New Concepts and Tools in Constraint-based Analysis of Metabolic Networks

Abdelhalim Larhlimi

Freie Universität Berlin
Berlin, Germany

New Concepts and Tools in Constraint-based Analysis of Metabolic Networks

New Concepts and Tools in Constraint-based Analysis of Metabolic Networks

Abdelhalim Larhlimi

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

July 2008

Freie Universität Berlin
Berlin, Germany

Datum des Kolloquiums: ***3 Dezember 2008***

Gutachter:
Prof. Dr. Alexander Bockmayr*, Freie Universität Berlin, Deutschland.*
Prof. Dr. Stefan Schuster*, Friedrich Schiller University Jena, Deutschland.*

TO MY WIFE MOUNIA, MY KID BILAL ALHABIB
AND MY PARENTS

CONTENTS

Acknowledgements

I thank Almighty God, Allah, Who gave me the confidence to pursue this work to its conclusion. To Him I extend my heartfelt thanks.

I wish to express my sincere and deepest heartfelt gratitude to my advisor Prof. Dr. Alexander Bockmayr, chair of Mathematics in Life Sciences, Freie Universität Berlin, for his unselfish commitment of time, trusted conscientious advice, comprehensive guidance, constant support and encouragement. I am grateful for his help in every step of the process, from my first day as a PhD student to the proofreading of the thesis.

I want to thank Prof. Dr. Stefan Schuster, head of the chair of Bioinformatics, Friedrich Schiller University Jena, for his kind acceptance to act as a reviewer for this thesis.

I would like to thank Prof. Dr. Knut Reinert, chair of Algorithms in Bioinformatics, Freie Universität Berlin, Prof. Dr. Martin Vingron, director of CMB Department at the MPI for Molecular Genetics, and Dr. Heike Siebert, research associate in the DFG Research Center Matheon, for their kind acceptance to be my committee members.

I want to express my sincere thanks to my colleagues in the Department of Mathematics and Computer Science of the Freie Universität Berlin, specially our "Mathematics in Life Sciences group". Thank you Amir, Corinna, Gunnar, Heike and Theo for your friendly talking, nice cooperation, good working atmosphere, and valuable discussions. A special gratitude is due to Katja for her kind and accurate support.

Special acknowledgements go to Dr. Damien Eveillard, Nantes Laboratory for Computer Science, France, for numerous scientific discussions and suggestions. We have had many interesting and illuminating discussions touching all aspects of life.

I would like to thank all my friends for their support and for the nice time we spent together. Among all of them I am particularly thankful to Abderrahim and Samer for countless hours spent discussing fruitful ideas over cups of coffee. A special gratitude is due to my friend Amin with whom I enjoyed talking about everything but work.

The research whose results are described in the thesis has been carried out at the Department of Mathematics and Computer Science of the Freie Universität Berlin. I thank the department for a wonderful working environment. The financial support from the Freie Universität Berlin and the DFG Research Center Matheon is greatly appreciated.

Last, but by far not least, my gratitude is addressed to my wife for her love, support and for putting up with my late hours and my spoiled weekends. To my pretty kid Bilal Alhabib for his lovely smile that can relieve any kinds of tiredness. I am also grateful for my family, specially my loving parents, without whose love, support and encouragement this work would not have been possible.

Abstract

Constraint-based approaches have proved successful in analyzing complex metabolic networks. They restrict the range of all possible behaviors that a metabolic system can display under governing constraints. The set of all possible flux distributions over a metabolic network at steady state defines a polyhedral cone, the *steady-state flux cone.* This cone can be analyzed using an *inner description*, based on sets of generating vectors such as *elementary modes* or *extreme pathways.* We present a new constraint-based approach to metabolic network analysis, characterizing a metabolic network by its *minimal metabolic behaviors* and the *reversible metabolic space.* Our method uses an *outer description* of the flux cone, based on sets of non-negativity constraints. The resulting description is minimal and unique. We then study the relationship between inner and outer descriptions of the cone. We give a generic procedure to show how inner descriptions can be computed from the outer one. We use this procedure to explain why the size of the inner descriptions may be several orders of magnitude larger than that of the outer description.

Our approach suggests a refined classification of reactions according to their reversibility type (*irreversible, pseudo-irreversible, and fully reversible*). Using these concepts, we improve an existing algorithm for identifying *blocked* and *coupled* reactions and devise a new algorithm for *flux coupling analysis.* We extend this analysis by introducing *minimal direction cuts (MDCs)* which prevent a target reaction from being performed in an undesired direction. We develop an algorithm which allows not only for computing MDCs in a metabolic network, but also for a direct calculation of *minimal cut sets (MCSs).* Based on our refined classification of reactions, we also provide a constraint-based approach for analyzing the changes in the overall capabilities of a metabolic network following a gene deletion.

Flux coupling and gene deletion analysis help for identifying important reactions in metabolic networks. Alternatively, the essentiality of reactions can be assessed using *control-effective flux (CEF) analysis*, which is based on elementary modes. We compare CEF analysis with the use of a minimal generating set of the flux cone to elucidate crucial reactions.

Zusammenfassung

In der Analyse metabolischer Netzwerke haben constraintbasierte Ansätze erfolgreiche Anwendung gefunden. Hierbei wird der Bereich des möglichen Verhaltens eines metabolischen Systems durch zusätzliche Anforderungen an das System eingeschränkt. Die resultierende Menge aller Flussverteilungen eines metabolischen Netzwerks im stationären Zustand hat die Gestalt eines polyedrischen Kegels, welcher *Flusskegel* genannt wird. Eine *innere Beschreibung* dieses Kegels basierend auf Mengen erzeugender Vektoren, wie etwa *Elementarmodi* oder *Extremalpfade*, ermöglicht eine effiziente Analyse. Wir haben einen neuen constraintbasierten Ansatz zur Analyse metabolischer Netzwerke entwickelt, in dem das System durch *minimale metabolische Verhaltensmuster* und den *reversiblen metabolischen Raum* charakterisiert wird. In unserer Methode kommt eine *äußere Beschreibung* des Flusskegels zur Anwendung, die wir durch Ausnutzung von Nicht-Negativitäts-Bedingungen erhalten. Diese Beschreibung ist minimal und eindeutig. Wir untersuchen die Beziehung zwischen innerer und äußerer Beschreibung des Kegels und stellen ein allgemeines Verfahren zur Herleitung der inneren aus der äußeren Beschreibung vor. Dieses Verfahren verdeutlicht, warum die äußere im Vergleich zur inneren Beschreibung eine meist sehr viel kompaktere, sogar bis zu mehreren Größenordungen kleinere Darstellung liefert.

In unserem Ansatz verwenden wir eine verfeinerte Klassifikation von Reaktionen des metabolischen Netzwerks entsprechend ihres Reversibilitäts-Typus (irreversibel, pseudo-irreversibel und vollständig reversibel). Diese Einteilung ermöglicht uns eine deutliche Verbesserung existierender Algorithmen zur Bestimmung von *blockierten* und *gekoppelten* Reaktionen und die Formulierung eines neuen, effizienten Algorithmus für die *Flusskopplungsanalyse*. Die von uns eingeführten *minimalen gerichteten Schnitte (MDCs)*, die die Ausführung einer Zielreaktion in eine ungewünschte Richtung verhindern, erweitern die klassische Flusskopplungsanalyse. Ein von uns entwickelter Algorithmus ermöglicht nicht nur die Berechnung von MDCs in einem metabolischen Netzwerk, sondern auch die direkte Ermittlung *minimaler Schnittmengen (MCSs)*. Basierend auf unserer verfeinerten Klassifizierung von Reaktionen stellen wir schließlich einen constraintbasierten Ansatz zur Analyse der durch Gen-Ausfall ausgelösten Beeinträchtigungen globaler Fähigkeiten eines metabolischen Netzwerks vor.

Flusskopplungs- und Gen-Ausfall-Analyse helfen bei der Identifikation essentieller Reaktionen im metabolischen System. Altenativ kann die Bedeutung von Reaktionen für die Netzwerkfunktionen mittels auf Elementarmodi basierender *control-effective Fluss-Analyse (CEF)* bewertet werden. Wir vergleichen CEF-Analyse mit der Verwendung eines minimalen Erzeugendensystems für die Bestimmung von Schlüsselreaktionen.

CHAPTER 1

Introduction

The unprecedented progress in molecular biology fuels the interest in system-level understanding of living systems, complementing the reductionist approach that prevails in molecular biology during the last century. System-level analysis has been a recurrent topic in biology since the days of Norbert Wiener [45], while defining the field of cybernetics [122]. The ultimate aim of such an integrative approach is to link the behaviour of living matter to system's structure and dynamics, helping for the comprehension of biological systems and for engineering and designing strains more appropriate for metabolite production purposes [10]. The behavior of a living system depends on its ability to import materials from the environment and convert them to the needed molecules. These conversions are carried out by metabolism.

1.1 Metabolic Network Modeling

1.1.1 Basic Concepts

Broadly speaking, a *metabolic reaction* refers to a chemical transformation that occurs in living organisms, allowing them to feed, grow and reproduce [31]. Metabolic reactions sustain several biological functions including the degradation of chemical substances for energy production or the assembly of cellular components. A *catabolic* reaction breaks down complex molecules into smaller components and yields energy. For instance, the breakdown of proteins, carbohydrates, and lipids in digestion are carried out by catabolic reactions. An *anabolic* reaction, on the other hand, uses energy to build cellular constituents such as proteins and nucleic acids. Fig. 1.1 is a schematic representation of anabolic and catabolic reactions.

The substances that will react in a metabolic reaction are called *substrates.* As a result of the reaction, the substrates will be converted into different molecules named *products. Metabolites* are the substances that are involved in metabolic reactions. They carry out most of the metabolic functions in a living system. The *stoichiometric coefficient* of a metabolite in a reaction is the amount of that metabolite involved in the reaction in terms of molecules (or moles of molecules). For instance, the *decomposition of hydrogen peroxide* converts two molecules of hydrogen peroxide (H_2O_2) into two molecules of water (H_2O) and one molecule of oxygen (O_2). The stoichiometric coefficient of water in this reaction is twice that of oxygen.

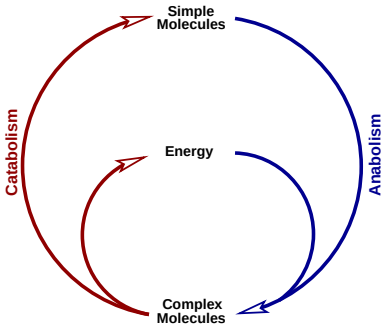


Figure 1.1: An overview of catabolic and anabolic reactions

A *reaction equation* is a symbolic representation which defines the stoichiometric conversion of substrates into products for some reaction. In such an equation, substrates are given on the left hand side, products are written on the right hand side, and the numbers next to the metabolites are the absolute values of the stoichiometric coefficients. For example, the reaction equation of the decomposition of hydrogen peroxide is

$$2\ H_2O_2\ \rightarrow\ 2\ H_2O\ +\ O_2.$$

A reaction is said to be operating in the *forward* (resp. *backward*) direction if that reaction converts metabolites given on the left (resp. right) hand side of its equation into ones given on the right (resp. left) hand side of its equation. Reactions that can proceed in either direction are called *reversible*, and the remaining reactions are named *irreversible*.

The *(net) flux* of a metabolic reaction is the rate of consumption of any substrate divided by the corresponding stoichiometric coefficient. This is equal to the rate of formation of any product divided by the corresponding stoichiometric coefficient. A reaction with a high flux operates at a faster speed than a reaction with a low flux. In addition, a flux is positive (resp. negative) if the forward (resp. backward) reaction is faster than the backward (resp. forward) reaction. While fluxes through reversible reactions may be negative, the convention is to consider that fluxes through irreversible ones are always non-negative.

For the rest of the thesis, we shall use the simpler term reaction for metabolic reaction.

1.1.2 Metabolic Networks

A metabolite is often transformed to another by a series of reactions, called a *metabolic pathway*. The products of one reaction serve as substrates for other reactions. A *metabolic network* refers then to an interconnected set of reactions that carry out step by step transformation of the initial metabolites to convert them into other products. Each step is catalyzed by special proteins called *enzymes*. These catalysts speed up reactions without being used up and their concentrations affect the rates of reactions. A living organism can then regulate the fluxes through its reactions by producing differ-

ent amounts of enzymes. This control allows biological systems to adapt and respond to their environments.

The significant advances in molecular biology have paved the way for a fast reconstruction of metabolic networks for an increasing number of microorganisms, including *S. cereviseae, H. influenzae, E. coli,* and *H. pylori* [32; 83; 86; 89]. Several genome-scale (reconstructed) networks are available from metabolic databases such as *KEGG [41], EcoCyc [43], MetaCyc [18],* and *BioCyc [55].* In general, a genome-scale metabolic network consists of hundred to thousands of metabolites linked by an even larger set of highly interconnected reactions. Knowledge and insight that may be gained from an isolated study of a single component in a metabolic network are then limited. Accordingly, it is becoming generally accepted that one has to move from a component-based view to a systems-level understanding of metabolic networks [36; 72]. This shift in paradigm allows for reconstructing the integrated behavior of living organisms from the underlying metabolic components. Such an integrative approach helps not only to elucidate the intrinsic biological properties that emerge from the whole metabolic system, but also to predict how these properties would change in response to alterations in environment or system components. These properties are usually qualified as either *qualitative* (e.g., how robust is the network? how many alternative ways are there to produce a particular metabolite?) or *quantitative* (e.g., what is the rate of glucose uptake or the concentration of oxygen?) [4]. Moreover, we often divide information required to predict or calculate such properties into two categories: *structural* and *kinetic.* The former describe the set of components that are involved in the network, and interactions or connections among these components. The latter contain differential equations that describe reactions, changes in metabolite concentrations and numerical values for the parameters used in those equations. In general, the choice of a particular method depends not only on whether the properties of interest are quantitative or qualitative, but also on the type of the available biological data (kinetic or structural). Despite the advances in experimental high-throughput techniques, quantitative modeling is often still hampered by incomplete knowledge of kinetic information.

In the following we give a short introduction about stoichiometry and reaction reversibility.

Stoichiometic Matrix

When modeling a metabolic system, we often distinguish between *internal* and *external* metabolites [38]. An internal metabolite should not accumulate or decrease in time, and so its rate of formation is equal to its rate of consumption. In contrast, external metabolites could be buffered as it can be assumed for water or many species from the environment. In general, classifying a metabolite as external or internal depends on the purpose of the model. It should be noted that this classification has also an impact on the algorithmic complexity of analyzing the network [24].

It is also common to distinguish between *internal* and *boundary* reactions in a metabolic network [98]. An internal reaction has the property that its substrates and

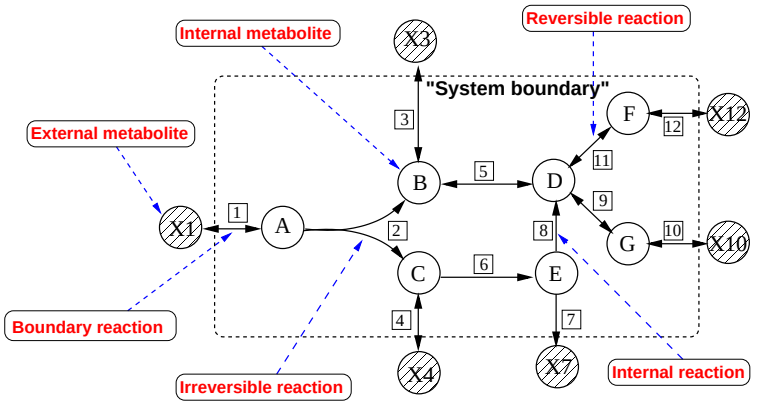


Figure 1.2: Network example ILLUSNET. Metabolites are depicted as nodes while reactions are depicted as arrows. Reversible reactions are indicated by double arrowheads.

Internal reactions		Boundary reactions	
Reaction 2 :	$1\,A \rightarrow 2\,B + 1\,C$	Reaction 1 :	$. \leftrightarrow 1\,A$
Reaction 5 :	$2\,B \leftrightarrow 1\,D$	Reaction 3 :	$. \leftrightarrow 1\,B$
Reaction 6 :	$1\,C \rightarrow 1\,E$	Reaction 4 :	$. \leftrightarrow 1\,C$
Reaction 8 :	$3\,E \rightarrow 1\,D$	Reaction 7 :	$1\,E \rightarrow .$
Reaction 9 :	$1\,D \leftrightarrow 1\,G$	Reaction 10 :	$1\,G \rightarrow .$
Reaction 11 :	$1\,D \rightarrow 1\,F$	Reaction 12 :	$1\,F \rightarrow .$

Table 1.1: List of reaction equations for the metabolic network depicted in Fig. 1.2. Internal metabolites are $A, \ldots, G$ and the external ones are replaced by dots.

products each contain at least one internal metabolite. On the other hand, all substrates consumed or all products formed by a boundary reaction are external. Accordingly, a boundary reaction, which is also called an *exchange reaction*, allows the transport of materials across the system boundary, and somehow is a connection between the metabolic system and its environment.

Example 1.1. *For illustration throughout this chapter, we consider the hypothetical network ILLUSNET depicted in Fig. 1.2. It consists of seven internal metabolites* $(A, \ldots, G)$*, and twelve reactions* $(1, \ldots, 12)$*, whereof four reactions are irreversible. Reversible reactions are indicated by double arrowheads. External metabolites are not represented in this network. Tab. 1.1 shows the list of reaction equations for this metabolic network.*

The stoichiometric matrix of a metabolic network is defined similar to the adjacency matrix of a directed graph. It enables to represent the structure of a metabolic network in terms of relationships between internal metabolites and reactions. For instance, given a reaction j that converts a molecules of metabolite A and c molecules of metabolite C into e molecules of metabolite E, the column in the stoichiometric matrix that corresponds to reaction j is given by

$$a\ A + c\ C \xrightarrow{j} e\ E$$

$$\begin{array}{c} \\ \mathbf{A} \\ B \\ \mathbf{C} \\ D \\ \mathbf{E} \end{array} \begin{array}{c} j \\ \begin{pmatrix} . & . & . & . & \text{-}a & . & . & . & . \\ . & . & . & . & \mathbf{0} & . & . & . & . \\ . & . & . & . & \text{-}c & . & . & . & . \\ . & . & . & . & \mathbf{0} & . & . & . & . \\ . & . & . & . & \mathbf{e} & . & . & . & . \end{pmatrix} \end{array}$$

The rows of the stoichiometric matrix correspond to the internal metabolites and its columns correspond to the reactions making up the metabolic network.

Definition 1.2 (Stoichiometric matrix)**.** *Given a metabolic network, let m be the number of internal metabolites and let n be the number of all reactions in the network. The corresponding stoichiometric matrix is an $m \times n$ matrix S, such that for each internal metabolite i and each reaction j*

$$S_{ij} = \begin{cases} +\alpha, & \text{if } j \text{ produces } \alpha \text{ molecules of } i \text{ in its forward direction.} \\ -\alpha, & \text{if } j \text{ consumes } \alpha \text{ molecules of } i \text{ in its forward direction.} \\ 0, & \text{if the reaction j neither produces nor consumes i.} \end{cases}$$

Example 1.3. *The stoichiometric matrix corresponding to the network ILLUSNET depicted in Fig. 1.2 reads*

$$S = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & -1 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & -3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \end{pmatrix}.$$

For instance, the second column of this stoichiometric matrix corresponds to reaction 2. This reaction consumes one molecule of metabolite A and produces two molecules of metabolite B and one molecule of metabolite C. In addition, we can deduce from the stoichiometric matrix that reactions $1, 3, 4, 7, 10, 12$ are boundary reactions, while reactions $2, 5, 6, 8, 9, 11$ are internal.

It should be noted that only internal metabolites are represented in the stoichiometric matrix, that is, each row in this matrix corresponds to an internal metabolite. In contrast, the columns in the stoichiometric matrix correspond to all reactions in the network. A distinction between the columns corresponding to internal reactions and those corresponding to boundary reactions is possible. Indeed, the former contain entries of opposite signs, while the entries in the latter have the same sign. Distinguishing between these two categories of columns allows for identifying *internal cycles*, which are metabolic pathways involving only internal reactions.

