Universidade do Minho
Escola de Engenharia

Carlos Daniel Moutinho Machado

Novel modeling formalisms and simulation tools in Computational Biosystems

Outubro de 2011

Novel modeling formalisms and simulation tools in Computational Biosystems
Carlos Daniel Moutinho Machado
UMinho | 2011

Universidade do Minho
Escola de Engenharia

Carlos Daniel Moutinho Machado

Novel modeling formalisms and simulation tools in Computational Biosystems

Doutoramento em Bioengenharia

Trabalho efectuado sob a orientação de:
Doutor Eugenio Manuel de Faria Campos Ferreira
Doutora Isabel Cristina de Almeida Pereira da Rocha
Doutor Bruce Tidor

Outubro de 2011

Autor
Carlos Daniel Moutinho Machado

Email: dmachado@deb.uminho.pt
Telefone: +351 253 604 400
BI: 12309439

Título da tese
Novel modeling formalisms and simulation tools in Computational Biosystems

Orientadores
Doutor Eugénio Manuel de Faria Campos Ferreira
Doutora Isabel Cristina de Almeida Pereira da Rocha
Doutor Bruce Tidor

Ano de conclusão: 2011
Doutoramento em Bioengenharia

Universidade do Minho, Outubro de 2011

Acknowledgements

Completing this PhD is an achievement that could not have been possible without the support from several people. Therefore, I begin by acknowledging the financial support from the portuguese government through my FCT grant from the MIT-Portugal Program. I also thank my supervisors for their guidance both in scientific and personal aspects. During the last 4 years I had the opportunity to work in different places and meet many people. The help of my colleagues from the first year of the doctoral program was fundamental for surviving those advanced bioengineering classes in topics unfamiliar to a computer science guy. I also thank everyone at the Mobile and Anthropomorphic Robotics lab, where I did my robotics project, for receiving me so warmly. My first year of research was spent in the graduate students office at DEB, where I got the chance to interact with a bunch of friendly people. During that time I also took the opportunity to sneak into some classes of the Bioinformatics masters. I thank my colleagues there for treating me as one of their own. I spent the last years in our Systems Biology group, where we had fruitful work discussions and also fun time together outside of the lab. This PhD also gave me the opportunity to live abroad for a while. It was a great experience that made me grow a lot. Many people contributed to make that experience so fulfilling, including everyone I met in the Yellow house, in the well-being seminars and the portuguese students in Boston. I thank everyone at the Tidor lab for all the good time working together. Finally, I thank my family and friends for their support throughout the years, specially my mother for bearing my absence during my time abroad. Last but not least, I thank Gisela, who I've met during this PhD, and has been my support ever since.

The work presented in this thesis was financially supported by a research grant (SFRH/BD/35215/2007) from the Fundação para a Ciência e a Tecnologia, and funded by FEDER in Programa Operacional Factores de Competitividade - COMPETE, through the project "Bridging Systems and Synthetic Biology for the development of Improved Microbial Cell Factories" - MIT-Pt/BS-BB/0082/2008 (FCOMP-01-0124-FEDER-008433).

Abstract

The goal of Systems Biology is to understand the complex behavior that emerges from the interaction among the cellular components. Industrial biotechnology is one of the areas of application, where new approaches for metabolic engineering are developed, through the creation of new models and tools for simulation and optimization of the microbial metabolism. Although whole-cell modeling is one of the goals of Systems Biology, so far most models address only one kind of biological network independently. This work explores the integration of different kinds of biological networks with a focus on the improvement of simulation of cellular metabolism. The bacterium *Escherichia coli* is the most well characterized model organism and is used as our case-study.

An extensive review of modeling formalisms that have been used in Systems Biology is presented in this work. It includes several formalisms, including Boolean networks, Bayesian networks, Petri nets, process algebras, constraint-based models, differential equations, rule-based models, interacting state machines, cellular automata and agent-based models. We compare the features provided by these formalisms and classify the most suitable ones for the creation of a common framework for modeling, analysis and simulation of integrated biological networks.

Currently, there is a separation between dynamic and constraint-based modeling of metabolism. Dynamic models are based on detailed kinetic reconstructions of central metabolic pathways, whereas constraint-based models are based on genome-scale stoichiometric reconstructions. Here, we explore the gap between both formulations and evaluate how dynamic models can be used to reduce the solution space of constraint-based models in order

to eliminate kinetically infeasible solutions.

The limitations of both kinds of models are leading to new approaches to build kinetic models at the genome-scale. The generation of kinetic models from stoichiometric reconstructions can be performed within the same framework as a transformation from discrete to continuous Petri nets. However, the size of these networks results in models with a large number of parameters. In this scope, we develop and implement structural reduction methods that adjust the level of detail of metabolic networks without loss of information, which can be applied prior to the kinetic inference to build dynamic models with a smaller number of parameters.

In order to account for enzymatic regulation, which is not present in constraint-based models, we propose the utilization of Extended Petri nets. This results in a better scaffold for the kinetic inference process. We evaluate the impact of accounting for enzymatic regulation in the simulation of the steady-state phenotype of mutant strains by performing knockouts and adjustment of enzyme expression levels. It can be observed that in some cases the impact is significant and may reveal new targets for rational strain design.

In summary, we have created a solid framework with a common formalism and methods for metabolic modeling. This will facilitate the integration with gene regulatory networks, as we have already addressed many issues also associated with these networks, such as the trade-off between size and detail, and the representation of regulatory interactions.

Resumo

O objectivo da Biologia de Sistemas é compreender os comportamentos que resultam das complexas interacções entre todos os componentes celulares. A biotecnologia industrial é uma das áreas de aplicação, onde novas abordagens para a engenharia metabólica são desenvolvidas através da criação de novos modelos e ferramentas de simulação e optimização do metabolismo microbiano. Apesar de um dos principais objectivos da Biologia de Sistemas ser a criação de um modelo completo de uma célula, até ao momento a maioria dos modelos desenvolvidos incorpora de forma separada cada tipo de rede biológica. Este trabalho explora a integração de diferentes tipos de redes biológicas, focando melhorar a simulação do metabolismo celular. A bactéria *Escherichia coli* é o organismo modelo que está melhor caracterizado e é usado como caso de estudo.

Neste trabalho é elaborada uma extensa revisão dos formalismos de modelação que têm sido utilizados em Biologia de Sistemas. São considerados vários formalismos tais como, redes Booleanas, redes Bayesianas, redes de Petri, álgebras de processos, modelos baseados em restrições, equações diferenciais, modelos baseados em regras, máquinas de interacção de estados, autómatos celulares e modelos baseados em agentes. As funcionalidades inerentes a estes formalismos são analisadas de forma a classificar os mesmos pelo seu potencial em servir de base à criação de uma plataforma para modelação, análise e simulação de redes biológicas integradas.

Actualmente, existe uma separação entre modelação dinâmica e modelação baseada em restrições para o metabolismo celular. Os modelos dinâmicos consistem em reconstruções cinéticas detalhadas de vias centrais do metabolismo, enquanto que os modelos baseados em restrições são construídos à

escala genómica com base apenas na estequiometria das reacções. Neste trabalho explora-se a separação entre os dois tipos de formulação e é avaliada a forma como os modelos dinâmicos podem ser utilizados para reduzir o espaço de soluções de modelos baseados em restrições de forma a eliminar soluções inalcançáveis.

As limitações impostas por ambos os tipos de modelos estão a conduzir à criação de novas abordagens para a construção de modelos cinéticos à escala genómica. A geração de modelos cinéticos a partir de reconstruções estequiométricas pode ser feita dentro de um mesmo formalismo através da transformação de redes de Petri discretas em redes de Petri contínuas. No entanto, devido ao tamanho destas redes, os modelos resultantes incluem um número extremamente grande de parâmetros. Neste trabalho são implementados métodos para a redução estrutural de redes metabólicas sem perda de informação, que permitem ajustar o nível de detalhe das redes. Estes métodos podem ser aplicados à inferência de cinéticas, de forma a gerar modelos dinâmicos com um menor número de parâmetros.

De forma a considerar efeitos de regulação enzimática, os quais não são representados em modelos baseados em restrições, propõe-se a utilização de redes de Petri complementadas com arcos regulatórios. Este formalismo é utilizado como base para o processo de inferência cinética. A influência da regulação enzimática na determinação do estado estacionário de estirpes mutantes é avaliada através da análise da remoção de reacções e da variação dos níveis de expressão enzimática. Observa-se que em alguns casos esta influência é significativa e pode ser utilizada para obter novas estratégias de manipulação de estirpes.

Em suma, neste trabalho foi criada uma plataforma sólida para modelação do metabolismo baseada num formalismo comum. Esta plataforma facilitará a integração com redes de regulação genética, pois foram abordados vários problemas que também se colocam nestas redes, tais como o ajuste entre o tamanho da rede e o seu nível de detalhe, bem como a representação de interacções regulatórias entre componentes da rede.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Systems Biology

The cell is the fundamental building block of life. From this basic unit, a myriad of life forms have emerged. Systems Biology is a recent field of study that focuses on the complex interactions that happen inside a cell. It represents a new paradigm when compared to classical biology, as it looks at the cell as a whole rather than enumerating its parts [40, 41].

A single cell is composed by thousands of components such as genes, proteins and metabolites. These components interact in several ways, forming complex biological networks. The behavior of the cell emerges not only from the structure but also from the dynamics of these networks. A common analogy in the community is the functioning of a radio. It is not possible to fix a radio, if all the parts are disassembled and we do not know how to put them together [42].

Unlike the radio, which has its own blueprint, the design of the cell has evolved in nature. Reverse engineering its *specification* involves collecting and analyzing vast amounts of data. The rise of Systems Biology is related to the development of high-throughput technologies in the past years, that have generated several of the so-called *omics*, including genomics [67], transcriptomics [83], proteomics [25], metabolomics [24], and fluxomics [71]. These data allow the quantification of the molecular species present in the

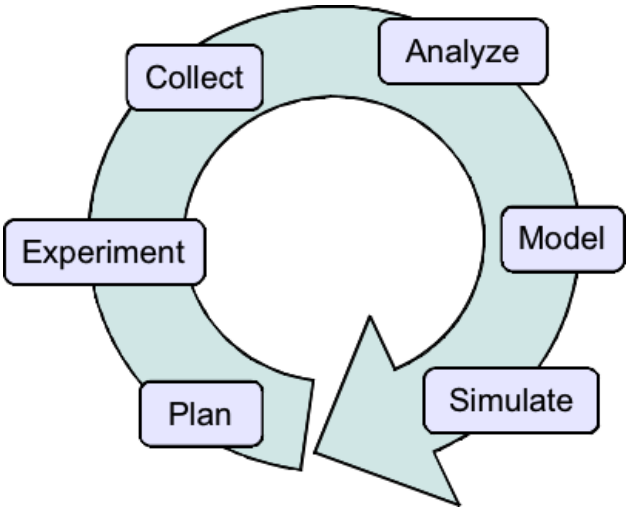

Figure 1.1: Model development cycle.

cell and the reconstruction of their map of interactions. With this information we can build models for *in silico* simulation. Models can also be used to drive new experiments that will allow model refinement. Hence, the model building cycle is an iterative process that alternates between *wet* and *dry* lab experiments (Figure 1.1).

The synergy between computational and experimental methods, comes from the multidisciplinary nature of this field. It requires cooperation among researchers from several areas such as computer science, mathematics, physics, chemistry, biology and systems engineering. In order to facilitate communication and sharing of resources, several standards and tools have been developed [37]. These include standards for representing experimental results (*e.g.,* MIAME [8], MIAPE [78]), and for representing models (*e.g.,* SBML [30], CellML [50]). Databases are also a fundamental resource for information sharing. Some examples that can be used to build a biological model are databases with gene annotation (*e.g.,* Entrez Gene [52]), pathway information (*e.g.,* KEGG [35]) and enzyme kinetics (*e.g.,* BRENDA [69], SABIO-RK [65]). Curated models can also be found in publicly available databases (*e.g.,* Biomodels.net [43]). Several tools have been developed to construct, analyze and simulate biological models (*e.g.,* CellDesigner [21], Cytoscape [72], COPASI [28], OptFlux [64]).

Despite all the resources and tools available, the ultimate goal of Systems Biology to build a fully detailed cell model is still far from complete. Although there are initiatives for whole cell simulation, such as the E-cell project [80, 79], the fact that the models developed are detailed at the biochemical kinetic level, has limited their scope to particular biological processes (*e.g.,* mitochondrial metabolism [84], circadian clocks [75]). Recently, models based on less detailed descriptions that integrate all kinds of biological networks, started to appear [12, 44].

Systems Biology has been growing both in academia and industry. The *in silico* approach provides a fast and inexpensive way to run experiments and test new hypothesis. Moreover, although it does not replace laboratory experiments completely, it can guide experimental design methods that save time and resources in the laboratory [16]. Computational models can be used to *understand* biological systems through simulations that predict the cellular behavior, and also to *redesign* these systems by finding the required manipulations towards a desired goal. Therefore, they have several applications both in science and engineering areas.

In biomedical research, computational models are used as a framework for studying disease mechanisms [29, 60] and for drug discovery [10]. Using *in silico* experiments, researchers can find specific drug targets and study the effect of new drugs at a system-wide level, avoiding potential side-effects.

Industrial biotechnology is another major area of application for Systems Biology [55]. The creation of microbial factories has a high impact in society in terms of economy and sustainability. These applications involve the redesign of the microbial metabolism for specific production goals. This will be the main focus of this work and will be explored in more detail in the next section.

1.2 Metabolic Engineering

The utilization of microbial factories began many centuries ago, even before microorganisms were discovered. Their first application was the production of alcoholic beverages such as wine and beer. Nowadays, biotechnology

has widespread use in industry for the production of commodity chemicals and materials [22]. Industrial biotechnology provides cost competitive, environmental friendly and sustainable alternatives to existing chemical-based production processes [55].

Among biotechnologicaly produced compounds are pharmaceutical drugs such as penicillin [59] and insulin [61]. Another major market is the production of nutrients such as vitamins [81, 77] and amino acids [31, 47]. The sustainability and environmental concerns in the recent years have led to a demand of renewable energy sources as an alternative to petroleum-based fuels, creating an important market for biofuel production [38, 39].

For biotechnological production to be cost competitive, it is necessary that the bioconversion of substrates into products has the highest possible yield. Metabolic engineering is the modification of cellular metabolism for optimization of a desired production objective [76]. Traditionally, this was done by directed evolution [3] or random mutagenesis [66]. Although these approaches have been successful (*e.g.,* over 1000-fold increase in penicillin production [53]), they do not elucidate the genetic changes that have occurred in the cells. Consequently, it is not possible to obtain knowledge of the target changes and to reapply them in further strain improvements.

Metabolic engineering, on the other hand, uses rational strain design to redirect the metabolic flux towards the desired objective. This can be achieved by changing the expression of the genes associated with the metabolic enzymes, such as gene knockout in key metabolic branch points and increased enzyme expression in the target pathways. Additionally, it is also possible to add new metabolic functions to an organism by heterologous gene expression. These manipulations became possible when molecular biology techniques such as recombinant DNA emerged [23].

In order to guide the rational design of mutant microbial strains it is necessary to have metabolic models that are able to simulate and predict the metabolic phenotype of such mutants. The earlier models were based on mechanistic kinetic descriptions of some of the central metabolic reactions. Metabolic Control Analysis (MCA) and Biochemical Systems Theory (BST) are two mathematical frameworks that were developed to analyze the key

parameters that control the metabolic flux [13]. However, the difficulty of obtaining kinetic data for building larger models, together with the recent development of genome sequencing technologies have decreased the popularity of these models in favor of genome-scale stoichiometric models [14]. In chapter 3 the connection between these two kind of models is explored in detail.

Genome sequencing is the first step for creating a genome-scale metabolic network reconstruction. The advances in DNA sequencing techniques have greatly decreased the time and cost required for sequencing the complete genome of an organism [70, 2]. For this reason, the number of sequenced genomes has been growing exponentially, with hundreds of complete sequences currently available [55]. The model reconstruction process begins with functional annotation of the genome sequences, by comparing against sequence databases such as GenBank [6] and Entrez Gene [52], followed by search in metabolic databases such as KEGG [35] and SEED [56] to obtain the gene-protein-reaction (GPR) associations. This process can be done automatically and complemented with manual curation using literature data and organism-specific knowledge [18].

There are currently over 50 genome-scale metabolic reconstructions available for a variety of organisms [54]. These allow the simulation of the metabolic phenotype under steady-state conditions, using constraint-based methods such as the popular Flux Balance Analysis (FBA) [82, 15]. In order to find the optimal changes for mutant strain design, methods have been developed that find optimal knockout sets [9, 58] and also target enzymes for the adjustment of expression levels [62].

1.3 *Escherichia coli* as a model organism

Perhaps the most fundamental aspect of an industrial biotechnological application is the selection of the ideal microorganism to perform as a cellular factory. *Escherichia coli* is a bacterium commonly found in the intestinal flora, and is the most well studied prokaryotic organism. It can be easily cultivated in the laboratory with inexpensive medium, making it a suitable

model organism for several applications. In particular, *E. coli* K-12 is the most widely used strain in laboratories worldwide and has been a workhorse for industrial biotechnology [63].

E. coli K-12 was one of the first organisms to be completely sequenced [7], and it currently has several dedicated projects and resource databases including the International *E. coli* Alliance (IECA) [27], EcoGene [68] and EcoCyc [36]. Also, some metabolic models of this organism have been published and are publicly available, including kinetic reconstructions of the central metabolism [11] and genome-scale stoichiometric reconstructions [17]. For these reasons, *E. coli* was selected as the case-study for all the work presented in this thesis.

There are many successful cases of metabolic engineering applications that use *E. coli* strains designed for overproduction of target metabolites [19]. These include the production of amino acids such as L-valine [57] and L-threonine [45] as well as other organic compounds for food industry like vanillin [49], lactic acid [20], succinic acid [46] and lycopene [1]. This microbe has also had a significant role in the production of different sources of biofuels [32, 34, 51, 5, 4].

1.4 Motivation for this work

Despite the advances in the area of Systems Biology, and the many successful cases of metabolic engineering applications, we have not yet reached a point where we can precisely predict all possible outcomes of the cellular behavior. The reason is that we have not yet built a fully detailed whole-cell model for any model organism. In fact, with few exceptions [12, 44], most models address only each of the main kinds of biological networks (signaling, gene regulatory and metabolic) individually. Also, given the heterogeneous background of the Systems Biology community, these models have been based on a myriad of different formalisms.

Integration of different kinds of biological networks is fundamental for accurate simulations of cellular behavior. For instance, the simulation of a gene knockout in a metabolic network is simulated by setting the flux carried

by the corresponding enzyme to zero. However, this approach disregards all possible consequent adjustments at the regulatory level. This is of utmost importance if we consider how cells are robustly designed [48]. Therefore, this field would benefit from the creation of a modeling framework with a common modeling formalism, analysis and simulation methods, that supports all kinds of biological networks. The work presented in this thesis, addresses the problem of integration of biological network models, with a focus on improving metabolic network simulation.

1.5 Thesis Outline

Chapter 1 (current chapter) gives a brief introduction to the field of Systems Biology, its application to metabolic engineering, and the motivation for this work.

Chapter 2 is a review on modeling formalisms that have been used in Systems Biology to model all major kinds of biological networks. The formalisms are compared and analyzed in terms of features provided and successful applications to different biological networks. The goal is to find the best candidate suitable for whole-cell modeling. One of the key conclusions is that Petri nets [26] are a solid candidate for such purpose. Another important observation is that there is a separation of two distinct approaches for modeling metabolic networks, which is analyzed in detail in the sequent chapter.

Chapter 3 explores the gap between dynamic and constraint-based models of metabolism. This is a gap that comes from two opposite model building approaches. The first is performed in a bottom-up fashion by putting together kinetic equations for reactions that have been experimentally characterized, while the second begins with a genome-scale stoichiometric reconstruction and constrains the metabolic phenotype in a top-down approach by adding equilibrium, thermodynamic and flux constraints. This gap can also be found in gene regulatory and signaling networks. Therefore, it is important to address this problem if a true integration of different biological networks is foreseen.

Chapter 4 presents a Petri net based framework for metabolic model

transformation. Given the limitations of constraint-based models to model dynamic behavior, there have been recent efforts to generate kinetic models from stoichiometric reconstructions [73, 74, 33]. Petri nets are a suitable formalism to integrate different kinds of biological networks. In this particular case, discrete and continuous Petri nets can be used to represent, respectively, stoichiometric and dynamic models. Therefore, we propose this formalism as an unifying framework for the kinetic inference process. Moreover, the usually large size of stoichiometric reconstructions results in dynamic models with a huge number of parameters. To address this problem, we have developed structural transformation methods for network reduction that can be used prior to the kinetic inference process, in order to reduce the size of the generated models.

Chapter 5 explores the influence of enzymatic regulation in the steady-state flux distribution simulated with dynamic models of metabolism. Current approaches for building kinetic metabolic models at the genome-scale do not account for this kind of regulation. We propose the utilization of extended Petri Nets as a suitable formalism for modeling metabolic networks that account for enzymatic regulation, and extend the kinetic inference process to this new framework. We compare the metabolic phenotype of mutant strains simulated with regulated and unregulated models and observe significant differences in many cases.

Chapter 6 wraps up with general conclusions derived from this work and elaborates on perspectives for future directions.

1.5.1 Publications derived from this work

During the development of this work, several publications were written. Chapters 2–5 are based on those publications, adapted with minor changes to the format of this thesis. The three supervisors of the thesis are co-authors of all publications and revised the final manuscripts. Rafael Costa contributed with ideas and discussion, mainly in the topics of dynamic modeling and parameter estimation, and is also co-author of all publications.

Chapter 2 is based on the review article *"Modeling formalisms in Systems*

Biology" (submitted), which is a extended version of the article *"A critical review on modelling formalisms and simulation tools in Computational Biosystems"* published in *Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living*, volume 5518, pages 1063–1070, 2009. Miguel Rocha helped reviewing the computational formalisms and is co-author of this publication.

Chapter 3 is based on the article *"Exploring the gap between dynamic and constraint-based models of metabolism"* (in preparation).

Chapter 4 is based on the article *"Model transformation of metabolic networks using a Petri net based framework"* published at the *International Workshop on Biological Processes & Petri Nets (BioPPN 2010)*, Braga, Portugal, 2010.

Chapter 5 is based on the article *"Accounting for enzymatic regulation in large-scale kinetic reconstructions of metabolism"* (in preparation).

Bibliography

[1] H. Alper, K. Miyaoku, and G. Stephanopoulos. Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets. *Nature Biotechnology*, 23(5):612–616, 2005.

[2] W.J. Ansorge. Next-generation DNA sequencing techniques. *New Biotechnology*, 25(4):195–203, 2009.

[3] F.H. Arnold. Design by directed evolution. *Accounts of Chemical Research*, 31(3):125–131, 1998.

[4] S. Atsumi, A.F. Cann, M.R. Connor, C.R. Shen, K.M. Smith, M.P. Brynildsen, K.J.Y. Chou, T. Hanai, and J.C. Liao. Metabolic engineering of *Escherichia coli* for 1-butanol production. *Metabolic Engineering*, 10(6):305–311, 2008.

[5] S. Atsumi and J.C. Liao. Metabolic engineering for advanced biofuels production from *Escherichia coli*. *Current Opinion in Biotechnology*, 19(5):414–419, 2008.

[6] D.A. Benson, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, and D.L. Wheeler. GenBank. *Nucleic Acids Research*, 36(suppl 1):D25–D30, 2008.

[7] F.R. Blattner, G. Plunkett, C.A. Bloch, N.T. Perna, V. Burland, M. Riley, J. Collado-Vides, J.D. Glasner, C.K. Rode, G.F. Mayhew, et al. The complete genome sequence of *Escherichia coli* K-12. *Science*, 277(5331):1453–1462, 1997.

[8] A. Brazma, P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, J. Aach, W. Ansorge, C.A. Ball, H.C. Causton, et al. Minimum information about a microarray experiment (MIAME) — toward standards for microarray data. *Nature Genetics*, 29(4):365–371, 2001.

[9] A.P. Burgard, P. Pharkya, and C.D. Maranas. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and Bioengineering*, 84(6):647–657, 2003.

[10] E.C. Butcher, E.L. Berg, and E.J. Kunkel. Systems biology in drug discovery. *Nature Biotechnology*, 22(10):1253–1259, 2004.

[11] C. Chassagnole, N. Noisommit-Rizzi, J.W. Schmid, K. Mauch, and M. Reuss. Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnology and Bioengineering*, 79(1):53–73, 2002.

[12] 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.

[13] R. Curto, A. Sorribas, and M. Cascante. Comparative Characterization of the Fermentation Pathway of *Saccharomyces cerevisiae* Using Biochemical Systems Theory and Metabolic Control Analysis: Model Definition and Nomenclature. *Mathematical Biosciences*, 130(1):25–50, 1995.

[14] M. Durot, P.Y. Bourguignon, and V. Schachter. Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS Microbiology Reviews*, 33(1):164–190, 2009.

[15] J.S. Edwards and B.O. Palsson. Metabolic flux balance analysis and the *in silico* analysis of *Escherichia coli* K-12 gene deletions. *BMC Bioinformatics*, 1(1):1, 2000.

[16] D. Faller, U. Klingmüller, and J. Timmer. Simulation Methods for Optimal Experimental Design in Systems Biology. *Simulation*, 79(12):717–725, 2003.

[17] A.M. Feist, C.S. Henry, J.L. Reed, M. Krummenacker, A.R. Joyce, P.D. Karp, L.J. Broadbelt, V. Hatzimanikatis, and B.Ø. Palsson. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*, 3(121), 2007.

[18] A.M. Feist, M.J. Herrgård, I. Thiele, J.L. Reed, and B.Ø. Palsson. Reconstruction of biochemical networks in microorganisms. *Nature Reviews Microbiology*, 7(2):129–143, 2008.

[19] A.M. Feist and B.Ø. Palsson. The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nature Biotechnology*, 26(6):659–667, 2008.

[20] S.S. Fong, A.P. Burgard, C.D. Herring, E.M. Knight, F.R. Blattner, C.D. Maranas, and B.O. Palsson. *In silico* design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnology and Bioengineering*, 91(5):643–648, 2005.

[21] A. Funahashi, M. Morohashi, H. Kitano, and N. Tanimura. CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. *BIOSILICO*, 1(5):159–162, 2003.

[22] M. Gavrilescu and Y. Chisti. Biotechnology–a sustainable alternative for chemical industry. *Biotechnology Advances*, 23(7-8):471–499, 2005.

[23] B.R. Glick and J.J. Pasternak. *Molecular Biotechnology: Principles and applications of recombinant DNA*. American Society for Microbiology, Washington, DC, 1994.

[24] R. Goodacre, S. Vaidyanathan, W.B. Dunn, G.G. Harrigan, and D.B. Kell. Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends in Biotechnology*, 22(5):245–252, 2004.

[25] A. Görg, W. Weiss, and M.J. Dunn. Current two-dimensional electrophoresis technology for proteomics. *Proteomics*, 4(12):3665–3685, 2004.

[26] M. Heiner, D. Gilbert, and R. Donaldson. Petri nets for systems and synthetic biology. *Formal Methods for Computational Systems Biology*, 5016:215–264, 2008.

[27] C. Holden. Alliance Launched to Model *E. coli*. *Science*, 297(5586):1459–1460, 2002.

[28] S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, and U. Kummer. COPASI — a COmplex PAthway SImulator. *Bioinformatics*, 22(24):3067–3074, 2006.

[29] J.J. Hornberg, F.J. Bruggeman, H.V. Westerhoff, and J. Lankelma. Cancer: A Systems Biology disease. *Biosystems*, 83(2–3):81–90, 2006.

[30] M. Hucka, A. Finney, HM Sauro, H. Bolouri, JC Doyle, H. Kitano, AP Arkin, BJ Bornstein, D. Bray, A. Cornish-Bowden, et al. The Systems Biology Markup Language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, 2003.

[31] M. Ikeda. Amino Acid Production Processes. *Advances in Biochemical Engineering/Biotechnology*, 79:1–35, 2003.

[32] LO Ingram, T. Conway, DP Clark, GW Sewell, and JF Preston. Genetic engineering of ethanol production in *Escherichia coli*. *Applied and Environmental Microbiology*, 53(10):2420–2425, 1987.

[33] N. Jamshidi and B.O. Palsson. Mass action stoichiometric simulation models: Incorporating kinetics and regulation into stoichiometric models. *Biophysical Journal*, 98:175–185, 2010.

[34] R. Kalscheuer, T. Stölting, and A. Steinbüchel. Microdiesel: *Escherichia coli* engineered for fuel production. *Microbiology*, 152(9):2529–2536, 2006.

[35] M. Kanehisa and S. Goto. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1):27–30, 2000.

[36] I.M. Keseler, J. Collado-Vides, S. Gama-Castro, J. Ingraham, S. Paley, I.T. Paulsen, M. Peralta-Gil, and P.D. Karp. EcoCyc: a comprehensive database resource for *Escherichia coli*. *Nucleic Acids Research*, 33(suppl 1):D334–D337, 2005.

[37] J.S. Kim, H.S. Yun, H.U. Kim, H.S. Choi, T.Y. Kim, H.M. Woo, and S.Y. Lee. Resources for Systems Biology Research. *Journal of Microbiology and Biotechnology*, 16(6):832–848, 2006.

[38] S. Kim and B.E. Dale. Global potential bioethanol production from wasted crops and crop residues. *Biomass and Bioenergy*, 26(4):361–375, 2004.

[39] S. Kim and B.E. Dale. Life cycle assessment of various cropping systems utilized for producing biofuels: Bioethanol and biodiesel. *Biomass and Bioenergy*, 29(6):426–439, 2005.

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

[41] H. Kitano. Systems Biology: A Brief Overview. *Science*, 295(5560):1662–1664, 2002.

[42] Y. Lazebnik. Can a biologist fix a radio?–Or, what I learned while studying apoptosis. *Cancer cell*, 2(3):179–182, 2002.

[43] N. Le Novere, B. Bornstein, A. Broicher, M. Courtot, M. Donizelli, H. Dharuri, L. Li, H. Sauro, M. Schilstra, B. Shapiro, et al. BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Research*, 34(suppl 1):D689, 2006.

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

[45] K.H. Lee, J.H. Park, T.Y. Kim, H.U. Kim, and S.Y. Lee. Systems metabolic engineering of *Escherichia coli* for L-threonine production. *Molecular Systems Biology*, 3(149), 2007.

[46] S.J. Lee, D.Y. Lee, T.Y. Kim, B.H. Kim, J. Lee, and S.Y. Lee. Metabolic engineering of *Escherichia coli* for enhanced production of succinic acid, based on genome comparison and *in silico* gene knockout simulation. *Applied and Environmental Microbiology*, 71(12):7880, 2005.

[47] W. Leuchtenberger, K. Huthmacher, and K. Drauz. Biotechnological production of amino acids and derivatives: current status and prospects. *Applied Microbiology and Biotechnology*, 69(1):1–8, 2005.

[48] F. Li, T. Long, Y. Lu, Q. Ouyang, and C. Tang. The yeast cell-cycle network is robustly designed. *Proceedings of the National Academy of Sciences of the United States of America*, 101(14):4781–4786, 2004.

[49] K. Li and J.W. Frost. Synthesis of Vanillin from Glucose. *Journal of the American Chemical Society*, 120(40):10545–10546, 1998.

[50] C.M. Lloyd, M.D.B. Halstead, and P.F. Nielsen. CellML: its future, present and past. *Progress in Biophysics and Molecular Biology*, 85(2–3):433–450, 2004.

[51] X. Lu, H. Vora, and C. Khosla. Overproduction of free fatty acids in *E. coli*: implications for biodiesel production. *Metabolic Engineering*, 10(6):333–339, 2008.

[52] D. Maglott, J. Ostell, K.D. Pruitt, and T. Tatusova. Entrez Gene: gene–centered information at NCBI. *Nucleic Acids Research*, 33(suppl 1):D54–D58, 2005.

[53] J. Nielsen. *Physiological Engineering Aspects of Penicillium chrysogenum*. World Scientific, 1997.

[54] M.A. Oberhardt, J. Puchałka, V.A.P.M. dos Santos, and J.A. Papin. Reconciliation of Genome-Scale Metabolic Reconstructions for Compar-

ative Systems Analysis. *PLoS Computational Biology*, 7(3):e1001116, 2011.

[55] J.M. Otero and J. Nielsen. Industrial Systems Biology. *Biotechnology and Bioengineering*, 105(3):439–460, 2010.

[56] R. Overbeek, T. Begley, R.M. Butler, J.V. Choudhuri, H.Y. Chuang, M. Cohoon, V. de Crécy-Lagard, N. Diaz, T. Disz, R. Edwards, et al. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Research*, 33(17):5691–5702, 2005.

[57] J.H. Park, K.H. Lee, T.Y. Kim, and S.Y. Lee. Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation. *Proceedings of the National Academy of Sciences*, 104(19):7797–7802, 2007.

