Constraint-based Analysis of Substructures of Metabolic Networks

Sayed-Amir Marashi

Dissertation
zur Erlangung des Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)
am Fachbereich Mathematik und Informatik
der Freien Universität Berlin

Mai 2011

Freie Universität Berlin
Berlin, Germany

To my mother, my father and my wife

Datum des Kolloquiums: *20. Mai 2011*

Gutachter:
***Prof. Dr. Alexander Bockmayr**, Freie Universität Berlin*
***Prof. Dr. Ina Koch**, Johann Wolfgang Goethe-Universität Frankfurt am Main*

Abstract

Constraint-based methods (CBMs) are promising tools for the analysis of metabolic networks, as they do not require detailed knowledge of the biochemical reactions. Some of these methods only need information about the stoichiometric coefficients of the reactions and their reversibility types, i.e., constraints for steady-state conditions. Nevertheless, CBMs have their own limitations. For example, these methods may be sensitive to missing information in the models. Additionally, they may be slow for the analysis of genome-scale metabolic models. As a result, some studies prefer to consider substructures of networks, instead of complete models. Some other studies have focused on better implementations of the CBMs.

In Chapter 2, the sensitivity of flux coupling analysis (FCA) to missing reactions is studied. Genome-scale metabolic reconstructions are comprehensive, yet incomplete, models of real-world metabolic networks. While FCA has proved an appropriate method for analyzing metabolic relationships and for detecting functionally related reactions in such models, little is known about the impact of missing reactions on the accuracy of FCA. Note that having missing reactions is equivalent to deleting reactions, or to deleting columns from the stoichiometric matrix. Based on an alternative characterization of flux coupling relations using elementary flux modes, we study the changes that flux coupling relations may undergo due to missing reactions. In particular, we show that two uncoupled reactions in a metabolic network may be detected as directionally, partially or fully coupled in an incomplete version of the same network. Even a single missing reaction can cause significant changes in flux coupling relations. In case of two consecutive *E. coli* genome-scale networks, many fully-coupled reaction pairs in the incomplete network become directionally coupled or even uncoupled in the more complete reconstruction. In this context, we found gene expression correlation values being significantly higher for the pairs that remained fully coupled than for the uncoupled or directionally coupled pairs. Our study clearly suggests that FCA results are indeed sensitive to missing reactions. Since the currently available genome-scale metabolic models are incomplete, we advise to use FCA results with care.

In Chapter 3, a different, but related problem is considered. Due to the large size of genome-scale metabolic networks, some studies suggest to analyze subsystems, instead of original genome-scale models. Note that analysis of a subsystem is equivalent to deletion of some rows from the stoichiometric matrix, or identically, assuming some internal metabolites to be external. We show mathematically that analysis of a subsystem instead of the original model can lead the flux coupling relations to undergo certain changes. In particular, a pair of (fully, partially or directionally) coupled reactions may be detected as uncoupled in the chosen subsystem. Interestingly, this behavior is the opposite of the flux coupling changes that may happen due to the existence of missing reactions, or equivalently, deletion of reactions. We also show that analysis of organelle subsystems

has relatively little influence on the results of FCA, and therefore, many of these subsystems may be studied independent of the rest of the network.

In Chapter 4, we introduce a rapid FCA method, which is appropriate for genome-scale networks. Previously, several approaches for FCA have been proposed in the literature, namely flux coupling finder algorithm, FCA based on minimal metabolic behaviors, and FCA based on elementary flux patterns. To the best of our knowledge none of these methods are available as a freely available software. Here, we introduce a new FCA algorithm FFCA (Feasibility-based Flux Coupling Analysis). This method is based on checking the feasibility of a system of linear inequalities. We show on a set of benchmarks that for genome-scale networks FFCA is faster than other existing FCA methods. Using FFCA, flux coupling analysis of genome-scale networks of *S. cerevisiae* and *E. coli* can be performed in a few hours on a normal PC. A corresponding software tool is freely available for non-commercial use.

In Chapter 5, we introduce a new concept which can be useful in the analysis of fluxes in network substructures. Analysis of elementary modes (EMs) is proven to be a powerful CBM in the study of metabolic networks. However, enumeration of EMs is a hard computational task. Additionally, due to their large numbers, one cannot simply use them as an input for subsequent analyses. One possibility is to restrict the analysis to a subset of interesting reactions, rather than the whole network. However, analysis of an isolated subnetwork can result in finding incorrect EMs, i.e. the ones which are not part of any steady-state flux distribution in the original network. The ideal set of vectors to describe the usage of reactions in a subnetwork would be the set of all EMs projected onto the subset of interesting reactions. Recently, the concept of "elementary flux patterns" (EFPs) has been proposed. Each EFP is a subset of the support (i.e. non-zero elements) of at least one EM. In the present work, we introduce the concept of ProCEMs (Projected Cone Elementary Modes). The ProCEM set can be computed by projecting the flux cone onto the lower-dimensional subspace and enumerating the extreme rays of the projected cone. In contrast to EFPs, ProCEMs are not merely a set of reactions, but from the mathematical point of view they are projected EMs. We additionally prove that the set of EFPs is included in the set of ProCEM supports. Finally, ProCEMs and EFPs are compared in the study of substructures in biological networks.

Zusammenfassung

Constraintbasierte Methoden (CBM) sind vielversprechende Werkzeuge für die Analyse von metabolischen Netzwerken, da sie keine detaillierte Kenntnis der biochemischen Reaktionen verlangen. Einige dieser Methoden verlangen nur Informationen über die stöchiometrischen Koeffizienten der Reaktionen und deren Reversibilitäts-Typus, d.h. Einschränkungen für *Steady-State*-Bedingungen. Dennoch haben CBM ihre eigenen Grenzen. Zum Beispiel können diese Methoden empfindlich auf fehlende Informationen in den Modellen reagieren. Darüber hinaus können sie bei der Analyse von genomweiten metabolischen Modellen langsam sein. Deshalb ziehen es einige Studien vor, statt kompletten Modellen Substrukturen von Netzwerken zu untersuchen. Andere Studien konzentrieren sich auf eine bessere Implementierung der CBM.

In Kapitel 2 wird die Empfindlichkeit der Flusskopplungsanalyse (FCA) auf fehlende Reaktionen untersucht. Genomweite metabolische Rekonstruktionen sind umfassende, aber dennoch unvollständige, Modelle von realen metabolischen Netzwerken. Während FCA sich als geeignete Methode zur Analyse von metabolischen Beziehungen und zur Erfassung funktionell verwandter Reaktionen in solchen Modellen bewährt hat, ist wenig über die Auswirkungen der fehlenden Reaktionen auf die Genauigkeit der FCA bekannt. Fehlende Reaktionen sind äquivalent mit dem Löschen von Reaktionen oder dem Löschen von Spalten der stöchiometrischen Matrix. Basierend auf einer alternativen Charakterisierung der Flusskopplungsbeziehungen mithilfe von elementaren Flussmodi untersuchen wir die Veränderungen, die fehlende Reaktionen in Flusskopplungsbeziehungen bewirken. Insbesondere zeigen wir, dass zwei ungekoppelte Reaktionen in einem metabolischen Netzwerk als gerichtet, teilweise oder vollständig gekoppelt in einer unvollständigen Version des gleichen Netzwerks wahrgenommen werden können. Sogar eine einzige fehlende Reaktion kann zu erheblichen Veränderungen der Flusskopplungsbeziehungen führen. Bei zwei aufeinander folgenden *E. coli*-genomweiten Netzwerken werden viele vollständig gekoppelte Reaktionen im unvollständigen Netzwerk zu gerichtet gekoppelten oder sogar ungekoppelten Paaren in kompletteren Rekonstruktionen. In diesem Zusammenhang haben wir festgestellt, dass die Genexpressionskorrelationswerte deutlich höher für solche Paare waren, die vollständig gekoppelt blieben, als bei ungekoppelten oder gerichtet gekoppelten Paaren. Unsere Studie zeigt eindeutig, dass FCA-Ergebnisse tatsächlich empfindlich auf fehlende Reaktionen reagieren. Da die derzeit verfügbaren genomweiten metabolischen Modelle unvollständig sind, empfehlen wir, FCA-Ergebnisse mit Vorsicht zu verwenden.

In Kapitel 3 wird ein verwandtes Problem betrachtet. Aufgrund der Größe von genomweiten metabolischen Netzwerken werden in einigen Studien Subsysteme analysiert, statt der ursprünglichen genomweiten Modelle. Dies entspricht der Streichung einiger Zeilen aus der stöchiometrischen Matrix oder auch der Behandlung einiger interner Metabolite als extern. Wir zeigen mathematisch, dass die Analyse eines Subsystems statt des ursprünglichen Modells zu bestimmten

Veränderungen der Flusskopplungsbeziehungen führen kann. Insbesondere ein Paar voll, teilweise oder direktional gekoppelter Reaktionen kann im gewählten Subsystem als ungekoppelt wahrgenommen werden. Interessanterweise ist dieses Verhalten das Gegenteil von den Veränderungen, die aufgrund von fehlenden Reaktionen oder Streichung von Reaktionen geschehen. Wir zeigen auch, dass die Analyse von Organellen-Subsystemen relativ wenig Einfluss auf die Ergebnisse der FCA hat, und daher viele dieser Subsysteme unabhängig vom Rest des Netzwerkes untersucht werden können.

In Kapitel 4 stellen wir eine schnelle FCA-Methode vor, die sich für genomweite Netzwerke eignet. Bisher wurden mehrere Ansätze für FCA vorgeschlagen, nämlich der Flusskopplungsfindungsalgorithmus, FCA basierend auf minimal metabolischem Verhalten und FCA basierend auf elementaren Flussmustern. Soweit wir wissen ist keine dieser Methoden frei als Software verfügbar. Hier stellen wir einen neuen FCA-Algorithmus, FFCA (Zulässigkeitsbasierte-Flusskopplungsanalyse), vor. Bei dieser Methode wird die Zulässigkeit eines Systems von linearen Ungleichungen geprüft. Wir zeigen an einer Reihe von Benchmarks, dass FFCA für genomweite Netzwerke schneller ist als andere bestehende FCA Methoden. Mit FFCA, kann die Flusskopplunganalyse von genomweiten Netzwerken von *S. cerevisiae* und *E. coli* in ein paar Stunden auf einem normalen PC durchgeführt werden. Ein entsprechendes Software-Tool ist zur nicht-kommerziellen Nutzung frei verfügbar.

In Kapitel 5 stellen wir ein neues Konzept vor, das bei der Analyse der Flüsse in Netzwerk-Substrukturen nützlich ist. Die Analyse der Elementarmodi (EMs) ist bewiesenermaßen eine mächtige CBM bei der Studie von metabolischen Netzwerken. Allerdings bedeutet die Aufzählung von EMs einen großen Rechenaufwand. Darüber hinaus kann man sie aufgrund ihrer großen Zahl nicht einfach als Input für spätere Analysen nutzen. Eine Möglichkeit ist es, die Analyse auf eine Teilmenge interessanter Reaktionen zu beschränken. Allerdings kann die Analyse eines isolierten Subnetzes inkorrekte EMs aufdecken, d.h., solche, die nicht Teil einer *Steady-State* Flussverteilung im ursprünglichen Netzwerk sind. Die ideale Menge von Vektoren, um die Nutzung von Reaktionen in einem Teilnetz zu beschreiben, ist die Menge aller EMs, projiziert auf die Teilmenge der interessanten Reaktionen. Vor kurzem wurde das Konzept der elementaren Flussmuster (EFP) vorgeschlagen. Jedes EFP ist eine Teilmenge der Träger (d.h. Nicht-Null-Elemente) von mindestens einem EM. In der vorliegenden Arbeit stellen wir das Konzept der ProCEMs (Projected Cone Elementary Modes) vor. Das ProCEM Set kann durch Projektion des Flusskegels auf den unteren Teilraum und der Aufzählung der extremen Strahlen der projizierten Kegel berechnet werden. Im Gegensatz zu EFPs sind ProCEMs nicht nur eine Reihe von Reaktionen, sondern sind, mathematisch gesehen, projizierte EMs. Wir weisen außerdem nach, dass jedes EFP gleich dem Träger von mindestens einem ProCEM ist. Schließlich werden ProCEMs und EFPs in der Analyse von Substrukturen in biologischen Netzwerken verglichen.

Acknowledgments

First of all, I give thanks to Almighty God who gave me the confidence, the wisdom and the means to follow this work to its conclusion.

I owe my deepest gratitude to my supervisor Prof. Dr. Alexander Bockmayr. As a student with a background in biological sciences, he provided me with the unique opportunity to pursue my PhD studies in his group, opening the doors of Mathematical Biology and Constraint-based Modeling to me. Without his continuous support, I could never finish this work.

My sincere thanks go to Prof. Dr. Ina Koch for reading and evaluating this thesis and also my defense presentation. I am also thankful to Prof. Dr. Martin Vingron and Dr. Heike Siebert for taking part as members of my thesis advisory committee.

I would like to thank the International Max Planck Research School for Computational Biology and Scientific Computing (IMPRS-CBSC) for financial support of my PhD study. Additionally, it is not an exaggeration if I say that the support of Prof. Dr. Vingron, the speaker of the IMPRS-CBSC, was far beyond expectation.

I am especially grateful to Dr. Hannes Luz and Dr. Kirsten Kelleher for all of their kind supports. Kirsten especially helped me by proofreading parts of my thesis and translating the Abstract to German.

I should certainly acknowledge Laszlo David and Dr. Abdelhalim Larhlimi. Abdelhalim helped me a lot during my PhD work, specially at the start, when I was a beginner in the subject. Laszlo was a good friend and a great colleague. I greatly enjoyed working together with him.

I would like to thank Dr. Peter Arndt, my mentor, for helpful discussions on different topics of my work. Dr. Balazs Papp kindly provided us with some datasets which were used in Chapter 2. I should also add here that the software FFCA (Chapter 4) is partly implemented by Bettina Mieth and Laszlo David.

I would not forget Dr. Gunnar Klau, Dr. Corinna Heldt, Shahrad Jamshidi, Hannes Klarner and Yaron Goldstein for for the supportive and friendly environment of our work group. Also, I would like to thank Katja Geiger and Marianne Braun for their kind assistance.

Finally, I wish to dedicate my gratefulness to my family, especially my parents and my wife, for their patience and encouragement throughout my studies to the completion of this thesis.

CONTENTS

1. Introduction 1
1.1 Systems Biology 1
1.2 A Formal Introduction to Metabolic Networks 2
1.2.1 The Basics of Constraint-based Modeling 3
1.2.2 The Constraints, Flux Cone and Flux Space 4
1.2.3 Elementary Modes 8
1.3 Reconstruction of Genome-scale Metabolic Networks 8
1.4 Constraint-based Approaches for Metabolic Network Analysis: An Overview 9
1.4.1 Gap Filling 9
1.4.2 Metabolic Pathway Analysis 11
1.4.3 Flux-balance Methods: FBA, FVA and FCA 12
1.4.4 Uniform Random Sampling of the Flux Space 14
1.4.5 Minimal Cut Sets and Drug Target Detection 15
1.5 Organization of this Thesis 15

2. Flux Coupling Analysis of Metabolic Networks is Sensitive to Missing Reactions 17
2.1 Background 17
2.1.1 Formal Definition of Concepts 18
2.2 Results and Discussion 18
2.2.1 Flux Coupling Definitions based on Elementary Modes . . 19
2.2.2 Impact of Missing Reactions on FCA 22
2.2.3 Biological Implications 31
2.2.4 Comparison of Flux Coupling Relations in two Versions of the *E. coli* Metabolic Network 32
2.3 Methods 35
2.3.1 Datasets: Metabolic Networks and Gene Expression Correlations 35
2.3.2 Study of Flux Coupling Changes by Removal of Single Reactions 36
2.3.3 Comparison of Flux Coupling Relations in two Versions of the *E. coli* Metabolic Network 36
2.3.4 Relationship between Gene Expression Correlations and Flux Coupling in a Recent *E. coli* Model 36

2.3.5 Statistical Analysis 37

3. On Flux Coupling Analysis of Metabolic Subsystems 39
3.1 Background 39
3.2 Formal Definition of the Concepts 40
3.2.1 Concepts from Linear Algebra 40
3.2.2 Concepts Related to Metabolic Networks 40
3.3 Results and Discussion 41
3.3.1 Impact of Redrawing the Network Boundaries on FCA . . 41
3.3.2 Mathematical Analysis of Flux Coupling Relations in Subsystems 42
3.3.3 Subsystem Selection vs. Reaction Deletion 48
3.3.4 Biological Implications 49
3.4 Methods 52
3.4.1 Datasets: Genome-scale Network Models and Organelle Subsystems 52
3.4.2 Flux Coupling Analysis 53

4. FFCA: A Feasibility-based Method for Flux Coupling Analysis 59
4.1 Background 59
4.1.1 Mathematical definitions 59
4.1.2 Approaches to Flux Coupling Analysis 60
4.1.3 Goals of the Present Chapter 61
4.2 Results and Discussion 62
4.2.1 FFCA: Feasibility-based Flux Coupling Analysis 62
4.2.2 Comparison of the four FCA approaches 63
4.3 Materials and Methods 65
4.3.1 Metabolic Network Models 65
4.3.2 Comparison of different FCA methods 65
4.3.3 Implementation Details 66
4.3.4 FCA and Random Sampling of the Flux Space 79

5. Analysis of Metabolic Subnetworks by Flux Cone Projection . 81
5.1 Introduction 81
5.2 Formal Definitions 82
5.3 The State of the Art 82
5.3.1 Computation of a Subset of EMs 82
5.3.2 Computation of EMs in Isolated Subsystems 83
5.3.3 Computation of Elementary Flux Patterns 84
5.3.4 Projection Methods 86
5.4 Method and Implementation 87
5.4.1 Computational Procedure 87
5.4.2 Implementation and Computational Experiments 89
5.4.3 Dataset 89

5.5 Results and Discussion 89
5.5.1 Mathematical Relationships among PEMs, EFPs and ProCEMs 89
5.5.2 Computing the Set of EFPs from the Set of ProCEMs . . 90
5.5.3 Comparing EFPs and ProCEMs 91

6. Conclusions and Further Research 95

Bibliography 97

A. Nomenclature 111

B. Curriculum Vitae 113

Chapter 1 Introduction

Success comes with improving the art of mixing heterogeneous components and working out elegant solutions to complex problems. Consequently, the focus is less on the ultimate components of matter than on the relations between them (Bensaude-Vincent, 2004).

1.1 Systems Biology

According to Kitano (2002), the importance of a systems-level understanding of biological processes was first noticed by Norbert Wiener (Wiener, 1948). In the 1960s, when theoretical biologists started to create computer models of biological systems, systems-level analysis of biological processes came to the center of attention (Spivey, 2004). In 1966, even an international conference on "Systems Theory and Biology" was launched (Mesarović, 1968). However, the term "systems biology" was only introduced to the context of modern biology in the late 1990s (Kitano et al., 1998; Hood, 1998). This was around the time that functional genomics was appearing. In these years, the first genome-scale network was being reconstructed based on the metabolism of *Mycoplasma genitalium* with some additional reactions (Tomita et al., 1997, 1999). Interestingly, the recent remarkable achievement of systems biology to create a new synthetic bacterium (Gibson et al., 2010) is in fact based on the same original idea.

Model reconstruction is the prerequisite of any study in systems biology. There are at least two types of model elements in all systems biology studies: biological *components*, and the *interactions* or *reactions* between them. The set of components together with their relationships form a *network*. For example, in a metabolic network, each metabolite is a component, while each reaction converts a set of reactant metabolites to a set of product metabolites. For network reconstruction, different sources of information, ranging from genomic data to the literature, might be used (Joyce and Palsson, 2006).

In systems biology, the main focus is to study the reactions or interactions, rather than the components. In other words, structures of the components are often ignored when biological systems are studied, although this is not a general rule (see e.g. Aloy and Russell, 2006; Zhang et al., 2009). A biological system can be studied at four levels (Kitano, 2000):

Structural level: The structural relationships among the components are studied at this level. For example, it is possible to identify those components that are connected with many other components of the system (the so-called 'hubs'). Such components potentially have a pivotal role in the function of the system.

Behavioral level: At this level, the goal is to understand or predict the behavior of the system based on its structure. For example, it might be interesting to know how some properties of the system are sensitive to the perturbations. Analysis of the system at this level is very important, as it directly tells us about the functions of the system.

Control level: When the behavior of a system is fairly well understood, it might be possible to apply changes to the system, such that it behaves differently. For example, it is interesting to have an *engineered* bacterial strain which is able to grow on a toxic growth medium and decompose the toxic compounds to non-toxic molecules.

Design level: When a system is well understood, it might be possible to apply major changes to the system, or even design a totally new system *de novo* to have the characteristics of interest. With the new advances in systems biology, biological system design (which is sometimes referred to as *synthetic biology*) becomes more and more popular in the scientific community (Andrianantoandro et al., 2006; Medema et al., 2011).

Unquestionably, the network reconstruction step is of central importance in systems biology. Analysis of the system at any level highly depends on the quality of the input data. Any error in the reconstruction step may result in wrong conclusions (Stelzl and Wanker, 2006; Mestres et al., 2008). Among the biochemical networks, metabolic networks are probably the best studied biological networks (Patil, 2003; Pfeiffer et al., 2005) and probably, have the highest level of correctness. Existence of high-quality genome-scale metabolic models (Thiele and Palsson, 2010) makes them a promising tool for systems biology research.

This dissertation is devoted to metabolic networks, unless stated otherwise.

1.2 A Formal Introduction to Metabolic Networks

Traditionally, biochemists have analyzed biochemical reactions *in vitro* to find the mechanisms and parameters of metabolic reactions (Beard and Qian, 2005). Throughout the last century, this type of study has been done for several enzymatic reactions in a variety of different organisms. If such information is available, one can model a biochemical system with ordinary differential equations (ODEs) to study its dynamical behavior. ODE models are deterministic models, that is,

if the state of the system is known at a certain time point, it is possible to predict the system behavior at any other time point (Wolkenhauer, 2001).

For the analysis of metabolic networks at the systems level, even in the absence of regulation, ODE modeling is not very appropriate for at least two reasons (Beard and Qian, 2005). Firstly, for genome-scale metabolic networks, mechanisms and parameters are unknown for a considerable number of reactions. Secondly, the reaction mechanism and parameters are usually determined by analyzing enzymes in isolation. Even if these data are available for all the reactions, there is no guarantee that the enzymes and reactions behave similarly *in vivo*.

In the analysis of biochemical reaction networks, some data are often available about each reaction without knowing the exact details. Based on this fact, a different strategy to model reaction networks is constraint-based modeling (CBM). By imposing the physicochemical constraints in a model, one can find out what is impossible, but the precise prediction of the system behavior is impossible (Palsson, 2000). Informally speaking, there is a trade-off between the information content of the results and the knowledge required for the modeling.

In CBM, no *a priori* knowledge (or assumptions) regarding the mechanisms and parameters of the reaction system is required. However, when such information is available, e.g. a certain intracellular flux is experimentally measured, this knowledge can be introduced into the model as an additional constraint (Price et al., 2004a; Beard and Qian, 2005).

1.2.1 The Basics of Constraint-based Modeling

Genetic networks and metabolic networks are the two systems controlling the fundamental mechanisms that govern biological systems (Kitano, 2000). While we are aware of the important interplay between these two networks (Shlomi et al., 2007), metabolic networks are usually studied without regulation. The reason lies in the fact that genome-scale metabolic networks are much easier to construct and analyze compared to genome-scale gene regulatory networks. Moreover, even in the absence of the regulation assumption, genome-scale metabolic networks have proven to be useful in predicting the intracellular fluxes and growth phenotypes of different organisms (Edwards et al., 2001; Lee et al., 2010).

As mentioned in Section 1.1, a metabolic network has two elements: the *metabolites*, and the *reactions*. A reaction determines what metabolites (and at what ratio) react with each other to produce other metabolites. A system boundary (which is often the cell membrane) separates the *internal* metabolites (i.e., those metabolites which are inside the system) from the *external* metabolites (those which are outside the system). *Boundary* reactions are those reactions which convert internal metabolites to external ones (or vice versa), while *internal* reactions are the ones which convert internal metabolites to each other.

For a metabolic network N with m internal metabolites and n reactions, the *stoichiometric matrix* S is an $m \times n$ matrix, where element S_{ij} is the stoichiometric coefficient of metabolite i in reaction j. The columns of S will be denoted by

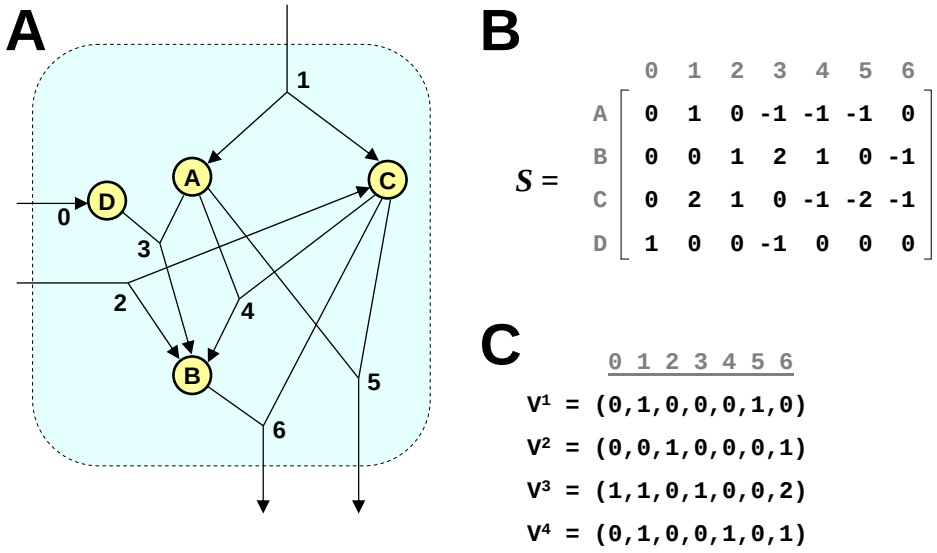


Figure 1.1: *A small hypothetical metabolic network (SmallNet). (A) This network has four (internal) metabolites A,B,C,D and seven reactions 0,...,6; (B) Stoichiometric matrix of the network; (C) List of elementary flux modes.*

$S_1, \ldots, S_n$. We assume $S_j \neq 0$, for all $j = 1, \ldots, n$. The set of metabolites and reactions are shown as M and R, respectively.

Fig. 1.1A shows a small metabolic network, *SmallNet*. The stoichiometric matrix of *SmallNet* is given in Fig. 1.1B.

1.2.2 The Constraints, Flux Cone and Flux Space

In a metabolic network, *flux* through the i-th reaction, $v_i(t)$, is equivalent to the rate of this reaction, which is a function of time. If $\dot{c}$ is the vector denoting changes of internal metabolite concentrations, then for every flux distribution vector $v(t)$, we have $S \cdot v(t) = \dot{c}$ (Schuster and Schuster, 1991).

In the analysis of metabolic networks, it is often assumed that the system is in *steady state*, which means that there is no net change in the concentration of the internal metabolites ($\dot{c} = 0$). Therefore, the flux balance equation can be written as $S \cdot v = 0$ (Fell and Small, 1986), where $v \in \mathbb{R}^n$ is the vector of flux distribution values.

Assuming that the metabolic network is in (quasi-)steady state is a good approximation, since the metabolic reactions are generally much faster than other biological functions, e.g. protein expression (Edwards and Palsson, 1998). Therefore, it is reasonable to assume the system to be in steady-state, unless the dynamics of the system are studied over long time intervals.

There are two types of reactions in metabolic networks: *Irr* is the set of *irreversible* reactions, for which the corresponding flux values are always non-negative; and *Rev* is the set of other reactions which are allowed to have both negative and non-negative values.

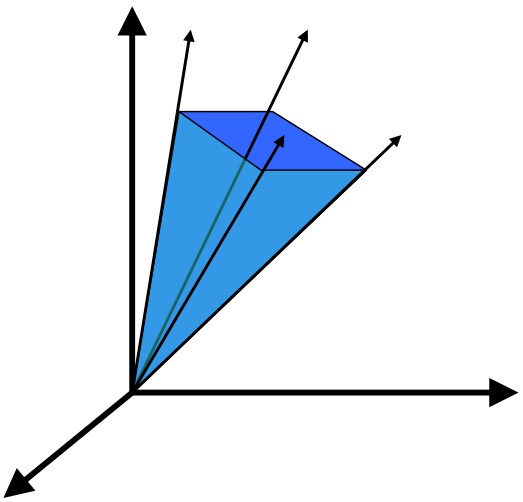


Figure 1.2: *A pointed polyhedral cone in three-dimensional space.*

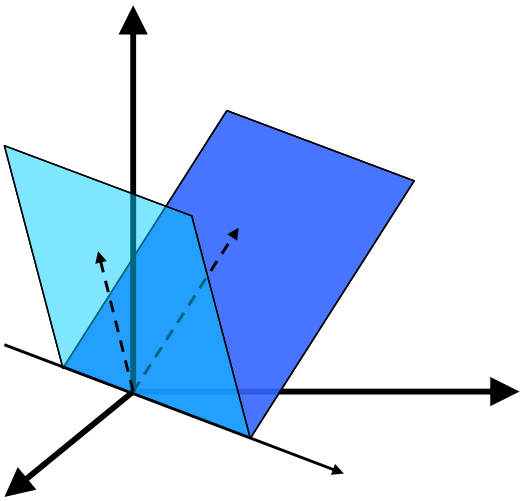


Figure 1.3: *A non-pointed polyhedral cone in three-dimensional space.*

The *flux cone* C of a metabolic network $N = (S, Irr)$, which is the set of all possible flux vectors in steady-state, is defined as:

$$C = \{v \in \mathbb{R}^n \mid S \cdot v = 0, v_i \geq 0 \text{ for all } i \in Irr\} \tag{1.1}$$

For a matrix $A \in \mathbb{R}^{m \times n}$ and a vector $b \in \mathbb{R}^m$, the solution set $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ of a system of linear inequalities $Ax \leq b$ is called a *polyhedron* (Bockmayr and Weispfenning, 2001). If $b = 0$, then P is called a *polyhedral cone*. It can be shown easily that the flux cone is a polyhedral cone. Two examples of polyhedral cones in three-dimensional space are shown in Figures 1.2 and 1.3.

Any non-zero element $r \in C$ is called a *ray* of C. Two rays r and r' are equivalent, written $r \cong r'$, if $r = \lambda r'$, for some $\lambda > 0$. A ray r is *extreme* if there do not exist rays $r', r'' \in C, r' \ncong r''$, such that $r = r' + r''$. In Figure 1.2, the

extreme rays are shown by thin arrows. Each vector v in a pointed cone can be written as $v = \sum \alpha_i g^i$, where the vectors g^i are the extreme rays and $\alpha_i \in \mathbb{R}_{\geq 0}$ for each i.

A subset Φ of a cone C is called a *face* of C (Larhlimi, 2008, page 13) if $\Phi = C$ or $\Phi = C \cap \{x \in \mathbb{R}^n \mid a^T x = 0\}$, where $a^T x \geq 0$ with $a \in \mathbb{R}^n \setminus \{0\}$ is a valid inequality for C.

The lineality space of C is given by:

$$lin.space(C) = \{v \in \mathbb{R}^n \mid S \cdot v = 0, v_i = 0 \text{ for all } i \in Irr\} \tag{1.2}$$

For a pointed cone (like in Figure 1.2), we have $lin.space(C) = \{0\}$. A *non-pointed cone* is a cone for which $dim(lin.space(C)) \geq 1$, where $dim(X)$ denotes dimension of X. If the cone is non-pointed, like in Figure 1.3, then $\{0\} \subsetneq lin.space(C)$. For non-pointed cones, instead of extreme rays we define *minimal proper faces*, which are the faces of dimension $dim(lin.space(C)) + 1$ (Schrijver, 1986). For a pointed cone $dim(lin.space(C)) = 0$, and minimal proper faces are the extreme rays of the cone. A non-pointed cone can be characterized by generating vectors in the lineality space (the thin vector in Figure 1.3) together with generating vectors in minimal proper faces (dashed vectors in Figure 1.3). Each vector v in the non-pointed cone can be written as $v = \sum \alpha_i g^i + \sum \beta_j h^j$, where the vectors g^i and h^j are the generating vectors in lineality space and minimal proper faces, respectively, and $\beta_j \in \mathbb{R}_{\geq 0}$ for each j.

Sometimes, lower bounds and upper bounds of the flux values are also given (*capacity* constraints). In this case, the *flux space*, which is a polyhedron in n-dimensional space, is defined as (Price et al., 2004b):

$$F = \{v \in \mathbb{R}^n \mid S \cdot v = 0, v_i \geq 0 \text{ for all } i \in Irr, l \leq v \leq u\} \tag{1.3}$$

where l and u are the lower and upper bound vectors, respectively. Figure 1.4 shows a polyhedron in three-dimensional space.

Equations 1.1 and 1.3 are the fundamental equations of constraint-based analysis of metabolic networks.

It should be emphasized here that the lower bound (resp. upper bound) of a reaction flux does not necessarily determine the minimum (resp. maximum) possible flux value through this reaction. Because of the flux balance assumption, constraints on one flux can result in constraining other fluxes.

Finding the minimum and maximum possible flux values is done by a technique called *linear programming*. A linear program (LP) is defined by a linear objective function and a set of linear constraints. Here, the objective function can be the flux through a certain reaction, or a linear combination of the fluxes through the reactions. For finding the optimal value of the objective function $k_1 v_1 + \ldots + k_n v_n$ in the flux space defined by Equation 1.3, the following LP should be solved:

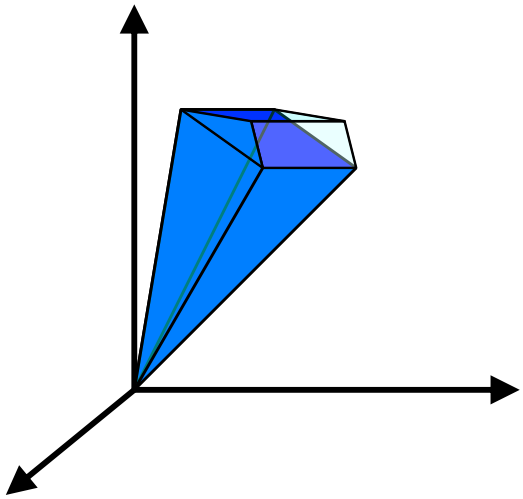


Figure 1.4: *A polyhedron in three-dimensional space. This polyhedron can be obtained by applying additional constraints on the polyhedral cone depicted in Figure 1.2.*

$$\begin{array}{rll} \text{maximize (or minimize)} & k_1 v_1 + \ldots + k_n v_n & k_1, \ldots, k_n \in \mathbb{R} \\ \text{subject to:} & S \cdot v = 0 & \\ & v_i \geq 0 & \text{for all } i \in Irr \\ & l \leq v \leq u & \end{array}$$

Similar LPs are solved in flux balance analysis (Orth et al., 2010), flux variability analysis (Mahadevan and Schilling, 2003) and flux coupling analysis (Burgard et al., 2004). The difference is in the selection of objective functions, l and u vectors, and the number of LPs that should be solved in each approach (see Section 1.4). Other approaches may consider additional constraints, e.g. by assuming some variables to be integer or Boolean. We will use the term *mixed integer linear programming* (MILP) for this type of optimization problems.

"Nonadjustable" vs. "Adjustable" Physicochemical Constraints

Sometimes, stoichiometric and reversibility constraints are the only constraints used in CBM of metabolic networks. These constraints are relatively easy to determine. These constraints are "nonadjustable" constraints (Palsson, 2000), in the sense that they are the intrinsic properties of the system. Some authors prefer to consider additional nonadjustable constraints. For example, Vazquez et al. (2008) incorporated the solvent capacity constraint into the CBM of metabolic networks. Beard et al. (2002) on the other hand, introduced to CBM the use of energy constraints, which are sometimes referred to as "thermodynamic" constraints and are not to be confused with the reversibility constraints.

Flux capacity constraints are generally "adjustable" (Palsson, 2000), i.e., one may adjust these constraints based on the environmental conditions or biochemical prior knowledge. Regulatory constraints (Shlomi et al., 2007) and metabolite concentration constraints (Hoppe et al., 2007) are other examples of adjustable constraints that can be used together with the nonadjustable constraints to improve the reliability and precision of the constraint-based models.

1.2.3 Elementary Modes

For every $v \in \mathbb{R}^n$, the *support* of v denoted by $supp(v) = \{i \in \{1, 2, ..., n\} \mid v_i \neq 0\}$, represents the non-zero components of v.