It is also worth mentioning that many stoichiometric matrices may correspond to the same metabolic network. For instance, let $S \in \mathbb{R}^{m \times n}$ be a stoichiometric matrix

and let $P \in \mathbb{R}^{m \times m}$ and $Q \in \mathbb{R}^{n \times n}$ be permutation matrices. Let $D \in \mathbb{R}^{n \times n}$ be a diagonal matrix such that $D_{ii} = -1$ for some reversible reaction i and $D_{jj} = 1$ for all $j \neq i$. Thus, S, PS, SQ and SD are stoichiometric matrices for the same metabolic network. The problem of deciding whether two stoichiometric matrices describe the same metabolic network is non-trivial to solve [52]. Dealing successfully with this problem is important for instance to find out whether two metabolic (sub-) networks share the same structure.

Reaction Reversibility

Thermodynamically, all metabolic reactions are reversible [22]. In fact, a reaction can proceed in either direction depending on its Gibbs free energy difference. A positive flux through a reaction implies a corresponding negative change of the Gibb's free reaction energy and vice versa [6; 53], i.e.,

$$sgn(v_i) = -sgn(\Delta G_i) \tag{1.1}$$

where 'sgn' is the sign function and v_i (resp. ΔG_i) is the flux (resp. Gibbs free energy difference) of reaction i. Accordingly, a reaction operates in the forward, i.e., $v_i > 0$ (resp. backward, i.e., $v_i < 0$) direction if its Gibbs free energy difference is negative (resp. positive). Furthermore, a reaction reverses its direction if the corresponding Gibbs free energy difference changes its sign. Reactions that have ΔG values closer to zero may involve free energy ranges that span both negative and positive values. These reactions are very likely reversible. However, under physiological conditions, a so-called irreversible reaction can only proceed in one direction because the corresponding Gibbs free energy difference is far from zero and its sign is always constant. This leads to classify reactions as irreversible or reversible depending on their Gibbs free energy difference values. Nevertheless, this strategy requires the computation of the Gibbs free energy differences for all reactions. Significant efforts have been made in doing this calculation which, however, is still a hard computational task. A recent approach has been proposed to assign reaction directions in metabolic networks on the basis of network topology considerations and thermodynamics-based heuristic rules [57]. Nonetheless, the reversibility of metabolic reactions is still largely a matter of convention or perspective. For instance, some reactions are irreversible because they are assumed to be so or this is in keeping with the biological function the system has to accomplish, e.g., ATP production by glycolysis [95]. In contrast, reactions shared by catabolic and anabolic pathways are often considered reversible [98]. This is justified by the fact that these reactions are both anabolic and catabolic. In the context of this work, we assume that irreversible reactions are defined as follows.

Definition 1.4 (Irreversible reaction)**.** *A reaction i is irreversible if and only if its flux is always non-negative.*

For the purpose of this thesis, a formal description of the structure and stoichiometry of a metabolic network will be given as follows:

- m: number of internal metabolites,
- n: number of reactions,
- S: stoichiometric matrix ($S \in \mathbb{R}^{m \times n}$),
- Irr: set of irreversible reactions ($Irr \subseteq \{1, \ldots, n\}$),
- Rev: set of reversible reactions ($Rev = \{1, \ldots, n\} \setminus Irr$).

1.2 Aims and Organization of the Thesis

The huge amount of genomic, transcriptomic and related data has allowed for a fast reconstruction of an increasing number of genome-scale metabolic networks. The latter have been recognized to be highly complex. In this respect, the need for mathematical and computational methods that focus on the systemic properties of metabolic networks is increasingly pressing. Various approaches have been developed, ranging from *metabolic control, stochastic, cybernetic, kinetic* to *constraint-based analysis* [29; 35; 38; 77; 81]. In the absence of detailed kinetic information, constraint-based modeling has recently attracted ample interest due to its ability to analyze genome-scale metabolic networks while using very few information. This approach is based on the application of a series of constraints that govern the operation of a metabolic network at steady state, including the stoichiometric and thermodynamic constraints. The latter limit the range of allowable behaviors of the metabolic network, each basically representing a possible metabolic phenotype. Applying these constraints leads to the formulation of the solution space, called the *steady-state flux cone* [20].

Several approaches have been proposed to describe the steady-state flux cone using an *inner description*, which is based on sets of generating vectors [20; 90; 98]. The number of these generators may be very large even for small networks and their calculation may need many resources in terms of time and memory. This limits the practical applicability of these methods. This thesis mainly aims to define a constraint-based approach that shrinks the size of the flux cone description to a more manageable level. A key idea is to use an *outer description* of the steady-state flux cone, based on sets of non-negativity constraints.

Chap. 2 is dedicated to a general introduction to polyhedral theory. The concepts introduced in this chapter constitute the basics for the analysis of the steady-state flux cone of metabolic networks developed in this work. The concept of steady-state flux cone is formally defined in Chap. 3. This notion proceeds directly from the application of the steady-state conditions governing the operation of a metabolic network. We conclude this chapter by a detailed discussion of the main approaches for exploring the flux cone.

In Chap. 4, we propose a new mathematical approach to metabolic network analysis, characterizing a metabolic network by its *minimal metabolic behaviors* and the *reversible metabolic space*. Our method uses an outer description of the steady-state flux cone, based on sets of irreversible reactions. This is different from existing approaches,

such as *elementary modes* [99] or *extreme pathways* [90], which use an inner description, based on sets of generating vectors. The resulting description of the flux cone is minimal, unique, and satisfies a simplicity condition similar to the one that holds for elementary modes. Our approach suggests a refined classification of reactions according to their reversibility type (*irreversible*, *pseudo-irreversible*, and *fully reversible*). While the irreversible and pseudo-irreversible reactions completely characterize minimal metabolic behaviors, the fully reversible reactions define the reversible metabolic space, which may contain useful biological information. This information is no longer explicit if we replace each reversible reaction with two irreversible ones.

Chap. 5 is devoted to the study of the relationship between inner and outer descriptions of the flux cone. By distinguishing pseudo-irreversible and fully reversible reactions, we analyze the impact of reconfiguring the metabolic network in terms of the size of the description of the reconfigured flux cone as well as the reversibility type of reactions. This leads to a generic procedure for computing inner descriptions from the outer one. This procedure makes clear why the size of the inner descriptions may be several orders of magnitude larger than that of the outer description.

In Chap. 6, we show that the reversibility type also provides a key to elucidate reaction dependencies. Indeed, coupling relationships can only hold between reactions of a certain reversibility type. In particular, (pseudo-) irreversible reactions cannot be coupled with fully reversible reactions, and all reactions in an enzyme subset [78] must have the same reversibility type. The mathematical results that have been obtained not only allow for improving an existing algorithm, but also lead to a new algorithm for identifying blocked and coupled reactions in a metabolic network.

In Chap. 7, we introduce the concept of *minimal direction cuts* that allow preventing a target reaction from being performed in an undesired direction. If the target reaction is irreversible, MDCs are not different from *minimal cut sets (MCSs)* [46]. However, if this reaction is reversible, MDCs allow it to operate in the desired direction, while MCSs make it inactive in both directions. In both cases, each MCS can be seen as the union of two MDCs. Therefore, all the useful applications of MCSs can also be done with MDCs. However, the computation of MDCs does not require that of elementary modes. MDCs may be determined based on the Farkas lemma for equality and inequality constraints. The mathematical results of this chapter lead to an iterative algorithm for computing MDCs in a metabolic network. Since MCSs can be obtained from MDCs, this algorithm also allows for a direct computation of MCSs, in the sense that we need not calculate beforehand the elementary modes. Finally, our algorithm gives the possibility of introducing additional constraints that may reduce the search space. This makes our approach applicable even for genome-scale metabolic networks.

In Chap. 8, we analyze the changes in the overall capabilities of a metabolic network caused by gene deletion. In particular, we show how to obtain, in a constraint-based approach, a description of the altered steady-state flux cone. The analysis is again based on our refined classification of reactions.

The importance of single reactions for the overall metabolic network performance can by assessed using knockout mutations. An important reaction can be identified using flux coupling analysis or gene deletion analysis discussed in Chap. 6 and Chap. 8,

respectively. Alternatively, the essentiality of some reaction could correlate with how this reaction participates in flexible and efficient operations of the metabolic network. In Chap. 9, we discuss *control-effective flux (CEF) analysis*, which has proved promising in assessing the importance of reactions. We formally explain why elementary modes are useful for CEF analysis. We also consider the use of a minimal generating set of the flux cone in such an analysis.

CHAPTER

2 Mathematical Preliminaries

In this chapter we give a short overview of mathematical concepts we are going to use throughout the thesis. In particular, we recall basic definitions from linear algebra and polyhedral theory. Also, we present linear programming and the Farkas lemma which provides an efficient way to give a certificate for the infeasibility of a linear program. Finally, to conclude this chapter, we shall review the double description method which is considered as the most efficient algorithm for enumerating the extreme rays of polyhedral cones. The reader familiar with the notions introduced here may skip this chapter and refer to it when necessary.

2.1 Linear Algebra

We start with some notations and terminology. We denote by $\mathbb{R}^n$ the n-dimensional vector space over $\mathbb{R}$. By convention, the vectors in $\mathbb{R}^n$ are column vectors. The superscript "T" denotes *transposition*. So for $x \in \mathbb{R}^n$, x^T is a row vector. Given two vectors $x, y \in \mathbb{R}^n$, $x^T y$ stands for the inner product of x and y, i.e., $\sum_{i=1}^{n} x_i y_i$. The *support* of a vector $x \in \mathbb{R}^n$, denoted by $Supp(x)$, is the index set of its non-zero components, i.e., $Supp(x) = \{i \in \{1, \ldots, n\} \mid x_i \neq 0\}$.

A vector $x \in \mathbb{R}^n$ is a *linear combination* of the vectors $x^1, \ldots, x^p \in \mathbb{R}^n$ if

$$x = \sum_{i=1}^{p} \lambda_i x^i, \text{ for some } \lambda_1, \ldots, \lambda_p \in \mathbb{R}.$$

If in addition, $\lambda_1, \ldots, \lambda_p \geq 0$ (resp. $\sum_{i=1}^{p} \lambda_i = 1$), x is a *conic* (resp. *affine*) combination of $x^1, \ldots, x^p$.

For a set $X \subseteq \mathbb{R}^n$, $X \neq \emptyset$, the *linear* (resp. *affine*, *conic*) *hull* of X, denoted by $\mathrm{lin}(X)$ (resp. $\mathrm{aff}(X)$, $\mathrm{cone}(X)$), is the set of all linear (resp. affine, conic) combinations of finitely many vectors of X.

A set $X \subseteq \mathbb{R}^n$, $X \neq \emptyset$, is called *linearly* (resp. *affinely*) *independent* if no vector $x \in X$ is expressible as a linear (resp. affine) combination of the vectors in $X \setminus \{x\}$, otherwise X is called *linearly* (resp. *affinely*) *dependent*. The cardinality of X is denoted by $|X|$. The *rank* of X (resp. *affine rank* of X), denoted $\mathrm{rank}(X)$ (resp. $\mathrm{arank}(X)$), is the cardinality of the largest linearly (resp. affinely) independent subset(s) of X. The *dimension* of X, denoted by $\dim(X)$, is the cardinality of the

largest affinely independent subset(s) of X minus one, i.e., $\dim(X) = \operatorname{arank}(X) - 1$. It is well known that, if $0 \in \operatorname{aff}(X)$ then $\operatorname{arank}(X) = \operatorname{rank}(X) + 1$, otherwise $\operatorname{arank}(X) = \operatorname{rank}(X)$.

Let $A \in \mathbb{R}^{m \times n}$ be a real matrix. For any row index $i \in \{1, \ldots, m\}$ and any subset $I \subseteq \{1, \ldots, m\}$, we denote by A_{i*} the row of A indexed by i and by A_{I*} the submatrix of A formed by the rows A_{i*} with $i \in I$. Likewise, for any column index $j \in \{1, \ldots, n\}$ and any subset $J \subseteq \{1, \ldots, n\}$, A_{*j} denotes the column of A indexed by j, while A_{*J} denotes the submatrix of A formed by the columns A_{*j}, with $j \in J$. The *column* (resp. *row*) *rank* of A is the rank of its column (resp. row) vectors. It is shown in linear algebra that the column rank and the row rank are always equal. Accordingly, they are simply called the *rank* of A and denoted by $\operatorname{rank}(A)$. The matrix A has *full rank* if $\operatorname{rank}(A) = \min(m, n)$. The *kernel* or *null space* of A, denoted by $kern(A)$, is the set of all vectors $x \in \mathbb{R}^n$ for which $Ax = 0$, i.e., $kern(A) = \{x \in \mathbb{R}^n \mid Ax = 0\}$.

2.2 Polyhedral Theory

Throughout this thesis polyhedral theory plays a central role. This section introduces some basic facts about polyhedral cones. For a comprehensive treatment of this subject the reader should refer to [8; 92; 125].

Definition 2.1 (Convex cone)**.** *A non-empty subset $C \subseteq \mathbb{R}^n$ is called a* (convex) cone *if $\lambda x + \mu y \in C$, whenever $x, y \in C$ and $\lambda, \mu \geq 0$.*

Definition 2.2 (Linear homogeneous inequality)**.** *A linear inequality is an expression of the form $a^T x \Diamond b$ with $a, x \in \mathbb{R}^n$, $b \in \mathbb{R}$ and $\Diamond \in \{\leq, \geq\}$. A linear inequality is homogeneous if $b = 0$.*

A *system* of homogeneous linear inequalities is a finite conjunction of homogeneous linear inequalities and can concisely be written in matrix form as $Ax \geq 0$, where $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$ and m is the number of inequalities in the system. The set of vectors satisfying a finite system of homogeneous linear inequalities is called a *polyhedral cone*.

Definition 2.3 (Polyhedral cone)**.** *A cone C is* polyhedral*, if C is the set of solutions of a finite system of linear homogeneous inequalities, i.e., $C = \{x \in \mathbb{R}^n \mid Ax \geq 0\}$, for some real matrix $A \in \mathbb{R}^{m \times n}$.*

For any $a \in \mathbb{R}^n \setminus \{0\}$, the vector subspace $H = \{x \in \mathbb{R}^n \mid a^T x = 0\}$ is called a *hyperplane*. H partitions the vector space $\mathbb{R}^n$ into two *halfspaces*: $H^+ = \{x \in \mathbb{R}^n \mid a^T x \geq 0\}$ and $H^- = \{x \in \mathbb{R}^n \mid a^T x \leq 0\}$. Accordingly, a cone $C = \{x \in \mathbb{R}^n \mid Ax \geq 0\}$ can be seen as the intersection of finitely many halfspaces. This observation is used in many algorithms on polyhedral cones (e.g., the double description method reviewed in Sect. 2.4).

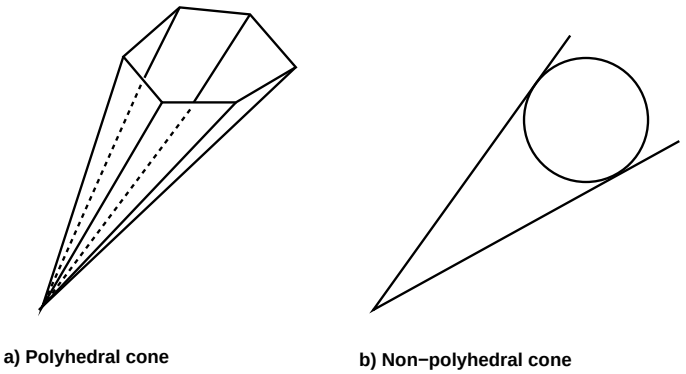


Figure 2.1: Polyhedral vs. non-polyhedral cone

Definition 2.4 (Lineality space)**.** *Let C be a polyhedral cone defined by $C = \{x \in \mathbb{R}^n \mid Ax \geq 0\}$. The* lineality space *of C, denoted by $\text{lin.space}(C)$, is defined by*

$$\text{lin.space}(C) = \{x \in \mathbb{R}^n \mid Ax = 0\}.$$

A cone C is *finitely generated* if there exist $g^1, \ldots, g^s \in \mathbb{R}^n$ such that $C = \text{cone}\{g^1, \ldots, g^s\} = \{\lambda_1 g^1 + \ldots + \lambda_s g^s \mid \lambda_1, \ldots, \lambda_s \geq 0\}$. If this is the case, the vectors g^i for $i = 1, \ldots, s$ are called *generating vectors* of the cone C. The set $B = \{g^1, \ldots, g^s\}$ is called a *minimal set of generating vectors* if no proper subset of B generates the cone C.

A fundamental theorem of Farkas-Minkowski-Weyl (see e.g., [92], p. 87) asserts that a convex cone is polyhedral if and only if it is finitely generated.

Theorem 2.5 (Farkas-Minkowski-Weyl)**.** *A convex cone is polyhedral if and only if it is finitely generated.*

This theorem states that every cone admits two types of representations, either as the solution set of a finite system of linear homogeneous inequalities or as the conic hull of a finite set of vectors. These are usually referred to as *external* and *internal representation*, respectively. For the rest of the thesis, we will consider only polyhedral cones and simply use the term cone.

An inequality $a^T x \geq 0, a \in \mathbb{R}^n \setminus \{0\}$, is *valid* for a cone C if $C \subseteq \{x \in \mathbb{R}^n \mid a^T x \geq 0\}$.

Definition 2.6 (Face)**.** *A subset F of a cone C is called a* face *of C if $F = C$ or $F = C \cap \{x \in \mathbb{R}^n \mid a^T x = 0\}$, where $a^T x \geq 0, a \in \mathbb{R}^n \setminus \{0\}$, is a valid inequality for C. The* dimension *of F is defined as the dimension of the linear hull of F, i.e., $\dim(F) = \dim(\text{lin}(F))$.*

Definition 2.7 (Minimal face)**.** *A* minimal *face of a cone C is a non-empty face which does not contain any other non-empty face of C.*

In other words, a minimal face of C is a face of smallest dimension. It is easy to see that the only minimal face of a cone C is its lineality space. This leads to the following definition of *minimal proper faces* of C.

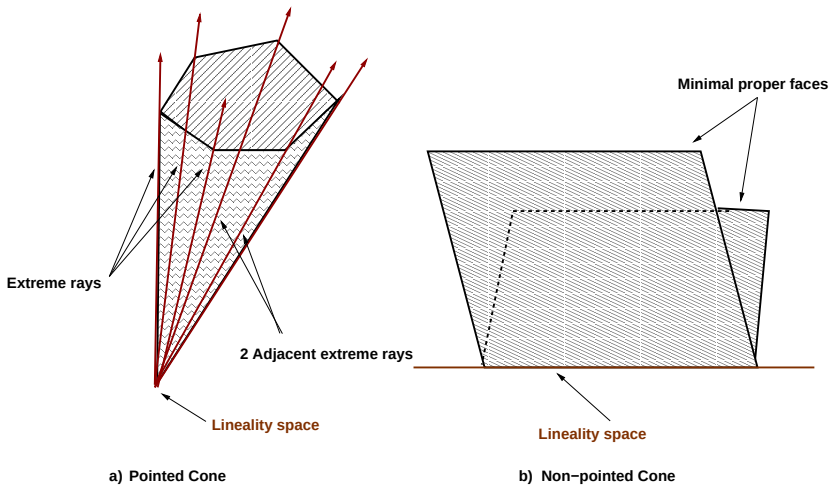


Figure 2.2: Pointed vs. non-pointed cone.

Definition 2.8 (Minimal proper face)**.** *Let C be a cone and let t be the dimension of its lineality space. A face of C of dimension $t+1$ is called a* minimal proper *face.*

Definition 2.9 (Adjacent minimal proper faces)**.** *Let C be a cone and let t be the dimension of its lineality space. Two minimal proper faces of C are called* adjacent *if they are contained in some face of dimension $t+2$.*

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 $\lambda > 0$.

Definition 2.10 (Extreme ray)**.** *Let C be a cone. A ray r is* extreme *if there do not exist rays $r', r'' \in C, r' \ncong r''$, such that $r = r' + r''$.*

Definition 2.11. *A cone C is called* pointed *if $\text{lin.space}(C) = \{0\}$ or equivalently $C = \{x \in \mathbb{R}^n \mid Ax \geq 0\}$ for some $A \in \mathbb{R}^{m \times n}$ with $rank(A) = n$.*

An example of a pointed cone is the intersection of the null space of a matrix with the positive orthant. Note that the extreme rays of a pointed cone are identical with its minimal proper faces. According to Definition 2.9, two extreme rays are adjacent if they are contained in one face of dimension 2.

Any pointed cone C has a *canonical* representation

$$C = \text{cone}\{r^1, \ldots, r^s\}, \tag{2.1}$$

where $r^1, \ldots, r^s$ are the (distinct) extreme rays of C. This representation, which is often used in metabolic network analysis (see Chap. 3 for more details), is minimal and unique up to multiplication by positive scalars.