[58] K. Patil, I. Rocha, J. Förster, and J. Nielsen. Evolutionary programming as a platform for *in silico* metabolic engineering. *BMC Bioinformatics*, 6(1):308, 2005.

[59] G.C. Paul and C.R. Thomas. A structured model for hyphal differentiation and penicillin production using *Penicillium chrysogenum*. *Biotechnology and Bioengineering*, 51(5):558–572, 1996.

[60] D. Petranovic and J. Nielsen. Can yeast systems biology contribute to the understanding of human disease? *Trends in Biotechnology*, 26(11):584–590, 2008.

[61] D. Petrides, E. Sapidou, and J. Calandranis. Computer-aided process analysis and economic evaluation for biosynthetic human insulin production — A case study. *Biotechnology and Bioengineering*, 48(5):529–541, 1995.

[62] P. Pharkya and C.D. Maranas. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. *Metabolic Engineering*, 8(1):1–13, 2006.

[63] M. Riley, T. Abe, M.B. Arnaud, M.K.B. Berlyn, F.R. Blattner, R.R. Chaudhuri, J.D. Glasner, T. Horiuchi, I.M. Keseler, T. Kosuge, et al. Escherichia coli K-12: a cooperatively developed annotation snapshot — 2005. *Nucleic Acids Research*, 34(1):1–9, 2006.

[64] I. Rocha, P. Maia, P. Evangelista, P. Vilaça, S. Soares, J.P. Pinto, J. Nielsen, K.R. Patil, E.C. Ferreira, and M. Rocha. OptFlux: An open-source software platform for in silico metabolic engineering. *BMC Systems Biology*, 4(1):45, 2010.

[65] I. Rojas, M. Golebiewski, R. Kania, O. Krebs, S. Mir, A. Weidemann, and U. Wittig. SABIO-RK: a database for biochemical reactions and their kinetics. *BMC Systems Biology*, 1(Suppl 1):S6, 2007.

[66] RT Rowlands. Industrial strain improvement: mutagenesis and random screening procedures. *Enzyme and Microbial Technology*, 6(1):3–10, 1984.

[67] G.M. Rubin, M.D. Yandell, J.R. Wortman, G.L. Gabor, et al. Comparative genomics of the eukaryotes. *Science*, 287(5461):2204–2215, 2000.

[68] K.E. Rudd. EcoGene: a genome sequence database for *Escherichia coli* K-12. *Nucleic Acids Research*, 28(1):60–64, 2000.

[69] I. Schomburg, A. Chang, and D. Schomburg. BRENDA, enzyme data and metabolic information. *Nucleic Acids Research*, 30(1):47–49, 2002.

[70] S.C. Schuster. Next-generation sequencing transforms today's biology. *Nature Methods*, 5(1):16–18, 2007.

[71] Y. Sekiyama and J. Kikuchi. Towards dynamic metabolic network measurements by multi-dimensional NMR-based fluxomics. *Phytochemistry*, 68(16-18):2320–2329, 2007.

[72] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, 13(11):2498–2504, 2003.

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

[74] K. Smallbone, E. Simeonidis, N. Swainston, and P. Mendes. Towards a genome-scale kinetic model of cellular metabolism. *BMC Systems Biology*, 4(1):6, 2010.

[75] P. Smolen, P.E. Hardin, B.S. Lo, D.A. Baxter, and J.H. Byrne. Simulation of Drosophila circadian oscillations, mutations, and light responses by a model with VRI, PDP-1, and CLK. *Biophysical Journal*, 86(5):2786–2802, 2004.

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

[77] H. Takeyama, A. Kanamaru, Y. Yoshino, H. Kakuta, Y. Kawamura, and T. Matsunaga. Production of antioxidant vitamins, β-carotene, vitamin C, and vitamin E, by two-step culture of *Euglena gracilis Z*. *Biotechnology and Bioengineering*, 53(2):185–190, 1997.

[78] C.F. Taylor, N.W. Paton, K.S. Lilley, P.A. Binz, R.K. Julian, A.R. Jones, W. Zhu, R. Apweiler, R. Aebersold, E.W. Deutsch, et al. The minimum information about a proteomics experiment (MIAPE). *Nature Biotechnology*, 25(8):887–893, 2007.

[79] M. Tomita. Whole-cell simulation: a grand challenge of the 21st century. *Trends in Biotechnology*, 19(6):205–210, 2001.

[80] M. Tomita, K. Hashimoto, K. Takahashi, T.S. Shimizu, Y. Matsuzaki, F. Miyoshi, K. Saito, S. Tanida, K. Yugi, J.C. Venter, et al. E-CELL: software environment for whole-cell simulation. *Bioinformatics*, 15(1):72–84, 1999.

[81] E.J. Vandamme. Production of vitamins, coenzymes and related biochemicals by biotechnological processes. *Journal of Chemical Technology & Biotechnology*, 53(4):313–327, 1992.

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

[83] Z. Wang, M. Gerstein, and M. Snyder. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1):57–63, 2009.

[84] K. Yugi and M. Tomita. A general computational model of mitochondrial metabolism in a whole organelle scale. *Bioinformatics*, 20(11):1795–1796, 2004.

Chapter 2

Modeling Formalisms

This chapter is based on the review article *"Modeling formalisms in Systems Biology"* (submitted), which is a extended version of the article *"A critical review on modelling formalisms and simulation tools in Computational Biosystems"* published in *Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living*, volume 5518, pages 1063–1070, 2009.

Abstract

The field of Systems Biology has taken advantage of computational tools and high-throughput experimental data to model several biological processes including signaling, gene regulatory, and metabolic networks. However, most of these models are specific to each kind of network. The interconnection between all biological processes demands a whole-cell modeling framework for a complete understanding of cellular systems. Here, we describe the main types of cellular processes and the features required by an integrated framework for modeling, analyzing and simulating such processes. We then review several modeling formalisms that have been used in Systems Biology including Boolean networks, Bayesian networks, Petri nets, process algebras, constraint-based models, differential equations, rule-based models, interacting state machines, cellular automata, and agent-based models. We compare the features provided by different formalisms, and discuss recent approaches in the conversion and integration of these formalisms. Considering that no formalism fits all demands, it may become common to use different formalisms for different stages of the modeling process. Support for different extensions, hierarchical structure, multi-scale and robust model inference are key features for a framework that will support increasingly complex models.

2.1 Introduction

Living organisms are complex systems that emerge from the fundamental building blocks of life. Systems Biology (SB) is a field of science that studies these complex phenomena currently, mainly at the cellular level [102]. Understanding the mechanisms of the cell is essential for research in several areas such as drug development and biotechnological production. In the latter case, metabolic engineering approaches are applied in the creation of microbial strains with increased productivity of compounds with industrial interest such as biofuels and pharmaceutical products [170]. Using mathematical models of cellular metabolism, it is possible to systematically test and predict manipulations, such as gene knockouts, that generate (sub)optimal phenotypes for specific applications [18, 129]. These models are typically built in an iterative cycle of experiment and refinement, by multidisciplinary research teams that include biologists, engineers and computer scientists.

The interconnection between different cellular processes, such as metabolism and genetic regulation, reflects the importance of the holistic approach introduced by the SB paradigm in replacement of traditional reductionist methods. Although most cellular components have been studied individually, the behavior of the cell emerges at the network-level and requires an integrative analysis.

Recent high-throughput experimental methods allow to generate the so-called *omics* data (*e.g.*: genomics, transcriptomics, proteomics, metabolomics, fluxomics) that have allowed the reconstruction of many biological networks [57]. However, despite the great advances in the area, we are still far from a whole-cell computational model that integrates and simulates all the components of a living cell. Due to the enormous size and complexity of intracellular biological networks, computational cell models tend to be partial and focused on the application of interest. Also, due to the multidisciplinarity of the field, these models are based on several different kinds of formalisms, including those based on graphs (*e.g.* Boolean networks) and equation-based ones (*e.g.* ordinary differential equations). This diversity can lead to the fragmentation of modeling efforts as it hampers the integration of models from

different sources. Therefore, the whole-cell simulation goals of SB would benefit with the development of a framework for modeling, analysis and simulation that is based on a single formalism. This formalism should be able to integrate the entities and their relationships, spanning all kinds of biological networks.

This work reviews several modeling formalisms that have been used in SB, comparing their features and relevant applications. We opted to focus on the formalisms rather than the tools as they are the essence of the modeling approach. For the software tools implementing the formalisms, the interested reader may use the respective references. This review is divided into three parts. Section 2.2 describes the main types of biological networks that the models try to represent. Section 2.3 describes the relevant features for modeling these networks and section 2.4 explores the modeling formalisms found in the literature. Section 2.5 compares the formalisms and discusses their potential from an integrative perspective. Section 2.6 presents some conclusions and future directions regarding the most suitable formalisms for an integrated whole-cell framework.

2.2 Biological Networks

Cells are composed by thousands of components that interact in a myriad of ways. Despite this intricate interconnection, it is usual to divide and classify these networks according to their biological function (Fig. 2.1). The main types of networks are signaling, gene regulatory and metabolic (although some authors also classify protein-protein interactions as another type of network). These main types of networks will be briefly described.

2.2.1 Signaling networks

Signal transduction is a process for cellular communication where the cell receives (and responds to) external stimuli from other cells and from the environment. It affects most of the basic cell control mechanisms such as differentiation and apoptosis. The transduction process begins with the binding

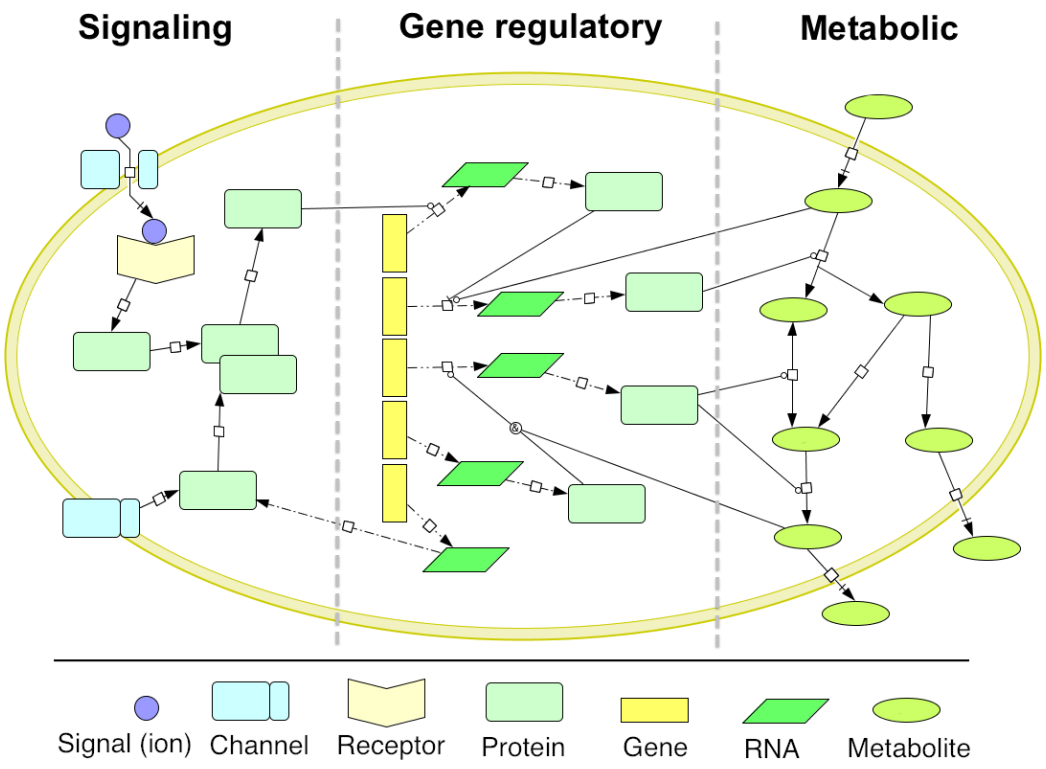

Figure 2.1: Conceptual representation of the main cellular processes. Signaling cascades receive external signals from the environment. Gene regulatory networks control the transcription level of genes. Metabolic networks obtain energy and carbon from external sources using internal conversion steps. (Figure created with the free software tool CellDesigner [64] that uses the graphical notations defined in [103].)

of an extracellular signaling molecule to a cell-surface receptor. The signal is then propagated and amplified inside the cell through signaling cascades that involve a series of trigger reactions such as protein phosphorylation. The output of these cascades is connected to gene regulation in order to control cell function. Signal transduction pathways are able to crosstalk, forming complex signaling networks [71, 3].

2.2.2 Gene regulatory networks

Gene regulation controls the expression of genes and, consequently, all cellular functions. Although all of the cell functionality is encoded in the genome by thousands of genes, it is essential for the survival of the cell that different functions are active in different stages of life and in the adaptation to different environments.

Gene expression is a process that involves transcription of the gene into mRNA, followed by translation to a protein, which may also be subject to post-translational modifications. The transcription process is controlled by transcription factors that can be activators or inhibitors. Transcription factors are themselves encoded by genes and subject to regulation, which altogether forms complex regulatory networks [156, 95].

2.2.3 Metabolic networks

Metabolism is a mechanism composed by a set of biochemical reactions, by which the cell sustains its growth and energy requirements. It includes several catabolic and anabolic pathways of enzyme–catalyzed reactions that import substrates from the environment and transform them into energy and building blocks required to build the cellular components. Metabolic pathways are interconnected through intermediate metabolites, forming complex networks. Gene regulation controls the production of enzymes and, consequently, directs the metabolic flux through the appropriate pathways in function of substrate availability and nutritional requirements [171, 127].

2.3 Modeling Requirements

Due to the different properties and behavior of the biological networks, they usually require different modeling features (although some desired features such as graphical visualization are common to all). For instance, features such as stochasticity and multi-state components may be important for signaling but not for metabolic networks. A summary of the major modeling features required by these networks is presented next.

2.3.1 Network visualization

SB is a multidisciplinary research field gathering biologists, computer scientists and engineers. Therefore, biological models should be expressed as intuitively as possible and easily interpreted by people from different areas. For that matter, graph and diagram based formalisms can be more appealing than mathematical or textual notations. Such formalisms can take advantage of state of the art network visualization tools, that when compared to traditional textbook diagrams, allow a much better understanding of the interconnections in large-scale networks, as well as the integration of heterogeneous data sources [131].

2.3.2 Topological analysis

A considerable amount of the work in this field is based on topological analysis of biological networks. In this case, graph-based representations also play a fundamental role. The analysis of the topological properties of these graphs, such as degree distribution, clustering coefficient, shortest paths or network motifs can reveal information from biological networks, including organization, robustness and redundancy [90, 10, 8].

2.3.3 Modularity and hierarchy

Despite its great complexity, the cell is organized as a set of connected modules with specific functions [79, 143]. Taking advantage of this modularity

can help to alleviate the complexity burden, facilitating its analysis. Compositionality is a related concept meaning that two modeling blocks can be aggregated together into one model without changes to any of the submodels. This property can be of special interest for applications in Synthetic Biology [7].

While modularity represents the horizontal organization of the cell, living systems also present vertical organization [30]. Molecules, cells, tissues, organs, organisms, populations and ecosystems reflect the hierarchical organization of life. A modeling formalism that supports hierarchical models and different levels of abstraction will cope with models that connect vertical organization layers. Also, it will have the required flexibility to cope with the different modeling approaches in SB, namely, top-down, bottom-up and middle-out [125].

2.3.4 Multi-state components

Some compounds may have multiple states, for example, a protein may be modified by phosphorylation. This is a very common case in signaling networks. The state of a protein can affect its functionality and consequently the reactions in which it participates. Therefore, different states are represented by different entities. However, a protein with n binding sites will have 2^n possible states, which results in a combinatorial explosion of entities and reactions [81, 15]. To avoid this problem, a suitable modeling formalism should consider entities with internal states and state-dependent reactions.

2.3.5 Spatial structure and compartmentalization

On its lowest level, the cell can be seen as a bag of mixed molecules. However, this bag is compartmentalized and requires transport processes for some species to travel between compartments. Furthermore, in some compartments, including the cytosol, the high viscosity, slow diffusion and amount of molecules may not be sufficient to guarantee a spatial homogeneity [172]. Spatial localization and concentration gradients are actually important mechanisms in biological processes such as morphogenesis [175].

2.3.6 Qualitative analysis

Experimental determination of kinetic parameters to build quantitative models is a cumbersome task. Furthermore, they are dependent on the experimental conditions, and there is generally no guarantee that the *in vitro* values will match the *in vivo* conditions [173]. Therefore, several models are only qualitative. Although these models do not allow for quantitative simulations, they allow us to ask qualitative questions about the system and to learn valuable knowledge. For instance, elementary mode analysis is used for calculating all possible pathways through a metabolic network [157].

2.3.7 Dynamic simulation

Dynamic simulation allows the prediction of the transient behavior of a system under different conditions. The simulation approach depends on the type of components included in the model, which depend not only on the nature of the involved interactions but also on the available information for their characterization. In regulatory networks, genes are activated and deactivated through the transcription machinery. Due to its complexity and the lack of kinetic information, the details of the machinery are usually not considered. Instead, genes are modeled by discrete (typically boolean) variables that change synchronously through discrete time steps. Synchronized simulation is the simplest simulation method and requires models with little detail.

In biological processes like signaling cascades, that are triggered by a low number of signaling molecules, it is important to take into consideration the inherent stochasticity in the diffusion of these molecules. Stochastic simulation is a common approach for simulation of signaling networks [36]. This approach requires the attribution of probability functions for each reaction in the model. Metabolic reactions, on the other hand, comprise large quantities of metabolites. Therefore, their behavior can be averaged and modeled by continuous variables governed by deterministic rate laws [27]. In both cases, experimental data is required to estimate the parameters in the models, which is a significant bottleneck in the modeling process.

2.3.8 Standardization

Biological models need to be represented in a common format for exchange between different tools. The Systems Biology Markup Language (SBML) has become the *de facto* standard of the SB community, and is currently supported by over two hundred tools [84]. It is an XML–based language for representation of species, compartments, reactions and their specific properties such as concentrations, volumes, stoichiometry and rate laws. It also facilitates the storage of tool specific data using appropriate tags. SBML was initially focused on biochemical reaction networks such as metabolic and signaling pathways, therefore it is not so well-suited for modeling other kinds of processes such as regulatory networks which are better described by logical models. Nevertheless, these and other limitations are being addressed in the development of future releases [59, 83].

CellML is another XML–based language with a similar purpose to SBML albeit more generic [117]. The Systems Biology Graphical Notation (SBGN) [111] is a standard that focuses on the graphical notation and may be seen as a complement to SBML. It addresses the visualization concerns discussed previously, specially the creation of graphical models with a common notation that can be shared and unambiguously interpreted by people from different areas.

2.4 Modeling Formalisms

Many formalisms have been used to approach the modeling of biological systems, in part due to the diversity of phenomena that occur in living systems, and also due to the multidisciplinarity of the research teams. Biologists may be more familiar with mathematical modeling and computer scientists may be religious to their computational formalism of choice. The dichotomy between mathematical and computational models has been discussed elsewhere [85]. Although they follow different approaches (denotational *vs* operational), it has been questioned if there is such a clear separation between mathematical and computational models. Therefore, in the following we will briefly

Table 2.1: Overview of some of the literature references on the applications of each formalism, classified by the type of process. (BN) Boolean networks; (Bay) Bayesian networks; (PN) Petri nets; (PA) Process algebras; (CB) Constraint-based models; (DE) Differential equations; (RB) Rule-based models; (ISM) Interacting state machines; (CA) Cellular automata; (AB) Agent-based models.

	Signaling	Gene regulatory	Metabolic
BN	[76, 153]	[2, 47, 97, 114, 164]	
Bay	[150, 151]	[50, 86, 100, 134, 195]	
PN	[17, 28, 77, 152]	[25, 26]	[106, 110, 144, 165, 192]
PA	[141, 146, 147, 148]		
CB	[112, 128]	[67, 112]	[56, 155, 157, 160, 162]
DE	[177]	[14, 29, 45, 177]	[27, 88, 149]
RB	[11, 12, 13, 16]		
ISM	[52, 61, 62, 94]		
CA	[98, 189]	[185]	[183, 185]
AB	[6, 72, 137, 138]		[105]

describe several formalisms regardless of such distinction.

There are other reviews on modeling formalisms in the literature [60, 120], some focusing on particular processes, such as gene regulatory networks [156, 95] or signaling pathways [4]. However, to the best of our knowledge, none covers the whole spectrum presented in this work. Note that besides the intracellular level, several studies in SB also address the cellular population level. Therefore, formalisms for modeling the dynamics of cellular populations have also received attention in the field and will be considered in this work. Table 2.1 summarizes some of the literature references reviewed herein, classified by type of intracellular process implemented. Toy examples of the formalisms with graphical notation are depicted in Figure 2.2.

2.4.1 Boolean networks

Boolean networks (Fig. 2.2a) were introduced by Kauffman in 1969 to model gene regulatory networks [97]. They consist on networks of genes, modeled by boolean variables that represent active and inactive states. At each time

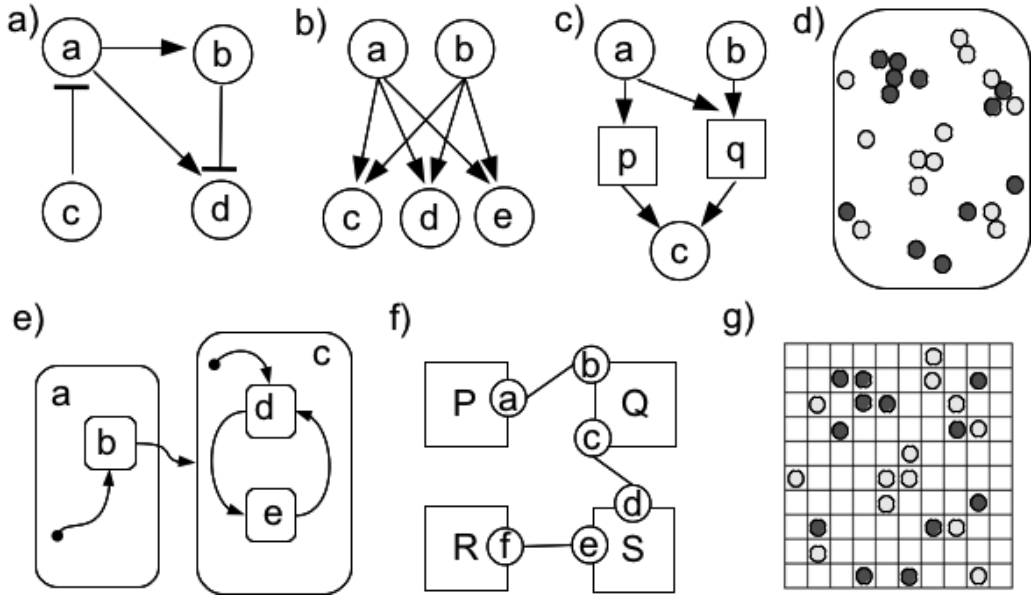

Figure 2.2: Toy examples of the formalisms with visual representation. a) Boolean network: node (gene) a is deactivated by node c, node b is activated by node a, node d is activated by node a and deactivated by node b; b) Bayesian network: the value of each node (gene) c, d, and e, is given by a probability function that is conditionally dependent on the values of nodes a and b; c) Petri net: transition (reaction) p consumes one token (molecule) from place (substance) a and produces one token in place c, transition q consumes tokens from places a and b and produces one token in place c; d) Agent-based model: two types of agents (light gray and dark gray), representing two different kinds of cell (or two kinds of molecules) move freely and interact within the containing space; e) Interacting state machine: a system that can move from state a to state c, where state a as an internal sub-state b and state c has two internal sub-states, d and e; f) Contact map of a rule-based model: agents (proteins) P, Q, R and S, contain several binding sites (a to f), edges between binding sites describe possible interactions (*e.g.* phosphorylation); g) Cellular automata: a 10×10 grid with three possible values (empty, light gray or dark gray), representing two different kinds of cells (or two kinds of molecules) that can change by interacting with their surrounding neighbors.

step, the state of each gene is determined by a logic rule which is a function of the state of its regulators. The state of all genes forms a global state that changes synchronously. For large network sizes (n nodes) it becomes impractical to explore all possible states (2^n). This type of model can be used to find steady-states (called attractors), and to analyze network robustness [114]. Boolean networks can be inferred directly from experimental gene expression time-series data [1, 47]. They have also been applied in some studies to model signaling pathways [76, 153]. To cope with the inherent noise and the uncertainty in biological processes, stochastic extensions like Boolean networks with noise [2] and Probabilistic Boolean networks [164] were introduced.

2.4.2 Bayesian networks

Bayesian networks (Fig. 2.2b) were introduced in the 80's by the work of Pearl [132]. They are a special type of probabilistic graphs. Their nodes represent random variables (discrete or continuous) and the edges represent conditional dependencies, forming a directed acyclic graph. Each node contains a probabilistic function that is dependent on the values of its input nodes. There are learning methods to infer both structure and probability parameters with support for incomplete data. This flexibility makes Bayesian networks specially interesting for biological applications. They have been used for inferring and representing gene regulatory [63, 134, 75, 9] and signaling networks [150, 151]. One disadvantage of Bayesian networks is the inability to model feedback loops, which is a common motif in biological networks. This limitation can be overcome by dynamic Bayesian networks [86, 100, 195, 50]. In this case, the variables are replicated for each time step and the feedback is modeled by connecting the nodes at adjacent time steps.

2.4.3 Petri nets

Petri nets (Fig. 2.2c) were created in the 60's by Carl Adam Petri for the modeling and analysis of concurrent systems [136]. They are bipartite graphs with two types of nodes, *places* and *transitions*, connected by directed arcs.

Places hold *tokens* that can be produced (respectively, consumed) when an input (respectively, output) transition *fires*. The execution of a Petri net is non-deterministic and specially suited for distributed systems with concurrent events. Their application to biological processes began in 1983, by the work of Reddy and coworkers, to overcome the limitations in quantitative analysis of metabolic pathways [144].

There are currently several Petri net extensions (*e.g.:* coloured, timed, stochastic, continuous, hybrid, hierarchical, functional), forming a very versatile framework for both qualitative and quantitative analysis. Due to this versatility, they have been used in metabolic [110, 192, 106], gene regulatory [25, 26], and signaling networks [152, 28, 17, 77]. Also, they are suited for integrating different types of networks, such as gene regulatory and metabolic [165].

2.4.4 Process algebras

Process algebras are a family of formal languages for modeling concurrent systems. They generally consist on a set of process primitives, operators for sequential and parallel composition of processes, and communication channels. The Calculus of Communicating Systems (CCS) was one of the first process algebras, developed during the 70's by Robin Milner [122], and later gave origin to the more popular π-calculus [123]. In SB the application of process algebras has been mainly focused on signaling pathways due to their similarity to communication processes. About a decade ago, Regev and coworkers published their pioneer work on the representation of signaling pathways with π-calculus [147, 148]. They later extended their work using stochastic π-calculus (BioSpi) to support quantitative simulations [141] and using Ambient calculus (BioAmbients) for representation of compartments [146]. Other relevant biological applications of process algebras include Bio-calculus [124], κ-calculus (for protein-protein interactions) [44], CCS-R [43], Beta binders [140], Brane Calculi [22], SpacePi [91], Bio-PEPA [31, 32] and BlenX [46, 139].

2.4.5 Constraint-based models

Constraint-based models for cellular metabolism began spreading during the 90's, mainly influenced by the work of Palsson and coworkers [178]. Assuming that cells rapidly reach a steady-state, these models overcome the limitations in lack of experimental data for parameter estimation inherent in fully detailed dynamic models. They are based on stoichiometric, thermodynamic and enzyme capacity constraints [145, 142]. Instead of a single solution, they define a space of possible solutions representing different phenotypes that comply with the constraints. The simplicity in this formulation allows its application to genome-scale metabolic models comprising thousands of reactions, such as the most recent metabolic reconstruction of *E. coli* [56].

Constraint-based models have been used in metabolic engineering strategies for the determination of flux distributions (metabolic flux analysis [184], flux balance analysis [96]), knockout phenotype predictions (minimization of metabolic adjustment [160], regulatory on/off minimization [162]) or enumerating all possible pathways (extreme pathways [155], elementary flux modes [157]). Although their main application has been on metabolic networks, there are recent efforts towards application on gene regulatory and signaling networks [128, 67, 112].

2.4.6 Differential equations

Differential equations describe the rate of change of continuous variables. They are typically used for modeling dynamical systems in several areas. Systems of non-linear ordinary differential equations (ODEs) have been used in SB to describe the variation of the amount of species in the modeled system as a function of time. They have been applied to all kinds of biological pathways [27, 149, 29, 177]. With a fully detailed kinetic model, one can perform time-course simulations, predict the response to different inputs and design system controllers. However, building ODE models requires insight into the reaction mechanisms to select the appropriate rate laws, and experimental data to estimate the kinetic parameters. The lack of kinetic data has limited the size of the modeled networks to pathway size, with exception for

the human red blood cell model [88].

Approximative rate laws such as generalized mass action (GMA) [82], S-systems [154], lin-log [179], and convenience kinetics [116], have compact standard formulations that can facilitate the development and analysis of large-scale models [80, 37]. This opens the possibility for kinetic modeling at the genome-scale [167].

Other types of differential equations, such as stochastic differential equations (SDEs) and partial differential equations (PDEs) can be used respectively to account for stochastic effects and spatial distribution [176]. Piecewise-linear differential equations (PLDEs) have been used to integrate discrete and continuous features in gene regulatory networks [45, 14].

2.4.7 Rule-based models

Rule-based (Fig. 2.2f) modeling comprises a recent approach to the problem of multi-state components in biological models. In rule-based formalisms the species are defined in a structured manner and support multiple states. The reaction rules are defined as transformations of classes of species, avoiding the need for specifying one reaction per each possible state of a species. This high-level specification is then automatically transformed into a biochemical network with the set of species and reactions generated by the specification. This kind of formalism is implemented in BioNetGen [15] which generates an ODE model or a stochastic simulation from the ruled-based specification. It has been applied in the modeling of different signaling pathways [16, 11, 12, 13]. A similar rule-based formalism used for this kind of pathways is the κ language, where the species are defined by agents that have a structured interface for interaction with other agents [41, 42, 58]. The possible interactions are defined by a set of rules, which can be visualized by a contact map. BIOCHAM implements a rule-based approach for model specification which is complemented with a temporal logic language for the verification of the properties the biological models [20].

The main advantage of the rule-based approach is that it can avoid the combinatorial explosion problem in the generation and simulation of the com-

plete reaction network by performing stochastic simulations that only instantiate the species and reactions as they become available [33, 34] or by the generation of coarse-grained ODE systems [58]. Spatial simulation has been addressed recently by the inclusion of geometric information as part of the structure of the species [74].

2.4.8 Interacting state machines

Interacting state (Fig. 2.2e) machines are diagram-based formalisms that describe the temporal behavior of a system based on the changes in the states of its parts. They are suited to model biological behavior in a qualitative way as they require little quantitative data. They differ from other approaches as they define a system in terms of its states rather than its components. They are typically used for model checking and interactive execution.

One such formalism is Statecharts, developed by David Harel during the 80's [78] that was first applied in biology for modeling the T-cell activation process [94, 52] and more recently in pancreatic organogenesis [161]. In this formalism, the state of a system may contain sub-states at multiple levels, allowing an hierarchical view of the system and the relation between events at smaller and larger scales. Other related formalisms are Reactive Modules [5] and Live Sequence Charts [40], which, along with the former, have been applied in the modelling of *C. elegans* vulval development [62, 61].

2.4.9 Cellular automata

Cellular automata (Fig. 2.2g) were created by von Neumann and Ulam in the 40's [181]. They are discrete dynamic models that consist on a grid of cells with a finite number of states. A cellular automaton has an initial configuration that changes at each time step through a predefined rule that calculates the state of each cell as a function of the state of its neighbors at the previous step. They are specially suited for modeling complex phenomena in a scale-free manner and have been used in biological studies for a long time [55]. Due to their spatial features their main applications are related to molecular dynamics and cellular population dynamics.

Application examples at the molecular level include enzyme reaction networks that account for spatial diffusion [183] and signaling pathways [189, 98]. At the cellular level they were used for models such as those of bacterial aggregation [168] and HIV infection [194, 35]. Dynamic cellular automata are a variation of cellular automata that allows for movement of the cell contents inside the grid, mimicking brownian motion. They were used to model enzyme kinetics, molecular diffusion and genetic circuits [185].

2.4.10 Agent-based models

Agent-based models (Fig. 2.2d) describe the interactions among multiple autonomous agents. They are similar in concept to cellular automata, except in this case, instead of using a grid and synchronized time steps, the agents move freely within the containing space. Likewise, they are used to study complex phenomena and emergent dynamics using populations of agents with simple rules. At the molecular level they have been mainly used to build models of signaling pathways that account for spatial distribution and the structural properties of the cell [72, 138, 137, 6]. Recently, they have also been applied to metabolic reactions [105]. However, their main application is at the multi-cellular level, where they have been used to study granuloma formation [159], tumor growth [193, 54], morphogenesis [73], chemotaxis [53], immune responses [118, 115], and several others [174, 121].