In the flux cone C, a flux vector e is an *elementary mode* (EM) (Schuster and Hilgetag, 1994; Schuster et al., 2000) if there is no vector $v \in C \setminus \{0\}$ such that $supp(v) \subsetneq supp(e)$. Thus, each EM represents a minimal set of reactions that can work together in steady-state.

The set of all non-equivalent EMs, $E = \{e^1, e^2, \ldots, e^s\}$, in which $e^i \not\cong e^j$ if $i \neq j$, is a set of generating vectors of C (Schuster and Hilgetag, 1994). This means that every flux vector in C can be written as a non-negative linear combination of the vectors in E. In Figure 1.1C, the four EMs of the network are also listed.

1.3 Reconstruction of Genome-scale Metabolic Networks

As described in Section 1.2.1, if in the analysis of a metabolic system the focus is not on the dynamics, a low level of detail is required for building a genome-scale metabolic model (Durot et al., 2009): for every reaction, the only required information is the reversibility type of reaction and the precise stoichiometric coefficients of the substrate(s) and product(s). The main challenge of reconstruction of a genome-scale metabolic network for a certain organism is to provide a comprehensive set of reactions in that organism.

There are at least two main sources of information which can be used in metabolic network reconstruction: *bibliomic* (literature) data and *genomic* data (Duarte et al., 2007).

For several decades, metabolic reactions in different organisms have been identified by biochemists. This is an invaluable and indispensable source of information, which can be retrieved mainly from certain databases, like BRENDA (Schomburg et al., 2002). Further information about the reactions and enzymes may be directly obtained from literature mining (Dickerson et al., 2001).

The second important source of information for metabolic network reconstruction is functional genomics (Fell et al., 2010). Sequence similarity search techniques are the basic tools to annotate sequenced genomes. When the sequence of a new genome is determined, the search techniques are used to find sequences

with 'high' similarity to known enzymes. High similarity in the sequences implies the same enzymatic function.

Although the reactions can be identified automatically with a variety of methods, the resulting metabolic networks are very error-prone. In practice, most of the published metabolic networks are hand-curated. This is a very difficult and time consuming task, which may take from several months to a few years of work for a research team (Thiele and Palsson, 2010). The manual curation and refinement of the model often relies on experimental, organism-specific information (Feist et al., 2009; Thiele and Palsson, 2010).

It should be noted here that sometimes additional "auxiliary" reactions may be added to the metabolic network. An example is the case of *biomass-producing* reaction (Durot et al., 2009). If the stoichiometric ratio of each component in the biomass is known, the biomass-producing reaction can be defined as a boundary reaction which produces an external metabolite (called 'Biomass') and consumes each of its precursors with the corresponding ratios. The biomass objective function, which is the flux through the biomass-producing reaction, is often used in flux balance analysis to model cell growth (Feist and Palsson, 2010).

1.4 Constraint-based Approaches for Metabolic Network Analysis: An Overview

Analysis of metabolism is one of the hot topics in systems biology. At all of the four levels of systems biology (cf. Section 1.1), namely structure, behavior, control and design, analysis of metabolism plays a central role.

Figure 1.5 shows some of the techniques, both *in silico* and *in vivo*, which are used in the analysis and/or manipulation of metabolic networks. In this section, we briefly review some of the CBM methods mentioned in Figure 1.5. These methods are either in the behavioral or control level. We will come back to some of these methods in the next chapters.

1.4.1 Gap Filling

Gap filling is the first technique which naturally comes to mind from the analysis of metabolic behavior (Orth and Palsson, 2010). If the metabolic model does not behave like the real-world network, then some network content may be missing. This inconsistency can suggest further refinement of the reconstructed model.

Constraint-based techniques are possible tools to deal with this problem. For example, Reed et al. (2006) used datasets of growth phenotype data and also databases of potential (enzymatic and exchange) reactions to find missing reactions. In their approach, linear optimization is used to find a minimum number of candidate reactions which, if added to the model, would reduce the inconsistencies between the observed growth phenotypes and the *in silico* predictions. A related optimization-based strategy was proposed by Satish Kumar et al. (2007). In this

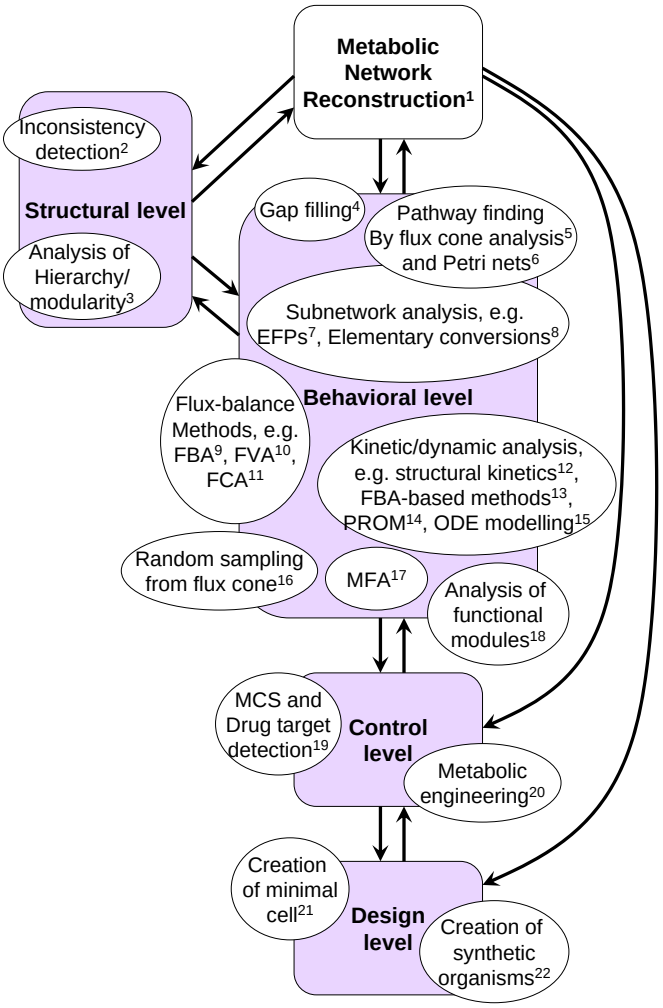


Figure 1.5: *Several techniques have been developed for systems-level analysis of metabolism. These techniques belong to different levels of system analysis (see Section 1.1). At the structural level, the questions are usually answered with in silico methods. The importance of incorporating wet-lab and in vivo experiments generally increases in the next levels, and reaches the highest level at the system manipulation and design level. An arrow from a first box to a second box means that the results of the analysis in the first box is useful at the level of the second box. Examples/reviews of the mentioned techniques can be found in the following references. 1: Thiele and Palsson (2010); 2: Gevorgyan et al. (2008); 3: Ravasz et al. (2002); 4: Orth and Palsson (2010); 5: Klamt and Stelling (2003); 6: Voss et al. (2003); Baldan et al. (2010); 7: Kaleta et al. (2009); 8: Urbanczik and Wagner (2005); 9: Orth et al. (2010); 10: Mahadevan and Schilling (2003); 11: Burgard et al. (2004); 12: Steuer et al. (2006); 13: Covert and Palsson (2002); Covert et al. (2008); Shlomi et al. (2007); 14: Chandrasekaran and Price (2010); 15: Kremling et al. (2007); 16: Price et al. (2004b); 17: Zamboni (2011); 18: Poolman et al. (2007); 19: Klamt and Gilles (2004); Perumal et al. (2011); 20: Unrean et al. (2010); 21: Yus et al. (2009); 22: Gibson et al. (2010).*

approach, growth phenotype information is not required. The first part of the algorithm, GapFind, works based on the assumption that all internal metabolites are produced by at least one reaction. Therefore, if a metabolite is not producible, then the network should be fixed. In order to fix the network, not only addition of new enzymatic and exchange reactions is taken into account, but also addition of compartment-exchange reactions and changing the reversibility type of reactions were considered. GapFind, which is based on MILP, discovers the metabolites that are not producible. Then, GapFill, which is another optimization-based procedure, tries to resolve the problem of missing reactions by applying a minimum number of modifications.

1.4.2 Metabolic Pathway Analysis

Metabolic pathway analysis is the finding and analysis of simple and biologically meaningful "routes" in metabolic networks (Schuster et al., 2000; Schilling et al., 2000; Klamt and Stelling, 2003). Each pathway explains a certain biochemical ability to convert a set of reactants to a set of products in steady-state. Each complicated biochemical behavior of a metabolic networks in steady-state is assumed to be a combination of these simple pathways.

There are several approaches for metabolic pathway analysis. A recent review can be found in Llaneras and Picó (2010). Extremal currents, Elementary modes (EMs), extreme pathways and minimal metabolic behaviors are the most famous concepts.

Extremal (or extreme) current (Clarke, 1980, 1988) is probably the oldest constraint-based concept for metabolic pathway analysis. To obtain extremal currents, it is necessary to split all the reversible reactions. Therefore, a post-processing step is required to remove the 2-cycles (i.e., cycles with two reactions) corresponding to the split reactions. Extremal currents are the extreme rays of the reconfigured cone. It has been shown that except for the 2-cycles, extremal currents and EMs are equivalent (Klamt and Stelling, 2003; Larhlimi and Bockmayr, 2008). It should be noted that this concept has not appeared frequently in the literature in recent years.

Extreme pathway (EXPA) analysis (Schilling et al., 2000) follows a similar logic, except for splitting only internal (and not boundary) reversible reactions, in contrast to splitting all reversible reactions in the case of extremal currents. A wide range of studies have used extreme pathway analysis, e.g. to study network redundancy (Price et al., 2002), reaction coparticipation (Papin et al., 2002) and in analyzing biologically meaningful pathways (Wiback and Palsson, 2002; Ding et al., 2008).

EM analysis is probably the most popular method to study pathways in metabolic networks. A mathematical definition of the EMs is presented in Section 1.2.3. EMs are defined in the flux cone, in contrast to extremal currents and extreme pathways which are defined in reconfigured cones obtained by splitting (some) reversible reactions. EMs are not only used for pathway identification, but

they are also used for analysis of minimal cut sets (see Section 1.4.5), control-effective fluxes (Stelling et al., 2002) and many other applications (Gagneur and Klamt, 2004). Efficient algorithms for computing EMs are reported in the literature (Terzer and Stelling, 2008, 2010) which makes the EM analysis even more attractive.

EMs, extremal currents and extreme pathways are all inner descriptions of the flux cone (Larhlimi and Bockmayr, 2008). Minimal metabolic behaviors (MMBs) provide a minimal outer description of the steady-state flux cone (Larhlimi and Bockmayr, 2009). A *metabolic behavior* is defined as a non-empty set of irreversible reactions D such that there exists a flux vector $v \in C$ with $D = supp(v)$. A metabolic behavior is said to be minimal if no proper subset $D' \subsetneq D$ is also a metabolic behavior. The reversible metabolic space is equal to the $lin.space(C)$, which can be generated by a finite number of (non-unique) generating vectors. Since the MMBs are in a one-to-one correspondence with the minimal proper faces of the flux cone (Larhlimi and Bockmayr, 2009), the full set of MMBs together with the $lin.space(C)$ yields a complete, minimal and unique description of the cone.

1.4.3 Flux-balance Methods: FBA, FVA and FCA

Flux-balance methods (Veeramani and Bader, 2010) are certainly the most commonly used techniques in the constraint-based analysis of metabolic networks. Several related flux-balance techniques are suggested in the literature.

Flux balance analysis (FBA) (Fell and Small, 1986; Varma and Palsson, 1993) is one of the oldest methods in CBM of metabolic networks. FBA is based on assuming the stoichiometric and capacity constraints on the flux values. The goal of FBA is to find the optimal value of an objective function, under the constraints.

For example, suppose that *E. coli* is cultured in a (continuous) growth medium with a certain known composition of nutrients (e.g. glucose, amino acids, etc). The question is how fast can the cells grow, which we model by taking the following steps. Initially, an appropriate genome-scale metabolic network model of *E. coli* should be selected. In such models, often a fictitious "biomass" reaction is included to model the cell growth (see Section 1.3). The lower- and upper-bounds of some fluxes can be determined from the composition of the growth medium and the reversibility type of reactions. Suppose that v_{obj} is the flux through the biomass reaction. For flux balance analysis, one should solve the following LP:

$$
\begin{aligned}
\max \quad & v_{obj} && \\
\text{subject to:} \quad & \sum_{r \in R} S_{mr} v_r = 0 && \forall m \in M \\
& v_r^{min} \leq v_r \leq v_r^{max} && \forall r \in R \\
& v_i \geq 0 && \forall i \in Irr
\end{aligned}
$$

For a well-defined set of constraints in FBA, the LP will be feasible and an optimal solution is obtained by using an LP solver software.

It is shown that the experimental growth rates of bacteria for a variety of different substrates can be predicted successfully by FBA (Edwards et al., 2001; Raghunathan et al., 2009). However, FBA has its own limitations. One important drawback of FBA is that the objective function of the cell should be known in advance. It has been shown that no single objective function describes the flux states under all growth conditions (Schuetz et al., 2007). Another shortcoming of FBA is that by solving an optimization problem, only one single flux distribution will be obtained. If multiple optimal solutions exist, which is often the case due to alternative biochemical pathways, there is no guarantee that this certain flux distribution is even close to the flux distributions *in vivo*.

Flux variability analysis (FVA) (Mahadevan and Schilling, 2003) is a method to study the possible range of each flux when the objective function of FBA has its optimal value. In FVA, for every reaction i the following LPs are solved:

$$\begin{aligned}
\text{max/min} \quad & v_i \\
\text{subject to:} \quad & \sum_{r \in R} S_{mr} v_r = 0 && \forall m \in M \\
& v_r^{min} \leq v_r \leq v_r^{max} && \forall r \in R \\
& v_i \geq 0 && \forall i \in Irr \\
& v_{obj} = v_{obj}^*
\end{aligned}$$

where $v_{obj}^* \in \mathbb{R}$ is the optimal value of the objective function of FBA.

Flux coupling analysis (FCA) (Burgard et al., 2004) considers a different problem. It is known that in steady-state conditions, the so-called *blocked* reactions never take a non-zero flux. For unblocked reactions, some of the fluxes are dependent on some other fluxes. For mathematical definitions of flux coupling relations, we refer to Chapter 2.

In FCA, the goal is to find the blocked reactions, and the flux coupling relation between every pair of unblocked reactions. For this, Burgard et al. (2004) suggest to construct a modified network by splitting every reversible reaction into two irreversible (forward and backward) reactions. In order to find blocked reactions, they suggest to find the following LP for every reaction j:

$$\begin{aligned}
\text{max} \quad & v_j \\
\text{subject to:} \quad & \sum_{r \in R^*} S_{mr} v_r = 0 && \forall m \in M \\
& v_r \geq 0 && \forall r \in R^* \\
& v_r^{uptake} \leq v_r^{uptake_max} && \forall r \in R^*_{trans}
\end{aligned}$$

where R^* is the set of reactions in the modified network, R_{trans} is the set of transport reactions for which an upper bound is known, and v_r^{uptake} is the set of boundary reactions for which an upper bound is known. Obviously, if the maximum possible value for v_j is zero, then reaction j cannot take any flux, and it is a blocked reaction.

In the next step, in order to find the coupling relations of unblocked reactions, the following two LPs should be solved for every pair of reactions i and j:

$$\begin{aligned}
\text{max/min} \quad & \widehat{v}_i \\
\text{subject to:} \quad & \sum_{r \in \mathcal{R}} S_{mr} \widehat{v}_r = 0 && \forall m \in M \\
& \widehat{v}_j = 1 \\
& \widehat{v}_r^{uptake} \leq \widehat{v}_r^{uptake_max} \cdot t && \forall r \in R_{trans} \\
& \widehat{v}_r \geq 0 && \forall r \in R \\
& t \geq 0
\end{aligned}$$

where the $\widehat{v}$ variables are the metabolic fluxes normalized by v_j. Based on the optimal values of the LPs, flux coupling relations are inferred (see Chapter 4).

FCA has attracted great attention since it was introduced. The main reason is that flux coupling relations can be seen as the "functional" relations of the metabolic reactions (Burgard et al., 2004). As an example, correlation between the expression levels of metabolic genes can also be explained by flux coupling (Notebaart et al., 2008).

1.4.4 Uniform Random Sampling of the Flux Space

A different strategy to analyze the functional relations of metabolic fluxes is to study their correlation coefficient (Price et al., 2004b). In the first step of this method, a parallelepiped is found that encloses the flux space, i.e., the polyhedron defined in Equation 1.3, as tightly as possible. Then, the generating vectors of the parallelepiped are used for uniform random sampling of flux vectors within the parallelepiped. Each of the generated vectors, which satisfies the stoichiometric and the capacity constraints (Equation 1.3) is within the flux space.

Such randomly generated vectors in the flux space can be used to compute the Pearson correlation coefficient of the fluxes in the flux space. If the correlation coefficient is ± 1, then the fluxes are said to be *perfectly correlated*. If the fluxes have a zero correlation coefficient, then they are *uncorrelated*. If the fluxes are fairly (and not perfectly) correlated, then they are described as *well-correlated* (Price et al., 2004b). A set of reactions with correlated fluxes is sometimes called a *correlated reaction set*, or simply, a *Co-Set*.

It has been shown that correlation between flux pairs can be used for functional analysis of metabolic networks (Schellenberger and Palsson, 2009). For

example, flux correlations are used to predict synthetic lethal genetic interactions (Veeramani and Bader, 2009) and to classify and correlate the causality of single nucleotide polymorphisms (Jamshidi and Palsson, 2006).

1.4.5 Minimal Cut Sets and Drug Target Detection

Minimal cut sets (MCSs) are the irreducible sets of interventions that will result in blocking a set of reactions in the manipulated network (Klamt and Gilles, 2004). Informally speaking, the set of MCSs is the "dual" concept for the set of EMs (Klamt, 2006) in the sense that both concepts are representations of network functions and additionally it is possible to compute each of them if we know the other one.

MCSs can be used in drug target detection. For example, if a certain vital metabolite is known to be specific to a certain pathogenic bacterium, it might be possible to find a drug target to reduce or block the production of this metabolite. MCSs determine all possible minimal sets of target reactions, which if inhibited by appropriate drug(s), will cause the concentration of that certain metabolite to decrease.

1.5 Organization of this Thesis

After the first introductory chapter on systems biology and constraint-based modeling of metabolic networks, five chapters are included to cover the main results.

Chapter 2 is about the sensitivity of FCA to missing reactions. Current metabolic network models are not representing a complete picture of the real networks. Therefore, it is important to know how sensitive the results of FCA are to missing reactions. In this chapter, we prove that due to missing reactions, coupling relations may change to other types of coupling relations and even to uncoupling relations. On the other hand, uncoupling relations are not sensitive to missing reactions. The importance of the results are shown for *E. coli* metabolic network models.

In Chapter 3, a related problem is studied. Since the genome-scale metabolic networks are not easy to study, sometimes authors "cut out" a subsystem of interest, i.e., a new network boundary is considered and all metabolites (and reactions) outside this boundary are assumed to be external. In this chapter, we investigate the impact of this reduction on the results of FCA. We mathematically prove that if a subsystem is analyzed instead of the complete network, coupling relations in the subsystem are certainly coupling relations in the complete network, although they might be a different type of coupling relation.

In Chapter 4, a novel approach for FCA is presented. In this approach, called FFCA, the feasibility of some LPs are checked to determine the flux coupling relation between each pair of reactions. We show that this method is much faster compared to the other FCA methods reported in the literature. Additionally,

we explain why we did not include the alleged FCA method based on random sampling of the flux space.

Often, it is not possible to enumerate the EMs of a metabolic network. When computation of all EMs is not necessary, e.g. when we focus on a certain subnetwork, enumeration of all EMs results in wasted time and effort. In such a case, having the set of projected EMs is enough. In Chapter 5, we present a new concept, the set of projected cone elementary modes (ProCEMs), which represents (a subset of) projected EMs. We prove that every elementary flux pattern is equal to the support of at least one ProCEM. Computational complexity of the problems are also compared.

Summary:

- Constraint-based modeling does not require detailed knowledge of the biochemical reactions, but it cannot precisely characterize the state of the biochemical system.
- In systems biology, a reconstructed model of the system can be studied at four levels: structure, behavior, control and design.
- Genome-scale metabolic networks are reconstructed based on functional genomics, databases of biochemical reactions, literature and sometimes other sources of information.
- A metabolic network can be characterized by its stoichiometric matrix, S, and the set of its irreversible reactions, Irr.
- When no additional constraint is assumed on fluxes through reactions, the set of all feasible steady-state flux vectors is called flux cone. The flux cone is a polyhedral cone.
- When lower- and upper-bounds are known for (some) fluxes, the set of all feasible steady-state flux vectors is called flux space. The flux space is a polyhedron.

Chapter 2

Flux Coupling Analysis of Metabolic Networks is Sensitive to Missing Reactions

Many authors have studied the effect of missing reactions on the analysis of metabolic networks. Additionally, it is important to know how deleting a reaction can influence the properties of a network. This chapter is devoted to investigating the effect of missing reactions (or equivalently, deleting reactions) on flux coupling analysis of metabolic networks. It should be noted that the main results of this chapter are published in Marashi and Bockmayr (2011).

2.1 Background

In the past decade, genome-scale metabolic models for several organisms have been proposed (for an overview see Terzer et al., 2009; Feist et al., 2009; Suthers et al., 2009; Mo and Palsson, 2009). By including hundreds if not thousands of reactions, a genome-scale metabolic network provides a fairly comprehensive overview of a species' metabolic properties (Feist and Palsson, 2008; Price et al., 2003). This allows us to study the functional dependencies among reactions in living organisms, to model physiological properties (see e.g. Liao et al., 1996; Grafahrend-Belau et al., 2009; de Figueiredo et al., 2009; Mo and Palsson, 2009), and to predict their performance in biotechnological applications (Schuster et al., 2001).

A meaningful metabolic pathway can be seen as a series of consecutive reactions that work together for converting substrates into products (Kholodenko et al., 1995). Different approaches have been developed to compute such pathways, the most popular being elementary flux modes (Schuster and Hilgetag, 1994) and extreme pathways (Schilling et al., 2000). However, due to combinatorial explosion (Klamt and Stelling, 2003; Yeung et al., 2007), it may be practically impossible to enumerate *all* elementary flux modes or extreme pathways for a genome-scale metabolic network. In such a case, flux coupling analysis (FCA) offers an alternative technique that does not require pathway computation (Burgard et al., 2004; Notebaart et al., 2008; Bundy et al., 2007; Notebaart et al., 2009; Suthers et al., 2010; Lee et al., 2008). FCA investigates dependencies be-

tween reactions that are active together in steady-state. The coupled reaction sets are sometimes referred to as "Co-Sets" (Papin et al., 2004). Three types of flux coupling relations, namely directional coupling, partial coupling and full coupling, have been introduced.

2.1.1 Formal Definition of Concepts

For a metabolic network $N = (S, Irr)$, if for some reaction r, we have $v_r = 0$ for all $v \in C$, then we say that r is *blocked* (Burgard et al., 2004). In other words, blocked reactions cannot take any non-zero flux in steady-state.

Originally, Burgard et al. (2004) introduced the following definitions for coupling relations between a pair of unblocked reactions i and j:

- *Directionally coupled reactions*: if for all $v \in C, v_i \neq 0$ implies $v_j \neq 0$, then we say that i is directionally coupled to j, or $i \longrightarrow j$ (or identically, $j \longleftarrow i$). Equivalently, we could require that for all $v \in C, v_j = 0$ implies $v_i = 0$. The latter definition has been denoted as $j \xrightarrow{=0} i$ by Larhlimi and Bockmayr (2006).

- *Partially coupled reactions*: if for all $v \in C, v_i \neq 0$ implies $v_j \neq 0$ and vice versa, then we say that i and j are partially coupled, or $i \longleftrightarrow j$. Obviously, partially coupled reactions are also directionally coupled.

- *Fully coupled reactions*: if for all $v \in C$, $v_i \neq 0$ implies $v_j \neq 0$ and vice versa, and additionally there exists a constant c such that for all $v \in C$, $v_i \neq 0$ implies $v_j/v_i = c$, then we say that i and j are fully coupled, or $i \Longleftrightarrow j$. This is equivalent to saying that i and j belong to the same enzyme subset (Pfeiffer et al., 1999). Obviously, fully coupled reactions are also directionally and partially coupled.

If two unblocked reactions i and j are not directionally coupled, i.e., neither $i \longrightarrow j$ nor $j \longrightarrow i$ holds, then they are said to be *uncoupled* (Burgard et al., 2004). This will be denoted by $i \xleftrightarrow{Un} j$.

At present, our knowledge of metabolic reactions in biological systems is not complete. Past studies have considered the effect of missing reactions on modeling these networks (reviewed in Feist et al., 2009). Yet, little is known about the impact of missing reactions on FCA. The objective of this work is to formally study this problem, and to discuss some possible biological implications.

2.2 Results and Discussion

In this section, we first provide an alternative characterization of flux coupling based on elementary modes. We then study the impact of missing reactions on

flux coupling relations. Finally, we will discuss the relevance of our results for the analysis of real metabolic networks.

Before we start, let us prove two useful lemmata. The first lemma formally proves a well-known property, which is often referred to as the "conservation property" of elementary modes (Schuster et al., 2002a; Klamt and Stelling, 2003).

Lemma 2.1. *Consider two metabolic networks $N = (S, Irr)$ and $N' = (S', Irr')$ such that N is a completion of N' with set of missing reactions R. Let C (resp. C') be the flux cone and E (resp. E') be the set of elementary modes of N (resp. N'). For a flux vector v' in N' let $v = (v', 0)$ be the vector in N obtained by adding zero components $v_r = 0$, for all missing reactions $r \in R$. Then $e = (e', 0) \in E$, for all $e' \in E'$. Conversely, if $e = (v', 0) \in E$, for some $v' \in C'$, then $v' \in E'$.*

Proof. Let $e' \in E'$ and suppose $e = (e', 0) \notin E$. Then there exists $v = (v', 0) \in C \setminus \{0\}$, with $supp(v) \subsetneq supp(e) = supp(e')$. Since $0 = Sv = S'v'$ and $Irr' \subseteq Irr$ we get $v' \in C' \setminus \{0\}$ and $supp(v') = supp(v) \subsetneq supp(e')$, in contradiction with e' being an elementary mode in N'.

Conversely, let $e \in E$ with $e = (v', 0)$ for some $v' \in C'$. Suppose $v' \notin E'$. Then there exists $u' \in C' \setminus \{0\}$ with $supp(u') \subsetneq supp(v')$. For $u = (u', 0)$ we get $Su = S'u' = 0$ and since $Irr' = Irr \setminus R$ it follows $u \in C \setminus \{0\}$, $supp(u) = supp(u') \subsetneq supp(v') = supp(e)$, in contradiction with e being an elementary mode in N. ∎

We also prove the following useful lemma:

Lemma 2.2. *Let E be the set of elementary modes in a metabolic network N. Given two reactions i and j suppose that for all $g, h \in E$ with $g_i, h_i \neq 0$ we have $g_i h_j = h_i g_j$. Then either there exists a constant $c \neq 0$ such that $e_j = c \cdot e_i$ for all $e \in E$ with $e_i \neq 0$, or $e_j = 0$ for all $e \in E$ with $e_i \neq 0$.*

Proof. Suppose there exists $e^* \in E$ with $e^*_i \neq 0$ and $e^*_j = 0$. Consider any $e \in E$ with $e_i \neq 0$. By our hypothesis, $e^*_i e_j = e_i e^*_j = 0$, which implies $e_j = 0$.

Suppose that for all $e \in E$ we have $e_i = 0$ or $e_j \neq 0$. For any $g, h \in E$ with $g_i, h_i \neq 0$ we have $g_j/g_i = h_j/h_i = c$, with c independent from g, h. This implies that for all $e \in E$ with $e_i \neq 0$, we have $e_j = ce_i$. Since $e_j \neq 0$, it follows $c \neq 0$. ∎

2.2.1 Flux Coupling Definitions based on Elementary Modes

Flux coupling is originally defined by considering all vectors in the flux cone C (Burgard et al., 2004). In addition, we know that such vectors can be written as non-negative linear combination of elementary flux modes in the network (Schuster and Hilgetag, 1994). The following theorem shows that flux coupling relations may also be defined in terms of elementary modes.

Theorem 2.1. *Let N be a metabolic network with flux cone C and set of elementary modes E. For any two reactions i and j, the following are equivalent:*

(i) For all $v \in C$, $v_i = 0$ implies $v_j = 0$.

(ii) For all $e \in E$, $e_i = 0$ implies $e_j = 0$.

Proof. Since $E \subseteq C$, we have immediately (i) $\Rightarrow$ (ii). To prove that (ii) $\Rightarrow$ (i), we consider the following three cases:

1. For all $e \in E$, we have $e_i \geq 0$: If (ii) holds, then for all $e \in E$ there exists $\lambda_e \in \mathbb{R}$ with $e_j = \lambda_e \cdot e_i$. Any $v \in C$ can be written as $v = \sum_{e \in E} \alpha_e e$, with $\alpha_e \geq 0$. From $v_i = \sum_{e \in E} \alpha_e e_i = 0$, it follows $\alpha_e e_i = 0$, for all $e \in E$. This implies $v_j = \sum_{e \in E} \alpha_e e_j = \sum_{e \in E} \alpha_e \lambda_e e_i = 0$.

2. For all $e \in E$, we have $e_i \leq 0$: similar

3. There exist $e, f \in E$ with $e_i > 0, f_i < 0$. We consider two cases:

   a) For all $e, f \in E$ with $e_i > 0, f_i < 0$, we have $f_i e_j = e_i f_j$. Since there exist $e, f \in E$ with $e_i > 0, f_i < 0$, we have $f_i e_j = e_i f_j$, for all $e, f \in E$, with $e_i, f_i \neq 0$. Based on Lemma 2.2, there are two possibilities:

      a1) There exists $c \neq 0$ such that $e_j = c \cdot e_i$, for all $e \in E$ with $e_i \neq 0$. Let $v = \sum_{e \in E} \alpha_e e$ in C and $v_i = 0$. It follows

$$\begin{aligned} 0 &= v_i = \sum_{e \in E} \alpha_e e_i = \sum_{e \in E,\, e_i \neq 0} \alpha_e e_i = \frac{1}{c} \cdot \sum_{e \in E,\, e_i \neq 0} \alpha_e e_j \\ &\overset{(ii)}{=} \frac{1}{c} \cdot \sum_{e \in E} \alpha_e e_j = \frac{1}{c} \cdot v_j, \end{aligned}$$

      which implies $v_j = 0$.

      a2) For all $e \in E$, $e_i \neq 0$ implies $e_j = 0$. With (ii) we get $e_j = 0$, for all $e \in E$, and so j is a blocked reaction. This implies that (i) also holds.

   (b) There exist $e, f \in E$ with $e_i > 0, f_i < 0$ and $f_i e_j \neq e_i f_j$. We define $v^* = e_i \cdot f - f_i \cdot e$. Since v^* is a positive linear combination of elementary modes, we get $v^* \in C$, with $v^*_i = 0$ and $v^*_j \neq 0$. Let N' be the network obtained from N by deleting reaction i. As before we denote by C' the flux cone of N', and by E' the corresponding set of elementary modes. Let v' be the vector in N' obtained from v^* by deleting the component $v^*_i = 0$. Then $v' \in C'$ and so $v' = \sum_{e' \in E'} \beta_{e'} e'$, for some $\beta_{e'} \geq 0$. Since $v^*_j = v'_j = \sum_{e' \in E'} \beta_{e'} e'_j \neq 0$, there exists $e' \in E'$ with $e'_j \neq 0$. According to Lemma 2.1, the vector $e = (e', 0)$, obtained from e' by adding the i-th component $e_i = 0$, is an elementary mode in N with $e_j = e'_j \neq 0$. This contradicts (ii).

■

Corollary 2.1. *Suppose the reversible reaction i can operate in both directions, i.e., there exist two elementary modes $e, f \in E$ such that $e_i > 0$ and $f_i < 0$. If for some reaction j, we have $i \longrightarrow j$ and $j \longrightarrow i$, then necessarily $i \Longleftrightarrow j$ (i.e., it is not possible to have $i \longleftrightarrow j$ and not $i \Longleftrightarrow j$).*

Proof. First note that j is unblocked, since i is unblocked and $i \longrightarrow j$. The hypotheses of the Corollary correspond to Case 3) in the proof of Theorem 2.1. Since $j \longrightarrow i$, Case 3b) is not possible. It follows that $f_i e_j = e_i f_j$, for all $e, f \in E$ with $e_i, f_i \neq 0$. By Lemma 2.2 and since j is unblocked, there exists a constant $c \neq 0$ such that $g_j = c \cdot g_i$, for all $g \in E$ with $g_i \neq 0$. Hence, $i \Longleftrightarrow j$. ∎

The next corollary gives an alternative characterization of flux coupling relations based on elementary modes. While this has no computational advantage (since the full set of EMs is difficult to compute in genome-scale metabolic networks), this characterization plays an essential role in our formal analysis of flux (un-)coupling relations.

Corollary 2.2. *Let i, j be two unblocked reactions in a metabolic network N with set of elementary modes E.*

- *i is directionally coupled to j ($i \longrightarrow j$) if and only if for all $e \in E$, $e_i \neq 0$ implies $e_j \neq 0$.*
- *i and j are partially coupled ($i \longleftrightarrow j$) if and only if for all $e \in E$, $e_i \neq 0$ implies $e_j \neq 0$ and vice versa.*
- *i and j are fully coupled ($i \Longleftrightarrow j$) if and only if there exists a constant $c \neq 0$ such that for all $e \in E$, $e_j = c \cdot e_i$.*

Proof. This follows directly from the definitions and Theorem 2.1. In the case $i \Longleftrightarrow j$, we also use that any $v \in C$ can be written in the form $v = \sum_{e \in E} \alpha_e e$, for some $\alpha_e \geq 0$. ∎

From the Introduction, we recall that two unblocked reactions i and j are uncoupled iff neither $i \longrightarrow j$ nor $j \longrightarrow i$. This means that there exist $f, g \in E$ with $f_i \neq 0, f_j = 0$ and $g_i = 0, g_j \neq 0$. Here, we define two types of uncoupling relations:

- *Sometimes coupled reactions*: if two unblocked reactions i and j are uncoupled and additionally there exists $e \in E$ such that $e_i \neq 0$ and $e_j \neq 0$, then i and j will be called *sometimes coupled* ($i \overset{S.C.}{\longleftrightarrow} j$).
- *Mutually exclusive reactions*: if two unblocked reactions i and j are uncoupled and additionally, for all $e \in E$, $e_i = 0$ or $e_j = 0$, i.e., they never appear in the same EM, then i and j will be called *mutually exclusive* ($i \overset{M.E.}{\longleftrightarrow} j$).

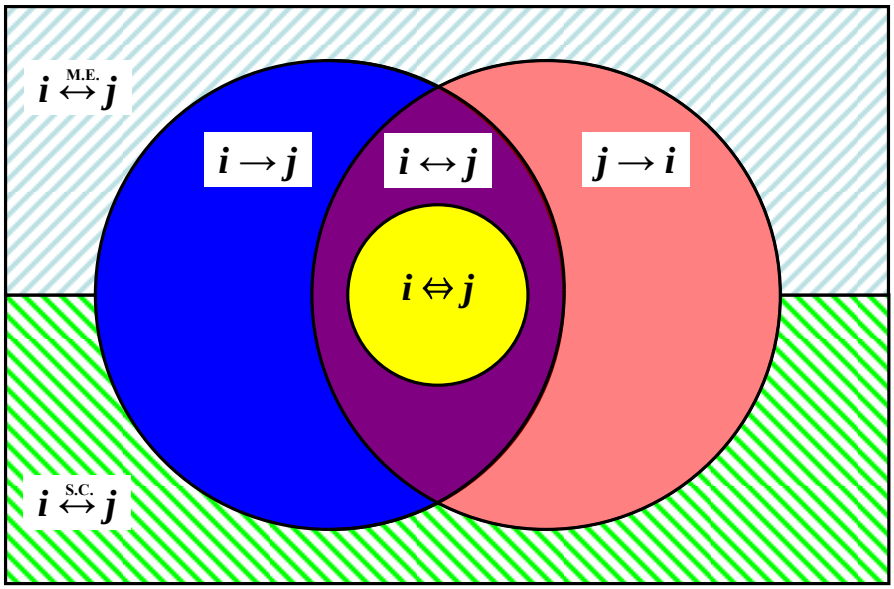


Figure 2.1: *Different types of flux (un-)coupling relations. In a metabolic network, every pair of unblocked reactions (i, j) lies in one of the regions of the diagram. Either i and j are coupled, i.e., inside the two circles, or uncoupled, i.e., outside the circles. If (i, j) lies in the intersection of the two circles, i.e., i is directionally coupled to j and vice versa, then they are partially coupled. Fully coupled reaction pairs form a subset of the partially coupled pairs. If i and j are uncoupled, they are either sometimes coupled or mutually exclusive.*

Fig. 2.1 briefly summarizes these relationships.