If C is not pointed, there is no longer such a unique minimal representation. Let t be the dimension of the lineality space of C. Instead of the extreme rays, we consider now the *minimal proper faces* $G^1, \ldots, G^s$ of C, which are the faces of C of dimension

$t+1$. Each G^i can be represented by a row vector a_i^T and a submatrix A'_i of A, with $\text{rank}\begin{pmatrix} A'_i \\ a_i^T \end{pmatrix} = n-t$, such that [92]

$$G^i = \{x \in C \mid a_i^T x \geq 0, A'_i x = 0\}, \tag{2.2}$$

and

$$\text{lin.space}(C) = \{x \in C \mid a_i^T x = 0, A'_i x = 0\}.$$

If we select for each $i = 1, \ldots, s$ a vector $g^i \in G^i \setminus \text{lin.space}(C)$, and vectors $b^0, \ldots, b^t \in \text{lin.space}(C)$ such that $\text{lin.space}(C) = \text{cone}\{b^0, \ldots, b^t\}$, we get

$$C = \text{cone}\{g^1, \ldots, g^s, b^0, \ldots, b^t\}. \tag{2.3}$$

For each minimal proper face $G^i, i = 1, \ldots, s$, we get

$$G^i = \text{cone}\{g^i\} + \text{lin.space}(C) = \{\lambda g^i + w \mid \lambda \geq 0, w \in \text{lin.space}(C)\} \tag{2.4}$$

For additional information, we refer to [92], p. 105-106.

(2.3) generalizes (2.1), but this representation is no longer unique. We may choose an arbitrary base of $\text{lin.space}(C)$, and arbitrary rays g^i in $G^i \setminus \text{lin.space}(C)$. However, it follows from (2.2) that G^i can also be characterized using constraints $a_i^T x \geq 0$, where a_i^T is a row vector from the matrix A that defines the cone. This observation will lead us to a new way for describing and analyzing the flux cone associated with a metabolic network (see Chap. 4 for more details).

Let us now illustrate the concepts that have been introduced above using the following example.

Example 2.12. *Consider the cone C defined by*

$$C = \{(x, y, z)^T \in \mathbb{R}^3 \mid -11x + 4z \geq 0 \text{ and } x + 3z \geq 0\}.$$

Let G^1 and G^2 be the faces defined by

$$G^1 = \{(x, y, z)^T \in \mathbb{R}^3 \mid -x + 4z = 0 \text{ and } 2x + z \geq 0\},$$
$$G^2 = \{(x, y, z)^T \in \mathbb{R}^3 \mid -x + 4z \geq 0 \text{ and } 2x + z = 0\}.$$

The linear hulls of G^1 and G^2 are respectively,

$$\text{lin}(G^1) = \{(x, y, z)^T \in \mathbb{R}^3 \mid -x + 4z = 0\},\ \dim(\text{lin}(G^1)) = 2,$$
$$\text{lin}(G^2) = \{(x, y, z)^T \in \mathbb{R}^3 \mid 2x + z = 0\},\ \dim(\text{lin}(G^2)) = 2.$$

Fig. 2.3 shows the polyhedral cone C and the faces G^1 and G^2. The y-axis is the lineality space of C, i.e., $\text{lin.space}(C) = \{(x, y, z)^T \in \mathbb{R}^n \mid x = z = 0\} = \{\lambda b^1 \mid \lambda \in \mathbb{R}\}$, with $b^1 = (0, 1, 0)^T$, and so $t = \dim(\text{lin.space}(C)) = 1$. Since $\dim(G^1) = \dim(\text{lin}(G^1)) = t + 1$ and $\dim(G^2) = \dim(\text{lin}(G^2)) = t + 1$, G^1 and G^2 are minimal proper faces of C that could be represented by the vectors $g^1 = (1, 0, 4)^T$ and $g^2 = (-1, 0, 2)^T$, respectively.

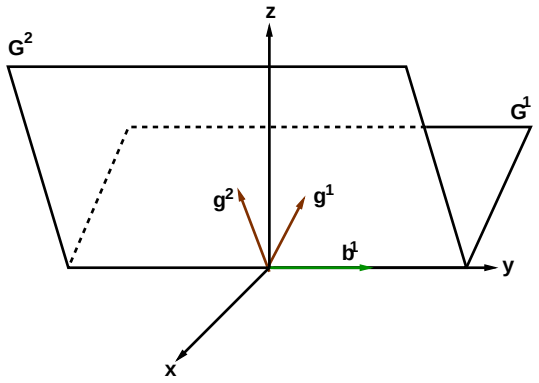


Figure 2.3: An illustration of the polyhedral cone C corresponding to the constraint system given in Example 2.12.

According to Theorem 2.5, each polyhedral cone can be seen either as the solution set of a finite system of linear homogeneous inequalities (external representation) or as the conic hull of a finite set of vectors (internal representation). The problem of converting an external representation to an internal representation is known as the *extreme ray enumeration problem* and the reverse is known as the *facet enumeration problem*. These problems have been well studied and have many equivalent formulations [44]. Clearly, the number of extreme rays of the cone $C = \{x \in \mathbb{R}^n \mid Ax \geq 0\}$, for some real matrix $A \in \mathbb{R}^{m \times n}$, can be (and typically is) exponential in n or m (see Sect. 4.4 for an example). Therefore, when we consider the computational complexity of the extreme ray enumeration problem, one can only seek *output-sensitive* algorithms whose running time depends not only on the size of the input but also on the size of the output. In general, no polynomial output-sensitive algorithm is known for the extreme ray enumeration problem [44; 112]. Nonetheless, the *double description method* [33], which will be reviewed in Sect. 2.4, is one of the most efficient algorithms for solving this problem.

2.3 Linear Programming

In this section we briefly describe the basic theory of linear programming. Details can be found e.g. in [19; 92].

Linear programming (LP) is the problem of maximizing or minimizing some linear function, called the *objective function*, subject to a set of linear inequalities. In the following, without loss of generality, we maximize the objective function.

More formally, given a matrix $A \in \mathbb{R}^{m \times n}$, a vector $b \in \mathbb{R}^m$ and a vector $c \in \mathbb{R}^n$, the corresponding LP problem, is denoted by

$$\max\ c^T x \text{ subject to } Ax \leq b,$$

or shortly

$$\max\{c^T x : Ax \leq b\}. \tag{2.5}$$

Each of the linear inequalities $A_{i*}x \leq b_i$ for $i = 1, \ldots, m$ is called a *linear constraint.* A vector $x^* \in \mathbb{R}^n$ is a *feasible solution* if x^* satisfies all the linear constraints, i.e., $Ax^* \leq b$. If in addition, $c^T x^* \geq c^T x$ for all feasible solutions x, x^* is called an *optimal solution.* The set of all feasible solutions is called the *feasible region.* A linear program is *infeasible* if its feasible region is empty, otherwise it is called *feasible.*

LP problems, in which some variables are required to be integers but others can be real valued, are known as *mixed integer linear programming (MILP) problems.*

Linear programming can be solved using the *simplex method,* which enumerates adjacent vertices of the feasible region such that at each new vertex the objective function improves or is unchanged. Another efficient polynomial time algorithm is the *interior point method.* The interested reader can refer to [92].

Every LP problem, referred to as a *primal* problem, can be converted into a *dual* problem by swapping the constraints and variables. For instance, the dual of the LP problem (2.5) can be written as

$$\min\{b^T y : A^T y = c,\ y \geq 0\}. \tag{2.6}$$

For a more detailed treatment, the reader is again referred to [92]. The following theorem is one of the main results in duality theory.

Theorem 2.13. *The linear program (2.5) has an optimal solution x^* if and only if the dual linear program (2.6) has an optimal solution y^* such that $c^T x^* = b^T y^*$.*

There are many practical applications of duality. For instance, it might be faster to run the simplex method on the dual linear program. More importantly, duality provides an efficient way to give a certificate for the infeasibility of a linear program. The following theorem, called the *Farkas lemma*, states that the unsolvability of a system of constraints can be established by finding a solution for a corresponding dual system.

Theorem 2.14 (Farkas lemma)**.** *Given a matrix $A \in \mathbb{R}^{m \times n}$ and a vector $b \in \mathbb{R}^m$, there exists a vector $x \in \mathbb{R}^n$ such that $Ax \leq b$ if and only if there does not exist a vector $y \in \mathbb{R}^m$ such that $y^T A = 0$, $y \geq 0$, and $y^T b < 0$.*

In the context of our work, we instead use the following variant of the Farkas lemma for equality and inequality constraints to characterize minimal directions cuts in metabolic networks (see Chap. 7 for more details).

Theorem 2.15 (Farkas lemma for equality and inequality constraints [125])**.** *Given matrices $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{p \times n}$ and $C \in \mathbb{R}^{q \times n}$ and vectors $x \in \mathbb{R}^m$, $y \in \mathbb{R}^p$ and $z \in \mathbb{R}^q$, **either** there exists a solution vector $v \in \mathbb{R}^n$ for*

$$Av = x,\ Bv \geq y,\ Cv \leq z$$

***or** there exist row vectors $a \in \mathbb{R}^m$, $b \in \mathbb{R}^p$ and $c \in \mathbb{R}^q$ with*

$$cC = aA + bB,\ b \geq 0,\ c \geq 0,\ -ax - by + cz < 0.$$

2.4 Double Description Method

The *double description method (DDM)* is a simple yet efficient algorithm for computing a minimal generating set of a polyhedral cone in $\mathbb{R}^n$. To understand this algorithm, we first recall the concept of a *double description pair (DDP)* [33].

Definition 2.16 (Double Description Pair (DDP))**.** *Given some matrices $A \in \mathbb{R}^{m\times n}$, $R \in \mathbb{R}^{n\times s}$, $B \in \mathbb{R}^{n\times t}$ and $G = \begin{pmatrix} R & B \end{pmatrix}$, the pair (A, G) is said to be a* double description pair *(DDP) if the following relationship holds:*

$$Ax \geq 0 \text{ if and only if } x = R\lambda + B\mu \text{ for some } \lambda \in \mathbb{R}^s_{\geq 0} \text{ and } \mu \in \mathbb{R}^t.$$

The term "double description" means that such a pair (A, G) contains two different descriptions of the same cone C. Namely, C represented by its *representation matrix* A as

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

is also given by its *generating matrix* $G = (R, B)$ as

$$C = \{x \in \mathbb{R}^n \mid x = R\lambda + B\mu \text{ for some } \lambda \in \mathbb{R}^s_{\geq 0} \text{ and } \mu \in \mathbb{R}^t\}.$$

By Theorem 2.5, each polyhedral cone admits a generating matrix. Clearly, it is more appropriate to construct a minimal generating matrix G so that no proper submatrix of G is generating the same cone. The strategy of the double description method is to iteratively build a minimal generating matrix (R', B') for the cone $C^k = \{x \in \mathbb{R}^n \mid A_{i*}x \geq 0 \text{ for } i = 1, \ldots, k\}$ from a minimal generating matrix (R, B) of the cone $C^{k-1} = \{x \in \mathbb{R}^n \mid A_{i*}x \geq 0 \text{ for } i = 1, \ldots, k-1\}$, such that $B = (b_1, \ldots, b_{t_{k-1}})$, with $t_{k-1} = \dim(\text{lin.space}(C^{k-1}))$, is a basis of the lineality space of C^{k-1} and $R = (r_1, \ldots, r_{s_{k-1}})$ is a set of representatives of the minimal proper faces (faces of dimension $t_{k-1} + 1$) of C^{k-1}. The incremental step introduces a constraint $A_{k*}x \geq 0$ that is not yet fulfilled by all the generators in (R, B). The generators in (R, B) that fulfill this constraint will be kept in the description of C^k, the others will be discarded and new ones are generated. The computation of the new generators relies on the concept of adjacent rays. Two generating vectors $r_i, r_j \in R$ are adjacent if they are contained in some $(t + 2)$-dimensional face of C^{k-1}. Fig. 2.4 gives an illustration of the k-th iteration. The constraint $A_{k*}x \geq 0$ partitions the set of generators R into three parts:

$$\begin{aligned} R^+ &= \{r \in R \mid A_{k*}r > 0\}, \\ R^0 &= \{r \in R \mid A_{k*}r = 0\}, \\ R^- &= \{r \in R \mid A_{k*}r < 0\}. \end{aligned}$$

A minimal set of generators (R', B') for the cone C^k is determined by the following rules [33]:

1. If $A_{k*} \perp B$, i.e., $\text{lin.space}(C^{k-1}) \subseteq \{x \in \mathbb{R}^n \mid A_{k*}x = 0\}$, then $B' = B$ and $R' = R^+ \cup R^0 \cup Adj$, with

   $$Adj = \{(A_{k*}r_i)\cdot r_j - (A_{k*}r_j)\cdot r_i \mid r_i \in R^+,\ r_j \in R^-,\ r_i \text{ and } r_j \text{ are adjacent in } C^{k-1}\}$$

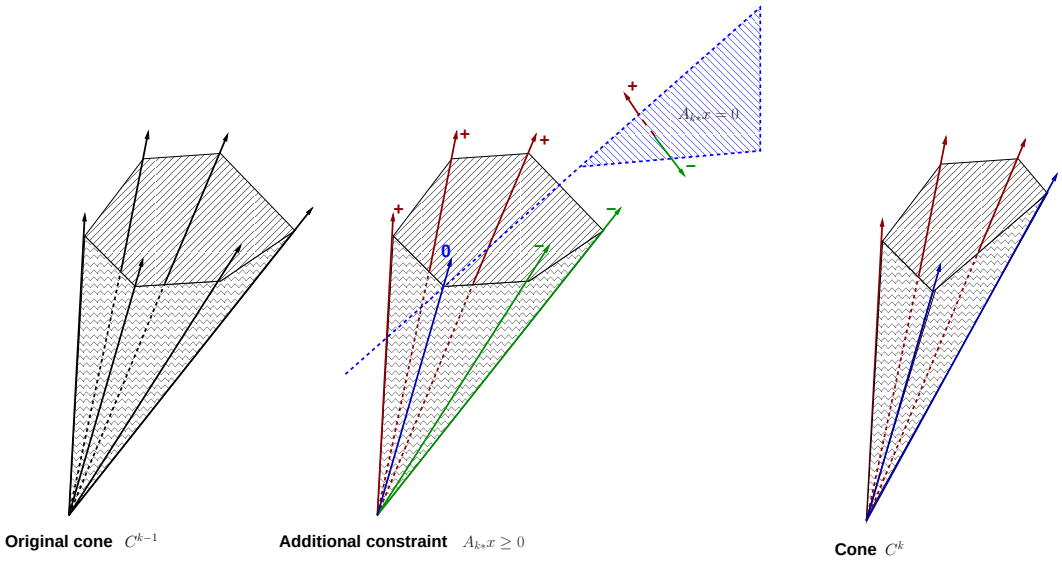


Figure 2.4: Illustration of the k-th iteration of the double description method. In this iteration, the constraint $A_{k*}x \geq 0$ partitions the generating vectors of the cone C^{k-1} into positive (red color), negative (green color) and zero (blue color) generators. All positive and zero generators of the cone C^{k-1} are kept in the description of the cone C^k. Other new generator of the cone C^k are obtained by combining a positive and a negative generator of C^{k-1} that are adjacent.

2. Else choose $B = \{b'_1, \ldots, b'_{t_{k-1}}\}$ such that $A_{k*}b'_i = 0$ for $i = 1, \ldots, t_{k-1} - 1$ and $A_{k*}b'_{t_{k-1}} > 0$, and set $B' = \{b'_1, \ldots, b'_{t_{k-1}-1}\}$ and $R' = \{r'_1, \ldots, , r'_{s_{k-1}}, b'_{t_{k-1}}\}$, with $r'_j = (A_{k*}b'_{t_{k-1}}) \cdot r_j - (A_{k*}r_j) \cdot b'_{t_{k-1}}$ for $j = 1, \ldots, s_{k-1}$.

At each iteration k, the DDM algorithm needs to check for each pair of generators r and r' of C^k with $A_{k*}r > 0$ and $A_{k*}r' < 0$ whether they are adjacent in C^k. Actually, enumerating adjacent rays is the most time consuming part of the DDM algorithm. Performing this enumeration would improve the DDM algorithm. We should also mention that this algorithm is sensitive to the ordering of rows of the representation matrix A. Many (fixed or dynamic) ordering strategies have been proposed and compared in [33].

CHAPTER

3 Metabolic Network Analysis

In this chapter, we give an overview of the main approaches for modeling and analyzing metabolic networks. We will focus on *constraint-based* approaches which, through the help of *constraints* such as *stoichiometric* and *thermodynamic* constraints, define the space of all attainable behaviors of a metabolic network at steady state. In the last part of this chapter, we shall discuss the well-known concept of *elementary modes*. While each elementary mode involves a minimal set of reactions, these modes span the steady-state flux cone in the sense that, each steady-state flux distribution can be expressed as a non-negative linear combination of the elementary modes.

3.1 Main Existing Approaches

The impressive biological data that is now available has allowed the reconstruction of an increasing number of genome-scale metabolic networks. However, this information is not sufficient to determine quantitatively the metabolic phenotypes that are expressed by biological systems under different environmental conditions. Intuitive reasoning for predicting and analyzing metabolic phenotypes can be inadequate, often giving incomplete or incorrect predictions. In this respect, rigorous mathematical and computational methods are strongly required to investigate the principles of metabolic behaviors.

In an attempt to make quantitative predictions about the dynamics of the metabolic behavior, *kinetic modeling* [38] determines the reaction rate by means of kinetic functions of metabolite concentrations and kinetic parameters such as equilibrium constants. A well-known kinetic function is the *Michaelis-Menten* rate equation, which defines the flux through some reaction i as a function of the concentration of a substrate A as follows:

$$v_i = v_i^{max} \frac{[A]}{[A] + K_m},$$

where v_i is the flux through reaction i, $[A]$ is the concentration of the substrate A, v_i^{max} is the maximal possible flux through reaction i and K_m is a Michaelis-Menten parameter, which is equal to the substrate concentration when

$$v_i = \frac{v_i^{max}}{2}.$$

Although kinetic modeling is most appropriate for fully characterizing metabolic systems, this approach has been hampered by the lack of detailed kinetic information. Indeed, it is difficult to determine kinetic parameters experimentally. Consequently, there is often not enough kinetic information in the literature to construct kinetic models. Moreover, the results that could be drawn from such models are strongly sensitive to the definition of both kinetic functions and parameters.

In view of the limits of kinetic modeling, growing attention is being paid to other approaches which require only topological and structural properties rather than detailed kinetic information. Indeed, using the language of graph theory, a metabolic network can be captured as a graph representation of the reactions taking place in that network. In this representation, the metabolic network is seen as either a *reaction graph* [40] or a *substrate graph* [118], allowing for the use of the tools of graph theory to investigate topological properties of metabolic networks. These properties include the *clustering coefficient* [85], as well as the *average path-length* between metabolites [1].

While graph-based methods have greatly contributed to the understanding of metabolic network topology, they are limited by a crucial difference between graphs and metabolic networks. Indeed, about 85% of reactions are bimolecular or more in either substrates, products or both [31]. For instance, the following is a reaction involving two substrates and two products,

$$\text{glucose} + \text{ATP} \rightarrow \text{glucose 6-phosphate} + \text{ADP}.$$

A graph-based representation does not explicitly consider the hypergraphical nature of metabolic networks. A *hypergraph* is a generalized graph in which hyperedges (reactions) may link more than two nodes (metabolites). Besides, graph-based approaches do not straightforwardly consider reaction stoichiometry. One topological approach to coping with this situation is to use *Petri Net* theory [124]. This approach has many parallels with pathway-based network analysis reviewed in Sect. 3.4. In addition to be a user-friendly means of visualization, Petri nets allows for characterizing several properties of metabolic networks (*trap*, *deadlock* and *liveness*, to name but a few) [71]. However, determining these properties requires a translation of Petri nets into linear constraint systems. This modeling can be obtained using constraint-based approaches.

Rather than attempting to predict exactly what a metabolic system does, one could narrow the range of all possible behaviors this system can display under certain physicochemical constraints [23; 72; 81]. In such an approach, we assume that these constraints define the space of all possible attainable behaviors of a metabolic network. Constraint-based approaches are simple but powerful [104]. Indeed, the few parameters used in such approaches enable models to be built and studied easily. Although such approaches do not strive to find exact behaviors of metabolic networks, the degree of freedom of the constrained system yields an indication of our level of understanding of the metabolic system. If all the constraints that govern a metabolic network are known, the allowable domain shrinks to a single feasible behavior. This is often not the case, and so many further flux distributions fulfill the governing constraints. Nonetheless, adding further constraints may reduce the number of allowable

flux distributions in the network. This is the strategy used by *metabolic flux analysis (MFA)* [64; 107]. In this approach, fluxes through many boundary reactions are measured to render a determined constraint system. Besides, consistency tests of governing constraints make possible improving metabolic reconstructions. For instance, some reaction could be unable of carrying a flux under constraints taken into account in the model. This could indicate the presence of errors or omissions during network reconstruction processes.