2.4.11 Other

There are other modeling formalisms that have been used in SB which are worth mentioning. Cybernetic modeling is one of the earliest approaches for dynamic modeling that was used in bioprocess applications [109, 48]. A recent approach combines cybernetic variables with elementary flux modes [191, 99]. Hybrid automata addressed the integration of discrete and continuous components in the Delta-Notch signaling pathway [65, 66]. Artificial neural networks were used to model gene expression [180]. Molecular interaction maps are a popular graph-based formalism created by Kohn in 1999 [107, 108, 119] that influenced the SBGN standard [111]. Other graph-based

formalisms include modular interaction networks [190] and logical interaction hypergraphs [104]. The P systems formalism created by Paun in 1998, inspired the area of membrane computing [130] and has been recently applied in SB [135, 21]. Chemical organization theory is a recent approach for modeling biochemical reaction networks that uses set theory to analyze how they can be decomposed into self-maintaining subnetworks called organizations, that reveal dynamic properties of the system [49]. It has been used to analyze different types of networks including signaling pathways and regulated metabolic networks [23, 24, 92, 93].

2.5 Discussion

The diversity of problems studied in SB gave rise to the application of several different types of formalisms. Some of the literature references for each formalism, classified by the type of biological process described, are shown in Table 2.1. We can observe that only four formalisms (Petri nets, constraint-based models, differential equations and cellular automata) have been applied to all three types of biological networks, which makes them potential candidates as a suitable integrative formalism for whole-cell modeling. Nevertheless, this does not automatically exclude other formalisms from this possibility as well. Another interesting observation is that metabolism is the biological process with the smaller number of formalisms applied. This is likely due to the fact that its two main frameworks (differential equations and constraint-based) are well suited for modeling metabolic networks. On the other hand, all of the formalisms have been applied to signaling pathways. One possible reason for this, is the fact that they have a larger number of modeling feature requirements, including spatial localization and multi-state components.

The modeling features provided by the formalisms reviewed in this work are compared in Table 2.2. Some of the features are only available in extensions of the formalisms. We can observe that no single formalism covers the whole spectrum of features desired for modeling all kinds of biological components. Petri nets and rule-based models are among the formalisms

Table 2.2: Modeling formalisms and implemented features. (BN) Boolean networks; (Bay) Bayesian networks; (PN) Petri nets; (PA) Process algebras; (CB) Constraint-based models; (DE) Differential equations; (RB) Rule-based models; (ISM) Interacting state machines; (CA) Cellular automata; (AB) Agent-based models; (+) Supported feature; (e) Available through extension.

	BN	Bay	PN	PA	CB	DE	RB	ISM	CA	AB
Visualization	+	+	+				+	+	+	+
Topology	+	+	+		+					
Modularity			+	+			+	+		
Hierarchy			e	e				+		
Multi-state			e				+	+	+	+
Compartments				e			+	+		+
Spatial						e	e		+	+
Qualitative	+	+	+	+	+		+			
Synchronized	+		e						+	
Stochastic	e	+	e	+		e	+	+	+	+
Continuous			e			+	+			

that cover most features. Petri nets have several extensions available, and although none of the extensions alone fulfills all requisites, altogether they form a very versatile modeling framework. Rule-based models present a high level of abstraction and can be used for stochastic simulation and automatic generation of lower level ODE-based representations. Therefore, they take advantage of the analytic power of abstract representations, preserving the ability to generate stochastic and deterministic simulations.

Although none of the formalisms implements all the required features, this is not necessarily a limitation, since different formalisms can be used at different stages of the modeling process. The model construction process begins with biochemical knowledge and experimental data that allow an enumeration of the components and connections in the system. Graph-based models, such as Boolean networks, Bayesian networks and Petri nets can be used for modeling this map of interactions. This allows a deeper understanding of the organization of the system through topological analysis, and drives new experiments by finding gaps in the models. This kind of models also allows qualitative descriptions of system behavior and coarse simulation

capabilities. If the reactions' stoichiometry and directionality are known, one may analyze the steady-states of the system using constraint-based models. Finally, if extensive experimental data is available to infer the kinetics of the reactions, probabilistic or deterministic rate laws can be used to create dynamic models. These are used to generate time-course simulations under different sets of initial conditions. Stochastic process algebras, stochastic Petri nets, continuous Petri nets, rule-based models and differential equations, would all be ideal candidates for this purpose.

Cellular automata and agent-based models account for the individual replicas of each component in the system. When applied at the molecular level, this paradigm provides accurate simulations of small sets of biochemical reactions that account for spatial diffusion. However, it becomes infeasible to perform simulations at the genome-scale network level, as this would imply modeling every copy of all substances present in the cell. Nevertheless, this approach is very convenient for modeling at the cell population level, as it allows to track changes in individual cells and to study the emergent properties of cellular communities.

The inability of the formalisms to fit all purposes has driven the development of methodologies to convert between different formalisms. Two different methods have been proposed to convert Boolean networks to Petri nets [25, 169]. Boolean networks have also been converted to constraint-based models [68] and to ODEs [186]. Other formalisms have also been converted to ODEs, including constraint-based models [166], Petri nets [69], process algebras [19] and rule-based models [58]. When the mappings are made from abstract to more detailed models they usually require some assumptions and insight into the reaction mechanisms. The language for biochemical systems (LBS) is a recent language that integrates a rule-based approach with process calculus, and supports the generation of Petri nets, ODEs and continuous time Markov chains [133].

Along with the conversion between formalisms, there is also a recent trend for developing methods that support integrated simulation of different formalisms in order to integrate different kinds of biological networks, where each network is modeled in its own formalism. Extensions of flux balance

analysis (FBA) [96], such as regulated FBA (rFBA) [38] and steady-state regulated FBA (SR-FBA) [163] incorporate boolean rules into constraint-based models for integrated simulation of regulatory and metabolic networks. Integrated FBA (iFBA) extends rFBA by integrating kinetic information from ODE models [39]. Integrated dynamic FBA (idFBA) aims to integrate signaling, regulatory and metabolic networks by modeling all networks in the constraint-based formulation [113]. Biochemical systems theory (BST) has been recently integrated with Hybrid Functional Petri Nets (HFPN) in order to integrate metabolic, regulatory and signaling networks, in a framework that accounts for different time-scales as well as discrete, stochastic and continuous effects [187, 188].

In search for a proper formalism perhaps the most important aspect to consider is the balance between simplicity and expressiveness. There is a price to pay for the amount of features provided by a formalism, which may come at the cost of increased model complexity. The complexity of the representation and the number of parameters determines the amount of experimental data required for model construction. This is the reason why the most simple formalisms such as Boolean networks and constraint-based models have been used to build, respectively, gene regulatory and metabolic networks at the genome scale. This concern is most critical when not only the parameters but also the network structure are unknown. Model inference (also known as *reverse engineering*) methods are applied in these cases. They have been used to infer Boolean networks [1, 47], Bayesian networks [63, 9], Petri nets [126, 51] and ODEs [101, 87] from experimental data. However, the scalability of these methods is greatly dependent on the simplicity of the underlying formalism.

2.6 Conclusions

With the myriad of formalisms that have been applied in SB, we face the challenge of choosing the proper formalism for the problem in hands. As more data become available for network reconstruction, we move towards integration of all kinds of biological networks, namely signaling, gene regulatory

and metabolic. Although some formalisms like Petri nets, constraint-based models and differential equations have been applied for all these networks, no single formalism covers the whole spectrum of functionalities available in the different formalisms reviewed in this work. Petri nets have several extensions available, covering most of the features analyzed (Table 2.2). However, they lack support for compartments and spatial localization. Rule-based models are another strong candidate as they also cover a great part of the modeling features. In general, formalisms with a visual format can be more appealing and reveal insight into the system functioning before simulation is possible. This is particularly important given the lack of kinetic data for larger models.

A common problem in the analysis of biological networks is the combinatorial explosion that originates from the complexity of large models. A typical example is the computation of elementary flux modes, which is currently still infeasible at the genome-scale, requiring modular decomposition of the networks [158]. This problem will aggravate as we get closer to whole-cell modeling. The solution may reside in the application of hierarchical formalisms to represent an intermediate level between the reaction and the cell. As stated elsewhere, one should not "model bulldozers with quarks" [70]. Hierarchical Petri nets, BioAmbients and Statecharts are formalisms that support hierarchical modeling.

Models of cell populations are also becoming more frequent. They are used to study scenarios like cell differentiation, chemotaxis, infections or tumor growth. This kind of models depends on the internal dynamics of the cells as well as population dynamics. Therefore, they require modeling of interactions across organizational scales [182]. It is possible that these multi-scale models will require the integration of different formalisms. For instance, the evolution of a population of cells could be modeled by an agent-based model, with each agent having a boolean network for internal representation of its gene expression.

In order to convert between different formalisms it is important to have a standard representation format that preserves most of the features in the models. SBML is the most popular standard in the SB community, currently supported by over two hundred tools [84]. Most of the modeling features cov-

ered herein have been proposed for future versions of SBML [59]. These include hierarchical model composition, rule-based modeling, spatial geometry and alternative mathematical representations.

Given the size and complexity of the biological networks operating inside the cell, the model building process is based on iterative steps of refinement and validation. Recent approaches for genome-scale kinetic modeling of metabolism, begin with the network topology, modeled in the constraint-based framework, and then refine the models by adding the kinetic structure in order to generate ODE models [89, 167]. Petri nets seem to be a promising formalism for this purpose, given that discrete Petri nets can model the network topology, and can then be used as a scaffold for the generation of dynamic models based on continuous or stochastic Petri nets. The fact that the same kind of formalism is used during the whole model refinement process, helps the creation of more straightforward and formal methods for automatic mapping and validation of the models.

Many of the proposed formalisms, such as Petri nets or process algebras, were originally created by the computational community for the specification of software systems, where the final system has to comply to the model. The biological community faces the opposite problem, where the model has to mimic the system's behavior, and where most components cannot even be measured directly. Therefore, a proper framework for SB must provide not only a suitable formalism with attractive features and simulation methods, but also methods for model inference and parameter estimation that are sufficiently robust to handle experimental data that are incomplete and prone to measurement error.

Bibliography

[1] T. Akutsu, S. Miyano, and S. Kuhara. Identification of genetic networks from a small number of gene expression patterns under the Boolean network model. In *Pacific Symposium on Biocomputing*, volume 4, pages 17–28, 1999.

[2] T. Akutsu, S. Miyano, and S. Kuhara. Inferring qualitative relations in genetic networks and metabolic pathways. *Bioinformatics*, 16(8):727–734, 2000.

[3] R. Albert and R.S. Wang. Discrete dynamic modeling of cellular signaling networks. *Methods in Enzymology*, pages 281–306, 2009.

[4] B.B. Aldridge, J.M. Burke, D.A. Lauffenburger, and P.K. Sorger. Physicochemical modelling of cell signalling pathways. *Nature Cell Biology*, 8(11):1195–1203, 2006.

[5] R. Alur and T.A. Henzinger. Reactive modules. *Formal Methods in System Design*, 15(1):7–48, 1999.

[6] G. An. A model of TLR4 signaling and tolerance using a qualitative, particle-event-based method: introduction of spatially configured stochastic reaction chambers (SCSRC). *Mathematical Biosciences*, 217(1):43–52, 2009.

[7] E. Andrianantoandro, S. Basu, D.K. Karig, and R. Weiss. Synthetic biology: new engineering rules for an emerging discipline. *Molecular Systems Biology*, 2(1), 2006.

[8] Y. Assenov, F. Ramirez, S.E. Schelhorn, T. Lengauer, and M. Albrecht. Computing topological parameters of biological networks. *Bioinformatics*, 24(2):282–284, 2008.

[9] C. Auliac, V. Frouin, X. Gidrol, F. D'Alché-Buc, et al. Evolutionary approaches for the reverse-engineering of gene regulatory networks: A study on a biologically realistic dataset. *BMC Bioinformatics*, 9(1):91, 2008.

[10] A.L. Barabási and Z.N. Oltvai. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2):101–113, 2004.

[11] D. Barua, J.R. Faeder, and J.M. Haugh. Structure-based kinetic models of modular signaling protein function: focus on Shp2. *Biophysical Journal*, 92(7):2290–2300, 2007.

[12] D. Barua, J.R. Faeder, and J.M. Haugh. Computational models of tandem Src homology 2 domain interactions and application to phosphoinositide 3-kinase. *Journal of Biological Chemistry*, 283(12):7338–7345, 2008.

[13] D. Barua, J.R. Faeder, and J.M. Haugh. A Bipolar Clamp Mechanism for Activation of Jak-Family Protein Tyrosine Kinases. *PLoS Computational Biology*, 5(4):e1000364, 2009.

[14] 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–i28, 2005.

[15] M.L. Blinov, J.R. Faeder, B. Goldstein, and W.S. Hlavacek. BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics*, 20(17):3289–3291, 2004.

[16] M.L. Blinov, J.R. Faeder, B. Goldstein, and W.S. Hlavacek. A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity. *BioSystems*, 83(2-3):136–151, 2006.

[17] R. Breitling, D. Gilbert, M. Heiner, and R. Orton. A structured approach for the engineering of biochemical network models, illustrated for signalling pathways. *Briefings in Bioinformatics*, 9(5):404–421, 2008.

[18] A.P. Burgard, P. Pharkya, and C.D. Maranas. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and Bioengineering*, 84(6):647–657, 2003.

[19] M. Calder, S. Gilmore, and J. Hillston. Automatically deriving ODEs from process algebra models of signalling pathways. In *Computational Methods in Systems Biology*, pages 204–215, 2005.

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

[21] H. Cao, F.J. Romero-Campero, S. Heeb, M. Cámara, and N. Krasnogor. Evolving cell models for systems and synthetic biology. *Systems and Synthetic Biology*, 4(1):55–84, 2010.

[22] L. Cardelli. Brane calculi. In *Computational Methods in Systems Biology*, pages 257–278. Springer, 2005.

[23] F. Centler, P.S. Fenizio, N. Matsumaru, and P. Dittrich. Chemical organizations in the central sugar metabolism of *Escherichia coli*. *Mathematical Modeling of Biological Systems*, 1(2):105–119, 2007.

[24] F. Centler, C. Kaleta, P.S. Di Fenizio, and P. Dittrich. Computing chemical organizations in biological networks. *Bioinformatics*, 24(14):1611–1618, 2008.

[25] C. Chaouiya, E. Remy, P. Ruet, and D. Thieffry. Qualitative modelling of genetic networks: From logical regulatory graphs to standard petri nets. *Applications and Theory of Petri Nets*, 3099:137–156, 2004.

[26] C. Chaouiya, E. Remy, and D. Thieffry. Petri net modelling of biological regulatory networks. *Journal of Discrete Algorithms*, 6(2):165–177, 2008.

[27] C. Chassagnole, N. Noisommit-Rizzi, J.W. Schmid, K. Mauch, and M. Reuss. Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnology and Bioengineering*, 79(1):53–73, 2002.

[28] L. Chen, G. Qi-Wei, M. Nakata, H. Matsuno, and S. Miyano. Modelling and simulation of signal transductions in an apoptosis pathway by using timed Petri nets. *Journal of Biosciences*, 32(1):113–127, 2007.

[29] T. Chen, H.L. He, and G.M. Church. Modeling Gene Expression with Differential Equations. In *Pacific Symposium on Biocomputing*, pages 29–40, 1999.

[30] C.Y. Cheng and Y.J. Hu. Extracting the abstraction pyramid from complex networks. *BMC Bioinformatics*, 11(1):411, 2010.

[31] F. Ciocchetta and J. Hillston. Bio-PEPA: an extension of the process algebra PEPA for biochemical networks. *Electronic Notes in Theoretical Computer Science*, 194(3):103–117, 2008.

[32] F. Ciocchetta and J. Hillston. Bio-PEPA: a framework for the modelling and analysis of biological systems. *Theoretical Computer Science*, 410(33–34):3065–3084, 2009.

[33] J. Colvin, M.I. Monine, J.R. Faeder, W.S. Hlavacek, D.D. Von Hoff, and R.G. Posner. Simulation of large-scale rule-based models. *Bioinformatics*, 25(7):910–917, 2009.

[34] J. Colvin, M.I. Monine, R.N. Gutenkunst, W.S. Hlavacek, D.D. Von Hoff, and R.G. Posner. RuleMonkey: software for stochastic simulation of rule-based models. *BMC Bioinformatics*, 11(1):404, 2010.

[35] D.W. Corne and P. Frisco. Dynamics of HIV infection studied with cellular automata and conformon-P systems. *BioSystems*, 91(3):531–544, 2008.

[36] M.N. Costa, K. Radhakrishnan, B.S. Wilson, D.G. Vlachos, J.S. Edwards, and H. Jonsson. Coupled stochastic spatial and non-spatial simulations of ErbB1 signaling pathways demonstrate the importance of spatial organization in signal transduction. *PLoS ONE*, 4:e6316, 2009.

[37] R.S. Costa, D. Machado, and I. Rocha. Hybrid dynamic modeling of *Escherichia coli* central metabolic network combining Michaelis-Menten and approximate kinetic equations. *BioSystems*, 100(2):150–157, 2010.

[38] M.W. Covert and B.Ø. Palsson. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *Journal of Biological Chemistry*, 277(31):28058–28064, 2002.

[39] 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.

[40] W. Damm and D. Harel. LSCs: Breathing life into message sequence charts. *Formal Methods in System Design*, 19(1):45–80, 2001.

[41] V. Danos, J. Feret, W. Fontana, and R. Harmer. Rule-Based Modelling of Cellular Signalling. In *CONCUR 2007 – Concurrency Theory*, volume 4703, pages 17–41. Springer, 2007.

[42] V. Danos, J. Feret, W. Fontana, R. Harmer, and J. Krivine. Rule-based modelling and model perturbation. *Transactions on Computational Systems Biology XI*, 5750:116–137, 2009.

[43] V. Danos and J. Krivine. Formal molecular biology done in CCS-R. *Electronic Notes in Theoretical Computer Science*, 180(3):31–49, 2007.

[44] V. Danos and C. Laneve. Formal molecular biology. *Theoretical Computer Science*, 325(1):69–110, 2004.

[45] 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. *Bulletin of Mathematical Biology*, 66(2):301–340, 2004.

[46] L. Dematte, C. Priami, A. Romanel, and O. Soyer. Evolving BlenX programs to simulate the evolution of biological networks. *Theoretical Computer Science*, 408(1):83–96, 2008.

[47] P. D'haeseleer, S. Liang, and R. Somogyi. Genetic network inference: from co-expression clustering to reverse engineering. *Bioinformatics*, 16(8):707–726, 2000.

[48] P. Dhurjati, D. Ramkrishna, MC Flickinger, and GT Tsao. A cybernetic view of microbial growth: modeling of cells as optimal strategists. *Biotechnology and Bioengineering*, 27(1):1–9, 1985.

[49] P. Dittrich and P.S. Di Fenizio. Chemical organisation theory. *Bulletin of Mathematical Biology*, 69(4):1199–1231, 2007.

[50] N. Dojer, A. Gambin, A. Mizera, B. Wilczyński, and J. Tiuryn. Applying dynamic Bayesian networks to perturbed gene expression data. *BMC Bioinformatics*, 7(1):249, 2006.

[51] M. Durzinsky, A. Wagler, and W. Marwan. Reconstruction of extended Petri nets from time series data and its application to signal transduction and to gene regulatory networks. *BMC Systems Biology*, 5(1):113, 2011.

[52] S. Efroni, D. Harel, and I.R. Cohen. Toward rigorous comprehension of biological complexity: modeling, execution, and visualization of thymic T-cell maturation. *Genome Research*, 13(11):2485–2497, 2003.

[53] T. Emonet, C.M. Macal, M.J. North, C.E. Wickersham, and P. Cluzel. AgentCell: a digital single-cell assay for bacterial chemotaxis. *Bioinformatics*, 21(11):2714, 2005.

[54] J.A. Engelberg, G.E.P. Ropella, and C.A. Hunt. Essential operating principles for tumor spheroid growth. *BMC Systems Biology*, 2(1):110, 2008.

[55] G.B. Ermentrout and L. Edelstein-Keshet. Cellular Automata Approaches to Biological Modeling. *Journal of Theoretical Biology*, 160(1):97–133, 1993.

[56] A.M. Feist, C.S. Henry, J.L. Reed, M. Krummenacker, A.R. Joyce, P.D. Karp, L.J. Broadbelt, V. Hatzimanikatis, and B.Ø. Palsson. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*, 3(121), 2007.

[57] A.M. Feist, M.J. Herrgård, I. Thiele, J.L. Reed, and B.Ø. Palsson. Reconstruction of biochemical networks in microorganisms. *Nature Reviews Microbiology*, 7(2):129–143, 2008.

[58] J. Feret, V. Danos, J. Krivine, R. Harmer, and W. Fontana. Internal coarse-graining of molecular systems. *Proceedings of the National Academy of Sciences of the United States of America*, 106(16):6453, 2009.

[59] A. Finney and M. Hucka. Systems biology markup language: Level 2 and beyond. *Biochemical Society Transactions*, 31:1472–1473, 2003.

[60] J. Fisher and T.A. Henzinger. Executable cell biology. *Nature Biotechnology*, 25(11):1239–1249, 2007.

[61] J. Fisher, N. Piterman, A. Hajnal, and T.A. Henzinger. Predictive Modeling of Signaling Crosstalk during *C. elegans* Vulval Development. *PLoS Computational Biology*, 3(5):e92, 2007.

[62] J. Fisher, N. Piterman, E. Hubbard, M.J. Stern, and D. Harel. Computational insights into *Caenorhabditis elegans* vulval development. *Proceedings of the National Academy of Sciences of the United States of America*, 102(6):1951–1956, 2005.

[63] N. Friedman. Inferring cellular networks using probabilistic graphical models. *Science*, 303(5659):799–805, 2004.

[64] A. Funahashi, M. Morohashi, H. Kitano, and N. Tanimura. CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. *BIOSILICO*, 1(5):159–162, 2003.

[65] R. Ghosh and C. Tomlin. Lateral inhibition through delta-notch signaling: A piecewise affine hybrid model. *Hybrid Systems: Computation and Control*, 2034:232–246, 2001.

[66] R. Ghosh and C. Tomlin. Symbolic reachable set computation of piecewise affine hybrid automata and its application to biological modelling: Delta-Notch protein signalling. *Systems Biology*, 1(1):170–183, 2004.

[67] E.P. Gianchandani, A.R. Joyce, B.Ø. Palsson, and J.A. Papin. Functional States of the Genome-Scale *Escherichia coli* Transcriptional Regulatory System. *PLoS Computational Biology*, 5(6):e1000403, 2009.

[68] E.P. Gianchandani, J.A. Papin, N.D. Price, A.R. Joyce, and B.O. Palsson. Matrix Formalism to Describe Functional States of Transcriptional Regulatory Systems. *PLoS Computational Biology*, 2(8):e101, 2006.

[69] D. Gilbert and M. Heiner. From Petri Nets to Differential Equations-An Integrative Approach for Biochemical Network Analysis. In *Petri nets and other models of concurrency, ICATPN 2006*, volume 4024, pages 181–200. Springer, 2006.

[70] N. Goldenfeld. Simple lessons from complexity. *Science*, 284(5411):87–89, 1999.

[71] B.D. Gomperts, I.M. Kramer, and P.E.R. Tatham. *Signal transduction*. Academic Press, 2009.

[72] P.P. Gonzalez, M. Cardenas, D. Camacho, A. Franyuti, O. Rosas, and J. Lagunez-Otero. Cellulat: an agent-based intracellular signalling model. *BioSystems*, 68(2–3):171–185, 2003.

[73] M.R. Grant, K.E. Mostov, T.D. Tlsty, and C.A. Hunt. Simulating Properties of *in vitro* Epithelial Cell Morphogenesis. *PLoS Computational Biology*, 2(10):e129, 2006.

[74] G. Gruenert, B. Ibrahim, T. Lenser, M. Lohel, T. Hinze, and P. Dittrich. Rule-based spatial modeling with diffusing, geometrically constrained molecules. *BMC Bioinformatics*, 11(1):307, 2010.

[75] M. Grzegorczyk, D. Husmeier, K.D. Edwards, P. Ghazal, and A.J. Millar. Modelling non-stationary gene regulatory processes with a non-homogeneous Bayesian network and the allocation sampler. *Bioinformatics*, 24(18):2071–2078, 2008.

[76] S. Gupta, S.S. Bisht, R. Kukreti, S. Jain, and S.K. Brahmachari. Boolean network analysis of a neurotransmitter signaling pathway. *Journal of Theoretical Biology*, 244(3):463–469, 2007.

[77] S. Hardy and P.N. Robillard. Petri net-based method for the analysis of the dynamics of signal propagation in signaling pathways. *Bioinformatics*, 24(2):209–217, 2008.

[78] D. Harel. Statecharts: A visual formalism for complex systems. *Science of computer programming*, 8(3):231–274, 1987.

[79] L.H. Hartwell, J.J. Hopfield, S. Leibler, and A.W. Murray. From molecular to modular cell biology. *Nature*, 402:C47–C52, 1999.

[80] J.J. Heijnen. Approximative kinetic formats used in metabolic network modeling. *Biotechnology and Bioengineering*, 91(5):534–545, 2005.

[81] W.S. Hlavacek, J.R. Faeder, M.L. Blinov, A.S. Perelson, and B. Goldstein. The complexity of complexes in signal transduction. *Biotechnology and Bioengineering*, 84(7):783–794, 2003.

[82] F. Horn and R. Jackson. General mass action kinetics. *Archive for Rational Mechanics and Analysis*, 47(2):81–116, 1972.

[83] M. Hucka, F.T. Bergmann, S. Hoops, S.M. Keating, S. Sahle, J.C. Schaff, L.P. Smith, and D.J. Wilkinson. The Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 1 Core. *Nature Precedings*, 2010.

[84] M. Hucka, A. Finney, HM Sauro, H. Bolouri, JC Doyle, H. Kitano, AP Arkin, BJ Bornstein, D. Bray, A. Cornish-Bowden, et al. The Systems Biology Markup Language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, 2003.

[85] C.A. Hunt, G.E.P. Ropella, S. Park, and J. Engelberg. Dichotomies between computational and mathematical models. *Nature Biotechnology*, 26(7):737–738, 2008.

[86] D. Husmeier. Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks. *Bioinformatics*, 19(17):2271–2282, 2003.

[87] H. Iba. Inference of differential equation models by genetic programming. *Information Sciences*, 178(23):4453–4468, 2008.

[88] N. Jamshidi, J.S. Edwards, T. Fahland, G.M. Church, and B.O. Palsson. Dynamic simulation of the human red blood cell metabolic network. *Bioinformatics*, 17(3):286–287, 2001.

[89] N. Jamshidi and B.O. Palsson. Mass action stoichiometric simulation models: Incorporating kinetics and regulation into stoichiometric models. *Biophysical Journal*, 98:175–185, 2010.

[90] H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, 2000.

[91] M. John, R. Ewald, and A.M. Uhrmacher. A spatial extension to the π-Calculus. *Electronic Notes in Theoretical Computer Science*, 194(3):133–148, 2008.

[92] C. Kaleta, F. Centler, P.S. Di Fenizio, and P. Dittrich. Phenotype prediction in regulated metabolic networks. *BMC Systems Biology*, 2(1):37, 2008.

[93] C. Kaleta, S. Richter, and P. Dittrich. Using chemical organization theory for model checking. *Bioinformatics*, 25(15):1915–1922, 2009.

[94] N. Kam, I.R. Cohen, and D. Harel. The Immune System as a Reactive System: Modeling T Cell Activation With Statecharts. In *Proceedings of the IEEE 2001 Symposia on Human Centric Computing Languages and Environments (HCC'01)*, pages 15–22. IEEE Computer Society, 2001.

[95] G. Karlebach and R. Shamir. Modelling and analysis of gene regulatory networks. *Nature Reviews Molecular Cell Biology*, 9(10):770–780, 2008.

[96] K.J. Kauffman, P. Prakash, and J.S. Edwards. Advances in flux balance analysis. *Current opinion in Biotechnology*, 14(5):491–496, 2003.

[97] S.A. Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, 22(3):437–467, 1969.

[98] L.B. Kier, D. Bonchev, and G.A. Buck. Modeling biochemical networks: a cellular-automata approach. *Chemistry & Biodiversity*, 2(2):233–243, 2005.

[99] J.I. Kim, J.D. Varner, and D. Ramkrishna. A hybrid model of anaerobic *E. coli* GJT001: Combination of elementary flux modes and cybernetic variables. *Biotechnology Progress*, 24(5):993–1006, 2008.

[100] S.Y. Kim, S. Imoto, and S. Miyano. Inferring gene networks from time series microarray data using dynamic Bayesian networks. *Briefings in Bioinformatics*, 4(3):228–235, 2003.

[101] S. Kimura, K. Ide, A. Kashihara, M. Kano, M. Hatakeyama, R. Masui, N. Nakagawa, S. Yokoyama, S. Kuramitsu, and A. Konagaya. Inference of S-system models of genetic networks using a cooperative coevolutionary algorithm. *Bioinformatics*, 21(7):1154–1163, 2005.

[102] H. Kitano. Systems Biology: A Brief Overview. *Science*, 295(5560):1662–1664, 2002.

[103] H. Kitano, A. Funahashi, Y. Matsuoka, and K. Oda. Using process diagrams for the graphical representation of biological networks. *Nature Biotechnology*, 23(8):961–966, 2005.

[104] S. Klamt, J. Saez-Rodriguez, J.A. Lindquist, L. Simeoni, and E.D. Gilles. A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinformatics*, 7(1):56, 2006.

[105] M. Klann, A. Lapin, and M. Reuss. Agent-based simulation of reactions in the crowded and structured intracellular environment: Influence of mobility and location of the reactants. *BMC Systems Biology*, 5(1):71, 2011.

[106] I. Koch, B.H. Junker, and M. Heiner. Application of Petri net theory for modelling and validation of the sucrose breakdown pathway in the potato tuber. *Bioinformatics*, 21(7):1219–1226, 2005.

[107] K.W. Kohn. Molecular interaction map of the mammalian cell cycle control and DNA repair systems. *Molecular Biology of the Cell*, 10(8):2703–2734, 1999.

[108] K.W. Kohn, M.I. Aladjem, S. Kim, J.N. Weinstein, and Y. Pommier. Depicting combinatorial complexity with the molecular interaction map notation. *Molecular Systems Biology*, 2(51), 2006.

[109] D.S. Kompala, D. Ramkrishna, and G.T. Tsao. Cybernetic modeling of microbial growth on multiple substrates. *Biotechnology and Bioengineering*, 26(11):1272–1281, 1984.

[110] R. Küffner, R. Zimmer, and T. Lengauer. Pathway analysis in metabolic databases via differential metabolic display (DMD). *Bioinformatics*, 16(9):825–836, 2000.

[111] N. Le Novère, M. Hucka, H. Mi, S. Moodie, F. Schreiber, A. Sorokin, E. Demir, K. Wegner, M.I. Aladjem, S.M. Wimalaratne, et al. The systems biology graphical notation. *Nature Biotechnology*, 27(8):735–741, 2009.

[112] J.M. Lee, E.P. Gianchandani, J.A. Eddy, and J.A. Papin. Dynamic Analysis of Integrated Signaling, Metabolic, and Regulatory Networks. *PLoS Computational Biology*, 4(5):e1000086, 2008.

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

[114] F. Li, T. Long, Y. Lu, Q. Ouyang, and C. Tang. The yeast cell-cycle network is robustly designed. *Proceedings of the National Academy of Sciences of the United States of America*, 101(14):4781–4786, 2004.

[115] N.Y.K. Li, K. Verdolini, G. Clermont, Q. Mi, E.N. Rubinstein, P.A. Hebda, and Y. Vodovotz. A patient-specific *in silico* model of inflammation and healing tested in acute vocal fold injury. *PLoS ONE*, 3(7):e2789, 2008.

[116] W. Liebermeister and E. Klipp. Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. *Theoretical Biology and Medical Modelling*, 3(1):1–13, 2006.

[117] C.M. Lloyd, M.D.B. Halstead, and P.F. Nielsen. CellML: its future, present and past. *Progress in Biophysics and Molecular Biology*, 85(2-3):433–450, 2004.

[118] P.L. Lollini, S. Motta, and F. Pappalardo. Discovery of cancer vaccination protocols with a genetic algorithm driving an agent based simulator. *BMC Bioinformatics*, 7(1):352, 2006.

[119] A. Luna, E. Karac, M. Sunshine, L. Chang, R. Nussinov, M. Aladjem, and K. Kohn. A formal MIM specification and tools for the common exchange of MIM diagrams: an XML-Based format, an API, and a validation method. *BMC Bioinformatics*, 12(1):167, 2011.