The concept of "mutually exclusive reactions" is not new. Based on the study of reaction participation in extreme pathways, Papin et al. (2002) classified reactions into sets that are always, sometimes or never used for target production. Later, Klamt and Stelling (2003) applied the concept of *mutually excluding* reactions to reactions that never appear in the same elementary mode.

2.2.2 Impact of Missing Reactions on FCA

We now investigate the effect of missing reactions in metabolic networks. Suppose that the stoichiometric matrix S' of a metabolic network $N' = (S', Irr')$ is obtained by deleting one or more columns (i.e., reactions) from the stoichiometric matrix S of a second network $N = (S, Irr)$. We will refer to the first network as the *incomplete network*, and to the second network as the *completion*. The set $R \neq \varnothing$ of reactions present only in the larger network N will be called *missing reactions* or *missing network content* of the smaller network N'. We further assume $Irr' = Irr \setminus R$.

Suppose that *SmallNet* (see Fig. 1.1) constitutes the actual complete metabolic network of a biological system. The four EMs instantly indicate that reactions 1

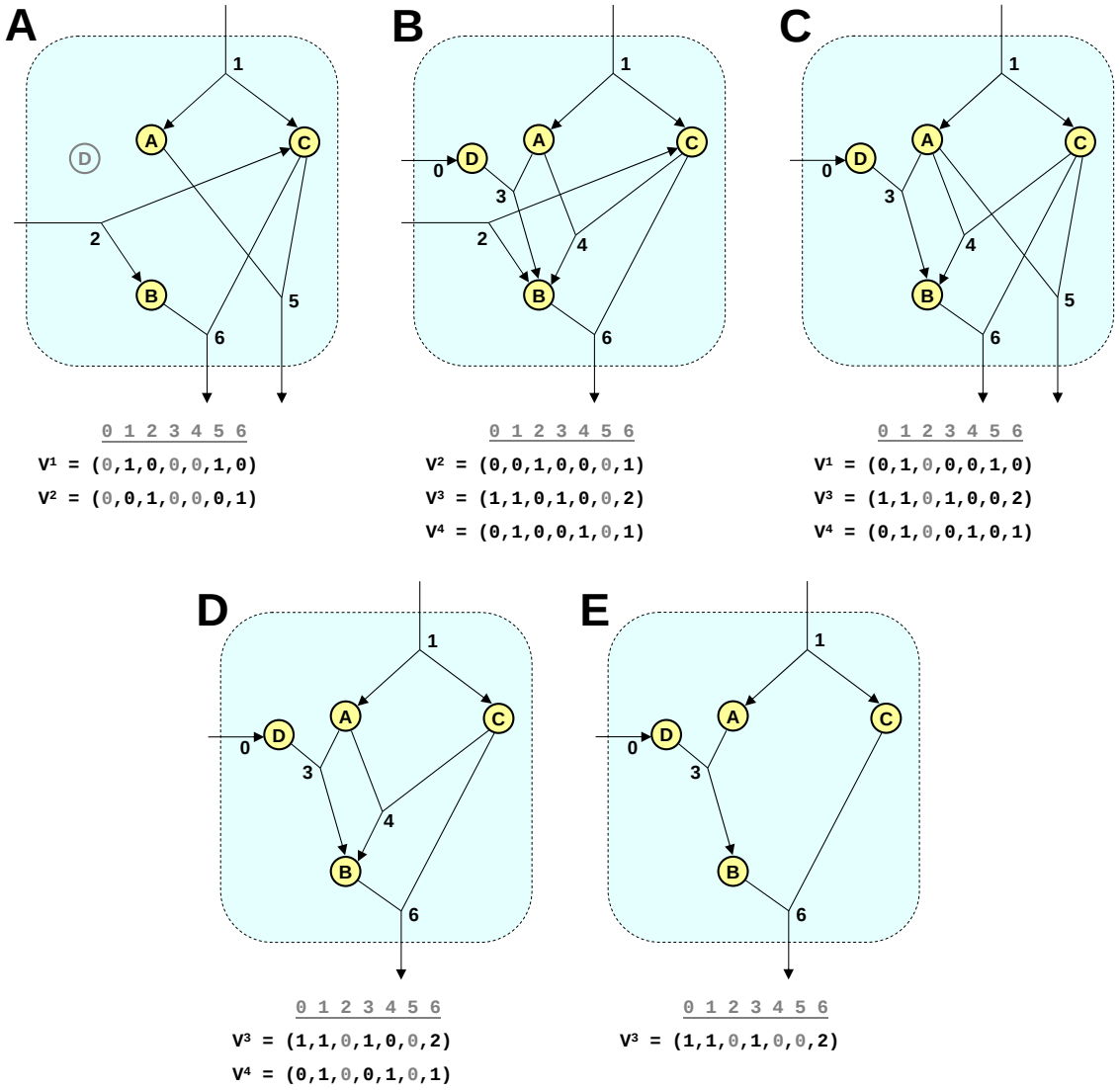


Figure 2.2: *Five possible scenarios for missing reactions in a metabolic network. Suppose SmallNet (Fig. 1.1) is the original complete network. We consider five incomplete reconstructions with the following missing reactions: (A) 0, 3 and 4; (B) 5; (C) 2; (D) 2 and 5; (E) 2, 4 and 5. In (A), metabolite D is not connected to the rest of the network, and therefore is shown in gray. The list of elementary modes is given for each network, with grey "zero" values corresponding to the missing reactions.*

and 6 are sometimes coupled. In Fig. 2.2, the effect of missing network content on the EMs is shown. Suppose that we want to reconstruct a model of *SmallNet*. Each part of the figure represents a possible incomplete network scenario, i.e., a reconstruction of the actual network *SmallNet*, but with some reactions missing due to incomplete knowledge. If, for example, reactions 0, 3 and 4 are missing from this network, then no EM containing both 1 and 6 exists in the resulting

network. Hence, they will be detected as mutually exclusive. On the other hand, if only reaction 5 or only reaction 2 is missing, reactions 1 and 6 will be detected as directionally coupled pairs. If both reactions 2 and 5 are missing, 1 and 6 become partially coupled. Finally if in addition to 2 and 5, reaction 4 is also missing, 1 and 6 will be detected as fully coupled. The effect of missing reactions on the relationship between reactions 1 and 6 is summarized in Table 2.1.

Missing reactions	flux coupling relation	related figure
None	$1 \overset{S.C.}{\longleftrightarrow} 6$	Fig. 1.1A
0,3,4	$1 \overset{M.E.}{\longleftrightarrow} 6$	Fig. 2.2A
5	$1 \longrightarrow 6$	Fig. 2.2B
2	$6 \longrightarrow 1$	Fig. 2.2C
2,5	$1 \longleftrightarrow 6$	Fig. 2.2D
2,4,5	$1 \Longleftrightarrow 6$	Fig. 2.2E

Table 2.1: *Some possible scenarios for missing reactions in a network and the effect on flux coupling relations.*

In the above example, we observe that it is possible to have a pair of uncoupled reactions in the original complete network, which is detected as a pair of coupled reactions due to missing reactions in the incomplete network. The following theorems study possible changes in flux coupling relations due to missing network content.

Lemma 2.3. *Let $i \neq j$ be two irreversible reactions in an incomplete network $N' = (S', Irr')$. Let $N = (S, Irr)$ be a completion constructed by adding an irreversible reaction $n+1$ to N', such that $S_{n+1} = -S'_i$. Then, we have:*

1. *$i \longrightarrow j$ does not hold in N.*
2. *If $i \longleftrightarrow j$ in N', then $j \longrightarrow i$ holds in N.*

Proof. Let C (resp. C') be the flux cone of N (resp. N').

1. Define $v \in \mathbb{R}^{n+1}$ by $v_i = v_{n+1} = 1$ and $v_r = 0$ for all $r \neq i, n+1$. Then $v \geq 0$ and $S \cdot v = 1 \cdot S_i + 1 \cdot S_{n+1} = 0$. It follows $v \in C$, with $v_i \neq 0$ and $v_j = 0$. This shows that $i \longrightarrow j$ does not hold in N.

2. If $i \longleftrightarrow j$ in N', there exists $u' \in C'$ such that $u'_i > 0$ and $u'_j > 0$. Define $u \in C$ by $u = (u', 0)$. Assume $j \longrightarrow i$ does not hold in N. Then there exists $w = (w', w_{n+1}) \in C$ such that $w_j > 0$ and $w_i = 0$. Since i and j are partially coupled in N', $w' \notin C'$ and therefore $w_{n+1} > 0$.

   Now define $v = (v', v_{n+1}) \in \mathbb{R}^{n+1}$ by $v_i = v_{n+1} = 0$, and $v_r = \frac{u_i}{w_{n+1}} \cdot w_r + u_r$, for $r \neq i, n+1$. Using $Su = \sum u_r S_r = 0$ and $Sw = \sum w_r S_r = 0$, we get

$v \in C$ because $Sv = \sum_{r\neq i,n+1}(\frac{u_i}{w_{n+1}} \cdot w_r + u_r)S_r = \frac{u_i}{w_{n+1}} \sum_{r\neq i,n+1} w_r S_r + \sum_{r\neq i,n+1} u_r S_r = \frac{u_i}{w_{n+1}}(-w_{n+1}S_{n+1}) - u_i S_i = 0$. Furthermore, $v_r \geq 0$, whenever $u_r, w_r \geq 0$. Since $v_{n+1} = 0$, we get $v' \in C'$, $v'_i = 0$ and $v'_j = u_i w_j / w_{n+1} + u_j > 0$, in contradiction with $i \longleftrightarrow j$ in N'.

■

Theorem 2.2. *In an incomplete network N' with no blocked reactions, the following statements hold for any pair of reactions i and j:*

1. *If i and j are irreversible and fully ($i \Longleftrightarrow j$) or partially coupled ($i \longleftrightarrow j$) in N', there exists a completion N in which $i \longleftrightarrow j$ does not hold. Instead, each of the following relations can hold: $i \longrightarrow j$ or $j \longrightarrow i$ or $i \overset{S.C.}{\longleftrightarrow} j$.*

2. *If i and j are irreversible and directionally coupled ($i \longrightarrow j$) in N', there exists a completion N in which $i \overset{S.C.}{\longleftrightarrow} j$.*

3. *If i and j are not "proportional" (i.e., $S_i \neq c \cdot S_j$, for all c) and mutually exclusive ($i \overset{M.E.}{\longleftrightarrow} j$) in N', there exists a completion in which they are sometimes coupled: $i \overset{S.C.}{\longleftrightarrow} j$.*

Proof. 1. If $i \Longleftrightarrow j$ in N', then also $i \longleftrightarrow j$ in N'. We construct a completion N_1 by adding an irreversible reaction $n+1$ to N', such that $S_{n+1} = -S_i$. According to Lemma 2.3, $i \longrightarrow j$ does not hold in N_1, while $j \longrightarrow i$ holds in N_1. Similarly, $i \longrightarrow j$ and not $j \longrightarrow i$ hold in a completion N_2 constructed by adding an irreversible reaction $n+1$ to N', such that $S_{n+1} = -S_j$.

If N is constructed from N' by adding two irreversible reactions $n+1$ and $n+2$, such that $S_{n+1} = -S_i$ and $S_{n+2} = -S_j$, then by Lemma 2.3, neither $j \longrightarrow i$ nor $i \longrightarrow j$ hold in N. This means that i and j are uncoupled in N. Since we have $i \Longleftrightarrow j$ in N', there exists $e' \in E'$ with $e'_i, e'_j \neq 0$. By Lemma 2.1, $e = (e', 0, 0)$ is an elementary mode in N with $e_i, e_j \neq 0$. Hence, $i \overset{S.C.}{\longleftrightarrow} j$ holds in N.

2. Similar to 1) (by adding reactions $n+1$ and $n+2$ such that $S_{n+1} = -S_i$ and $S_{n+2} = -S_j$).

3. Let N be the network constructed by adding a reaction $n+1$ to N', such that $S_{n+1} = -S_i - S_j$. Define $e \in C$ by $e_i = e_j = e_{n+1} = 1$ and $e_r = 0$, for all other reactions r.

   We claim that e is an elementary mode in N. Suppose there exists $v \in C \setminus \{0\}$ such that $supp(v) \subsetneq supp(e)$. If $v_{n+1} \neq 0$ and $v_r = 0$, for $r \neq n+1$, we get $S_{n+1} = 0$. It follows $S_j = -S_i$ in contradiction to the hypothesis that i, j are not proportional. If $v_i, v_{n+1} \neq 0$ and $v_r = 0$, for $r \neq i, n+1$, we get $S_j = c \cdot S_i$, for some $c \neq 0$, which is again a contradiction. The case $v_j, v_{n+1} \neq 0$ and $v_r = 0$, for $r \neq j, n+1$, is analogous. This shows $e \in E$.

Since i, j are uncoupled in N', there exist $f', g' \in E'$ with $f'_i \neq 0, f'_j = 0$ and $g'_i = 0, g'_j \neq 0$. By Lemma 2.1, the vectors $f = (f', 0)$ and $g = (g', 0)$ are elementary modes in N, i.e., i, j remain uncoupled in N. Since at the same time there exists an elementary mode e such that $e_i, e_j \neq 0$, we conclude that $i \overset{S.C.}{\longleftrightarrow} j$ in N.

■

Under some additional hypotheses, we can prove the following:

Theorem 2.3. *Let N' be a metabolic network with irreversible reactions only and no blocked reactions. Consider a pair of reactions i and j with $i \Longleftrightarrow j$ in N' such that for all $v \in C' \setminus \{0\}$, $\{i, j\} \subsetneq supp(v)$, where C' is the flux cone of N'. Then there exists a completion N in which $i \longleftrightarrow j$ holds, but not $i \Longleftrightarrow j$.*

Proof. Since $i \Longleftrightarrow j$ in N', there exists $\lambda > 0$ such that for all $v' \in C'$, we have $v'_j = \lambda v'_i$. We construct the completion N by adding an irreversible reaction $n+1$ to N' with $S_{n+1} = -S_i - (\lambda + 1)S_j$. We claim that $i \longleftrightarrow j$ holds in N, but not $i \Longleftrightarrow j$.

a) Define $v \in \mathbb{R}^{n+1}$ by $v_i = v_{n+1} = 1$, $v_j = \lambda + 1$, and $v_r = 0$ for all $r \neq i, j, n+1$. Then $S \cdot v = S_i + (\lambda + 1)S_j + S_{n+1} = 0$. Therefore, $v \in C$ and $v_j/v_i = \lambda + 1 \neq \lambda$. This shows that $i \Longleftrightarrow j$ does not hold in N.

b) To prove $i \longrightarrow j$, we assume that there exists $u \in C$ for which $u_i > 0$ and $u_j = 0$. From Lemma 2.1, we get $u_{n+1} > 0$, because if $u_{n+1} = 0$, then $(u_1, \ldots, u_n) \in C'$, in contradiction with $i \Longleftrightarrow j$ in N'. Since $i \Longleftrightarrow j$ in N', there exists $v' \in C'$, and a corresponding $v = (v', 0) \in C$, such that $v_j = \lambda v_i > 0$. Using $S_{n+1} = -S_i - (\lambda + 1)S_j$, it follows that

$$\begin{aligned}
0 &= \frac{\lambda}{u_{n+1}} \cdot S \cdot u + \frac{\lambda + 1}{v_i} \cdot S \cdot v \\
&= \frac{\lambda}{u_{n+1}} \left[u_i S_i + u_{n+1}(-S_i - (\lambda + 1)S_j) + \sum_{r \notin \{i,j,n+1\}} u_r S_r \right] + \\
&\qquad \frac{\lambda + 1}{v_i} \left[v_i S_i + \lambda v_i S_j + \sum_{r \notin \{i,j,n+1\}} v_r S_r \right] \\
&= \left[\lambda \cdot \frac{u_i}{u_{n+1}} + 1 \right] S_i + \sum_{r \notin \{i,j,n+1\}} \left[\lambda \cdot \frac{u_r}{u_{n+1}} + (\lambda + 1)\frac{v_r}{v_i} \right] S_r.
\end{aligned}$$

Define $w \in \mathbb{R}^{n+1}$ by $w_i = \lambda u_i/u_{n+1} + 1 > 0$, $w_j = w_{n+1} = 0$ and $w_r = \lambda u_r/u_{n+1} + (\lambda + 1)v_r/v_i$, for all $r \notin \{i, j, n+1\}$. Then $Sw = 0$ and $w_r \geq 0$, for every reaction r, which shows $w \in C$. Using $w_{n+1} = 0$, we get $(w_1, \ldots, w_n) \in C'$. Since $w_j = 0$, this is a contradiction to the assumption of $i \Longleftrightarrow j$ in N'.

c) To prove $j \longrightarrow i$, we assume that there exists $u \in C$ with $u_j > 0$ and $u_i = 0$. As before, we get $u_{n+1} > 0$. Furthermore, there exists $v' \in C'$, and a

corresponding $v = (v', 0) \in C$, such that $v_j = \lambda v_i > 0$. It follows that

$$0 = \frac{1}{u_{n+1}} \cdot S \cdot u = \frac{u_j}{u_{n+1}} \cdot S_j + S_{n+1} + \sum_{r \notin \{i,j,n+1\}} \frac{u_r}{u_{n+1}} \cdot S_r \tag{2.1}$$

and

$$0 = \frac{1}{v_i} \cdot S \cdot v = S_i + \lambda S_j + \sum_{r \notin \{i,j,n+1\}} \frac{v_r}{v_i} \cdot S_r. \tag{2.2}$$

Adding the two equations and using $S_{n+1} = -S_i - (\lambda + 1)S_j$ we get

$$0 = \left[\frac{u_j}{u_{n+1}} - 1\right] S_j + \sum_{r \notin \{i,j,n+1\}} \left[\frac{u_r}{u_{n+1}} + \frac{v_r}{v_i}\right] S_r. \tag{2.3}$$

Now define $w \in \mathbb{R}^{n+1}$ by $w_i = w_{n+1} = 0$, $w_j = u_j/u_{n+1} - 1$ and $w_r = u_r/u_{n+1} + v_r/v_i$, for $r \notin \{i, j, n+1\}$.

Then $Sw = 0$ and $w_r \geq 0$, for $r \neq j$.

Assume $w_r = 0$, for all $r \notin \{i, j, n+1\}$. Then $v_r/v_i = 0$, for all $r \notin \{i, j, n+1\}$. With (2.2), we get $S_i + \lambda S_j = 0$. If we define $\tilde{w} \in \mathbb{R}^n$ by $\tilde{w}_i = 1, \tilde{w}_j = \lambda$, and $\tilde{w}_r = 0$, for $r \neq i, j$, we get $\tilde{w} \in C'$ with $supp(\tilde{w}) = \{i, j\}$, in contradiction to the hypothesis. We conclude $w_r > 0$, for some $r \notin \{i, j, n+1\}$.

If $w_j \geq 0$, we get $w' = (w_1, \ldots, w_n) \in C' \setminus \{0\}$ with $w'_i = 0$, in contradiction to the hypothesis $v'_i \neq 0$, for all $v' \in C' \setminus \{0\}$.

If $w_j < 0$, we have with (2.3)

$$S_j = \sum_{r \notin \{i,j,n+1\}} -\frac{w_r}{w_j} \cdot S_r$$

Using (2.2) this implies

$$S_i + \sum_{r \notin \{i,j,n+1\}} \left[-\lambda \cdot \frac{w_r}{w_j} + \frac{v_r}{v_i}\right] \cdot S_r = 0$$

Define $\overline{w} \in \mathbb{R}^{n+1}$ by $\overline{w}_i = 1$, $\overline{w}_j = \overline{w}_{n+1} = 0$, and $\overline{w}_r = -\lambda \cdot \frac{w_r}{w_j} + \frac{v_r}{v_i} \geq 0$, for $r \notin \{i, j, n+1\}$. Then $\overline{w}' = (\overline{w}_1, \ldots, \overline{w}_n) \in C' \setminus \{0\}$, and $w'_j = 0$, in contradiction with $i \Longleftrightarrow j$ in N'. ∎

Fig. 2.3 helps to better understand why these additional hypotheses should be considered in Theorem 2.3. In Fig. 2.3A, one can see the case where all the assumptions of Theorem 2.3 are satisfied. In the incomplete network, i and j are fully coupled. However, after addition of reaction k, reactions i and j become directionally (and not fully coupled). Fig. 2.3B shows an example where there are some reversible reactions in the network. In this case, we have $i \Longleftrightarrow j$ in the incomplete network. If the prescribed reaction k is added to this network, however, we get $i \longrightarrow j$ and i and j are not partially coupled in the completion.

Fig. 2.3C shows an example where i and j are fully coupled in the incomplete network, but there exist some flux distribution v in which $v_i = v_j = 0$. Again, in the completion, we have $i \longrightarrow j$ and not $i \longleftrightarrow j$. Finally, Fig. 2.3D–E show examples in which there exist a flux distribution v such that $\{i, j\} = supp(v)$. Addition of reaction k again results in $i \longrightarrow j$, rather than a partial coupling relation in the completion.

The last theorem summarizes all possible cases.

Theorem 2.4. *Let N' be an incomplete metabolic network and let N be a completion. All possible changes in flux coupling relations that can occur for two unblocked reactions i, j are the following:*

1. $i \Longleftrightarrow j$ *in* N' *vs.* $i \longleftrightarrow j$ *in* N
2. $i \Longleftrightarrow j$ *in* N' *vs.* $i \longrightarrow j$ *(or* $j \longrightarrow i$*) in* N
3. $i \Longleftrightarrow j$ *in* N' *vs.* $i \overset{S.C.}{\longleftrightarrow} j$ *in* N
4. $i \longleftrightarrow j$ *in* N' *vs.* $i \longrightarrow j$ *(or* $j \longrightarrow i$*) in* N
5. $i \longleftrightarrow j$ *in* N' *vs.* $i \overset{S.C.}{\longleftrightarrow} j$ *in* N
6. $i \longrightarrow j$ *in* N' *vs.* $i \overset{S.C.}{\longleftrightarrow} j$ *in* N
7. $i \overset{M.E.}{\longleftrightarrow} j$ *in* N' *vs.* $i \overset{S.C.}{\longleftrightarrow} j$ *in* N

Proof. Theorem 2.2 and 2.3 show the possibility of all these seven changes in flux coupling relations. It remains to prove that no other changes can happen.

We first show that it is not possible to have a pair of uncoupled reactions in N' that become coupled in N. If i and j are uncoupled, then there exist elementary modes $f, g \in E'$ such that $f_i \neq 0, f_j = 0$ and $g_i = 0$ and $g_j \neq 0$. Now, if we add one or more reactions to the network, all EMs of the incomplete network will be included in the set of EMs in the completion. This means that not for all EMs e in the completion, $e_i \neq 0$ implies $e_j \neq 0$ (and also $e_j \neq 0$ implies $e_i \neq 0$). Hence, it is impossible for i and j to become coupled in the completion.

If $i \longrightarrow j$ and not $i \longleftrightarrow j$ holds in N', then there exists $e' \in E'$ such that $e'_j \neq 0$ and $e'_i = 0$. Therefore, after making a completion, $e_j \neq 0$ does not imply $e_i \neq 0$ for all EMs e in the completion. Thus, it is impossible to have $j \longrightarrow i$, $i \longleftrightarrow j$ and $i \Longleftrightarrow j$ in N.

If $i \longleftrightarrow j$ and not $i \Longleftrightarrow j$ holds in N', then there exist $f', g' \in E'$ such that $f'_j/f'_i \neq g'_j/g'_i$. Clearly, i and j cannot become fully coupled in the completion, because $f = (f', 0)$ and $g = (g', 0)$ are both included in the set E of elementary modes in N.

Finally, changing a pair of coupled or sometimes coupled reactions to a pair of mutually exclusive reactions in the completion is also impossible, since there exists at least one $e' \in E'$ and a corresponding $e = (e', 0) \in E$ such that $e_i \neq 0$ and $e_j \neq 0$. ∎

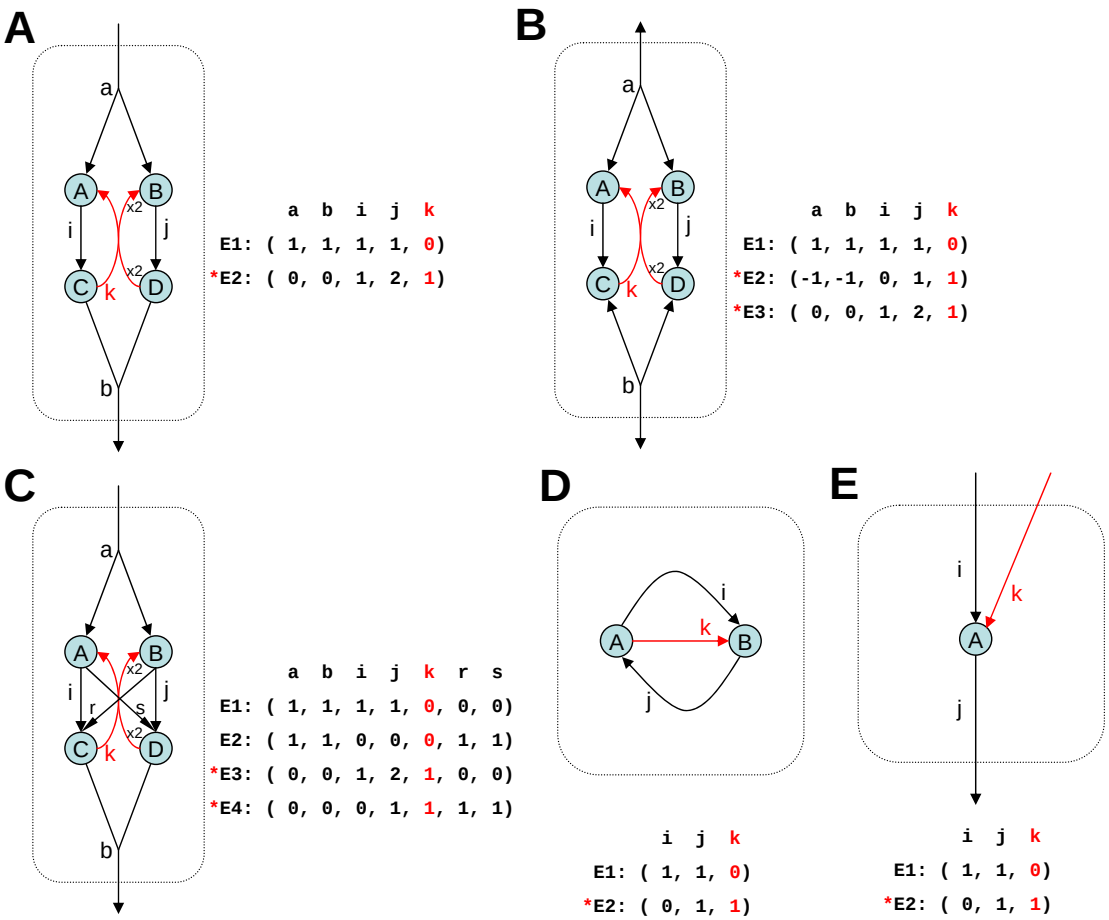


Figure 2.3: *Different networks when the prescribed reaction in Theorem 2.3 (shown here as reaction k) is added to the network. When the stoichiometric coefficient is not 1, the coefficient is indicated on the hyperarchs by "$\times$" sign. Elementary modes of each network are also indicated. Those EMs which appear only in the completion are asterisked. (A) A network in which all reactions are irreversible, and additionally, i and j are non-zero in all flux distributions, but there is no EM including only i and j in its support; (B) A network in which some reactions are reversible, and additionally, i and j are non-zero in all flux distributions, but there is no EM including only i and j in its support; (C) A network in which all reactions are irreversible, but i and j are zero in some flux distributions. (D) A network in which i and j form a (futile) cycle. (E) A network in which there is a (noncyclic) EM including only i and j in its support.*

Possible changes in flux coupling relations due to missing reactions are summarized in Fig. 2.4.

From these theorems we immediately conclude that the results of FCA are not always reliable. More specifically, if two reactions are uncoupled in an incomplete model, they are necessarily uncoupled in the original complete network. On the other hand, two reactions that are coupled in an incomplete network may be

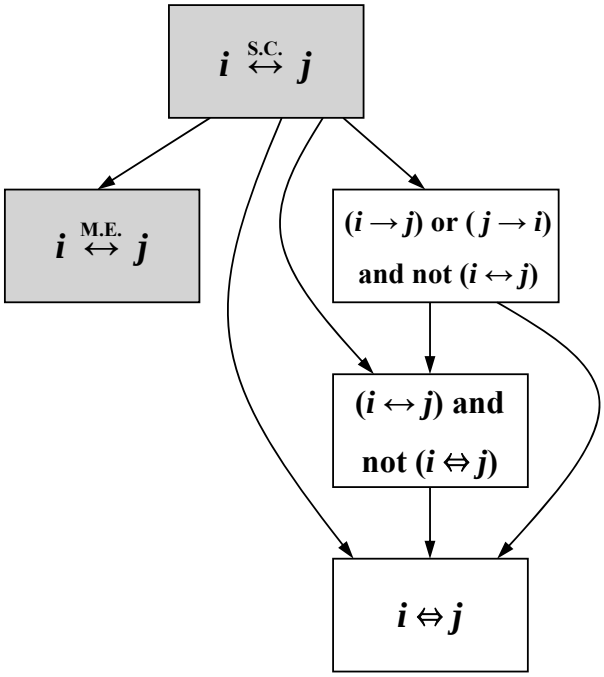


Figure 2.4: *Possible changes in flux coupling relations due to missing reactions. An arrow from relation A to relation B indicates that while relation A holds in the completion, relation B may be observed in the incomplete network. Grey boxes stand for uncoupling, white boxes for coupling relations.*

uncoupled in the complete network. Burgard et al. (2004) accurately report such examples, e.g. in case of ZWF and PGL reactions.

Note that Theorem 2.4 considers unblocked reactions. Obviously, a blocked reaction in an incomplete network may participate in different types of (un-)coupling reactions in the completion, depending on the additional reactions. Furthermore, if reaction i is directionally coupled to reaction j in the original complete network and j is missing in the incomplete reconstructed network, then the flux through i is always zero. This means that i is blocked in the incomplete network. If some reactions belong to the same set of partially or fully coupled reactions in the complete network, then one missing reaction in the reconstructed model results in all other reactions being blocked. This can formally explain the prevalence of blocked reactions in genome-scale metabolic networks (Feist et al., 2007; Puchałka et al., 2008; Suthers et al., 2009). If further constraints are applied to the network, e.g., if certain uptake reactions cannot work because their corresponding substrates are not available in the growth medium, we may have even more blocked reactions.

2.2.3 Biological Implications

How Important is the Effect of Missing Reactions?

As mentioned in the previous section, missing reactions in a metabolic network can influence flux coupling relations. How prevalent is this influence? In order to investigate this question, different metabolic networks were chosen. For each model, we temporarily removed one reaction from the complete network and counted how many reaction pairs changed their coupling type in the incomplete network (cf. Fig. 2.4). This was repeated for all reactions in the network. The results are summarized in Table 2.2. We can see that there exist reactions in the network whose omission results in little or no change. However, there exist also some reactions that if deleted, influence flux-coupling relations of many other reactions. We may call these reactions "important" because they play a pivotal role in shaping the coupling type of other reactions. Table 2.2 shows that such important reactions exist in all three models. Since the average ratio of the observed changes to the total number of reaction pairs is considerable, one can conclude that "important" reactions are not exceptional. Note that the set of these "important" reactions is not necessarily equal to the set of "hubs" in reaction maps (Burgard et al., 2004), which can be defined as reactions whose omission results in blocking many other reactions.

Model Name	# Reactions	# Reaction pairs	Min	Max	Average	Average ratio
Calv	23	231	0	119	54.7	0.24
RBC	44	903	0	440	77.4	0.09
EC core	76	2775	0	1011	248.0	0.09

Table 2.2: *Number of changes in flux coupling types due to removal of one reaction. # Reactions is the number of reactions in each network. # Reaction pairs is the number of reaction pairs after removing one reaction. In each iteration, one reaction is deleted from the complete network and the number of changed relations is determined. Min denotes the minimum number of changes among all reaction pairs, while Max is the corresponding maximum value. Only those pairs of reactions are considered that are both unblocked after the single reaction removal. Also shown is the average number of changes (average) and its proportion to the number of all pairs (average ratio).*

A possible explanation for the existence of "important" reactions is that all three models are greatly simplified. Therefore, many alternative pathways are removed from the network, resulting in a model in which many of the reactions are coupled to each other, and therefore more or less "important". In genome-scale models with many alternative pathways and a large number of uncoupled reaction pairs, we expect that "important" reactions will be less frequent, since numerous

alternative pathways and therefore many uncoupled reaction pairs exist in these networks. Here, individual removal of one reaction at a time will probably have little effect on the coupling relations between other reaction pairs.

In Table 2.3, the frequencies of different types of changes in flux coupling relations are compared to each other. The first four changes (columns 2-5) pertain to scenarios where, after removal of a single reaction, at least one reaction has become blocked in the incomplete network. The majority of observed changes fall into these categories. Second most frequent are changes relating to the transformation of an uncoupled pair to a directionally coupled pair. This shows again why the results of FCA in many cases may not be reliable.

Model	S⇒B	D⇒B	P⇒B	F⇒B	S⇒F	D⇒F	P⇒F	S⇒P	D⇒P	S⇒D	S⇒M	Total
Calv	1292	414	0	71	310	37	0	63	9	564	276	3036
RBC	8286	2734	35	1025	99	297	27	0	307	2634	41	15 485
EC core	27 104	9626	48	1004	583	892	72	143	582	6946	9628	56 628

Table 2.3: *Total number of changes in flux coupling relations due to removal of one reaction. $X \Rightarrow Y$ indicates that the coupling type changes from X to Y. In each iteration, one reaction is deleted from the complete network and the number of changes is determined. The numbers give the sum of changes over all reactions. S: sometimes coupled; M: mutually exclusive; D: directionally coupled; P: partially coupled; F: fully coupled; B: both reactions or at least one of them are blocked.*

2.2.4 Comparison of Flux Coupling Relations in two Versions of the *E. coli* Metabolic Network

In a more comprehensive study, we compared the flux coupling relationships of two versions of the *E. coli* metabolic model. The original model (Reed et al., 2003) contains 1075 reactions, while in the more recent model (Feist et al., 2007) the number has nearly doubled to 2077 with newly added reactions accounting for most of the differences between the two models. Only 784 reaction IDs can be found in both versions, constituting what we call the "common" set, while 291 reaction IDs from the original model seem to have no equivalent in the new model. On closer inspection, the majority actually do appear in the newer model but under a different ID. A number of reactions have been deleted in the 2007 model, either because they were incorrect or because they were decomposed into more discrete enzymatic steps (Feist et al., 2007). Since it is difficult to determine a possible "equivalent" for these 291 reactions, we limited our analysis to the "common" set. Flux coupling analysis was performed for both versions and the results for the "common" set were compared, see Fig 2.5. As expected, numerous changes in coupling types can be observed. Overall, in this analysis, we study $(784 \times 783)/2 = 306\,936$ reaction pairs. In the 2003 version of the *E. coli* network, in 144 477 pairs either one or both of the reactions are blocked. For these, we have

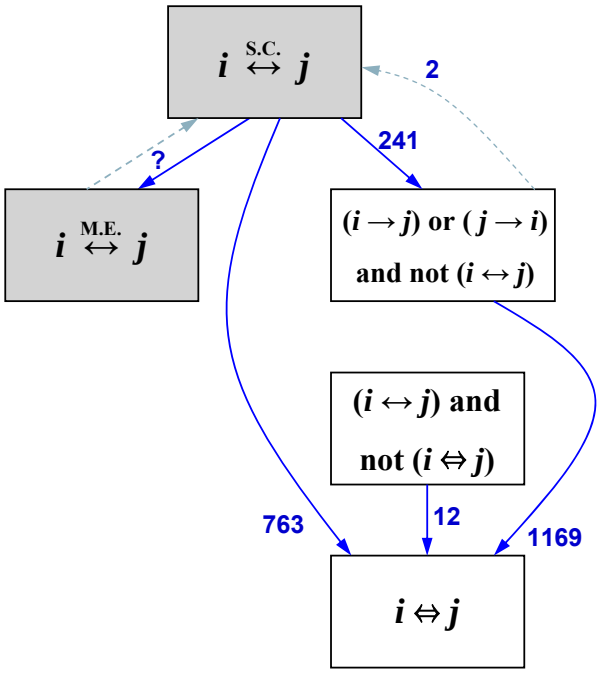


Figure 2.5: *Changes in flux (un-)coupling relations for two E. coli models. An arrow from relation A to relation B indicates that in the more recent reconstruction (Feist et al., 2007) relation A holds, while in the former model (Reed et al., 2003) relation B was observed. The numbers indicate how many reaction pairs have changed their relation. Note that the 2003 model does not contain partially coupled reactions that are not fully coupled. Grey boxes represent uncoupling relations, while white boxes show coupling relations. The dashed arrows correspond to changes that cannot be caused by missing reactions only. Since the FCF algorithm (Burgard et al., 2004) is used for FCA, it is not known how many S.C. relations are changed to M.E. (and possibly, vice versa).*

no coupling or uncoupling relation. The number of uncoupled pairs is 159 772. There are only 2687 fully or directionally and no partially coupled reaction pairs in this model. In the 2007 model, 2185 pairs (i.e., 12+1169+763+241 pairs) have changed their coupling type either to another type of coupling, or to uncoupling. This means that more than 81% of coupling relations have changed. Again, these results indicate the importance of using more complete metabolic models for drawing biological conclusions from the results of FCA.