In the context of the present work, we are specifically concerned with deriving all possible flux distributions subject to the two most important constraints that have to hold in a metabolic network at steady state: the *stoichiometric* and *thermodynamic* constraints. As we will see in the next section, these constraints define the steady-state flux cone that contains all possible flux distributions in a metabolic network subsisting at steady state.

3.2 Steady-State Flux Cone

According to kinetic theory, the difference between the rate of formation and consumption of a particular internal metabolite is equal to the change in concentration of that metabolite over time. Mathematically, the behavior of a metabolic network can be captured as a system of ordinary differential equations [38]. A compact expression of this equation system is

$$\frac{dx}{dt} = Sv, \tag{3.1}$$

where S is the stoichiometric matrix, x stands for the $m-$dimensional vector of internal metabolite concentrations and v denotes the *flux distribution* with elements corresponding to the n fluxes through reactions. Actually, the flux vector v consists of nonlinear functions of metabolite concentrations x as well as of a set of kinetic parameters (e.g., Michaelis constants K_m, maximal reaction rates v_m and equilibrium constants K_{eq}). For all except very simple cases, the constraint system (3.1) cannot be solved analytically, but it can be investigated numerically with appropriate nonlinear solvers. Nonetheless, if we consider only the fluxes through reactions, then the constraint system (3.1) is linear in them.

Identification of steady states plays a crucial role in the analysis of metabolic networks [98]. At steady state, the change in the concentration of a compound x over time t across all reactions within the system becomes zero. This assumption is relevant for most metabolic reactions since they are typically much faster than environmental changes [34; 115]. The steady-state assumption is expressed by a zero time derivative of the concentration, leading to the *flux balance equation*

$$Sv = 0. \tag{3.2}$$

This equation defines the *stoichiometric constraints* which state that the total rate of formation for any internal metabolite must equal the total rate of consumption for that

metabolite. Remember that the flux balance equation holds only for internal metabolites, since each internal metabolite must not accumulate or decrease in time. In contrast, external metabolites could be buffered, and so these metabolites are not involved in the flux balance equation.

Stoichiometric constraints are mandatory for understanding the behaviour of metabolic networks, but they are far from sufficient. These constraints allow a wide range of possible steady-state flux distributions, namely all flux distributions that are situated in the null space of the stoichiometric matrix (see Sect. 2.1 for a definition of the null space of a matrix). Additional constraints imposed by thermodynamic considerations are often crucial to reduce the range of possible flux distributions. Indeed, since each irreversible reaction can proceed only in the forward direction, fluxes through irreversible reactions must be greater than or equal to zero. This is stated by the following constraint system

$$v_i \geq 0, \text{ for all } i \in Irr. \tag{3.3}$$

According to [78], a set of reactions forms a *functionally coherent set* in metabolism if the flux distribution v realizable by these reactions obeys both the stoichiometric and thermodynamic constraints, i.e., v fulfills the following linear constraint system

$$Sv = 0, \;\; v_i \geq 0, \text{ for all } i \in Irr, \tag{3.4}$$

wherein the number of constraints ($m + |Irr|$) is often far less than the number n of unknown fluxes. Consequently, this set of linear constraints is, in general, underdetermined. Hence, multiple steady-state solutions are possible, each representing a possible flux distribution over the network at steady state. In addition, owing to the linear inequalities (3.3), the mathematical problem (3.4) is beyond the scope of standard linear algebra. In polyhedral theory, it is shown that the solutions of the constraint system (3.4) form a polyhedral cone in the flux space (see Chap. 2 for a definition of a polyhedral cone).

Definition 3.1 (Flux cone [20], p. 20-21)**.** *The set of all solutions of the constraint system (3.4), which corresponds to the set of all possible flux distributions over the network at steady state, defines a polyhedral cone,*

$$C = \{v \in \mathbb{R}^n \mid Sv = 0, \; v_i \geq 0, \textit{for all } i \in Irr\}, \tag{3.5}$$

which is called the steady-state flux cone.

For the rest of the thesis, we shall use the simpler term flux cone for steady-state flux cone.

Since the flux cone in general contains infinitely many possible steady-state flux distributions, it is interesting to find out which of these feasible flux distributions are actually displayed by the metabolic network under consideration. Currently, constraint-based approaches have attempted to analyze metabolic networks by use of different mathematical and computational tools (linear algebra, polyhedral theory and

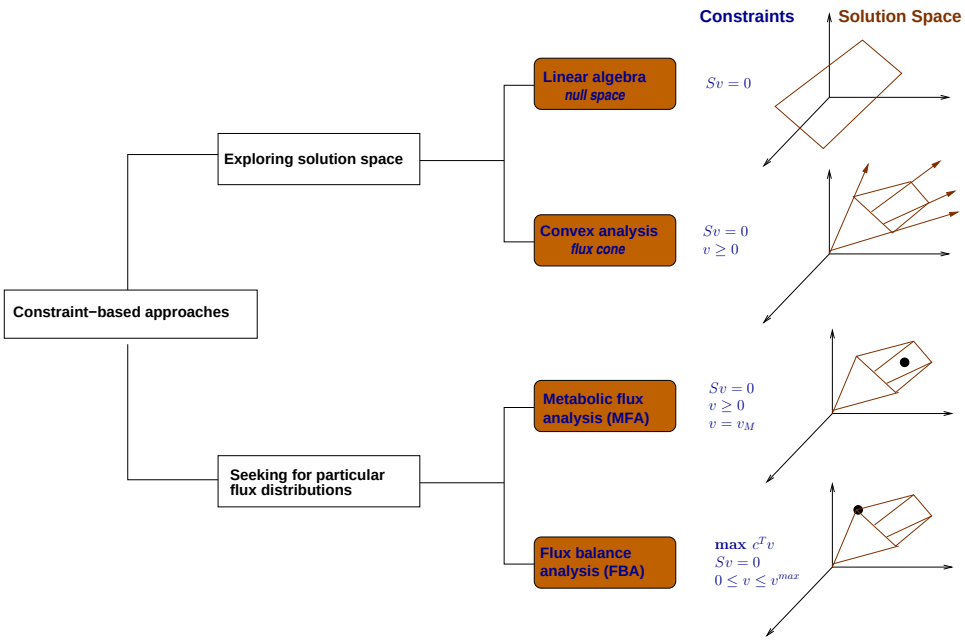


Figure 3.1: Constraint-based analysis of metabolic networks. The stoichiometric and thermodynamic constraints that have to hold in a metabolic network define the space of all possible flux distributions over the network. Optimization-based approaches seek to identify single behaviors that optimize a predefined objective function. Alternatively, the whole capabilities of the metabolic network could be assessed using pathway-based network analysis.

linear programming, to name but a few). There are two main strategies to analyze a metabolic network: searching for single optimal behaviors using *optimization-based approaches*, or assessing the whole capabilities of a metabolic network by means of *pathway-based network analysis*. Fig. 3.1 gives an overview of these two categories of constraint-based approaches.

3.3 Optimization-based Approaches

Optimization-based approaches assume that metabolic networks behave optimally, driven by an objective. For applying such methods, we first need to determine a most likely physiologically meaningful objective of a living system. This is interesting, as it may allow us to identify rules that govern the operation of a metabolic network under different environmental conditions. These governing rules are important not only for enhancing our understanding of the metabolic system, but also for engineering and designing strains more appropriate for metabolite production purposes [10; 12].

In an attempt to determine an optimal flux distribution, one should state an objective function and seek its maximal value within the feasible domain. This approach is referred to as *flux balance analysis (FBA)* [42; 63]. In addition to the stoichiometric and thermodynamic constraints, the optimization strategy employed by FBA uses flux

capacity constraints that place bounds on the values of a given flux, and possibly other physicochemical constraints to further limit the space of possible flux distributions. These additional constraints are necessary because using optimization techniques requires that the feasible domain is bounded in the direction of the objective function. Assuming the objective function is linear, FBA constitutes the following linear programming problem (see Sect. 2.3 about linear programming)

$$\begin{aligned} &\max\ c^T v \\ &\text{subject to:} \\ &Sv = 0, \\ &v_i^{min} \leq v_i \leq v_i^{max} \text{ for all } i \in \{1, \ldots, n\}, \end{aligned} \tag{3.6}$$

where c denotes the vector that defines the objective function by means of costs of or benefits derived from the fluxes [84]. Note that also minimization can be performed by simply finding the maximum of the negative of the objective function. The bounds v_i^{min} and v_i^{max} are the minimum and maximum flux capacities. In particular, $v_i^{min} = 0$ for each irreversible reaction i. By changing the vector c in the linear problem (3.6), we could test various objective functions, each capturing different information on which rules are governing metabolic networks. Many optimization-based approaches assume that a well-suitable objective function for an optimal operation of metabolic networks is the maximization of biomass production (growth) [28; 115]. They consider that microorganisms have evolved in such a way that their metabolic networks guarantee the most efficient conversion of resources to produce more cells. This simple optimization principle has been widely used for many studies, such as predicting the optimal performance of a metabolic network under a range of growth conditions, studying gene essentiality and identifying targets for metabolic engineering [10; 27; 82]. On the other hand, some microorganisms may not necessarily have evolved solely to optimize biomass production. They instead may be driven by other objective functions including [26]:

- *Minimizing ATP production* to guarantee optimal metabolic energy efficiency.
- *Minimizing nutrient uptake* to minimize the amount of available nutrients that are needed by the living system to perform its metabolic functions.
- *Minimizing the overall flux* used to efficiently channel the metabolites through the metabolic network.
- *Maximizing the production* of a chosen metabolite to determine the production capabilities of a given metabolic network.

No single objective function completely describes the optimal operation of metabolic networks under all environmental conditions. Accordingly, it is still mandatory to verify whether a hypothesized objective function is consistent with experimental flux data. Recent work systematically evaluates the relevance of eleven objective functions to predict fluxes in *E. coli* under six environmental conditions [93]. Another

approach uses an optimization-based framework (*ObjFind*) for inferring the most plausible objective function given observed experimental data [13].

A knocked-out organism may not be accurately described with the objectives used for wild-type systems. A recent approach, called *minimization of metabolic adjustment* (MOMA), assumes that a knocked-out organism displays a flux distribution closest to the optimal flux distribution prior to the gene deletion [103].

MOMA determines a flux distribution with the smallest euclidean distance to the optimal wild-type flux distribution obtained by FBA using the following quadratic program:

$$\begin{aligned}
&\text{min } (v-w)^T(v-w)\\
&\text{subject to:}\\
&Sv = 0,\\
&v_i^{min} \leq v_i \leq v_i^{max} \text{ for all } i \in \{1,\ldots,n\},\\
&v_j = 0, \text{ for all } j \in A,
\end{aligned}$$

where w is the wild-type flux distribution and A is the set of reactions corresponding to the deleted genes. On the other hand, *regulatory on-off minimization (ROOM)* identifies the metabolic flux state of mutants by minimizing the number of significant flux changes from the wild-type flux distribution [105]. For this, ROOM assigns a cost to each flux, defines a range $[w^l, w^u]$ around the wild-type flux distribution w and determines a feasible flux distribution v with a minimal number of components v_i such that $v_i \notin [w_i^l, w_i^u]$. The resulting optimization problem involves a mixed-integer linear problem that can be written as:

$$\begin{aligned}
&\text{min } \textstyle\sum_{i=1}^{n} c_i y_i\\
&\text{subject to:}\\
&Sv = 0,\\
&v_i - y_i(v_i^{max} - w_i^u) \leq w_i^u \text{ for all } i \in \{1,\ldots,n\},\\
&v_i - y_i(v_i^{min} - w_i^l) \geq w_i^l \text{ for all } i \in \{1,\ldots,n\},\\
&v_j = 0, \text{ for all } j \in A,\\
&y_i \in \{0,1\},
\end{aligned}$$

where c_i denotes the cost for a change in the flux through reaction i, and for each reaction $i \in \{1,\ldots,n\}$, $y_i = 1$ for a significant flux change in v_i, i.e., $v_i \notin [w_i^l, w_i^u]$, and $y_i = 0$ otherwise.

In analogy with FBA, both MOMA and ROOM require the definition of an objective function to compute their minimal metabolic adjustment. In addition, these approaches require the solution of FBA for the wild-type organism. Many other optimization-based methods have been developed to analyze metabolic networks including automated curation of metabolic reconstructions [56], recovering metabolic pathways via optimization [7], analysis of gene essentiality [2; 15; 82] and metabolic engineering [10; 12].

While optimization-based approaches have proved successful in analyzing optimal capabilities of several microorganisms, their results are sensitive to the definition of the objective function. Moreover, these methods assume that metabolic systems operate

according to a single rule of optimization. A recent study has shown, however, that a microorganism could use different optimization criteria depending on the environmental conditions [93]. The exploration of all suitable objective functions is still a difficult task.

Furthermore, an optimal solution with respect to a suitable objective function need not be unique. Optimization-based techniques often return an arbitrary chosen flux distribution from the set of all optimal flux distributions. In analogy with the flux cone, the set of all optimal flux distributions is, in general, an infinite convex set and requires an adequate description. A recursive mixed-integer linear programming algorithm has been developed to find all alternative optima [65]. This algorithm was, however, applied only on small networks. Finally, optimization-based approaches consider only optimal states, which form a restricted subset of all possible behaviors of the living system. An interesting alternative to optimization-based modeling is to describe all possible steady-state flux distributions in the metabolic network using pathway-based network analysis.

3.4 Pathway-based Network Analysis

Pathway-based network analysis [51; 75; 91] has been recognized as an important approach in computational biology. This analysis is concerned with describing the infinite flux cone C (defined in equation (3.5)) by means of a finite set of generating vectors. A key distinction to be made is whether the flux cone is pointed or not. By definition, the flux cone is pointed if its lineality space

$$\mathrm{lin.space}(C) = \{v \in C \mid v_i = 0, \text{for all } i \in Irr\} \tag{3.7}$$

is reduced to the origin, i.e., no steady-state flux distribution involves only reversible reactions. In particular, if all reactions are irreversible, i.e., $Irr = \{1, \ldots, n\}$, then $\mathrm{lin.space}(C) = \{0\}$ and so the flux cone is pointed. In this case, the flux cone is generated by a unique (up to multiplication by positive scalars) and minimal set of flux vectors that correspond to its extreme rays.

In the presence of reversible reactions, the situation is more involved. Indeed, if some reactions are reversible, the flux cone may be non-pointed and thus has no longer a unique and minimal representation by its extreme rays. To deal with this situation, some approaches propose to reconfigure the metabolic network in order to render the flux cone pointed [20; 90]. For this, they consider a subset SR $\subseteq Rev$ of reversible reactions and split each reversible reaction $j \in$ SR into a forward and a backward reaction, which both are constrained to be irreversible. Let $s = |\text{SR}|$ and SR $= \{j_1, \ldots, j_s\}$. For convenience, the stoichiometric matrix $S' \in \mathbb{R}^{m \times (n+s)}$ of the reconfigured network can be written as follows:

$$\begin{array}{ll} S'_{*j} = S_{*j} & \text{for all } j \in \{1, \ldots, n\}, \\ S'_{*(n+k)} = -S_{*j_k} & \text{for all } k \in \{1, \ldots, s\}. \end{array}$$

The set of irreversible reactions in the reconfigured network is given by

$$Irr' = Irr \cup \text{SR} \cup \{n+1, \ldots, n+s\}.$$

The *reconfigured flux cone* C', which contains all possible steady-state flux distributions in the reconfigured network, is defined by

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

As a result of this reconfiguration, for a well-chosen set SR of split reactions, the reconfigured flux cone C' is pointed and can be described by a unique and minimal set of extreme rays. This reconfiguration has, however, undesirable consequences. On the one hand, the number of variables and constraints increases by s and $2s$, respectively. This renders more complex the constraint system that defines the reconfigured flux cone. On the other hand, a significant number of rays in the reconfigured cone are extreme for the only reason that the split reversible reactions have been decomposed into forward and backward reactions. In the initial cone, these extreme rays are conically dependent. Accordingly, the number of extreme rays increases by this reconfiguration, which limits the practical applicability of this strategy. A more detailed study of the network reconfiguration is given in Chap. 5.

Three main approaches have been proposed to analyze metabolic networks using inner descriptions of the flux cone [20; 90; 98]. They all determine flux distributions corresponding to a convex basis of the flux cone, but use a different set of reactions that have to be split [120]. If the latter includes all reversible reactions, the reconfigured flux cone is pointed and generated by its extreme rays called *extremal currents* [20]. If only internal reversible reactions are split, the reconfigured flux cone is again pointed and the extreme rays are termed *extreme pathways* [90]. Note that if all boundary reactions are irreversible, both concepts are identical. We should also mention that the extremal current and the extreme pathway approach require a reconfiguration of the network even if the initial cone is pointed. Also in this case, the set of extreme rays of the reconfigured cone is much larger than that of the initial cone.

Schuster and Hilgetag [98; 99] have proposed a description of the flux cone without any reconfiguration, using *elementary modes (EMs)*. An elementary mode corresponds to a steady-state flux distribution involving a minimal set of reactions. This concept is related to that of a *minimal T-invariant* in Petri net theory [37; 100] and has also been used for analyzing signaling and regulatory networks [49]. From a biological viewpoint, each EM converts a set of metabolites into each other by means of a minimal set of reactions. Since reactions are catalyzed by enzymes, each EM corresponds to a minimal set of enzymes that must be expressed by genes. The simplicity property of EMs is of great interest because the effort provided by a biological system to maintain a metabolic pathway increases with the number of enzymes expressed [74]. In addition to the simplicity property, it has been shown that elementary modes span the steady-state flux cone. In other words, each steady-state flux distribution can be expressed as a non-negative linear combination of elementary modes. For a more detailed explanation of the similarities and differences between the three inner descriptions, we refer

Inner description	Split reactions	Characteristics of the reconfigured flux cone C'			Characteristics of ID	
ID	SR	Dimension	Number of constraints	$\mathrm{lin.space}(C')$	Uniqueness	Minimality
Extreme pathways	Rev_{int}	$n + \lvert Rev_{int}\rvert$	$m + \lvert Irr\rvert + 2\lvert Rev_{int}\rvert$	$\{0\}$	$\surd$	yes
Extremal currents	Rev	$n + \lvert Rev\rvert$	$m + \lvert Irr\rvert + 2\lvert Rev\rvert$	$\{0\}$	$\surd$	yes
Elementary modes	$\emptyset$	n	$m + \lvert Irr\rvert$	$\mathrm{lin.space}(C)$	$\surd$	no

Table 3.1: Inner descriptions of the flux cone, with the set of split reversible reactions SR, the characteristics of the reconfigured flux cone C' and of the three inner descriptions. Contrary to elementary modes, the sets of extreme pathways and extremal currents correspond to the extreme rays of their corresponding reconfigured flux cone and so are minimal. However, possibly many of these generating vectors could be in the interior of the original flux cone.

to [51; 73; 75; 76; 120]. Tab. 3.1 summarizes the characteristics of the different inner descriptions.

Elementary modes are defined in the original n-dimensional flux space. In contrast, to define extreme pathways (resp. extremal currents), the dimension of the flux space is increased by p (resp. q), the number of internal reversible reactions (resp. the number of all reversible reactions). In the following, we formally characterize the relationships between the three inner descriptions.