[120] W. Materi and D.S. Wishart. Computational systems biology in drug discovery and development: methods and applications. *Drug Discovery Today*, 12(7-8):295–303, 2007.

[121] E. Merelli, G. Armano, N. Cannata, F. Corradini, M. d'Inverno, A. Doms, P. Lord, A. Martin, L. Milanesi, S. Moller, et al. Agents in bioinformatics, computational and systems biology. *Briefings in Bioinformatics*, 8(1):45–59, 2007.

[122] R. Milner. *A calculus of communicating systems*. Springer, 1980.

[123] R. Milner, J. Parrow, and D. Walker. A calculus of mobile processes, i. *Information and Computation*, 100(1):1–40, 1992.

[124] M. Nagasaki, S. Onami, S. Miyano, and H. Kitano. Bio-calculus: Its Concept and Molecular Interaction. In *Genome informatics. Workshop on Genome Informatics*, volume 10, pages 133–143, 1999.

[125] D. Noble. The rise of computational biology. *Nature Reviews Molecular Cell Biology*, 3(6):459–463, 2002.

[126] J. Nummela and B.A. Julstrom. Evolving petri nets to represent metabolic pathways. In *Proceedings of the 2005 conference on genetic and evolutionary computation*, pages 2133–2139. ACM, 2005.

[127] B.O. Palsson. *Systems Biology: Properties of Reconstructed Networks*. Cambridge University Press, 2006.

[128] J.A. Papin, T. Hunter, B.O. Palsson, and S. Subramaniam. Reconstruction of cellular signalling networks and analysis of their properties. *Nature Reviews Molecular Cell Biology*, 6(2):99–111, 2005.

[129] K.R. Patil, I. Rocha, J. Forster, and J. Nielsen. Evolutionary programming as a platform for *in silico* metabolic engineering. *BMC Bioinformatics*, 6(1):308, 2005.

[130] G. Paun. Computing with membranes. *Journal of Computer and System Sciences*, 61(1):108–143, 2000.

[131] G.A. Pavlopoulos, A.L. Wegener, and R. Schneider. A survey of visualization tools for biological network analysis. *Biodata mining*, 1(1):12, 2008.

[132] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., 1988.

[133] M. Pedersen and G. Plotkin. A language for biochemical systems: Design and formal specification. *Transactions on Computational Systems Biology XII*, 5945:77–145, 2010.

[134] JM Pena, J. Bjorkegren, and J. Tegnér. Growing Bayesian network models of gene networks from seed genes. *Bioinformatics*, 21(Suppl 2):224–229, 2005.

[135] M. Pérez-Jiménez and F. Romero-Campero. P systems, a new computational modelling tool for systems biology. *Transactions on Computational Systems Biology VI*, 4220:176–197, 2006.

[136] C.A. Petri. *Kommunikation mit Automaten*. PhD thesis, Rheinisch-Westfälisches Institut f. instrumentelle Mathematik an d. Univ., 1962.

[137] M. Pogson, M. Holcombe, R. Smallwood, and E. Qwarnstrom. Introducing Spatial Information into Predictive NF-κB Modelling–An Agent-Based Approach. *PLoS ONE*, 3(6):e2367, 2008.

[138] M. Pogson, R. Smallwood, E. Qwarnstrom, and M. Holcombe. Formal agent-based modelling of intracellular chemical interactions. *BioSystems*, 85(1):37–45, 2006.

[139] C. Priami, P. Ballarini, and P. Quaglia. BlenX4Bio–BlenX for Biologists. In *Computational Methods in Systems Biology*, pages 26–51. Springer, 2009.

[140] C. Priami and P. Quaglia. Beta binders for biological interactions. In *Computational Methods in Systems Biology*, pages 20–33. Springer, 2005.

[141] C. Priami, A. Regev, E. Shapiro, and W. Silverman. Application of a stochastic name-passing calculus to representation and simulation of molecular processes. *Information Processing Letters*, 80(1):25–31, 2001.

[142] N.D. Price, J.A. Papin, C.H. Schilling, and B.O. Palsson. Genome-scale microbial *in silico* models: the constraints-based approach. *Trends in Biotechnology*, 21(4):162–169, 2003.

[143] E. Ravasz, A.L. Somera, D.A. Mongru, Z.N. Oltvai, and A.L. Barabási. Hierarchical organization of modularity in metabolic networks. *Science*, 297(5586):1551–1555, 2002.

[144] V.N. Reddy, M.L. Mavrovouniotis, and M.N. Liebman. Petri Net Representations in Metabolic Pathways. In *Proceedings of the 1st International Conference on Intelligent Systems for Molecular Biology*, pages 328–336. AAAI Press, 1993.

[145] J.L. Reed and B.O. Palsson. Thirteen years of building constraint-based *in silico* models of *Escherichia coli*. *Journal of Bacteriology*, 185(9):2692–2699, 2003.

[146] A. Regev, E.M. Panina, W. Silverman, L. Cardelli, and E. Shapiro. BioAmbients: an abstraction for biological compartments. *Theoretical Computer Science*, 325(1):141–167, 2004.

[147] A. Regev, W. Silverman, and E. Shapiro. Representing biomolecular processes with computer process algebra: π-calculus programs of sig-

nal transduction pathways. In *Proceedings of Pacific Symposium of Biocomputing*, 2000.

[148] A. Regev, W. Silverman, and E. Shapiro. Representation and simulation of biochemical processes using the-calculus process algebra. In *Pacific Symposium on Biocomputing*, volume 6, pages 459–470, 2001.

[149] M. Rizzi, M. Baltes, U. Theobald, and M. Reuss. *In vivo* analysis of metabolic dynamics in *Saccharomyces cerevisiae*: II. mathematical model. *Biotechnology and Bioengineering*, 55(4):592–608, 1997.

[150] K. Sachs, D. Gifford, T. Jaakkola, P. Sorger, and D.A. Lauffenburger. Bayesian network approach to cell signaling pathway modeling. *Sci STKE*, 2002(148):pe38, 2002.

[151] K. Sachs, O. Perez, D. Pe'er, D.A. Lauffenburger, and G.P. Nolan. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721):523–529, 2005.

[152] A. Sackmann, M. Heiner, and I. Koch. Application of Petri net based analysis techniques to signal transduction pathways. *BMC Bioinformatics*, 7(1):482, 2006.

[153] J. Saez-Rodriguez, L. Simeoni, J.A. Lindquist, R. Hemenway, U. Bommhardt, B. Arndt, U.U. Haus, R. Weismantel, E.D. Gilles, S. Klamt, and B. Schraven. A logical model provides insights into T cell receptor signaling. *PLoS Computational Biology*, 3(8):e163, 2007.

[154] M.A. Savageau and E.O. Voit. Recasting nonlinear differential equations as S-systems: a canonical nonlinear form. *Mathematical Biosciences*, 87(1):83–115, 1987.

[155] C.H. Schilling, D. Letscher, and B.O. Palsson. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *Journal of Theoretical Biology*, 203(3):229–248, 2000.

[156] T. Schlitt and A. Brazma. Current approaches to gene regulatory network modelling. *BMC Bioinformatics*, 8(Suppl 6):S9, 2007.

[157] 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 in Biotechnology*, 17(2):53–60, 1999.

[158] S. Schuster, T. Pfeiffer, F. Moldenhauer, I. Koch, and T. Dandekar. Exploring the pathway structure of metabolism: decomposition into subnetworks and application to *Mycoplasma pneumoniae*. *Bioinformatics*, 18(2):351–361, 2002.

[159] J.L. Segovia-Juarez, S. Ganguli, and D. Kirschner. Identifying control mechanisms of granuloma formation during *M. tuberculosis* infection using an agent-based model. *Journal of Theoretical Biology*, 231(3):357–376, 2004.

[160] D. Segrè, D. Vitkup, and G.M. Church. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(23):15112–15117, 2002.

[161] Y. Setty, I.R. Cohen, Y. Dor, and D. Harel. Four-dimensional realistic modeling of pancreatic organogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 105(51):20374–20379, 2008.

[162] T. Shlomi, O. Berkman, and E. Ruppin. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proceedings of the National Academy of Sciences of the United States of America*, 102(21):7695–7700, 2005.

[163] T. Shlomi, Y. Eisenberg, R. Sharan, and E. Ruppin. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Molecular Systems Biology*, 3(1), 2007.

[164] I. Shmulevich, E.R. Dougherty, S. Kim, and W. Zhang. Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics*, 18(2):261–274, 2002.

[165] E. Simao, E. Remy, D. Thieffry, and C. Chaouiya. Qualitative modelling of regulated metabolic pathways: application to the tryptophan biosynthesis in *E. coli*. *Bioinformatics*, 21(suppl 2):190–196, 2005.

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

[167] K. Smallbone, E. Simeonidis, N. Swainston, and P. Mendes. Towards a genome-scale kinetic model of cellular metabolism. *BMC Systems Biology*, 4(1):6, 2010.

[168] O. Sozinova, Y. Jiang, D. Kaiser, and M. Alber. A three-dimensional model of myxobacterial aggregation by contact-mediated interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 102(32):11308–11312, 2005.

[169] L.J. Steggles, R. Banks, O. Shaw, and A. Wipat. Qualitatively modelling and analysing genetic regulatory networks: a Petri net approach. *Bioinformatics*, 23(3):336–343, 2007.

[170] G. Stephanopoulos. Metabolic engineering. *Biotechnology and Bioengineering*, 58(2-3):119–120, 1998.

[171] R. Steuer and B.H. Junker. *Computational models of metabolism: Stability and regulation in metabolic networks*, volume 142. John Wiley & Sons, Inc, 2008.

[172] K. Takahashi, S.N.V. Arjunan, and M. Tomita. Space in systems biology of signaling pathways-towards intracellular molecular crowding *in silico*. *FEBS Letters*, 579(8):1783–1788, 2005.

[173] B. Teusink, J. Passarge, C.A. Reijenga, E. Esgalhado, C.C. Van Der Weijden, M. Schepper, M.C. Walsh, B.M. Bakker, K. Van Dam, H.V. Westerhoff, et al. Can yeast glycolysis be understood in terms of *in vitro* kinetics of the constituent enzymes? Testing biochemistry. *European Journal of Biochemistry*, 267(17):5313–5329, 2000.

[174] B.C. Thorne, A.M. Bailey, and S.M. Peirce. Combining experiments with multi-cell agent-based modeling to study biological tissue patterning. *Briefings in Bioinformatics*, 8(4):245–257, 2007.

[175] AM Turing. The Chemical Basis of Morphogenesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 237(641):37–72, 1952.

[176] T.E. Turner, S. Schnell, and K. Burrage. Stochastic approaches for modelling *in vivo* reactions. *Computational Biology and Chemistry*, 28(3):165–178, 2004.

[177] J.J. Tyson, K.C. Chen, and B. Novak. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Current Opinion in Cell Biology*, 15(2):221–231, 2003.

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

[179] D. Visser and J.J. Heijnen. Dynamic simulation and metabolic redesign of a branched pathway using linlog kinetics. *Metabolic Engineering*, 5(3):164–176, 2003.

[180] J. Vohradsky. Neural network model of gene expression. *The FASEB Journal*, 15(3):846–854, 2001.

[181] J. Von Neumann and A.W. Burks. *Theory of self-reproducing automata*. University of Illinois Press, 1966.

[182] D.C. Walker and J. Southgate. The virtual cell — a candidate coordinator for 'middle-out' modelling of biological systems. *Briefings in Bioinformatics*, 10(4):450, 2009.

[183] J. Weimar. Cellular automata approaches to enzymatic reaction networks. *Cellular Automata*, 2493:294–303, 2002.

[184] W. Wiechert. 13C metabolic flux analysis. *Metabolic Engineering*, 3(3):195–206, 2001.

[185] D.S. Wishart, R. Yang, D. Arndt, P. Tang, and J. Cruz. Dynamic cellular automata: an alternative approach to cellular simulation. *In Silico Biology*, 5(2):139–161, 2005.

[186] D.M. Wittmann, J. Krumsiek, J. Saez-Rodriguez, D.A. Lauffenburger, S. Klamt, and F.J. Theis. Transforming Boolean models to continuous models: methodology and application to T-cell receptor signaling. *BMC Systems Biology*, 3(1):98, 2009.

[187] J. Wu and E. Voit. Hybrid modeling in biochemical systems theory by means of functional petri nets. *Journal of Bioinformatics and Computational Biology*, 7(1):107–34, 2009.

[188] J. Wu and E. Voit. Integrative biological systems modeling: challenges and opportunities. *Frontiers of Computer Science in China*, 3(1):92–100, 2009.

[189] J.U. Wurthner, A.K. Mukhopadhyay, and C.J. Peimann. A cellular automaton model of cellular signal transduction. *Computers in Biology and Medicine*, 30(1):1–21, 2000.

[190] A. Yartseva, H. Klaudel, R. Devillers, and F. Képès. Incremental and unifying modelling formalism for biological interaction networks. *BMC Bioinformatics*, 8(1):433, 2007.

[191] J.D. Young, K.L. Henne, J.A. Morgan, A.E. Konopka, and D. Ramkrishna. Integrating cybernetic modeling with pathway analysis provides

a dynamic, systems-level description of metabolic control. *Biotechnology and Bioengineering*, 100(3):542–559, 2008.

[192] I. Zevedei-Oancea and S. Schuster. Topological analysis of metabolic networks based on Petri net theory. *In Silico Biology*, 3(3):323–345, 2003.

[193] L. Zhang, C.A. Athale, and T.S. Deisboeck. Development of a three-dimensional multiscale agent-based tumor model: simulating gene-protein interaction profiles, cell phenotypes and multicellular patterns in brain cancer. *Journal of Theoretical Biology*, 244(1):96–107, 2007.

[194] R.M. Zorzenon dos Santos and S. Coutinho. Dynamics of HIV infection: A cellular automata approach. *Physical Review Letters*, 87(16):168102, 2001.

[195] M. Zou and S.D. Conzen. A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. *Bioinformatics*, 21(1):71–79, 2005.

Chapter 3

Dynamic *vs* Constraint-based modeling

This chapter is based on the article *"Exploring the gap between dynamic and constraint-based models of metabolism"* (in preparation).

Abstract

Systems biology provides new approaches for metabolic engineering through the development of models and methods for simulation and optimization of microbial metabolism. Currently, there are two different modeling frameworks in common use, dynamic and constraint-based models. The construction of dynamic models with detailed kinetic rate laws has been limited to central pathways due to the large volume of experimental data required for parameter estimation. On the other hand, constraint-based models that define a space of solutions for the steady-state flux distribution, have been used for genome-scale stoichiometric reconstruction. In this work, we explore the relationship between these two types of model by comparing and analyzing the dynamic and constraint-based formulations of the same model of the central carbon metabolism of *E. coli*. Our results show that, if the kinetic parameters of the dynamic model are unconstrained, the space of steady states described by both types of model is the same. However, the imposition of parameter ranges can be mapped into kinetically feasible regions of the solution space. Therefore, if at least some of the kinetic parameters are known, dynamic models can be used to generate constraints that reduce the solution space of constraint-based models, eliminating infeasible solutions and increasing the accuracy of simulation and optimization methods.

3.1 Introduction

The prevalence of systems approaches to biological problems has renewed interest in mathematical models as fundamental research tools for performing *in silico* experiments of biological systems [8]. In the context of metabolic engineering, models of metabolism play an important role in the simulation of cellular behavior under different genetic and environmental conditions [22]. Typical experiments include knockout simulations to study how metabolic flux distributions readjust throughout a given network. With the selection of an optimal set of knockouts or changes in enzyme expression levels, it is desirable to optimize the production of compounds of industrial interest [1, 13].

Systems of ordinary differential equations (ODEs) have been applied in different areas to model dynamical systems. In the context of metabolic networks, they describe the rate of change of metabolite concentrations. These dynamic models contain rate law equations for the reactions as well as their kinetic parameters and initial metabolite concentrations. Building this type of model requires insight into enzyme mechanism to select appropriate rate laws, as well as experimental data for parameter estimation. Therefore, their application has been more limited, but areas of application include central metabolic pathways of well-studied organisms such as *E. coli* [2] and *S. cerevisiae* [14]. There are, however, some recent efforts to overcome these limitations in the reconstruction of large-scale dynamic models, such as through the hybrid dynamic/static approach [27], the ensemble modeling approach [23], and the application of approximative kinetic formats using stoichiometric models as a scaffold [20, 7]. Nevertheless, these techniques have so far been applied to very few organisms.

On the other hand, advances in genome sequencing have facilitated the reconstruction of genome-scale metabolic networks for several organisms, with over 50 reconstructions available to date [12]. Due to the lack of kinetic data at the genome scale, this type of model only accounts for reaction stoichiometry and reversibility. Analysis is performed under the assumption of steady state using a constraint-based formulation that is underdetermined, resulting

in a continuous space of solutions for the reaction flux distributions. This uncertainty of the flux distributions requires additional conditions to determine unique solutions and predictions. Often this takes the form of an optimization based on a particular assumption, such as optimal biomass growth for wild-type [4] and minimization of cellular adjustments for knock-out strains [18, 19]. The inclusion of regulatory constraints, introduced by [3], is a current approach to reduce the size of the solution space and eliminate infeasible solutions.

The two most common model types in use, therefore, represent two extremes. The dynamic ODE formulation contains detailed mechanistic information that gives solutions of the transient dynamic approach to equilibrium from any given set of initial conditions (generally concentrations of enzymes and metabolites), as well as the steady state specified by metabolite concentrations that depend on total enzyme concentrations (for the usual case where they are treated as fixed) but often do not depend on the initial metabolite concentrations. Steady-state fluxes are readily computed from the steady-state concentrations and the rate laws. The constraint-based formulation seems minimalistic by comparison: it has no mechanistic knowledge of any of the chemical reactions beyond their stoichiometry, its solutions have fluxes at steady state, but no information regarding concentrations, or dynamics, and rather than giving a unique solution, it produces a high-dimensional continuum of steady-state solutions (referred to as the flux cone). The dynamic formulation needs significant information (parameters in term of rate constants and total enzyme concentrations, as well as reaction mechanisms to give rate laws), but generally rewards that effort with unique and detailed solutions. The constraint-based formulation requires less (no parameters except maximum fluxes) but delivers less.

Because of these significant differences between dynamic and constraint-based formulations, they treat the effects of network perturbations that might be undertaken as part of a metabolic engineering study very differently. A dynamic formulation will make very specific predictions about the response to a gene knockout, for example, but generally such models lack information about gene regulatory changes that accompany metabolic changes, and so

without foreknowledge to adjust relative enzyme concentrations, such predictions can be significantly in error. Constraint-based formulations can access all possible steady-state solutions but can only rely on relatively simple heuristics to select among them, and are uncertain how to include specific information on gene regulatory changes.

Here we explore further the relationship between these formulations by essentially considering the continuous ensemble of dynamic formulations obtained by varying parameters (principally rate constants and enzyme concentrations) and compare the steady-state solutions to those from the corresponding constraint-based formulation. We find an equivalence between the sets of steady states when only maximum flux constraints are present, but that more specific constraints and enzyme concentrations can be directly incorporated to define a reduced dynamic ensemble that is significantly more informative regarding possible steady-state solutions than the constraint-based formulation.

3.2 Methods

3.2.1 Models

We have used a dynamic model of the central carbon metabolism of *E. coli* [2] available at the Biomodels database [9]. The model was converted from its original SBML format into a MATLAB (The Mathworks; Natick, MA, USA) file which was used for all the computations in this work. The model consists of a total of 18 metabolites and 31 reactions, including several enzymatic reactions, one exchange reaction, and a few lumped versions of biosynthetic pathways. Several types of rate laws are used, including constant rate, mass-action, Hill cooperativity, allosteric regulation, and Michaelis-Menten with its variants for reversibility and inhibition, with a total of 125 parameters. We have not considered metabolite dilution or algebraic rules for co-metabolite variation, as they cannot be represented in the constraint-based model. Also, we changed the rate law of MurSynth from constant rate to Michaelis-Menten, as it leads to inconsistencies when its substrate (f6p) de-

pletes. The model maintained its original steady-state despite these changes.

A constraint-based version of the model was built by accounting only for the stoichiometry and reversibility constraints. The glucose uptake rate was allowed to vary between 0 and the maximum value in the dynamic model. The dynamic model also contains two other inputs (TrpSynth, MethSynth), with a constant rate, that were considered in the constraint-based version with constant fluxes.

3.2.2 Hit-and-Run sampler

We implemented an algorithm for random sampling (Fig. 3.1a) adapted to this problem following the concept of hit-and-run methods [21]. The solution space of the constraint-based model is contained within the null space of the stoichiometric matrix. Starting with a point inside this coordinate space, the sampler started generating new points by iterative steps in one direction. Each point was then projected into the flux space and tested by checking the flux boundary constraints. Each time the test failed, meaning that it crossed the boundary of the flux cone, the point was discarded and a new direction was randomly chosen. Otherwise, the point's projection in the flux space was stored. To facilitate uniform sampling of the whole space, the sampler only stores one point every 1000 iterations. Also, in order to adapt to cones of different sizes, it used a variable step size that increased (decreased) in case of successful (failed) iterations, which quickly converged to an average size.

3.2.3 Geometric sampler

Given the poor results obtained by the hit-and-run method at the edges of the cone, we designed and implemented a geometric sampler (Fig. 3.1b) that started by searching the corners of the flux cone. It found the corners by solving linear programing problems within the model with randomized objective functions using the GLPK library [11]. After finding the corners, it sampled along all possible edges between the corners, which defined the bounds of the cone. Then, it iteratively sampled from all edges in the direction of the center of the cone, defined as the mean of all corners. This method facili-

tated the visualization of the flux cone. However, in this case, the probability distribution of the points did not have any statistical meaning.

3.2.4 Parameter sampler

Metabolite concentrations and kinetic parameters are theoretically defined in an infinite semi-positive space. Therefore, in order to sample this type of space without constraints, we scaled each element individually (concentration or parameter) by a random factor with log-normal distribution ($\log_{10}(X) \sim \mathcal{N}(0,1)$). This distribution is defined over $\mathbb{R}^+$, with nearly all values (99.73 %) within 3 orders of magnitude above or bellow unit. This resulted in variation of the original values by several orders of magnitude. In order to perform constrained parameter variation within well-defined ranges, specified in terms of orders of magnitude (m), we scaled each parameter by a factor with uniform distribution in logarithmic scale ($log_{10}(X) \sim \mathcal{U}(-m/2, m/2)$). All kinetic parameters associated with binding and rate constants were varied, while other parameters such as Hill coefficients, co-metabolite levels and dilution rate, were kept fixed.

3.2.5 Calculating steady states

For each simulation of the dynamic model, the steady state was calculated by numerically integrating the differential equations from time zero toward infinity with a stop condition when the steady state was reached. To avoid non-halting computations when the system diverged or was oscillatory, a second stop condition, based on a computational time limit, was also added.

3.2.6 Relative volume estimation

In order to estimate the volume of the cone after imposition of the kinetic parameter constraints, we started by sampling the dynamic model under those constraints. In this way the kinetic parameter ranges could be mapped to flux ranges (Fig. 3.1e). Then, we used a random sample of the constraint-based model (obtained with the hit-and-run sampler) and calculated the

fraction of points of that sample that were contained within the generated flux ranges. This fraction determined the relative volume of the subspace compared to the original space [25].

3.3 Results

In order to explore the gap between both types of formulations, we analyzed and compared the dynamic and constraint-based formulations of the same model of the central carbon metabolism of *E. coli* [2] (see Methods section for model formulation).

Our goal is to compare the steady states achievable by the two model types. Intuitively the dynamic formulation has more constraints than the constraint-based one because the later only enforces the steady-state condition and maximum flux constraints. Therefore, any set of steady-state fluxes achieved by the dynamic formulation and that do not violate the maximum flux constraints will automatically be a solution of the constraint-based formulation. Thus, here we focus on mapping solutions in the opposite direction: Is every solution of the constraint-based formulation also a steady-state solution of the dynamic one? Or, instead, does the extra information in the dynamic formulation effectively reduce the steady-state solution space so that it is a proper subset of the constraint-based formulation.

3.3.1 Solution space of the constraint-based model

We implemented a Monte-Carlo based random sampler, which is a variation of the hit-and-run method [21] (see Methods) and applied it to the constraint-based model. The sampling distribution for each reaction (Fig. 3.2, diagonal) forms skewed gaussian shaped curves, very similar to the results obtained by [25] for the human red blood cell model. However, more insight into the shape of the solution space can be revealed by plotting the sample two-dimensionally for every pair of reactions (Fig. 3.2). It is possible to observe that, due to the random nature of this method, the edges of the flux cone are not sharply defined due to the low probability of samples in the tails of

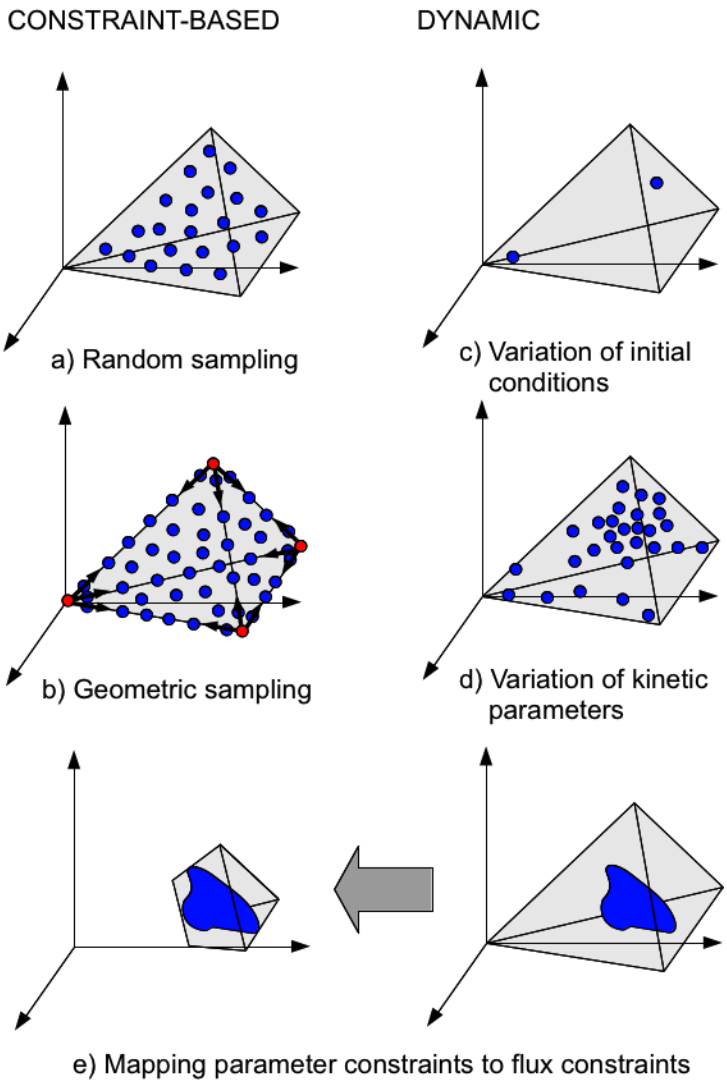

Figure 3.1: Overview of the methods applied in this work to the constraint-based and the dynamic model. The solution space of the constraint-based model has been sampled by (a) random sampling using a Hit-and-Run algorithm, and (b) geometric sampling using the corners of the flux cone as starting points. The solution space of the dynamic model has been sampled by (c) varying the initial metabolite concentrations, and (d) the kinetic parameters. (e) By constraining the kinetic parameters of the dynamic model we can delimit kinetically feasible flux regions and transfer them to the constraint-based model.

the gaussians. To obtain a clearer delineation of the borders of the space, we implemented a geometric sampling approach that systematically identified first the vertices of the flux cone through solution of linear programs, then the edges through vertex connection, and finally explored the interior of the flux cone (see Methods). The full solution space of the constraint-based model is now clear (Fig 3.2), and it can be compared to that from the dynamic model.

3.3.2 Solution space of the dynamic model

Whereas the constraint-based model has no adjustable parameters, the dynamic model has a large number of parameters that describe the specific chemistry being modeled, consisting of the rate laws, kinetic parameters (in which we include a fixed total concentration for each protein), and initial metabolite concentrations. This results in a single deterministic steady-state solution. To examine how this solution is influenced by the extra information, we varied the initial conditions and kinetic parameters, again by random sampling (see Methods).

If the system has a unique steady state, then simulations will converge to the same steady state, independent of the initial concentrations. This network exhibits multistability; two distinct steady states were identified when the initial concentrations of metabolites were varied (Fig. 3.3). This bi-stability is caused by a positive feedback loop that is formed when phosphoenolpyruvate (PEP), a product of glycolysis, is used as an energy source to import external glucose through the phosphotransferase system (PTS). During the transient phase of the system, the concentration of PEP may reach a critical level, where it gets depleted before re-entering PTS. If this happens, the cell is unable to capture its external substrate, and all internal metabolites eventually deplete as well, leading to a network with residual activity. We can observe that this steady state (referred here as secondary) occurs much less frequently than the steady state obtained with the original conditions (Fig. 3.3, diagonal).

A random procedure was used to vary the kinetic model parameters, including binding and rate constants (because all enzyme concentrations are

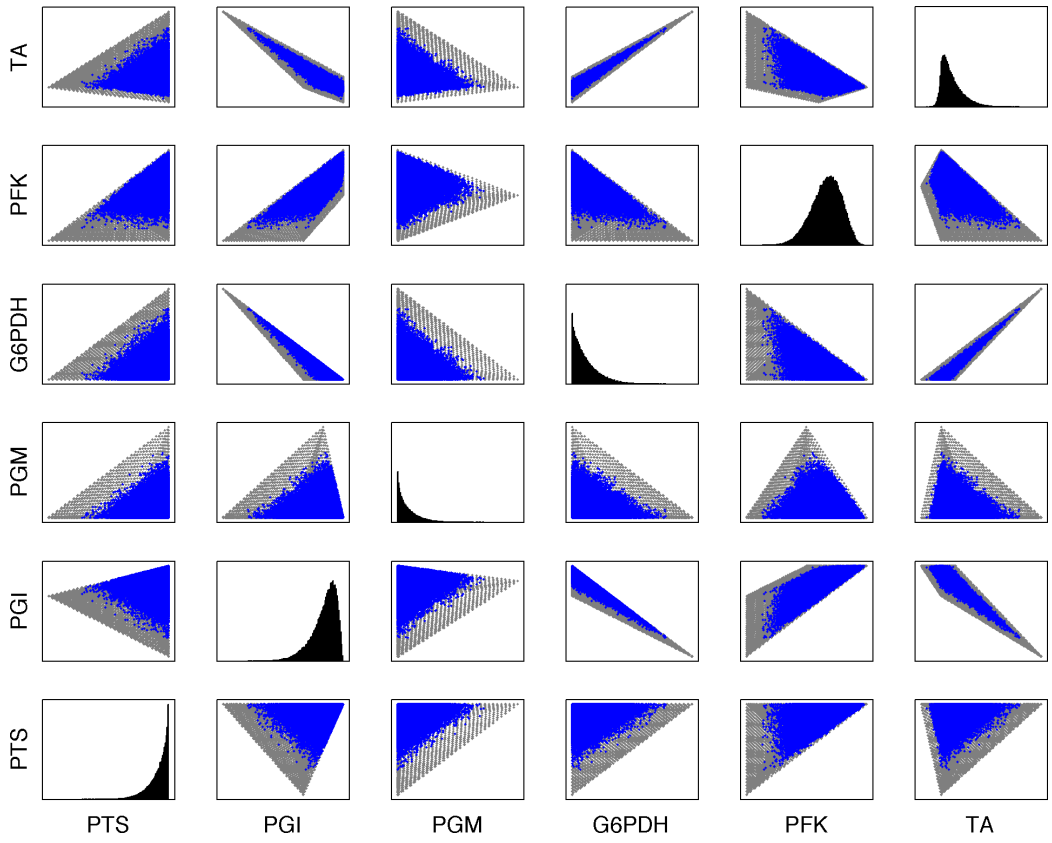

Figure 3.2: Pairwise projection of the sampling of the constraint-based solution space using the hit-and-run sampler (blue) and the geometric sampler (gray). The diagonal shows the probability distribution for each reaction relative to the hit-and-run sampling. Only the first six reactions are shown. Note that the gray points are plotted underneath the blue ones, and that the geometric sampler delineates all of the space region covered by the hit-and-run sampler, plus the additional spaces seen here.

