the experimental literature based on smoothing and subsequent processing of the primary data. We apply the methods to the analysis of fluorescent reporter gene data acquired in kinetic experiments with Escherichia coli. The methods are capable of reliably reconstructing time-course profiles of growth rate, promoter activity and protein concentration from weak and noisy signals at low population volumes. Moreover, they capture critical features of those profiles, notably rapid changes in gene expression during growth transitions.

[Berthoumieux et al., 2013] **Shared control of gene expression in bacteria by transcription factors and global physiology of the cell.** Sara Berthoumieux, Hidde de Jong, Guillaume Baptist, Corinne Pinel, Caroline Ranquet, Delphine Ropers, Johannes Geiselmann. *Molecular Systems Biology*, 9:634, 2013.
Gene expression is controlled by the joint effect of (i) the global physiological state of the cell, in particular the activity of the gene expression machinery, and (ii) DNA-binding transcription factors and other specific regulators. We present a model-based approach to distinguish between these two effects using time-resolved measurements of promoter activities. We demonstrate the strength of the approach by analyzing a circuit involved in the regulation of carbon metabolism in *E. coli*. Our results show that the transcriptional response of the network is controlled by the physiological state of the cell and the signaling metabolite cyclic AMP (cAMP). The absence of a strong regulatory effect of transcription factors suggests that they are not the main coordinators of gene expression changes during growth transitions, but rather that they complement the effect of global physiological control mechanisms. This change of perspective has important consequences for the interpretation of transcriptome data and the design of biological networks in biotechnology and synthetic biology.

[Izard et al., 2015] **A synthetic growth switch based on controlled expression of RNA polymerase.** Jérôme Izard, Cindy Gomez Balderas, Delphine Ropers, Stephan Lacour, Xiaohu Song, Yifan Yang, Ariel B. Lindner, Johannes Geiselmann, Hidde de Jong, *Molecular Systems Biology*, 11(11):840, 2015.
The ability to control growth is essential for fundamental studies of bacterial physiology and biotechnological applications. We have engineered an *Escherichia coli* strain in which the transcription of a key component of the gene expression machinery, RNA polymerase, is under the control of an inducible promoter. By changing the inducer concentration in the medium, we can adjust the RNA polymerase

concentration and thereby switch bacterial growth between zero and the maximal growth rate supported by the medium. We show that our synthetic growth switch functions in a medium-independent and reversible way, and we provide evidence that the switching phenotype arises from the ultrasensitive response of the growth rate to the concentration of RNA polymerase. We present an application of the growth switch in which both the wild-type *E. coli* strain and our modified strain are endowed with the capacity to produce glycerol when growing on glucose. Cells in which growth has been switched off continue to be metabolically active and harness the energy gain to produce glycerol at a twofold higher yield than in cells with natural control of RNA polymerase expression. Remarkably, without any further optimization, the improved yield is close to the theoretical maximum computed from a flux balance model of *E. coli* metabolism. The proposed synthetic growth switch is a promising tool for gaining a better understanding of bacterial physiology and for applications in synthetic biology and biotechnology.

[Morin et al., 2016] **The post-transcriptional regulatory system CSR controls the balance of metabolic pools in upper glycolysis of *Escherichia coli***. Manon Morin, Delphine Ropers, Fabien Letisse, Sandrine Laguerre, Jean-Charles Portais, Muriel Cocaign-Bousquet, Brice Enjalbert. *Molecular Microbiology*, 100(4), 686-700, 2016.
Metabolic control in *Escherichia coli* is a complex process involving multilevel regulatory systems but the involvement of post-transcriptional regulation is uncertain. The post-transcriptional factor CsrA is stated as being the only regulator essential for the use of glycolytic substrates. A dozen enzymes in the central carbon metabolism (CCM) have been reported as potentially controlled by CsrA, but its impact on the CCM functioning has not been demonstrated. Here, a multiscale analysis was performed in a wild-type strain and its isogenic mutant attenuated for CsrA (including growth parameters, gene expression levels, metabolite pools, abundance of enzymes and fluxes). Data integration and regulation analysis showed a coordinated control of the expression of glycolytic enzymes. This also revealed the imbalance of metabolite pools in the csrA mutant upper glycolysis, before the phosphofructokinase PfkA step. This imbalance is associated with a glucose–phosphate stress. Restoring PfkA activity in the *csrA* mutant strain suppressed this stress and increased the mutant growth rate on glucose. Thus, the carbon storage regulator system is essential for the effective functioning of the upper glycolysis mainly through its control of PfkA. This work demonstrates the pivotal role of post-transcriptional regulation to shape the carbon metabolism.

[Morin et al., 2017] **The Csr system regulates *Escherichia coli* fitness by controlling glycogen accumulation and energy levels**. Manon Morin, Delphine Ropers, Eugenio Cinquemani, Jean-Charles Portais, Brice Enjalbert and Muriel Cocaign-Bousquet. *mBio*, 8(5), 2017.
In the bacterium *Escherichia coli*, the post-transcriptional regulatory system Csr was postulated to influence the transition from glycolysis to gluconeogenesis. Here, we explored the role of the Csr system in the glucose-acetate transition as a model of the glycolysis-to-gluconeogenesis switch. Mutations in the Csr system influence the reorganization of gene expression after glucose exhaustion and disturb the timing of acetate consumption after glucose exhaustion. Analysis of metabolite concentrations during the transition revealed that the Csr system has a major effect on the energy levels of the cells after glucose exhaustion. This influence was demonstrated to result directly from the effect of the Csr system on glycogen accumulation. Mutation in glycogen metabolism was also demonstrated to hinder metabolic adaptation after glucose exhaustion because of insufficient energy. This work explains how the Csr system influences *E. coli* fitness during the glycolysis-gluconeogenesis switch and demonstrates the role of glycogen in maintenance of the energy charge during metabolic adaptation.