Let Rev_{int} stand for the set of reversible internal reactions. Suppose $Rev_{int} = \{j_1, \ldots, j_p\}$ and $Rev = \{j_1, \ldots, j_q\}$. Let $\pi : C \to \mathbb{R}^{n+p}$ (resp. $\phi : C \to \mathbb{R}^{n+q}$) be the function that maps each flux vector $v \in C$ to $v' = \pi(v)$ (resp. $v' = \phi(v)$) such that $v'_j = v_j$ for all $j \in \{1, \ldots, n\} \setminus Rev_{int}$ (resp. $j \in \{1, \ldots, n\} \setminus Rev$), and for each $k \in \{1, \ldots, p\}$ (resp. $k \in \{1, \ldots, q\}$)

$$\begin{array}{llll} v'_{j_k} = v_{j_k} & \text{and} & v'_{n+k} = 0 & \text{if } v_{j_k} \geq 0, \\ v'_{j_k} = 0 & \text{and} & v'_{n+k} = -v_{j_k} & \text{if } v_{j_k} < 0. \end{array}$$

The function π (resp. ϕ) formally defines the reconfiguration of a flux vector $v \in C$ by splitting each free variable v_{j_k} with $j_k \in Rev_{int}$ (resp. $j_k \in Rev$) into two non-negative variables v'_{j_k} and v'_{n+k} with $v_{j_k} = v'_{j_k} - v'_{n+k}$. This operation is similar to standard form transformation in linear programming. To define the reverse operation, let $\pi^r : \mathbb{R}^{n+p} \to C$ (resp. $\phi^r : \mathbb{R}^{n+q} \to C$) be the function that maps each vector $v' \in \mathbb{R}^{n+p}$ (resp. $v' \in \mathbb{R}^{n+q}$) to $v = \pi^r(v')$ (resp. $v = \phi^r(v')$) such that $v_j = v'_j$ for all $j \in \{1, \ldots, n\} \setminus Rev_{int}$ (resp. $j \in \{1, \ldots, n\} \setminus Rev$) and $v_{j_k} = v'_{j_k} - v'_{n+k}$ for all $k \in \{1, \ldots, p\}$ (resp. $k \in \{1, \ldots, q\}$).

Finally, let $\Pi \subseteq \mathbb{R}^{n+p}$ and $\Phi \subseteq \mathbb{R}^{n+q}$ be the sets of 2-cycles corresponding to the split reversible reactions, i.e.,

$$\begin{array}{l} \Pi = \{x \in \mathbb{R}^{n+p} \mid x_j = 0 \text{ for all } j \in \{1, \ldots, n+p\} \setminus \{j_k, n+k\} \\ \text{and } x_{j_k} = x_{n+k} = 1, \text{ for some } j_k \in Rev_{int}\}, \end{array}$$

$$\begin{array}{l} \Phi = \{x \in \mathbb{R}^{n+q} \mid x_j = 0 \text{ for all } j \in \{1, \ldots, n+q\} \setminus \{j_k, n+k\} \\ \text{and } x_{j_k} = x_{n+k} = 1, \text{ for some } j_k \in Rev\}. \end{array}$$

The following proposition reformulates the relationship between extreme pathways and elementary modes given in [51]. Except the 2-cycles corresponding to the split reactions, each extreme pathway completely defines a unique elementary mode.

Proposition 3.2 ([51])**.** *If $x \notin \Pi$ is an extreme pathway, then there exists a unique elementary mode $e \in C$ such that $x = \pi(e)$ and $e = \pi^r(x)$.*

According to the proposition above, the set of extreme pathways corresponds to a subset of elementary modes. Next, we restate the equivalence between elementary modes and extremal currents given in [34].

Proposition 3.3 ([34])**.** *Let $e \in C$ be a steady-state flux distribution. The following are equivalent:*

- *e is an elementary mode.*
- *There exists a unique extremal current $x \notin \Phi$ such that $x = \phi(e)$ and $e = \phi^r(x)$.*

It follows that up to the 2-cycles corresponding to the split reactions, extremal currents and elementary modes are equivalent. Accordingly, an algorithm for computing extremal currents could also be used to calculate elementary modes and vice versa.

Thus all three approaches are concerned with describing a pointed reconfigured flux cone C' by means of its extreme rays. There may exist many generating vectors of the reconfigured flux cone C' lying in the interior of the original flux cone C. This observation is important because the number of these generators may be very large, making a complete analysis of the whole metabolic network impossible and limiting the practical applicability of these methods.

3.5 Elementary Flux Modes

In this section, we limit ourselves to further characterize elementary modes. All the properties we will discuss here hold also for the extreme pathways and extremal currents.

The notion of an elementary vector was first introduced in [87] (see also [88], p. 205) where it was defined as a vector having a minimal support, i.e., a minimal set of non-zero components.

Definition 3.4 (Elementary vector [87])**.** *Let $L \subseteq \mathbb{R}^n$ be a vector subspace. A vector $e \in L \setminus \{0\}$ is an elementary vector if its support $Supp(e)$ is minimal, i.e., there exists no vector $e' \in L \setminus \{0\}$ such that $Supp(e') \subsetneq Supp(e)$.*

It has been shown that a subspace $L \subseteq \mathbb{R}^n$ can be considered as a linear hull of a finite set of elementary vectors [87] (see Sect. 2.1 for a definition of a linear hull). Building on earlier ideas [21; 30; 66], Schuster et al. introduced the concept of *elementary modes* to provide a finite set of vectors that span the flux cone by non-negative combination [97; 98; 99].

Definition 3.5 (Elementary mode [98])**.** *A flux distribution $e \in \mathbb{R}^n \setminus \{0\}$ represents an elementary (flux) mode if, and only if, e fulfils the two following conditions:*

- *Feasibility: e obeys the stoichiometric and thermodynamic constraints, i.e., $e \in C$.*
- *Simplicity (non-decomposability): e cannot be represented as a positive linear combination*

$$e = \lambda_1 e' + \lambda_2 e'' \text{ for some } \lambda_1, \lambda_2 > 0 \tag{3.9}$$

of two flux vectors $e', e'' \neq 0$ with the following properties:

1. *e' and e'' obey the stoichiometric and thermodynamic constraints, i.e.,*

$$e' \in C \text{ and } e'' \in C.$$

2. *e' and e'' contain zero elements wherever e does and they include at least one additional zero component each,*

$$Supp(e') \subsetneq Supp(e) \text{ and } Supp(e'') \subsetneq Supp(e). \tag{3.10}$$

In addition to the simplicity property, elementary modes span the steady-state flux cone. Each steady-state flux distribution can be expressed as a non-negative linear combination of elementary modes.

Proposition 3.6 ([99])**.** *Let $e^1, \ldots, e^p$ be the elementary modes of the flux cone C. Each possible flux distribution $v \in C$ is a non-negative linear combination of $e^1, \ldots, e^p$*

$$v = \sum_{k=1}^{p} \lambda_k e^k \text{ for some } \lambda_k \geq 0.$$

The next proposition restates an equivalent formulation of the simplicity property characterizing elementary modes. An EM is a flux distribution having a minimal support, i.e., a minimal set of active reactions (non-zero components).

Proposition 3.7 ([99])**.** *For any pair of vectors $e, e' \in C$, with e representing an elementary mode and e' having zero components wherever e has zero components, i.e.,*

$$Supp(e') \subseteq Supp(e), \tag{3.11}$$

e' either represents the same elementary mode as e or the same elementary mode as $-e$, which implies $Supp(e') = Supp(e)$.

By the proposition above, each EM can be defined by its set of active reactions, which is minimal. From a biological viewpoint, each EM converts certain metabolites into each other by means of a minimal set of reactions. Since reactions are catalyzed by enzymes, each EM corresponds to a minimal set of enzymes that must be expressed by genes. The simplicity property of EMs is of great interest because the effort provided

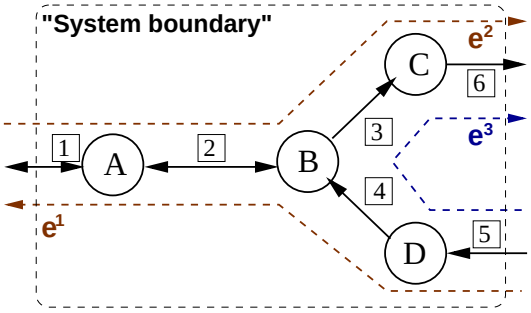


Figure 3.2: A hypothetic network with the corresponding elementary modes. This network contains three elementary modes e^1, e^2 and e^3. The flux cone is pointed and is spanned only by two extreme rays e^1 and e^2.

by a biological system to maintain a metabolic route increases with the number of enzymes expressed [74]. From a mathematical point of view, the simplicity property guarantees the uniqueness of the set of EMs. However, given an EM e, we could find two flux vectors e' and e'' in the flux cone that fulfill equation (3.9) but not condition (3.10). Accordingly, if the flux cone is pointed, an EM is not necessarily an extreme ray. On the other hand, an extreme ray cannot be decomposed into two other feasible flux vectors, and so each extreme ray defines an EM. This is the reason why, in many situations, more EMs exist than are needed to span the flux cone.

Example 3.8. *For illustration, consider the network depicted in Fig. 3.2. It consists of four internal metabolites* $(A, \ldots, D)$ *and six reactions* $(1, \ldots, 6)$*, where reactions* $3, 4, 5$ *and* 6 *are irreversible. The flux cone corresponding to this network is pointed, i.e., no steady-state flux distribution involving only reversible reactions is possible. Two extreme rays generate the flux cone. They correspond to the elementary modes* $e^1 = (-1, -1, 0, 1, 1, 0)$ *and* $e^2 = (1, 1, 1, 0, 0, 1)$ *depicted in Fig. 3.2. This network contains another elementary mode* $e^3 = (0, 0, 1, 1, 1, 1)$*, which can be written* $e^3 = e^1 + e^2$*. Although* e^3 *fulfills the simplicity condition, this mode does not correspond to an extreme ray of the flux cone.*

Elementary modes are also useful for studying reaction deletions, i.e., the removal of one or more reactions from the metabolic network. The following proposition restates the conservation property characterizing EMs.

Proposition 3.9 (Conservation property [51])**.** *If a set of reactions are removed from the metabolic network, all elementary modes not involving these reactions form the complete set of elementary modes in the altered network.*

Elementary-mode analysis has been used to investigate several features of metabolic networks. Tab. 3.2 shows the main applications of this approach.

Various algorithms have been developed for computing elementary modes. The main algorithm [99], which is based on an earlier algorithm by Nozicka [70], proceeds

Application	Reference
Investigation of viable pathways	[80]
Inference of mutant viability	[16; 106]
Detection of pathways with maximal molar yields	[54; 96; 97; 102]
Control-effective flux analysis	[16; 17; 106]
Identification of correlated reactions	[78]
Detection of thermodynamically infeasible cycles	[34]
Computation of minimal cut sets	[46]
Analysis of dynamical capabilities of a metabolic system	[108]

Table 3.2: Applications of elementary-mode analysis for elucidating network properties.

in an iterative way by computing the EMs for a series of cones $C^0, \ldots, C^m$ given by

$$\begin{aligned} C^0 &= \{v \in \mathbb{R}^n \mid v_j \geq 0 \text{ for all } j \in Irr\}, \\ C^i &= \{v \in \mathbb{R}^n \mid S_{I_i *} v = 0;\ v_j \geq 0 \text{ for all } j \in Irr\}, \\ &\quad \text{with } I_i = \{1, \ldots, i\}, \text{ for all } i = 1, \ldots, m. \end{aligned}$$

Obviously, the EMs of the cone C^0 corresponds to the vectors of the canonical basis of the euclidean space $\mathbb{R}^n$. These initial EMs fulfill the thermodynamic but not the stoichiometric constraints. At each iteration $i \in \{1, \ldots, m\}$, the EMs of the cone C^i are computed from those of the cone C^{i-1} in two steps:

1. The EMs of C^{i-1} fulfilling the stoichiometric constraint $S_{i*}v = 0$ are also EMs of the cone C^i.

2. The remaining EMs of C^{i-1}, which do not fulfill $S_i v = 0$, are combined with each other to compute the EMs of C^i that lie within the intersection between C^{i-1} and the hyperplane $H^i = \{v \in \mathbb{R}^n \mid S_{i*}v = 0\}$. However, many of these combinations do not lead to EMs of C^i and need to be discarded. There are two methods that allow for discarding these combinations. The first is based on the simplicity property of EMs. Indeed, a combination of a pair of two EMs of C^{i-1} does not lead to an EM of C^i if the set of non-zero components involved in that combination includes the set of non-zero components involved in an already computed EM of C^i. The second uses a simple criterion on the rank of some submatrix of the stoichiometric matrix [114].

Since the cone C^m and the flux cone C are identical, after iteration m, the algorithm terminates having computed all the EMs of the flux cone C. The most time consuming part in this algorithm are the computations needed to check the simplicity property in the second step of each iteration. Several variants of this algorithm have been developed to reduce the cost of these computations [34; 110], whereas others have attempted to reduce the number of iterations [114; 119]. While efficient in analyzing metabolic networks of small sizes, these algorithms are hampered by the combinatorial explosion of the number of EMs in genome-scale networks [50].

A natural question is that of defining methods which shrink the size of the flux cone description to a more manageable level. In particular, one might wish that the description of the flux cone is minimal. Finding a description which fulfills this requirement essentially amounts to determining a minimal set of generating vectors [78; 120]. However, this strategy is not ideal in the context of metabolic networks. These generating vectors may not be unique, and so it is difficult to attribute a biological interpretation to such a non unique description. Alternatively, some authors have suggested projecting the flux cone onto the subspace spanned by the boundary reactions. This allows to consider instead of the flux cone a simpler cone, called the *conversion cone* [114]. Several elementary modes are then equivalent with respect to the boundary reactions. They differ only in the active internal reactions. While the description of the conversion cone is much smaller than that of the flux cone, this approach abstracts away the operation of internal reactions.

CHAPTER

4 Minimal Metabolic Behaviors and the Reversible Metabolic Space

This chapter is devoted to providing a deeper insight into the mathematics underlying our new constraint-based approach to metabolic network analysis published in [58]. A key idea is to use an outer description of the steady-state flux cone, based on sets of non-negativity constraints. These can be identified with irreversible reactions and therefore have a direct biochemical interpretation. Our method is thus different from existing approaches, such as elementary modes or extreme pathways, which use an inner description. We characterize a metabolic network by two new concepts: *minimal metabolic behaviors* (MMBs) and the *reversible metabolic space* (RMS). Like elementary modes or extreme pathways, these are uniquely determined by the network. The set of all MMBs together with the RMS yields a complete description of the flux cone, which is minimal, unique, and satisfies a simplicity condition similar to the one that holds for elementary modes. Moreover, our approach leads to a new classification of reactions (irreversible, pseudo-irreversible, fully reversible), which may be used for a refined analysis of the network. We demonstrate the reliability of our new approach by studying the metabolic functions of the human red blood cell.

4.1 Minimal Metabolic Behaviors

In the context of metabolic pathway analysis, the set of all possible flux distributions over a metabolic network at steady state defines the steady-state flux cone (see Definition 3.1 in Chap. 3). Already in [20], we can find the distinction between inner and outer descriptions of this cone, which are called there internal and external representations. The external representation gives a test for determining whether a given flux vector belongs to the cone, while the internal representation allows one to construct flux vectors from a set of generators.

If the metabolic network does not contain any irreversible reaction, the steady-state flux cone becomes a linear subspace of $\mathbb{R}^n$, which can be analyzed by standard methods from linear algebra. Therefore, we assume for the rest of this chapter that the metabolic network contains at least one irreversible reaction.

4.1.1 Characterizing Minimal Proper Faces

We start by characterizing the minimal proper faces of the flux cone through irreversible reactions of the network.

Definition 4.1. *Let G be a minimal proper face of the flux cone C and let $j \in Irr$ be an irreversible reaction. We say that G is* characterized by j *if there exists $I_j \subseteq Irr$ such that $G = \{v \in C \mid v_j \geq 0,\ v_i = 0,\ \text{for all } i \in I_j\}$, and $\text{lin.space}(C) = \{v \in C \mid v_j = 0,\ v_i = 0,\ \text{for all } i \in I_j\}$.*

It follows from equation (2.2) in Chap. 2 that each minimal proper face G of C is characterized by at least one irreversible reaction. However, this reaction need not be unique. In general, there will be several irreversible reactions satisfying the conditions of Definition 4.1. The following proposition provides a simple criterion to identify the irreversible reactions that characterize a given minimal proper face. In particular, given a minimal proper face G, if an irreversible reaction j is involved in some flux vector $v \in G \setminus \text{lin.space}(C)$, i.e., $v_j > 0$, then j is involved in all flux vectors $v \in G \setminus \text{lin.space}(C)$ and so G is characterized by j.

Proposition 4.2. *Let G be a minimal proper face of C and let $j \in Irr$ be an irreversible reaction. Then the following statements are equivalent:*

1. *G is characterized by j.*
2. *$v_j > 0$, for some $v \in G \setminus \text{lin.space}(C)$.*
3. *$v_j > 0$, for all $v \in G \setminus \text{lin.space}(C)$.*

Proof. (1) $\Rightarrow$ (2): Since $\dim(G) = 1 + \dim(\text{lin.space}(C))$, we have $G \setminus \text{lin.space}(C) \neq \emptyset$. So there exists $v \in G \setminus \text{lin.space}(C)$, with $v_j > 0$.

(2) $\Rightarrow$ (3): Suppose $g \in G \setminus \text{lin.space}(C)$ with $g_j > 0$. By equation (2.4), for any $v \in G \setminus \text{lin.space}(C)$ there exists $\lambda > 0$ and $w \in \text{lin.space}(C)$ such that $v = \lambda \cdot g + w$. It follows that $v_j = \lambda \cdot g_j > 0$.

(3) $\Rightarrow$ (1): It follows from equation (2.2) that G is characterized by at least one irreversible reaction. So there exists $k \in Irr$ and $I_k \subset Irr$ with $G = \{v \in C \mid v_k \geq 0,\ v_i = 0,\ \text{for all } i \in I_k\}$, and $\text{lin.space}(C) = \{v \in C \mid v_k = 0,\ v_i = 0,\ \text{for all } i \in I_k\}$. To prove (1), we set $I_j = I_k$ and claim that the same equations hold for k replaced with j. Consider $v \in C$ with $v_i = 0$, for all $i \in I_j = I_k$. Since $v \in C$, we have $v_j \geq 0$ and $v_k \geq 0$. If $v_k > 0$, then $v \in G \setminus \text{lin.space}(C)$ and by (3) $v_j > 0$. If $v_k = 0$, then $v \in \text{lin.space}(C)$ and so $v_j = 0$. Together this shows $v_j = 0$ if and only if $v_k = 0$. It follows $G = \{v \in C \mid v_k \geq 0, v_i = 0, i \in I_k\} = \{v \in C \mid v_j \geq 0, v_i = 0, i \in I_j\}$ and $\text{lin.space}(C) = \{v \in C \mid v_k = 0, v_i = 0, i \in I_k\} = \{v \in C \mid v_j = 0, v_i = 0, i \in I_j\}$. □

Now, we define the *characteristic set* of a minimal proper face G as the set of all irreversible reactions characterizing G. As a consequence of Proposition 4.2, this set is equal to the set of irreversible reactions involved in some flux vectors in G. Let us state this fact in the following definition.

Definition 4.3. *Given a minimal proper face G of the flux cone C, the set*

$$D = \{j \in Irr \mid v_j > 0, \text{ for some } v \in G\}$$

of all irreversible reactions characterizing G is called the characteristic set *of G.*

Note that the characteristic set is uniquely determined by the corresponding minimal proper face.

As the next theorem shows, all flux vectors $v \in G \setminus \mathrm{lin.space}(C)$ have the following common property: the flux through all irreversible reactions belonging to D is positive, i.e., $v_j > 0$, for $j \in D$, while the flux through all the other irreversible reactions is zero, i.e., $v_j = 0$, for $j \in Irr \setminus D$.

Since $\mathrm{lin.space}(C) = \{v \in C \mid v_i = 0 \text{ for all } i \in Irr\}$, flux vectors in $\mathrm{lin.space}(C)$ involve only reversible reactions and so $-v \in \mathrm{lin.space}(C)$. For this reason, a flux vector $v \in \mathrm{lin.space}(C)$ will be called reversible. Note that information about reversible pathways is lost if the network is reconfigured in order to obtain a pointed cone.

Theorem 4.4. *Let G be a minimal proper face of the flux cone C and D its characteristic set. Then*

$$G = \{v \in C \mid v_j > 0, \text{ for all } j \in D,\ v_i = 0, \text{ for all } i \in Irr \setminus D\} \cup \mathrm{lin.space}(C).$$

Proof. Suppose $j \in D$. Then $G = \{v \in C \mid v_j \geq 0,\ v_i = 0, \text{ for } i \in I_j\}$ and $\mathrm{lin.space}(C) = \{v \in C \mid v_j = 0,\ v_i = 0, \text{ for } i \in I_j\}$, for some $I_j \subset Irr$. From Proposition 4.2, we see that $I_j \subset Irr \setminus D$. It follows that $\{v \in C \mid v_j > 0, \text{ for all } j \in D,\ v_i = 0, \text{ for all } i \in Irr \setminus D\} \cup \mathrm{lin.space}(C) \subseteq G$. To show the reverse inclusion, suppose $v \in G \setminus \mathrm{lin.space}(C)$. With Proposition 4.2, $v_j > 0$, for all $j \in D$. Suppose $v_i > 0$, for some $i \in Irr \setminus D$. From Definition 4.3, we would get $i \in D$, which is a contradiction.