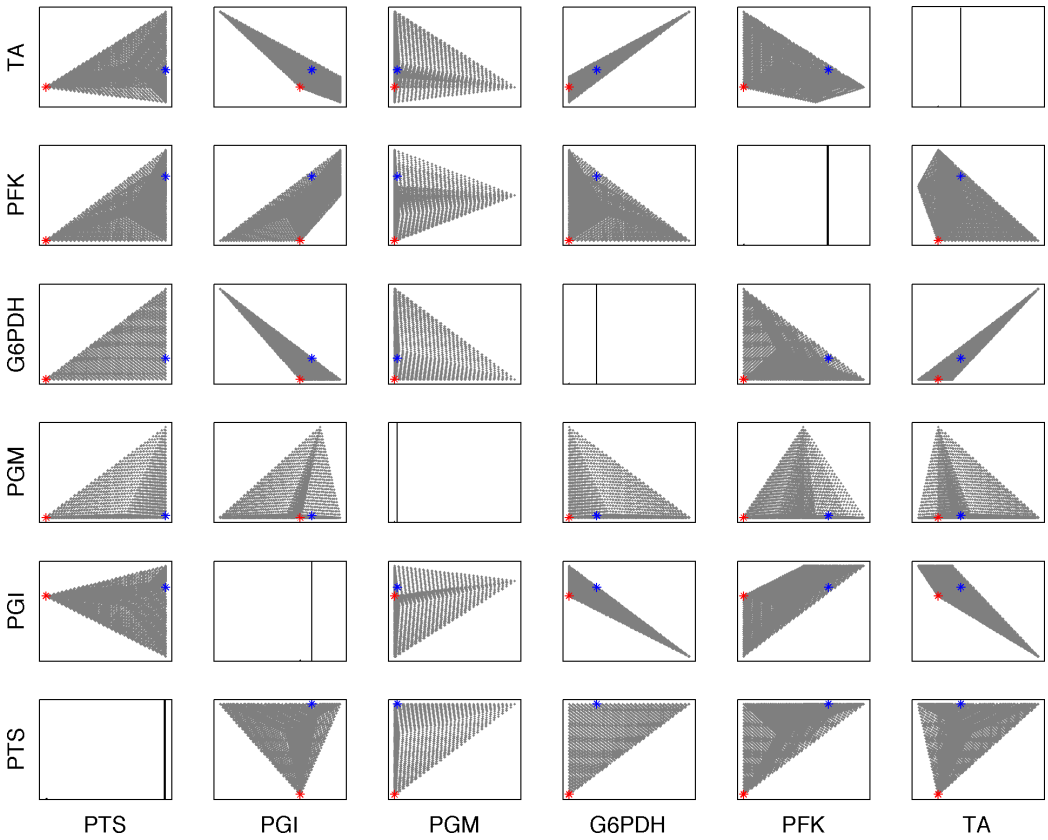

Figure 3.3: Pairwise projection of the sampling of the solution space obtained for the dynamic model by sampling the initial metabolite concentrations, overlapping the complete solution space (gray) for better visualization. The dark blue dot shows the location of the original steady state. The red dot shows the location of the secondary steady state. Only the first six reactions are shown. The diagonal gives the relative probabilities of the steady-state flux distribution.

included in $V_{\max}$, which was varied, effectively enzyme concentrations were varied as well), but not Hill coefficients, co-metabolite concentrations or the dilution rate (see Methods). A single set of initial concentrations was used (that for which the unperturbed model goes to the higher probability steady state). A projection of the resulting steady-state concentrations shows that the dynamic model, through parameter variation, appears to be able to produce the same steady states as the constraint-based model, but no additional steady-states. This situation is tempered by two issues: (i) there are areas of light coverage in Figure 3.4 that one presumes are truly occupied, and (ii) even if the two-dimensional projection overlaps, this does not confirm that the full-dimensional flux cones for the two models overlap. To more stringently test the notion that the polytopes are identical, we generated a procedure that would optimize parameters for the dynamic model to reproduce any desired steady-state solution (see Methods). We applied this to 10,000 randomly selected solutions from the constraint-based model and the resulting parameters recovered the desired steady state when run in the dynamic model every time. Thus, operationally the steady-state flux cones for ODE and constraint-based models are the same.

3.3.3 Kinetically feasible solution space

An ODE kinetic model of central carbon metabolism has exactly the same set of possible steady-state solutions as the corresponding flux balance model, as demonstrated in the previous section. The ODE model maps out the solution space through systematic variation of model parameters (binding constants, rate constants, and enzyme concentrations) with no constraints beyond non-negativity. Knowledge of actual parameter values or ranges, from experimental measurement or physical constraints, would lead to further constraints on the feasible parameter space. To explore how constraints on the feasible parameter space affect the range of steady-state solutions achievable in the ODE kinetic model, we sampled parameter combinations from constrained spaces and computed the steady states of the resulting models. The fluxes in those steady states are plotted in Figure 3.5 for pa-

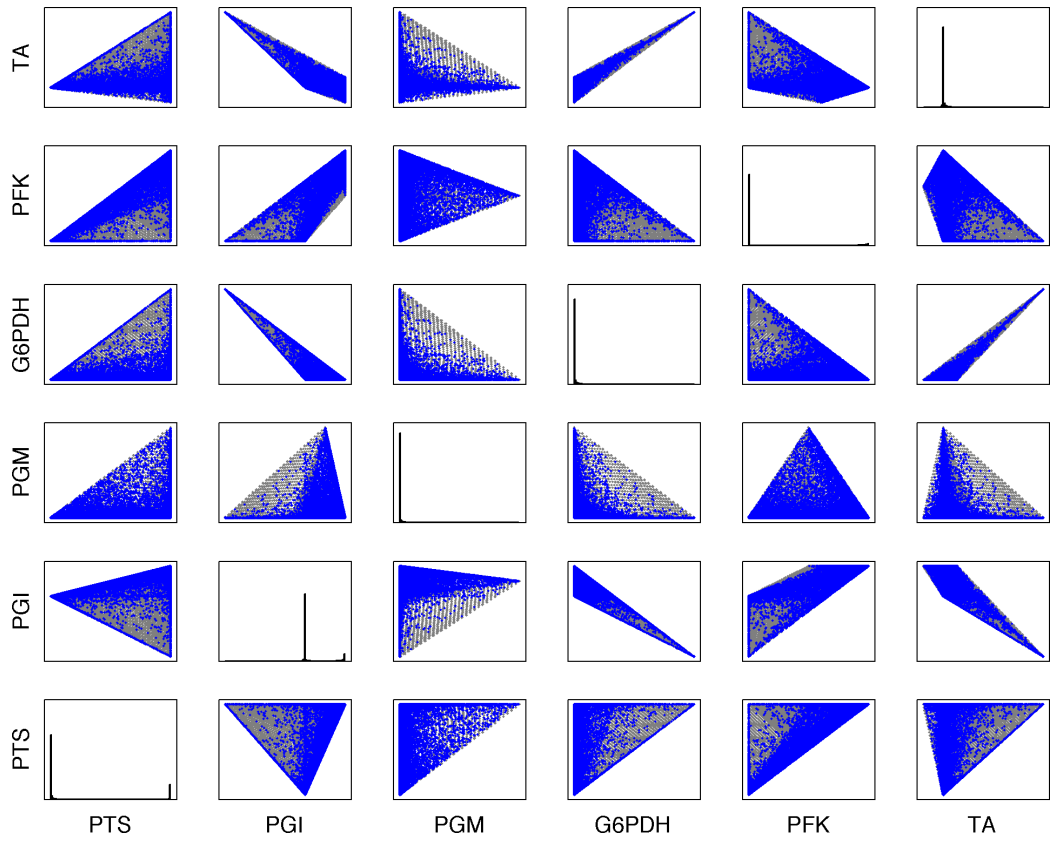

Figure 3.4: Pairwise projection of the sampling of the steady-state solution space for the dynamic model obtained by sampling the kinetic parameters. The corresponding space overlaps the solution space given by the stoichiometric model (gray). The diagonal shows the probability distribution for each reaction. Only the first six reactions are shown.

rameter ranges from 10^0 up to 10^4-fold around the base parameter values. The results show that parameter variation of 100-fold or greater appears to produce the full set of steady-state flux solutions observed from the unconstrained non-negative parameters in the ODE model, which corresponds to the flux-balance steady states. Parameter constraints leading to less than 100-fold variation produced significant restriction of the steady-state fluxes.

The solution-space volume reduction due to parameter constraints is plotted quantitatively in Figure 3.6. The ratio of the solution flux cone with constrained and unconstrained parameters is shown as a function of the constrained parameter ranges. The results (labeled "normal uptake"), show that reduction of parameter uncertainty to a 10-fold range leads to a reduction in the solution flux space to 10% of its unconstrained volume. Moreover, because the size of the original space depends on a control variable of the system, namely the glucose uptake rate, we increased glucose uptake from 1.28 mmol/gDW/h, the value in the original model, to 10.50 mmol/gDW/h, the maximum value for *E. coli* under aerobic conditions [24]. The results, shown in Figure 3.6 as "maximum uptake" show a similar sigmoid shape, but shifted toward greater parameter variation. Under these condition, the flux cone of solutions is reduced to 10% of its unconstrained volume with 300-fold parameter variation.

3.4 Discussion

We have analyzed and compared dynamic and constraint-based formulations of the same model for the central carbon metabolism of *E. coli* [2]. The constraint-based version does not account for metabolite concentrations, and it does not express transient behavior. Therefore, the formulations can only be compared at their common domain, which is the steady-state flux distribution.

The constraint-based model defines a solution space for the steady-state flux distribution (usually called the flux cone). This space is difficult to visualize due to its high dimensionality. We addressed this problem by developing sampling and projection approaches that facilitate the visualization

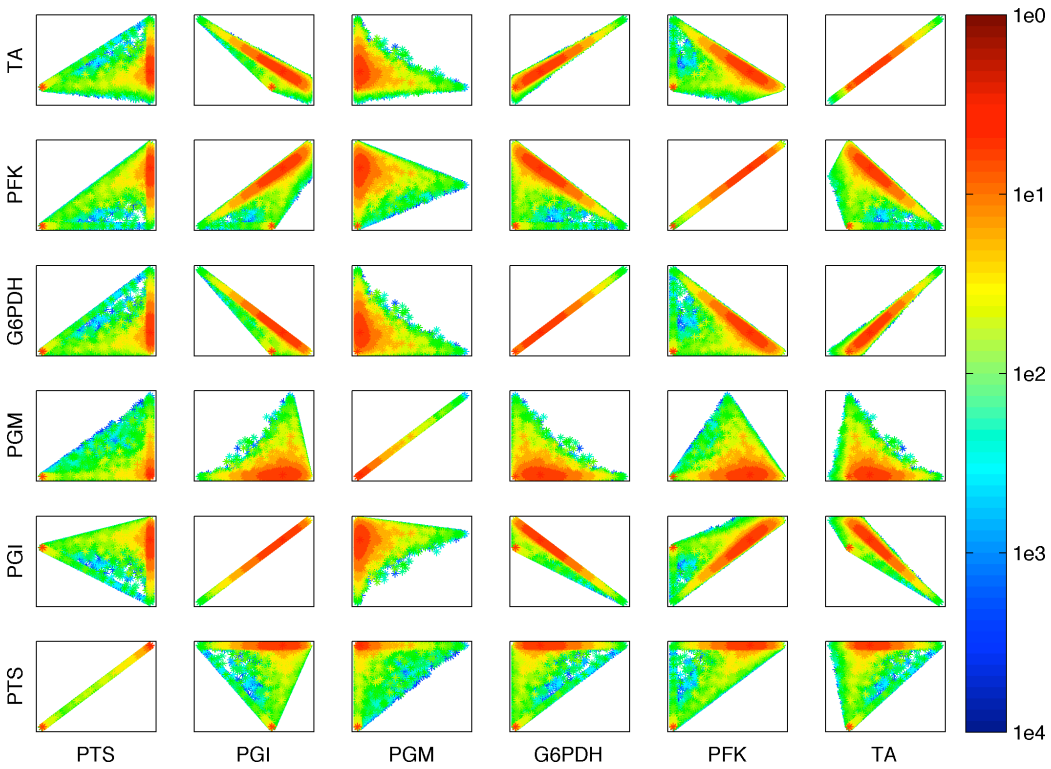

Figure 3.5: Pairwise projection, in heat-map form, of the solution space reachable by the dynamic model as a function of the variation, in terms of orders of magnitude, of the kinetic parameter space. The diagonal shows the variation for each flux independently. Only the first six reactions are shown.

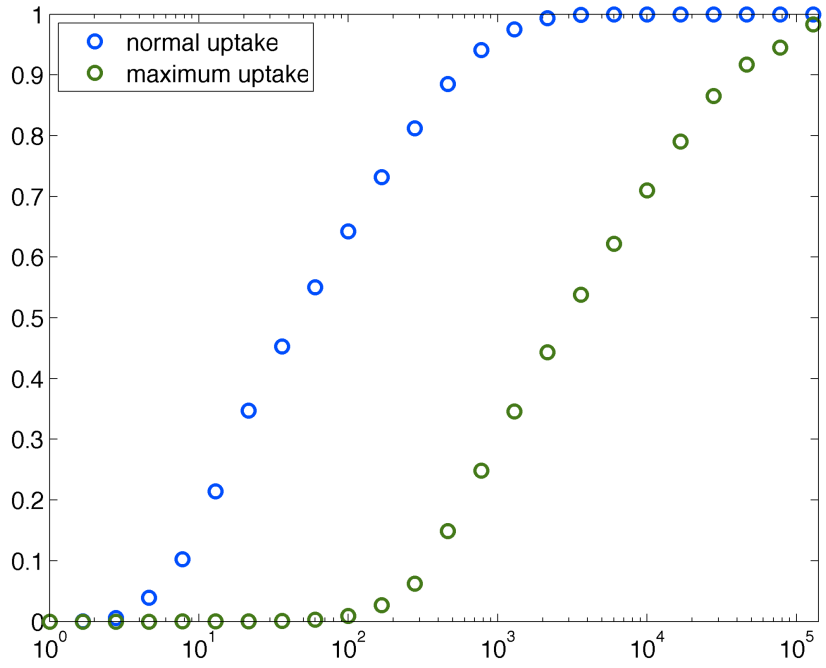

Figure 3.6: Relative volume of the kinetically feasible solution space, compared to the original space, as a function of the parameter variation, in terms of orders of magnitude. The volume was calculated for the original glucose uptake rate in the model and also for the maximum uptake rate.

of the shape of solution space.

The steady state of the dynamic model contains the same constraints as the constraint-based model (stoichiometry, thermodynamic reversibility, and maximum uptake rates) and also any additional constraints imposed by the kinetic rate laws, kinetic parameters, and initial metabolite concentrations. Therefore, its solution space is a subset of the constraint-based solution space.

For a predefined set of initial conditions and parameter values, the dynamic model usually determines one steady-state solution. In fact, the initial metabolite concentrations of dynamic models determine their transient behavior, but, for the steady-state flux determination, they serve only to determine which steady-state is chosen in case of multiple stabilities in the model. In this case, sampling the metabolite concentration space revealed a second steady-state characterized by a flux distribution with lower values of the fluxes and an accumulation of external glucose.

Instead, we also verified, as expected, that the location of the steady-state solution(s) inside the solution space is determined by the kinetic parameters, because by varying the kinetic parameters, the solution moves inside the solution space. The sampling of the kinetic parameter space revealed that, with unconstrained parameter values, the solutions of the dynamic model cover the whole steady-state solution space. This overlapping may seem unintuitive, as one would expect the rate laws to impose one additional layer of constraint into the steady-state solution space. However, besides having observed this with our sampling approaches, we also observe that, given any valid steady-state flux distribution, one can find kinetic parameter values that make the rate laws produce those steady state flux values by solving each equation separately. This separation is only possible because the parameters are specific for each rate law, which defines a partition over the parameter set. The running example contains an average of 4 parameters per rate law, yielding many degrees of freedom for each equation. Thus, it is not surprising that, generally, parameter values can be found that satisfy the equations.

One interesting result is that, in general, tuning the maximum rate constants is sufficient to make the rate laws fit any given steady-state flux. This is valid for irreversible reactions given any metabolite concentrations. And,

it is also valid for reversible reactions with the additional constraint that the ratio between the given product and substrate concentrations exceeds the equilibrium constant. In fact, the maximum rate constant is the only parameter that the cell can control by adjusting the enzyme concentration levels (recall that $V_{\max} = k_{\text{cat}}[E]_0$). This means that disregarding the details of the transcriptional regulation (e.g. the fact that different genes share the same transcription factor, or that same genes are associated with different reactions) and the practical limitations in enzyme concentration, the cell can in theory adjust its steady-state flux distribution within the admissible space imposed by the topology of the metabolic network. This hypothesis reflects the adaptability of the cell under different conditions and is in agreement with the observations that microorganisms can undergo adaptive evolution to attain their optimal theoretical yields when placed under conditions where they originally performed sub-optimally [6].

The observations stated above show that, in theory, a dynamic model can be fitted to any steady-state flux distribution inside the constraint-based solution space. However, there are physical limitations to the values of the kinetic parameters. Also, by querying parameter databases such as BRENDA [17] and SABIO-RK [16], it is possible to observe that for each kinetic parameter there is a range of possible values determined by experimental conditions (such as temperature and pH) in which the cells are able to grow. Therefore, we evaluated how the imposition of parameter ranges map into flux ranges within the steady-state solution space. Although the rate laws do not constrain the solution space by themselves, they influence the probability distribution of the steady-state solutions. This is evidenced by the imposition of the kinetic parameter constraints. As the constraints become tighter, the solutions of lower probability disappear and the reachable solution space becomes smaller. Our results show that the impact of these constraints depends on the size of the solution space of the genome-scale model, which is mainly determined by the uptake rate of the limiting substrates, and on the allowable ranges of the kinetic parameters in the dynamic model.

3.5 Conclusions

In this work we have explored the solution spaces of both dynamic and constraint-based models in order to bring together top-down and bottom-up approaches, and we have proposed methods of treating each as well as their interrelation.

Dynamic model reconstruction is a bottom-up approach for iteratively building large-scale metabolic pathways with kinetic detail. Due to lack of experimental data, differences in experimental conditions, and error prone measurements, the kinetic parameters are often unavailable or defined within certain ranges.

On the other hand, genome-scale reconstruction is a top-down approach that takes advantage available high-throughput data to build models of metabolic networks that account for stoichiometry and thermodynamic constraints. These models are analyzed under a steady-state assumption through the constraint-based approach. Furthermore, they can be iteratively refined by imposition of new constraints that shrink the size of the solution space. One such approach is the imposition of regulatory constraints, which can result in significant reductions [3].

Taking advantage of the information available in dynamic models of central pathways can increase the accuracy of genome-scale constraint-based models by imposition of kinetic feasibility constraints even if the dynamic model is not fully determined. Furthermore, sampling the solution space of the dynamic model can be used as an experimental design tool to determine which kinetic parameters have greater influence in defining the volume of the solution space.

Increasing the accuracy of constraint-based models can influence simulation methods such as metabolic flux analysis (MFA) [26], flux balance analysis (FBA) [4], minimization of metabolic adjustment (MOMA) [18] and regulatory on/off minimization (ROOM) [19]. Tools that implement these methods [15] can be extended to include kinetic constraints.

The constraint-based approach has been recently applied to other kinds of biological networks, namely gene regulatory and signaling networks [5, 10].

The availability of models for all kinds of networks will facilitate the creation of integrated cellular models that account for all types of intracellular phenomena under the same mathematical framework. Because those models can be either constraint-based or dynamic, understanding relationships between the two as discussed in this work will have an even greater impact. In fact, although the use of common frameworks (either constraint-based or dynamic) for representing different kinds of biological phenomena is a step towards the use of integrated models, the development of tools that promote the integration of the two most important representation frameworks is also necessary for true integration. The current contribution is a step in that direction.

Bibliography

[1] A.P. Burgard, P. Pharkya, and C.D. Maranas. OptKnock: A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain Optimization. *Biotechnology and Bioengineering*, 84(6):647–657, 2003.

[2] C. Chassagnole, N. Noisommit-Rizzi, J.W. Schmid, K. Mauch, and M. Reuss. Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnology and Bioengineering*, 79(1):53–73, 2002.

[3] M.W. Covert and B.O. Palsson. Constraints-based models: Regulation of gene expression reduces the steady-state solution space. *Journal of Theoretical Biology*, 221(3):309–325, 2003.

[4] J.S. Edwards and B.O. Palsson. Metabolic flux balance analysis and the *in silico* analysis of *Escherichia coli* K-12 gene deletions. *BMC Bioinformatics*, 1(1):1, 2000.

[5] E.P. Gianchandani, J.A. Papin, N.D. Price, A.R. Joyce, and B.O. Palsson. Matrix Formalism to Describe Functional States of Transcriptional Regulatory Systems. *PLoS Computational Biology*, 2(8):e101, 2006.

[6] 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.

[7] N. Jamshidi and B.O. Palsson. Mass action stoichiometric simulation models: Incorporating kinetics and regulation into stoichiometric models. *Biophysical Journal*, 98:175–185, 2010.

[8] H. Kitano. Systems Biology: A Brief Overview. *Science*, 295(5560):1662–1664, 2002.

[9] N. Le Novere, B. Bornstein, A. Broicher, M. Courtot, M. Donizelli, H. Dharuri, L. Li, H. Sauro, M. Schilstra, B. Shapiro, et al. BioModels Database: A free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Research*, 34(suppl 1):D689, 2006.

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

[11] A. Makhorin. GLPK - The GNU Linear Programming Toolkit, 2006.

[12] M.A. Oberhardt, J. Puchałka, V.A.P.M. dos Santos, and J.A. Papin. Reconciliation of Genome-Scale Metabolic Reconstructions for Comparative Systems Analysis. *PLoS Computational Biology*, 7(3):e1001116, 2011.

[13] K.R. Patil, I. Rocha, J. Forster, and J. Nielsen. Evolutionary programming as a platform for in silico metabolic engineering. *BMC Bioinformatics*, 6(1):308, 2005.

[14] M. Rizzi, M. Baltes, U. Theobald, and M. Reuss. In vivo analysis of metabolic dynamics in *Saccharomyces cerevisiae*: II. Mathematical model. *Biotechnology and Bioengineering*, 55(4):592–608, 1997.

[15] I. Rocha, P. Maia, P. Evangelista, P. Vilaça, S. Soares, J.P. Pinto, J. Nielsen, K.R. Patil, E.C. Ferreira, and M. Rocha. OptFlux: An open-source software platform for in silico metabolic engineering. *BMC Systems Biology*, 4(1):45, 2010.

[16] I. Rojas, M. Golebiewski, R. Kania, O. Krebs, S. Mir, A. Weidemann, and U. Wittig. SABIO-RK: A database for biochemical reactions and their kinetics. *BMC Systems Biology*, 1(Suppl 1):S6, 2007.

[17] I. Schomburg, A. Chang, and D. Schomburg. BRENDA, enzyme data and metabolic information. *Nucleic Acids Research*, 30(1):47–49, 2002.

[18] D. Segrè, D. Vitkup, and G.M. Church. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(23):15112–15117, 2002.

[19] T. Shlomi, O. Berkman, and E. Ruppin. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proceedings of the National Academy of Sciences of the United States of America*, 102(21):7695–7700, 2005.

[20] K. Smallbone, E. Simeonidis, N. Swainston, and P. Mendes. Towards a genome-scale kinetic model of cellular metabolism. *BMC Systems Biology*, 4(1):6, 2010.

[21] R.L. Smith. Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research*, 32(6):1296–1308, 1984.

[22] G. Stephanopoulos. Metabolic engineering. *Biotechnology and Bioengineering*, 58(2-3):119–120, 1998.

[23] L.M. Tran, M.L. Rizk, and J.C. Liao. Ensemble modeling of metabolic networks. *Biophysical Journal*, 95(12):5606–5617, 2008.

[24] A. Varma and B.O. Palsson. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Applied and Environmental Microbiology*, 60(10):3724–3731, 1994.

[25] S.J. Wiback, I. Famili, H.J. Greenberg, and B.O. Palsson. Monte Carlo sampling can be used to determine the size and shape of the steady-state flux space. *Journal of Theoretical Biology*, 228(4):437–447, 2004.

[26] W. Wiechert. 13C metabolic flux analysis. *Metabolic Engineering*, 3(3):195–206, 2001.

[27] K. Yugi, Y. Nakayama, A. Kinoshita, and M. Tomita. Hybrid dynamic/static method for large-scale simulation of metabolism. *Theoretical Biology and Medical Modelling*, 2(1):42, 2005.

Chapter 4

A Framework for Model Transformation

This chapter is based on the article *"Model transformation of metabolic networks using a Petri net based framework"* published at the *International Workshop on Biological Processes & Petri Nets (BioPPN 2010)*, Braga, Portugal, 2010.

Abstract

The different modeling approaches in Systems Biology create models with different levels of detail. The transformation techniques in Petri net theory can provide a solid framework for zooming between these different levels of abstraction and refinement. This work presents a Petri net based approach to Metabolic Engineering that implements model reduction methods to reduce the complexity of large-scale metabolic networks. These methods can be complemented with kinetics inference to build dynamic models with a smaller number of parameters. The central carbon metabolism model of *E. coli* is used as a test-case to illustrate the application of these concepts. Model transformation is a promising mechanism to facilitate pathway analysis and dynamic modeling at the genome-scale level.

4.1 Introduction

Systems Biology provides a new perspective in the study of living systems and embraces the complexity emerging of interactions among all biological components. Combining theory and experiments, scientists build models to explain and predict the behavior of the systems under study. Metabolic Engineering is one of the fields where this perspective has proven useful through the optimization of metabolic processes for industrial applications [29, 2].

Modeling in Systems Biology is an iterative process as the life-cycle of a model is comprised of successive refinements using experimental data. Different approaches, such as top-down, bottom-up or middle-out [18] are used depending on the purpose of the model and the type of data available for its construction. In Metabolic Engineering there are macroscopic kinetic models that consider the cell as a black-box converting substrates into biomass and products, which are typically used for bioprocess control. On the other hand, there are reaction-network-level models, either medium-scale dynamic models with detailed kinetic information derived from literature and experimental data [3], or genome-scale stoichiometric reconstructions derived from genome annotation complemented with literature review [5].

Although the ultimate goal of Systems Biology is a complete understanding of the cell as a whole, not only it is extremely difficult to collect all the kinetic information necessary to build a fully detailed whole-cell model due to the lack of experimental data and model identifiability concerns, but also the computational cost of simulating the dynamics of a system with such detail would be tremendous. Therefore, there is a need to fit the level of detail of a model to the specific problem at hand. For instance, Metabolic Pathway Analysis (MPA) has been useful in the analysis of metabolism as a way to determine, classify and optimize the possible pathways throughout a metabolic network. However, due to the combinatorial explosion of pathways with increasing number of reactions, it is still infeasible to apply these methods in genome-scale metabolic reconstructions without decomposing the network into connected modules [24, 25]. This zooming in and out between

different levels of abstraction and connecting parts with different levels of detail is a feature where formal methods and particularly Petri nets may play an important role. Concepts such as subnetwork abstraction, transition refinement or node fusion, among others, have been explored in Petri net theory [8] and may provide the theoretical background for method development.

In previous work, we reviewed different modeling formalisms used in Systems Biology from a Metabolic Engineering perspective and concluded that Petri nets are a promising formalism for the creation of a common framework of methods for modeling, analysis and simulation of biological networks [15]. They are a mathematical and graphical formalism, therefore intuitive and amenable to analysis. The different extensions available (*e.g.:* stochastic, continuous, hybrid) provide the flexibility required to model and integrate the diversity of phenomena occurring in the main types of biological networks (metabolic, regulatory and signaling). Moreover, one may find biological meaning in several concepts in Petri net theory; for instance, the incidence matrix of a Petri net is the equivalent of the stoichiometric matrix, and the minimal *t-invariants* correspond to the elementary flux modes (EFMs).

In this work, we explore strategies of model reduction for Petri net representations of metabolic networks, and the integration of this methodology with recent approaches such as genome-scale dynamic modeling. This chapter is organized as follows. Section 4.2 explores the motivation for the work. Section 4.3 presents the model reduction and kinetics inference methods, Section 4.4 discusses their application to *E. coli* and Section 4.5 elaborates on conclusions and future work.

4.2 Background

There are different examples of model reduction in the literature. One such method was developed in [17], based on timescale analysis for classification of metabolite turnover time using experimental data. The fast metabolites are removed from the differential equations and their surrounding reactions are lumped. In [20] the EFMs of a reaction network are calculated in order to create a macroscopic pathway network, where each EFM is a

macro-reaction connecting extracellular substrates and products. A simple Michaelis–Menten rate law is assumed for each macro-reaction and the parameters are inferred from experimental data. The method is applied in a network with 18 reactions and a total of 7 EFMs. However it does not scale well to larger networks because, in the worst case, the number of EFMs grows exponentially with the network size.

The combinatorial pathway explosion problem is well known; there are methods for network decomposition in the literature that address this issue. In [24] the authors perform a genome-scale pathway analysis on a network with 461 reactions. After estimating the number of extreme pathways (EPs) to be over a million, the network is decomposed into 6 subsystems according to biological criteria and the set of EPs is computed separately for each subsystem. A similar idea in [25] consists on automatic decomposition based on topological analysis. The metabolites with higher connectivity are considered as external and connect the formed subnetworks. An automatic decomposition approach based on Petri nets is the so-called *maximal common transition sets* (MCT-sets) [23], and consists on decomposing a network into modules by grouping reactions by participation in the minimal *t-invariants* (equivalent to EFMs). A related approach relies on clustering of *t-invariants* for network modularization [9]. A very recent network coarsening method based on so-called *abstract dependent transition sets* (ADT-sets) is formulated without the requirement of pre-computation of the *t-invariants* and thus may be a promising tool for larger networks [12].

Another problem in genome-scale metabolic modeling is the study of dynamic behavior. Genome-scale metabolic reconstructions are stoichiometric and usually analyzed under steady-state assumption using constraint-based methods, such as flux balance analysis (FBA) [1]. Dynamic flux balance analysis (dFBA) allows variation of external metabolite concentrations, and simulates the network dynamics assuming an internal pseudo steady-state at each time step [16]. It is used in [19] to build a genome-scale dynamic model of *L. lactis* that simulates fermentation profiles. However, this approach gives no insight into intracellular dynamics, neither it integrates reaction kinetics. In [27] the authors build a kinetic genome-scale model of *S. cerevisiae* us-

ing linlog kinetics, where the reference steady-state is calculated using FBA. Some of the elasticity parameters and metabolite concentrations are derived from available kinetic models, while the majority use default values. Using the stoichiometric coefficients as elasticity values is a rough estimation of the influence of the metabolites on the reaction rates. Moreover, no time-course simulation is performed. Mass action stoichiometric simulation (MASS) models are introduced in [14] as a way to incorporate kinetics into stoichiometric reconstructions. Parameters are estimated from metabolomic data. Regulation can be included by incorporating the mechanistic metabolite/enzyme interactions. A limitation of these models is that mass-action kinetics do not reflect the usual non-linearity of enzymatic reactions and the incorporation of regulation leads to a significant increase in network size.

4.3 Methods

The idea of this work is closer to the reduction concepts of [17, 20] than the modularization concepts in [24, 25]. In the latter cases a large model is decomposed into subunits to ease its processing by analyzing the parts individually. Instead, our objective is to facilitate the visualization, analysis and simulation of a large-scale model as a whole by abstracting its components. This reduction is to be attained by reaction lumping in a way that maintains biological meaning and valid application of current analysis and simulation tools. The Michaelis–Menten kinetics is a typical example of abstraction, where the small network of mass-action reactions are lumped into one single reaction.

The overall idea of the model reduction method is depicted in Fig. 4.1. A large-scale stoichiometric model can be structurally reduced into a simplified version that can be more easily analyzed by methods such as MPA. Also, one may infer a kinetic structure to build a dynamic version of the reduced model. Due to the smaller size, a lower number of parameters has to be estimated. The data used for estimation may be experimental data found in the literature, or pseudo-experimental data from dynamic simulations if part of the system has been kinetically characterized.

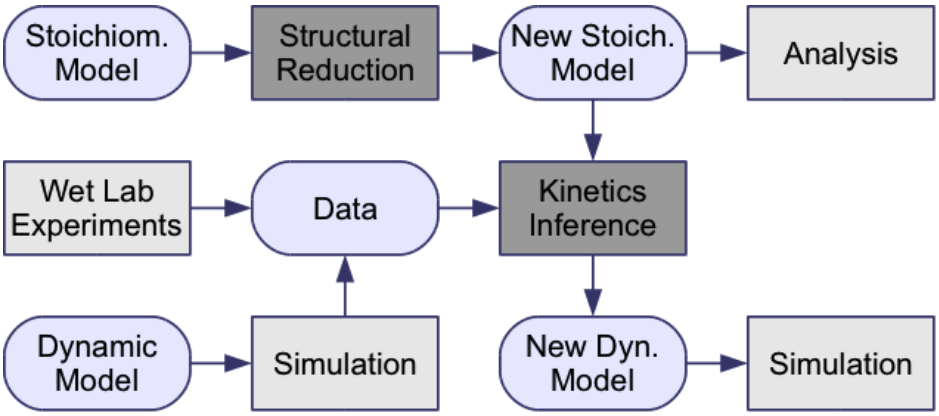

Figure 4.1: Overall concept of model reduction and kinetics inference.