From our mathematical results we can conclude that only certain changes in flux (un-)coupling relations are possible if reactions are merely added to a network. Since in addition to the many new reactions in the 2007 model, some reactions from the 2003 model were also omitted, we can observe two cases in which an uncoupling relation in the 2003 model becomes a directional coupling relation in the 2007 model.

Relationship between Gene Expression Correlation and Flux Coupling for a Pair of Reactions

Recently, Notebaart et al. (2008) demonstrated that gene expression correlation (GEC) between metabolic genes can be reasonably explained by flux coupling. They used the 2003 model of *E. coli* (Reed et al., 2003) to show that fully coupled, directionally coupled and uncoupled reaction pairs exhibit distinct GEC values in their corresponding gene expression levels. Although high GEC values are expected for enzymes that catalyze fully coupled reactions, in this study the average GEC values of these reactions was about 0.27, which is relatively low. We hypothesized that this low value is related to reaction pairs that are in fact directionally coupled or uncoupled, but were mistaken for fully coupled pairs due to some missing reactions. Therefore, we expected some of these "falsely" fully coupled reactions to become uncoupled or partially coupled in a more complete model.

To test our hypothesis, we compared the recent *E. coli* model (Feist et al., 2007) to the original model (Reed et al., 2003) used by Notebaart et al. (2008). We chose all reaction pairs identified as fully coupled by Notebaart et al. (2008). From these 947 reaction pairs, 933 pairs had a reported GEC value. For these, we re-calculated the flux coupling relations based on the recent *E. coli* metabolic model. Interestingly, 379 pairs became directionally coupled and 204 pairs became uncoupled in this model, while only 350 pairs remained fully coupled. Again, we observe that the results of flux coupling are not reliable in incomplete metabolic models.

Using the model of Feist et al. (2007), we compared gene expression levels of the following two categories: reaction pairs that are still fully coupled, and reaction pairs that have changed their coupling relationship to directionally coupled or uncoupled. Results are illustrated in Fig. 2.6. In Fig. 2.6A, the two categories are compared based on their average GEC values. The average GEC of reaction pairs that remained fully coupled is about 0.30, while it is less than 0.25 for directionally coupled or uncoupled reaction pairs. Although the difference is not big, it is statistically significant ($P < 0.01$). Reaction pairs that were "falsely" fully coupled in the smaller model are now in a category with significantly less correlation values in gene expression levels.

In the next step, based on their GEC values, we classified the gene pairs into five different categories: N: negatively correlated; U: uncorrelated; S: slightly correlated; M: moderately correlated; and H: highly correlated. For each category, the proportion of reaction pairs that are fully coupled in the recent *E. coli* model is given in Fig. 2.6B. In this diagram, the average GEC of the last category, i.e., highly correlated pairs, is significantly greater than the other categories ($P < 0.01$ for *H-N*, *H-U* and *H-S* comparisons; and $P < 0.05$ for *H-M* comparison). The pairwise difference in average GECs of the other categories is not significant ($P > 0.05$ in all cases). The results indicate that highly correlated genes have a greater chance of remaining fully coupled as the network becomes

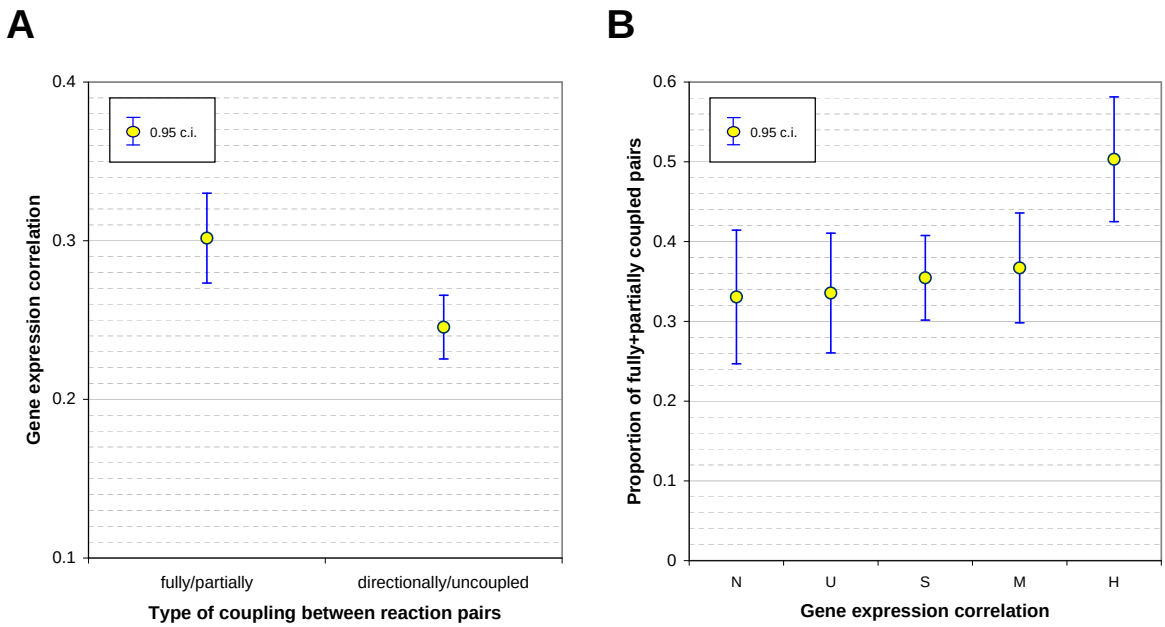


Figure 2.6: *Repeating FCA for fully coupled reaction pairs in E. coli reported by Notebaart et al. (2008). Since fully coupled reaction pairs in (Notebaart et al., 2008) are based on an "incomplete" E. coli model (Reed et al., 2003), we performed FCA on a more recent and more "complete" network of E. coli (Feist et al., 2007). (A) Comparing GEC distributions for reaction pairs that are still fully coupled in the recent E. coli model and reaction pairs that are uncoupled or directionally coupled in this model; (B) Comparing the proportion of fully coupled pairs in the recent E. coli model for reaction pairs with different levels of GEC. N: negatively correlated pairs with* $GEC \leq -0.10$*; U: uncorrelated pairs with* $-0.10 < GEC \leq 0.10$*; S: slightly correlated pairs with* $0.10 < GEC \leq 0.35$*; M: moderately correlated pairs with* $0.35 < GEC \leq 0.60$*; and H: highly correlated pairs with* $0.60 < GEC$*.*

more complete.

2.3 Methods

2.3.1 Datasets: Metabolic Networks and Gene Expression Correlations

Five metabolic networks are used in this study: *Calv*: photosynthate metabolic network in the chloroplast stroma (Poolman et al., 2003) (this model has been chosen from default metabolic networks associated with CellNetAnalyzer (Klamt et al., 2007) and differs slightly from the original model); *RBC*: metabolic network of red blood cell (Çakır et al., 2004b); *EC core*: central metabolic network of *E. coli* (without the "Biomass" reaction) Palsson (2006); the 2003 *E. coli model*

(Reed et al., 2003); and the 2007 *recent E. coli model* (Feist et al., 2007). The last three models are available from `http://gcrg.ucsd.edu/In_Silico_Organisms/E_coli`.

We used the flux coupling and gene expression correlation data reported by Notebaart et al. (2008) (kindly provided by B. Papp). The flux coupling data are related to the *E. coli* model published in 2003 (Reed et al., 2003). The gene expression correlation values are based on the data reported elsewhere (Price et al., 2006).

2.3.2 Study of Flux Coupling Changes by Removal of Single Reactions

In order to study the effect of missing reactions in a metabolic network, three small models, *Calv*, *RBC*, and *EC core*, were used. For each model, the following procedure was performed for all reactions in the network. First of all, the set of EMs of the network was computed by METATOOL (Pfeiffer et al., 1999). In the next step, one reaction was chosen in each iteration and the elementary modes in which this reaction has a non-zero flux were temporarily removed. Then, based on the remaining EMs, coupling relations between each pair of reactions were determined.

2.3.3 Comparison of Flux Coupling Relations in two Versions of the *E. coli* Metabolic Network

In order to investigate how flux coupling relations may change by adding new reactions, two versions of the *E. coli* genome-scale metabolic network were studied (Reed et al., 2003; Feist et al., 2007). Only reactions with identical reaction IDs in both models were considered. The FCF algorithm (Burgard et al., 2004) was applied to determine the flux coupling relations for any reaction pair in both networks. Then, the FCA results of both networks were compared.

2.3.4 Relationship between Gene Expression Correlations and Flux Coupling in a Recent *E. coli* Model

Notebaart et al. (2008) recently reported that fully coupled reaction pairs have positive, yet rather low, gene expression correlations. Their work was based on a metabolic model of *E. coli* published in 2003 (Reed et al., 2003). By applying the FCF algorithm (Burgard et al., 2004) on a more recent metabolic model of *E. coli* (Feist et al., 2007), we determined the flux coupling relations between those reaction pairs that were found to be fully coupled by Notebaart et al. (2008). Like in Notebaart et al. (2008), partially coupled reaction pairs are included in the same category as the fully coupled pairs. Moreover, similar modifications were applied to the 2007 *E. coli* network, e.g. we allow inflow for all external

metabolites and we remove the biomass reaction. Then, distribution of GEC values of those pairs that are fully coupled in both *E. coli* models were compared to GEC values of those pairs that are not fully coupled in the recent model. In addition, we classified these gene pairs based on their GEC values into five categories: negatively correlated (N), uncorrelated (U), slightly (S), moderately (M) and highly (H) correlated pairs (see Fig. 2.6). The proportion of reaction pairs that remained fully coupled in the recent model were then compared over the five categories.

2.3.5 Statistical Analysis

In order to compare the distributions of GECs (see Fig. 2.6A), a two-sample t-test was used. Confidence intervals in this plot are based on one-sample t-test. In order to compare two proportions (in Fig. 2.6B), we performed a "test of two binomial proportions" in each case. Since in all cases, the two population sizes were large enough, a normal approximation of the binomial distribution was used to simplify the comparisons (Hogg and Tanis, 1988), i.e., a Z-test on the following statistic (which is assumed to be normally distributed):

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} \tag{2.4}$$

where

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} \tag{2.5}$$

and where n_1 and n_2 are sample sizes, x_1 and x_2 are the frequency of "success" (i.e., fully coupled pairs) in samples 1 and 2, and $\hat{p}_1$ and $\hat{p}_2$ are estimated probabilities of "success" in the corresponding samples.

The corresponding 0.95 confidence intervals were also computed based on the normal approximation, i.e.:

$$\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/n} \tag{2.6}$$

where $\hat{p}$ is the proportion estimated from the statistical sample, and n is the population size.

Summary:

- We mathematically prove the "conservation property" of elementary modes (Lemma 2.1).

- Flux coupling relations are defined based on the set of vectors in the flux cone. We prove that flux coupling relations can be characterized based on the set of EMs in a similar way (Corollary 2.2).

- When reactions are missing in a reconstructed metabolic model, only certain changes in flux coupling relations are possible (Theorem 2.4).
- A general method is introduced to construct examples of partially coupled reactions (in the constructive proof of Theorem 2.3).
- We show that in two consecutive genome-scale metabolic networks of *E. coli*, most of flux coupling relations are different (because of the incompleteness of the older model).
- We show that GECs between enzyme pairs are better explained by FCA when the recent *E. coli* model is used instead of the older one.

Chapter 3 On Flux Coupling Analysis of Metabolic Subsystems

3.1 Background

Genome-scale metabolic networks are useful models for the analysis of metabolism at the systems level (Oberhardt et al., 2009). However, due to the existence of hundreds to thousands of reactions in each genome-scale network, it is not easy to analyze such networks. For this reason, some authors prefer to study only some interesting "subsystems" within these networks.

A mathematical definition of subsystems will be presented in the next section. Informally speaking, a subsystem can be chosen by "cutting out" the uninteresting components, i.e., by redrawing the boundary of the network to include only a subset of reactions and metabolites (see Figure 3.1). By choosing a new boundary, some of the internal reactions are converted to exchange reactions of the subsystem (i.e., reaction 3 in Figure 3.1). Other reactions may become "external" (like reactions 6 to 9 in Figure 3.1). External reactions convert only external metabolites to each other.

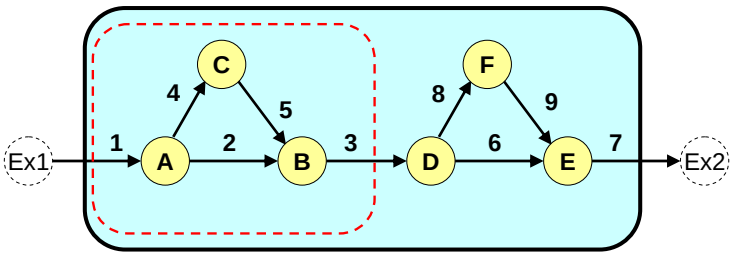


Figure 3.1: *A metabolic network with nine reactions and six internal metabolites. The system boundary is shown as a solid black line. A subsystem of the original network can be selected by including only three metabolites (A–C) and five reactions (1–5). The new boundary is shown as dashed line. Note that reaction 3 is a boundary reaction in the selected subsystem, while it is an internal reaction in the original network.*

Subsystems may appear naturally in biological networks. For example, due to compartmentalization of eukaryotic cells, organelle networks can be considered as subsystems within the whole-cell network. Many authors prefer to study subnetworks, e.g. metabolic networks of eukaryotic organelles in isolation (Poolman

et al., 2003; Vo et al., 2004). Similarly, some authors have studied other subsystems in isolation (e.g. see Teusink et al., 2009). A related approach is to split a genome-scale metabolic network into smaller subsystems and study them in isolation (Schilling and Palsson, 2000; Schilling et al., 2002; Schwarz et al., 2005; Verwoerd, 2010, 2011).

Analysis of metabolic subsystems has the advantage that considerable savings can be achieved in computation time. However, this "simplification" of the network may sometimes result in wrong computational conclusions. For example, instead of the original elementary modes of the network, one may obtain "pathway fragments" in a subsystem (Imielinski and Belta, 2008), which may not be part of any steady-state flux distribution in the original network (Kaleta et al., 2009). Additionally, interdependencies between pairs of fluxes, which can be determined by flux coupling analysis (FCA), can be missed when subsystems are analyzed. It has been previously shown that FCA of subsystems results in smaller sets of flux-coupled reactions compared to the FCA in genome-scale networks (see Burgard et al., 2004, page 308).

In the present chapter, we focus on subsystem-based vs. genome-scale flux coupling analysis. We first formally introduce the concepts used in this study. Then, we mathematically explore the effect of subsystem selection on FCA.

3.2 Formal Definition of the Concepts

3.2.1 Concepts from Linear Algebra

The i-th *unit vector* $\mathbf{e}_i$ is defined as a column vector with element i equals one and all other elements equal zero, i.e., $\mathbf{e}_i = (0, \cdots, 0, 1, 0, \cdots, 0)$, where only the i-th element is 1. Therefore, for a column vector v, we have: $\mathbf{e}_i^T \cdot v = v_i$, which is the i-th element of v.

A *zero vector* $\mathbf{0}$ is defined as a column vector with all elements equal to zero. Additionally, $[n]$ will be used as an abbreviation for the set $\{1, \ldots, n\}$.

For an $m \times n$ matrix M, $P \subseteq [m]$ and $Q \subseteq [n]$, we denote by $M_{P;Q}$ the submatrix of M induced by rows in P and columns in Q.

3.2.2 Concepts Related to Metabolic Networks

An irreversible boundary reaction which can produce internal metabolites is called an *uptake* reaction. Reactions that both produce and consume only external metabolites are considered external to the metabolic network and are not included in its stoichiometric matrix.

We adopt the definition of subsystems suggested in Imielinski and Belta (2008). Suppose that in a network $N = (S, Irr)$, the set of metabolites is equal to $[m]$. Without loss of generality, suppose that the stoichiometric matrix S^* of a network $N^* = (S^*, Irr)$ is obtained by deleting the set of rows $\{r+1, \ldots, m\}$

(with $1 \leq r < m$) from S . We will refer to the second network N^* as the *subsystem* and to the first network N as the *extension* or the *extended network*. By selecting a subsystem, it is possible to get a set of zero columns, Z. Such columns in the stoichiometric matrix correspond to those reactions that were internal in the extended network, but are considered external to the subsystem. By convention, one can redefine the stoichiometric matrix of the subsystem $S^\star = S_{[r];[n]\setminus Z}$, which is an $r \times (n - |Z|)$ matrix in which only the non-zero columns of S^* are kept. $Irr^\star = Irr \setminus Z$ is the subset of Irr including the irreversible reactions in the subsystem.

3.3 Results and Discussion

In this section, we first observe how selecting a subsystem (or equivalently, redrawing network boundaries) can affect the results of flux coupling analysis. We then mathematically study the impact of subsystem selection on flux coupling relations. Afterwards, we briefly compare the effect of deleting reactions (cf. Chapter 2) with subsystem selection. Finally, we study the relevance of our results for the analysis of real metabolic networks.

3.3.1 Impact of Redrawing the Network Boundaries on FCA

Figure 3.2A shows a metabolic network. In Appendix 3.1, the corresponding METATOOL input file is presented. Using the METATOOL software (Pfeiffer et al., 1999; von Kamp and Schuster, 2006), only one EM (which includes all reactions) is found in this network. Therefore, reactions i and j (like other reactions) are fully coupled.

In Appendix 3.1, a number of scenarios are shown, where different subsystems are selected. When the subsystem in Figure 3.2B is chosen, metabolites A and B are assumed as external. In this subsystem, two EMs are found, and reactions i and j are partially coupled to each other. Choosing the subsystem in Figure 3.2C results in a system with three EMs, where j is directionally coupled to i. If the network boundaries are redrawn to include the subsystem in Figure 3.2D, i and j become uncoupled, or more precisely, sometimes coupled. Finally, keeping only metabolites C and H in the subsystem will result in Figure 3.2E, where i and j are mutually exclusive.

Figure 3.3A shows another small example network. In this network no reaction consumes metabolite D. As a result, reactions 3 and 4 are blocked under steady-state conditions and reactions i and j are mutually exclusive. However, if we select the subsystem shown in Figure 3.3B, metabolite D is considered as an external metabolite. In this case, reactions 3 and 4 are not blocked. Therefore, $i \overset{S.C.}{\longleftrightarrow} j$ in this subsystem. In Appendix 3.2, the METATOOL input files of the two networks are presented.

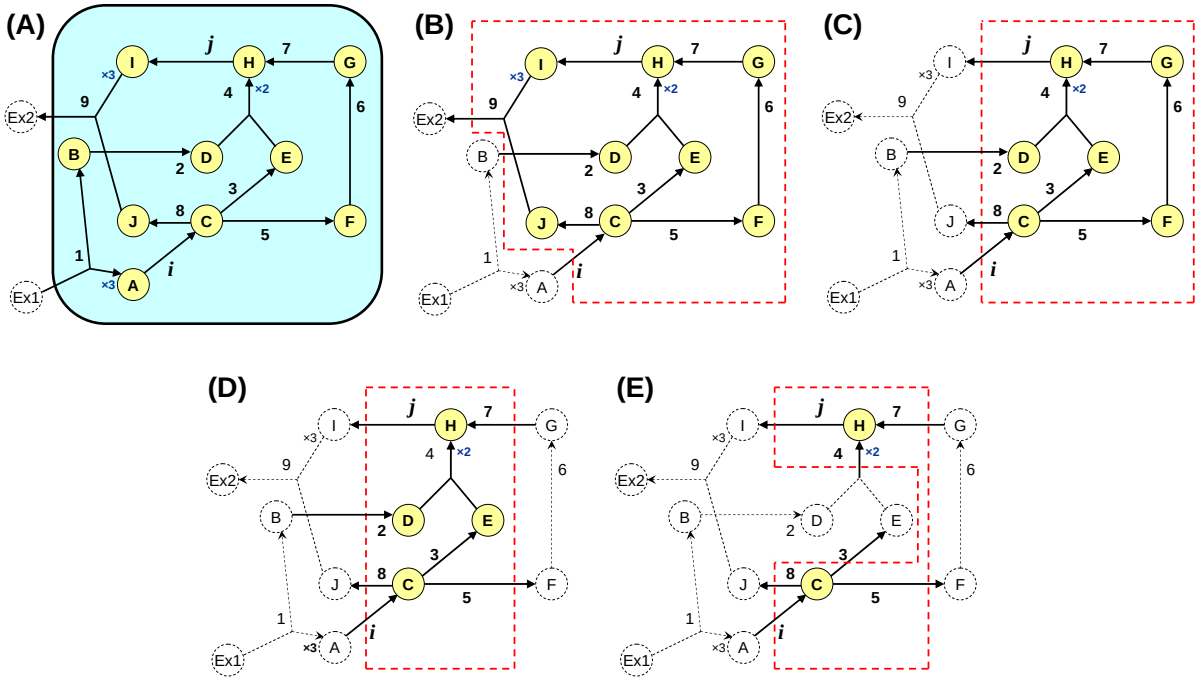


Figure 3.2: *Flux coupling between reactions i and j depends on subsystem selection. (A) In the original network, the network boundary is shown as solid line. In this network, $i \Longleftrightarrow j$. However, depending on the selected subsystem boundaries (dashed lines), different flux coupling relations are observed: (B) $i \longleftrightarrow j$; (C) $j \longrightarrow i$; (D) $i \overset{S.C.}{\longleftrightarrow} j$; (E) $i \overset{M.E.}{\longleftrightarrow} j$. See Appendix 3.1 for more details.*

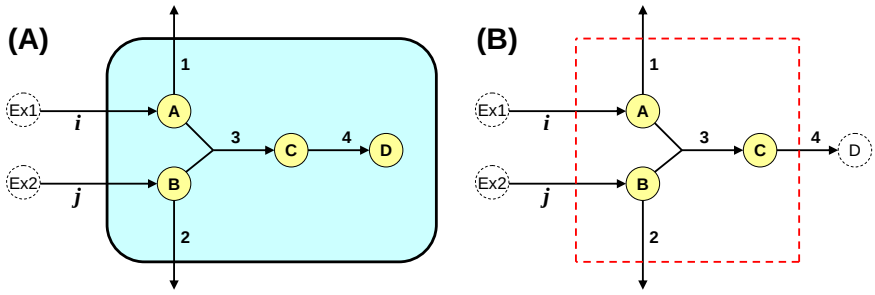


Figure 3.3: *(A): A metabolic network with six reactions and four internal metabolites. The system boundary is shown as a solid black line. In this network, we have $i \overset{M.E.}{\longleftrightarrow} j$; (B): A subsystem of the network is selected. The new boundary is shown as dashed line. In this subsystem, we have $i \overset{S.C.}{\longleftrightarrow} j$. See Appendix 3.2 for more details. This figure can be seen also as a schematic representation for understanding Theorem 3.2: N^* is the network in (B); $\Lambda = \{4\}$; $supp(d) = \{i, j, 3, 4\}$; $supp(f) = \{i, 1\}$; $supp(g) = \{j, 2\}$; and the extension N, is shown in (A).*

3.3.2 Mathematical Analysis of Flux Coupling Relations in Subsystems

In the previous section, we observed that it is possible to have certain changes in flux coupling relation of two reactions, depending on the selection of the subsys-

tem. The following theorems summarize these changes:

Theorem 3.1. *For any pair of unblocked uptake reactions i and j in a metabolic network N^*, the following holds:*

1. *There exists an extension N in which $i \Longleftrightarrow j$.*
2. *If i and j are not fully coupled in N^*, then there exists an extension N in which $i \longleftrightarrow j$ and not $i \Longleftrightarrow j$.*
3. *If i and j are uncoupled in N^*, then there exists an extension N in which $j \longrightarrow i$ (or $i \longrightarrow j$) holds.*
4. *If i and j are mutually exclusive in N^*, then there exists an extension N in which i and j are sometimes coupled.*

Proof. We denote by C^* the flux cone of N^*. Let $m+1$ and $m+2$ be the two distinct (external) metabolites consumed by reactions i and j, respectively.

1. Since i and j are unblocked, there exists $u, v \in C^*$ such that $u_i > 0$ and $v_j > 0$. Since i and j are irreversible, for $\widetilde{w} = u + v$, we have $\widetilde{w}_i, \widetilde{w}_j > 0$. Hence, there exists a constant $\lambda \in \mathbb{R}$ such that $\widetilde{w}_i = \lambda \widetilde{w}_j \neq 0$.

   We construct an extension N by adding an irreversible reaction $n+1$ that produces $m+1$ and $m+2$ (see Figure 3.4A). The flux cone of N is defined as:

   $$C = \{v \in \mathbb{R}^{n+1} \mid S \cdot v = 0, v_i \geq 0, \text{ for all } i \in Irr\} \tag{3.1}$$

   where $Irr = Irr^* \cup \{n+1\}$, and

   $$S = \begin{pmatrix} S^* & \mathbf{0} \\ -\mathbf{e}_i^T & \widetilde{w}_i \\ -\mathbf{e}_j^T & \widetilde{w}_j \end{pmatrix}$$

   Note that $\begin{pmatrix} \widetilde{w} \\ 1 \end{pmatrix} \in C$. Hence, i, j and $n+1$ are unblocked reactions in N. Additionally, $Sv = 0$ implies that $-v_i + \widetilde{w}_i v_{n+1} = 0$ and $-v_j + \widetilde{w}_j v_{n+1} = 0$. Thus, for all $v \in C$, $v_j > 0$ implies $v_i > 0$ (and $v_{n+1} > 0$) and vice versa, with $v_i / v_j = (\widetilde{w}_i v_{n+1}) / (\widetilde{w}_j v_{n+1}) = \lambda$. Therefore, $i \Longleftrightarrow j$.

2. If i and j are unblocked and not fully coupled, then there exists $\lambda_1, \lambda_2 \in \mathbb{R}$ with $\lambda_1 \neq \lambda_2$ and flux distributions $\widetilde{w}, \widetilde{u} \in C^*$ such that $\widetilde{w}_i = \lambda_1 \widetilde{w}_j > 0$ and $\widetilde{u}_i = \lambda_2 \widetilde{u}_j > 0$.

   We construct an extension N by adding two irreversible reactions $n+1$ and $n+2$ which both produce $m+1$ and $m+2$ but with different stoichiometric coefficients (see Figure 3.4B). The flux cone of N is defined as:

   $$C = \{v \in \mathbb{R}^{n+2} \mid S \cdot v = 0, v_i \geq 0, \text{ for all } i \in Irr\} \tag{3.2}$$

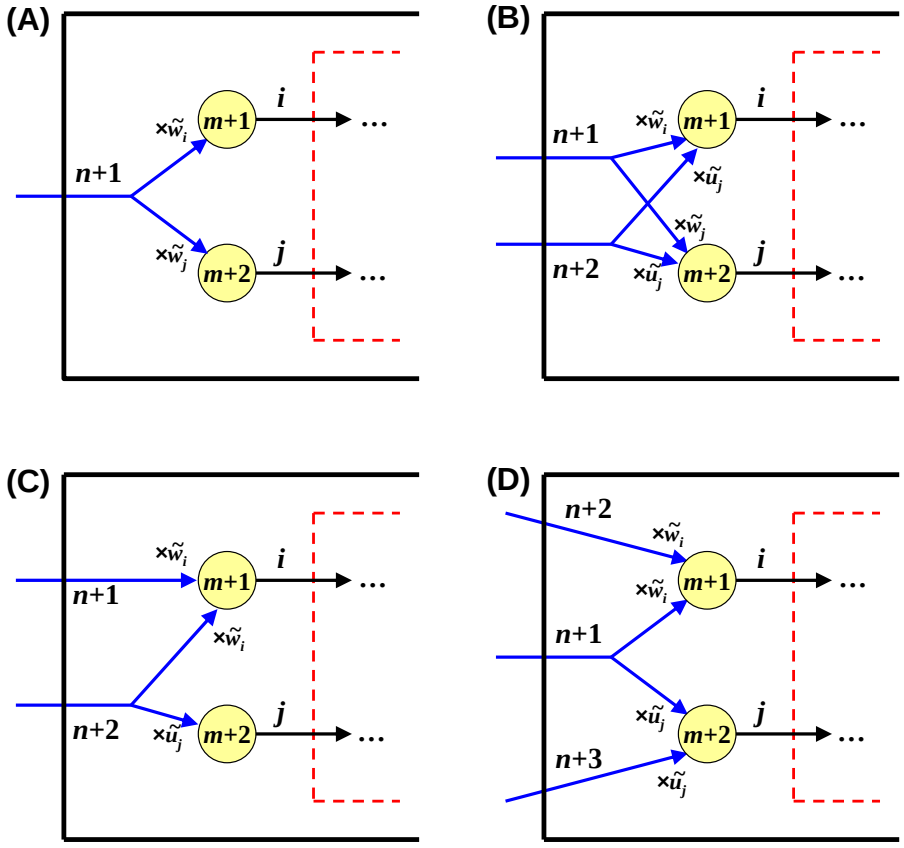


Figure 3.4: *(A-D): Schematic representations for understanding parts 1-4 of Theorem 3.1. The dashed line represents the boundary of the subsystem, while the solid line represents the boundary of the extension. In each case, i and j are uptake reactions in the subsystem, while they become internal reactions in the extension network.*

where $Irr = Irr^* \cup \{n+1, n+2\}$, and

$$S = \begin{pmatrix} S^* & \mathbf{0} & \mathbf{0} \\ -\mathbf{e}_i^T & \widetilde{w}_i & \widetilde{u}_i \\ -\mathbf{e}_j^T & \widetilde{w}_j & \widetilde{u}_j \end{pmatrix}$$

First, note that $\begin{pmatrix} \widetilde{w} \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} \widetilde{u} \\ 0 \\ 1 \end{pmatrix} \in C$. Thus i, j, $n+1$ and $n+2$ are unblocked reactions in N. Additionally, since $\widetilde{w}_i/\widetilde{w}_j \neq \widetilde{u}_i/\widetilde{u}_j$, i and j cannot be fully coupled. Now we prove that for all $v \in C$, $v_i = 0$ implies $v_j = 0$ and vice versa. Since $Sv = 0$, we have $-v_i + v_{n+1}\widetilde{w}_i + v_{n+2}\widetilde{u}_i = 0$ and $-v_j + v_{n+1}\widetilde{w}_j + v_{n+2}\widetilde{u}_j = 0$. If $v_i = 0$, we have $v_{n+1}\widetilde{w}_i + v_{n+2}\widetilde{u}_i = 0$. However, $\widetilde{w}_i, \widetilde{u}_i > 0$ and $v_{n+1}, v_{n+2} \geq 0$, which means that both v_{n+1} and v_{n+2} must be zero. Hence, $v_j = 0$. Similarly, for all $v \in C$, $v_j = 0$ implies

$v_i = 0$. Therefore, $i \longleftrightarrow j$.

3. If i and j are unblocked and uncoupled, then there exists $\widetilde{w}, \widetilde{u} \in C^*$ such that $\widetilde{w}_i > 0$, $\widetilde{w}_j = 0$ and $\widetilde{u}_i = 0$, $\widetilde{u}_j > 0$.

   We construct an extension N by adding two irreversible reactions $n+1$ and $n+2$. The first reaction only produces $m+1$, while the second reaction produces both $m+1$ and $m+2$ (see Figure 3.4C). The flux cone of N is defined as:

$$C = \{v \in \mathbb{R}^{n+2} \mid S \cdot v = 0, v_i \geq 0, \text{ for all } i \in Irr\} \tag{3.3}$$

   where $Irr = Irr^* \cup \{n+1, n+2\}$, and

$$S = \begin{pmatrix} S^* & \mathbf{0} & \mathbf{0} \\ -\mathbf{e}_i^T & \widetilde{w}_i & \widetilde{w}_i \\ -\mathbf{e}_j^T & 0 & \widetilde{u}_j \end{pmatrix}$$

   Note, that $\begin{pmatrix} \widetilde{w} + \widetilde{u} \\ 0 \\ 1 \end{pmatrix} \in C$. Therefore, i, j and $n+2$ are unblocked. Moreover, for all $v \in C$, we have $-v_i + \widetilde{w}_i(v_{n+1} + v_{n+2}) = 0$ and $-v_j + \widetilde{u}_j v_{n+2} = 0$. From these two equations, it can be seen that $v_j > 0$ implies $v_i > 0$. Therefore, $j \longrightarrow i$. However, since $\begin{pmatrix} \widetilde{w} \\ 1 \\ 0 \end{pmatrix} \in C$, $i \longrightarrow j$ cannot hold.

4. If i and j are unblocked and uncoupled in N^*, then there exists $\widetilde{w}, \widetilde{u} \in C^*$ such that $\widetilde{w}_i > 0$, $\widetilde{w}_j = 0$ and $\widetilde{u}_i = 0$, $\widetilde{u}_j > 0$.

   We extend the network by adding three new reactions $n+1$, $n+2$ and $n+3$ to N^*, as shown in Figure 3.4D. The extended network, N, has the following flux cone:

$$C = \{v \in \mathbb{R}^{n+3} \mid S \cdot v = 0, v_i \geq 0, \text{ for all } i \in Irr\} \tag{3.4}$$

   where $Irr = Irr^* \cup \{n+1, n+2, n+3\}$, and

$$S = \begin{pmatrix} S^* & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{e}_i^T & \widetilde{w}_i & \widetilde{w}_i & 0 \\ -\mathbf{e}_j^T & \widetilde{u}_j & 0 & \widetilde{u}_j \end{pmatrix} \tag{3.5}$$

   Note that, $\begin{pmatrix} \widetilde{w} + \widetilde{u} \\ 1 \\ 0 \\ 0 \end{pmatrix} \in C$. As a result, reactions i, j and $n+1$ are unblocked. It can be easily seen that for all $v \in C$, we have $-v_i + \widetilde{w}_i(v_{n+1} + v_{n+2}) = 0$

and $-v_j + \widetilde{u}_j(v_{n+1} + v_{n+3}) = 0$. Therefore, for all $v \in C$, $v_{n+1} > 0$ implies that $v_i, v_j > 0$. Thus, there exists an EM, f, with $f_i, f_j > 0$. Additionally, $\begin{pmatrix} \widetilde{w} \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} \widetilde{u} \\ 0 \\ 0 \\ 1 \end{pmatrix} \in C$. Therefore, there exists two EMs, say g, h, with $g_i > 0, g_j = 0$ and $h_i = 0, h_j > 0$. We conclude that i and j are sometimes coupled.

■

Theorem 3.2. *Let i and j are a pair of unblocked uptake reactions with $i \overset{S.C.}{\longleftrightarrow} j$ in a metabolic network N^*. Suppose there exists a subset Λ of boundary reactions that satisfies the following three conditions: (a) for each elementary mode d, with $d_i, d_j \neq 0$, there exists $r \in \Lambda$ such that $d_r \neq 0$; (b) there exists an EM f in N^* with $f_i \neq 0, f_j = 0$ such that $f_r = 0$ for all $r \in \Lambda$; and (c) there exists an EM g in N^* with $g_i = 0, g_j \neq 0$, such that $g_r = 0$ for all $r \in \Lambda$. Then, there exists an extension N in which $i \overset{M.E.}{\longleftrightarrow} j$.*

Please see Figure 3.3 for an illustrative example.