□

If $G^1, \ldots, G^s$ are the minimal proper faces of the flux cone C, the corresponding characteristic sets $D^1, \ldots, D^s$ together with the lineality space $\mathrm{lin.space}(C)$ completely describe C.

The next result shows that inside a minimal proper face G, the fluxes through the irreversible reactions in D are proportional to each other.

Corollary 4.5. *Let D be the characteristic set of the minimal proper face G. Then for all $j, k \in D$, there exists $\alpha > 0$ such that $v_k = \alpha \cdot v_j$, for all $v \in G$. In particular, $v_j = 0$ implies $v_k = 0$, and $v_j > 0$ implies $v_k > 0$, for all $v \in G$.*

Proof. Consider $g \in G \setminus \mathrm{lin.space}(C)$. Since $j, k \in D$, Proposition 4.2 implies $g_j > 0$ and $g_k > 0$. By equation (2.4), for all $v \in G \setminus \mathrm{lin.space}(C)$, there exist $\lambda > 0$ and $w \in \mathrm{lin.space}(C)$ such that $v = \lambda \cdot g + w$. It follows that $v_j = \lambda \cdot g_j > 0, v_k = \lambda \cdot g_k > 0$, and therefore $v_j / v_k = g_j / g_k \stackrel{\text{def}}{=} \alpha > 0$, independently from v. This shows that $v_j = \alpha \cdot v_k > 0$, for all $v \in G \setminus \mathrm{lin.space}(C)$. For all $v \in \mathrm{lin.space}(C)$, we have $v_j = v_k = 0$. It follows for all $v \in G$ that $v_j = \alpha \cdot v_k$. □

4.1.2 Minimal Metabolic Behaviors and the Reversible Metabolic Space

We are now ready to define the key notions of this chapter.

Definition 4.6. *A* metabolic behavior *is a set of irreversible reactions $D \subseteq Irr, D \neq \emptyset$, such that there exists a flux vector $v \in C$ with*

$$D = \{i \in Irr \mid v_i \neq 0\}. \tag{4.1}$$

A metabolic behavior D is minimal*, if there is no metabolic behavior D' strictly contained in D. The set*

$$\{v \in C \mid v_i = 0, \text{ for all } i \in Irr\} \tag{4.2}$$

is called the reversible metabolic space*.*

Remember that elementary modes correspond to flux vectors $v \in C$ involving a minimum set of reactions, i.e., the set $Supp(v) = \{i \in Rev \cup Irr \mid v_i \neq 0\}$ is minimal [99]. Similarly, a minimal metabolic behavior corresponds to a minimal set of *irreversible* reactions involved in a flux vector $v \in C \setminus \{0\}$, i.e., the set $D = \{i \in Irr \mid v_i \neq 0\}$ is minimal.

One may ask whether a set of irreversible reactions $D \subseteq Irr$ is a metabolic behavior (MB). This is the same as asking whether there exists a steady-state flux distribution $v \in C$ with $D = \{i \in Irr \mid v_i \neq 0\}$. The existence of such a flux distribution could be verified using linear programming (LP). Indeed, consider the following LP problem

$$\begin{array}{l} \max \alpha \\ \text{subject to:} \\ Sv = 0, \\ v_i = 0 \text{ for all } i \in Irr \setminus D, \\ 0 \leq \alpha \leq v_i \leq 1 \text{ for all } i \in D. \end{array} \tag{4.3}$$

The set D is a metabolic behavior if and only if the optimal value of the LP problem above is strictly positive. If this is the case, one might wonder whether this metabolic behavior D is minimal (MMB). To deal with this question, we use the following statement, which is a straightforward consequence of the definition of a minimal metabolic behavior.

Proposition 4.7. *A set $D \subseteq Irr, D \neq \emptyset$, is a minimal metabolic behavior if and only if the following two conditions hold:*

1. *There exists $v \in C$ with $v_i > 0$, for all $i \in D$, and $v_i = 0$, for all $i \in Irr \setminus D$.*
2. *For any $v \in C$ with $v_i = 0$ for all $i \in Irr \setminus D$, if $v_j = 0$ for some $j \in D$, then $v_j = 0$ for all $j \in D$.*

Proof. "⇒": Suppose $D \subseteq Irr$ is a minimal metabolic behavior. Since D is a metabolic behavior, there exists $v \in C$ such that $D = \{i \in Irr \mid v_i \neq 0\}$. Since $v \in C$, we get $v_i > 0$, for all $i \in D$. By definition of D, $v_i = 0$, for all $i \in Irr \setminus D$. This shows (1). Now let $v \in C$ with $v_i = 0$ for all $i \in Irr \setminus D$. Suppose $v_j = 0$ for some $j \in D$ and $v_k \neq 0$, for some $k \in D, k \neq j$. Then $D' = \{i \in Irr \mid v_i \neq 0\}$ is a metabolic behavior strictly contained in D, contradicting the minimality of D.

"⇐": Consider $\emptyset \neq D \subseteq Irr$ such that (1) and (2) hold. By (1), there exists $v \in C$ with $v_i > 0$, for all $i \in D$, and $v_i = 0$, for all $i \in Irr \setminus D$. Then $D = \{i \in Irr \mid v_i \neq 0\}$ and so D is a metabolic behavior. Suppose D is not minimal. Then there exists a metabolic behavior $\emptyset \neq D' \subsetneq D$ strictly contained in D. Let $D' = \{i \in Irr \mid v'_i \neq 0\}$, for a suitable $v' \in C$. Then $v'_i = 0$, for $i \in Irr \setminus D' \supsetneq Irr \setminus D$. Since $D' \subsetneq D$, there exists $j \in D$ with $v'_j = 0$. From (2) we get $v'_j = 0$, for all $j \in D$. This is a contradiction, since $D' \neq \emptyset$ implies that there exists $i \in D' \subseteq D$ with $v'_i \neq 0$. □

Given a metabolic behavior D, the above proposition states that D is minimal if and only if for any $j \in D$ and any $v \in C \cap \{v \in \mathbb{R}^n \mid v_j = 0 \text{ and } v_i = 0 \text{ for all } i \in Irr \setminus D\}$, we have $\sum_{i \in D \setminus \{j\}} v_i = 0$, or equivalently, the optimal value of the following LP problem

$$\begin{array}{l} \max \ \sum_{i \in D \setminus \{j\}} v_i \\ \text{subject to:} \\ Sv = 0, \\ v_i = 0 \text{ for all } i \in Irr \setminus D, \\ v_i \geq 0 \text{ for all } i \in D \setminus \{j\}, \\ v_j = 0, \\ \sum_{i \in D \setminus \{j\}} v_i \leq 1. \end{array} \tag{4.4}$$

is equal to zero. Accordingly, we need to solve at most $|D|$ LP problems to check the minimality of a metabolic behavior D.

The following theorem shows that the MMBs are in a 1-1 correspondence with the minimal proper faces of the flux cone. Indeed, each minimal metabolic behavior is identical to the characteristic set of a minimal proper face.

Theorem 4.8. *Let $D \subseteq Irr$ be a set of irreversible reactions. Then, the following two statements are equivalent:*

- *D is a minimal metabolic behavior.*
- *There exists a minimal proper face G whose characteristic set is D.*

Proof. "⇒": Suppose D is a minimal metabolic behavior and let $G = \{v \in C \mid v_j \geq 0, \text{ for all } j \in D, \ v_i = 0, \text{ for all } i \in Irr \setminus D\}$. Since $G = \{v \in C \mid v_i = 0, \text{ for all } i \in Irr \setminus D\}$, G is a face of C (cf. [92], p. 101). Let $G' \subseteq G$ be a minimal proper face of C and D' its characteristic set. Since $G' \subseteq G$, we get $D' \subseteq D$. Suppose there exists $k \in D \setminus D'$. Then $v_k = 0$ for all $v \in G'$ and $G' \subseteq G \cap \{v \in \mathbb{R}^n \mid v_k = 0\}$. Since D is minimal, $v \in G$ and $v_k = 0$ implies $v_j = 0$, for all $j \in D$, and therefore $G \cap \{v \in \mathbb{R}^n \mid v_k = 0\} = \text{lin.space}(C)$. It follows that $G' \subseteq \text{lin.space}(C)$, in

contradiction to the assumption that G' is a minimal proper face. We conclude $D = D'$. Applying Theorem 4.4 to G', we get $G = G'$ and G is a minimal proper face of C.

"$\Leftarrow$": Let G be a minimal proper face with characteristic set D. We use Proposition 4.7 to show that D is a minimal metabolic behavior. If $v \in G \setminus \mathrm{lin.space}(C)$, then by Theorem 4.4, we get $v_i > 0$ for all $i \in D$ and $v_i = 0$ for all $i \in \mathit{Irr} \setminus D$, i.e., v satisfies condition (1) of Proposition 4.7. To check condition (2), let $v \in C$ such that $v_i = 0$ for all $i \in \mathit{Irr} \setminus D$. From Theorem 4.4, we get $v \in G$. If $v_j = 0$ for some $j \in D$, Corollary 4.5 yields $v_k = 0$ for all $k \in D$. □

The next proposition provides an algebraic characterization of a minimal metabolic behavior.

Proposition 4.9. *Let $D \subseteq \mathit{Irr}$ be a set of irreversible reactions. Then, the following two statements are equivalent:*

- *D is a minimal metabolic behavior.*
- $|D| = \mathrm{rank}(S_{*D\cup \mathit{Rev}}) - \mathrm{rank}(S_{*\mathit{Rev}}) + 1.$

Proof. Let $D \subseteq \mathit{Irr}$ be a set of irreversible reactions and let G be the face defined by

$$G = \{v \in C \mid v_i = 0, \text{ for all } i \in \mathit{Irr} \setminus D\}.$$

According to Theorem 4.8, D is a minimal metabolic behavior if and only if G is a minimal proper face, or equivalently,

$$\dim(\mathrm{lin}(G)) = \dim(\mathrm{lin.space}(C)) + 1. \tag{4.5}$$

Let $I \in \mathbb{R}^{n\times n}$ be the identity matrix and let $\overline{D} = \mathit{Irr} \setminus D$. Since $\mathrm{lin}(G)$ and $\mathrm{lin.space}(C)$ are the null spaces of the matrices

$$\begin{pmatrix} S \\ I_{\overline{D}*} \end{pmatrix} \text{ and } \begin{pmatrix} S \\ I_{\mathit{Irr}*} \end{pmatrix},$$

respectively, statement (4.5) is equivalent to the assertion

$$\mathrm{rank}\left(\begin{pmatrix} S \\ I_{\overline{D}*} \end{pmatrix}\right) = \mathrm{rank}\left(\begin{pmatrix} S \\ I_{\mathit{Irr}*} \end{pmatrix}\right) - 1. \tag{4.6}$$

Using row operations, we get the following two equations

$$\mathrm{rank}\left(\begin{pmatrix} S \\ I_{\overline{D}*} \end{pmatrix}\right) = |\overline{D}| + \mathrm{rank}(S_{*D\cup \mathit{Rev}}), \tag{4.7}$$

$$\mathrm{rank}\left(\begin{pmatrix} S \\ I_{\mathit{Irr}*} \end{pmatrix}\right) = |\mathit{Irr}| + \mathrm{rank}(S_{*\mathit{Rev}}). \tag{4.8}$$

Combining equations (4.6), (4.7) and (4.8), we obtain

$$|D| = \mathrm{rank}(S_{*D\cup \mathit{Rev}}) - \mathrm{rank}(S_{*\mathit{Rev}}) + 1. \tag{4.9}$$

□

Note that if t is the dimension of the lineality space of the flux cone C, we have

$$\mathrm{rank}\left(\begin{array}{c} S \\ I_{Irr*} \end{array}\right) = n - t,\ n = |Rev| + |Irr|,$$

and so $\mathrm{rank}(S_{*Rev}) = |Rev| - t$. Accordingly, if a set $D \subseteq Irr$ is a minimal metabolic behavior, we have

$$|D| = \mathrm{rank}(S_{*D \cup Rev}) - |Rev| + t + 1. \tag{4.10}$$

In addition, we know that

$$\mathrm{rank}(S_{*D \cup Rev}) \leq \mathrm{rank}(S). \tag{4.11}$$

Using statements (4.10) and (4.11), we obtain

$$|D| \leq \mathrm{rank}(S) - |Rev| + t + 1. \tag{4.12}$$

Corollary 4.10. *Let $D \subseteq Irr$ be a set of irreversible reactions and let $t = \dim(\mathrm{lin.space}(C))$. D can be a minimal metabolic behavior only if*

$$|D| \leq \mathrm{rank}(S) - |Rev| + t + 1.$$

The corollary above defines an upper bound on the cardinality of all minimal metabolic behaviors. This upper bound is particularly convenient for testing whether a set of irreversible reactions $D \subseteq Irr$ is a candidate to be a minimal metabolic behavior. More concisely, this test could be done in two steps. First, we check whether D satisfies inequality (4.12). If this is not the case, D is not an MMB. If D fulfils inequality (4.12), we then use a second test based on equation (4.10) or the linear programs (4.3) and (4.4).

According to Theorem 4.8, each MMB completely defines its corresponding minimal proper face and vice versa. This important feature guarantees the minimality property of the set of irreversible reactions defining an MMB as well as the uniqueness of the set of MMBs. Moreover, Theorem 4.8 states that the MMBs are in a 1-1 correspondence with the minimal proper faces of the flux cone. Therefore, the set of MMBs is minimal in the sense that no strict subset of MMBs could completely describe the flux cone. Hence, there are two minimality properties that hold for minimal metabolic behaviors: the minimality of each MMB and the minimality of the set of MMBs.

Example 4.11. *In the network ILLUSNET from Fig. 4.1, the MMBs and the corresponding minimal proper faces are as follows:*

$$D^1 = \{2\}, \quad D^2 = \{6, 7\}, \quad D^3 = \{6, 8\},$$
$$G^k = \{v \in C \mid v_j \geq 0, j \in D^k,\ v_i = 0, i \in Irr \setminus D^k\}, \quad k = 1, 2, 3.$$

Note that the irreversible reaction 6 is participating in the definition of two minimal proper faces, G^2 and G^3. Fig. 4.1 shows three pathways

$$\begin{aligned} g^1 &= (1,\ 1,\ 2,\quad 1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0), \\ g^2 &= (0,\ 0,\ 0,\ -1,\ 0,\ 1,\ 1,\ 0,\ 0,\ 0,\ 0,\ 0), \\ g^3 &= (0,\ 0,\ 0,\ -3,\ 0,\ 3,\ 0,\ 1,\ 1,\ 1,\ 0,\ 0), \end{aligned}$$

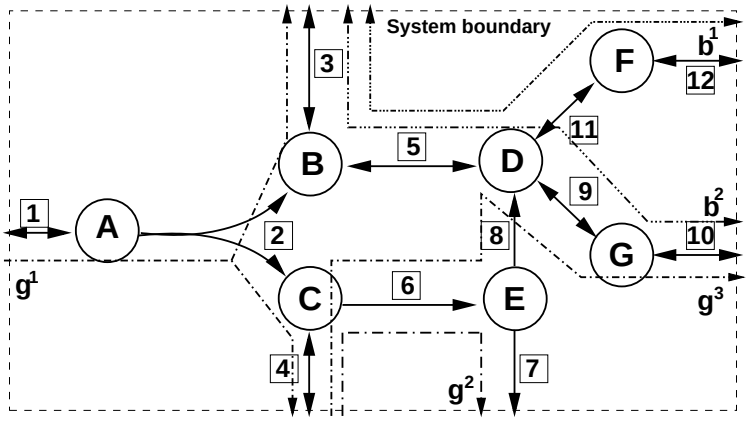


Figure 4.1: Representative pathways in ILLUSNET

representing the minimal proper faces G^1, G^2, and G^3, respectively. The reversible metabolic space $\mathrm{lin.space}(C) = \{v \in C \mid v_i = 0, i \in Irr\}$ *has dimension 2. As a vector space, it can be generated by the pathways*

$$\begin{aligned} b^1 &= (0,\ 0,\ -2,\ 0,\ 1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 1,\ 1), \\ b^2 &= (0,\ 0,\ -2,\ 0,\ 1,\ 0,\ 0,\ 0,\ 1,\ 1,\ 0,\ 0). \end{aligned}$$

An arbitrary flux vector $v \in C$ *can be written as linear combination* $v = \sum_{k=1}^{3} \lambda_k g^k + \sum_{l=1}^{2} \mu_l b^l$, *with* $\lambda_k \geq 0$ *and* $\mu_l \in \mathbb{R}$.

In this example, the number of elementary modes is 18, while the number of extreme pathways (after reconfiguration) is 14.

The irreversible reactions defining an MMB D cannot necessarily operate on their own. However, for each minimal metabolic behavior D, there exists at least one elementary mode involving exactly the irreversible reactions from D. For $v \in C$, let $D(v) = \{i \in Irr \mid v_i \neq 0\}$.

Proposition 4.12. *Let D be a minimal metabolic behavior. Then there exists an elementary mode f such that $D(f) = D$.*

Proof. Let D be a minimal metabolic behavior. According to Theorem 4.8, there exists a minimal proper face G whose characteristic set is D. Suppose $g \in G \setminus \mathrm{lin.space}(C)$. According to Proposition 3.6 in Chap. 3, $g = \sum_k \lambda_k f^k$ is a linear combination of elementary modes f^k, for some $\lambda_k \geq 0$. Since $g \neq 0$, there exists at least one elementary mode f^l such that $\lambda_l > 0$. For each $i \in D(f^l)$, we have $g_i = \sum_k \lambda_k f_i^k \geq \lambda_l f_i^l > 0$ and so $i \in D(g)$. This shows $D(f^l) \subseteq D(g)$. Since $g \in G$, it follows from Theorem 4.4 that $D(g) = D$. Finally, since D is a minimal metabolic behavior and $D(f^l) \subseteq D(g) = D$, we get $D(f^l) = D$ and the result follows. □

In general, there can be more than one elementary mode f with $D(f) = D$. If this is the case, there are different elementary modes (possibly many) that all belong to the same minimal proper face. In addition, there may exist elementary modes lying in the interior of the flux cone C. We refer to Sect. 4.3 for computational results illustrating these remarks.

4.2 Pseudo-irreversible and Fully Reversible Reactions

In this section, we classify reactions according to their reversibility type. We obtain a unique sign pattern for each MMB. These sign patterns can be used to decompose the network into minimal functional subnetworks.

4.2.1 Classification of Reactions

We start by distinguishing two classes of reversible reactions.

Definition 4.13. *Given the flux cone C, the set*

$$Prev_0 = \{i \in Rev \mid v_i = 0, \text{ for all } v \in \text{lin.space}(C)\}$$

is called the set of pseudo-irreversible reactions*. Reactions in $Frev = Rev \setminus Prev_0$ are called* fully reversible*.*

Example 4.14. *In the ILLUSNET network from Fig. 4.1, there is no reversible flux distribution involving reaction 4. We have $v_4 = 0$, for all $v \in \text{lin.space}(C)$. Therefore, reaction 4 is pseudo-irreversible. On the other hand, reaction 3 is involved in the reversible flux distribution b^1 which belongs to the lineality space $\text{lin.space}(C)$. Thus, reaction 3 is fully reversible.*

The next proposition shows that pseudo-irreversible reactions become irreversible inside minimal proper faces. Within each minimal proper face G, any pseudo-irreversible reaction with non-zero flux will take a unique direction, which is imposed by the MMB D associated with G. By taking the conical hull of the corresponding faces, we can identify a subspace of the cone in which the given pseudo-irreversible reaction takes only one direction.

Proposition 4.15. *Let G be a minimal proper face of C and let $i \in Prev_0$ be a pseudo-irreversible reaction. Then exactly one of the following three conditions holds:*

1. *$v_i > 0$, for all $v \in G \setminus \text{lin.space}(C)$.*
2. *$v_i = 0$, for all $v \in G \setminus \text{lin.space}(C)$.*
3. *$v_i < 0$, for all $v \in G \setminus \text{lin.space}(C)$.*

Proof. Suppose $g \in G \setminus \text{lin.space}(C)$. For any $v \in G \setminus \text{lin.space}(C)$ there exists $\lambda > 0$ and $w \in \text{lin.space}(C)$ such that $v = \lambda \cdot g + w$. Since $i \in Prev_0$, it follows that $sign(v_i) = sign(\lambda \cdot g_i) = sign(g_i)$, independently from v. □

Let G be a minimal proper face whose characteristic set is D. Now, we define the *auxiliary set* of D as the set of all pseudo-irreversible reactions that are involved in all flux vectors in G. As a consequence of Proposition 4.15, this set is equal to the set of pseudo-irreversible reactions involved in some flux vectors in G. Let us state this fact in the following definition.