When abstracting a reaction subnetwork into one or more macro-reactions, it is important to consider the assumptions created by such abstraction. As in Michaelis–Menten kinetics, these simplifications result in a pseudo-steady-state assumption for the intermediate species that disappear. While this may not be a problem for flux balance models, it changes the transient behavior of dynamic models because the buffering effect of intermediates in a pathway is neglected. The selection of metabolites to be removed depends on the purpose of the reduction. The network may have different levels of granularity based on the availability of experimental data, topological properties, or simply in order to aggregate pathways according to biological function.

4.3.1 Basic definitions

The proposed method for model reduction uses several Petri net concepts from the literature. We will use the following definition of an unmarked continuous Petri net (adapted from [4]) for modeling a stoichiometric metabolic network:

$$Pn = < P, T, Pre, Post >$$
$$Pre : P \times T \to \mathbb{R}^+$$
$$Post : P \times T \to \mathbb{R}^+$$

where the set of places P represents the metabolites, the set of transitions T represents the reactions and $Pre, Post$ are, respectively, the substrate and product stoichiometries of the reactions. Note that for the representation of a stoichiometric network, a discrete Petri net usually suffices; however, because some models may contain non-integer stoichiometric coefficients, the continuous version was adopted. Moreover, we will assume that reversible reactions are split into irreversible reaction pairs. We will also use the following definitions:

$$\begin{aligned} loc(x) &= \{x\} \cup {}^\bullet x \cup x^\bullet \\ In(p) &= \sum_{t \in {}^\bullet p} Post[p,t] \cdot v(t) \\ Out(p) &= \sum_{t \in p^\bullet} Pre[p,t] \cdot v(t) \end{aligned}$$

where ${}^\bullet x, x^\bullet$ are sets representing the input and output nodes of a node x, the set $loc(x) \subseteq P \cup T$ is called the locality of x, function $v : T \to \mathbb{R}_0^+$ is a given flux distribution (or the so-called instantaneous firing rate), and $In, Out : P \to \mathbb{R}_0^+$ are, respectively, the feeding and draining rates of the metabolites.

The method for network reduction consists of eliminating a set of selected metabolites from the network. For each removed metabolite its surrounding reactions are lumped in order to maintain the fluxes through the pathways. This reduction assumes a steady-state condition for the metabolite, *i.e.* $In(p) = Out(p)$.

4.3.2 Model reduction: Conjunctive fusion

There are two options for lumping the reactions depending on the transformation method applied. The first approach is based on a transformation called *conjunctive transition fusion* [8] and it results in an abstraction that replaces the transition-bordered subnet $loc(p)$ by a single macro-reaction. The drawback of this method is that the flux ratios between the internal reactions are lost. If a known steady-state flux distribution (v) is given,

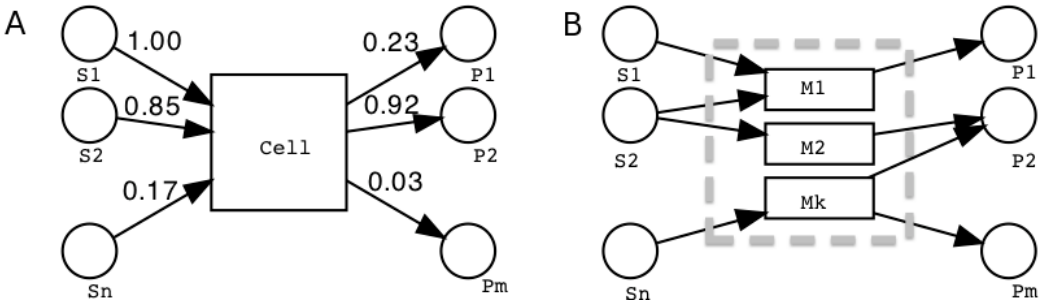

Figure 4.2: Exemplification of limit scenarios where all the internal metabolites are removed. (A) In the conjunctive reduction case the result is one single macro-reaction converting substrates into products with the respective yields specified in the stoichiometry. (B) In the disjunctive reduction method, all possible pathways connecting substrates and products are enumerated.

then the stoichiometric coefficients can be adjusted to preserve the ratios for that distribution; however, the space of solutions of the flux balance formulation becomes restricted to a particular solution. In the limiting case, if all the internal metabolites are removed, the cell is represented by one single macro-reaction connecting extracellular substrates and products with the stoichiometric yields inferred from the network topology for one particular steady-state (Fig 4.2A). The transformation method for removing metabolite p in Pn given a flux distribution v is described as follows:

$$
\begin{aligned}
Pn' &= < P', T', Pre', Post' > \\
P' &= P \setminus \{p\} \\
T' &= T \setminus ({}^\bullet p \cup p^\bullet) \cup \{t_p\} \\
Pre' &= \{(p_i, t_j) \mapsto Pre(p_i, t_j) \mid (p_i, t_j) \in dom(Pre) \setminus (P \times ({}^\bullet p \cup p^\bullet))\} \\
&\quad \cup \{(p_i, t_p) \mapsto f_{in}(p_i) \mid p_i \in {}^\bullet({}^\bullet p \cup p^\bullet), p_i \neq p, v'(t_p) \neq 0, f_{in}(p_i) \neq 0\} \\
Post' &= \{(p_i, t_j) \mapsto Post(p_i, t_j) \mid (p_i, t_j) \in dom(Post) \setminus (P \times ({}^\bullet p \cup p^\bullet))\} \\
&\quad \cup \{(p_i, t_p) \mapsto f_{out}(p_i) \mid p_i \in ({}^\bullet p \cup p^\bullet)^\bullet, p_i \neq p, v'(t_p) \neq 0, f_{out}(p_i) \neq 0\} \\
v' &= \{t \mapsto v(t) \mid t \in T \setminus ({}^\bullet p \cup p^\bullet)\} \cup \{t_p \mapsto In(p)\}.
\end{aligned}
$$

where

$$f_{in}(p_i) = \frac{\sum_{t \in p_i^\bullet \cap ({}^\bullet p \cup p^\bullet)} Pre(p_i, t) \cdot v(t)}{v'(t_p)}$$
$$f_{out}(p_i) = \frac{\sum_{t \in {}^\bullet p_i \cap ({}^\bullet p \cup p^\bullet)} Post(p_i, t) \cdot v(t)}{v'(t_p)}$$

The stoichiometric coefficients of the new reaction may be very high or low, depending on $v'(t_p)$ and so, optionally, one may also normalize them with some scalar λ, such that $Pre''(p_i, t_p) = \frac{1}{\lambda} \cdot Pre'(p_i, t_p)$, $Post''(p_i, t_p) = \frac{1}{\lambda} \cdot Post'(p_i, t_p)$ and $v''(t_p) = \lambda \cdot v'(t_p)$. This will also make the final result independent of the order of the metabolites removed. A good choice for λ is:

$$\lambda = \max\left(\{Pre(p_i, t_p) \mid p_i \in {}^\bullet t_p\} \cup \{Post(p_i, t_p) \mid p_i \in t_p{}^\bullet\}\right)$$

4.3.3 Model reduction: Disjunctive fusion

The second approach is based on a transformation called *disjunctive transition fusion* [8], where every combination of input and output reaction pairs connected by the removed metabolite is replaced by one macro-reaction. Although this approach does not constrain the steady-state solution space of the flux distribution, it has the drawback of increasing the number of transitions, if the metabolite is highly connected, due to the combinatorial procedure. Note that applying this reduction step to metabolite p_i is equivalent to performing one iteration of the *t-invariant* calculation algorithm to remove column i of the transposed incidence matrix. Therefore, in the limiting case where all internal metabolites are removed, the cell is represented by the set of all possible pathways connecting extracellular substrates and products (Fig. 4.2B), as was done in [20]. The definition, similar to the previous one,

is as follows:

$$
\begin{aligned}
Pn' &= < P', T', Pre', Post' > \\
P' &= P \setminus \{p\} \\
T' &= T \setminus ({}^\bullet p \cup p^\bullet) \cup \{t_{xy} \mid (x,y) \in ({}^\bullet p \times p^\bullet)\} \\
Pre' &= \{(p_i, t) \mapsto Pre(p_i, t) \mid (p_i, t) \in dom(Pre) \setminus (P \times ({}^\bullet p \cup p^\bullet)\} \\
&\quad \cup \{(p_i, t_{xy}) \mapsto Pre_0(p_i, x) \cdot Pre(p, y) + Pre_0(p_i, y) \cdot Post(p, x) \\
&\quad\quad \mid (x,y) \in ({}^\bullet p \times p^\bullet), p_i \in {}^\bullet\{x, y\}\} \\
Post' &= \{(p_i, t) \mapsto Post(p_i, t) \mid (p_i, t) \in dom(Post) \setminus (P \times ({}^\bullet p \cup p^\bullet)\} \\
&\quad \cup \{(p_i, t_{xy}) \mapsto Post_0(p_i, x) \cdot Pre(p, y) + Post_0(p_i, y) \cdot Post(p, x) \\
&\quad\quad \mid (x,y) \in ({}^\bullet p \times p^\bullet), p_i \in \{x, y\}^\bullet\}
\end{aligned}
$$

where

$$
Pre_0(p,t) = \begin{cases} Pre(p,t) & \text{if } (p,t) \in dom(Pre) \\ 0 & \text{if } (p,t) \notin dom(Pre) \end{cases}
$$

$$
Post_0(p,t) = \begin{cases} Post(p,t) & \text{if } (p,t) \in dom(Post) \\ 0 & \text{if } (p,t) \notin dom(Post) \end{cases}
$$

Whenever there are pathways with the same net stoichiometry, these can be removed by checking the columns of the incidence (stoichiometric) matrix and eliminating repeats. It should also be noted that in both methods, if a metabolite acts both as substrate and product in a lumped reaction, it will create a redundant cycle that is not reflected in the incidence matrix. If these cycles are not removed, they propagate through the reduction steps; therefore, they should be replaced by a single arc containing the overall

stoichiometry. The procedure works as follows:

$$\begin{aligned}
Pre' =& \{(p,t) \mapsto Pre(p,t) \mid (p,t) \in dom(Pre) \setminus dom(Post)\} \\
& \cup \{(p,t) \mapsto Pre(p,t) - Post(p,t) \\
& \quad \mid (p,t) \in dom(Pre) \cap dom(Post), Pre(p,t) > Post(p,t)\} \\
Post' =& \{(p,t) \mapsto Post(p,t) \mid (p,t) \in dom(Post) \setminus dom(Pre)\} \\
& \cup \{(p,t) \mapsto Post(p,t) - Pre(p,t) \\
& \quad \mid (p,t) \in dom(Pre) \cap dom(Post), Post(p,t) > Pre(p,t)\}
\end{aligned}$$

The previous arc removing procedure may cause isolation of some nodes when $Pre(p,t) = Post(p,t)$; therefore, the isolated nodes should be removed:

$$\begin{aligned}
P' &= \{p \mid p \in P, loc(p) \neq \{p\}\} \\
T' &= \{t \mid t \in T, loc(t) \neq \{t\}\}
\end{aligned}$$

4.3.4 Kinetics inference

Given a stoichiometric model, if metabolomic or fluxomic data are available for parameter estimation, one may try to build a dynamic model by inferring appropriate kinetics for the reactions. In [26] the authors propose that this is performed by assuming linlog kinetics for all reactions using an FBA solution as the reference state and the stoichiometries as elasticity parameters. An integration of Biochemical Systems Theory (BST) with Hybrid Functional Petri Nets (HFPN) is presented in [30], where general mass action (GMA) kinetics is assumed for each transition. The review of kinetic rate formulations is out of the scope of this work and may be found elsewhere [10].

Assuming that all metabolites with unknown concentration were removed, we will extend our definition to a marked continuous Petri net:

$$Pn =< P, T, Pre, Post, m_0 >$$

where $m_0 : P \to \mathbb{R}_0^+$ denotes the initial marking (concentration) of the metabolites. The kinetics inference process consists on defining a firing rate

$v : T \to \mathbb{R}_0^+$, which will be dependent on the current marking (m) and the specific kinetic parameters (see [7] for an introduction on marking-dependent firing rates). As we assumed irreversible reactions, each rate will only vary with substrate concentration. The rates can be easily derived from the net topology. In case of GMA kinetics v is given by:

$$v(t) = k_t \prod_{p \in {}^\bullet t} m(p)^{a_{p,t}}$$

where k_t is the kinetic rate of t and $a_{p,t}$ is the kinetic order of metabolite p in reaction t. A usual first approximation for $a_{p,t}$ is $Pre(p,t)$.

Linlog kinetics are formulated based on a reference rate v_0, and defined by:

$$v(t) = v_0(t) \left(1 + \sum_{p \in {}^\bullet t} \varepsilon^0_{p,t} \ln \left(\frac{m(p)}{m_0(p)} \right) \right)$$

where $\varepsilon^0_{p,t}$ is called the elasticity of metabolite p in reaction t, reflecting the influence of the concentration change of the metabolite in the reference reaction rate. As in the previous case, $Pre(p,t)$ can be chosen as an initial approximation for $\varepsilon^0_{p,t}$. The relative enzyme activity term (e/e_0) commonly present in linlog rate laws to account for regulatory effects at larger time scales will not be considered.

4.4 Results and Discussion

4.4.1 Central carbon metabolism of *E. coli*

The proposed methods were tested using the dynamic central carbon metabolism model of *E. coli* [3], where the stoichiometric part was used for the application of the reduction methods, and the dynamic profile was used to generate pseudo-experimental data sets for parameter estimation and validation of the kinetics inference method. A Petri net representation of this model (Fig. 4.3) was built using the Snoopy tool [22]. All reversible reactions were split into irreversible pairs. The net contains a total of 18 places, 44 transitions and is covered by 95 *semipositive t-invariants*, computed with the

Integrated Net Analyzer [28].

In the application of the conjunctive method (Fig 4.4A), the metabolites were classified as in [17] based on their timescale (Table 4.1), by calculating their turnover time ($\tau : P \rightarrow \mathbb{R}_0^+$) using the reference steady-state of the dynamic model, where:

$$\tau(p) = \frac{m_0(p)}{In(p)}$$

Metabolites with small turnover time are considered fast. In this case, all metabolites except the slowest 5 (*glcex, pep, g6p, pyr, g1p*) were removed.

For the application of the disjunctive method (Fig 4.4B), the metabolites were classified based on their topology (Table 4.1). We conveniently opted to remove the metabolites with lower connectivity to avoid the combinatorial explosion problem. All metabolites except 5 (*g6p, pyr, f6p, gap, xyl5p*) were removed. This reduction assumes steady-state for the removed metabolites. However, it makes no assumptions on the ratios between the fluxes, therefore preserving the flux-balance solution space.

Because we are assuming that the reference steady-state is known, the conjunctive reduced model was chosen for the application of the kinetics inference method assuming linlog kinetics at the reference state. The elasticity parameters were estimated using COPASI [13]. The pseudo-experimental data was generated from simulation with the original model after a 1 mM extracellular glucose pulse with the addition of Gaussian noise ($std = 0.05$ mM) (Fig. 4.5A). The fitted model was then validated using pseudo-experimental data from a 2 mM pulse (Fig. 4.5B). It is possible to observe an instantaneous increase in *pyr* (from 2.67 to 3.93) and an instantaneous decrease *pep* (from 2.69 to 1.26) which the model is unable to reproduce. The poor fitting in some of the intracellular metabolites is expected given the significant reduction to the model. However, the extracellular glucose consumption profile is remarkably good, both in the fitting and validation cases.

Although both reducing methods can be combined with kinetics inference, the conjunctive version seems more suitable if a steady-state distribution is known, because it generates smaller models, hence less parameters. The disjunctive version is appropriate for analyzing all elementary pathways between

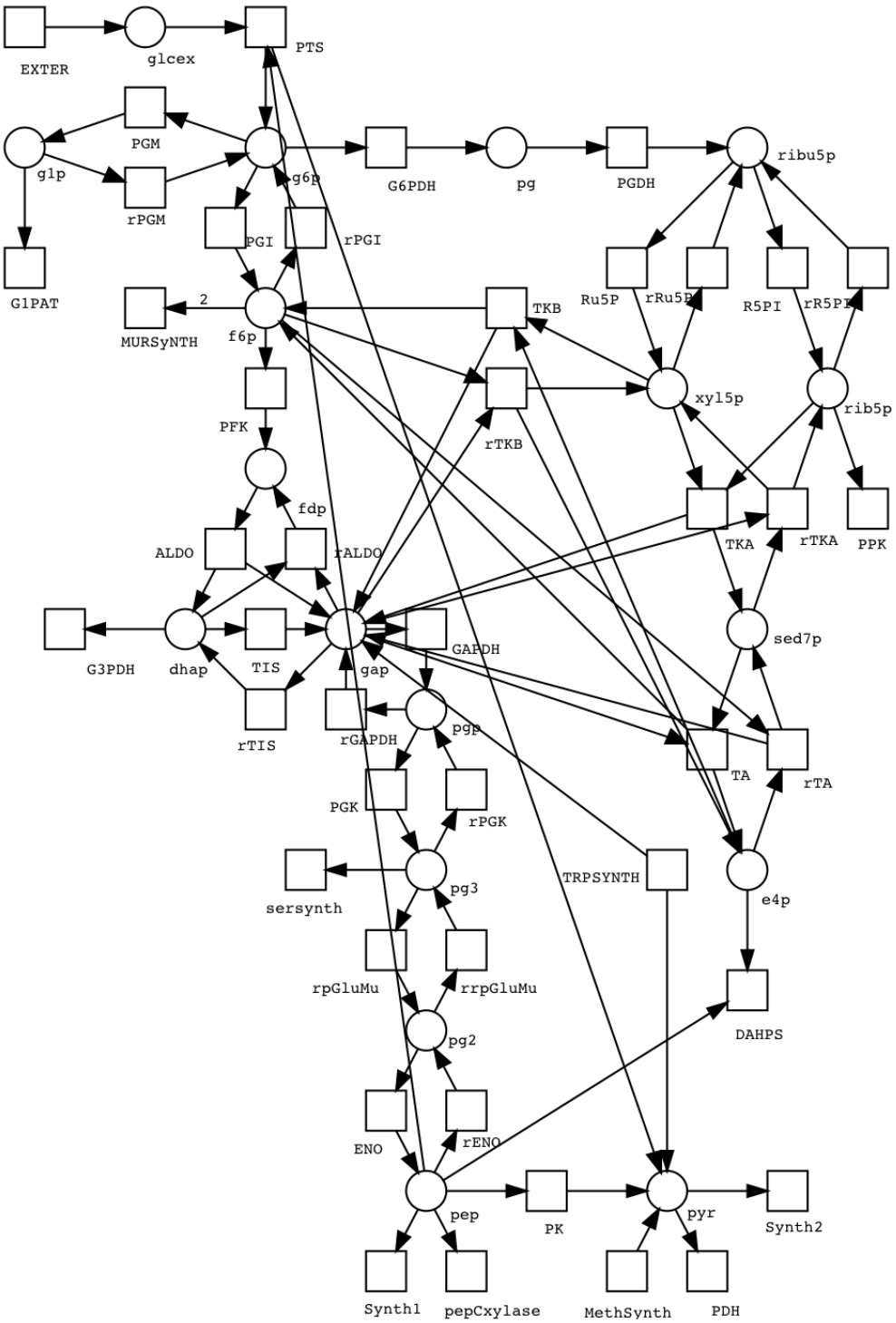

Figure 4.3: Petri net model of the dynamic central carbon metabolism model of *E. coli* with reversible reactions split into irreversible pairs.

Table 4.1: Metabolite topological properties: input reactions, output reactions, connectivity; and dynamic properties: concentration (mM), flux (mM/s), turnover time (s) at the reference steady-state.

Metabolite	$\#({}^\bullet p)$	$\#(p^\bullet)$	$\#({}^\bullet p \times p^\bullet)$	m_0	In	τ
glcex	1	1	1	0.0558	0.0031	**18.099**
pep	1	6	6	2.6859	0.3031	**8.8603**
g6p	3	3	**9**	3.4882	0.2004	**17.406**
pyr	4	2	**8**	2.6710	0.2418	**11.044**
f6p	3	5	**15**	0.6014	0.1423	4.2266
g1p	1	2	2	0.6539	0.0023	**278.62**
pg	1	1	1	0.8092	0.1397	5.7929
fdp	2	1	2	0.2757	0.1414	1.9495
sed7p	2	2	4	0.2761	0.0454	6.0757
gap	7	6	**42**	0.2196	0.3661	0.5997
e4p	2	3	6	0.0986	0.0454	2.1684
xyl5p	3	3	**9**	0.1385	0.0839	1.6503
rib5p	2	3	6	0.3994	0.0558	7.1626
dhap	2	3	6	0.1682	0.1414	1.1892
pgp	2	2	4	0.0080	0.3207	0.0251
pg3	2	3	6	2.1437	0.3207	6.6851
pg2	2	2	4	0.4014	0.3031	1.3241
ribu5p	3	2	6	0.1114	0.1397	0.7974

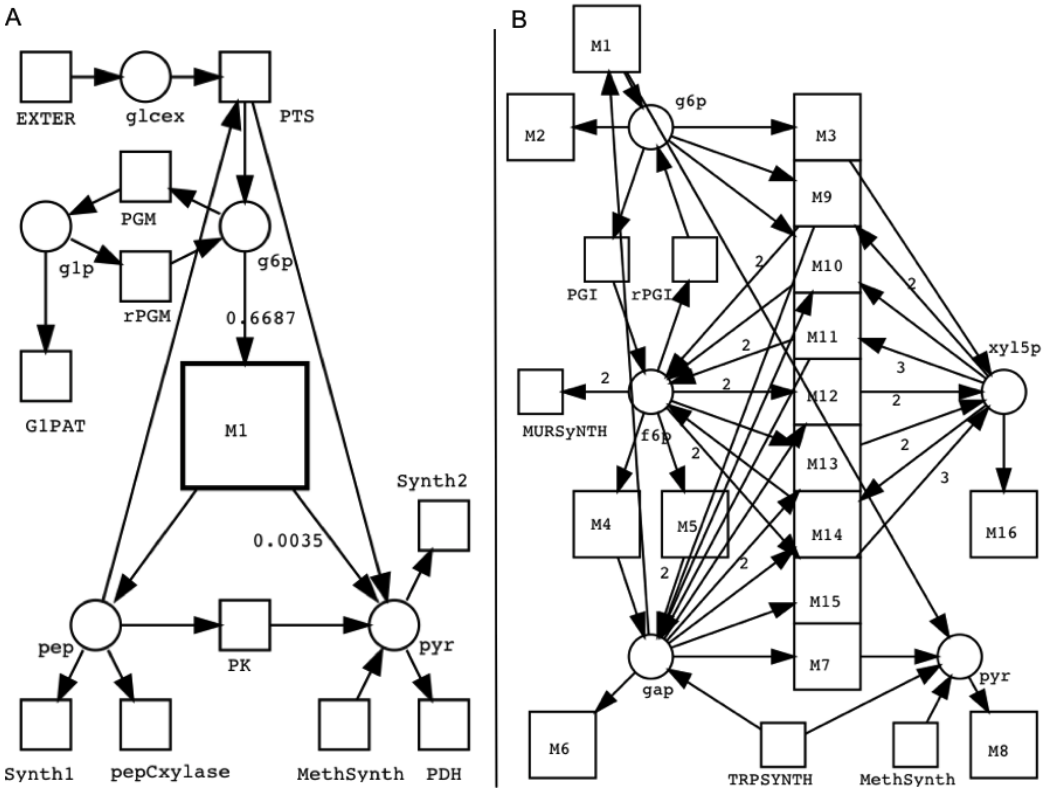

Figure 4.4: Reduced versions of the original network. (A) Conjunctive reduction method. (B) Disjunctive reduction method.

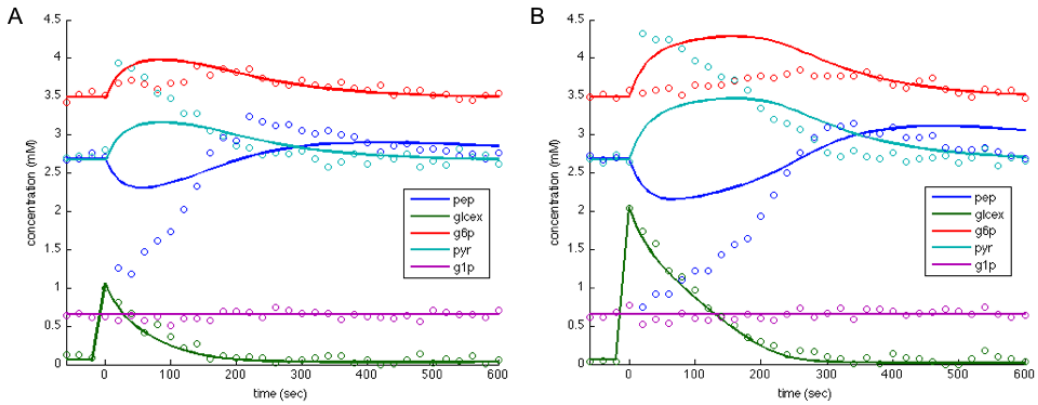

Figure 4.5: (A) Results of parameter estimation with pseudo-experimental data with 1 mM extracellular glucose pulse. (B) Validation of the model with a 2 mM extracellular glucose pulse. In both cases, the circles represent the experimental data and the lines represent time-course simulations generated by the reduced model.

a set of metabolites without the burden of calculating the set of EFMs of the whole model. For instance, the macro-reactions *M4* ($ALDO + G3PDH$) and *M5* ($ALDO + TIS$), with net stoichiometries of, respectively, [$fdp \rightarrow gap$] and [$fdp \rightarrow 2\ gap$], are two unique pathways between these two metabolites.

4.4.2 Transforming a genome-scale model

In order to test the proposed framework at the genome-scale level, we used a genome-scale metabolic model of *E. coli* [21], which includes 625 metabolites and 931 biochemical reactions. Genome-scale models can be simulated using FBA, but the results are hard to visualize graphically, and they do not provide a suitable starting point for inferring kinetic models, due to the complexity of the generated models. However, a reduced version of this model would provide a suitable scaffold for building a dynamic model. The disjunctive method is not feasible at the genome scale due to its combinatorial nature. Therefore, the conjunctive approach is clearly the only option in this case.

As a proof of concept, we created a reduced genome-scale model that preserves all the metabolites and reactions in common with the dynamic central carbon model. The remainder intermediate metabolites are removed, resulting in a lumped version of all other metabolic pathways. Note that some of the reactions on the central carbon model already represent lumped versions of some biosynthetic pathways (*e.g. mursynth, trpsynth, methsynth, sersynth*). However they were not deduced from the genome-scale network and may not be accurate abstractions of these pathways. An FBA simulation of the genome-scale model was performed in order to obtain the reference steady-state flux distribution, where the fluxes of the common reactions were constrained to their reference value in the central carbon model. Figure 4.6 shows the resulting condensed network. The nomenclature and visual layout from the central carbon model was preserved in order to facilitate comparison. It is possible to observe that almost all reactions that are not part of the central carbon metabolism were lumped together with the biomass reaction to form a lumped macro-reaction that reflects the contribution of all internal

metabolites to the biomass formation. This picture also expresses the how the external substrates and products contribute to the overall metabolic activity.

4.5 Conclusions

This work presents strategies for model reduction of metabolic networks based on a Petri net framework. Two approaches, conjunctive and disjunctive reduction are presented. The conjunctive approach allows the abstraction of a subnetwork into one lumped macro-reaction, however limited to one particular flux distribution of the subnetwork. The disjunctive approach on the other hand, makes no assumptions on the flux distribution by replacing the removed subnetwork with macro-reactions for all possible pathways through the subnetwork, therefore not constraining the steady-state solution space. In both cases, the reduced model may be transformed into a dynamic model using kinetics inference and parameter estimation if experimental data is available. Using the reduced model, instead of the original, facilitates this process because it significantly decreases the number of parameters to be estimated.

We have shown how our framework can be applied in the creation of condensed genome-scale metabolic models that preserve the stoichiometry of the original models. In future work, we intend to create a dynamic model based on the generated condensed model. This model can reuse all the information already available in the dynamic central carbon. It will only be necessary to find suitable rate laws and kinetic parameters for the new reactions in the model.

Among the extensions available to Petri nets are the addition of different types of arcs, such as read-arcs and inhibitor-arcs, which could be use to represent activation and inhibition in biochemical reactions. They could also be used to integrate metabolic and regulatory networks. Optimization in metabolic processes is usually based on knockout simulations in metabolic networks. However, these simulations do not take into consideration the possible regulatory effects caused by the knockouts. In our transformation methods we removed the arcs with the same stoichiometry in both directions,

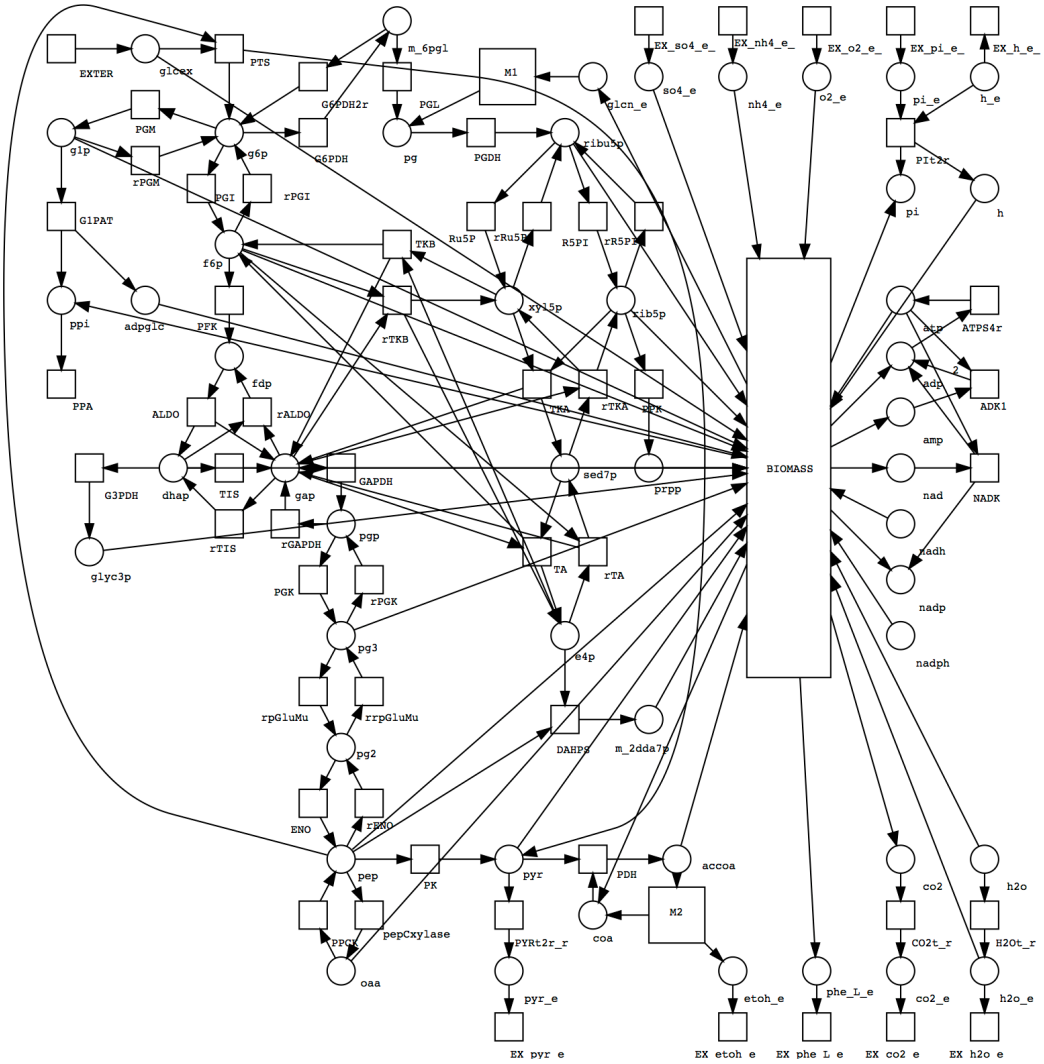

Figure 4.6: Condensed version of the genome-scale metabolic model of *E. coli*, extending the previous example of the central carbon model. To avoid saturating the image, the participation of the cofactors in most of the reactions is not represented. The stoichiometric coefficients of the biomass reaction are also not displayed. The reaction is the following: glcex + 6.21 pep + 29.3 pyr + 0.72 f6p + 0.218 g1p + 0.277 sed7p + 0.0769 rib5p + 0.03 dhap + 38.2 pg3 + 0.462 ribu5p + 669 atp + 251 nadph + 238 nadh + 1.88 glyc3p + 10.3 prpp + 5.45 2dda7p + 2.56 so4(e) + 59.1 accoa + 64.8 h2o + 1.69 adpglc + 1.95e+03 h + 25.3 oaa + 236 o2(e) + 118 nh4(e) + h2o(e) $\rightarrow$ 0.522 gap + 655 adp + 15 amp + 251 nadp + 238 nad + 59.1 coa + glcn(e) + 35.2 ppi + 90.7 co2 + 0.934 phe_L(e) + 2.1e+03 h(e) + 687 pi .