Proof. We denote by C^* the flux cone of N^*. We extend the network by adding a new metabolite for each reaction in $\Lambda = \{r_1, \cdots, r_{|\Lambda|}\}$. The extended network, N, has the following flux cone:

$$C = \{v \in \mathbb{R}^n \mid S \cdot v = 0, v_i \geq 0, \text{ for all } i \in Irr^*\} \tag{3.6}$$

where

$$S = \begin{pmatrix} S^* \\ -\mathbf{e}_{r_1}^T \\ \cdots \\ -\mathbf{e}_{r_{|\Lambda|}}^T \end{pmatrix}$$

Therefore, $C = C^* \cap \{v \in \mathbb{R}^n \mid v_r = 0 \text{ for all } r \in \Lambda\}$. Based on conservation property of elementary modes (see Lemma 2.1), the EMs of N are a subset of EMs of N^*. Additionally, $f, g \in C$. However, for each elementary mode d of N^*, if $d_i, d_j \neq 0$ then there exists $r \in \Lambda$ such that $d_r \neq 0$, which implies that $d \notin C$. We conclude that $i \overset{M.E.}{\longleftrightarrow} j$.

■

The possible changes in flux coupling relations due to subsystems selection (Theorems 3.1 and 3.2) are summarized in Figure 3.5.

Theorems 3.1 and 3.2 discuss the flux coupling relation changes only for a pair of boundary reactions i and j. However, these subsystem-induced changes are the only possible changes that can happen (independent of whether the reactions i and j are boundary reactions or internal reactions). This is shown in Theorem 3.3.

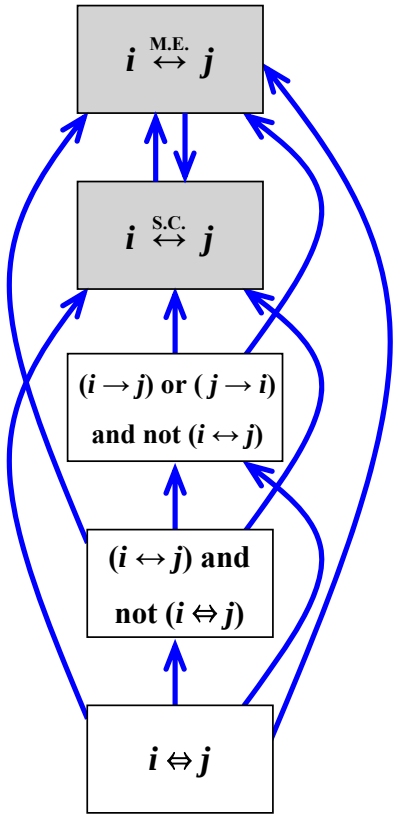


Figure 3.5: *Possible changes in flux coupling relations due to subsystem selection. An arrow from relation A to relation B indicates that while relation A holds in the extended network, relation B may be observed in the subsystem. Gray and white boxes indicate uncoupling and coupling relations, respectively.*

Theorem 3.3. *Let i and j be reactions in a metabolic subsystem N^*, and let N be an extension of N^*. All possible changes in flux coupling relations which can occur for a pair of unblocked reactions i, j are the following:*

1. $i \Longleftrightarrow j$ *in* N *vs. any other flux (un)coupling relation in* N^*
2. $i \longleftrightarrow j$ *in* N *vs.* $i \longrightarrow j$ *(or* $j \longrightarrow i$*) in* N^*
3. $i \longleftrightarrow j$ *in* N *vs.* $i \overset{S.C.}{\longleftrightarrow} j$ *in* N^*
4. $i \longleftrightarrow j$ *in* N *vs.* $i \overset{M.E.}{\longleftrightarrow} j$ *in* N^*
5. $i \longrightarrow j$ *(or* $j \longrightarrow i$*) in* N *vs.* $i \overset{S.C.}{\longleftrightarrow} j$ *in* N^*
6. $i \longrightarrow j$ *(or* $j \longrightarrow i$*) in* N *vs.* $i \overset{M.E.}{\longleftrightarrow} j$ *in* N^*
7. $i \overset{S.C.}{\longleftrightarrow} j$ *in* N *vs.* $i \overset{M.E.}{\longleftrightarrow} j$ *in* N^*
8. $i \overset{M.E.}{\longleftrightarrow} j$ *in* N *vs.* $i \overset{S.C.}{\longleftrightarrow} j$ *in* N^*

Proof. The flux cone of the subsystem can be defined as

$$C^* = \{v \in \mathbb{R}^n \mid S^* \cdot v = 0, v_i \geq 0, \text{ for all } i \in Irr^*\} \tag{3.7}$$

Let N be an extension of N^* obtained by adding q metabolites (and possibly, r reactions). Then, N has the following flux cone:

$$C = \{v \in \mathbb{R}^{n+r} \mid S \cdot v = 0, v_i \geq 0, \text{ for all } i \in Irr\} \tag{3.8}$$

where

$$S = \begin{pmatrix} S^* & \mathbf{0} \\ M & P \end{pmatrix}$$

with $M \in \mathbb{R}^{q \times n}$ and $P \in \mathbb{R}^{q \times r}$. Clearly, for all $u \in \mathbb{R}^n$ and $v \in \mathbb{R}^r$, $\begin{pmatrix} u \\ v \end{pmatrix} \in C$ implies $u \in C^*$. Now, we show that the only possible changes in the flux coupling relations are the ones mentioned in the theorem:

i) If i and j are uncoupled in N, then there exists $\begin{pmatrix} u \\ v \end{pmatrix} \in C$ such that $u_i = 0$ and $u_j \neq 0$. Also, there exists $\begin{pmatrix} u' \\ v' \end{pmatrix} \in C$ such that $u'_i \neq 0$ and $u'_j = 0$. Therefore, i and j cannot be directionally, partially, or fully coupled in N^*.

ii) If $i \longrightarrow j$ (and not $i \longleftrightarrow j$) in N, then there exists $\begin{pmatrix} u \\ v \end{pmatrix} \in C$ such that $u_i = 0$ and $u_j \neq 0$. Therefore, i and j cannot be partially, or fully coupled in N^*. Similarly, $j \longrightarrow i$ (and not $i \longleftrightarrow j$) in N implies the same result in N^*.

iii) If $i \longleftrightarrow j$ (and not $i \Longleftrightarrow j$) in N, then there exists $\begin{pmatrix} u \\ v \end{pmatrix}, \begin{pmatrix} u' \\ v' \end{pmatrix} \in C$ such that $u_i, u_j, u'_i, u'_j \neq 0$ and $u_i/u_j \neq u'_i/u'_j$. Therefore, i and j cannot be fully coupled in N^*.

■

3.3.3 Subsystem Selection vs. Reaction Deletion

In Chapter 2, we showed that flux coupling relations can change if some reaction(s) are deleted from the network. Deleting a reaction (or equivalently, existence of a missing reaction) is equivalent to deletion/absence of a column from the stoichiometric matrix (see Section 2.2.2). It was shown that pairs of uncoupled reactions in the original network may become coupled in the resulting incomplete network (but not vice versa). Additionally, deletion of reactions results in a decrease in the number of EMs.

In the present chapter, we observed that restriction to metabolic subsystems has very different consequences. Choosing a subsystem is equivalent to deletion of

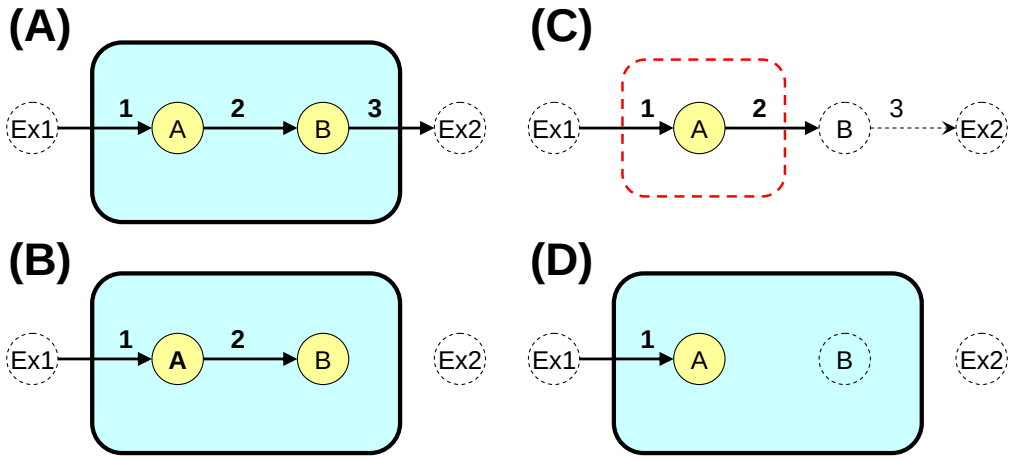


Figure 3.6: *Reaction deletion vs. subsystem selection. (A) A small metabolic network with three reactions and two internal metabolites; (B) the same metabolic network, when reaction 3 is deleted (or missing); (C) the same metabolic network, when metabolite B (and reaction 3) are assumed to be external; (D) the same metabolic network, when reaction 2 and 3 (which produce and consume metabolite B) are deleted.*

a subset of rows from the stoichiometric matrix. In a subsystem, coupled reaction pairs in the original network may become uncoupled in the resulting subsystem (but not vice versa). Moreover, the number of EMs in a subsystem can be greater than the number of EMs in the original network.

It is important to notice that subsystem selection and reaction deletion are very different concepts. When a reaction is not included in the selected subsystem, this is *not* equivalent to "deleting" the reaction from the network. Figure 3.6 shows an example. One can see here that deleting reaction 3 from the network in Figure 3.6A blocks all other reactions in steady-state (Figure 3.6B), while considering a subsystem excluding metabolite B and reaction 3 does not have the same effect (Figure 3.6C). This is due to the fact that by deleting a reaction, the corresponding internal metabolites are not deleted, and therefore, they can become unbalanced (Dandekar et al., 2003). Consequently, such metabolites will become "dead-ends", and the associated reactions will be blocked. Note that considering the above subsystem (by excluding metabolite B and reaction 3) is also different from deleting all reactions that are involved with metabolite B (Figure 3.6D).

3.3.4 Biological Implications

It is believed that some organelles like mitochondria and plastids originated from free-living bacteria which were endosymbionts of the ancestral eukaryotic cells (Sagan, 1967; Gross and Bhattacharya, 2009). These organelles are enclosed by two membranes, which strictly control the inflow and outflow of metabolites

and proteins. Therefore, one may expect the metabolic subsystems of plastids and mitochondria to be almost independent from the rest of the network. As mentioned in Section 3.1, some authors have studied these subsystems in isolation (e.g. Poolman et al., 2003; Vo et al., 2004; Urbanczik, 2007).

In contrast, one may argue that only a limited number of the proteins in mitochondria and plastids are encoded in their genomes. In fact, most of these proteins are encoded in the nuclear genome, synthesized in cytosol, and then transported to these organelles. There is a possibility that during evolution, some of these enzymatic functions are replaced by their cytosolic counterparts. As a result, the metabolism in these organelles might be highly interconnected to the cytosolic enzymatic activities. Therefore, the question is: can we study the organelle subsystems in isolation, without (much) influencing the dependencies and couplings among the fluxes?

In order to answer this question, we studied eight organelles (see Section 3.4.1). Plastid subsystems were selected from the genome-scale metabolic networks of *Hordeum vulgare* (barley), *Arabidopsis thaliana* and *Chlamydomonas reinhardtii*. Moreover, five mitochondrial subsystems were selected from the genome-scale metabolic network of barley, *A. thaliana*, *C. reinhardtii*, *homo sapiens* (human) and *Saccharomyces cerevisiae* (baker's yeast). We considered all pairs of unblocked reactions in each subsystem, and computed pairwise flux coupling relations: (i) when the genome-scale network is analyzed; and (ii) when the isolated subsystem is analyzed. The results of cases (i) and (ii) are compared in Figure 3.7.

Table 3.1 shows the frequencies of changes in flux coupling relations due to analysis of organelle subsystems instead of the complete network. The ratio of the changed flux coupling relations to the total coupling relations is shown in this table. This ratio can be seen as a "measure" of the independence of the metabolic subsystem: very small ratio means that most of the flux coupling relations are not changed, and therefore, isolation of the subsystem does not much change the functional dependencies of the metabolic fluxes. On the other hand, an increased ratio means that a higher number of functional dependencies of the metabolic fluxes are changed due to analysis of the subsystem in isolation, and therefore, the subsystem is more dependent on the fluxes outside the subsystem.

From this table, one can see that the mitochondrial subsystems are relatively independent of the rest of the metabolic networks, with change ratios ranging between 0.1% to 4.6%. This observation suggests that mitochondrial metabolic functions do not depend much on the "extra-organelle" metabolic fluxes. Thus, these subsystems can be studied in isolation without losing much information about the functional dependencies. In the plastids of modern plants, however, the change ratio is much higher, $11.2\% - 22.7\%$. The change ratio is very low in case of *C. reinhardtii*, which is a very simple single-celled green alga. This observation suggests that the analysis of plastid subsystems in isolation can result in the loss of functional dependencies between metabolic fluxes.

The discrepancy between the level of independence in the plastids and mitochondria subsystems might be simply an artefact of metabolic network reconstruc-

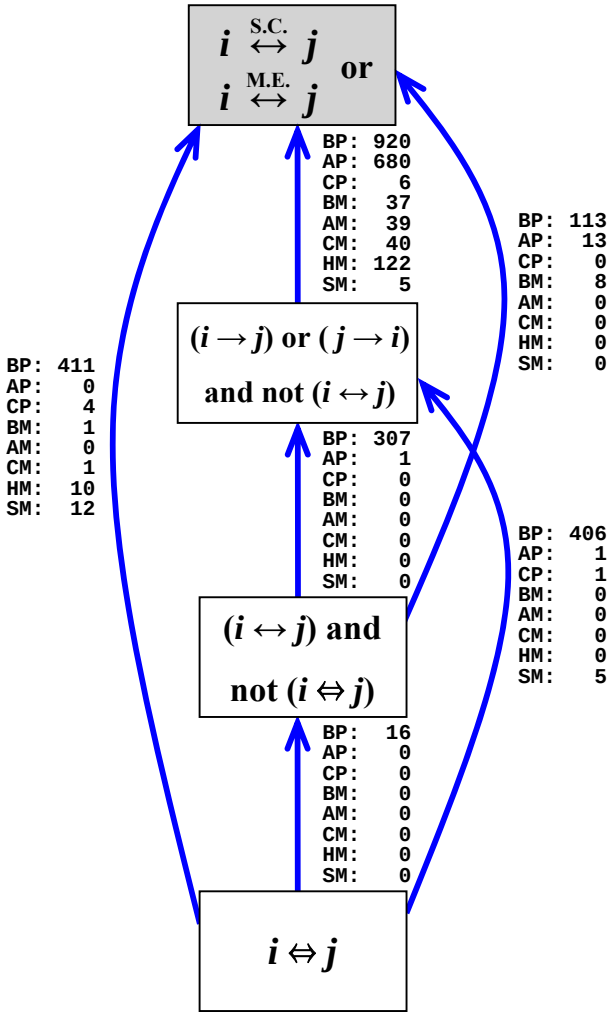


Figure 3.7: *Observed changes in flux (un)coupling relations due to subsystem selection for the four subsystems: AP: plastid of A. thaliana; AM: mitochondrion of A. thaliana; BP: plastid of barley; BM: mitochondrion of barley; CP: plastid of C. reinhardtii; CM: mitochondrion of C. reinhardtii; HM: mitochondrion of human; and SM: mitochondrion of S. cerevisiae. An arrow from relation A to relation B indicates that while relation A holds in the extended network, relation B may be observed in the subsystem. Gray and white boxes indicate uncoupling and coupling relations, respectively.*

tion. If this is not the case, one possible explanation for the difference between the plastids and mitochondria could be that during evolution, some pathways might have evolved in modern plants which increase the dependencies between the internal and external fluxes of the plastids. Such pathways, however, might be missing in a simple unicellular alga like *C. reinhardtii*. Another possibility is that such pathways are not included in the *C. reinhardtii* metabolic model as a result of its specific reconstruction method. The reconstruction procedure relies on verification of transcripts in *C. reinhardtii* when grown in constant light.

Model name	Organelle subsystem	subsystem size	Total number of coupling relations	Number of changes	Change ratio
Barley	plastid	139	9591	2173	22.7%
A. thaliana	plastid	112	6216	695	11.2%
C. reinhardtii	plastid	81	3240	11	0.34%
Barley	mitochondrion	45	990	46	4.6%
A. thaliana	mitochondrion	51	2250	39	1.5%
C. reinhardtii	mitochondrion	49	1176	41	3.5%
Human	mitochondrion	484	116 886	132	0.11%
Yeast	mitochondrion	161	12 880	22	0.17%

Table 3.1: *Frequencies of changes in flux coupling relations due to analysis of organelle subsystems instead of the complete network. Subsystem size is the number of reactions of a subsystem which are unblocked in the complete genome-scale network.*

Therefore, the "chloroplast" reactions (and possibly not all plastid reactions) are included. If more comprehensive genome-scale metabolic networks of plants are available in future, one can test these hypotheses.

3.4 Methods

3.4.1 Datasets: Genome-scale Network Models and Organelle Subsystems

Five genome-scale metabolic network models are used in this study: the network of *Hordeum vulgare* (barley) (Grafahrend-Belau et al., 2009); *AraGEM*, the network of *Arabidopsis thaliana*, (de Oliveira Dal'Molin et al., 2010); *iAM303*, the network of *Chlamydomonas reinhardtii* (Manichaikul et al., 2009); *Recon 1*, the network of *Homo sapiens* (human) (Duarte et al., 2007); and *iND750*, the network of *Saccharomyces cerevisiae* (baker's yeast) (Duarte et al., 2004).

Eight organelle subsystems were considered: plastid subsystems from barley (140 reactions including 139 unblocked ones), *A. thaliana* (123 reactions including 112 unblocked ones), and *C. reinhardtii* (99 reactions including 81 unblocked ones); and mitochondrial subsystems from barley (45 reactions, all unblocked), *A. thaliana* (58 reactions including 51 unblocked ones), *C. reinhardtii* (58 reactions including 49 unblocked ones), human (600 reactions including 484 unblocked ones) and baker's yeast (263 reactions including 161 unblocked ones). To select each organelle subsystem from the stoichiometric matrix of the genome-scale network, we considered a submatrix which includes all the rows corresponding to the metabolites in the organelle subsystem.

3.4.2 Flux Coupling Analysis

Flux coupling analysis was performed by FFCA software (see Chapter 4). We considered all pairs of unblocked reactions in each subsystem, and computed pairwise flux coupling relations: (i) when the genome-scale network is analyzed; and (ii) when the isolated subsystem is analyzed.

Summary:

- Selecting a subsystem of reactions from a metabolic network, i.e., redrawing network boundaries, is equivalent to selecting a row submatrix of the stoichiometric matrix (in contrast to deleting reactions, which is equivalent to selecting a column submatrix).

- Two coupled reactions in an isolated subsystem, if remaining unblocked, will stay coupled in the extended network (possibly with a different coupling relation). However, uncoupled reaction pairs in an isolated subsystem may or may not remain uncoupled.

- As mentioned, some pairs of reactions in a certain subsystem may change their flux coupling relations if the subsystem is studied in isolation. The change ratio, i.e., ratio of changed coupling relations to all coupling relations can be a measure of the dependence of the subsystem to the rest of the network.

- Plastids of *Hordeum vulgare* (barley) and *Arabidopsis thaliana* show a relatively high change ratio compared to the plastid of *Chlamydomonas reinhardtii* or mitochondria of any studied organism.

Appendix 3.1

(A) Original network:

METATOOL input:

```
-ENZREV

-ENZIRREV
1 2 3 4 5 6 7 8 9 i j

-METINT
a b c d e f g h ii jj

-METEXT

-CAT
i : 1 a = 1 c .
j : 1 h = 1 ii .
1 : = 1 b + 3 a .
2 : 1 b = 1 d .
3 : 1 c = 1 e .
4 : 1 e + 1 d = 2 h .
5 : 1 c = 1 f .
6 : 1 f = 1 g .
7 : 1 g = 1 h .
8 : 1 c = 1 jj .
9 : 1 jj + 3 ii = .
```

Elementary mode(s):

```
1 1 1 1 1 1 1 1 1 3 3
```

(B) Setting "A" and "B" as external (and removing "1"):

METATOOL input:

```
-ENZREV

-ENZIRREV
2 3 4 5 6 7 8 9 i j

-METINT
c d e f g h ii jj

-METEXT
a b

-CAT
i : 1 a = 1 c .
j : 1 h = 1 ii .
2 : 1 b = 1 d .
3 : 1 c = 1 e .
4 : 1 e + 1 d = 2 h .
5 : 1 c = 1 f .
6 : 1 f = 1 g .
7 : 1 g = 1 h .
8 : 1 c = 1 jj .
9 : 1 jj + 3 ii = .
```

Elementary mode(s):

```
3 3 3 0 0 0 2 2 5 6
0 0 0 3 3 3 1 1 4 3
```

(C) Setting "A", "B", "I" and "J" as external (and removing "1" and "9"):

METATOOL input:

```
-ENZREV

-ENZIRREV
2 3 4 5 6 7 8 i j

-METINT
c d e f g h

-METEXT
a b ii jj

-CAT
i : 1 a = 1 c .
j : 1 h = 1 ii .
2 : 1 b = 1 d .
3 : 1 c = 1 e .
4 : 1 e + 1 d = 2 h .
5 : 1 c = 1 f .
6 : 1 f = 1 g .
7 : 1 g = 1 h .
8 : 1 c = 1 jj .
```

Elementary mode(s):

```
0 0 0 0 0 0 1 1 0
1 1 1 0 0 0 0 1 2
0 0 0 1 1 1 0 1 1
```

(D) Setting "A", "B", "F", "G" , "I" and "J" as external (and removing "1", "6" and "9"):

METATOOL input:

```
-ENZREV

-ENZIRREV
2 3 4 5 7 8 i j

-METINT
c d e h

-METEXT
a b ii jj f g

-CAT
i : 1 a = 1 c .
j : 1 h = 1 ii .
2 : 1 b = 1 d .
3 : 1 c = 1 e .
4 : 1 e + 1 d = 2 h .
5 : 1 c = 1 f .
7 : 1 g = 1 h .
8 : 1 c = 1 jj .
```

Elementary mode(s):

```
0 0 0 1 0 0 1 0
0 0 0 0 0 1 1 0
0 0 0 0 1 0 0 1
1 1 1 0 0 0 1 2
```

(E) Setting "A", "D", "E", "F", "G", "I" and "J" as external (and removing "1", "2", "6" and "9"). Additionally, metabolite "B" is removed because it is not involved in any reaction:

METATOOL input:

```
-ENZREV

-ENZIRREV
3 4 5 7 8 i j

-METINT
c h
-METEXT
a ii jj f g d e

-CAT
i : 1 a = 1 c .
j : 1 h = 1 ii .
3 : 1 c = 1 e .
4 : 1 e + 1 d = 2 h .
5 : 1 c = 1 f .
7 : 1 g = 1 h .
8 : 1 c = 1 jj .
```

Elementary mode(s):

```
1 0 0 0 0 1 0
0 0 1 0 0 1 0
0 0 0 0 1 1 0
0 1 0 0 0 0 2
0 0 0 1 0 0 1
```

Appendix 3.2

(A) Original network:

METATOOL input:

```
-ENZREV

-ENZIRREV
1 2 3 4 i j

-METINT
a b c d

-METEXT

-CAT
i : = 1 a .
j : = 1 b .
1 : 1 a = .
2 : 1 b = .
3 : 1 a + 1 b = 1 c .
4 : 1 c = 1 d .
```

Elementary mode(s):

```
1 0 0 0 1 0
0 1 0 0 0 1
```

(B) Setting "D" as external:

METATOOL input:

```
-ENZREV

-ENZIRREV
1 2 3 4 i j

-METINT
a b c

-METEXT
d

-CAT
i : = 1 a .
j : = 1 b .
1 : 1 a = .
2 : 1 b = .
3 : 1 a + 1 b = 1 c .
4 : 1 c = 1 d .
```

Elementary mode(s):

```
1 0 0 0 1 0
0 1 0 0 0 1
0 0 1 1 1 1
```

Chapter 4

FFCA: A Feasibility-based Method for Flux Coupling Analysis

4.1 Background

Constraint-based analysis of metabolic networks has become an important technique to describe and predict the behavior of living organisms (Reed, 2009; Fell et al., 2010). While a growing number of metabolic network reconstructions has become available during the last years, the computational analysis of genome-scale networks with hundreds or thousands of reactions may still be very time-consuming. Therefore, there is a need for more efficient algorithms and tools (see e.g. Terzer and Stelling, 2008; Haus et al., 2008; Gudmundsson and Thiele, 2010)).

Flux coupling analysis (FCA) (Burgard et al., 2004) is a useful method to find dependencies between fluxes of a metabolic network at steady-state. Several studies have used FCA for exploring various biological questions such as network evolution (Notebaart et al., 2009; Pál et al., 2005*a*; Seshasayee et al., 2009), gene essentiality (Notebaart et al., 2009), analysis of experimentally measured fluxes (Suthers et al., 2010; Bundy et al., 2007) or gene regulation (Notebaart et al., 2008). Having a time efficient implementation of FCA is important in such studies.

In the rest of this chapter, we first recall some basic definitions. Then we briefly review the previously proposed FCA methods.

4.1.1 Mathematical definitions

Basic preliminaries

In a metabolic network, if a reversible reaction has a positive (resp. negative) flux, we say that it is working in *forward* (resp. *backward*) direction. *Splitting* a reversible reaction i means making reaction i irreversible and adding one more irreversible reaction $i+n$ to the network, which works in the backward direction. Without loss of generality, we assume that in the numbering of reactions, the first $|Rev|$ reactions are the reversible ones.

The set of unblocked reactions can be further partitioned based on the *re-*

versibility type of reactions (Larhlimi and Bockmayr, 2006). We define $Irev$ as the set of all reactions that can work only in one direction, i.e., those reactions that take only non-negative or only non-positive flux values at steady-state. The set of reactions that can work in both directions at steady-state is further divided into two subsets: $Prev$, the set of *pseudo-irreversible* reactions, and $Frev$, the set of *fully reversible* reaction. A reaction i is in $Prev$ if for all flux vectors v in the lineality space of the flux cone, we have $v_i = 0$. Accordingly, we define $Frev$ as the set of those reactions that can have a non-zero flux value if fluxes through all irreversible reactions are set to zero.

Elementary flux patterns

Suppose $v \in C$ is a flux vector and $A \subseteq R$ is a subnetwork. The flux pattern of v for A is defined as $A \cap supp(v)$, which is the set of those reactions in A which have non-zero values in v (Kaleta et al., 2009). A flux pattern is called an *elementary flux pattern* (EFP) if it cannot be written as the union of other flux patterns. For studying EFPs, Kaleta et al. (2009) assume that the network contains only irreversible reactions. To achieve this, every reversible reaction should be split into two irreversible reactions (forward and backward).

4.1.2 Approaches to Flux Coupling Analysis

In this section, we briefly introduce different approaches to flux coupling analysis. For additional information and the technical details on the implementation of these algorithms, see Section 4.3.3.

Flux Coupling Finder Algorithm (FCF)

The most widely used method for FCA is the Flux Coupling Finder (FCF) algorithm (Burgard et al., 2004). This approach is based on solving linear programming (LP) problems, and therefore, is an optimality-based method. After finding blocked reactions and splitting reversible reactions, for every pair of unblocked reactions i and j, two LP problems are solved. Depending on the optimal values obtained, the coupling relation between i and j is determined. There is a post-processing step in FCF. Since the reversible reactions have been split, flux coupling relations for these reactions have to be obtained from the coupling relations for the corresponding irreversible forward and backward reactions.

The FCF algorithm has been successfully used for finding coupling relations in a number of metabolic networks (Burgard et al., 2004; Notebaart et al., 2008, 2009; Suthers et al., 2010; Pál et al., 2005*a*; Bundy et al., 2007; Seshasayee et al., 2009). However, this approach is rather time-consuming for genome-scale metabolic networks with thousands of reactions, although it is still one of the fastest FCA methods. To the best of our knowledge, no implementation of the FCF algorithm is publicly available at the moment.

FCA based on Minimal Metabolic Behaviors (MMB-FCA)

Larhlimi and Bockmayr (2006, 2009) have proposed a different strategy for flux coupling analysis. In this approach, a minimal set of generating vectors of the flux cone is computed. Then, the coupling relation for any pair of reactions is inferred based on the co-appearance of non-zero fluxes in the generating vectors. Additionally, they show that depending on the reversibility type of the reactions, only certain flux coupling relations can occur. The authors suggest that this may result in a considerable speed-up of any FCA method, including their own approach (Larhlimi and Bockmayr, 2006).

FCA based on elementary flux patterns (EFP-FCA)

Recently, Kaleta et al. (2009) introduced the concept of elementary flux patterns (EFPs) for analysis of minimal active reactions in a "subnetwork", which account for possible steady-state flux distributions in a (possibly big) metabolic network. They also presented a method based on mixed integer linear programming (MILP) to compute EFPs. Kaleta et al. suggested that EFPs can be used for characterizing flux coupling relations (see Supplemental Material in Kaleta et al., 2009). Consider a subnetwork including two unblocked reactions i and j. If each of these reactions can have a non-zero flux independently of the other (i.e., $i \overset{Un}{\longleftrightarrow} j$), $\{i\}$ and $\{j\}$ are the only EFPs in this subnetwork. On the other hand, if we assume (without loss of generality) that i is directionally coupled to j, then the EFPs of this subnetwork are $\{i, j\}$ and $\{j\}$. Finally, if the reactions are partially coupled, we will have only one EFP, which is $\{i, j\}$. With this method, it is not possible to distinguish between partial and full coupling, since flux patterns only contain the information about the activity or inactivity of the fluxes, but not the flux values.

4.1.3 Goals of the Present Chapter

Although FCA is a promising tool for metabolic network analysis, to the best of our knowledge there is currently no publicly-available tool for performing genome-scale FCA. Several approaches for FCA have been proposed in the literature. It is not known which of these methods is the fastest in practice. In this chapter, we present a novel "feasibility-based" flux coupling analysis method (FFCA) and compare it to previously existing approaches. A corresponding software tool will be freely available for non-commercial use.

4.2 Results and Discussion

4.2.1 FFCA: Feasibility-based Flux Coupling Analysis

We introduce a new approach for flux coupling analysis which is based on feasibility testing (a system of linear inequalities $Ax \leq b$ is *feasible* if the polyhedron $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ is not empty).

According to Larhlimi and Bockmayr (2006), two unblocked reactions i and j can be coupled only if one of the following four cases holds (note that initially $3 \times 3 = 9$ reversibility types for the pair (i, j) would be possible):

1. $i, j \in Irev$: In this case, i and j can be directionally, partially or fully coupled.
2. $i \in Irev$ and $j \in Prev$: The only possibility is $j \longrightarrow i$.
3. $i, j \in Prev$: In this case, we can only have $i \Longleftrightarrow j$.
4. $i, j \in Frev$: In this case, we can only have $i \Longleftrightarrow j$.

Therefore, it is enough to determine the reversibility type of i and j, and then check if the corresponding coupling relation holds. This will be referred to as "Reversibility-Type prunings" (or simply, RT-prunings).

Taking this observation into account, we propose the following procedure for FFCA:

1. $i, j \in Irev$: In this case, we check the feasibility of two systems of linear inequalities:

$$v_i = 1,\ \ v_j = 0,\ \ Sv = 0,\ \ v_r \geq 0,\ \text{for all } r \in Irr, \tag{P1}$$

and

$$v_i = 0,\ \ v_j = 1,\ \ Sv = 0,\ \ v_r \geq 0,\ \text{for all } r \in Irr. \tag{P2}$$

If (P1) and (P2) are both feasible, then i and j are not coupled to each other ($i \overset{Un}{\longleftrightarrow} j$). If (P1) (resp. (P2)) is infeasible, then $i \longrightarrow j$ (resp. $j \longrightarrow i$). If (P1) and (P2) are both infeasible, then i and j are partially (and maybe fully) coupled. To check whether they are fully coupled, one has to use other methods, e.g. computing enzyme subsets (Pfeiffer et al., 1999) or solving two LPs as in the FCF algorithm (Burgard et al., 2004).

2. $i \in Irev$ and $j \in Prev$: The only possible coupling relation is $j \longrightarrow i$ (but not $i \longrightarrow j$). Hence, (P1) will be always feasible, because feasibility of (P1) means that $i \longrightarrow j$ does not hold. However, we need to check the feasibility of (P2). Additionally, since j can take negative values, one more system should be checked for feasibility:

$$v_i = 0,\ \ v_j = -1,\ \ Sv = 0,\ \ v_r \geq 0,\ \text{for all } r \in Irr. \tag{P3}$$

It can be easily shown that if (P2) and (P3) are both infeasible, then $j \longrightarrow i$. Otherwise, i and j are uncoupled.

3. $i, j \in Prev$: If (P2) and (P3) are both infeasible, then $i \Longleftrightarrow j$ (because feasibility of (P2) and (P3) implies $j \longrightarrow i$, which in turn implies $i \Longleftrightarrow j$ based on Proposition 2 in Larhlimi and Bockmayr (2006)). If (P2) or (P3) are feasible, then i and j are uncoupled.

4. $i, j \in Frev$: Similar to the latter case.

To perform FFCA, a method is needed to check the feasibility of a system of linear inequalities. In practice this can be done by solving an LP constructed by the system of inequalities together with a constant objective function. Any feasible solution will be an optimal solution of the LP, and therefore, the LP solver will finish after finding the first feasible solution. For example, for checking the feasibility of (P1), one can solve the following LP:

$$\begin{array}{lll} \max & c & \text{a constant value} \\ s.t. & \sum_{r \in R} S_{mr} v_r = 0 & \forall m \in M \\ & v_r \geq 0 & \forall r \in Irr \\ & v_i = 1 & \\ & v_j = 0 & \end{array}$$

Since v_i is constant, the optimal value exists iff this problem is feasible. Similar LPs can be solved to check the feasibility of (P2) and (P3).

In Table 4.1, the characteristics of the FFCA approach are compared to the other FCA methods studied in this article.

4.2.2 Comparison of the four FCA approaches

To compare the different approaches, namely FCF, MMB-FCA, EFP-FCA and FFCA, we implemented all of them in Matlab (see Section 4.3.3). A benchmark set of six metabolic network models was used to evaluate the running times. The number of unblocked reactions in these models ranges from 18 to 765. Table 4.2 summarizes the results. One can see that in all cases FFCA is 2 to 3 times faster than FCF and orders of magnitude faster than EFP-FCA. The table also shows that FFCA is more appropriate for FCA in genome-scale networks. MMB-FCA is the fastest method for the three smallest networks. However, already for the middle-sized *H. pylori* network and even more for the large networks of *S. cerevisiae* and *E. coli*, FFCA proves to be faster than MMB-FCA. The computation time required for MMB-FCA rapidly grows when the number of reactions increases. This is possibly due to the exponential size of the set of generating vectors which has to be computed before finding the coupled reactions (see Section 4.4 in Larhlimi, 2008). EFP-FCA, which is based on solving mixed integer

Method name	Preprocessing	Main procedure: Type of linear program and solution	Main procedure: Further distinguishing of partial and full coupling required?	Postprocessing for reversible reactions?
MMB-FCA	computing MMBs + reaction classification	n/a	No	No
EFP-FCA	splitting reversible reactions	MILP, optimal	Yes	Yes
FCF	splitting reversible reactions	LP, optimal	No	Yes
FCF_{woS}	n/a	LP, optimal	No	No
$FCF_{RT,woS}$	reaction classification	LP, optimal	No	No
FFCA	reaction classification	LP, feasible	Yes	No

Table 4.1: *General comparison of different approaches to flux coupling analysis. In the three methods which use RT-prunings, namely MMB-FCA, $FCF_{RT,woS}$ and FFCA, blocked reactions are determined within the preprocessing step. In the other methods, blocked reactions are found in the main procedure.*

	number of unblocked reactions	MMB-FCA	EFP-FCA	FCF	FCF_{woS}	$FCF_{RT,woS}$	FFCA
ILLUSNET	18	0.01	26.3	0.25	0.14	0.09	0.08
RBC	38	0.05	152	1.39	0.80	0.68	0.63
EC core	63	0.22	585	6.58	3.03	3.13	2.50
H. pylori	217	69.8	$>$ 1 day	196	83.6	67.0	60.9
S. cerevisiae	639	$>$ 1 day	$>$ 1 day	8.5×10^3	4.0×10^3	3.4×10^3	3.1×10^3
E. coli	765	$>$ 1 day	$>$ 1 day	1.2×10^4	7.4×10^3	6.3×10^3	5.6×10^3

Table 4.2: *CPU running time (in seconds) required for flux coupling analysis of the benchmark networks. See the text for more details.*

linear programs, turns out to be much slower than other methods. Although the concept of elementary flux patterns is very useful in the analysis of subnetworks, its applicability in full FCA therefore seems to be limited.