[Etienne et al., 2020] **Competitive effects in bacterial mRNA decay**. Thibault A. Etienne, Muriel Cocaign-Bousquet, Delphine Ropers. *Journal of Theoretical Biology*, 504: 110333, 2020.
In living organisms, the same enzyme catalyses the degradation of thousands of different mRNAs, but the possible influence of competing substrates has been largely ignored so far. We develop a simple

mechanistic model of the coupled degradation of all cell mRNAs using the total quasi-steady-state approximation of the Michaelis-Menten framework. Numerical simulations of the model using carefully chosen parameters and analyses of rate sensitivity coefficients show how substrate competition alters mRNA decay. The model predictions reproduce and explain a number of experimental observations on mRNA decay following transcription arrest, such as delays before the onset of degradation, the occurrence of variable degradation profiles with increased non linearities and the negative correlation between mRNA half-life and concentration. The competition acts at different levels, through the initial concentration of cell mRNAs and by modifying the enzyme affinity for its targets. The consequence is a global slow down of mRNA decay due to enzyme titration and the amplification of its apparent affinity. Competition happens to stabilize weakly affine mRNAs and to destabilize the most affine ones. We believe that this mechanistic model is an interesting alternative to the exponential models commonly used for the determination of mRNA half-lives. It allows analysing regulatory mechanisms of mRNA degradation and its predictions are directly comparable to experimental data.

[Etienne et al., In preparation] **A mechanistic model informed by dynamic omics data reveals the physiological control of bacterial mRNA decay**. Thibault A. Etienne, Eugenio Cinquemani, Laurence Girbal, Muriel Cocaign-Bousquet, Delphine Ropers. In preparation.

Bibliography

W. Abou-Jaoudé, P. Traynard, P. T. Monteiro, J. Saez-Rodriguez, T. Helikar, D. Thieffry, and C. Chaouiya. Logical modeling and dynamical analysis of cellular networks. *Frontiers Genet*, 7: 94, 2016.

R. Agren, S. Bordel, A. Mardinoglu, N. Pornputtapong, I. Nookaew, and J. Nielsen. Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using init. *PLoS Comput. Biol.*, 8(5):e1002518, 2012.

M. Åkesson, J. Förster, and J. Nielsen. Integration of gene expression data into genome-scale metabolic models. *Metab. Eng.*, 6(4):285–293, 2004.

K. R. Albe, M. H. Butler, and B. E. Wright. Cellular concentrations of enzymes and their substrates. *J. Theor. Biol.*, 143(2):163–195, 1990.

T. Ali Azam, A. Iwata, A. Nishimura, S. Ueda, and A. Ishihama. Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J. Bacteriol.*, 181(20):6361–70, 1999.

A. P. Arkin. A wise consistency: engineering biology for conformity, reliability, predictability. *Curr. Opin. Chem. Biol.*, 17(6):893–901, 2013.

C. Auffray and L. Nottale. Scale relativity theory and integrative systems biology: 1: founding principles and scale laws. *Prog Biophys Mol Biol*, 97(1):79–114, 2008.

V. Baldazzi, D. Ropers, Y. Markowicz, D. Kahn, J. Geiselmann, and H. de Jong. The carbon assimilation network in *Escherichia coli* is densely connected and largely sign-determined by directions of metabolic fluxes. *PLoS Comput. Biol.*, 6:e1000812, 2010.

V. Baldazzi, D. Ropers, J. Geiselmann, D. Kahn, and H. de Jong. Importance of metabolic coupling for the dynamics of gene expression following a diauxic shift in *Escherichia coli*. *J. Theor. Biol.*, 295:100–115, 2012.

G. Batt, D. Ropers, H. de Jong, J. Geiselmann, R. Mateescu, M. Page, and D. Schneider. Validation of qualitative models of genetic regulatory networks by model checking: analysis of the nutritional stress response in *Escherichia coli*. *Bioinformatics*, 21(suppl 1):i19–28, 2005.

G. Batt, B. Besson, P.-E. Ciron, H. de Jong, E. Dumas, J. Geiselmann, R. Monte, P. T. Monteiro, M. Page, F. Rechenmann, and D. Ropers. Genetic network analyzer: a tool for the qualitative modeling and simulation of bacterial regulatory networks. In *Bacterial Molecular Networks*, pages 439–462. Springer, 2012.

D. A. Beard, E. Babson, E. Curtis, and H. Qian. Thermodynamic constraints for biochemical networks. *J. Theor. Biol.*, 228(3):327–333, 2004.

S. A. Becker, A. M. Feist, M. L. Mo, G. Hannum, B. Ø. Palsson, and M. J. Herrgard. Quantitative prediction of cellular metabolism with constraint-based models: the cobra toolbox. *Nat. protoc.*, 2(3):727–738, 2007.

I. Belgacem, S. Casagranda, E. Grac, D. Ropers, and J.-L. Gouzé. Reduction and stability analysis of a transcription–translation model of rna polymerase. *Bull. Math. Biol.*, 80(2):294–318, 2018.

C. Bernard. *Introduction à l'étude de la médecine expérimentale*. JB Baillière et fils, 1865.

M. Bersanelli, E. Mosca, D. Remondini, E. Giampieri, C. Sala, G. Castellani, and L. Milanesi. Methods for the integration of multi-omics data: mathematical aspects. *BMC Bioinform.*, 17(S2): S15, 2016.

S. Berthoumieux, H. de Jong, G. Baptist, C. Pinel, C. Ranquet, D. Ropers, and J. Geiselmann. Shared control of gene expression in bacteria by transcription factors and global physiology of the cell. *Mol. Syst. Biol.*, 9:634, 2013.

H. P. Bonarius, G. Schmid, and J. Tramper. Flux analysis of underdetermined metabolic networks: the quest for the missing constraints. *Trends Biotechnol*, 15(8):308–314, 1997.

A. Bordbar, J. Monk, Z. King, and B. Palsson. Constraint-based models predict metabolic and associated cellular functions. *Nature Rev Genet*, 15(2):107–120, 2014.

J. A. M. Borghans, R. J. de Boer, and L. A. Segel. Extending the quasi-steady state approximation by changing variables. *Bull. Math. Biol.*, 58(1): 43–63, Jan 1996. ISSN 1522-9602.

F. Boyer, B. Besson, G. Baptist, J. Izard, C. Pinel, D. Ropers, J. Geiselmann, and H. de Jong. WellReader: a MATLAB program for the analysis of fluorescence and luminescence reporter gene data. *Bioinformatics*, 26:1262–1263, 2010.

I. Brigandt. Beyond reduction and pluralism: Toward an epistemology of explanatory integration in biology. *Erkenntnis*, 73(3):295–311, 2010.