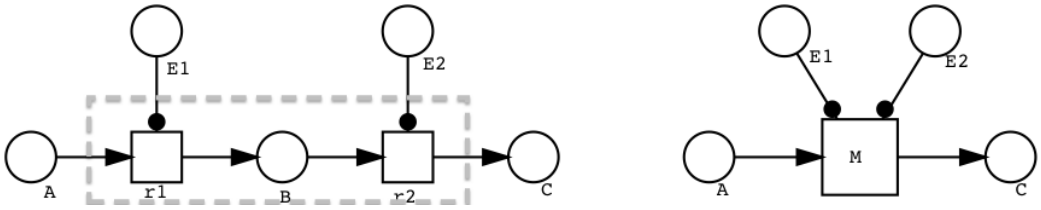

Figure 4.7: Reduction step conserving the read-arcs associated with the enzymes of the original reactions.

because these are not reflected in the stoichiometric matrix. In the Michaelis–Menten example this results in removing the enzyme from the network. The proposed methods can be extended to consider read-arcs for these situations, which should be preserved during the reduction steps, therefore establishing connection places to the integration of a regulatory network (Fig 4.7).

An alternative to the reduction of the models would be to consider their representation using hierarchical Petri nets. In this case, each macro-reaction would be connected to its detailed subnetwork. Although this would not reduce the number of kinetic parameters in the case of kinetics inference, it would be extremely useful for facilitated modeling and visualization of large-scale networks without compromising detail. It could also be the solution for genome-scale pathway analysis, if it is performed independently at each hierarchical level. The hierarchical model composition proposed for SBML [6] may facilitate the implementation of this alternative. See [11] for an automatic network coarsening algorithm based on hierarchical petri nets applied to different kinds of biological networks.

Bibliography

[1] 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. *Nature Protocols*, 2(3):727–738, 2007.

[2] A.P. Burgard, P. Pharkya, and C.D. Maranas. Optknock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and Bioengineering*, 84(6):647–657, 2003.

[3] C. Chassagnole, N. Noisommit-Rizzi, J.W. Schmid, K. Mauch, and M. Reuss. Dynamic modeling of the central carbon metabolism of Escherichia coli. *Biotechnology and Bioengineering*, 79(1):53–73, 2002.

[4] R. David and H. Alla. *Discrete, continuous, and hybrid Petri nets*. Springer Verlag, 2005.

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

[6] A. Finney. Developing SBML beyond level 2: proposals for development. In *Computational Methods in Systems Biology*, pages 242–247. Springer, 2005.

[7] D. Gilbert and M. Heiner. From Petri nets to differential equations-an integrative approach for biochemical network analysis. *Petri Nets and Other Models of Concurrency-ICATPN 2006*, pages 181–200, 2006.

[8] C. Girault and R. Valk. *Petri Nets for System Engineering: A Guide to Modeling, Verification, and Applications*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2001.

[9] E. Grafahrend-Belau, F. Schreiber, M. Heiner, A. Sackmann, B.H. Junker, S. Grunwald, A. Speer, K. Winder, and I. Koch. Modularization of biochemical networks based on classification of Petri net t-invariants. *BMC bioinformatics*, 9(1):90, 2008.

[10] J.J. Heijnen. Approximative kinetic formats used in metabolic network modeling. *Biotechnology and bioengineering*, 91(5):534–545, 2005.

[11] M. Heiner. Understanding Network Behavior by Structured Representations of Transition Invariants. *Algorithmic Bioprocesses*, page 367, 2009.

[12] M. Heiner and K. Sriram. Structural analysis to determine the core of hypoxia response network. *PloS one*, 5(1):e8600, 2010.

[13] S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, and U. Kummer. COPASI—a COmplex PAthway SImulator. *Bioinformatics*, 22(24):3067–3074, 2006.

[14] N. Jamshidi and B.Ø. Palsson. Mass action stoichiometric simulation models: Incorporating kinetics and regulation into stoichiometric models. *Biophysical Journal*, 98:175–185, 2010.

[15] D. Machado, R.S. Costa, M. Rocha, I. Rocha, B. Tidor, and E.C. Ferreira. A critical review on modelling formalisms and simulation tools in computational biosystems. In *Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living*, pages 1063–1070. Springer, 2009.

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

[17] I.E. Nikerel, W.A. van Winden, P.J.T. Verheijen, and J.J. Heijnen. Model reduction and a priori kinetic parameter identifiability analysis using metabolome time series for metabolic reaction networks with linlog kinetics. *Metabolic Engineering*, 11(1):20–30, 2009.

[18] D. Noble. The rise of computational biology. *Nature Reviews Molecular Cell Biology*, 3(6):459–463, 2002.

[19] G.M. Oddone, D.A. Mills, and D.E. Block. A dynamic, genome-scale flux model of Lactococcus lactis to increase specific recombinant protein expression. *Metabolic Engineering*, 2009.

[20] A. Provost and G. Bastin. Dynamic metabolic modelling under the balanced growth condition. *Journal of Process Control*, 14(7):717–728, 2004.

[21] J.L. Reed, T.D. Vo, C.H. Schilling, and B.O. Palsson. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biology*, 4(9):R54, 2003.

[22] C. Rohr, W. Marwan, and M. Heiner. Snoopy-a unifying Petri net framework to investigate biomolecular networks. *Bioinformatics*, 2010.

[23] A. Sackmann, M. Heiner, and I. Koch. Application of Petri net based analysis techniques to signal transduction pathways. *BMC bioinformatics*, 7(1):482, 2006.

[24] C.H. Schilling and B.Ø. Palsson. Assessment of the metabolic capabilities of *Haemophilus influenzae Rd* through a genome-scale pathway analysis. *Journal of Theoretical Biology*, 203(3):249–283, 2000.

[25] S. Schuster, T. Pfeiffer, F. Moldenhauer, I. Koch, and T. Dandekar. Exploring the pathway structure of metabolism: decomposition into sub-

networks and application to *Mycoplasma pneumoniae*. *Bioinformatics*, 18(2):351–361, 2002.

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

[27] K. Smallbone, E. Simeonidis, N. Swainston, and P. Mendes. Towards a genome-scale kinetic model of cellular metabolism. *BMC Systems Biology*, 4(1):6, 2010.

[28] P.H. Starke. INA: Integrated Net Analyzer. *Reference Manual*, 1992.

[29] G. Stephanopoulos. Metabolic engineering. *Biotechnology and Bioengineering*, 58, 1998.

[30] J. Wu and E. Voit. Hybrid modeling in biochemical systems theory by means of functional Petri nets. *Journal of bioinformatics and computational biology*, 7(1):107, 2009.

Chapter 5

Accounting for Enzymatic Regulation

This chapter is based on the article *"Accounting for enzymatic regulation in large-scale kinetic reconstructions of metabolism"* (in preparation).

Abstract

The current limitations in constraint-based models and mechanistic kinetic models of metabolism are leading to new approaches for building kinetic models at the genome-scale. These models are built by combining constraint-based models and approximative kinetic formats. However, they lack the effects of enzymatic regulation, as it is not accounted for in the underlying network topology. In this work, we propose the utilization of Extended Petri nets as scaffold for the generation of large-scale kinetic models in order to account for enzymatic regulation during the kinetic inference process. We generate kinetic models for the central carbon metabolism of *E. coli*, with and without enzymatic regulation. We then evaluate the impact of accounting for this kind of regulation in metabolic reconstructions by performing several knockouts and changes in enzyme expression levels and comparing the results with those generated with the original dynamic model of this pathway. Our results show that accounting for enzymatic regulation has an influence on the determination of the steady-state flux distribution of mutants, and allows the prediction of changes that would otherwise be unforeseen. We conclude that the inclusion of enzymatic regulation in metabolic reconstructions is an important step that can be performed if flexible model representations like Petri nets are applied, and it can be used to reveal new manipulation targets for strain optimization

.

5.1 Introduction

During the past years, Systems Biology has pushed the frontier of our understanding of the complex phenomena of life, with the creation of mathematical and computational models of the cell [24]. The predictive capability provided by these models is fundamental for the improvement of several areas such as biomedical research and industrial biotechnology. In particular, the field of Metabolic Engineering [51], takes advantage of mathematical models of cellular metabolism, in order to discover optimal sets of genetic manipulations for the design of mutant microbial strains that efficiently produce compounds of industrial interest [4, 34].

One of the current approaches for modeling metabolic pathways is the development of mechanistic kinetic models based on systems of ordinary differential equations (ODEs). This kind of models includes detailed kinetic rate laws that describe the details of the enzymatic mechanisms and usually contain several kinetic parameters. Due to the large amount of experimental data required to estimate these parameters, this modeling approach has been limited to central pathways of well-studied organisms, such as *E. coli* [8] and *S. cerevisiae* [40]. On the other hand, constraint-based modeling is an alternative approach that describes the admissible steady-state flux distributions in terms of stoichiometric and thermodynamic constraints. Given the simplicity in the formulation, and the fact that no kinetic parameters are required to instantiate this kind of models, they have been applied in most genome-scale metabolic reconstructions [39, 36].

Although constraint-based models have a clear advantage in terms of scalability when compared to kinetic models, they contain several limitations. Besides not describing intracellular transient behavior, they do not take into account metabolite concentration and enzymatic regulation effects. To overcome the limitations of both approaches, a more recent approach is emerging [49, 50, 22]. It consists on the automatic generation of approximative kinetic models, using constraint-based models as a scaffold. Approximative

kinetic formats abstract from the enzymatic mechanism details, hence they usually require fewer parameters than mechanistic rate laws [19]. Therefore, this kind of models can scale to larger metabolic networks when compared to mechanistic models. Also, they determine a single solution rather than a space of solutions, and they are able to integrate high-throughput data from several *omics* (proteomics, fluxomics and metabolomics).

The rate of a reaction can be controlled at the gene regulatory level by controlling the amount of enzyme that is produced (transcriptional regulation), and also by regulating enzyme activity through metabolic activators and inhibitors (enzymatic regulation). Gene regulatory networks operate at a larger time-scale than metabolic networks, therefore enzyme concentration is usually considered constant in metabolic models. However, enzymatic regulation is part of the metabolism itself and it is responsible for the regulation of many metabolic pathways. A current limitation in the process of generating kinetic models from constraint-based models is the fact that the later do not express the enzymatic regulation relationships between metabolites and enzymes.

Petri nets are a graphical and mathematical formalism that have been applied in the modeling of several biological pathways, including metabolic [38, 27, 57, 26], gene regulatory [6, 7], and signaling [43, 9, 3, 18]. They are very similar to the constraint-based formulation of metabolic models, as they both determine the topology of the network in terms of consumption and production of metabolites. Extended Petri nets are extensions to the original formalism that include special types of arcs that model the effect of components that participate in a process without being consumed or produced [10, 1].

In this work we propose the utilization of Extended Petri nets as a scaffold for the kinetic inference process, in order to build large-scale kinetic reconstructions that account for enzymatic regulation. We test the potential of this approach by generating a kinetic model of the central carbon metabolism of *E. coli* and evaluating the impact of enzymatic regulation in the prediction of mutant phenotypes by comparing with the simulation results obtained from the available dynamic model for this organism that as been experimentally

validated [8]. See Fig. 5.1 for an overview of our proposed procedure.

5.2 Methods

5.2.1 Central carbon metabolism model of *E. coli*

In order to evaluate the predictive capabilities of the models automatically generated by kinetic inference, we used a published and validated dynamic model of the central carbon metabolism of *E. coli* as reference [8]. The model is available in SBML format [21] at the Biomodels database [29]. It contains a total of 18 metabolites and 31 reactions, including several enzymatic reactions, one exchange reaction and a few lumped versions of biosynthetic pathways, holding a total of 125 kinetic parameters. We did not consider metabolite dilution and the contribution of cofactors both to the topology and dynamics of the network.

The model's topology was used to build the topology of the Petri net models used in this work. The enzymatic regulation effects present in the kinetic equations were used to define the topology of the regulatory interactions in the Extended Petri net model. The original model was also used to generate pseudo-experimental data, both for parameter estimation and validation of results.

5.2.2 Petri net models

Petri nets are bipartite graphs with two types of nodes, *places* and *transitions* that, within the biochemical context, respectively represent substances and reactions. Arcs between places and transitions define consumption and production relationships. The notation to define Petri nets can vary among the literature. We will adopt the following definition (adapted from the definition of an unmarked generalized Petri net of [15]). A Petri net (Pn) is a

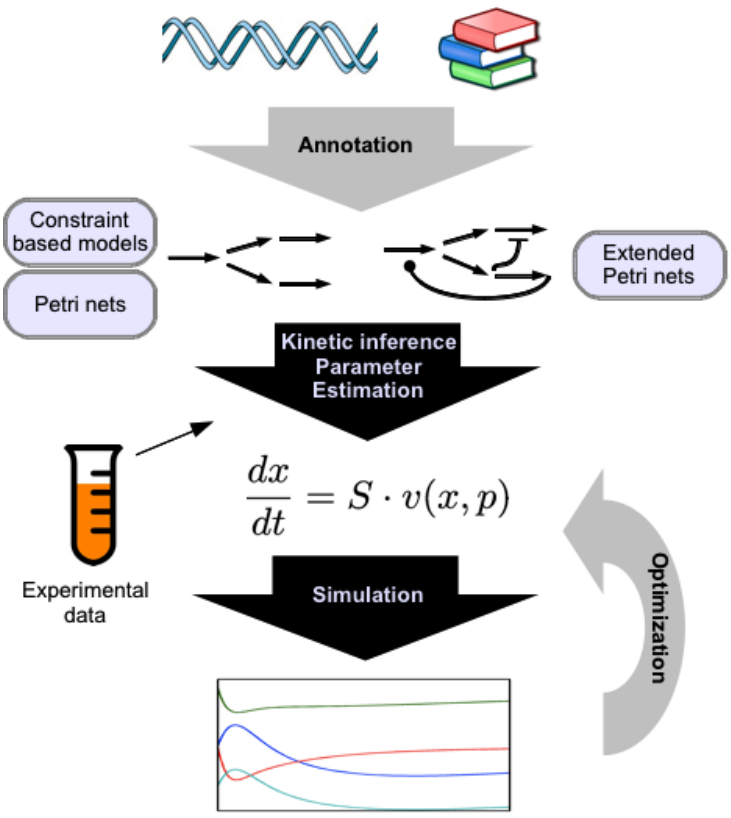

Figure 5.1: Overview of the main steps of a metabolic model reconstruction process: Genome annotation and literature data are used to reconstruct metabolic networks; using Extended Petri nets as a scaffold, rather than simple Petri nets or constraint-based models, allows the inclusion of enzymatic regulation; the kinetic inference uses the network's topology to build a kinetic model, using experimental data for parameter estimation; the kinetic models can be used for time-course and steady-state simulation; optimization methods can be used to find optimal targets for rational strain design. This work focuses only on the kinetic inference and simulation steps (dark arrows).

4-tuple:

$$Pn = < P, T, Pre, Post >$$
$$Pre : P \times T \rightarrow \mathbb{N}$$
$$Post : P \times T \rightarrow \mathbb{N}$$

where P is the set of places, T is the set of transitions, Pre and $Post$ are the arcs representing, respectively, substrate and product stoichiometries.

We created a Petri net model based on the topology of the original dynamic model. The model is able to account for the stoichiometric information that is present in the original model. However, because transitions are unidirectional, the reversible reactions are decomposed into irreversible reaction pairs. This model is essentially equivalent to a constraint-based model (without flux capacity constraints).

Extended Petri nets are extensions to the original Petri net formalism, that include special types of arcs, such as *activator* (also know as *read* or *test*) and *inhibitor* arcs [10, 1]. In the biochemical context these arcs represent regulatory interactions that modulate enzymatic activity. We will use the following definition of an Extended Petri net (En):

$$En = < P, T, Pre, Post, A, I >$$
$$Pre : P \times T \rightarrow \mathbb{N}$$
$$Post : P \times T \rightarrow \mathbb{N}$$
$$A : P \times T \rightarrow \mathbb{N}$$
$$I : P \times T \rightarrow \mathbb{N}$$

where $P, T, Pre, Post$ have the same meaning as in the previous case, and A, I respectively represent activator and inhibitor arcs. We will use the following

notation:

$$
\begin{aligned}
{}^{\bullet}r &= \{p \mid p \in P, (p,t) \in dom(Pre)\} \\
t^{\bullet} &= \{p \mid p \in P, (p,t) \in dom(Post)\} \\
{}^{\uparrow}t &= \{p \mid p \in P, (p,t) \in dom(A)\} \\
{}^{\downarrow}t &= \{p \mid p \in P, (p,t) \in dom(I)\}
\end{aligned}
$$

to respectively represent the sets of substrates, products, activators and inhibitors of a reaction modeled by transition t.

Similarly to the previous case, we created an Extended Petri net model based on the original dynamic model. However, in this case, the model not only accounts for the stoichiometry but also for the enzymatic regulation relationships that exist in the original model. Figure 5.2 shows a graphical representation of the model, built with the Snoopy Petri net editor [41]. It is possible to observe a regulatory layer that is not considered in the simple Petri net model (or equivalently, in a constraint-based model).

5.2.3 Kinetic inference

The kinetic inference process consists on the generation of kinetic rate laws for the reactions in the model and the instantiation of the kinetic parameters and the initial metabolite concentrations. Approximative kinetic formats facilitate this process as they do not require insight into the enzymatic mechanism details. Commonly used formats include generalized mass action (GMA) [20], lin-log [55], and convenience kinetics [31]. For a review in this topic see [19].

Within the Petri net framework, the kinetic inference process can be performed by a transformation from the discrete Petri net model into a continuous Petri net. We will adopt the following definition of a continuous Petri net (adapted from the definition of a marked continuous Petri net of [15]).

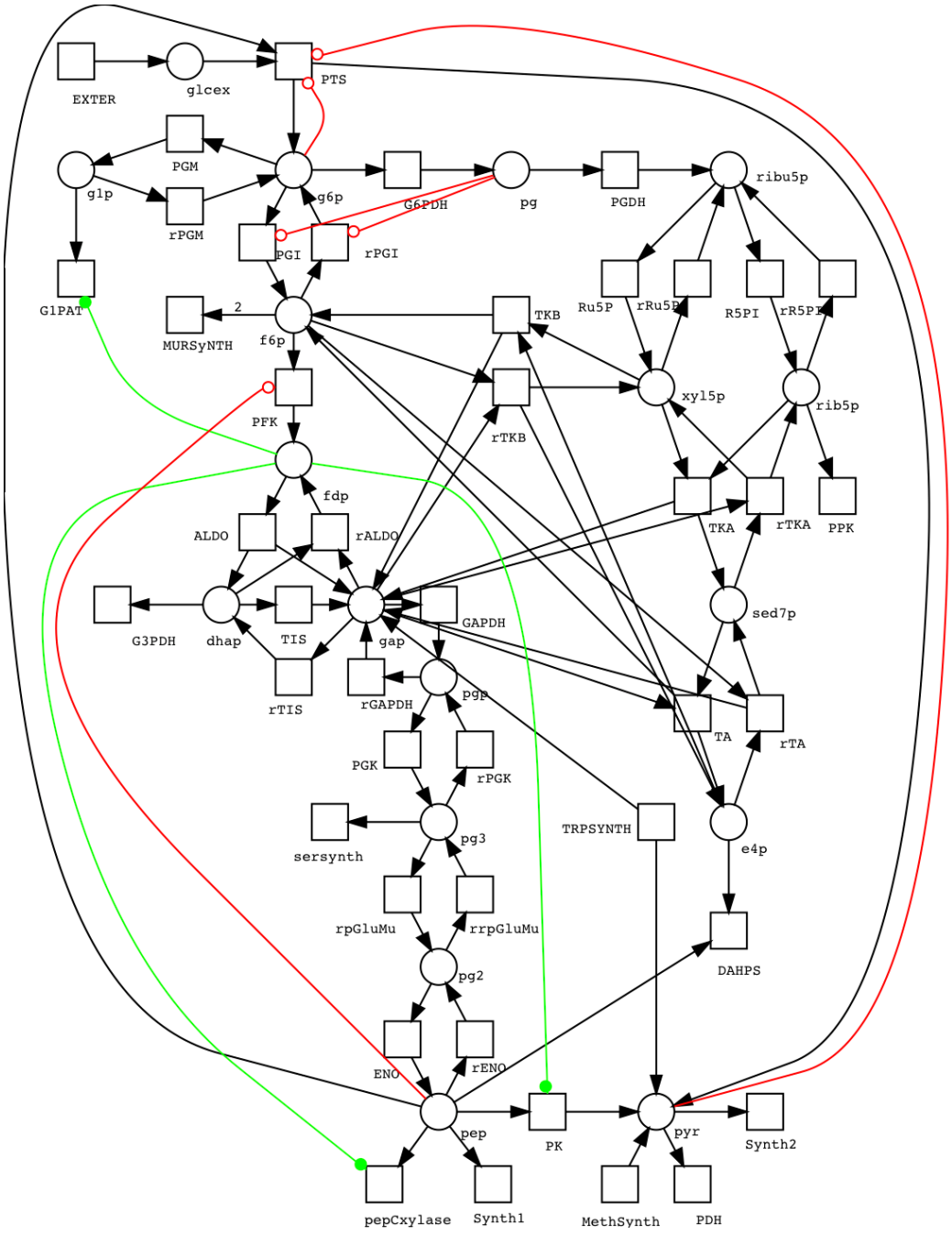

Figure 5.2: Petri net representation of the central carbon metabolism model of *E. coli* built with the Snoopy Petri net editor [41]. Reversible reactions are decomposed into irreversible pairs. Edges with arrows represent production and consumption of metabolites. Edges with closed circle (red) and edges with open circle (green), respectively represent activation and inhibition.

A continuous Petri net (Cn) is a 6-tuple:

$$\begin{aligned} Cn &= < P, T, Pre, Post, m_0, v > \\ Pre &: P \times T \to \mathbb{R}^+ \\ Post &: P \times T \to \mathbb{R}^+ \\ m_0 &: P \to \mathbb{R}^+ \\ v &: T \to \mathbb{R}^+ \end{aligned}$$

where $P, T, Pre, Post$ represents the network topology of the discrete Petri net, m_0 is the initial metabolite concentration and v is the *firing rate* function, which is defined according to the approximative kinetic format chosen. In this case, we have adopted the GMA format:

$$v(i) = k_i \prod_{j \in {}^\bullet i} \left(m(j)^{f_{j,i}} \right)$$

where k_i is the kinetic rate of reaction i and $f_{j,i}$ is the kinetic order of metabolite j in reaction i. The Extended Petri net version can also be mapped into a continuous version, by adopting the following definition:

$$\begin{aligned} Cn &= < P, T, Pre, Post, A, I, m_0, v > \\ Pre &: P \times T \to \mathbb{R}^+ \\ Post &: P \times T \to \mathbb{R}^+ \\ A &: P \times T \to \mathbb{R} \\ I &: P \times T \to \mathbb{R} \\ m_0 &: P \to \mathbb{R}^+ \\ v &: T \to \mathbb{R}^+ \end{aligned}$$

where all the elements have the same meaning as in the previous case. However, in this case the firing rate function needs to take into consideration the regulatory effects of the activator and inhibitor arcs. We adopted the

multiplication by a regulatory factor as suggested in [44]:

$$v(i) = k_i \prod_{j \in \bullet i} \left(m(j)^{f_{j,i}}\right) \prod_{j \in \uparrow i} \left(\frac{m(j)}{K_{A,j} + m(j)}\right) \prod_{j \in \downarrow i} \left(\frac{K_{I,j}}{K_{I,j} + m(j)}\right)$$

where $K_{A,j}$ and $K_{I,j}$ respectively represent activation and inhibition constants. Note that when $K_I \to \infty$ (never inhibited) and $K_A \to 0$ (always active) the regulatory model is equivalent to the non-regulatory model.

5.2.4 Parameter estimation

After generating the kinetic models it is necessary to estimate the values of the kinetic parameters. We used the original dynamic model to generate steady-state fluxomic and metabolomic data. For the reversible reactions, the flux was decomposed into forward and reverse rates. These data were used to estimate the kinetic parameters in both models. For the kinetic order parameters it is common to assume the stoichiometry of the metabolite as default value [49, 12]. Therefore, in the non-regulatory model, given the steady-state fluxes (v_{ss}) and metabolite concentrations (m_{ss}), we can estimate the kinetic rate constants for each reaction i as :

$$k_i = \frac{v_{\text{ss}}(i)}{\prod_{j \in \bullet i} \left(m(j)^{Pre(j,i)}\right)}.$$

The case is different for the regulatory model, as there can be many parameters per equation. We opted to set a default value of 1 mM for all regulatory constants to be in the same order of magnitude as the metabolite concentrations. Again, the kinetic constants can be calculated from the given data:

$$k_i = \frac{v_{\text{ss}}(i)}{\prod_{j \in \bullet i} \left(m(j)^{Pre(j,i)}\right) \prod_{j \in \uparrow i} \left(\frac{m(j)}{K_{A,j}+m(j)}\right) \prod_{j \in \downarrow i} \left(\frac{K_{I,j}}{K_{I,j}+m(j)}\right)}.$$

5.2.5 Simulation

Our models are generated using our code and stored in SBML format. Note that in SBML there is no distinction between activation and inhibition as all

regulators are simply referred to as *modifiers*. We circumvented this problem by manually adding annotations to the reactions with this information, which are then recognized by our kinetic generation code. This problem is also discussed in [16] where the authors suggest the utilization of Systems Biology Ontology (SBO) [28] to annotate the models.

For simulation purposes, the SBML files are converted to Matlab files, and we use Matlab's *ode15s* function to integrate the generated system of ordinary differential equations (ODEs). The integration is performed in the time range of zero to infinity, with additional conditions to stop the computation when the system reaches a steady-state or to abort when it exceeds a given CPU utilization time limit. Knockouts and changes in enzyme expression levels are simulated by premultiplying the respective equations with a factor (e/e_0). After simulation, the fluxes of the decomposed reversible reactions are combined in order to facilitate comparison with the original model.

5.3 Results

We evaluated the predictive capability of the generated models to determine mutant phenotypes, by performing single knockouts of the enzymes in the model. We tested both the regulatory and non-regulatory models and compared with the pseudo-experimental data generated with the original model. We also tested under and over-expression of enzyme concentration levels by 5-fold decrease and increase, respectively. Table 5.1 shows the results of the simulations given by the error (ε) of the steady-state flux distribution of the mutant (v_{m}) compared to the pseudo-experimental data generated with the original model under the same perturbation (v_{m}^*), normalized by the wild-type flux distribution (v_{w}^*):

$$\varepsilon = \frac{||v_{\mathrm{m}} - v_{\mathrm{m}}^*||}{||v_{\mathrm{w}}^*||}.$$

In both models it is possible to observe a low error for most mutations. However, in the cases with higher errors, there is no clear dominance of any

Table 5.1: Normalized errors of the simulations performed with the non-regulatory (-r) and regulatory (+r) models, for all single enzyme perturbations including knockout (KO), under-expression (UE) and over-expression (OE). In some cases (*) the simulation could not reach a steady-state.

Enzyme	KO (-r)	KO (+r)	UE (-r)	UE (+r)	OE (-r)	OE (+r)
PGI	0.010	0.010	0.000	0.000	0.000	0.000
PGM	0.003	0.003	0.001	0.001	0.000	0.000
G6PDH	0.024	0.024	0.019	0.022	1.112	1.080
PFK	0.018	0.021	0.478	0.361	0.230	0.219
TA	(*)	(*)	0.005	0.004	0.001	0.001
TKA	0.379	0.816	0.006	0.006	0.001	0.001
TKB	(*)	0.048	0.004	0.002	0.001	0.000
ALDO	(*)	(*)	0.003	0.024	0.000	0.006
GAPDH	0.005	0.005	0.080	0.098	0.037	0.203
TIS	0.036	0.100	0.002	0.020	0.000	0.005
G3PDH	0.002	0.001	0.002	0.001	0.009	0.005
PGK	0.005	0.005	0.002	0.002	0.000	0.000
PGluMu	0.005	0.005	0.001	0.001	0.000	0.000
ENO	0.005	0.005	0.001	0.001	0.000	0.000
PK	0.049	0.029	0.041	0.021	0.134	0.059
PEPC	0.095	0.035	0.066	0.023	0.092	0.053
DAHPS	0.022	0.008	0.017	0.006	0.044	0.033
PDH	1.011	1.010	0.147	0.147	0.060	0.060
PGDH	(*)	(*)	0.000	0.000	0.000	0.000
R5PI	(*)	(*)	0.010	0.008	0.002	0.002
Ru5P	0.377	0.823	0.012	0.010	0.002	0.002
PPK	0.012	0.012	0.010	0.009	0.035	0.020
G1PAT	0.003	0.003	0.002	0.002	0.011	0.009

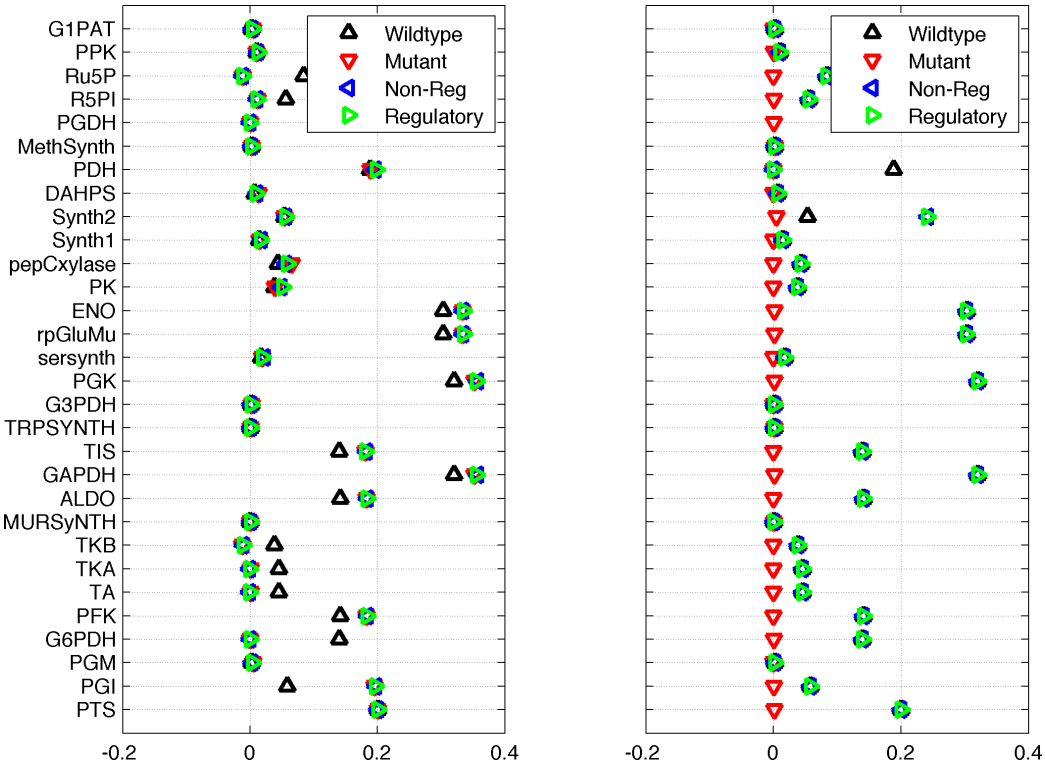

Figure 5.3: Comparison of the steady-state flux distributions (mM/s) of the generated models ("Non-Reg", "Regulatory") with the pseudo-experimental data ("Wildtype", "Mutant") after knockout of: G6PDH (left side), PDH (right side).

of the models over the other.

In some of the knockout mutations the simulations were not able to reach a steady-state. This happened in both models with the exception of the knockout of *TKB*, where the simulation was able to reach a steady-state in the regulatory model but not in the non-regulatory one. In Figures 5.3–5.5 it is possible to observe some cases that will be analyzed in more detail.

The knockout of *G6PDH* was predicted with great accuracy by both models (Fig. 5.3). In this case, the flux from *G6PDH* is redirected through *PGI*. It is possible to observe an inversion in the fluxes of *Ru5P* and *TKB*, which were correctly predicted in both cases.