FFCA vs. FCF

Both the FCF algorithm and the current implementation of FFCA solve LPs for flux coupling analysis. One may ask why FFCA is faster than the classical FCF method. There are at least four major differences:

- When an LP is solved in FFCA, finding the first feasible solution is enough, while the LPs should be solved to optimality in case of the FCF algorithm.
- In the FCF method, in contrast to FFCA, every reversible reaction is split

into two (forward and backward) irreversible reactions. This step slows down the procedure and increases the size of the LPs to be solved.

– For computing the flux coupling relation between any pair of reactions, we always need two LPs in FFCA, while in FCF sometimes more LPs have to be solved. For example, for computing the coupling relation between an irreversible and a reversible reaction (after splitting), four LPs are solved (see Section 4.3.3).

– Only in FFCA we consider the Reversibility-Type prunings (Larhlimi and Bockmayr, 2006) to reduce the number of possible coupled reaction pairs.

The first difference is because of the natures of FFCA and FCF. Nevertheless, one can think of implementing FCF without splitting reversible reactions and/or with the RT-prunings. In order to assess the importance of these issues, two improved versions of FCF were implemented as suggested by Larhlimi (2008) (see Section 4.3.3 and also Table 4.1): (*i*) FCF was re-implemented without splitting reactions (FCF_{woS}); and (*ii*) FCF was re-implemented without splitting reactions and with the RT-prunings ($\text{FCF}_{\text{RT,woS}}$).

In Table 4.2 the computational running times of these methods are also shown. As expected, the two versions of the improved FCF algorithm are better than the classical FCF algorithm, while they are still slower that FFCA.

4.3 Materials and Methods

4.3.1 Metabolic Network Models

Six metabolic networks were used in this study: (i) *ILLUSNET* network from Larhlimi and Bockmayr (2006); (ii) *RBC*: metabolic network of red blood cell (Wiback and Palsson, 2002); (iii) *EC core*: central metabolic network of *E. coli* (Palsson, 2006); (iv) *H. pylori* genome-scale metabolic network (Thiele et al., 2005); (v) *yeast* (*S. cerevisiae*) genome-scale metabolic network (Duarte et al., 2004); and (vi) *E. coli* genome-scale metabolic network (Reed et al., 2003).

For FCA, all uptake reactions were assumed to be able to carry non-zero fluxes.

4.3.2 Comparison of different FCA methods

All computations were performed on a 64-bit Debian Linux system with Intel Xeon 3.0 GHz processor. The running times include the CPU time for pre-processing, computation of flux coupling relations, and post-processing (where necessary).

4.3.3 Implementation Details

To the best of our knowledge, no implementation of flux coupling methods is publicly available. Therefore, we implemented the different approaches to compare their time efficiency. Unless indicated otherwise, all tools were implemented in Matlab v7.4. MMB-FCA, EFP-FCA and FCF are the previous approaches which were (re-)implemented in this study. Additionally, our new method called FFCA, together with two improved versions of the FCF method, namely FCF_{woS} (FCF without splitting reversible reactions) and $\text{FCF}_{\text{RT,woS}}$ (FCF with Reversibility-Type prunings and without splitting reversible reactions) were implemented in order to get a better picture about the efficiency of the different approaches.

Implementation of MMB-FCA

In this approach, a convex basis of the flux cone is needed. We use the software *cdd* (Fukuda and Prodon, 1996), a tool based on the double description method, to compute a minimum set of generating vectors. These correspond to the lineality space and the minimal proper faces (or minimal metabolic behaviors, MMBs) of the flux cone (see Larhlimi and Bockmayr, 2009, for more details). Next the reversibility type of the reactions is determined by computing the sets $Blk, Irev, Prev$, and $Frev$. The flux coupling relations are then obtained as described in Larhlimi and Bockmayr (2006). The following pseudo-code (Algorithm 1) summarizes the procedure. $CouplingRelation$ is a matrix where the entry (i, j), with $i < j$, describes the coupling relation between two unblocked reactions i and j.

Algorithm 1: MMB-FCA, an FCA approach based on the generators of flux cone (Larhlimi and Bockmayr, 2006)

Input:
- S (the $m \times n$ stoichiometric matrix)
- $Irr \subseteq \{1, \ldots, n\}$ (the set of irreversible reactions)

Output:
- Blk (the set of blocked reactions)
- $CouplingRelation$

Initialization:
$Rev := \{1, \ldots, n\} \setminus Irr$
$Blk := \varnothing; Irev := \varnothing; Prev := \varnothing; Frev := \varnothing$

Preprocessing:
`/* Classification of reactions based on their reversibility type */`
$B :=$ ComputeLinealitySpace$(S \cdot v = 0, v_i \geq 0,$ for all $i \in Irr)$
$b :=$ NumberOfRows(B)
$G :=$ ComputeMinimalProperFaces$(S \cdot v = 0, v_i \geq 0,$ for all $i \in Irr)$
$g :=$ NumberOfRows(G)
for $i \in R$ **do**
  if $B_{[b],\{i\}} \neq 0$ **then**
    $Frev := Frev \cup \{i\}$
  else if $G_{[g],\{i\}} = 0$ **then**
    $Blk := Blk \cup \{i\}$
  else if $G_{[g],\{i\}} \geq 0$ *or* $G_{[g],\{i\}} \leq 0$ **then**
    $Irev := Irev \cup \{i\}$
  else
    $Prev := Prev \cup \{i\}$
  end
end

Main procedure:
foreach $i, j \in Prev$ *with* $i < j$ **do**
  if $\exists \lambda \in \mathbb{R}$ *such that* $g_i = \lambda g_j$ *for all* $g \in G$ **then**
    $CouplingRelation[i, j] :=$ "$\Longleftrightarrow$"
  else
    $CouplingRelation[i, j] :=$ "$\overset{Un}{\longleftrightarrow}$"
  end
end
foreach $i, j \in Frev$ *with* $i < j$ **do**
  if $\exists \lambda \in \mathbb{R}$ *such that* $g_i = \lambda g_j$ *for all* $g \in G$ *and* $b_i = \lambda b_j$ *for all* $b \in B$ **then**
    $CouplingRelation[i, j] :=$ "$\Longleftrightarrow$"
  else
    $CouplingRelation[i, j] :=$ "$\overset{Un}{\longleftrightarrow}$"
  end
end

Algorithm 1: MMB-FCA, an FCA approach based on the generators of flux cone, continued

foreach $i, j \in Irev \cup Prev$ *with* $i \neq j$ *and* $\{i, j\} \nsubseteq Prev$ **do**
  if $g_i = 0$ *or* $g_j \neq 0$ *for all* $g \in G$ **then**
    $CouplingRelation[i, j] :=$ "$\longrightarrow$"
  else
    $CouplingRelation[i, j] :=$ "$\overset{Un}{\longleftrightarrow}$"
  end
end
foreach $i, j \in Irev$ *with* $i < j$ **do**
  if $CouplingRelation[i, j] =$"$\longrightarrow$" *and* $CouplingRelation(j, i) =$ "$\longrightarrow$" **then**
    if $\exists \lambda \in \mathbb{R}$ *such that* $g_i = \lambda g_j$ *for all* $g \in G$ **then**
      $CouplingRelation[i, j] :=$ "$\Longleftrightarrow$"
    else
      $CouplingRelation[i, j] :=$ "$\longleftrightarrow$"
    end
  end
end

Implementation of EFP-FCA

We first introduce some auxiliary functions. *minimize* and *maximize* are two functionals used to solve a linear program (LP), i.e., to optimize a linear objective function subject to a set of linear constraints. For example, if $Con = \{Ax \leq b, Cx = d\}$ and c defines the objective function, then $maximize(c^T x \mid Con)$ computes the maximum value of $c^T x$ subject to the constraints Con. If Con admits no real solution, then both *minimize* and *maximize* return the special value $\perp$ (i.e., infeasible). If Con is feasible, but the optimum is unbounded, then *minimize* (resp. *maximize*) return $-\infty$ (resp. $+\infty$).

Any available LP solver can be used to perform these computations. Here, we used CLP, the LP solver from the COIN-OR package (Lougee-Heimer, 2003).

Another function is CouplingRelationRecompute. In some of the FCA approaches, it is necessary to split reversible reactions i in two irreversible reactions, namely i^+ (forward direction) and i^- (backward direction). Suppose that for an irreversible reaction j, we have ⓐ = $CouplingRelation(i^+, j)$ and ⓑ = $CouplingRelation(i^-, j)$. Then a function is needed to compute the flux coupling relation between i and j, based on ⓐ and ⓑ. The aim of CouplingRelationRecompute is to take such an ⓐ, ⓑ pair and compute the flux coupling relation for the original (non-split) reversible reaction. For computing the coupling relation of two reversible reactions after splitting, this function should be called three times, since $CouplingRelation[i, j]$= CouplingRelationRecompute($CouplingRelation(i^+, j)$,$CouplingRelation(i^-, j)$).

Algorithm 2: Procedure CouplingRelationRecompute

Input:
- ⓐ $\in \{\longleftarrow, \longrightarrow, \longleftrightarrow, \Longleftrightarrow, \overset{Un}{\longleftrightarrow}\}$ (the first coupling relation)
- ⓑ $\in \{\longleftarrow, \longrightarrow, \longleftrightarrow, \Longleftrightarrow, \overset{Un}{\longleftrightarrow}\}$ (the second coupling relation)

Output: ⓝ $\in \{\longleftarrow, \longrightarrow, \longleftrightarrow, \Longleftrightarrow, \overset{Un}{\longleftrightarrow}\}$ (a new coupling relation)

switch (ⓐ, ⓑ) **do**
  case ("$\Longleftrightarrow$", "$\longleftrightarrow$")
    ⓝ :="$\longleftrightarrow$"
  case ("$\Longleftrightarrow$", "$\longrightarrow$")
    ⓝ :="$\longrightarrow$"
  case ("$\Longleftrightarrow$", "$\longleftarrow$")
    ⓝ :="$\longleftarrow$"
  case ("$\Longleftrightarrow$", "$\overset{Un}{\longleftrightarrow}$")
    ⓝ :="$\overset{Un}{\longleftrightarrow}$"
  case ("$\longleftrightarrow$", "$\longrightarrow$")
    ⓝ :="$\longrightarrow$"
  case ("$\longleftrightarrow$", "$\longleftarrow$")
    ⓝ :="$\longleftarrow$"
  case ("$\longleftrightarrow$", "$\overset{Un}{\longleftrightarrow}$")
    ⓝ :="$\overset{Un}{\longleftrightarrow}$"
  case ("$\longrightarrow$", "$\longleftarrow$")
    ⓝ :="$\overset{Un}{\longleftrightarrow}$"
  case ("$\longrightarrow$", "$\overset{Un}{\longleftrightarrow}$")
    ⓝ :="$\overset{Un}{\longleftrightarrow}$"
  case ("$\longleftarrow$", "$\longrightarrow$")
    ⓝ :="$\overset{Un}{\longleftrightarrow}$"
  case ("$\longleftarrow$", "$\overset{Un}{\longleftrightarrow}$")
    ⓝ :="$\overset{Un}{\longleftrightarrow}$"
  otherwise
    ⓝ := ⓐ
  end
end

We also introduce the function BlkFinderRevCorrector(S, Irr) to find the blocked reactions and correct the reversiblity types of those reactions which can work either in forward or in backward direction. Algorithm 3 summarizes this procedure.

Algorithm 3: Procedure BlkFinderRevCorrector

Input:
- S (the $m \times n$ stoichiometric matrix)
- $Irr \subseteq \{1, \ldots, n\}$ (the set of irreversible reactions)

Output:
- $Blk^\star$ (the set of blocked reactions)
- $S^\star$ (the stoichiometric matrix after correcting the reversibility of reactions)
- $Irr^\star$ (the set of irreversible reactions after correcting the reversibility of reactions)

Initialization:
$Blk^\star := \varnothing; Irr^\star := Irr; S^\star := S$

foreach $i \in \{1, \ldots, n\}$ **do**
  $Con := \{Sv = 0, v_r \geq 0 \text{ for all } r \in Irr\}$
  $max := maximize(v_i \mid Con)$
  if $i \in Rev$ **then**
    $min := minimize(v_i \mid Con)$
    if $min = max = 0$ **then**
      $Blk^\star = Blk^\star \cup \{i\}$
      $S := [S_{[m],[i-1]},\ \ 0,\ \ S_{[m],\{i+1,\ldots,n\}}]$
    end
    else if $min = 0, max > 0$ **then**
      $Irr^\star := Irr^\star \cup \{i\}$
    end
    else if $min < 0, max = 0$ **then**
      $Irr^\star := Irr^\star \cup \{i\}$
      $S := [S_{[m],[i-1]},\ \ -S_{[m],\{i\}},\ \ S_{[m],\{i+1,\ldots,n\}}]$
    end
  else
    if $max = 0$ **then**
      $Blk^\star = Blk^\star \cup \{i\}$
      $S := [S_{[m],[i-1]},\ \ 0,\ \ S_{[m],\{i+1,\ldots,n\}}]$
    end
  end
end

EFP-FCA is a recent FCA method suggested by Kaleta et al. (2009). In EFP-FCA we first split the reversible reactions. Subsequently, we use the standard implementation of the EFPTools package (Kaleta, 2009), which uses CLP to solve MILPs, for computing EFPs. Then directional and partial coupling, and also uncoupling relations are inferred. The whole procedure is summarized in Algorithm 4. A new function, ComputeElementaryFluxPatterns$(w, S^\star)$ is used, which computes the EFPs of a network (with stoichiometric matrix $S^\star$) for a subnetwork of selected reactions in w.

Algorithm 4: EFP-FCA, an FCA approach based on elementary flux patterns (Kaleta et al., 2009)

Input:
- S (the $m \times n$ stoichiometric matrix)
- $Irr \subseteq \{1, \ldots, n\}$ (the set of irreversible reactions)

Output:
- Blk (the set of blocked reactions)
- $CouplingRelation$

Preprocessing:

`/* Finding blocked reactions and updating` Irr `and splitting reversible reactions */`

$(Blk^\star, S^\star, Irr^\star) :=$ BlkFinderRevCorrector(S, Irr)
$Rev^\star := [n] \setminus (Irr^\star \cup Blk^\star)$
$S' := \left[S, \ -S_{[m],Rev^\star}\right]$

Main procedure:

foreach $i, j \in [n + |Rev^\star|] \setminus Blk^\star$ *with* $i < j$ **do**
- $S'' := S'$
- **if** $j \in Rev^\star, j \leq n$ **then** replace the $(j+n)$-th column of S'' with zero column
- **if** $j \in Rev^\star, j \geq n+1$ **then** replace the $(j-n)$-th column of S'' with zero column
- **if** $i \in Rev^\star, i \leq n$ **then** replace the $(i+n)$-th column of S'' with zero column
- **if** $i \in Rev^\star, i \geq n+1$ **then** replace the $(i-n)$-th column of S'' with zero column
- $E :=$ ComputeElementaryFluxPatterns$((i, j), S'')$
- **switch** E **do**
  - **case** $\{(1,1)\}$
    - $CouplingRelation[i,j] :=$ "$\longleftrightarrow$"
    - $min := minimize(v_i \mid \{S''v = 0, v \geq 0, v_j = 1\})$
    - $max := maximize(v_i \mid \{S''v = 0, v \geq 0, v_j = 1\})$
    - **if** $min = max$ **then** $CouplingRelation[i,j] :=$ "$\Longleftrightarrow$"
  - **case** $\{(1,1),(0,1)\}$
    - $CouplingRelation[i,j] :=$ "$\longrightarrow$"
  - **case** $\{(1,1),(1,0)\}$
    - $CouplingRelation[i,j] :=$ "$\longleftarrow$"
  - **case** $\{(1,0),(0,1)\}$
    - $CouplingRelation[i,j] :=$ "$\overset{Un}{\longleftrightarrow}$"
  - **end**
- **end**

end

Algorithm 4: EFP-FCA, an FCA approach based on elementary flux patterns, continued

Postprocessing:
for *every* i, j *with* $i \in Rev^\star$ *and* $j \in Irr^\star$ **do**
  ⓐ := $CouplingRelation[i, j]$
  ⓑ := $CouplingRelation(i + n, j)$
  $CouplingRelation[i, j]$:= CouplingRelationRecompute(ⓐ, ⓑ)
end
for *every* $i, j \in Rev^\star$ *with* $i < j$ **do**
  ⓐ := $CouplingRelation[i, j]$
  ⓑ := $CouplingRelation(i + n, j)$
  ⓝ := CouplingRelationRecompute(ⓐ, ⓑ)
  ⓐ := $CouplingRelation(i, j + n)$
  ⓑ := $CouplingRelation(i + n, j + n)$
  ⓟ := CouplingRelationRecompute(ⓐ, ⓑ)
  $CouplingRelation[i, j]$:= CouplingRelationRecompute(ⓝ, ⓟ)
end

Implementation of FCF

FCF is based on linear programming (LP) (Burgard et al., 2004). For every pair of irreversible reactions, two LPs are solved, in which the flux through one reaction is fixed and the flux through the other reaction is maximized or minimized. Based on the optimal values, the flux coupling relation between these two reactions can be inferred. Again we used CLP, the LP solver from the COIN-OR package (Lougee-Heimer, 2003). After computing all coupling relations, a post-processing step is required to obtain the coupling relations for the fluxes in the original network. Algorithm 5 summarizes the procedure.

Algorithm 5: FCF, the classical Flux Coupling Finder algorithm (Burgard et al., 2004)

Input:
- S (the $m \times n$ stoichiometric matrix)
- $Irr \subseteq \{1, \ldots, n\}$ (the set of irreversible reactions)

Output:
- Blk (the set of blocked reactions)
- $CouplingRelation$

Preprocessing:
$(Blk^\star, S^\star, Irr^\star)$:= BlkFinderRevCorrector(S, Irr)
$Rev^\star := [n] \setminus (Irr^\star \cup Blk^\star)$
$S' := \left[S,\ -S_{[m],Rev^\star}\right]$

Algorithm 5: FCF, the classical Flux Coupling Finder algorithm, cont'd

Main procedure:

foreach $i, j \in [n + |Rev^\star|] \setminus Blk^\star$ **do**
  $S'' := S'$
  if $j \in Rev^\star, j \leq n$ **then** replace the $(j+n)$-th column of S'' with zero column
  if $j \in Rev^\star, j \geq n+1$ **then** replace the $(j-n)$-th column of S'' with zero column
  if $i \in Rev^\star, i \leq n$ **then** replace the $(i+n)$-th column of S'' with zero column
  if $i \in Rev^\star, i \geq n+1$ **then** replace the $(i-n)$-th column of S'' with zero column
  $min := minimize(v_i \mid \{S''v = 0, v \geq 0, v_j = 1\})$
  $max := maximize(v_i \mid \{S''v = 0, v \geq 0, v_j = 1\})$
end
switch (min, max) **do**
  case $(0, \infty)$
    $CouplingRelation[i,j]$:="$\xleftarrow{Un}$"
  case $(> 0, \infty)$
    $CouplingRelation[i,j]$:="$\longleftarrow$"
  case $(0, > 0)$
    $CouplingRelation[i,j]$:="$\longrightarrow$"
  case $(> 0, > 0)$
    if $(min = max)$ **then**
      $CouplingRelation[i,j]$:="$\Longleftrightarrow$"
    else
      $CouplingRelation[i,j]$:="$\longleftrightarrow$"
    end
  end
end

Postprocessing:

for *every* i, j *with* $i \in Rev^\star$ *and* $j \in Irr^\star$ **do**
  ⓐ := $CouplingRelation[i,j]$
  ⓑ := $CouplingRelation[i+n,j]$
  $CouplingRelation[i,j]$:= CouplingRelationRecompute(ⓐ, ⓑ)
end
for *every* $i, j \in Rev^\star$ *with* $i < j$ **do**
  ⓐ := $CouplingRelation[i,j]$
  ⓑ := $CouplingRelation[i+n,j]$
  ⓝ := CouplingRelationRecompute(ⓐ, ⓑ)
  ⓐ := $CouplingRelation[i,j+n]$
  ⓑ := $CouplingRelation[i+n,j+n]$
  ⓟ := CouplingRelationRecompute(ⓐ, ⓑ)
  $CouplingRelation[i,j]$:= CouplingRelationRecompute(ⓝ, ⓟ)
end

Implementation of FCF$_{woS}$

Splitting reactions slows down the FCF algorithm for a variety of reasons, including the bigger size of reconfigured network, the larger number of LPs to be solved, and the need for post-processing. FCF$_{\text{woS}}$ (and also FCF$_{\text{RT,woS}}$) have been implemented as splitting-free FCF methods to get a fair comparison between FFCA and the optimality-based FCF. FCF$_{\text{woS}}$ uses a new function RevToIrrevCoupling to compute the flux coupling relations between a reversible reaction and an irreversible reaction. This function reads the maximum and minimum flux value through a reversible reaction, when the flux through the irreversible reaction is set to a constant. The output of the function is the flux coupling relation between the two reactions. It should be noted that the same function can be used for computing the coupling relation between a pair of reversible reactions i and j. First the coupling relation between i and j^+, and also between i and j^- are computed as explained before. Then, CouplingRelationRecompute can be used to obtain the flux coupling relation between i and j.

Algorithm 6: The RevToIrrevCoupling procedure

Input:
- min (minimum value of the reversible flux when the irreversible flux is constant)
- max (maximum value of the reversible flux when the irreversible flux is constant)

Output:
- ⓡ (the coupling relation between the reversible and the irreversible flux)

switch (min, max) **do**
  case $(-\infty, < 0)$
    ⓡ :="$\longleftarrow$"
  case $(-\infty, \geq 0)$
    ⓡ :="$\overset{Un}{\longleftrightarrow}$"
  case $(> 0, > 0)$
    if $(min = max)$ **then**
      ⓡ :="$\Longleftrightarrow$"
    else
      ⓡ :="$\longleftrightarrow$"
    end
  case $(< 0, < 0)$
    if $(min = max)$ **then**
      ⓡ :="$\Longleftrightarrow$"
    else
      ⓡ :="$\longleftrightarrow$"
    end
  case $(\leq 0, \geq 0)$
    ⓡ :="$\longrightarrow$"
  case $(< 0, \infty)$
    ⓡ :="$\overset{Un}{\longleftrightarrow}$"
  case $(-\infty, \infty)$
    ⓡ :="$\overset{Un}{\longleftrightarrow}$"
  case $(> 0, \infty)$
    ⓡ :="$\longleftarrow$"
  end
end

To find flux coupling relations between a pair of reactions, FCF_{woS} considers three cases (in contrast to the classical FCF algorithm). If both reactions are irreversible, the procedure is similar to the FCF method. However, if one reaction is reversible and the other is irreversible, the new RevToIrrevCoupling function is used. Finally, when both reactions are reversible, the coupling relations between one reaction and the forward or backward direction of the other is computed with RevToIrrevCoupling. The final coupling relation is inferred by CouplingRelationRecompute. Algorithm 7 summarizes the procedure.

Algorithm 7: FCF_{woS}, modified FCF algorithm without splitting reactions

Input:
- S (the $m \times n$ stoichiometric matrix)
- $Irr \subseteq \{1, \ldots, n\}$ (the set of irreversible reactions)

Output:
- Blk (the set of blocked reactions)
- $CouplingRelation$

Preprocessing:

`/* Finding blocked reactions and updating` Irr `*/`
$(Blk^\star, S^\star, Irr^\star) :=$ BlkFinderRevCorrector(S, Irr)
$Rev^\star := [n] \setminus (Irr^\star \cup Blk^\star)$

Main procedure:

foreach $i, j \in Irr^\star$ *with* $i < j$ **do**
  $min := minimize(v_i \mid \{S^\star v = 0, v_j = 1, v_r \geq 0 \text{ for all } r \in Irr^\star\})$
  $max := maximize(v_i \mid \{S^\star v = 0, v_j = 1, v_r \geq 0 \text{ for all } r \in Irr^\star\})$
  switch (min, max) **do**
    case $(0, \infty)$
      $CouplingRelation[i, j] :=$ "$\overset{Un}{\longleftrightarrow}$"
    case $(> 0, \infty)$
      $CouplingRelation[i, j] :=$ "$\longleftarrow$"
    case $(0, > 0)$
      $CouplingRelation[i, j] :=$ "$\longrightarrow$"
    case $(> 0, > 0)$
      if $(min = max)$ **then**
        $CouplingRelation[i, j] :=$ "$\Longleftrightarrow$"
      else
        $CouplingRelation[i, j] :=$ "$\longleftrightarrow$"
      end
    end
  end
end

Algorithm 7: FCF_{woS}, modified FCF algorithm without splitting reactions, continued

foreach $i \in Irr^\star$ *and* $j \in Rev^\star$ **do**
  $min := minimize(v_i \mid \{S^\star v = 0, v_j = 1, v_r \geq 0 \text{ for all } r \in Irr^\star\})$
  $max := maximize(v_i \mid \{S^\star v = 0, v_j = 1, v_r \geq 0 \text{ for all } r \in Irr^\star\})$
  $CouplingRelation[i,j] :=$ RevToIrrevCoupling(min, max)
end
foreach $i, j \in Rev^\star$ *with* $i < j$ **do**
  $min_1 := minimize(v_i \mid \{S^\star v = 0, v_j = 1, v_r \geq 0 \text{ for all } r \in Irr^\star\})$
  $max_1 := maximize(v_i \mid \{S^\star v = 0, v_j = 1, v_r \geq 0 \text{ for all } r \in Irr^\star\})$
  $min_2 := minimize(v_i \mid \{S^\star v = 0, v_j = -1, v_r \geq 0 \text{ for all } r \in Irr^\star\})$
  $max_2 := maximize(v_i \mid \{S^\star v = 0, v_j = -1, v_r \geq 0 \text{ for all } r \in Irr^\star\})$
  ⓐ := RevToIrrevCoupling(min_1, max_1)
  ⓑ := RevToIrrevCoupling(min_2, max_2)
  $CouplingRelation[i,j] :=$ CouplingRelationRecompute(ⓐ, ⓑ)
end

Implementation of $\text{FCF}_{RT,woS}$

$\text{FCF}_{\text{RT,woS}}$ is another improved version of the FCF algorithm, which has been designed and implemented in the present work. As before, reversible reactions are not split. In addition, the number of LPs to be solved is reduced by applying the Reversibility-Type prunings (Larhlimi and Bockmayr, 2006). The Algorithm 8 summarizes the procedure. The function SetToOrderedVector(A) receives a set of numbers and transforms it into an ordered vector. For example, SetToOrderedVec$(\{5,1,6,3\}) = (1,3,5,6)$.

Algorithm 8: $\text{FCF}_{\text{RT,woS}}$, modified FCF algorithm with RT-prunings and without splitting reactions

Input:
- S (the $m \times n$ stoichiometric matrix)
- $Irr \subseteq \{1, \ldots, n\}$ (the set of irreversible reactions)

Output:
- Blk (the set of blocked reactions)
- $CouplingRelation$

Preprocessing:

`/* Finding blocked reactions and updating` Irr `*/`
$(Blk^\star, S^\star, Irr^\star) :=$ BlkFinderRevCorrector(S, Irr)
$Rev^\star := [n] \setminus (Irr^\star \cup Blk^\star)$; $\omega :=$ SetToOrderedVec$(Rev^\star)$
$Irev := Irr^\star$
$K =$ NullSpace$(S_{[m],Rev^\star})$; $c :=$ NumberOfColumns(K)
foreach $i \in Rev^\star$ **do**
  if $K_{\{i\},[c]} = 0^T$ **then**
    $Prev := Prev \cup \{\omega_i\}$
  else $Frev := Frev \cup \{\omega_i\}$
end

Algorithm 8: $\text{FCF}_{\text{RT,woS}}$, modified FCF algorithm with RT-prunings and without splitting reactions, continued

Main Procedure:
foreach $i, j \in Irev$ *with* $i < j$ **do**
  $min := minimize(v_i \mid \{S^\star v = 0, v_j = 1, v_r \geq 0 \text{ for all } r \in Irr^\star\})$
  $max := maximize(v_i \mid \{S^\star v = 0, v_j = 1, v_r \geq 0 \text{ for all } r \in Irr^\star\})$
  switch (min, max) **do**
    case $(0, \infty)$
      $CouplingRelation[i, j] :=$ "$\overset{Un}{\longleftrightarrow}$"
    case $(> 0, \infty)$
      $CouplingRelation[i, j] :=$ "$\longleftarrow$"
    case $(0, > 0)$
      $CouplingRelation[i, j] :=$ "$\longrightarrow$"
    case $(> 0, > 0)$
      if $(min = max)$ **then**
        $CouplingRelation[i, j] :=$ "$\Longleftrightarrow$"
      else
        $CouplingRelation[i, j] :=$ "$\longleftrightarrow$"
      end
    end
  end
end
foreach $i \in Irev$ *and* $j \in Prev$ **do**
  $min := minimize(v_i \mid \{S^\star v = 0, v_j = 1, v_r \geq 0 \text{ for all } r \in Irr^\star\})$
  $max := maximize(v_i \mid \{S^\star v = 0, v_j = 1, v_r \geq 0 \text{ for all } r \in Irr^\star\})$
  $CouplingRelation[i, j] :=$ RevToIrrevCoupling(min, max)
end
foreach $i, j \in Prev$ *or* $i, j \in Frev$, *with* $i < j$ **do**
  $min_1 := minimize(v_i \mid \{S^\star v = 0, v_j = 1, v_r \geq 0 \text{ for all } r \in Irr^\star\})$
  $max_1 := maximize(v_i \mid \{S^\star v = 0, v_j = 1, v_r \geq 0 \text{ for all } r \in Irr^\star\})$
  $min_2 := minimize(v_i \mid \{S^\star v = 0, v_j = -1, v_r \geq 0 \text{ for all } r \in Irr^\star\})$
  $max_2 := maximize(v_i \mid \{S^\star v = 0, v_j = -1, v_r \geq 0 \text{ for all } r \in Irr^\star\})$
  ⓐ := RevToIrrevCoupling(min_1, max_1)
  ⓑ := RevToIrrevCoupling(min_2, max_2)
  $CouplingRelation[i, j]$:= CouplingRelationRecompute(ⓐ, ⓑ)
end

Implementation of FFCA

In FFCA (Feasibility-based Flux Coupling Analysis), we firstly classify the reactions into different subsets based on their reversibility types. Based on this classification, the feasibility of two LPs is checked in those cases where flux coupling is possible (see Section 4.2.1). Algorithm 9 summarizes the procedure.

Algorithm 9: FFCA, Feasibility-based flux coupling analysis

Input:
- S (the $m \times n$ stoichiometric matrix)
- $Irr \subseteq \{1, \ldots, n\}$ (the set of irreversible reactions)

Output:
- Blk (the set of blocked reactions)
- $CouplingRelation$

Preprocessing:

`/* Finding blocked reactions and updating` Irr `*/`

$(Blk^\star, S^\star, Irr^\star) := \textsc{BlkFinderRevCorrector}(S, Irr)$

$Rev^\star := [n] \setminus (Irr^\star \cup Blk^\star)$

$\omega := \textsc{SetToOrderedVec}(Rev^\star)$

$Irev := Irr^\star; Prev := \varnothing; Frev := \varnothing$

$K = \textsc{NullSpace}(S_{[m],Rev^\star})$

$c := \textsc{NumberOfColumns}(K)$

foreach $i \in Rev^\star$ **do**
  if $K_{\{i\},[c]} = 0^T$ **then**
    $Prev := Prev \cup \{\omega_i\}$
  else $Frev := Frev \cup \{\omega_i\}$
end

Main procedure:

foreach $i, j \in Irev$ *with* $i < j$ **do**
  $m_1 := maximize(0 \mid \{v_i = 1, v_j = 0, S^\star v = 0, v_r \geq 0 \text{ for all } r \in Irr^\star\})$
  $m_2 := maximize(0 \mid \{v_i = 0, v_j = 1, S^\star v = 0, v_r \geq 0 \text{ for all } r \in Irr^\star\})$
  if $m_1, m_2 \in \mathbb{R}$ **then**
    $CouplingRelation[i, j] :=$ "$\overset{Un}{\longleftrightarrow}$"
  else if $m_1 \in \mathbb{R}$ *and* $m_2 = \perp$ **then**
    $CouplingRelation[i, j] :=$ "$\longleftarrow$"
  else if $m_1 = \perp$ *and* $m_2 \in \mathbb{R}$ **then**
    $CouplingRelation[i, j] :=$ "$\longrightarrow$"
  else
    $CouplingRelation[i, j] :=$ "$\longleftrightarrow$"
    $min := minimize(v_i \mid \{v_j = 1, S^\star v = 0, v_r \geq 0 \text{ for all } r \in Irr^\star\})$
    $max := maximize(v_i \mid \{v_j = 1, S^\star v = 0, v_r \geq 0 \text{ for all } r \in Irr^\star\})$
    if $min = max$ **then** $CouplingRelation[i, j] :=$ "$\Longleftrightarrow$"
  end
end
foreach $i \in Prev$ *and* $j \in Irev$ **do**
  $m_1 := maximize(0 \mid \{v_i = 1, v_j = 0, S^\star v = 0, v_r \geq 0 \text{ for all } r \in Irr^\star\})$
  $m_2 := maximize(0 \mid \{v_i = -1, v_j = 0, S^\star v = 0, v_r \geq 0 \text{ for all } r \in Irr^\star\})$
  if $m_1 = \perp$ *and* $m_2 = \perp$ **then**
    $CouplingRelation[i, j] :=$ "$\longrightarrow$"
  else
    $CouplingRelation[i, j] :=$ "$\overset{Un}{\longleftrightarrow}$"
  end
end

Algorithm 9: FFCA, Feasibility-based flux coupling analysis, continued

foreach $i, j \in Prev$ *or* $i, j \in Frev$ **do**
  $m_1 := maximize(0 \mid \{v_i = 1, v_j = 0, S^\star v = 0\})$
  if $m_1 = \perp$ **then**
    $CouplingRelation[i, j] := \text{“}\Longleftrightarrow\text{”}$
  else
    $CouplingRelation[i, j] := \text{“}\overset{Un}{\longleftrightarrow}\text{”}$
  end
end

4.3.4 FCA and Random Sampling of the Flux Space

As explained in Chapter 1, functional relations of metabolic fluxes can be studied by uniform random sampling of the flux space (Price et al., 2004b). The random sampling is usually performed by COBRA toolbox (Becker et al., 2007).

Algorithm 10, which is suggested in Xi et al. (2011), is introduced as an FCA procedure based on randomly sampled vectors. Xi et al. (2011), correctly pointed out that although fully coupled reaction pairs always have $|\rho_{i,j}| = 1$, such a perfect correlation does not always imply $i \Longleftrightarrow j$. Therefore, this algorithm has a perfect sensitivity, but it may fail in finding with perfect specificity the set of fully coupled reactions. In the authors' words, "this approach is complete but not sound when it is used to calculate perfect Co-Sets" (Xi et al., 2011).

Nevertheless, one can think of a modified version of Algorithm 10 by devising a different strategy for computing the set of fully coupled reactions, e.g. the enzyme subsets method (Pfeiffer et al., 1999). However, even such a modified algorithm is not able to correctly compute the flux coupling relations for at least two reasons. Firstly, it should be emphasized that all FCA algorithms reported in Chapter 4 are exact methods, while Algorithm 10 is based on a finite set of random flux distributions. Therefore, this method does not guarantee that the computed flux coupling relations are valid for all possible feasible flux distributions. Secondly, at least with the current implementation of random sampling of vectors in the COBRA toolbox, (almost) always non-zero flux values are generated if the corresponding reactions are unblocked. Therefore, for (almost) every sampled vector, we will have $v_i \neq 0$ for all $i \in R \setminus Blk$. Therefore, there is a large chance that in the main procedure, both I and J sets become empty sets, and therefore, we incorrectly infer $i \longleftrightarrow j$ for every pair of reactions which are not fully coupled.