G. E. Briggs and J. B. S. Haldane. A note on the kinetics of enzyme action. *Biochem J*, 19(2):338, 1925.

F. J. Bruggeman and H. V. Westerhoff. The nature of systems biology. *Trends Microbiol.*, 15 (1):45–50, 2007.

J. Calvert and J. H. Fujimura. Calculating life? duelling discourses in interdisciplinary systems biology. *Stud Hist Philos Biol Biomed Sci*, 42(2): 155–163, 2011.

W. B. Cannon. Organization for physiological homeostasis. *Phys Rev*, 9(3):399–431, 1929.

W. B. Cannon. *The wisdom of the body*. Norton & Co., 1939.

R. Carlson and F. Srienc. Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: creation of overall flux states. *Biotechnol. Bioeng.*, 86(2):149–162, 2004a.

R. Carlson and F. Srienc. Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: identification of reactions. *Biotechnol. Bioeng.*, 85(1):1–19, 2004b.

S. Casagranda, D. Ropers, and J.-L. Gouzé. Model reduction and process analysis of biological models. In *2015 23rd Mediterranean Conference on Control and Automation (MED)*, pages 1132–1139. IEEE, 2015.

S. Casagranda, S. Touzeau, D. Ropers, and J.-L. Gouzé. Principal process analysis of biological models. *BMC Syst. Biol.*, 12(1):68, 2018.

S. Chandrasekaran and N. D. Price. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proc Nat Acad Sciences USA*, 107(41):17845–17850, 2010.

H. Chen, K. Shiroguchi, H. Ge, and X. Xie. Genome-wide study of mRNA degradation and transcript elongation in *Escherichia coli*. *Mol Syst Biol*, 11(1):781, 2015.

W. Chen, M. Niepel, and P. Sorger. Classic and contemporary approaches to modeling biochemical reactions. *Genes Dev.*, 24(17):1861–1875, 2010.

B. Choi, G. A. Rempala, and J. K. Kim. Beyond the Michaelis-Menten equation: Accurate and efficient estimation of enzyme kinetic parameters. *Sci Rep*, 7(1):17018, 2017.

A. Ciliberto, F. Capuani, and J. J. Tyson. Modeling networks of coupled enzymatic reactions using the total quasi-steady state approximation. *PLoS Comput. Biol.*, 3(3):e45, 2007.

E. Cinquemani, V. Laroute, M. Cocaign-Bousquet, H. de Jong, and D. Ropers. Estimation of time-varying growth, uptake and excretion rates from dynamic metabolomics data. *Bioinformatics*, 33 (14):i301–i310, 2017.

C. Colijn, A. Brandes, J. Zucker, D. S. Lun, B. Weiner, M. R. Farhat, T.-Y. Cheng, D. B. Moody, M. Murray, and J. E. Galagan. Interpreting expression data with metabolic flux models: predicting *Mycobacterium tuberculosis* mycolic acid production. *PLoS Comput Biol*, 5(8): e1000489, 2009.

S. J. Cooper. From claude bernard to walter cannon. emergence of the concept of homeostasis. *Appetite*, 51(3):419–427, 2008.

F. Corblin, S. Tripodi, E. Fanchon, D. Ropers, and L. Trilling. A declarative constraint-based method for analyzing discrete genetic regulatory networks. *Biosystems*, 98(2):91–104, 2009.

A. Cornish-Bowden. One hundred years of michaelis–menten kinetics. *Perspectives Sci*, 4: 3–9, 2015.

C. Cotten and J. L. Reed. Mechanistic analysis of multi-omics datasets to generate kinetic parameters for constraint-based metabolic models. *BMC Bioinf.*, 14(1):32, 2013.

M. W. Covert, C. H. Schilling, and B. Palsson. Regulation of gene expression in flux balance models of metabolism. *J. Theor. Biol.*, 213(1):73–88, 2001.

M. W. Covert, N. Xiao, T. J. Chen, and J. R. Karr. Integrating metabolic, transcriptional regulatory and signal transduction models in *Escherichia coli*. *Bioinformatics*, 24(18):2044–2050, 2008.

H. de Jong. Modeling and simulation of genetic regulatory systems: a literature review. *J. Comput. Biol.*, 9(1):69–105, 2002.

H. de Jong and D. Ropers. Qualitative approaches to the analysis of genetic regulatory networks. *System Modeling in Cellular Biology: From Concepts to Nuts and Bolts*, pages 125–147, 2006a.

H. de Jong and D. Ropers. Strategies for dealing with incomplete information in the modeling of molecular interaction networks. *Briefings Bioinform.*, 7(4):354–363, 2006b.

H. de Jong, J. Geiselmann, C. Hernandez, and M. Page. Genetic Network Analyzer: qualitative simulation of genetic regulatory networks. *Bioinformatics*, 19:336–44, 2003.

H. de Jong, J.-L. Gouzé, C. Hernandez, M. Page, T. Sari, and J. Geiselmann. Qualitative simulation of genetic regulatory networks using piecewise-linear models. *Bull. Math. Biol.*, 66 (2):301–40, 2004.

H. de Jong, C. Ranquet, D. Ropers, C. Pinel, and J. Geiselmann. Experimental and computational validation of models of fluorescent and luminescent reporter genes in bacteria. *BMC Syst Biol*, 4:55, 2010.

H. de Jong, S. Casagranda, N. Giordano, E. Cinquemani, D. Ropers, J. Geiselmann, and J.-L. Gouzé. Mathematical modeling of microbes: Metabolism, gene expression, and growth. *J R Soc Interface*, 14:20170502, 2017a.

H. de Jong, J. Geiselmann, and D. Ropers. Resource reallocation in bacteria by reengineering the gene expression machinery. *Trends Microbiol*, 25(6):480–493, 2017b.

P. Dennis and H. Bremer. Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *EcoSal Plus*, 3(1):1–49, 2008.

P. Dennis, M. Ehrenberg, and H. Bremer. Control of rRNA synthesis in *Escherichia coli*: a systems biology approach. *Microbio.l Mol. Biol. Rev.*, 68 (4):639–68, 2004.

H. Dourado and M. Lercher. An analytical theory of balanced cellular growth. *Nat. Commun.*, 11: 1226, 2020.

R. G. Duggleby and R. B. Clarke. Experimental designs for estimating the parameters of the michaelis-menten equation from progress curves of enzyme-catalyzed reactions. *Biochim Biophys Acta*, 1080(3):231–236, 1991.

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

J. S. Edwards and B. O. Palsson. Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. *J Biol Chem*, 274(25):17410–17416, 1999.