Definition 4.16. *Let G be a minimal proper face whose characteristic set is D. The set*

$$Aux = \{j \in Rev \mid j \text{ is pseudo-irreversible and } v_j \neq 0, \text{ for some } v \in G\}$$

of all pseudo-irreversible reactions involved in some flux vectors in G is called the auxiliary set *of G (and of D).*

Example 4.17. *In the ILLUSNET network, we have $Prev_0 = \{1, 4\}$, and $Frev = \{3, 5, 9, 10, 11, 12\}$. In the context of the MMB D^1, the pseudo-irreversible reaction 4 becomes positive, i.e., $v_4 > 0$, for all $v \in G^1 \setminus \text{lin.space}(C)$, while it becomes negative in the context of D^2 and D^3. The flux through the pseudo-irreversible reaction 1 is positive in D^1, and zero in D^2 and D^3. The auxiliary sets of D^1, D^2 and D^3 are the sets $\{1, 4\}$, $\{4\}$ and $\emptyset$, respectively.*

The next result shows that inside a minimal proper face G, the fluxes through the pseudo-irreversible reactions involved in flux vectors $v \in G \setminus \text{lin.space}(C)$ are proportional to each other.

Corollary 4.18. *Let G be a minimal proper face and Aux its auxiliary set. Let $j, k \in Prev_0$ be two pseudo-irreversible reactions. If $j, k \in Aux$, then there exists $\alpha \neq 0$ with $v_k = \alpha \cdot v_j$, for all $v \in G$.*

Proof. Suppose there exist $g, g' \in G \setminus \text{lin.space}(C)$ such that $g_j \neq 0$ and $g'_k \neq 0$. According to Proposition 4.15, we have $g_k \neq 0$. By equation (2.4), for all $v \in G \setminus \text{lin.space}(C)$, there exist $\lambda > 0$ and $w \in \text{lin.space}(C)$ such that $v = \lambda \cdot g + w$. Since $j, k \in Prev_0$, we get $v_j = \lambda \cdot g_j \neq 0, v_k = \lambda \cdot g_k \neq 0$, and therefore $v_k/v_j = g_k/g_j \stackrel{\text{def}}{=} \alpha \neq 0$, independently from v. This shows that $v_k = \alpha \cdot v_j \neq 0$, for all $v \in G \setminus \text{lin.space}(C)$. For all $v \in \text{lin.space}(C)$, we have $v_j = v_k = 0$. It follows for all $v \in G$ that $v_k = \alpha \cdot v_j$. □

Traditionally, there are two classes of reactions in a metabolic network: reversible and irreversible ones. Following our analysis, we may refine this classification and distinguish three types of reactions:

- Irreversible reactions $j \in Irr$: for all minimal proper faces G, we have either $v_j > 0$, for all $v \in G \setminus \text{lin.space}(C)$, or $v_j = 0$, for all $v \in G$. By definition, $w_j = 0$, for all $w \in \text{lin.space}(C)$.

- Pseudo-irreversible reactions $j \in Prev_0$: inside each minimal proper face, the flux v_j through j has a unique sign ($+, -$, or 0). For all $w \in \text{lin.space}(C)$, we have again $w_j = 0$.

- Fully reversible reactions $j \in Frev$: by definition, there exists $w \in \text{lin.space}(C)$ such that $w_j \neq 0$. This implies that we can find in each minimal proper face G flux vectors $v, v', v'' \in G \setminus \text{lin.space}(C)$ with $v_j > 0$, $v'_j < 0$ and $v''_j = 0$.

Altogether, this means that each MMB D can be characterized by a unique sign pattern P_D for the (pseudo-) irreversible reactions in the network ('+', '-', or '0'), while the flux through the fully reversible reactions may be arbitrary ('·').

Example 4.19. *In the network ILLUSNET, where* $Irr \cup Prev_0 = \{1, 2, 4, 6, 7, 8\}$*, the sign patterns of the MMBs* D^1, D^2, D^3 *are the following:*

$$\begin{aligned} P^1 &= (+,\ +,\ \cdot,\ +,\ \cdot,\ 0,\ 0,\ 0,\ \cdot,\ \cdot,\ \cdot,\ \cdot), \\ P^2 &= (0,\ 0,\ \cdot,\ -,\ \cdot,\ +,\ +,\ 0,\ \cdot,\ \cdot,\ \cdot,\ \cdot), \\ P^3 &= (0,\ 0,\ \cdot,\ -,\ \cdot,\ +,\ 0,\ +,\ \cdot,\ \cdot,\ \cdot,\ \cdot). \end{aligned}$$

4.2.2 Decomposing the Network

Minimal metabolic behaviors can also be used to decompose a given metabolic network $\mathcal{N}$ into different subnetworks. Indeed, given the sign pattern P_D of an MMB D, the set of fully reversible reactions together with the (pseudo-) irreversible reactions having a non-zero sign in P_D defines a subnetwork $\mathcal{N}_D$ of $\mathcal{N}$ with the following properties:

- The set of possible flux distributions in $\mathcal{N}_D$ includes the reversible metabolic space RMS.
- All pseudo-irreversible reactions become irreversible inside $\mathcal{N}_D$, i.e., each pseudo-irreversible reaction operates only in one direction, which is dictated by the MMB D and given in the sign pattern P_D.
- $\mathcal{N}_D$ is *minimal* in the sense that it includes all fully reversible reactions and a non-empty minimal set of (pseudo-) irreversible reactions that are capable of carrying flux under steady-state conditions.
- $\mathcal{N}_D$ is *functional* in the sense that the set of possible flux distributions over $\mathcal{N}_D$ includes at least one irreversible elementary mode. The latter exists because, according to Proposition 4.12, there is at least one elementary mode involving exactly the irreversible reactions from D.

Using the sign patterns of all the different MMBs, each corresponding to one particular minimal proper face of the flux cone, the overall metabolic network may be understood as a combination of these minimal functional subnetworks. Indeed, each possible flux distribution over the full network is a non-negative combination of possible flux distributions over the corresponding minimal functional subnetworks.

4.3 Computational Results

In this section, we discuss how one can compute minimal metabolic behaviors, and present a number of computational results.

4.3.1 Computing Minimal Metabolic Behaviors

A simple algorithm to determine the MMBs of a metabolic network is as follows. First we compute a set of generators of the flux cone C, using existing software for polyhedral computations such as `cdd` [3; 33]. Note that we do not have to reconfigure the cone by splitting reversible reactions as this is done in the extreme pathway approach. If the cone C is pointed, this will be detected automatically during computation. The number of MMBs of C (which is equal to the number of extreme rays of C if C is pointed) is typically much smaller than the number of extreme rays of the reconfigured cone. Second, for each minimal proper face G of C, represented by some generator $g \in G \setminus \mathrm{lin.space}(C)$, we identify the set D of irreversible reactions $j \in Irr$, with $g_j > 0$. Another possible approach is to apply the Fourier-Motzkin algorithm [92] to eliminate the variables corresponding to the internal reversible reactions. This results in a constraint system where all internal reactions are irreversible.

If we are interested in a minimal set of generators for the flux cone C, we have to choose for each minimal proper face G^k a vector $g^k \in G^k \setminus \mathrm{lin.space}(C)$, together with a generating set $\{b^0, \ldots, b^t\}$ of $\mathrm{lin.space}(C)$. If we decompose $g^k = (g^k_{Irr}, g^k_{Prev_0}, g^k_{Frev})$ into components corresponding to irreversible, pseudo-irreversible, and fully reversible reactions, then the components in $(g^k_{Irr}, g^k_{Prev_0})$ are uniquely determined up to multiplication by positive scalars. There remains some freedom in the choice of the components in g^k_{Frev}. Choosing $g^k \in \mathrm{lin.space}(C)^{\perp}$, i.e., $g^k b^T = 0$ for all $b \in \mathrm{lin.space}(C)$, yields the *orthogonal representation* of the flux cone C, which is unique but often very dense. Alternatively, work in the context of the software `cdd` [3; 79] discusses how to obtain a sparser representation of C, which is called a *lexico-smallest* representation of C. In such a representation, the generators are with a maximum number of zeroes and correspond to a subset of elementary modes.

4.3.2 Comparison with Existing Approaches

We now compare the different approaches on some example networks taken from the KEGG pathway database (`http://www.genome.ad.jp/kegg/pathway.html`). We suppose for these models that there is an unconstrained exchange flux for each metabolite that is not consumed or not produced by some internal reaction in the network. The computation of the extreme pathways, the minimal metabolic behaviors and the reversible metabolic space was done using the software `cdd` [33]. For computing the elementary modes, we used METATOOL [117]. The results are given in Tab. 4.1 and Tab. 4.2.

Tab. 4.1 shows the number of internal metabolites in the network, the number of irreversible/reversible internal reactions, the number of elementary modes/extreme pathways/MMBs, and the dimension of the RMS. We can see that the size of our representation, given as the sum of the number of MMBs and $\dim$(RMS), is typically much smaller than the number of extreme pathways or elementary modes. In various examples, the reduction is by several orders of magnitude.

Tab. 4.2 describes the distribution of the elementary modes and the extreme path-

Metabolic network	Met	Irr	Rev	EM	EP	MMB	RMS
Glycolysis / Gluconeogenesis	32	18	29	19464	1745	16	13
Citrate cycle (TCA cycle)	22	4	25	3870	1588	4	12
Pentose phosphate pathway	34	19	24	5155	1630	19	8
Pentose and glucuronate	50	13	46	2258	231	7	23
Fructose and mannose	46	37	31	2411	2102	30	6
Galactose	41	22	28	623	524	13	9
Starch and sucrose	47	35	30	2097	1718	30	5
Pyruvate	28	40	29	47708	27390	37	16
Propanoate	34	20	29	877	233	17	13
Butanoate	40	23	30	2138	541	18	11
Nitrogen	41	53	14	601	612	44	9
Sulfur	18	26	4	321	326	28	1

Table 4.1: Metabolic networks, with the number of internal metabolites (Met), the number of irreversible (Irr) and reversible (Rev) internal reactions, the number of elementary modes (EM), extreme pathways (EP), minimal metabolic behaviors (MMB), and the dimension of the reversible metabolic space (RMS). Contrary to the calculation of EMs, the calculation of EPs required a reconfiguration of the network. Except for the two-cycle extreme pathways made from a forward and a backward reaction, the set of EPs is always a subset of the set of EMs [51].

ways inside the steady-state flux cone. We can see that a very large number of elementary modes and extreme pathways lie in the interior of the cone. In addition, the number of elementary modes/extreme pathways belonging to the minimal proper faces (see column MMB) is much larger than the number of MMBs in Tab. 4.1. This means that many elementary modes/extreme pathways belong to the same minimal proper face, which mathematically can be represented by a single vector, resp. one MMB. Similarly, the number of elementary modes/extreme pathways belonging to the reversible metabolic space is much larger than its dimension, so that there are many dependencies.

4.4 On the Complexity of the MMB&RMS Approach

While the dimension of the lineality space of a metabolic network is smaller than or equal to the number of reversible reactions, the number of MMBs is, in the worst case, exponential in the number of reactions. This is particularly the case when all the reactions in the metabolic network are irreversible and the number of elementary modes (EMs) is very large. In such a case, MMBs are in a 1-1 correspondence with EMs and so their number is very large as well.

Consider the hypothetic network depicted in Fig. 4.2. This network contains $m = 2p + 1$ metabolites and $n = 3p + 2$ reactions for some $p \geq 1$. Each metabolite A_i, with $i \in \{2, \ldots, p + 1\}$, is obtained either by the conversion of metabolite A_{i-1} or

Metabolic network	MMB EM/EP	RMS EM/EP	Interior EM/EP
Glycolysis/Gluconeogenesis	1226/529	48/46	18190/1095
Citrate cycle (TCA cycle)	1608/568	502/258	1760/480
Pentose phosphate pathway	489/340	25/25	4641/1217
Pentose and glucuronate	1076/32	642/76	540/3
Fructose and mannose	154/148	14/14	2243/1895
Galactose	212/152	45/45	366/254
Starch and sucrose	108 /107	8/8	1981/1565
Pyruvate	2016/1776	146/146	45546/25293
Propanoate	449/93	133/32	295/50
Butanoate	357/244	35/34	1746/201
Nitrogen	183/171	22/22	396/384
Sulfur	44/44	1/1	276/276

Table 4.2: The distribution of the elementary modes (EM) and the extreme pathways (EP) in the three parts of the steady-state flux cone: the minimal proper faces (MMB), the lineality space (RMS), and the interior of the cone. Each pair of opposite extreme pathways is considered as one reversible pathway belonging to the RMS. The two-cycle extreme pathways made from a forward and a backward reaction are not taken into account.

by the conversion of metabolite B_{i-1}, which is in turn obtained by the conversion of metabolite A_{i-1}. We assume that all the stoichiometric coefficients are equal to one and all reactions are irreversible. Let $\xi : \{1, \ldots, p+1\} \to \mathbb{R}^n$ be the function that maps each $i \in \{1, \ldots, p+1\}$ to $\xi(i) = 3(i-1)+2$. Each steady-state flux distribution $v \in \mathbb{R}^n$ obeys, in addition to the thermodynamic constraints, the following stoichiometric constraints:

$$v_1 = v_2 + v_3, \; v_{\xi(p)} + v_{\xi(p)+2} = v_{\xi(p+1)}, \tag{4.13}$$

$$\text{For all } i \in \{2, \ldots, p\} \quad v_{\xi(i-1)} + v_{\xi(i-1)+2} = v_{\xi(i)} + v_{\xi(i)+2}, \tag{4.14}$$

$$\text{For all } i \in \{1, \ldots, p\} \quad v_{\xi(i)+1} = v_{\xi(i)+2}. \tag{4.15}$$

The constraints (4.13) and (4.14) (resp. (4.15)) express the flux balances around metabolites A_1, A_{p+1} and A_i (resp. B_i) for $i = 1, \ldots, p$. Combining these constraints, we obtain

$$\text{For all } i \in \{2, \ldots, p\} \quad v_1 = v_{\xi(i-1)} + v_{\xi(i-1)+1} = v_{\xi(i)} + v_{\xi(i)+1} = v_{\xi(p+1)}. \tag{4.16}$$

Accordingly, the following proposition states that the hypothetic network depicted in Fig. 4.2 contains 2^p elementary modes, and so the number of MMBs is exponential in the number of reactions. In general, metabolic networks contain reversible reactions and so the number of MMBs can be much smaller than the number of EMs.

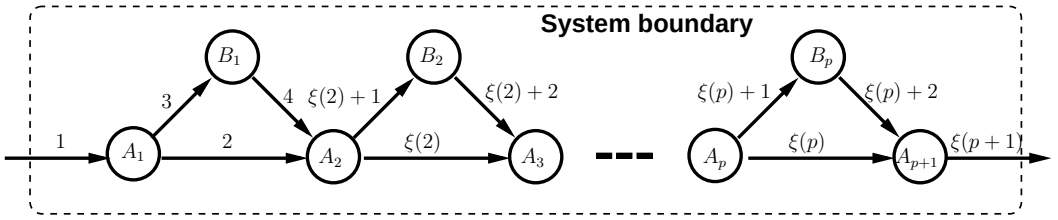


Figure 4.2: A hypothetic network for which the number of MMBs is exponential in the number of reactions.

Proposition 4.20. *There are 2^p elementary modes in the hypothetic network depicted in Fig. 4.2.*

Proof. By induction on p. □

4.5 Red Blood Cell Metabolism

In an attempt to illustrate the reliability and usefulness of our new approach, we propose to investigate the metabolic functions of the human red blood cell. Due to its relative accessibility and medical relevance, the human erythrocyte has been investigated not only by constraint-based approaches [16; 39; 109; 121], but also by several kinetic studies [68; 67; 94].

The main function of a red blood cell is the transport of oxygen and carbon dioxide. In addition, the cell needs to produce cofactors (ATP, NADPH, and NADH) for its own survival. The erythrocyte has to balance osmotic pressure while maintaining electroneutrality on both sides of the membrane [121].

The red blood cell metabolic network depicted in Fig. 4.3 contains three major parts: *glycolysis* including the *Rapoport-Leubering shunt, pentose phosphate pathway* and *adenosine nucleotide metabolism* [16]. It consists of 43 internal metabolites and 50 reactions, whereof 30 reactions are irreversible (Tab. 4.3). The main energy source for the red blood cell is glucose (GLC). Like many cellular systems, erythrocytes can also exchange adenine (ADE) and hypoxanthine (HYPX) through the membrane. Lactate (LAC), pyruvate (PYR), $2,3$-Diphosphoglycerate (D23PG) and carbon dioxide (CO_2) are excreted. The good correlation between in silico results given in [16; 109] and experimental evidence for human erythrocytes shows the reliability of this in silico model.

Table 4.3: The red blood cell metabolism adapted from [16].

Reaction abbreviation	Reaction name	Reaction equation
Glycolysis		
HK	hexokinase	GLC + ATP → G6P + ADP
PGI	glucose-6-phosphate isomerase	G6P ↔ F6P
PFK	phosphofructokinase	F6P + ATP → FDP + ADP
ALD	aldolase	FDP ↔ DHAP + GA3P
TPI	triosephosphate isomerase	DHAP ↔ GA3P

Continued on next page

Table 4.3 – continued from previous page

Reaction abbreviation	Reaction name	Reaction equation
GAPDH	glyceraldehyde-3-phosphate dehydrogenase	GA3P + NAD ↔ D13PG + NADH
PGK	phosphoglycerate kinase	D13PG + ADP ↔ P3G + ATP
PGM	phosphoglycerate mutase	P3G ↔ P2G
EN	enolase	P2G ↔ PEP
PK	pyruvate kinase	PEP + ADP → PYR + ATP
LDH	lactate dehydrogenase	PYR + NADH → LAC + NAD
Rapport-Luebering shunt		
DPGM	diphosphoglycerate mutase	D13PG → D23PG
DPGase	diphosphoglycerate phosphatase	D23PG → P3G
Pentose phosphate pathway		
G6PD	glucose-6-phosphate dehydrogenase	G6P + NADP → GL6P + NADPH
PGLase	phosphogluconolactonase	GL6P ↔ GO6P
GL6PDH	phosphogluconate dehydrogenase	GO6P + NADP → RU5P + NADPH + CO2
R5PI	ribose-5-phosphate isomerase	RU5P ↔ R5P
Xu5PE	ribulose phosphate epimerase	RU5P ↔ X5P
TKI	transketolase	X5P + R5P ↔ GA3P + S7P
TKII	transketolase	E4P + X5P ↔ GA3P + F6P
TA	transaldolase	S7P + GA3P ↔ E4P + F6P
Nucleotide metabolism		
AMPase	adenosine monophosphate phosphohydrolase	AMP → ADO
ADA	adenosine deaminase	ADO → INO
AK	adenosine kinase	ATP + ADO → ADP + AMP
ApK	adenylate kinase	2 ADP ↔ ATP + AMP
AMPDA	adenosine monophosphate deaminase	AMP → IMP
AdPRT	adenine phosphoribosyltransferase	ADE + PRPP → AMP
PRM	phosphoribomutase	RIP ↔ R5P
PRPPsyn	phosphoribosylpyrophosphate synthetase	R5P + ATP → PRPP + AMP
HGPRT	hypoxanthine phosphoribosyltransferase	HYPX + PRPP → IMP
IMPase	inosine monophosphate phosphohydrolase	IMP → INO
PNPase	purine-nucleoside phosphorylase	INO ↔ HYPX + RIP
Cellular functions		
MemPhos	membrane phosphorylation	ATP → ADP
GSSGR	glutathione-disulfide reductase	NADPH + GSSG ↔ NADP + 2 GSH
GSHox	glutathione oxidase	2 GSH + O2 → GSSG + H2O2
GSHpox	glutathione peroxidase	2 GSH + H2O2 → GSSG
NaKATPase	sodium-potassium cation pump	ATP + 3 Na + 2 Kext → ADP + 3 Naext + 2 K
D23PGdrain	2,3-diphosphoglyerate drain	D23PG + Hb → D23PGext
MetHbRed	methemoglobin reductase	MetHb + NADH → Hb + NAD
Boundary reactions		
KLeak	. ↔ K	
NaLeak	. ↔ Na	
HXtrans	HYPX ↔ .	
PYRex	PYR → .	
LACex	LAC → .	
CO2out	CO2 → .	
GLCin	. → GLC	
ADEin	. → ADE	
Hbout	Hb → .	
O2in	. → O2	
MetHbin	. → MetHb	