Algorithm 10: Alleged FCA algorithm based on randomly sampled flux distributions; based on 'Algorithm 2' in Xi et al. (2011)

Input:
- S (the $m \times n$ stoichiometric matrix)
- $Irr \subseteq \{1, \ldots, n\}$ (the set of irreversible reactions)
- $l, u \in \mathbb{R}^n$ (the lower and upper bound vectors)
- $\lambda \in \mathbb{N}$ (the number of fluxes to be sampled)

Output:
- Blk (the set of blocked reactions)
- $CouplingRelation$

Initialization:
$Blk := \varnothing; X := \varnothing$

Main procedure:
$X :=$ RandomSampling$(\lambda; S, Irr, l, u)$
foreach $r \in \{1, \ldots, n\}$ **do**
  if *for all* $w \in X$, $w_r = 0$ **then**
    $Blk := Blk \cup \{r\}$
  end
end
foreach $i, j \in R \setminus Blk$ *with* $i < j$ **do**
  $\rho_{i,j} :=$ CorrelationCoeff(i, j); ▷based on vectors in X
  if $|\rho_{i,j}| = 1$ **then**
    $CouplingRelation[i, j] :=$"$\Longleftrightarrow$"
  else
    $I := \{v \in X \mid v_i = 0\}$
    $J := \{v \in X \mid v_j = 0\}$
    if $I{=}J$ **then**
      $CouplingRelation[i, j] :=$"$\longleftrightarrow$"
    else if $I \supsetneq J$ **then**
      $CouplingRelation[i, j] :=$"$\longrightarrow$"
    else if $I \subsetneq J$ **then**
      $CouplingRelation[i, j] :=$"$\longleftarrow$"
    else $CouplingRelation[i, j] :=$"$\overset{Un}{\longleftrightarrow}$"
  end
end

Summary:

- FFCA, a feasibility-based FCA method is introduced. A corresponding software tool is also available for non-commercial use.
- On a set of benchmark metabolic networks, FFCA was compared to other FCA approaches, namely MMB-FCA, EFP-FCA and FCF. FFCA is the fastest FCA method for genome-scale networks (*H. pylori*, *E. coli* and *S. cerevisiae*), and the second best method (after MMB-FCA) for small-scale networks.
- We showed that FFCA is also faster than new improved implementations of the FCF algorithms with reversibility-type prunings and without splitting reversible reactions.

Chapter 5

Analysis of Metabolic Subnetworks by Flux Cone Projection

5.1 Introduction

Metabolic pathway analysis is the study of meaningful minimal pathways or routes of connected reactions in metabolic network models (Klamt and Stelling, 2003; Terzer et al., 2009). Two closely related concepts are often used for explaining such pathways: elementary modes (EMs) (Schuster and Hilgetag, 1994; Schuster et al., 2000) and extreme pathways (EXPAs) (Schilling et al., 2000). Mathematically speaking, EMs and EXPAs are generating sets of the flux cone (Klamt and Stelling, 2003; Jevremovic et al., 2010). Several approaches have been proposed for the computation of such pathways in metabolic models (Pfeiffer et al., 1999; Wagner, 2004; Gagneur and Klamt, 2004; Klamt et al., 2005; von Kamp and Schuster, 2006; Terzer and Stelling, 2008, 2010; Bell and Palsson, 2005).

EM and EXPA analysis are promising approaches for studying metabolic networks (Schilling et al., 1999; Trinh et al., 2009). However, due to the combinatorial explosion of the number of such pathways (Klamt and Stelling, 2002), this kind of analysis cannot be performed for "large" networks. Recent advances in the computation of EMs and extreme rays of polyhedral cones (Terzer and Stelling, 2008, 2010) has made it possible to compute tens of millions of EMs, but computing all EMs for large genome-scale networks may still be impossible. Additionally, we are usually interested a certain subset of reactions, not all of them. Therefore, even if the EMs are computable, possibly many of them are not interesting pathways because they are not related to the interesting reactions.

The goal of the present chapter is to introduce a new concept which can be useful in the analysis of fluxes in metabolic subnetworks. The chapter is organized as follows. Firstly, the mathematical concepts used in the chapter are formally defined. Secondly, we review the studies which have tried to investigate (some of) the EMs or EXPAs of large-scale networks. In the next step, we present the concept of Projected Cone EMs (ProCEMs) and we propose a method to compute them. We also compare ProCEMs with EFPs from the mathematical and computational point of view.

5.2 Formal Definitions

A polyhedron $Q \subseteq \mathbb{R}^n$ is a set of the form:

$$Q = \{v \in \mathbb{R}^n \mid A \cdot v \leq b\} \tag{5.1}$$

with a matrix $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$.

Let $E = \{1, \ldots, m\}$ and $F = \{1, \ldots, n\}$. The inequality $A_{\{i\};F} \cdot v \leq b_{\{i\}}$ is called redundant if $\{v \mid A_{E\setminus\{i\};F} \cdot v \leq b_{E\setminus\{i\}}\} = Q$.

Suppose that $\mathbb{A} \subseteq \mathbb{R}^p$ and $\mathbb{B} \subseteq \mathbb{R}^q$ for some $p, q \in \mathbb{N}$. Given a polyhedron $Q \subseteq \mathbb{A} \times \mathbb{B}$ the *projection* of Q onto $\mathbb{A}$ is defined as:

$$\mathcal{P}_{\mathbb{A}}(Q) = \{x \in \mathbb{A} \mid \exists y \in \mathbb{B}, (x, y) \in Q\} \tag{5.2}$$

Consider a metabolic network N with irreversible reactions only, and the corresponding flux cone $C \subseteq \mathbb{A} \times \mathbb{B}$. The set $\mathcal{P}_{\mathbb{A}}(C)$ is called the *projected cone* onto $\mathbb{A}$. Each generating vector j of $\mathcal{P}_{\mathbb{A}}(C)$ will be called a *projected cone elementary mode* (ProCEM). In the rest of this text, we assume that all reactions in metabolic networks are irreversible. This means that every reversible reactions should be split into two irreversible reactions.

Similarly, the projection of a vector $w \in \mathbb{A} \times \mathbb{B}$ onto $\mathbb{A}$ can be defined as:

$$\mathcal{P}_{\mathbb{A}}(w) = \{x \in \mathbb{A} \mid \exists y \in \mathbb{B}, (x, y) = w\} \tag{5.3}$$

A *projected elementary mode* (PEM) is the projection of an EM onto the subspace of interest. If a subnetwork (and not the complete network) is studied, PEMs might be more relevant than EMs, as they are in lower dimensions and easier to study. Currently, the only known method to compute PEMs is to enumerate the complete set of EMs and then project them onto the subspace of interest.

5.3 The State of the Art

As mentioned in the Section 5.1, the set of EMs of a genome-scale network may be large, and in general, it cannot be computed with the available tools. Even if this is possible, one cannot simply extract interesting information from it. Therefore, a subset of EMs (or in case that we are interested in a subset of reactions, the set of PEMs) should be computed to reduce the computation time and/or the output size of the pathway computation method. Several approaches to this problem have been proposed in the literature. These strategies can be classified into four main categories:

5.3.1 Computation of a Subset of EMs

The first strategy is to constrain the complete set of EMs (or EXPAs) to a subset describing a phenotype space or a set of phenotypic data. For example, Covert

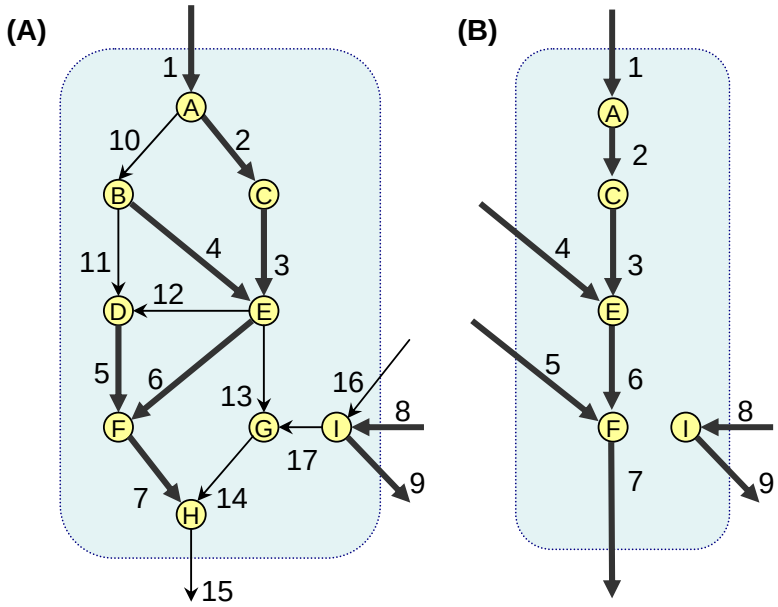
and Palsson (2003) showed that consideration of regulatory constraints in the analysis of a small "core metabolism" model can reduce the set of 80 EXPAs to a set of 2 to 26 EXPAs, depending on the applied regulatory constraints. On the other hand, Urbanczik (2007) suggested to compute "constrained" elementary modes which satisfy certain optimality criteria. As a result, instead of a full enumeration of EMs, only a subset of them should be computed, which results in a big computational gain. The idea of reducing the set of EMs is used recently by Song and Ramkrishna (2009) in an approach called *yield analysis*. In this approach, the yield space (or solution space) is defined as a bounded convex hull. Then, the minimal generating set spanning the yield space is recalculated, and therefore, all EMs with negligible contribution to the yield space can be excluded. The authors show that their method results in 91% reduction of the EM set for glucose/xylose-fermenting yeast.

5.3.2 Computation of EMs in Isolated Subsystems

A second strategy to focus on the EMs (or EXPAs) of interest is to select a (possibly disconnected) subnetwork, rather than the complete metabolic model, by assuming all other reactions and metabolites to be "external", and computing the EMs (or EXPAs) of this selected subsystem. This idea, i.e., cutting out subsystems or splitting big networks into several subsystems, is broadly used in the literature (e.g. see Nuño et al., 1997; Schuster et al., 1999; Schilling and Palsson, 2000; Schilling et al., 2002; Schuster et al., 2002b; Stelling et al., 2002; Çakır et al., 2004a; Schwarz et al., 2005; Verwoerd, 2007, 2011; Kim et al., 2008; Imielinski and Belta, 2008; Teusink et al., 2009; Kenanov et al., 2010). In some of these studies, not only the network boundary is redrawn, but also some reactions may be removed for further simplifying the network.

Although this strategy is useful, it can result in serious errors in the computational analysis of network properties (Kaleta et al., 2009). For example, dependencies and coupling relationships between reactions can be influenced by redrawing the system boundaries. As mentioned in Chapter 3, Burgard et al. (2004) showed that subsystem-based coupling analysis of the *H. pylori* network (Schilling et al., 2002) results in an incomplete detection of coupled reactions. Kaleta et al. (2009) suggest that neglecting such a coupling can lead to fluxes which are not part of any feasible EM in the original complete network. Existence of such infeasible "pathway fragments" (Imielinski and Belta, 2008) can result in incorrect conclusions.

To better understand this problem, we consider Figure 5.1A as an example. Let us assume that we are interested in a subnetwork composed of reactions $1, \ldots, 9$. This subnetwork is called SuN. If we simply assume "non-interesting" reactions and metabolites to be external reactions and metabolites, we will obtain the subsystem shown in Figure 5.1B. This subnetwork has only four EMs, two of which are not part of any feasible steady-state flux vector in the complete network. For example, the EM composed of reactions 5 and 7 in Figure 5.1B cannot appear



(A): A small metabolic network with 17 reactions. Metabolites are shown as nodes, while reactions are shown by arrows. Reactions 1, 8, 9, 15 and 16 are boundary reactions, while all other reactions are internal reactions. We might be interested only in a subnetwork containing nine reactions: $1, \ldots, 9$*, which are shown by thick arrows. This subnetwork will be called SuN. (B): The reduced subsystem comprising only the nine interesting reactions.*

in steady-state in the original complete network, because the coupling between reaction 1 and reaction 5 is broken. Therefore, analyzing this subnetwork instead of the complete original network can result in false conclusions.

5.3.3 Computation of Elementary Flux Patterns

We observed that some errors may appear in the analysis of isolated subsystems. One possible solution to this problem is to compute a "large" subset of PEMs, or alternatively, as suggested by Kaleta et al. (2009), to compute the support of a subset of PEMs. These authors proposed a procedure to compute the elementary flux patterns (EFPs) of a subnetwork within a genome-scale network. A flux pattern is defined as a set of reactions in a subnetwork that is included in the support of some steady-state flux vector of the entire network (Kaleta et al., 2009). A flux pattern is called an elementary flux pattern if it cannot be generated by combination of two or more different flux patterns. Each EFP is the support of (at least) one PEM. It is suggested that in many applications, the set of EFPs can be used instead of EMs (Kaleta et al., 2009).

Although EFPs are promising tools for the analysis of metabolic pathways, they also have their own shortcomings. The first important drawback of EFPs is that they cannot be used in place of EMs in certain applications (Gagneur and

EFPs	EFP set	ProCEMs	PEMs	vector
$E1$	$\{9\}$	$u1$	$p1$	(0, 0, 0, 0, 0, 0, 0, 0, 1)
$E2$	$\{8\}$	$u2$	$p2$	(0, 0, 0, 0, 0, 0, 0, 1, 0)
$E3$	$\{1, 4\}$	$u3$	$p3$	(1, 0, 0, 1, 0, 0, 0, 0, 0)
$E4$	$\{1, 2, 3\}$	$u4$	$p4$	(1, 1, 1, 0, 0, 0, 0, 0, 0)
$E5$	$\{1, 5, 7\}$	$u5$	$p5$	(1, 0, 0, 0, 1, 0, 1, 0, 0)
$E6$	$\{1, 4, 6, 7\}$	$u6$	$p6$	(1, 0, 0, 1, 0, 1, 1, 0, 0)
$E7$	$\{1, 2, 3, 6, 7\}$	$u7$	$p7$	(1, 1, 1, 0, 0, 1, 1, 0, 0)
–	–	$u8$	$p8$	(1, 1, 1, 0, 1, 0, 1, 0, 0)
–	–	$u9$	$p9$	(1, 0, 0, 1, 1, 0, 1, 0, 0)
–	–	–	$p10$	(0, 0, 0, 0, 0, 0, 0, 1, 1)

Table 5.1: *List of elementary flux patterns, projected elementary modes and projected cone elementary modes of SuN. Flux through reactions* $1, \ldots, 9$*, respectively, are the elements of the shown vectors. Zero vector and also the empty set are excluded.*

Klamt, 2004), where the precise flux values are required. For example, in the identification of all pathways with optimal yield (Schuster et al., 1999, 2001) and in the analysis of control-effective fluxes (Stelling et al., 2002; Çakır et al., 2004a; Zhao and Kurata, 2009), the flux values of the respective reactions in the EMs should be taken into account.

Another important shortcoming of EFP analysis is that it is possible to have very different EMs represented by the same EFP, since flux values are ignored in EFPs. For example, consider the case that two reactions i and j are partially coupled (Burgard et al., 2004). This means that there exist at least two EMs, say e and f, such that $e_i/e_j \neq f_i/f_j$ (see Chapter 2). However, if we consider a subnetwork composed of these two reactions, then we will only have one EFP, namely $\{i, j\}$. From the theoretical point of view, finding all EMs that correspond to a certain EFP can be NP-hard (see Theorem 2.7 in Acuña et al., 2010).

Every EFP is related to at least one EM in the original metabolic network. However, one of the limitations of EFP analysis is that EFPs are activity patterns of some EMs, not necessarily all of them. We will show this by an example. In Figure 5.1A, the flux cone is a subset of $\mathbb{R}^{17}$, while the subnetwork SuN induces a 9-dimensional subspace $\mathbb{A} = \mathbb{R}^9$. If G is the set of EMs in Figure 5.1A, then the set of PEMs can be computed as $P = \{\mathcal{P}_{\mathbb{A}}(e) \mid e \in G\}$. The set of the 10 PEMs of SuN in Figure 5.1A is shown in Table 5.1.

For the same network and subnetwork, we used EFPTools (Kaleta, 2009) to compute the set of the EFPs. The resulting 7 EFPs are presented in Table 5.1. If we compare the PEMs and EFPs in this table, we will find out that the support of each of the first 7 PEMs is equal to one of the EFPs. However, for the last three PEMs no corresponding EFP can be found in Table 5.1. This is due to the fact

that $supp(p8) = E4 \cup E5$, $supp(p9) = E3 \cup E5$, and $supp(p10) = E1 \cup E2$. Hence, the flux patterns corresponding to these PEMs are not elementary. Therefore, some EMs may exist in the network which have no corresponding EFP on a certain subnetwork. This means that by EFP analysis possibly many EMs of the original network can not be recovered. Informally speaking, the question is whether the set of EFPs is the largest set of PEM supports which can be computed without enumerating all EMs.

5.3.4 Projection Methods

A possible strategy to simplify the analysis of networks is to project the flux cone onto a lower-dimensional space of interest. In other words, if we are interested in a subnetwork, one can project the flux cone onto the low-dimensional subspace defined by the "interesting" reactions. Note that the projection of the flux cone is in general different from removing reactions from the network. Consider the simple network shown in Figure 5.2A and its corresponding flux cone in Figure 5.2B (i.e., the open triangle shown in light gray). This network has two EMs, which are the generating vectors of the flux cone, g_1 and g_2. Now, if we are interested in a subnetwork composed of reactions 1 and 2, then we can project the flux cone to the 2D subspace produced by these two reactions. This is comparable to light projection on a 3D object to make 2D shadows. The projected cone is shown in dark gray. When the flux cone is projected onto the low-dimensional space, new generating vectors may appear. In this example, g_1 and g_3 (in 2D space) are the generating vectors of the projected cone. This projected flux cone is certainly different from the flux cone of a network made by deleting reaction 3 (Figure 5.2C). Such a network has only one EM, and its corresponding flux cone can be generated by only one vector, namely, g_1.

Historically, the idea of flux cone projection has been considered in a few studies. Wiback and Palsson (2002) suggested that the space of cofactor production of red blood cell can be studied by projecting the cell-scale metabolic network onto a 2D subspace corresponding to ATP and NADPH production. A similar approach was used by Covert and Palsson (2003) and also by Wagner and Urbanczik (2005) to analyze the relationship between carbon uptake, oxygen uptake and biomass production. All the above studies considered very small networks. Therefore, the authors computed the extreme rays of the flux cone and then projected them onto the subspace of interest, without really projecting the flux cone. Urbanczik and Wagner (2005) later introduced the concept of elementary conversion modes (ECMs), which are in principle the extreme rays of the cone obtained by projecting the original flux cone onto the subspace of boundary reactions. They suggest that the extreme rays of this "conversion" cone, i.e., the ECMs, can be computed even for large networks (Urbanczik, 2006).

Following this idea, we introduce the ProCEM set ("Projected Cone Elementary Mode" set), which is the set of EMs of the projected flux cone. In contrast to Urbanczik and Wagner (2005), we formulate the problem in a way that any

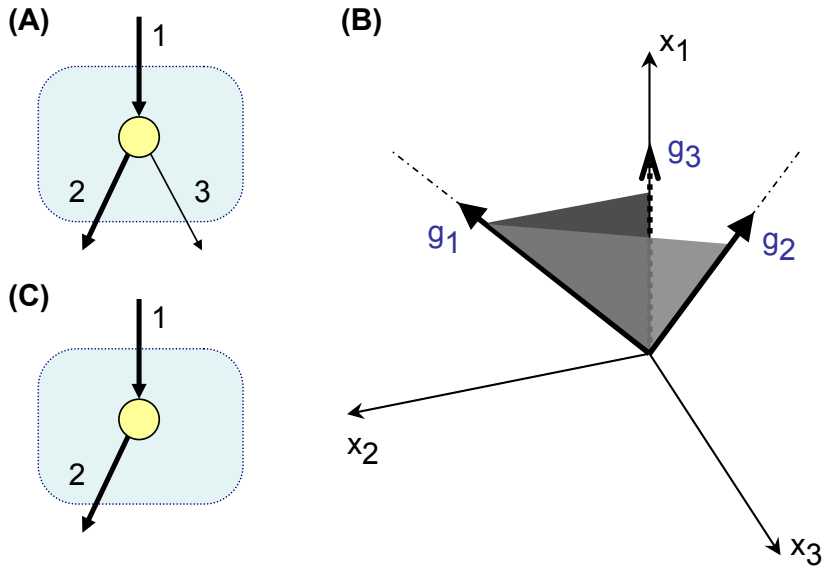


Figure 5.2: *(A): A small metabolic network. The reactions in the interesting subnetwork are shown as thick arrows. (B): The flux cone of this network, shown in light gray, can be generated by vectors g_1 and g_2. The projected cone is shown in dark gray. The projected cone can be generated by g_1 and g_3 in a 2D plane. (C): the same metabolic network as in A, but with reaction 3 removed. The flux cone of this network is generated by only one vector, namely g_1.*

subnetwork can be chosen, not only the boundary reactions. Additionally, we compare the closely related concepts of ProCEMs, PEMs and EFPs.

5.4 Method and Implementation

5.4.1 Computational Procedure

For computing ProCEMs of a certain subnetwork, we use block elimination algorithm (Balas and Pulleyblank, 1983), which is based on the "projection lemma" (see Section 5.4.2 in Breutel, 2004). The algorithm needs three input objects: the stoichiometric matrix (S) of the network; the binary vector of the reversibility type of reactions (Irr); and the set of reactions in the subnetwork of interest ($\Sigma \subseteq [n]$). The algorithm performs the following steps for computing ProCEMs:

Preprocessing: At the beginning, based on Σ we sort the columns of S in the form:

$$\bar{S} = \begin{pmatrix} \bar{A} & \bar{B} \end{pmatrix} \tag{5.4}$$

where the reaction corresponding to the i-th column belongs to Σ iff i is in $\bar{A}$. Then, the blocked reactions are removed. Finally, each of the reversible

reactions is split into two irreversible "forward" and" backward" reactions. The final stoichiometric matrix will be in the form:

$$S' = \begin{pmatrix} A & B \end{pmatrix} \tag{5.5}$$

where the columns of A represent the interesting reactions after splitting reversible reactions and removing the blocked reactions.

Cone Projection: For cone projection with the block elimination method, we consider the following (pointed) cone:

$$X = \{x \mid x \geq 0, H^T \cdot x = 0\} \tag{5.6}$$

where:

$$H = \begin{pmatrix} -B \\ B \\ I \end{pmatrix} \tag{5.7}$$

For the cone X, we enumerate the set of extreme rays, say $\{w_1, w_2, ..., w_k\}$. The projected cone P is given by (Breutel, 2004):

$$P = \{x \mid W \cdot G \cdot x \leq 0\} \tag{5.8}$$

where

$$W = \begin{pmatrix} w_1^T \\ w_2^T \\ ... \\ w_k^T \end{pmatrix} \tag{5.9}$$

and

$$G = \begin{pmatrix} -A \\ A \\ I \end{pmatrix} \tag{5.10}$$

This representation may contain a large number of redundant inequalities (Jones et al., 2008). The redundant inequalities are removed at this step.

Finding ProCEMs: In the final step, the extreme rays of the projected cone, i.e., the ProCEMs, are enumerated.

5.4.2 Implementation and Computational Experiments

Block elimination algorithm is implemented in MATLAB v7.5. In our implementation, *polco* tool v4.7.1 (Terzer and Stelling, 2008, 2010) is used for the enumeration of extreme rays (both for cone projection and for finding ProCEMs). For removing redundant inequalities, the software *redund* from the *lrslib* v4.2 is used (Avis, 2000). All computations were performed on a 64-bit Debian Linux system with Intel Core 2 Duo 3.0 GHz processor.

5.4.3 Dataset

The metabolic network model of red blood cell (RBC) (Wiback and Palsson, 2002) was used in this study. The network was taken from the example metabolic networks associated with CellNetAnalyzer (Klamt et al., 2007) and differs slightly from the original model.

5.5 Results and Discussion

5.5.1 Mathematical Relationships among PEMs, EFPs and ProCEMs

From Table 5.1, one can observe that the set of ProCEMs in Figure 1A is included in the set of PEMs. Additionally, the set of EFPs is included in the set of ProCEM supports. Here, we prove that these two properties are true in general. This means that analysis of ProCEMs has at least two advantages compared to the analysis of EFPs. Firstly, ProCEMs can tell us about the flux ratio of different reactions in an elementary mode, while EFPs can only tell us whether the reaction has a non-zero value in that mode. Secondly, enumeration of ProCEMs may result in modes which cannot be obtained by EFP analysis.

Theorem 5.1. *In a metabolic network N with irreversible reactions only, let J (resp. P) be the set of ProCEMs (resp. PEMs) for a given subspace $\mathbb{A}$. Then $J \subseteq P$.*

Proof. We have to show that for every $u \in J$ there exists an elementary mode $e \in C$ in N such that $\mathcal{P}_{\mathbb{A}}(e) \cong u$. We know that for any $u \in J$ there exists $v \in C$ such that $\mathcal{P}_{\mathbb{A}}(v) = u$.

Any $v \in C$ can be written in the form $v = \sum_{k=1}^{r} c_k \cdot e^k$, where $e^1, \ldots, e^r$ are elementary modes of N and $c_1, \ldots, c_r > 0$. It follows that $u = \mathcal{P}_{\mathbb{A}}(v) = \sum_{k=1}^{r} c_k \cdot \mathcal{P}_{\mathbb{A}}(e^k)$.

If all the vectors $\mathcal{P}_{\mathbb{A}}(e^k)$ are pairwise equivalent, u is a PEM.

Otherwise, u is a linear combination of at least two non-equivalent PEMs, which are vectors in $\mathcal{P}_{\mathbb{A}}(C)$. This implies that u is not an extreme ray of $\mathcal{P}_{\mathbb{A}}(C)$, in contradiction with Lemma 1 in Gagneur and Klamt (2004) saying that in

a metabolic network with irreversible reactions only, the EMs are exactly the extreme rays. ∎

Theorem 5.2. *In a metabolic network N with irreversible reactions only, let E (resp. J) be the set of EFPs (resp. ProCEMs) for a given subspace $\mathbb{A}$. Then, $E \subseteq \{supp(u) \mid u \in J\}$.*

Proof. Suppose that for some $F \in E$, there exists no $v \in J$ such that $F = supp(v)$. Since F is an EFP, there exists $p \in P$ such that $F = supp(p)$. It follows $p \notin J$, but $p \in \mathcal{P}_{\mathbb{A}}(C)$, where C is the flux cone. Therefore, there exist $r \geq 2$ different ProCEMs, say $u^1, \ldots, u^r \in J$, such that $p = \sum_{k=1}^{r} c_k \cdot u^k$, with $c_k > 0$ for all k. Since $u^k \geq 0$, for all k, we have $supp(p) = \bigcup_{k=1}^{r} supp(u^k)$, with $supp(u^k) \neq supp(p)$ for all k. Since $supp(u^k)$ is a flux pattern for all k, this is a contradiction with F being an EFP. ∎

5.5.2 Computing the Set of EFPs from the Set of ProCEMs

Here, we present a simple algorithm to show that it is possible to compute the set of EFPs when the set of ProCEMs is known. Algorithm 11 summarizes this procedure.

Algorithm 11: Computing the set of EFPs based on the set of ProCEMs

Input:
- J (the set of ProCEMs)

Output:
- E (the set of EFPs)

Initialization:
$E := \varnothing$

Main procedure:
foreach $u \in J$ **do**
  $Z := supp(u)$
  foreach $v \in J$ **do**
    if $supp(v) \subsetneq supp(u)$ **then**
      $Z := Z \setminus supp(v)$
    end
  end
  if $Z \neq \varnothing$ **then**
    $E := E \cup \{supp(u)\}$
  end
end

We know that the support of every ProCEM u is a flux pattern Z. In the main procedure, we check whether every such flux pattern is elementary or not. If Z is *not* elementary, then it is equal to the union of some other flux patterns. Therefore, if all other flux patterns which are subsets of $supp(u)$ are subtracted from Z, this set becomes empty. This algorithm has the complexity $\mathcal{O}(nq^2)$, where q is the number of ProCEMs and n is the number of reactions.

5.5.3 Comparing EFPs and ProCEMs

Analysis of Subnetworks in the Metabolic Network of RBC

In order to compare our approach (computation of ProCEMs) with enumeration of the EFPs, we tested these methods for the analysis of subnetworks in the RBC model (Wiback and Palsson, 2002). For the analysis of this network, we split every reversible reactions into one forward and one backward irreversible reaction. The resulting network contains 67 reactions, including 20 boundary reactions, and a total number of 811 EMs. For comparing the methods, the set of all boundary reactions was considered as the interesting subsystem, resulting in 502 PEMs.

When we computed the EFPs of this network by EFPTools (Kaleta, 2009), only 90 EFPs are determined. However, for the same subnetwork, we computed 252 ProCEMs. This means that the ProCEMs set covers more than half of the PEMs, while the EFPs set covers less than one fifth of the PEMs. These results confirm the relevance of using ProCEMs for the analysis of subnetworks.

In order to compare the computation of EFPs and ProCEMs, the following task was performed on the RBC model (Wiback and Palsson, 2002). In each iteration, a random subnetwork containing r reactions was selected. Then, EFPs and ProCEMs were computed. The task was repeated for different subnetwork sizes. The computational results can be found in Figure 5.3.

From Figure 5.3, it can be seen that EFP computation is faster than ProCEM computation for small subnetworks. However, when the subnetwork size r increases, computation of ProCEMs does not become slower, while computation of EFPs significantly slows down. This is an important observation, because the difference between the number of EFPs and ProCEMs also increases with r.

Analysis of Subnetworks in the Plastid Metabolic Network of *A. thaliana*

ProCEM analysis becomes important when PEMs cannot be computed. This may happen frequently in the analysis of large-scale metabolic networks, as memory consumption is a major challenge in computation of EMs Terzer and Stelling (2008). In such cases, cone projection might still be feasible.

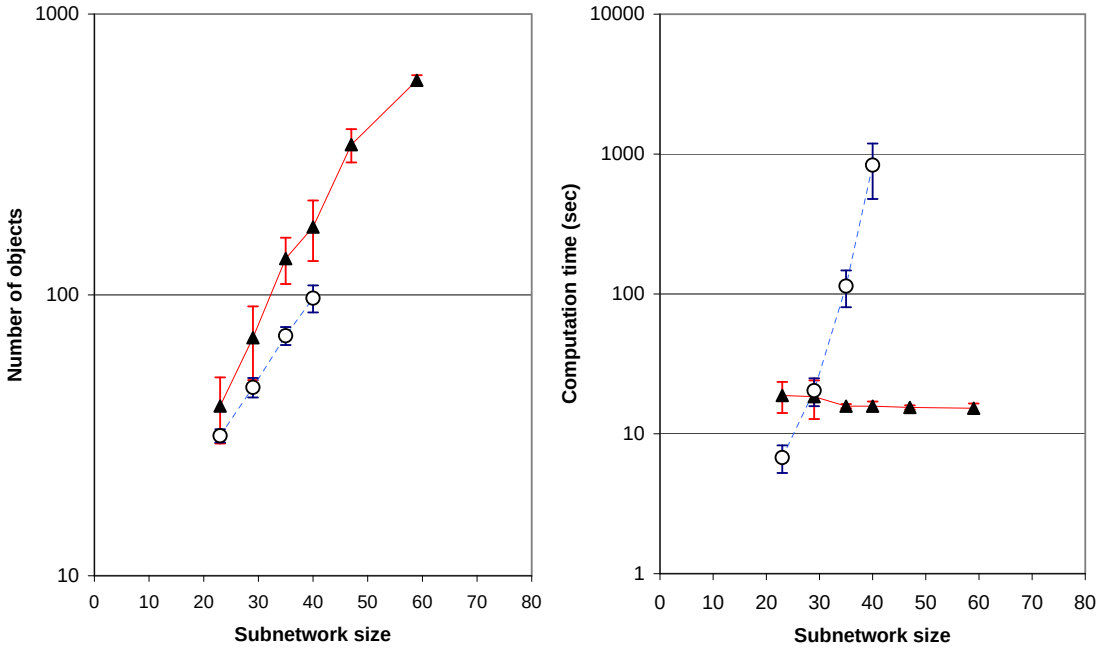


Figure 5.3: *Left: Number of ProCEMs and EFPs computed for random subnetworks of different sizes. Right: The computation times (per second) required for computing the ProCEMs and EFPs in the left chart. ▲: ProCEMs; ○: EFPs. Confidence intervals in this plot are based on one-sample t-test (95% c.i.). For large subnetworks ($r > 40$), we did not compute the EFPs because the program was very slow.*

As an example, the metabolic network of *A. thaliana* plastid was studied (Additional file 1). This network contains 102 metabolites and 123 reactions (205 reactions after splitting reversible reactions). Using efmtool (and also polco) Terzer and Stelling (2008), even after specifying 2 GB of memory, computation of EMs was not possible due to running out of memory. Therefore, for no subnetwork of the plastid network PEMs are computable. However, if the analysis is restricted to the 57 reactions involved in sugar and starch metabolism (see Additional file 1), one can compute the ProCEMs or EFPs of this subnetwork. We computed the ProCEMs as described in the Method and Implementation section, using a projection step size of 5 reactions. The complete set of 1310 ProCEMs was computed in approximately 15 minutes. However, when we tried to compute the set of EFPs using EFPTools Kaleta et al. (2009); Kaleta (2009), only 279 EFPs were computed after 4 days of running the program (270 EFPs were computed in the first two days). On the other hand, using a Matlab implementation of Algorithm 11, the complete set of 1054 EFPs was computed in 30 seconds. In conclusion, in metabolic networks for which the set of EMs cannot be enumerated, ProCEMs prove to be a useful concept to get insight into reaction activities.

Summary:

- When we are interested in a subnetwork, computation of all EMs is unnecessary. The interesting set of vectors is the set of projected EMs (PEMs). However, the only known method for computation of PEMs is to enumerate the EMs first.

- We introduce the concept of projected cone elementary modes (ProCEMs). This set is a subset of the PEMs set, and it can computed without enumeration of all EMs.

- The set of EFPs is a subset of the set of supports of ProCEMs. Additionally, in contrast to EFPs, ProCEMs contain the information about the flux values. Therefore, ProCEMs contain more information than EFPs.

- For large subnetworks, ProCEMs can be computed faster than EFPs.

Chapter 6 Conclusions and Further Research

Constraint-based methods are useful in the analysis of metabolic networks. Some of these methods, however, are not fast enough for being applied to genome-scale metabolic models. Additionally, the results may change if subsystems or subnetworks are studied instead of the complete metabolic models.

In **Chapter 2**, we show that when reactions are missing in a reconstructed metabolic model, flux coupling analysis may result in wrong conclusions about (un-)coupling relations for certain reaction pairs. We also prove that only certain changes in flux coupling relations are possible. For example, it is impossible to have two coupled reactions in the original network that are uncoupled in the incomplete reconstructed model, while the opposite is possible. Therefore, the results of flux coupling analysis should be used with care.

Since flux uncoupling is not influenced by missing network contents, we believe that further research on the properties of flux uncoupling is needed. For instance, if two reactions are coupled to each other, they cannot appear in a minimal cut set (Klamt and Gilles, 2004). Therefore, it is interesting to see how cut sets are related to flux uncoupling. While the relationship between evolution of metabolic genes and flux coupling has already been addressed by a number of authors (Pál et al., 2005*a*,*b*; Notebaart et al., 2009; Seshasayee et al., 2009) it would be interesting to see how genes evolve when their corresponding fluxes are uncoupled.

We believe that our findings are useful for both, the reconstruction of genome-scale metabolic networks and the validation of current models. For example, imagine two reactions that are expected to be (fully) coupled in a genome-scale metabolic model, while the corresponding genes show a very low gene expression correlation. In this case, some alternative pathway may exist in the real system which, however, is missing in the reconstructed model. If we find two reactions that are assumed to be uncoupled but have a high gene expression correlation, it is fair to assume that the corresponding genes are wrongly annotated and/or wrongly added to the metabolic model. We are currently studying how such a "model-driven gap-filling approach" (Feist et al., 2009) may lead to genome-scale models with better predictive capabilities.

In **Chapter 3**, we show that the analyzing a subsystem instead of the complete network can lead to certain changes in flux coupling. In particular, a pair of (fully, partially or directionally) coupled reactions may be detected as uncoupled

in the chosen subsystem, but not vice versa. Interestingly, this behavior is the opposite of the flux coupling changes that may happen due to the existence of missing reactions, or equivalently, deletion of reactions (Chapter 2). We emphasize that it is better to perform FCA in genome-scale networks, even when we focus on the reactions in a certain subsystem.

With the analysis of real-world metabolic models, we observed that a non-negligible number of reaction pairs in the plastids of modern plants have altered flux coupling relations when the plastids are studied in isolation. We suggested that the fraction of altered coupling relations can be seen as a measure of "dependence" of the subsystem. In contrast to the previous cases, we observed that in case of *C. reinhardtii* plastid, and also in case of all mitochondrial subsystems, a relatively small fraction of the reaction pairs will have different flux coupling relations if these subsystems are studied in isolation. This small change suggests that these organelles are relatively independent of the rest of the metabolic network. One may ask how small is this fraction for a randomly selected subsystem. This question is yet to be answered.