J. S. Edwards, R. U. Ibarra, and B. O. Palsson. In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat. Biotechnol.*, 19(2):125–130, Feb 2001.

T. Esquerré, S. Laguerre, C. Turlan, A. Carpousis, L. Girbal, and M. Cocaign-Bousquet. Dual role of transcription and transcript stability in the regulation of gene expression in *Escherichia coli* cells cultured on glucose at different growth rates. *Nucleic Acids Res*, 42(4):2460–2472, 2014.

T. Esquerré, A. Moisan, H. Chiapello, L. Arike, R. Vilu, C. Gaspin, M. Cocaign-Bousquet, and L. Girbal. Genome-wide investigation of mRNA lifetime determinants in *Escherichia coli* cells cultured at different growth rates. *BMC Genomics*, 16(1):275, 2015.

T. Esquerré, M. Bouvier, C. Turlan, A. J. Carpousis, L. Girbal, and M. Cocaign-Bousquet. The Csr system regulates genome-wide mRNA stability and transcription and thus gene expression in *Escherichia coli*. *Sci. Rep.*, 6:25057, 2016.

T. Etienne, M. Cocaign-Bousquet, and D. Ropers. Competitive effects in bacterial mRNA decay. *J. Theor. Biol.*, 504:110333, 2020.

T. Etienne, E. Cinquemani, L. Girbal, M. Cocaign-Bousquet, and D. Ropers. A mechanistic model informed by dynamic omics data reveals the physiological control of bacterial mRNA decay. In preparation.

D. Eveillard, D. Ropers, H. De Jong, C. Branlant, and A. Bockmayr. Multiscale modeling of alternative splicing regulation. In *International Conference on Computational Methods in Systems Biology*, pages 75–87. Springer, 2003.

D. Eveillard, D. Ropers, H. de Jong, C. Branlant, and A. Bockmayr. A multi-scale constraint programming model of alternative splicing regulation. *Theor. Comput. Sci.*, 325(1):3–24, 2004.

A. M. Feist and B. O. Palsson. The biomass objective function. *Curr Opin Microbiol*, 13(3):344–349, 2010.

R. M. Fleming, I. Thiele, and H. Nasheuer. Quantitative assignment of reaction directionality in constraint-based models of metabolism: application to *Escherichia coli*. *Biophys Chem*, 145(2-3): 47–56, 2009.

R. M. Fleming, I. Thiele, G. Provan, and H. Nasheuer. Integrated stoichiometric, thermodynamic and kinetic modelling of steady state metabolism. *J. Theor. Biol.*, 264(3):683–692, 2010.

J. Geiselmann, H. de Jong, D. Ropers, and J. Izard. Method for producing metabolites, peptides and recombinant proteins, 2015. Also published as EP3047031 (A1) US2016222428 (A1) WO2015036622 (A1).

L. Gerosa, K. Kochanowski, M. Heinemann, and U. Sauer. Dissecting specific and global transcriptional regulation of bacterial gene expression. *Mol. Syst. Biol.*, 9:658, 2013.

L. Glass and S. Kauffman. The logical analysis of continuous, non-linear biochemical control networks. *J. Theor. Biol.*, 39(1):103–29, 1973.

A. Gonzalez, J. Uhlendorf, J. Schaul, E. Cinquemani, G. Batt, and G. Ferrari-Trecate. Identification of biological models from single-cell data: a comparison between mixed-effects and moment-based inference. In *2013 European Control Conference (ECC)*, pages 3652–3657. IEEE, 2013.

B. C. Goodwin et al. Temporal organization in cells. a dynamic theory of cellular control processes. *Temporal organization in cells. A dynamic theory of cellular control processes.*, 1963.

S. Gopalakrishnan, S. Dash, and C. Maranas. K-FIT: An accelerated kinetic parameterization algorithm using steady-state fluxomic data. *Metabol. Eng.*, 2020.

C. Gu, G. B. Kim, W. J. Kim, H. U. Kim, and S. Y. Lee. Current status and applications of genome-scale metabolic models. *Genome Biol*, 20(1):121, 2019.

S. Gudmundsson, L. Agudo, and J. Nogales. Applications of genome-scale metabolic models of microalgae and cyanobacteria in biotechnology. In *Microalgae-Based Biofuels and Bioproducts*, pages 93–111. Elsevier, 2017.

H. Hallay, N. Locker, L. Ayadi, D. Ropers, E. Guittet, and C. Branlant. Biochemical and NMR study on the competition between proteins SC35, SRp40, and heterogeneous nuclear ribonucleoprotein A1 at the HIV-1 Tat exon 2 splicing site. *J. Biol. Chem.*, 281(48):37159–37174, 2006.

T. J. Hanly and M. A. Henson. Dynamic flux balance modeling of microbial co-cultures for efficient batch fermentation of glucose and xylose mixtures. *Biotechnol. Bioeng.*, 108(2):376–385, 2011.

H. S. Haraldsdóttir, B. Cousins, I. Thiele, R. M. Fleming, and S. Vempala. Chrr: coordinate hit-and-run with rounding for uniform sampling of constraint-based models. *Bioinformatics*, 33(11): 1741–1743, 2017.

W. R. Harcombe, N. F. Delaney, N. Leiby, N. Klitgord, and C. J. Marx. The ability of flux balance analysis to predict evolution of central metabolism scales with the initial distance to the optimum. *PLoS Comput Biol*, 9(6):e1003091, 2013.

J. Heijnen. Approximative kinetic formats used in metabolic network modeling. *Biotechnol. Bioeng.*, 91(5):534–545, 2005.

R. Heinrich and T. A. Rapoport. A linear steady-state treatment of enzymatic chains: general properties, control and effector strength. *Eur. J. Biochem.*, 42(1):89–95, 1974.

R. Heinrich and S. Schuster. *The Regulation of Cellular Systems*. Chapman and Hall, New-York, 1996.

L. Heirendt, S. Arreckx, T. Pfau, S. N. Mendoza, A. Richelle, A. Heinken, H. S. Haraldsdóttir, J. Wachowiak, S. M. Keating, V. Vlasov, et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v. 3.0. *Nat. protoc.*, 14(3):639–702, 2019.

C. S. Henry, M. D. Jankowski, L. J. Broadbelt, and V. Hatzimanikatis. Genome-scale thermodynamic analysis of *Escherichia coli* metabolism. *Biophysical J*, 90(4):1453–1461, 2006.