The knockout of *PDH* resulted in a blocked metabolism that does not carry any flux (Fig. 5.3). This is a consequence of an accumulation of pyru-

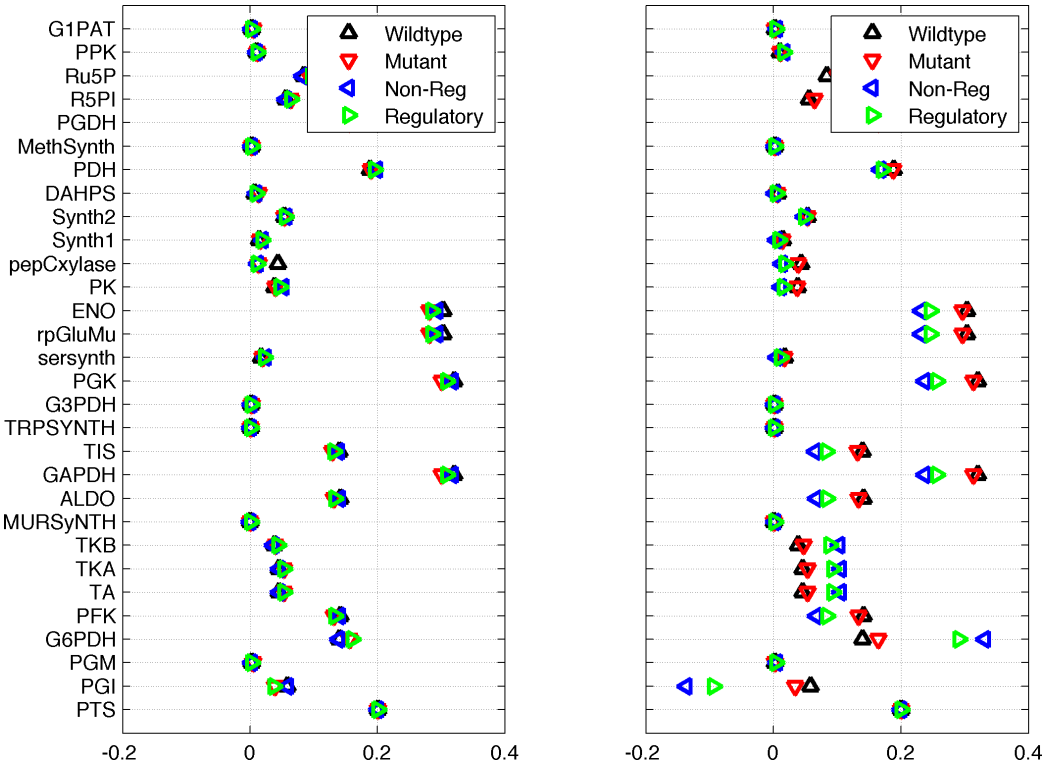

Figure 5.4: Comparison of the steady-state flux distributions (mM/s) of the generated models ("Non-Reg", "Regulatory") with the pseudo-experimental data ("Wildtype", "Mutant") after under-expression (0.2×) of: PEPC (left side), PFK (right side).

vate which causes the inhibition of *PTS*. This inhibition is only temporary but is however sufficient to allow a depletion of *pep* by drain reactions, breaking the cycle where it re-enters in *PTS*. We would expect this consequence to be predicted by the regulatory model. However, due to the fact that we have adopted GMA kinetics, the model does not account for the saturation of *Synth2*, which was able to completely consume the excess of pyruvate.

The under-expression of *PEPC* caused a redirection of a small part of the flux from glycolysis to the pentose-phosphate pathway (Fig. 5.4). This was correctly predicted by the regulatory model but not by the non-regulatory model. The reason is that the accumulation of *pep* causes the inhibition of *PFK* and consequently a decrease in the whole glycolytic pathway, favoring the pentose-phosphate pathway.

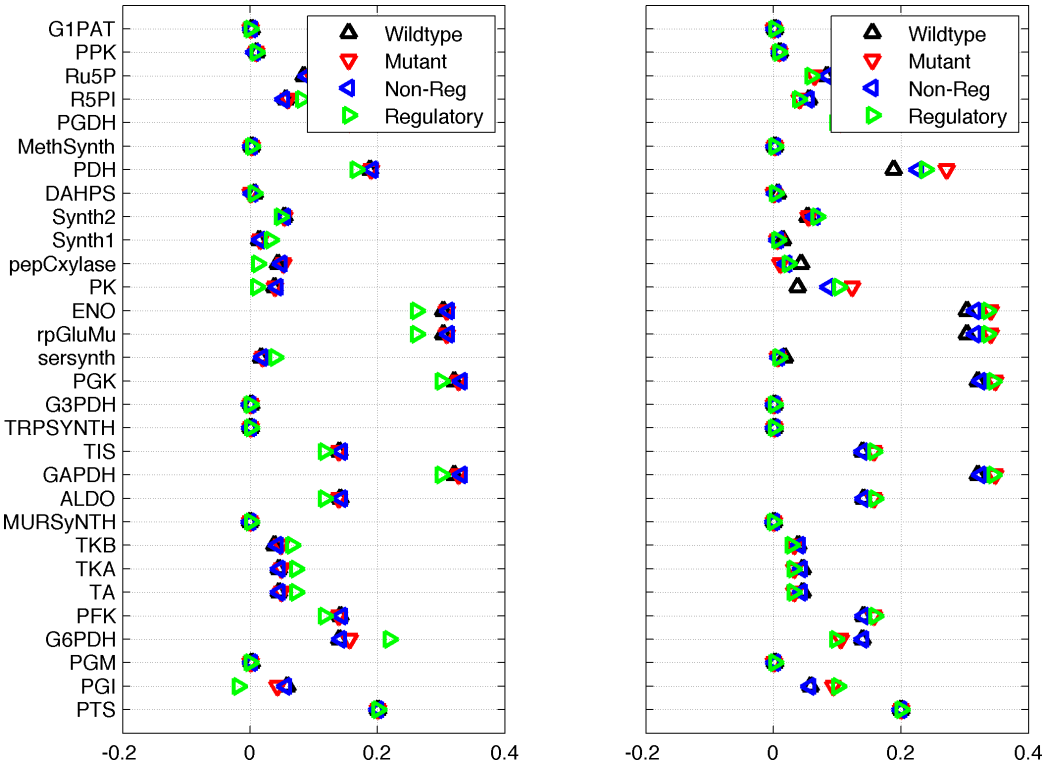

Figure 5.5: Comparison of the steady-state flux distributions (mM/s) of the generated models ("Non-Reg", "Regulatory") with the pseudo-experimental data ("Wildtype", "Mutant") after over-expression (5×) of: GAPDH (left side), PK (right side).

The under-expression of *PFK* has an effect similar to the previous case (Fig. 5.4). However, in this case both models predict an exaggerated response to the redirection of the flux, with an inversion of *PGI* coupled with a substantial flux increase in all the reactions of the pentose-phosphate pathway.

The over-expression of *GAPDH* was one of the cases where the regulatory model had a larger error than the non-regulatory model (Fig. 5.5). In this case there was a small shift of the flux from glycolysis to the pentose-phosphate pathway. The non-regulatory model did not account for this shift, whereas the regulatory model had an exaggerated response and incorrectly predicted an inversion of *PGI*.

The over-expression of *PK* resulted in a shift of the flux from the pentose-

phosphate pathway to the glycolytic pathway (Fig. 5.5). One of the causes for this change is that *PK* was one of the bottlenecks in this pathway. In this case the regulatory model predicted the resulting steady-state in good agreement with the pseudo-experimental data, whereas the non-regulatory model failed to predict these changes.

5.4 Discussion

5.4.1 Advantages of kinetic modeling

The construction of kinetic models of metabolism has been limited by the amount of experimental data required to describe in detail the enzymatic mechanisms and to estimate their respective kinetic parameters. Instead, constraint-based modeling has become the *de facto* framework for modeling the metabolism at the genome scale. However, there are several limitations with this approach, such as not accounting for metabolite concentrations, not defining a single steady-state solution and not describing transient behavior. The increasing availability of *omics* data (proteomics, metabolomics, fluxomics) is driving the development of new approaches for kinetic modeling at the genome scale.

Mass action stoichiometric simulation (MASS) models were introduced in [22]. It is a modeling approach that consists on adding mass action kinetics to a stoichiometric network and using metabolomic and fluxomic data to estimate the kinetic parameters. The method was tested on a stoichiometric reconstruction of the human red blood cell model. The authors also evaluate the impact of enzymatic regulation by adding the regulatory interactions to the model. However, this requires the decomposition of each enzymatic mechanism into its elementary steps, which more than doubled the size of the network. This decomposition step is avoided in our approach by the introduction of regulatory arcs.

In [50] the authors built a genome-scale kinetic model of *S. cerevisiae*. Using the available constraint-based as a scaffold, they generated a kinetic model using lin-log kinetics. The reference steady-state flux distribution is

found with the constraint-based model using flux balance analysis and the stoichiometric coefficients are used as estimates of the elasticity parameters. Since this kinetic model is based on the underlying constraint-based model, it does not account for enzymatic regulation.

There are advantages in using kinetic rather than constraint-based models for steady-state simulation. First, they define an unique steady-state solution. In this manner, it is possible to calculate the steady-state phenotype of mutants without requiring any extra assumptions, which is the case in constraint-based methods such as flux balance analysis (FBA) [54], minimization of metabolic adjustment (MOMA) [46] and regulatory on/off minimization (ROOM) [47]. Also, it is possible to account for the enzyme expression levels in a straightforward manner. In previous work we analyzed how the inclusion of proteomic data in kinetic models can improve the simulation of knockout strains [11]. There are some approaches to account for enzyme expression levels in simulations of constraint-based models, but their formulation is not straightforward and usually requires the introduction of new assumptions or the inclusion of a considerable amount of experimental data [35, 56, 53].

5.4.2 The effects of enzymatic regulation

In this work we propose a new approach for large-scale kinetic reconstruction that accounts for enzymatic regulation. We evaluate the impact of accounting for this kind of regulation in metabolic reconstructions by generating kinetic models with and without regulation and comparing the results with those obtained from the original dynamic model of the central carbon metabolism of *E. coli* [8]. In particular, we evaluate the effect of enzymatic regulation in the determination of the steady-state flux distribution of mutant strains by performing enzyme knockouts and changes in expression levels.

Our results show that enzymatic regulation has an influence on the steady-state flux distribution obtained with the kinetic model. We can observe that there is not a clear gain in the precision of the predictions. This is most likely due to the fact the regulatory model has more parameters than the non-

regulatory model, which were given a default value rather than estimated. Nonetheless, it is possible to observe that in some cases there were changes in the flux distribution which could not be predicted without accounting for the regulatory effects (*e.g.*: under-expression of *PEPC*, over-expression of *PK*).

Along with simulation, metabolic models are also used for optimization purposes. Optimization algorithms search for optimal modification targets in order to improve the production of compounds of industrial interest [4, 34]. The inclusion of enzymatic regulation in kinetic reconstructions can reveal new optimization targets. This idea is also explored in [33] where the author proposes an optimization method for kinetic models that not only adjusts the enzyme expression levels but also the regulatory parameters. The method is applied in a model of the central carbon metabolism of *E. coli* in order to improve glucose uptake and serine synthesis.

5.4.3 Limitations and directions for improvement

In order to improve the simulation results it is important to find an appropriate combination of rate laws and strategies for parameter estimation. Since this is not the main focus of the work, we adopted GMA kinetics which is one of the most simple formats. Default parameter values were used wherever possible, and the rest were fitted to the steady-state flux distribution of the original model.

One interesting remark is that the simulation results were reasonably accurate in most cases although we have used default values for several parameters. This indicates that although these models may not be able to accurately simulate the transient behavior of the system, they are good enough for steady-state prediction.

For the selection of the rate laws, convenience kinetics were also considered. They have a semi-mechanistic format, which gives a closer description to the enzymatic mechanisms and accounts for enzyme saturation [31]. However, preliminary tests showed no evidence of significant improvement when compared to GMA kinetics. This is likely due to the fact that it uses a higher number of parameters and requires a more sophisticate strategy for parame-

ter estimation. The lin-log format is also a strong candidate [55]. It is based on a reference steady-state, and has the advantage that only the elasticity coefficients need to be estimated if the wild-type steady-state is given. However, it is important to keep in mind that increasing the complexity of the rate laws not only increases the number of parameters but also the computational cost of the simulations performed. This is an fundamental aspect to consider in order to apply this approach at the genome scale.

Regarding parameter estimation, one hypothesis is to use time-course fluxomic and metabolomic data if available, to find better estimates for the activation and inhibition constants. It is also possible to search these parameters in databases such as BRENDA [45] and SABIO-RK [42] to use as initial estimates. However, the computational cost of time-course fitting strategies becomes problematic for large models.

Steady-state parameter fitting with fluxomic and metabolomic data may become mandatory at the genome scale. However, since there are typically many parameters per equation, there is an underdetermined solution space for the parameters. The most suitable values can be found through non-linear optimization. The objective functions and conditions should rely on generalized biological principles. For instance, it has been observed that substrate concentrations are typically above the K_M values to optimize enzyme efficiency [2], and that the enzyme concentrations are optimally distributed along the metabolic pathways [25]. The stability and robustness of the system at the solution point should also be a required condition in the formulation of the problem [5, 33].

Methods such as the ensemble modeling approach [52] may also be suitable in these cases. Rather than fitting particular parameter values, we may sample the parameter space, and iteratively refine the ensemble by comparing the simulation results with given experimental data.

5.5 Conclusions

The creation of kinetic models at the genome scale is an important step towards the whole-cell simulation goals of Systems Biology. Several recent

efforts in this field are also addressing the integration of metabolic, regulatory and signaling networks in order to account for the control that these exert over the cellular metabolism [13, 48, 14, 30]. This work addressed the integration of another layer of control which comes from enzymatic regulation. In fact, it has been shown that this kind of regulation precedes transcriptional regulation in situations where an immediate response is required [37].

Accounting for enzymatic regulation in large-scale kinetic reconstructions of metabolism yields more accurate descriptions of metabolic networks by expressing interactions that would otherwise be unforeseen. Moreover, it can reveal new sets of targets for strain optimization in biotechnological applications.

The computational cost of performing simulations with kinetic models of increasingly larger sizes may become a bottleneck in this kind of approaches. Model reduction strategies will have an important role to solve this issue. In previous work we suggested an approach for structural network reduction prior to the kinetic inference process, in order to reduce the network size, and consequently, the complexity of the generated models [32].

In this work we used a dynamic model of the central carbon metabolism of *E. coli* as a case-study in order to validate our approach. In the future, this will be applied to larger network sizes for which dynamic model reconstructions are not available. However, the latest genome-scale metabolic reconstructions [17] do not account for the enzymatic regulation interactions. Therefore, it will be necessary to annotate these models with such information, which can be obtained from databases such as EcoCyc [23] and BRENDA [45], before generating the kinetic models.

Bibliography

[1] P. Baldan, N. Busi, A. Corradini, and GM Pinna. Domain and event structure semantics for petri nets with read and inhibitor arcs. *Theoretical Computer Science*, 323(1-3):129–189, 2004.

[2] B.D. Bennett, E.H. Kimball, M. Gao, R. Osterhout, S.J. Van Dien, and J.D. Rabinowitz. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nature Chemical Biology*, 5(8):593–599, 2009.

[3] R. Breitling, D. Gilbert, M. Heiner, and R. Orton. A structured approach for the engineering of biochemical network models, illustrated for signalling pathways. *Briefings in Bioinformatics*, 9(5):404–421, 2008.

[4] A.P. Burgard, P. Pharkya, and C.D. Maranas. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and Bioengineering*, 84(6):647–657, 2003.

[5] Y.J. Chang and N.V. Sahinidis. Optimization of metabolic pathways under stability considerations. *Computers & Chemical Engineering*, 29(3):467–479, 2005.

[6] C. Chaouiya, E. Remy, P. Ruet, and D. Thieffry. Qualitative modelling of genetic networks: From logical regulatory graphs to standard petri nets. *Applications and Theory of Petri Nets*, 3099:137–156, 2004.

[7] C. Chaouiya, E. Remy, and D. Thieffry. Petri net modelling of biological regulatory networks. *Journal of Discrete Algorithms*, 6(2):165–177, 2008.

[8] C. Chassagnole, N. Noisommit-Rizzi, J.W. Schmid, K. Mauch, and M. Reuss. Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnology and Bioengineering*, 79(1):53–73, 2002.

[9] L. Chen, G. Qi-Wei, M. Nakata, H. Matsuno, and S. Miyano. Modelling and simulation of signal transductions in an apoptosis pathway by using timed Petri nets. *Journal of Biosciences*, 32(1):113–127, 2007.

[10] S. Christensen and N. Hansen. Coloured petri nets extended with place capacities, test arcs and inhibitor arcs. *Application and Theory of Petri Nets*, 691:186–205, 1993.

[11] R.S. Costa, D. Machado, E.C. Ferreira, and I. Rocha. Evaluating the integration of proteomic data for the prediction of intracellular fluxes after knockout experiments. In *11th International Symposium on Computer Applications in Biotechnology (CAB 2010)*, pages 114–119, 2010.

[12] R.S. Costa, D. Machado, I. Rocha, and E.C. Ferreira. Hybrid dynamic modeling of *Escherichia coli* central metabolic network combining Michaelis-Menten and approximate kinetic equations. *BioSystems*, 100(2):150–157, 2010.

[13] M.W. Covert and B.Ø. Palsson. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *Journal of Biological Chemistry*, 277(31):28058–28064, 2002.

[14] 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.

[15] R. David and H. Alla. *Discrete, Continuous, and Hybrid Petri Nets*. Springer Verlag, 2004.

[16] A. Dräger, A. Schröder, and A. Zell. Automating mathematical modeling of biochemical reaction networks. *Systems Biology for Signaling Networks*, pages 159–205, 2010.

[17] A.M. Feist, C.S. Henry, J.L. Reed, M. Krummenacker, A.R. Joyce, P.D. Karp, L.J. Broadbelt, V. Hatzimanikatis, and B.Ø. Palsson. A genome-scale metabolic reconstruction for escherichia coli k-12 mg1655 that accounts for 1260 orfs and thermodynamic information. *Molecular Systems Biology*, 3, 2007.

[18] S. Hardy and P.N. Robillard. Petri net-based method for the analysis of the dynamics of signal propagation in signaling pathways. *Bioinformatics*, 24(2):209–217, 2008.

[19] J.J. Heijnen. Approximative kinetic formats used in metabolic network modeling. *Biotechnology and bioengineering*, 91(5):534–545, 2005.

[20] F. Horn and R. Jackson. General mass action kinetics. *Archive for Rational Mechanics and Analysis*, 47(2):81–116, 1972.

[21] M. Hucka, A. Finney, HM Sauro, H. Bolouri, JC Doyle, H. Kitano, AP Arkin, BJ Bornstein, D. Bray, A. Cornish-Bowden, et al. The Systems Biology Markup Language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, 2003.

[22] N. Jamshidi and B.O. Palsson. Mass action stoichiometric simulation models: Incorporating kinetics and regulation into stoichiometric models. *Biophysical Journal*, 98:175–185, 2010.

[23] I.M. Keseler, C. Bonavides-Martínez, J. Collado-Vides, S. Gama-Castro, R.P. Gunsalus, D.A. Johnson, M. Krummenacker, L.M. Nolan, S. Paley, I.T. Paulsen, et al. Ecocyc: a comprehensive view of escherichia coli biology. *Nucleic Acids Research*, 37(suppl 1):D464, 2009.

[24] H. Kitano. Systems Biology: A Brief Overview. *Science*, 295(5560):1662–1664, 2002.

[25] E. Klipp and R. Heinrich. Competition for enzymes in metabolic pathways: Implications for optimal distributions of enzyme concentrations and for the distribution of flux control. *BioSystems*, 54:1–14, 1999.

[26] I. Koch, B.H. Junker, and M. Heiner. Application of Petri net theory for modelling and validation of the sucrose breakdown pathway in the potato tuber. *Bioinformatics*, 21(7):1219–1226, 2005.

[27] R. Küffner, R. Zimmer, and T. Lengauer. Pathway analysis in metabolic databases via differential metabolic display (DMD). *Bioinformatics*, 16(9):825–836, 2000.

[28] N. Le Novère. Model storage, exchange and integration. *BMC neuroscience*, 7(Suppl 1):S11, 2006.

[29] N. Le Novere, B. Bornstein, A. Broicher, M. Courtot, M. Donizelli, H. Dharuri, L. Li, H. Sauro, M. Schilstra, B. Shapiro, et al. BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Research*, 34(suppl 1):D689, 2006.

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

[31] W. Liebermeister and E. Klipp. Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. *Theoretical Biology and Medical Modelling*, 3(1):1–13, 2006.

[32] D. Machado, R.S. Costa, M. Rocha, I. Rocha, B. Tidor, and E.C. Ferreira. Model transformation of metabolic networks using a petri net based framework. In *International Workshop on Biological Processes & Petri Nets (BioPPN)*, pages 101–115, 2010.

[33] E.V. Nikolaev. The elucidation of metabolic pathways and their improvements using stable optimization of large-scale kinetic models of cellular systems. *Metabolic engineering*, 12(1):26–38, 2010.

[34] K.R. Patil, I. Rocha, J. Forster, and J. Nielsen. Evolutionary programming as a platform for in silico metabolic engineering. *BMC Bioinformatics*, 6(1):308, 2005.

[35] P. Pharkya and C.D. Maranas. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. *Metabolic engineering*, 8(1):1–13, 2006.

[36] N.D. Price, J.A. Papin, C.H. Schilling, and B.O. Palsson. Genome-scale microbial in silico models: the constraints-based approach. *Trends in Biotechnology*, 21(4):162–169, 2003.

[37] M. Ralser, M.M.C. Wamelink, S. Latkolik, E.E.W. Jansen, H. Lehrach, and C. Jakobs. Metabolic reconfiguration precedes transcriptional regulation in the antioxidant response. *Nature Biotechnology*, 27(7):604–605, 2009.

[38] V.N. Reddy, M.L. Mavrovouniotis, and M.N. Liebman. Petri Net Representations in Metabolic Pathways. In *Proceedings of the 1st International Conference on Intelligent Systems for Molecular Biology*, pages 328–336. AAAI Press, 1993.

[39] J.L. Reed and B.O. Palsson. Thirteen years of building constraint-based in silico models of *Escherichia coli*. *Journal of Bacteriology*, 185(9):2692–2699, 2003.

[40] M. Rizzi, M. Baltes, U. Theobald, and M. Reuss. In vivo analysis of metabolic dynamics in *Saccharomyces cerevisiae*: II. mathematical model. *Biotechnology and Bioengineering*, 55(4):592–608, 1997.

[41] C. Rohr, W. Marwan, and M. Heiner. Snoopy-a unifying Petri net framework to investigate biomolecular networks. *Bioinformatics*, 2010.

[42] I. Rojas, M. Golebiewski, R. Kania, O. Krebs, S. Mir, A. Weidemann, and U. Wittig. SABIO-RK: a database for biochemical reactions and their kinetics. *BMC Systems Biology*, 1(Suppl 1):S6, 2007.

[43] A. Sackmann, M. Heiner, and I. Koch. Application of Petri net based analysis techniques to signal transduction pathways. *BMC Bioinformatics*, 7(1):482, 2006.

[44] R. Schauer et al. Quasi-steady-state approximation in the mathematical modeling of biochemical reaction networks. *Mathematical biosciences*, 65(2):155–170, 1983.

[45] I. Schomburg, A. Chang, and D. Schomburg. BRENDA, enzyme data and metabolic information. *Nucleic acids research*, 30(1):47–49, 2002.

[46] D. Segrè, D. Vitkup, and G.M. Church. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(23):15112–15117, 2002.

[47] T. Shlomi, O. Berkman, and E. Ruppin. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proceedings of the National Academy of Sciences of the United States of America*, 102(21):7695–7700, 2005.

[48] T. Shlomi, Y. Eisenberg, R. Sharan, and E. Ruppin. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Molecular Systems Biology*, 3(1), 2007.

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

[50] K. Smallbone, E. Simeonidis, N. Swainston, and P. Mendes. Towards a genome-scale kinetic model of cellular metabolism. *BMC Systems Biology*, 4(1):6, 2010.

[51] G. Stephanopoulos. Metabolic engineering. *Biotechnology and Bioengineering*, 58(2-3):119–120, 1998.

[52] L.M. Tran, M.L. Rizk, and J.C. Liao. Ensemble modeling of metabolic networks. *Biophysical journal*, 95(12):5606–5617, 2008.

[53] R.J.P. van Berlo, D. de Ridder, J.M. Daran, P.A.S. Daran-Lapujade, B. Teusink, and M.J.T. Reinders. Predicting metabolic fluxes using gene expression differences as constraints. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 8(1):206–216, 2011.

[54] A. Varma and B.O. Palsson. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Applied and Environmental Microbiology*, 60(10):3724–3731, 1994.

[55] D. Visser and J.J. Heijnen. Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics. *Metabolic Engineering*, 5(3):164–176, 2003.

[56] 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, 2010.

[57] I. Zevedei-Oancea and S. Schuster. Topological analysis of metabolic networks based on Petri net theory. *In Silico Biology*, 3(3):323–345, 2003.

Chapter 6

Conclusions

The work developed during this thesis addressed the creation of a solid modeling framework for metabolic networks that will facilitate the integration with other kinds of networks, namely gene regulatory and signaling. The bacterium *Escherichia coli* was used as our case-study as it is the most well characterized model organism and an important microbe for biotechnological production processes [4].

Ideally, one would immediately attempt to integrate the available gene regulatory and metabolic reconstructions. However, there is no well-established framework for this integration so far. Current integration approaches are based on constraint-based methods which present several shortcomings [2, 8]. Gene regulatory networks are usually modeled with boolean networks, and the integration with constraint-based metabolic models involves translating these rules into constraints to be imposed in the solution space [5, 11]. However, such translation may not be trivial or intuitive.

A very thorough review on the modeling formalisms that have been used in Systems Biology shows that Petri nets are a suitable candidate for the creation of a modeling framework that supports all kinds of biological networks. It is a graphical and mathematically sound formalism, therefore simultaneously intuitive and highly expressive. The integration of gene regulatory and metabolic networks can be attained by "gluing" together both networks with activation and inhibition arcs, rather than adapting the regulatory rules to

a reaction-based scheme.

There is a current separation of approaches in the modeling of metabolic networks. On one hand, there are dynamic model reconstructions of central pathways that include fully detailed and parameterized kinetic equations [1]. On the other hand, there are genome-scale models based on more abstract representations that only account for stoichiometry and reversibility constraints [3]. We explored the gap between both formulations for the same metabolic network, and concluded that it is possible to take advantage of the availability of dynamic models, even if incomplete, to refine the solution space of constraint-based models.

The modeling trade-off between size and detail is also present in gene regulatory networks, where genome-scale models with abstract representation (boolean) and small kinetic models coexist separately [7]. Therefore the previous study may also be applied to this kind of networks, and consequently to integrated regulatory and metabolic networks.

Given the limitations associated with the constraint-based modeling approach, different authors have proposed the automatic generation of kinetic models based on stoichiometric reconstructions [12, 13, 6]. Considering that stoichiometric models and dynamic models are essentially equivalent to discrete and continuous Petri nets respectively, we implemented the kinetic inference process within a Petri net framework as a transformation from discrete to continuous Petri nets. The implementation of this procedure within the same formalism provides an intuitive and straightforward approach.

One of the consequences of applying this procedure to genome-scale models is the complexity of the generated models which comprise hundreds or even thousands of kinetic equations and parameters. The size of metabolic networks is also well known to be problematic when performing metabolic pathway analysis as the number of possible pathways suffers from combinatorial explosion given a large number of reactions [10]. In order to reduce the complexity of metabolic networks we implemented network transformation methods within this framework that merge together biochemical reactions by eliminating intermediate metabolites. These transformations can be applied prior to the kinetic inference process, hence reducing the complexity

and number of parameters in the generated dynamic models.

The top-down construction of kinetic models based on the available genome-scale stoichiometric reconstructions is a promising approach for genome-scale kinetic modeling. It allows for a straightforward incorporation of high-throughput data from several *omics*, including proteomics, metabolomics and fluxomics. However, the utilization of constraint-based models as a scaffold for this process results in the generation of kinetic models that do not account for enzymatic regulation. This kind of regulation is used by the metabolism for self-control and acts on a faster time-scale when compared to transcriptional regulation. We proposed the utilization of Extended Petri nets, that include activation and inhibition arcs, as a suitable formalism to model this kind of regulation, and as a better scaffold for the kinetic inference process.

Accounting for enzymatic regulation results in more realistic metabolic models, and may reveal new manipulation targets for rational strain design. Although there has been much focus in the integration of transcriptional regulation into metabolic networks [2, 8, 11], it seemed more urgent to address the integration of enzymatic regulation, has it is part of the metabolism itself. Moreover, it has been shown that the control imposed by this kind of regulation precedes transcriptional regulation in adaptive responses [9]. Nevertheless, the approach used to incorporate enzymatic regulation should support the integration of transcriptional regulation with minor adaptation efforts.

As more heterogeneous high-throughput data becomes available, the demand for a whole-cell simulation framework increases. From a metabolic engineering perspective, the integration of gene regulatory and metabolic networks is an important step for the creation of better models for rational strain design. Building genome-scale kinetic models presents several advantages compared to the popular constraint-based approach, such as accounting for enzymatic regulation and intracellular dynamics, and determining unique steady-state phenotypes. It remains unclear how to integrate gene regulatory networks with such models. One strong possibility is to similarly generate kinetic regulatory models from their boolean representation, or alternatively, to consider hybrid network representations that account for discrete and con-

tinuous nodes. In either case, Petri nets (with all their extensions) present a solid framework for such purpose.

The top-down and bottom-up approaches for building kinetic models may converge by replacing the central pathways of the automatically generated models with available dynamic reconstructions of these pathways. To cope with the size and complexity of the generated models, network modularity and hierarchy concepts will become even more important. The methods here developed for lumping selected modules into macroscopic reactions can play an important role, specially if one takes advantage of the well characterized central pathways to create genome-scale models with a detailed core complemented with lumped versions of secondary pathways.

Two aspects of our framework have not yet been explored in detail and will require further emphasis in the future. The first concerns the selection of the ideal set of intermediary nodes to be eliminated in the model reduction methods. Considering that this step is performed prior to kinetic inference, the criteria for selection should be based on topological properties. Further studies could elucidate how the structural changes of the network reflect in its dynamic properties. The second aspect involves the parameter estimation after the kinetic inference process. The large number of parameters in genome-scale kinetic reconstructions, and the genome-scale *omics* data available mainly at steady-state, will require new methods for parameter estimation rather than the traditional time-course fitting. Working with sets of parameter samples instead of unique values, and iteratively refining these sets of samples, as in the ensemble modeling approach [14], may be a possible direction to address this issue.

Bibliography

[1] C. Chassagnole, N. Noisommit-Rizzi, J.W. Schmid, K. Mauch, and M. Reuss. Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnology and Bioengineering*, 79(1):53–73, 2002.

[2] 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.

[3] A.M. Feist, C.S. Henry, J.L. Reed, M. Krummenacker, A.R. Joyce, P.D. Karp, L.J. Broadbelt, V. Hatzimanikatis, and B.Ø. Palsson. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*, 3(121), 2007.

[4] A.M. Feist and B.Ø. Palsson. The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nature Biotechnology*, 26(6):659–667, 2008.

[5] E.P. Gianchandani, J.A. Papin, N.D. Price, A.R. Joyce, and B.O. Palsson. Matrix Formalism to Describe Functional States of Transcriptional Regulatory Systems. *PLoS Computational Biology*, 2(8):e101, 2006.

[6] N. Jamshidi and B.O. Palsson. Mass action stoichiometric simulation models: Incorporating kinetics and regulation into stoichiometric models. *Biophysical Journal*, 98:175–185, 2010.

[7] G. Karlebach and R. Shamir. Modelling and analysis of gene regulatory networks. *Nature Reviews Molecular Cell Biology*, 9(10):770–780, 2008.

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

[9] M. Ralser, M.M.C. Wamelink, S. Latkolik, E.E.W. Jansen, H. Lehrach, and C. Jakobs. Metabolic reconfiguration precedes transcriptional regulation in the antioxidant response. *Nature Biotechnology*, 27(7):604–605, 2009.

[10] S. Schuster, T. Pfeiffer, F. Moldenhauer, I. Koch, and T. Dandekar. Exploring the pathway structure of metabolism: decomposition into subnetworks and application to *Mycoplasma pneumoniae*. *Bioinformatics*, 18(2):351–361, 2002.

[11] T. Shlomi, Y. Eisenberg, R. Sharan, and E. Ruppin. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Molecular Systems Biology*, 3(101), 2007.

[12] K. Smallbone, E. Simeonidis, D.S. Broomhead, and D.B. Kell. Something from nothing — bridging the gap between constraint-based and kinetic modelling. *Febs Journal*, 274(21):5576–5585, 2007.

[13] K. Smallbone, E. Simeonidis, N. Swainston, and P. Mendes. Towards a genome-scale kinetic model of cellular metabolism. *BMC Systems Biology*, 4(1):6, 2010.

[14] L.M. Tran, M.L. Rizk, and J.C. Liao. Ensemble Modeling of Metabolic Networks. *Biophysical journal*, 95(12):5606–5617, 2008.