In **Chapter 4** we introduce FFCA as a new method for flux coupling analysis, and proved it to be faster than any other available approach. Our implementation of FFCA is fast enough to perform FCA for every pair of reactions in *S. cerevisiae* and *E. coli* genome-scale networks in a few hours. A corresponding software tool is available for non-commercial use and we recommend it for FCA of genome-scale networks.

The only known method for computation of projected EMs is to enumerate the EMs first. In **Chapter 5**, we introduce the concept of projected cone elementary modes (ProCEMs). The set of ProCEMs is a set of vectors which covers more PEMs than EFPs. Therefore, ProCEMs contain more information compared to EFPs. The ProCEMs set is computable without enumerating all EMs. Is there a bigger set of vectors that covers more PEMs and does not require full enumeration of EMs? This is yet to be answered.

One possible extension to this work is to use a more efficient implementation of polyhedral projection. With such an implementation, analysis of different subnetworks in genome-scale network models using ProCEMs is an interesting possibility for further research. For example, the ProCEMs can be used in the identification of all pathways with optimal yield (Schuster et al., 1999) and in the analysis of control-effective fluxes (Stelling et al., 2002).

BIBLIOGRAPHY

Acuña, V., Marchetti-Spaccamela, A., Sagot, M. and Stougie, L. (2010), 'A note on the complexity of finding and enumerating elementary modes', *BioSystems* **99**, 210–214.

Aloy, P. and Russell, R. B. (2006), 'Structural systems biology: modelling protein interactions', *Nature Reviews Molecular Cell Biology* **7**, 188–197.

Andrianantoandro, E., Basu, S., Karig, D. K. and Weiss, R. (2006), 'Synthetic biology: new engineering rules for an emerging discipline', *Molecular Systems Biology* **2**, 2006.0028.

Avis, D. (2000), *lrs*: A revised implementation of the reverse search vertex enumeration algorithm, *in* G. Kalai and G. Ziegler, eds, 'Polytopes - Combinatorics and Computation, Oberwolfach Seminars', Birkhäuser-Verlag, pp. 177–198.

Balas, E. and Pulleyblank, W. (1983), 'The perfectly matchable subgraph polytope of a bipartite graph', *Networks* **13**, 495–516.

Baldan, P., Cocco, N., Marin, A. and Simeoni, M. (2010), 'Petri nets for modelling metabolic pathways: a survey', *Natural Computing* **9**, 955–989.

Beard, D. A., Liang, S.-d. and Qian, H. (2002), 'Energy balance for analysis of complex metabolic networks', *Biophysical Journal* **83**, 79–86.

Beard, D. A. and Qian, H. (2005), Constraint-based modeling of metabolomic systems, *in* M. Dunn, ed., 'Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics', Vol. 3, John Wiley & Sons, Inc, Chichester, p. g308203.

Becker, S. A., Feist, A. M., Mo, M. L., Hannum, G., Palsson, B. O. and Herrgard, M. J. (2007), 'Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox', *Nature Protocols* **2**, 727–738.

Bell, S. L. and Palsson, B. O. (2005), 'Expa: a program for calculating extreme pathways in biochemical reaction networks', *Bioinformatics* **21**, 1739–1740.

Bensaude-Vincent, B. (2004), 'Two Cultures of Nanotechnology?', *HYLE - International Journal for Philosophy of Chemistry* **10**, 65–82.

Bockmayr, A. and Weispfenning, V. (2001), Solving numerical constraints, *in* A. Robinson and A. Voronkov, eds, 'Handbook of Automated Reasoning', Elsevier B.V., Amsterdam, pp. 751–842.

Breutel, S. W. (2004), Analysing the Behaviour of Neural Networks, PhD thesis, Queensland University of Technology.

Bundy, J. G., Papp, B., Harmston, R., Browne, R. A., Clayson, E. M., Burton, N., Reece, R. J., Oliver, S. G. and Brindle, K. M. (2007), 'Evaluation of predicted network modules in yeast metabolism using NMR-based metabolite profiling', *Genome Research* **17**, 510–519.

Burgard, A. P., Nikolaev, E. V., Schilling, C. H. and Maranas, C. D. (2004), 'Flux coupling analysis of genome-scale metabolic network reconstructions', *Genome Research* **14**, 301–312.

Çakır, T., Kırdar, B. and Ülgen, K. O. (2004a), 'Metabolic pathway analysis of yeast strengthens the bridge between transcriptomics and metabolic networks', *Biotechnology and Bioengineering* **86**, 251–260.

Çakır, T., Tacer, C. S. and Ülgen, K. Ö. (2004b), 'Metabolic pathway analysis of enzyme-deficient human red blood cells', *BioSystems* **78**, 49–67.

Chandrasekaran, S. and Price, N. D. (2010), 'Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*', *Proceedings of the National Academy of Sciences of the United States of America* **107**, 17845–17850.

Clarke, B. L. (1980), 'Stability of complex reaction networks', *Advances in Chemical Physics* **43**, 1–216.

Clarke, B. L. (1988), 'Stoichiometric network analysis', *Cell Biophysics* **12**, 237–253.

Covert, M. W. and Palsson, B. O. (2002), 'Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*', *Journal of Biological Chemistry* **277**, 28058–64.

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

Covert, M. W., Xiao, N., Chen, T. J. and Karr, J. R. (2008), 'Integrating metabolic, transcriptional regulatory and signal transduction models in *Escherichia coli*', *Bioinformatics* **24**, 2044–2050.

Dandekar, T., Moldenhauer, F., Bulik, S., Bertram, H. and Schuster, S. (2003), 'A method for classifying metabolites in topological pathway analyses based on minimization of pathway number', *BioSystems* **70**, 255–270.

de Figueiredo, L. F., Schuster, S., Kaleta, C. and Fell, D. A. (2009), 'Erratum: Can sugars be produced from fatty acids? a test case for pathway analysis tools', *Bioinformatics* **25**, 152–158.

de Oliveira Dal'Molin, C. G., Quek, L.-E., Palfreyman, R. W., Brumbley, S. M. and Nielsen, L. K. (2010), 'AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis.', *Plant Physiology* **152**, 579–589.

Dickerson, J. A., Berleant, D., Cox, Z., Qi, W. and Wurtele, E. (2001), Creating metabolic network models using text mining and expert knowledge, *in* C. Wu, P. Wang and J. Wang, eds, 'Proceedings of the Atlantic Symposium on Computational Biology and Genome Information Systems and Technology, GBGIST 2001; Durham, NC', pp. 26–30.

Ding, D., Ding, Y., Cai, Y., Chen, S. and Xu, W. (2008), 'Exploring poly-beta-hydroxy-butyrate metabolism through network-based extreme pathway analysis', *Rivista di Biologia* **101**, 67–80.

Duarte, N. C., Becker, S. A., Jamshidi, N., Thiele, I., Mo, M. L., Vo, T. D., Srivas, R. and Palsson, B. O. (2007), 'Global reconstruction of the human metabolic network based on genomic and bibliomic data', *Proceedings of the National Academy of Sciences of the United States of America* **104**, 1777–1782.

Duarte, N. C., Herrgård, M. J. and Palsson, B. O. (2004), 'Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model', *Genome Research* **14**, 1298–1309.

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

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

Edwards, J. S. and Palsson, B. O. (1998), 'How will bioinformatics influence metabolic engineering?', *Biotechnology and Bioengineering* **58**, 162–169.

Feist, A., Herrgård, M., Thiele, I., Reed, J. and Palsson, B. (2009), 'Reconstruction of biochemical networks in microorganisms', *Nature Reviews Microbiology* **7**, 129–143.

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

Feist, A. M. and Palsson, B. O. (2008), 'The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*', *Nature Biotechnology* **26**, 659–667.

Feist, A. M. and Palsson, B. O. (2010), 'The biomass objective function', *Current Opinion in Microbiology* **13**, 344–349.

Fell, D. A., Poolman, M. G. and Gevorgyan, A. (2010), 'Building and analysing genome-scale metabolic models', *Biochemical Society Transactions* **38**, 1197–1201.

Fell, D. A. and Small, J. R. (1986), 'Fat synthesis in adipose tissue. an examination of stoichiometric constraints', *Biochemical Journal* **238**, 781–786.

Fukuda, K. and Prodon, A. (1996), Double description method revisited, *in* 'Combinatorics and Computer Science: 8th Franco-Japanese and 4th Franco-Chinese Conference. Brest, France', Vol. 1120 of *Lecture Notes in Computer Science*, pp. 91–111.

Gagneur, J. and Klamt, S. (2004), 'Computation of elementary modes: a unifying framework and the new binary approach', *BMC Bioinformatics* **5**, 175.

Gevorgyan, A., Poolman, M. G. and Fell, D. A. (2008), 'Detection of stoichiometric inconsistencies in biomolecular models', *Bioinformatics* **24**, 2245–51.

Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R. Y., Algire, M. A., Benders, G. A., Montague, M. G., Ma, L., Moodie, M. M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N., Andrews-Pfannkoch, C., Denisova, E. A., Young, L., Qi, Z. Q., Segall-Shapiro, T. H., Calvey, C. H., Parmar, P. P., Hutchison, C. A., Smith, H. O. and Venter, J. C. (2010), 'Creation of a bacterial cell controlled by a chemically synthesized genome', *Science* **52**, 52–56.

Grafahrend-Belau, E., Schreiber, F., Koschützki, D. and Junker, B. H. (2009), 'Flux balance analysis of barley seeds: A computational approach to study systemic properties of central metabolism', *Plant Physiology* **149**, 585–598.

Gross, J. and Bhattacharya, D. (2009), 'Mitochondrial and plastid evolution in eukaryotes: an outsiders' perspective', *Nature Reviews Genetics* **10**, 495–505.

Gudmundsson, S. and Thiele, I. (2010), 'Computationally efficient flux variability analysis', *BMC Bioinformatics* **11**, 489.

Haus, U.-U., Klamt, S. and Stephen, T. (2008), 'Computing knock-out strategies in metabolic networks', *Journal of Computational Biology* **15**, 259–268.

Hogg, R. V. and Tanis, E. A. (1988), *Probability and Statistical Inference*, Prentice-Hall, New Jersey.

Hood, L. (1998), 'Systems biology: new opportunities arising from genomics, proteomics and beyond', *Experimental Hematology* **26**, 681.

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

Imielinski, M. and Belta, C. (2008), 'Exploiting the pathway structure of metabolism to reveal high-order epistasis', *BMC Systems Biology* **2**, 40.

Jamshidi, N. and Palsson, B. O. (2006), 'Systems biology of SNPs', *Molecular Systems Biology* **2**, 38.

Jevremovic, D., Trinh, C. T., Srienc, F. and Boley, D. (2010), 'On algebraic properties of extreme pathways in metabolic networks', *Journal of Computational Biology* **17**, 107–119.

Jones, C. N., Kerrigan, E. C. and Maciejowski, J. M. (2008), 'On polyhedral projection and parametric programming', *Journal of Optimization Theory and Applications* **138**, 207–220.

Joyce, A. R. and Palsson, B. O. (2006), 'The model organism as a system: integrating 'omics' data sets', *Nature Reviews Molecular Cell Biology* **7**, 198–210.

Kaleta, C. (2009), 'EFPTools, for computing elementary flux patterns (EFPs)', `http://users.minet.uni-jena.de/~m3kach/EFPA/`.

Kaleta, C., de Figueiredo, L. F. and Schuster, S. (2009), 'Can the whole be less than the sum of its parts? Pathway analysis in genome-scale metabolic networks using elementary flux patterns', *Genome Research* **19**, 1872–1883.

Kenanov, D., Kaleta, C., Petzold, A., Hoischen, C., Diekmann, S., Siddiqui, R. A. and Schuster, S. (2010), 'Theoretical study of lipid biosynthesis in wild-type *Escherichia coli* and in a protoplast-type L-form using elementary flux mode analysis', *FEBS Journal* **277**, 1023–1034.

Kholodenko, B. N., Schuster, S., Rohwer, J. M., Cascante, M. and Westerhoff, H. V. (1995), 'Composite control of cell function: Metabolic pathways behaving as single control units', *FEBS Letters* **368**, 1–4.

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

Kitano, H. (2000), 'Perspectives on systems biology', *New Generation Computing* **18**, 199–216.

Kitano, H. (2002), 'Systems biology: a brief overview', *Science* **295**, 1662–1664.

Kitano, H., Hamahashi, S. and Luke, S. (1998), 'The perfect *C. elegans* project: an initial report', *Artificial Life* **4**, 141–156.

Klamt, S. (2006), 'Generalized concept of minimal cut sets in biochemical networks', *BioSystems* **83**, 233–247.

Klamt, S., Gagneur, J. and von Kamp, A. (2005), 'Algorithmic approaches for computing elementary modes in large biochemical reaction networks', *IEE Proceedings - Systems Biology* **152**, 249–255.

Klamt, S. and Gilles, E. D. (2004), 'Minimal cut sets in biochemical reaction networks', *Bioinformatics* **20**, 226–234.

Klamt, S., Saez-Rodriguez, J. and Gilles, E. D. (2007), 'Structural and functional analysis of cellular networks with CellNetAnalyzer', *BMC Systems Biology* **1**, 2.

Klamt, S. and Stelling, J. (2002), 'Combinatorial complexity of pathway analysis in metabolic networks', *Molecular Biology Reports* **29**, 233–236.

Klamt, S. and Stelling, J. (2003), 'Two approaches for metabolic pathway analysis?', *Trends in Biotechnology* **21**, 64–69.

Kremling, A., Bettenbrock, K. and Gilles, E. D. (2007), 'Analysis of global control of *Escherichia coli* carbohydrate uptake', *BMC Systems Biology* **1**, 42.

Larhlimi, A. (2008), New concepts and tools in constraint-based analysis of metabolic networks, PhD thesis, Freie Universität Berlin.

Larhlimi, A. and Bockmayr, A. (2006), A new approach to flux coupling analysis of metabolic networks, *in* 'Computational Life Sciences II, Second International Symposium (CompLife 2006), Cambridge, UK', Vol. 4216 of *Lecture Notes in Computer Science*, pp. 205–215.

Larhlimi, A. and Bockmayr, A. (2008), On inner and outer descriptions of the steady-state flux cone of a metabolic network, *in* 'Proceedings of the 6th International Conference on Computational Methods in Systems Biology (CMSB 2008)', Vol. 5307 of *Lecture Notes in Computer Science*, Rostock, Germany, pp. 308–327.

Larhlimi, A. and Bockmayr, A. (2009), 'A new constraint-based description of the steady-state flux cone of metabolic networks', *Discrete Applied Mathematics* **157**, 2257–2266.

Lee, D. S., Park, J., Kay, K. A., Christakis, N. A., Oltvai, Z. N. and Barabasi, A. L. (2008), 'The implications of human metabolic network topology for disease comorbidity', *Proceedings of the National Academy of Sciences of the USA* **105**, 9880–9885.

Lee, K. Y., Park, J. M., Kim, T. Y., Yun, H. and Lee, S. Y. (2010), 'The genome-scale metabolic network analysis of *Zymomonas mobilis* ZM4 explains physiological features and suggests ethanol and succinic acid production strategies', *Microbial Cell Factories* **9**, 94.

Liao, J. C., Hou, S. Y. and Chao, Y. P. (1996), 'Pathway analysis, engineering, and physiological considerations for redirecting central metabolism', *Biotechnology and Bioengineering* **52**, 129–140.

Llaneras, F. and Picó, J. (2010), 'Which metabolic pathways generate and characterize the flux space? A comparison among elementary modes, extreme pathways and minimal generators', *Journal of Biomedicine and Biotechnology* **2010**, 753904.

Lougee-Heimer, R. (2003), 'The Common Optimization INterface for Operations Research: Promoting open-source software in the operations research community', *IBM Journal of Research and Development* **47**, 57–66.

Mahadevan, R. and Schilling, C. H. (2003), 'The effects of alternate optimal solutions in constraint-based genome-scale metabolic models', *Metabolic Engineering* **5**, 264–276.

Manichaikul, A., Ghamsari, L., Hom, E. F. Y., Lin, C., Murray, R. R., Chang, R. L., Balaji, S., Hao, T., Shen, Y., Chavali, A. K., Thiele, I., Yang, X., Fan, C., Mello, E., Hill, D. E., Vidal, M., Salehi-Ashtiani, K. and Papin, J. A. (2009), 'Metabolic network analysis integrated with transcript verification for sequenced genomes.', *Nature Methods* **6**, 589–592.

Marashi, S.-A. and Bockmayr, A. (2011), 'Flux coupling analysis of metabolic networks is sensitive to missing reactions', *BioSystems* **103**, 57–66.

Medema, M. H., Breitling, R., Bovenberg, R. and Takano, E. (2011), 'Exploiting plug-and-play synthetic biology for drug discovery and production in microorganisms', *Nature Reviews Microbiology* **9**, 131–137.

Mesarović, M. D., ed. (1968), *Systems Theory and Biology*, Springer-Verlag, New York.

Mestres, J., Gregori-Puigjané, E., Valverde, S. and Solé, R. V. (2008), 'Data completeness — the Achilles heel of drug-target networks', *Nature Biotechnology* **26**, 983–984.

Mo, M. L. and Palsson, B. O. (2009), 'Understanding human metabolic physiology: a genome-to-systems approach', *Trends in Biotechnology* **27**, 37–44.

Notebaart, R. A., Kensche, P. R., Huynen, M. A. and Dutilh, B. E. (2009), 'Asymmetric relationships between proteins shape genome evolution', *Genome Biology* **10**, R19.

Notebaart, R. A., Teusink, B., Siezen, R. J. and Papp, B. (2008), 'Co-regulation of metabolic genes is better explained by flux coupling than by network distance', *PLoS Computational Biology* **4**, e26.

Nuño, J. C., Sánchez-Valdenebro, I., Pérez-Iratxeta, C., Meléndez-Hevia, E. and Montero, F. (1997), 'Network organization of cell metabolism: monosaccharide interconversion', *Biochemical Journal* **324**, 103–111.

Oberhardt, M. A., Palsson, B. O. and Papin, J. A. (2009), 'Applications of genome-scale metabolic reconstructions', *Molecular Systems Biology* **5**, 320.

Orth, J. D. and Palsson, B. O. (2010), 'Systematizing the generation of missing metabolic knowledge', *Biotechnology and Bioengineering* **107**, 403–412.

Orth, J. D., Thiele, I. and Palsson, B. O. (2010), 'What is flux balance analysis?', *Nature Biotechnology* **28**, 245–8.

Pál, C., Papp, B. and Lercher, M. J. (2005*a*), 'Adaptive evolution of bacterial metabolic networks by horizontal gene transfer', *Nature Genetics* **37**, 1372–1375.

Pál, C., Papp, B. and Lercher, M. J. (2005*b*), 'Horizontal gene transfer depends on gene content of the host', *Bioinformatics* **21**, ii222–ii223.

Palsson, B. (2000), 'The challenges of in silico biology', *Nature Biotechnology* **18**, 1147–1150.

Palsson, B. O. (2006), *Systems Biology: Properties of Reconstructed Networks*, Cambridge University Press, New York.

Papin, J. A., Price, N. D. and Palsson, B. O. (2002), 'Extreme pathway lengths and reaction participation in genome-scale metabolic networks', *Genome Research* **12**, 1889–1900.

Papin, J. A., Reed, J. L. and Palsson, B. O. (2004), 'Hierarchical thinking in network biology: The unbiased modularization of biochemical networks', *Trends in Biochemical Sciences* **29**, 641–647.

Patil, K. R. (2003), Systems biology of metabolic networks: uncovering regulatory and stoichiometric principles, PhD thesis, Technical University of Denmark.

Perumal, D., Samal, A., Sakharkar, K. R. and Sakharkar, M. K. (2011), 'Targeting multiple targets in *Pseudomonas aeruginosa* PAO1 using flux balance analysis of a reconstructed genome-scale metabolic network', *Journal of Drug Targeting* **19**, 1–13.

Pfeiffer, T., Sánchez-Valdenebro, I., Nuño, J. C., Montero, F. and Schuster, S. (1999), 'METATOOL: for studying metabolic networks', *Bioinformatics* **15**, 251–257.

Pfeiffer, T., Soyer, O. S. and Bonhoeffer, S. (2005), 'The evolution of connectivity in metabolic networks', *PLoS Biology* **3**(7), e228.

Poolman, M. G., Fell, D. A. and Raines, C. A. (2003), 'Elementary modes analysis of photosynthate metabolism in the chloroplast stroma', *European Journal of Biochemistry* **270**, 430–439.

Poolman, M. G., Sebu, C., Pidcock, M. K. and Fell, D. A. (2007), 'Modular decomposition of metabolic systems via null-space analysis', *Journal of Theoretical Biology* **249**, 691–705.

Price, M. N., Arkin, A. P. and Alm, E. J. (2006), 'The life-cycle of operons', *PLoS Genetics* **2**, e96.

Price, N. D., Papin, J. A. and Palsson, B. O. (2002), 'Determination of redundancy and systems properties of the metabolic network of *Helicobacter pylori* using genome-scale extreme pathway analysis', *Genome Research* **12**, 760–769.

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

Price, N. D., Reed, J. L. and Palsson, B. O. (2004a), 'Genome-scale models of microbial cells: evaluating the consequences of constraints', *Nature Reviews Microbiology* **2**, 886–897.

Price, N. D., Schellenberger, J. and Palsson, B. O. (2004b), 'Uniform sampling of steady-state flux spaces: means to design experiments and to interpret enzymopathies', *Biophysical Journal* **87**, 2172–2186.

Puchałka, J., Oberhardt, M. A., Godinho, M., Bielecka, A., Regenhardt, D., Timmis, K. N., Papin, J. A. and Martins Dos Santos, V. A. P. (2008), 'Genome-scale reconstruction and analysis of the *Pseudomonas putida* KT2440 metabolic network facilitates applications in biotechnology', *PLoS Computational Biology* **4**, e1000210.

Raghunathan, A., Reed, J., Shin, S., Palsson, B. O. and Daefler, S. (2009), 'Constraint-based analysis of metabolic capacity of *Salmonella typhimurium* during host-pathogen interaction', *BMC Systems Biology* **3**, 38.

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. and Barabási, A. L. (2002), 'Hierarchical organization of modularity in metabolic networks', *Science* **297**, 1551–5.

Reed, J. L. (2009), Descriptive and predictive applications of constraint-based metabolic models, *in* 'Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2009)', Vol. 2009, pp. 5460–5463.

Reed, J. L., Patel, T. R., Chen, K. H., Joyce, A. R., Applebee, M. K., Herring, C. D., Bui, O. T., Knight, E. M., Fong, S. S. and Palsson, B. O. (2006), 'Systems approach to refining genome annotation', *Proceedings of the National Academy of Sciences of the United States of America* **103**, 17480–4.

Reed, J. L., Vo, T. D., Schilling, C. H. and Palsson, B. O. (2003), 'An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR)', *Genome Biology* **4**, R54.

Sagan, L. (1967), 'On the origin of mitosing cells', *Journal of Theoretical Biology* **14**, 225–274.

Satish Kumar, V., Dasika, M. S. and Maranas, C. D. (2007), 'Optimization based automated curation of metabolic reconstructions', *BMC Bioinformatics* **8**, 212.

Schellenberger, J. and Palsson, B. O. (2009), 'Use of randomized sampling for analysis of metabolic networks', *Journal of Biological Chemistry* **284**, 5457–5461.

Schilling, C. H., Covert, M. W., Famili, I., Church, G. M., Edwards, J. S. and Palsson, B. O. (2002), 'Genome-scale metabolic model of *Helicobacter pylori* 26695', *Journal of Bacteriology* **184**, 4582–4593.

Schilling, C. H., Letscher, D. and Palsson, B. O. (2000), '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**, 229–248.

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

Schilling, C. H., Schuster, S., Palsson, B. O. and Heinrich, R. (1999), 'Metabolic pathway analysis: Basic concepts and scientific applications in the post-genomic era', *Biotechnology Progress* **15**, 296–303.

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

Schrijver, A. (1986), *Theory of Linear and Integer Programming*, John Wiley & Sons, Inc., New York.

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

Schuster, S., Dandekar, T. and Fell, D. A. (1999), 'Detection of elementary flux modes in biochemical networks: A promising tool for pathway analysis and metabolic engineering', *Trends in Biotechnology* **17**, 53–60.

Schuster, S., Fell, D. A. and Dandekar, T. (2000), 'A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks', *Nature Biotechnology* **18**, 326–332.

Schuster, S. and Hilgetag, C. (1994), 'On elementary flux modes in biochemical reaction systems at steady state', *Journal of Biological Systems* **2**, 165–182.

Schuster, S., Hilgetag, C., Woods, J. H. and Fell, D. A. (2002a), 'Reaction routes in biochemical reaction systems: algebraic properties, validated calculation procedure and example from nucleotide metabolism', *Journal of Mathematical Biology* **45**, 153–181.

Schuster, S., Klamt, S., Weckwerth, W., Moldenhauer, F. and Pfeiffer, T. (2001), 'Use of network analysis of metabolic systems in bioengineering', *Bioprocess and Biosystems Engineering* **24**, 363–372.

Schuster, S., Pfeiffer, T., Moldenhauer, F., Koch, I. and Dandekar, T. (2002b), 'Exploring the pathway structure of metabolism: decomposition into subnetworks and application to *Mycoplasma pneumoniae*', *Bioinformatics* **18**, 351–361.

Schuster, S. and Schuster, R. (1991), 'Detecting strictly detailed balanced subnetworks in open chemical reaction networks', *Journal of Mathematical Chemistry* **6**, 17–40.

Schwarz, R., Musch, P., von Kamp, A., Engels, B., Schirmer, H., Schuster, S. and Dandekar, T. (2005), 'YANA - a software tool for analyzing flux modes, gene-expression and enzyme activities', *BMC Bioinformatics* **6**, 135.

Seshasayee, A. S. N., Fraser, G. M., Babu, M. M. and Luscombe, N. M. (2009), 'Principles of transcriptional regulation and evolution of the metabolic system in *E. coli*', *Genome Research* **19**, 79–91.

Shlomi, T., Eisenberg, Y., Sharan, R. and Ruppin, E. (2007), 'A genome-scale computational study of the interplay between transcriptional regulation and metabolism', *Molecular Systems Biology* **3**, 101.

Song, H.-S. and Ramkrishna, D. (2009), 'Reduction of a set of elementary modes using yield analysis', *Biotechnology and Bioengineering* **102**, 554–568.

Spivey, A. (2004), 'Systems biology: the big picture', *Environmental Health Perspectives* **112**, A938–A943.

Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S. and Gilles, E. D. (2002), 'Metabolic network structure determines key aspects of functionality and regulation', *Nature* **420**, 190–193.

Stelzl, U. and Wanker, E. E. (2006), 'The value of high quality protein-protein interaction networks for systems biology', *Current Opinion in Chemical Biology* **10**, 551–558.

Steuer, R., Gross, T., Selbig, J. and Blasius, B. (2006), 'Structural kinetic modeling of metabolic networks', *Proceedings of the National Academy of Sciences of the United States of America* **103**, 11868–73.

Suthers, P. F., Chang, Y. J. and Maranas, C. D. (2010), 'Improved computational performance of MFA using elementary metabolite units and flux coupling', *Metabolic Engineering* **12**, 123–128.

Suthers, P. F., Dasika, M. S., Kumar, V. S., Denisov, G., Glass, J. I. and Maranas, C. D. (2009), 'A genome-scale metabolic reconstruction of *Mycoplasma genitalium*, *i*ps189', *PLoS Computational Biology* **5**, e1000285.

Terzer, M., Maynard, N. D., Covert, M. W. and Stelling, J. (2009), 'Genome-scale metabolic networks', *WIREs Systems Biology and Medicine* **1**, 285–297.

Terzer, M. and Stelling, J. (2008), 'Large-scale computation of elementary flux modes with bit pattern trees', *Bioinformatics* **24**, 2229–2235.

Terzer, M. and Stelling, J. (2010), Parallel extreme ray and pathway computation, *in* 'Proceedings of the 8th International Conference on Parallel Processing and Applied Mathematics (PPAM 2009)', Vol. 6068 of *Lecture Notes in Computer Science*, Wroclaw, Poland, pp. 300–309.

Teusink, B., Wiersma, A., Jacobs, L., Notebaart, R. A. and Smid, E. J. (2009), 'Understanding the adaptive growth strategy of *Lactobacillus plantarum* by in silico optimisation', *PLoS Computational Biology* **5**, e1000410.

Thiele, I. and Palsson, B. O. (2010), 'A protocol for generating a high-quality genome-scale metabolic reconstruction', *Nature Protocols* **5**, 93–121.

Thiele, I., Vo, T. D., Price, N. D. and Palsson, B. O. (2005), 'Expanded metabolic reconstruction of *Helicobacter pylori* (iIT341 GSM/GPR): an in silico genome-scale characterization of single- and double-deletion mutants', *Journal of Bacteriology* **187**, 5818–5830.

Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, S., Yugi, K., Venter, J. and Hutchison, C. (1997), E-cell: software environment for whole cell simulation, *in* 'Proceedings of the Eighth Workshop on Genome Informatics (GIW 1997)', Vol. 8 of *Genome Informatics*, Tokyo, Japan, pp. 147–155.

Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T. S., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, S., Yugi, K., Venter, J. and Hutchison, C. A.

(1999), 'E-CELL: software environment for whole-cell simulation', *Bioinformatics* **15**, 72–84.

Trinh, C. T., Wlaschin, A. and Srienc, F. (2009), 'Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism', *Applied Microbiology and Biotechnology* **81**, 813–826.

Unrean, P., Trinh, C. T. and Srienc, F. (2010), 'Rational design and construction of an efficient *E. coli* for production of diapolycopendioic acid.', *Metabolic Engineering* **12**, 112–122.

Urbanczik, R. (2006), 'SNA–a toolbox for the stoichiometric analysis of metabolic networks.', *BMC Bioinformatics* **7**, 129.

Urbanczik, R. (2007), 'Enumerating constrained elementary flux vectors of metabolic networks', *IET Systems Biology* **1**(5), 274–279.

Urbanczik, R. and Wagner, C. (2005), 'Functional stoichiometric analysis of metabolic networks', *Bioinformatics* **21**, 4176–4180.

Varma, A. and Palsson, B. O. (1993), 'Metabolic capabilities of *Escherichia coli*: I. synthesis of biosynthetic precursors and cofactors', *Journal of Theoretical Biology* **165**, 477–502.

Vazquez, A., Beg, Q. K., Demenezes, M. a., Ernst, J., Bar-Joseph, Z., Barabási, A.-L., Boros, L. G. and Oltvai, Z. N. (2008), 'Impact of the solvent capacity constraint on *E. coli* metabolism', *BMC Systems Biology* **2**, 7.

Veeramani, B. and Bader, J. S. (2009), 'Metabolic flux correlations, genetic interactions, and disease', *Journal of Computational Biology* **16**, 291–302.

Veeramani, B. and Bader, J. S. (2010), 'Predicting functional associations from metabolism using bi-partite network algorithms.', *BMC Systems Biology* **4**, 95.

Verwoerd, W. (2007), Identifying coherent subnetworks in genome scale metabolic networks, *in* 'MODSIM07', Christchurch, New Zealand, pp. 2013–2019.

Verwoerd, W. S. (2010), Dissecting Metabolic Networks into Functional Subnets, *in* G. Fung, ed., 'A Practical Guide to Bioinformatics Analysis', iConcept Press, pp. 79–98.

Verwoerd, W. S. (2011), 'A new computational method to split large biochemical networks into coherent subnets', *BMC Systems Biology* **5**, 25.

Vo, T. D., Greenberg, H. J. and Palsson, B. O. (2004), 'Reconstruction and functional characterization of the human mitochondrial metabolic network based on proteomic and biochemical data', *Journal of Biological Chemistry* **279**, 39532–39540.

von Kamp, A. and Schuster, S. (2006), 'Metatool 5.0: fast and flexible elementary modes analysis', *Bioinformatics* **22**, 1930–1931.

Voss, K., Heiner, M. and Koch, I. (2003), 'Steady state analysis of metabolic pathways using Petri nets', *In Silico Biology* **3**, 367–387.

Wagner, C. (2004), 'Nullspace Approach to Determine the Elementary Modes of Chemical Reaction Systems', *Journal of Physical Chemistry B* **108**, 2425–2431.

Wagner, C. and Urbanczik, R. (2005), 'The geometry of the flux cone of a metabolic network', *Biophysical Journal* **89**, 3837–3845.

Wiback, S. J. and Palsson, B. O. (2002), 'Extreme pathway analysis of human red blood cell metabolism', *Biophysical Journal* **83**, 808–818.

Wiener, N. (1948), *Cybernetics: or control and communication in the animal and the machine*, MIT Press, Cambridge.

Wolkenhauer, O. (2001), 'Systems biology: the reincarnation of systems theory applied in biology?', *Briefings in Bioinformatics* **2**, 258–270.

Xi, Y., Chen, Y.-P. P., Qian, C. and Wang, F. (2011), 'Comparative study of computational methods to detect the correlated reaction sets in biochemical networks', *Briefings in Bioinformatics* **12**, 132–150.

Yeung, M., Thiele, I. and Palsson, B. O. (2007), 'Estimation of the number of extreme pathways for metabolic networks', *BMC Bioinformatics* **8**, 363.

Yus, E., Maier, T., Michalodimitrakis, K., van Noort, V., Yamada, T., Chen, W.-H., Wodke, J. A. H., Güell, M., Martínez, S., Bourgeois, R., Kühner, S., Raineri, E., Letunic, I., Kalinina, O. V., Rode, M., Herrmann, R., Gutiérrez-Gallego, R., Russell, R. B., Gavin, A.-C., Bork, P. and Serrano, L. (2009), 'Impact of genome reduction on bacterial metabolism and its regulation', *Science* **326**, 1263–8.

Zamboni, N. (2011), '^{13}C metabolic flux analysis in complex systems', *Current Opinion in Biotechnology* **22**, 103–108.

Zhang, Y., Thiele, I., Weekes, D., Li, Z., Jaroszewski, L., Ginalski, K., Deacon, A. M., Wooley, J., Lesley, S. A., Wilson, I. A., Palsson, B. O., Osterman, A. and Godzik, A. (2009), 'Three-dimensional structural view of the central metabolic network of *Thermotoga maritima*', *Science* **325**, 1544–1549.

Zhao, Q. and Kurata, H. (2009), 'Genetic modification of flux for flux prediction of mutants', *Bioinformatics* **25**, 1702–1708.

Appendix A

Nomenclature

Abbreviations

CBM : constraint-based modeling

c.i. : confidence interval

Co-Set : correlated reaction set

ECM : elementary conversion mode

EM : elementary (flux) mode

EFP : elementary flux pattern

EFP-FCA : flux coupling analysis based on elementary flux patterns

EXPA : extreme pathway

FBA : flux balance analysis

FCA : flux coupling analysis

FCF : flux coupling finder (algorithm)

FFCA : feasibility-based flux coupling analysis

FVA : flux variability analysis

GEC : gene expression correlation

LP : linear program (or programming)

MCS : minimal cut set

M.E. : mutually exclusive

MILP : mixed integer linear program (or programming)

MMB : minimal metabolic behavior

MMB-FCA : flux coupling analysis based on minimal metabolic behaviors

ODE : ordinary differential equation

PEM : projected elementary mode

ProCEM : projected cone elementary mode

RBC : red blood cell

RT : reversibility-type

S.C. : sometimes coupled

woS : without splitting reversible reactions

Mathematical Notations

$\mathbf{0}$: zero vector

$\dot{c}$: rate of changes in c

$\mathbf{e}_i$: the i-th unit vector

$[n]$: the set $\{1, \ldots, n\}$

$M_{P;Q}$: submatrix of M induced by rows in P and columns in Q

$dim(X)$: dimension of X

$supp(v)$: the set of nonzero elements of v

$\mathcal{P}_{\mathbb{A}}(Q)$: projection of Q onto $\mathbb{A}$

$\cong$: equivalent (vectors)

$\longrightarrow$ **(or** $\longleftarrow$**)** : directionally coupled

$\longleftrightarrow$: partially coupled

$\Longleftrightarrow$: Fully coupled

$\overset{S.C.}{\longleftrightarrow}$: sometimes coupled

$\overset{M.E.}{\longleftrightarrow}$: mutually exclusive

$\overset{Un}{\longleftrightarrow}$: uncoupled

$\perp$: infeasible (linear program)

Appendix B

Curriculum Vitae

For reasons of data protection, the Curriculum Vitae is not included in the online version.