C. S. Henry, L. J. Broadbelt, and V. Hatzimanikatis. Thermodynamics-based metabolic flux analysis. *Biophys. J.*, 92(5):1792–1805, 2007.

A. Hoppe, S. Hoffmann, and H.-G. Holzhütter. Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks. *BMC Syst. Biol.*, 1(1):23, 2007.

R. U. Ibarra, J. S. Edwards, and B. O. Palsson. *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature*, 420(6912):186–189, 2002.

T. Ideker, T. Galitski, and L. Hood. A new approach to decoding life: systems biology. *Annu Rev Genomics Hum Genet*, 2(1):343–372, 2001.

J. Izard, C. Gomez Balderas, D. Ropers, S. Lacour, X. Song, Y. Yang, A. Lindner, J. Geiselmann, and H. de Jong. A synthetic growth switch based on controlled expression of RNA polymerase. *Mol Syst Biol*, 11(11):840, 2015.

S. Jacquenet, D. Ropers, P. S. Bilodeau, L. Damier, A. Mougin, C. M. Stoltzfus, and C. Branlant. Conserved stem–loop structures in the HIV-1 RNA region containing the A3 3' splice site and its cis-regulatory element: possible involvement in RNA splicing. *Nucleic Acids Res.*, 29(2):464–478, 2001.

M. L. Jenior, T. J. Moutinho Jr, B. V. Dougherty, and J. A. Papin. Transcriptome-guided parsimonious flux analysis improves predictions with metabolic networks in complex environments. *PLOS Computational Biology*, 16(4):e1007099, 2020.

P. A. Jensen and J. A. Papin. Functional integration of a metabolic network model and expression data without arbitrary thresholding. *Bioinformatics*, 27(4):541–547, 2011.

K. A. Johnson. A century of enzyme kinetic analysis, 1913 to 2013. *FEBS lett*, 587(17):2753–2766, 2013.

S. Jun, F. Si, R. Pugatch, and M. Scott. Fundamental principles in bacterial physiology-history, recent progress, and the future with focus on cell size control: a review. *Rep. Prog. Phys.*, 81(5): 056601, 2018.

H. Kacser. The control of flux. In *Symp. Soc. Exp. Biol.*, volume 27, pages 65–104, 1973.

T. Kailath, A. H. Sayed, and B. Hassibi. *Linear Estimation*. Prentice Hall, 2000.

S. Kauffman. Gene regulation networks: A theory for their global structure and behaviors. In *Curr. Top. Dev. Biol.*, volume 6, pages 145–182. Elsevier, 1971.

S. A. Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.*, 22(3):437–467, 1969.

D. E. Kaufman and R. L. Smith. Direction choice for accelerated convergence in hit-and-run sampling. *Oper. Res.*, 46(1):84–95, 1998.

T. C. Keaty and P. A. Jensen. Gapsplit: Efficient random sampling for non-convex constraint-based models. *Bioinformatics*, 36(8):2623–2625, 2020.

D. Kell, M. Brown, H. Davey, W. Dunn, I. Spasic, and S. Oliver. Metabolic footprinting and systems biology: the medium is the message. *Nat. Rev. Microbiol.*, 3(7):557–65, 2005.

D. B. Kell and S. G. Oliver. Here is the evidence, now what is the hypothesis? the complementary roles of inductive and hypothesis-driven science in the post-genomic era. *Bioessays*, 26(1):99–105, 2004.

L. Keren, O. Zackay, M. Lotan-Pompan, U. Barenholz, E. Dekel, V. Sasson, G. Aidelberg, A. Bren, D. Zeevi, A. Weinberger, U. Alon, R. Milo, and E. Segal. Promoters maintain their relative activity levels under different growth conditions. *Mol. Syst. Biol.*, 9:701, 2013.

G. Khoury, L. Ayadi, J. M. Sailou, S. Sanglier, D. Ropers, and C. Branlant. New actors in regulation of HIV-1 tat mRNA production. *Retrovirology*, 6(2):1–1, 2009.

Z. A. King, J. Lu, A. Dräger, P. Miller, S. Federowicz, J. A. Lerman, A. Ebrahim, B. O. Palsson, and N. E. Lewis. BiGG models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.*, 44(D1):D515–D522, 2016.

M. W. Kirschner. The meaning of systems biology. *Cell*, 121(4):503–504, 2005.

H. Kitano. Systems biology: toward system-level understanding of biological systems. *Foundations of Systems Biology*, pages 1–36, 2001.

H. Kitano. Looking beyond the details: a rise in system-oriented approaches in genetics and molecular biology. *Curr. Genet.*, 41(1):1–10, 2002.

E. Klipp, W. Liebermeister, C. Wierling, and A. Kowald. *Systems biology: a textbook*. John Wiley & Sons, 2016.

S. Klumpp and T. Hwa. Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proc Natl Acad Sci U S A*, 105:20245–20250, 2008.

K. Kochanowski, B. Volkmer, L. Gerosa, B. R. Haverkorn van Rijsewijk, A. Schmidt, and M. Heinemann. Functioning of a metabolic flux sensor in *Escherichia coli*. *Proc Natl Acad Sci USA*, 110(3):1130–1135, Jan 2013.

K. Kohn. Molecular interaction maps as information organizers and simulation guides. *Chaos*, 11 (1):84–97, 2001.

O. Kotte, J. B. Zaugg, and M. Heinemann. Bacterial adaptation through distributed sensing of metabolic fluxes. *Mol. Syst. Biol.*, 6:355, 2010.

A. Kümmel, S. Panke, and M. Heinemann. Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Mol. Syst. Biol.*, 2(1): 2006–0034, 2006a.

A. Kümmel, S. Panke, and M. Heinemann. Systematic assignment of thermodynamic constraints in metabolic network models. *BMC Bioinf.*, 7(1): 1–12, 2006b.

S. Laguerre, I. González, S. Nouaille, A. Moisan, N. Villa-Vialaneix, C. Gaspin, M. Bouvier, A. J. Carpousis, M. Cocaign-Bousquet, and L. Girbal. Large-scale measurement of mRNA degradation in *Escherichia coli*: To delay or not to delay. *Methods Enzymol.*, 612:47–66, 2018.

K. Larrabee, J. Phillips, G. Williams, and A. Larrabee. The relative rates of protein synthesis and degradation in a growing culture of *Escherichia coli*. *J Biol Chem*, 255(9):4125–4130, 1980.

M. Lavielle. *Mixed effects models for the population approach: models, tasks, methods and tools*. CRC press, 2014.

D. Lee, K. Smallbone, W. B. Dunn, E. Murabito, C. L. Winder, D. B. Kell, P. Mendes, and N. Swainston. Improving metabolic flux predictions using absolute gene expression data. *BMC Syst Biol*, 6(1):73, 2012.

J. M. Lee, E. P. Gianchandani, J. A. Eddy, and J. A. Papin. Dynamic analysis of integrated signaling, metabolic, and regulatory networks. *PLoS Comput Biol*, 4(5):e1000086, 2008.

N. E. Lewis, K. K. Hixson, T. M. Conrad, J. A. Lerman, P. Charusanti, A. D. Polpitiya, J. N. Adkins, G. Schramm, S. O. Purvine, D. Lopez-Ferrer, et al. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol. Syst. Biol.*, 6(1):390, 2010.

D. Lipson. The complex relationship between microbial growth rate and yield and its implications for ecosystem processes. *Front. Microbiol.*, 6:615, 2015.

D. Machado and M. Herrgård. Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLoS Comp. Biol.*, 10(4):e1003580, 2014.

R. Mahadevan and C. Schilling. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabol. Eng.*, 5(4):264–276, 2003.

R. Mahadevan, J. S. Edwards, and F. J. Doyle III. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys. J.*, 83(3):1331–1340, 2002.

A. Marguet, M. Lavielle, and E. Cinquemani. Inheritance and variability of kinetic gene expression parameters in microbial cells: modeling and inference from lineage tree data. *Bioinformatics*, 35(14):i586–i595, 2019.

Y. Martin, M. Page, C. Blanchet, and H. de Jong. WellInverter: a web application for the analysis of fluorescent reporter gene data. *BMC Bioinf*, 20(1):309, 2019.

W. Megchelenbrink, M. Huynen, and E. Marchiori. optGpSampler: an improved tool for uniformly sampling the solution-space of genome-scale metabolic networks. *PLoS one*, 9(2): e86587, 2014.

M. D. Mesarović. Systems theory and biology—view of a theoretician. In *Systems theory and biology*, pages 59–87. Springer, 1968.

T. Mestl, E. Plahte, and S. Omholt. A mathematical framework for describing and analysing gene regulatory networks. *J. Theor. Biol.*, 176(2):291–300, 1995.

A. Métris, S. M. George, and D. Ropers. Piecewise linear approximations to model the dynamics of adaptation to osmotic stress by food-borne pathogens. *Int J Food Microbiol*, 240:63–74, 2017.

L. Michaelis and M. L. Menten. The kinetics of the inversion effect. *Biochem. Z*, 49:333–369, 1913.

P. Mitchell. Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature*, 191(4784):144–148, 1961.

M. L. Mo, B. Ø. Palsson, and M. J. Herrgård. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst. Biol.*, 3(1):37, 2009.

D. Molenaar, R. Van Berlo, D. De Ridder, and B. Teusink. Shifts in growth strategies reflect tradeoffs in cellular economics. *Molecular systems biology*, 5(1):323, 2009.

P. Monteiro, D. Ropers, R. Mateescu, A. Freitas, and H. de Jong. Temporal logic patterns for querying dynamic models of cellular interaction networks. *Bioinformatics*, 24(16):i227–33, 2008.

P. T. Monteiro, P. J. Dias, D. Ropers, A. L. Oliveira, I. Sa-Correia, M. C. Teixeira, and A. T. Freitas. Qualitative modelling and formal verification of the FLR1 gene mancozeb response in *Saccharomyces cerevisiae*. *IET Syst Biol*, 5:308–316, 2011.

M. Morin, D. Ropers, F. Letisse, S. Laguerre, J.-C. Portais, M. Cocaign-Bousquet, and B. Enjalbert. The post-transcriptional regulatory system CSR controls the balance of metabolic pools in upper glycolysis of *Escherichia coli*. *Mol Microbiol*, 100 (4):686–700, 2016.

M. Morin, D. Ropers, E. Cinquemani, J. Portais, B. Enjalbert, and M. Cocaign-Bousquet. The Csr system regulates *Escherichia coli* fitness by controlling glycogen accumulation and energy levels. *mBio*, 8(5), 2017.

M. Morin, B. Enjalbert, D. Ropers, L. Girbal, and M. Cocaign-Bousquet. Genomewide stabilization of mRNA during a "feast-to-famine" growth transition in *Escherichia coli*. *Msphere*, 5(3), 2020.

A. C. Müller and A. Bockmayr. Fast thermodynamically constrained flux variability analysis. *Bioinformatics*, 29(7):903–909, 2013.

E. Noor, S. Cherkaoui, and U. Sauer. Biological insights through omics data integration. *Curr. Opin. Syst. Biol.*, 15:39–47, 2019.

C. J. Norsigian, N. Pusarla, J. L. McConn, J. T. Yurkovich, A. Dräger, B. O. Palsson, and Z. King. BiGG models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Res.*, 48(D1): D402–D406, 2020.

S. Nouaille, S. Mondeil, A.-L. Finoux, C. Moulis, L. Girbal, and M. Cocaign-Bousquet. The stability of an mRNA is influenced by its concentration: a potential physical mechanism to regulate gene expression. *Nucleic Acids Res*, 45(20): 11711–11724, 2017.

E. O'Brien, J. Lerman, R. Chang, D. Hyduke, and B. Palsson. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol. Syst. Biol.*, 9:693, 2013.

M. Okino and M. Mavrovouniotis. Simplification of mathematical models of chemical reaction systems. *Chemical Rev*, 98(2):391–408, 1998.

R. E. O'Malley. *Singular perturbation methods for ordinary differential equations*, volume 89. Springer, 1991.

J. D. Orth, I. Thiele, and B. Ø. Palsson. What is flux balance analysis? *Nature Biotechnol.*, 28(3): 245–248, 2010.

J. Papin, J. Stelling, N. Price, S. Klamt, S. Schuster, and B. Palsson. Comparison of network-based pathway analysis methods. *Trends Biotechnol*, 22(8):400–405, 2004.

E. Pecou. Splitting the dynamics of large biochemical interaction networks. *J. Theor. Biol.*, 232(3): 375–384, 2005.

M. G. Pedersen, A. M. Bersanib, and E. Bersanic. The total quasi-steady-state approximation for fully competitive enzyme reactions. *Bull. Math. Biol.*, 69(1):433, 2007.

M. G. Pedersen, A. M. Bersani, and E. Bersani. Quasi steady-state approximations in complex intracellular signal transduction networks–a word of caution. *J Math Chem*, 43(4):1318–1344, 2008a.

M. G. Pedersen, A. M. Bersani, E. Bersani, and G. Cortese. The total quasi-steady-state approximation for complex enzyme reactions. *Math. Comput. Simul.*, 79(4):1010 – 1019, 2008b.

S. Pinhal, D. Ropers, J. Geiselmann, and H. de Jong. Acetate metabolism and the inhibition of bacterial growth by acetate. *J. Bacteriol.*, 201(13):e00147–19, 2019.

C. Pourciau, Y.-J. Lai, M. Gorelik, P. Babitzke, and T. Romeo. Diverse mechanisms and circuitry for global regulation by the RNA-binding protein CsrA. *Front Microbiol*, 11:2709, 2020.

N. D. Price, J. L. Reed, and B. Ø. Palsson. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Rev. Microbiol.*, 2(11):886–897, 2004.

I. Prigogine and G. Nicolis. Biological order, structure and instabilities. *Q. Rev. Biophys.*, 4(2-3): 107–148, 1971.

T. Pusa, M. Ferrarini, R. Andrade, A. Mary, A. Marchetti-Spaccamela, L. Stougie, and M. Sagot. MOOMIN–mathematical exploration of'omics data on a metabolic network. *Bioinformatics*, 36(2):514–523, 2020.

H. Qian and D. A. Beard. Thermodynamics of stoichiometric biochemical networks in living systems far from equilibrium. *Biophys. Chem.*, 114 (2-3):213–220, 2005.

O. Radulescu, A. N. Gorban, A. Zinovyev, and V. Noel. Reduction of dynamical biochemical reactions networks in computational biology. *Front. Genet.*, 3:131, 2012.

C. Ramon, M. G. Gollub, and J. Stelling. Integrating–omics data into genome-scale metabolic network models: principles and challenges. *Essays Biochem.*, 62(4):563–574, 2018.

T. Romeo, M. Gong, M. Y. Liu, and A.-M. Brun-Zinkernagel. Identification and molecular characterization of *csrA*, a pleiotropic gene from *Escherichia coli* that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties. *J. Bacteriol.*, 175(15):4744–4755, 1993.

M. Ronen, R. Rosenberg, B. I. Shraiman, and U. Alon. Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. *Proc Natl Acad Sci USA*, 99:10555–10560, 2002.

D. Ropers and A. Métris. Data for the qualitative modeling of the osmotic stress response to NaCl in *Escherichia coli*. *Data Brief*, 9:606–612, 2016.

D. Ropers, L. Ayadi, R. Gattoni, S. Jacquenet, L. Damier, C. Branlant, and J. Stévenin. Differential effects of the SR proteins 9G8, SC35, ASF/SF2, and SRp40 on the utilization of the A1 to A5 splicing sites of HIV-1 RNA. *J. Biol. Chem.*, 279(29):29963–29973, 2004.

D. Ropers, H. de Jong, M. Page, D. Schneider, and J. Geiselmann. Qualitative simulation of the carbon starvation response in *Escherichia coli*. *Biosystems*, 84(2):124–152, 2006.

D. Ropers, H. de Jong, J.-L. Gouzé, M. Page, D. Schneider, and J. Geiselmann. *Piecewise-Linear Models of Genetic Regulatory Networks: Analysis of the Carbon Starvation Response in Escherichia coli*, chapter 8, pages 83–96. Springer, 2007. Proceedings of the Fifth European Conference on Mathematical and Theoretical Biology (ECMTB05), Dresden, Germany.

D. Ropers, V. Baldazzi, and H. de Jong. Model reduction using piecewise-linear approximations preserves dynamic properties of the carbon starvation response in *Escherichia coli*. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, 8(1):166–181, 2011.

R. Rosen. A relational theory of biological systems. *Bull Math Biophys*, 20(3):245–260, 1958.

H. Rottenberg, S. Caplan, and A. Essig. Stoichiometry and coupling: theories of oxidative phosphorylation. *Nature*, 216(5115):610–611, 1967.

M. R. Roussel and S. J. Fraser. Invariant manifold methods for metabolic model reduction. *Chaos*, 11(1):196–206, 2001.

P. Salvy and V. Hatzimanikatis. The ETFL formulation allows multi-omics integration in thermodynamics-compliant metabolism and expression models. *Nature Comm*, 11(1):1–17, 2020.

U. Sauer, D. C. Cameron, and J. E. Bailey. Metabolic capacity of *Bacillus subtilis* for the production of purine nucleosides, riboflavin, and folic acid. *Biotechnol Bioeng*, 59(2):227–238, 1998.

M. Savageau. Design principles for elementary gene circuits: Elements, methods, and examples. *Chaos*, 11(1):142–159, 2001.

M. A. Savageau. Biochemical systems analysis. a study of function and design in molecular biology. In *ADDISON WESLEY PUBL.* 1976.

C. H. Schilling, D. Letscher, and B. Ø. Palsson. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J. Theor. Biol.*, 203(3):229–248, 2000.

S. Schnell and P. Maini. Enzyme kinetics at high enzyme concentration. *Bull. Math. Biol.*, 62(3): 483–499, 2000.

S. Schnell and P. K. Maini. A century of enzyme kinetics. should we believe in the Km and Vmax estimates? *Comments Theor. Biol*, 8:169–187, 2003.

R. Schuetz, L. Kuepfer, and U. Sauer. Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Molecular systems biology*, 3(1):119, 2007.

S. Schuster, T. Dandekar, and D. A. Fell. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol*, 17(2): 53–60, 1999.

I. Segel. *Enzyme kinetics: behavior and analysis of rapid equilibrium and steady state enzyme systems.* Wiley & Sons, 1993.

L. A. Segel. On the validity of the steady state assumption of enzyme kinetics. *Bull. Math. Biol.*, 50(6):579–593, 1988.

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

M. Shamir, Y. Bar-On, R. Phillips, and R. Milo. Snapshot: timescales in cell biology. *Cell*, 164 (6):1302–1302, 2016.

T. Shlomi, M. N. Cabili, M. J. Herrgård, B. Ø. Palsson, and E. Ruppin. Network-based prediction of human tissue-specific metabolism. *Nature Biotechnol.*, 26(9):1003–1010, 2008.

K. Smallbone, E. Simeonidis, D. S. Broomhead, and D. B. Kell. Something from nothing- bridging the gap between constraint-based and kinetic modelling. *The FEBS journal*, 274(21):5576–5585, 2007.

C. D. Smolke and P. A. Silver. Informing biological design by integration of systems and synthetic biology. *Cell*, 144(6):855–859, 2011.

D. Stefan, C. Pinel, S. Pinhal, E. Cinquemani, J. Geiselmann, and H. de Jong. Inference of quantitative models of bacterial promoters from time-series reporter gene data. *PLoS Comput Biol*, 11(1):e1004028, 2015.

G. Stephanopoulos, A. Aristidou, and J. Nielsen. *Metabolic Engineering: Principles and Methodologies*. Academic Press, San Diego, CA, 1998.

W. Stroberg and S. Schnell. On the estimation errors of KM and V from time-course experiments using the michaelis–menten equation. *Biophys Chem*, 219:17–27, 2016.

J. Tang and W. Riley. A total quasi-steady-state formulation of substrate uptake kinetics in complex networks and an example application to microbial litter decomposition. *Biogeosciences*, 10 (12):8329–8351, 2013.

B. H. ter Kuile and H. V. Westerhoff. Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway. *FEBS Lett*, 500(3):169–171, 2001.

I. Thiele, S. Sahoo, A. Heinken, J. Hertel, L. Heirendt, M. K. Aurich, and R. M. Fleming. Personalized whole-body models integrate metabolism, physiology, and the gut microbiome. *Mol. Syst. Biol.*, 16(5):e8982, 2020.

R. Thomas and R. d'Ari. *Biological Feedback*. CRC Press, Boca Raton, FL, 1990.

M. Tian and J. L. Reed. Integrating proteomic or transcriptomic data into metabolic models using linear bound flux balance analysis. *Bioinformatics*, 34(22):3882–3888, 2018.

J. Timmermans and L. Van Melderen. Conditional essentiality of the i gene in *Escherichia coli*. *J. Bacteriol.*, 191(5):1722–1724, 2009.

K. Tummler, T. Lubitz, M. Schelker, and E. Klipp. New types of experimental data shape the use of enzyme kinetics for dynamic network modeling. *FEBS J*, 281(2):549–571, 2014.

A. Turing. The chemical theory of morphogenesis. *Phil. Trans. Roy. Soc*, 13(1), 1952.

A. Tzafriri. Michaelis-Menten kinetics at high enzyme concentrations. *Bull. Math. Biol.*, 65(6): 1111–1129, 2003.

A. Tzafriri, M. Bercovier, and H. Parnas. Reaction diffusion model of the enzymatic erosion of insoluble fibrillar matrices. *Biophys J*, 83(2):776–793, 2002.

R. J. van Berlo, D. de Ridder, J.-M. Daran, P. A. Daran-Lapujade, B. Teusink, and M. J. Reinders. Predicting metabolic fluxes using gene expression differences as constraints. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, 8(1):206–216, 2009.

K. van Eunen, S. Rossell, J. Bouwman, H. V. Westerhoff, and B. M. Bakker. Quantitative analysis of flux regulation through hierarchical regulation analysis. In *Methods Enzymol.*, volume 500, pages 571–595. Elsevier, 2011.

A. Varma and B. Palsson. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* w3110. *Appl Environ Microbiol*, 60 (10):3724–3731, 1994a.

A. Varma and B. O. Palsson. Metabolic flux balancing: basic concepts, scientific and practical use. *Nature Biotechnol*, 12(10):994–998, 1994b.

E. Voit. *A first course in systems biology*. Garland Science, 2017.

E. O. Voit, H. A. Martens, and S. W. Omholt. 150 years of the mass action law. *PLoS Comput. Biol.*, 11(1):e1004012, 2015.

S. Volkova, M. R. Matos, M. Mattanovich, and I. Marín de Mas. Metabolic modelling as a framework for metabolomics data integration and analysis. *Metabolites*, 10(8):303, 2020.

L. von Bertalanffy. Der organismus als physikalisches system betrachtet. *Naturwissenschaften*, 28 (33):521–531, 1940.

L. Von Bertalanffy. The theory of open systems in physics and biology. *Science*, 111(2872):23–29, 1950.

L. Von Bertalanffy. General system theory: Foundations, development, applications. Technical report, Georges Braziller, Inc., 1969.

G. Wahba. *Spline models for observational data*. SIAM, 1990.

Y. Wang, J. A. Eddy, and N. D. Price. Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Syst Biol*, 6(1):153, 2012.

A. Y. Weiße, D. A. Oyarzún, V. Danos, and P. S. Swain. Mechanistic links between cellular trade-offs, gene expression, and growth. *Proc Natl Acad Sci USA*, 112(9):E1038–E1047, 2015.

H. V. Westerhoff and B. O. Palsson. The evolution of molecular biology into systems biology. *Nature Biotechnol.*, 22(10):1249–1252, 2004.

S. J. Wiback, I. Famili, H. J. Greenberg, and B. Ø. Palsson. Monte Carlo sampling can be used to determine the size and shape of the steady-state flux space. *J. Theor. Biol.*, 228(4):437–447, 2004.

W. Wiechert. 13c metabolic flux analysis. *Metab. Eng.*, 3(3):195–206, 2001.

N. Wiener. *Cybernetics or Control and Communication in the Animal and the Machine*. MIT press, 1948.

A. J. Wolfe. The acetate switch. *Microbio.l Mol. Biol. Rev.*, 69(1):12–50, 2005.

O. Wolkenhauer. Systems biology: the reincarnation of systems theory applied in biology? *Brief. Bioinform.*, 2(3):258–270, 2001.

K. Yizhak, T. Benyamini, W. Liebermeister, E. Ruppin, and T. Shlomi. Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. *Bioinformatics*, 26(12):i255–i260, 2010.

V. Zulkower, M. Page, D. Ropers, J. Geiselmann, and H. de Jong. Robust reconstruction of gene expression profiles from reporter gene data using linear inversion. *Bioinformatics*, 31(12):i71–i79, 2015.

C. Zupke and G. Stephanopoulos. Modeling of isotope distributions and intracellular fluxes in metabolic networks using atom mapping matrixes. *Biotechnol. Prog.*, 10(5):489–498, 1994.