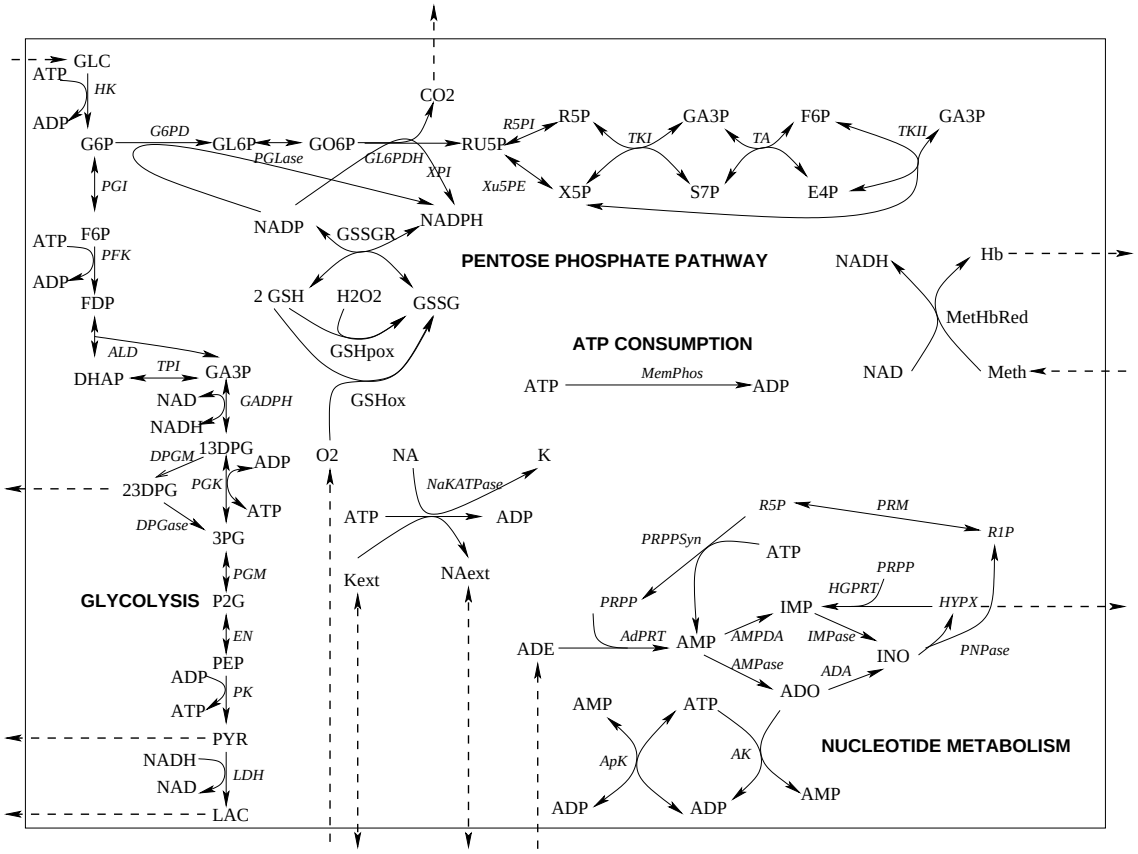


Figure 4.3: The red blood cell metabolism adapted from [16]. Dashed arrows correspond to boundary reactions while reversible reactions are indicated by double arrowheads. Reaction equations are listed in Tab. 4.3.

Table 4.4: Minimal metabolic behaviors (MMBs) for human red blood cell. The irreversible reactions defining each MMB are given in bold. The underlined reactions define the specific purpose for which serves the corresponding MMB. The numbers next to the reaction names denote related fluxes carried by the corresponding reactions.

MMB number	MMBs sorted with respect to their metabolic functions
	Glycolysis Pathway
	Pyruvate producing MMBs
1.	**2 MemPhos**
	1 GLCin 1 HK 2 PK, 1 PFK, 2 MetHbin 2 MetHbRed 2 Hbout 2 PYRex
	1 PGI 2 GAPDH 2 PGM 2 EN 1 ALD 1 TPI 2 PGK
2.	**2 NaKATPase**
	1 GLCin 1 HK 2 PK, 1 PFK, 2 MetHbin 2 MetHbRed 2 Hbout 2 PYRex
	1 PGI 2 GAPDH 2 PGM 2 EN 1 ALD 1 TPI 2 PGK, -4 KLeak 6 NaLeak
3.	**1 HGPRT 1 IMPase 1 PRPPsyn**
	1 GLCin 1 HK 2 PK, 1 PFK, 2 MetHbin 2 MetHbRed 2 Hbout 2 PYRex
	1 PGI 2 GAPDH 2 PGM 2 EN 1 ALD 1 TPI 2 PGK, -1 ApK 1 PNPase 1 PRM
4.	**2 AMPase 2 AK**
	1 GLCin 1 HK 2 PK, 1 PFK, 2 MetHbin 2 MetHbRed 2 Hbout 2 PYRex
	1 PGI 2 GAPDH 2 PGM 2 EN 1 ALD 1 TPI 2 PGK
5.	**1 AdPRT 1 PRPPsyn 1 ADEin, 1 AMPase 1 ADA**
	1 GLCin 1 HK 2 PK, 1 PFK, 2 MetHbin 2 MetHbRed 2 Hbout 2 PYRex
	1 PGI 2 GAPDH 2 PGM 2 EN 1 ALD 1 TPI 2 PGK, 1 PRM 1 PNPase 1 HXtrans -1 ApK
6.	**1 AdPRT 1 PRPPsyn 1 ADEin, 1 IMPase 1 AMPDA**
	1 GLCin 1 HK 2 PK, 1 PFK, 2 MetHbin 2 MetHbRed 2 Hbout 2 PYRex
	1 PGI 2 GAPDH 2 PGM 2 EN 1 ALD 1 TPI 2 PGK, 1 PRM 1 PNPase 1 HXtrans -1 ApK

Continued on next page

Table 4.4 – continued from previous page

MMB number	MMBs sorted with respect to their metabolic functions
7.	**2 DPGM 2 DPGase** **1 GLCin 1 HK 2 PK, 1 PFK, 2 MetHbin 2 MetHbRed 2 Hbout 2 PYRex** *1 PGI 2 GAPDH 2 PGM 2 EN 1 ALD 1 TPI*
8.	**1 DPGM 1 D23PGdrain** **1 GLCin 1 HK 2 PK, 1 PFK, 2 MetHbin 2 MetHbRed 1 Hbout 1 PYRex** *1 PGI 2 GAPDH 1 PGM 1 EN 1 ALD 1 TPI 1 PGK*
	Lactate producing MMBs
9.	**2 MemPhos** **1 GLCin 1 HK 2 PK, 1 PFK, 2 LDH 2 LACex** *1 PGI 2 GAPDH 2 PGM 2 EN 1 ALD 1 TPI 2 PGK*
10.	**2 NaKATPase** **1 GLCin 1 HK 2 PK, 1 PFK, 2 LDH 2 LACex** *1 PGI 2 GAPDH 2 PGM 2 EN 1 ALD 1 TPI 2 PGK, -4 KLeak 6 NaLeak*
11.	**1 HGPRT 1 IMPase 1 PRPPsyn** **1 GLCin 1 HK 2 PK, 1 PFK, 2 LDH 2 LACex** *1 PGI 2 GAPDH 2 PGM 2 EN 1 ALD 1 TPI 2 PGK, -1 ApK 1 PNPase 1 PRM*
12.	**2 AMPase 2 AK** **1 GLCin 1 HK 2 PK, 1 PFK, 2 LDH 2 LACex** *1 PGI 2 GAPDH 2 PGM 2 EN 1 ALD 1 TPI 2 PGK*
13.	**1 AdPRT 1 PRPPsyn 1 ADEin, 1 AMPase 1 ADA** **1 GLCin 1 HK 2 PK, 1 PFK, 2 LDH 2 LACex** *1 PGI 2 GAPDH 2 PGM 2 EN 1 ALD 1 TPI 2 PGK, 1 PRM 1 PNPase 1 HXtrans -1 ApK*
14.	**1 AdPRT 1 PRPPsyn 1 ADEin, 1 IMPase 1 AMPDA** **1 GLCin 1 HK 2 PK, 1 PFK, 2 LDH 2 LACex** *1 PGI 2 GAPDH 2 PGM 2 EN 1 ALD 1 TPI 2 PGK, 1 PRM 1 PNPase 1 HXtrans -1 ApK*

Continued on next page

Table 4.4 – continued from previous page

MMB number	MMBs sorted with respect to their metabolic functions
15.	**2 DPGM 2 DPGase**
	1 GLCin 1 HK 2 PK, 1 PFK, 2 LDH 2 LACex
	1 PGI 2 GAPDH 2 PGM 2 EN 1 ALD 1 TPI
16.	**1 DPGM 1 D23PGdrain**
	1 GLCin 1 HK 1 PK, 1 PFK, 1 LDH 1 LACex
	1 PGI 2 GAPDH 1 PGM 1 EN 1 ALD 1 TPI 1 PGK
Pentose Phosphate Pathway	
Pyruvate producing MMBs	
17.	**2 MemPhos**
	1 GLCin 1 HK 1 PK, 3 GSHox 3 GSHpox 3 G6PD 3 GL6PDH 3 CO2out 3 O2in, 1 MetHbin 1 MetHbRed 1 Hbout 1 PYRex
	-2 PGI 1 GAPDH 1 PGM 1 EN 3 PGLase 6 GSSGR 1 R5PI 2 Xu5PE 1 TKI 1 TKII 1 TA 1 PGK
18.	**2 NaKATPase**
	1 GLCin 1 HK 1 PK, 3 GSHox 3 GSHpox 3 G6PD 3 GL6PDH 3 CO2out 3 O2in, 1 MetHbin 1 MetHbRed 1 Hbout 1 PYRex
	-2 PGI 1 GAPDH 1 PGM 1 EN 3 PGLase 6 GSSGR 1 R5PI 2 Xu5PE 1 TKI 1 TKII 1 TA 1 PGK, -2 KLeak 3 NaLeak
19.	**1 HGPRT 1 IMPase 1 PRPPsyn**
	2 GLCin 2 HK 2 PK, 6 GSHox 6 GSHpox 6 G6PD 6 GL6PDH 6 CO2out 6 O2in, 2 MetHbin 2 MetHbRed 2 Hbout 2 PYRex
	-4 PGI 2 GAPDH 2 PGM 2 EN 6 PGLase 12 GSSGR 2 R5PI 4 Xu5PE 2 TKI 2 TKII 2 TA 2 PGK, -1 ApK 1 PNPase 1 PRM
20.	**2 AMPase 2 AK**
	1 GLCin 1 HK 1 PK, 3 GSHox 3 GSHpox 3 G6PD 3 GL6PDH 3 CO2out 3 O2in, 1 MetHbin 1 MetHbRed 1 Hbout 1 PYRex
	-2 PGI 1 GAPDH 1 PGM 1 EN 3 PGLase 6 GSSGR 1 R5PI 2 Xu5PE 1 TKI 1 TKII 1 TA 1 PGK
21.	**1 AdPRT 1 PRPPsyn 1 ADEin, 1 AMPase 1 ADA**
	2 GLCin 2 HK 2 PK, 6 GSHox 6 GSHpox 6 G6PD 6 GL6PDH 6 CO2out 6 O2in, 2 MetHbin 2 MetHbRed 2 Hbout 2 PYRex
	-4 PGI 2 GAPDH 2 PGM 2 EN 6 PGLase 12 GSSGR 2 R5PI 4 Xu5PE 2 TKI 2 TKII 2 TA 2 PGK, -1 ApK 1 PNPase 1 PRM 1 HXtrans

Continued on next page

Table 4.4 – continued from previous page

MMB number	MMBs sorted with respect to their metabolic functions
22.	**1 AdPRT 1 PRPPsyn 1 ADEin, 1 IMPase 1 AMPDA** **2 GLCin 2 HK 2 PK, 6 GSHox 6 GSHpox 6 G6PD 6 GL6PDH 6 CO2out 6 O2in, 2 MetHbin 2 MetHbRed 2 Hbout 2 PYRex** *-4 PGI 2 GAPDH 2 PGM 2 EN 6 PGLase 12 GSSGR 2 R5PI 4 Xu5PE 2 TKI 2 TKII 2 TA 2 PGK, -1 ApK 1 PNPase 1 PRM 1 HXtrans*
23.	**2 DPGM 2 DPGase** **1 GLCin 1 HK 1 PK, 3 GSHox 3 GSHpox 3 G6PD 3 GL6PDH 3 CO2out 3 O2in, 1 MetHbin 1 MetHbRed 1 Hbout 1 PYRex** *-2 PGI 1 GAPDH 1 PGM 1 EN 3 PGLase 6 GSSGR 1 R5PI 2 Xu5PE 1 TKI 1 TKII 1 TA*
24.	**1 DPGM 1 D23PGdrain** **2 GLCin 2 HK 1 PK, 6 GSHox 6 GSHpox 6 G6PD 6 GL6PDH 6 CO2out 6 O2in, 2 MetHbin 2 MetHbRed 1 Hbout 1 PYRex** *-4 PGI 2 GAPDH 1 PGM 2 EN 6 PGLase 12 GSSGR 2 R5PI 4 Xu5PE 2 TKI 2 TKII 2 TA 1 PGK*
25.	**2 MemPhos** **1 GLCin 1 HK 1 PK, 3 GSHox 3 GSHpox 3 G6PD 3 GL6PDH 3 CO2out 3 O2in, 1 LDH 1 LACex** *-2 PGI 1 GAPDH 1 PGM 1 EN 3 PGLase 6 GSSGR 1 R5PI 2 Xu5PE 1 TKI 1 TKII 1 TA 1 PGK*
26.	**2 NaKATPase** **1 GLCin 1 HK 1 PK, 3 GSHox 3 GSHpox 3 G6PD 3 GL6PDH 3 CO2out 3 O2in, 1 LDH 1 LACex** *-2 PGI 1 GAPDH 1 PGM 1 EN 3 PGLase 6 GSSGR 1 R5PI 2 Xu5PE 1 TKI 1 TKII 1 TA 1 PGK, -2 KLeak 3 NaLeak*
27.	**1 HGPRT 1 IMPase 1 PRPPsyn** **2 GLCin 2 HK 2 PK, 6 GSHox 6 GSHpox 6 G6PD 6 GL6PDH 6 CO2out 6 O2in, 2 LDH 2 LACex** *-4 PGI 2 GAPDH 2 PGM 2 EN 6 PGLase 12 GSSGR 2 R5PI 4 Xu5PE 2 TKI 2 TKII 2 TA 2 PGK, -1 ApK 1 PNPase 1 PRM*
28.	**2 AMPase 2 AK** **1 GLCin 1 HK 1 PK, 3 GSHox 3 GSHpox 3 G6PD 3 GL6PDH 3 CO2out 3 O2in, 1 LDH 1 LACex** *-2 PGI 1 GAPDH 1 PGM 1 EN 3 PGLase 6 GSSGR 1 R5PI 2 Xu5PE 1 TKI 1 TKII 1 TA 1 PGK*
29.	**1 AdPRT 1 PRPPsyn 1 ADEin, 1 AMPase 1 ADA** **2 GLCin 2 HK 2 PK, 6 GSHox 6 GSHpox 6 G6PD 6 GL6PDH 6 CO2out 6 O2in, 2 LDH 2 LACex** *-4 PGI 2 GAPDH 2 PGM 2 EN 6 PGLase 12 GSSGR 2 R5PI 4 Xu5PE 2 TKI 2 TKII 2 TA 2 PGK, -1 ApK 1 PNPase 1 PRM 1 HXtrans*

Continued on next page

Table 4.4 – continued from previous page

MMB number	MMBs sorted with respect to their metabolic functions
30.	**1 AdPRT 1 PRPPsyn 1 ADEin, 1 IMPase 1 AMPDA**
	2 GLCin 2 HK 2 PK, 6 GSHox 6 GSHpox 6 G6PD 6 GL6PDH 6 CO2out 6 O2in, 2 LDH 2 LACex
	-4 PGI 2 GAPDH 2 PGM 2 EN 6 PGLase 12 GSSGR 2 R5PI 4 Xu5PE 2 TKI 2 TKII 2 TA 2 PGK, -1 ApK 1 PNPase 1 PRM 1 HXtrans
31.	**2 DPGM 2 DPGase**
	1 GLCin 1 HK 1 PK, 3 GSHox 3 GSHpox 3 G6PD 3 GL6PDH 3 CO2out 3 O2in, 1 LDH 1 LACex
	-2 PGI 1 GAPDH 1 PGM 1 EN 3 PGLase 6 GSSGR 1 R5PI 2 Xu5PE 1 TKI 1 TKII 1 TA
32.	**1 DPGM 1 D23PGdrain**
	2 GLCin 2 HK 1 PK, 6 GSHox 6 GSHpox 6 G6PD 6 GL6PDH 6 CO2out 6 O2in, 1 MetHbin 1 MetHbRed 1 LDH 1 LACex
	-4 PGI 2 GAPDH 1 PGM 2 EN 6 PGLase 12 GSSGR 2 R5PI 4 Xu5PE 2 TKI 2 TKII 2 TA 1 PGK

Analyzing the red blood cell metabolic network results in 32 minimal metabolic behaviors (MMBs) listed in Tab. 4.4. The corresponding flux cone is pointed, and so there is no steady-state flux distribution involving only reversible reactions. In general, the irreversible reactions defining an MMB and their associated pseudo-irreversible reactions cannot necessarily operate on their own. They may need some fully reversible reactions to define a complete metabolic pathway. Here, since the flux cone is pointed, each MMB and its auxiliary set define a metabolic pathway, which is in fact an elementary mode.

Although many MMBs partly overlap each other, each MMB serves a specific purpose carried out by a set of characteristic reactions. MMB1 is characterized by the use of reaction MemPhos which allows the cell to maintain the plasticity of its membrane [16]. MMB2 could be used by the cell to control its volume through the sodium-potassium cation pump NaKATPase [39]. In addition to the transport of oxygen and its delivery, the red blood cell is also responsible for the carriage of purine bases [123]. This function is performed by MMB5 and MMB6 in which adenine (ADE) is taken up into the cell and Hypoxanthine (HYPX) is excreted. Although both MMBs have the same overall stoichiometry and serve the transport of purine bases, they slightly differ in the usage of AMPASE and ADA versus IMPASE and AMPDA, showing a certain network redundancy in the red blood cell metabolism.

MMB7 utilizes the D23PG shunt, namely reactions DPGM and DPGase, instead of the ATP producing reaction PGK and backs into main glycolysis. The cell could use this MMB to regulate its ATP production [121]. MMB8 is responsible for the formation of $2,3$-diphosphoglycerate (D23PG) for use in the regulation of the oxygen affinity of hemoglobin [39; 121]. In MMB4, reactions AMPASE and AK form a cycle that consumes ATP repeatedly. Accordingly, MMB4 serves to dissipate excess ATP [16]. The same function is realized by MMB3 as well. This MMB cycles through the nucleotide metabolism, causing dissipation of ATP through reactions ApK and PRPPsyn.

All the MMBs1-8 metabolize glucose through the glycolysis pathway with the production of pyruvate as an end product. Each of these eight MMBs has a nearly identical "twin" MMB among MMBs9-16, the only difference being in the end product. Each of the MMBs9-16 takes its corresponding MMB a step further and converts pyruvate (PYR) into lactate (LAC). This conversion serves to balance all NADH produced by the cell [121]. Since there is no load on NADH in the red blood cell, the MMBs9-16 could be used to completely balance the NAD/NADH ratio.

There are two features which distinguish the MMBs1-16 from the remaining MMBs. First, all the MMBs1-16 involve reaction TPI, which plays an important role in several metabolic networks. In fact, this reaction prevent the accumulation of dihydroxyacetone phosphate (DHAP), which is reported to be toxic for cellular functions and leads to hemolytic anemia with neurological dysfunction [16]. Second, the MMBs1-16 are merely made up of glycolysis and nucleotide reactions and do not involve any reaction from the pentose phosphate pathway (PPP).

The main functions of PPP in the human red blood cell are to generate NADPH and to provide the cell with ribose-5-phosphate (R5P) for the synthesis of the nu-

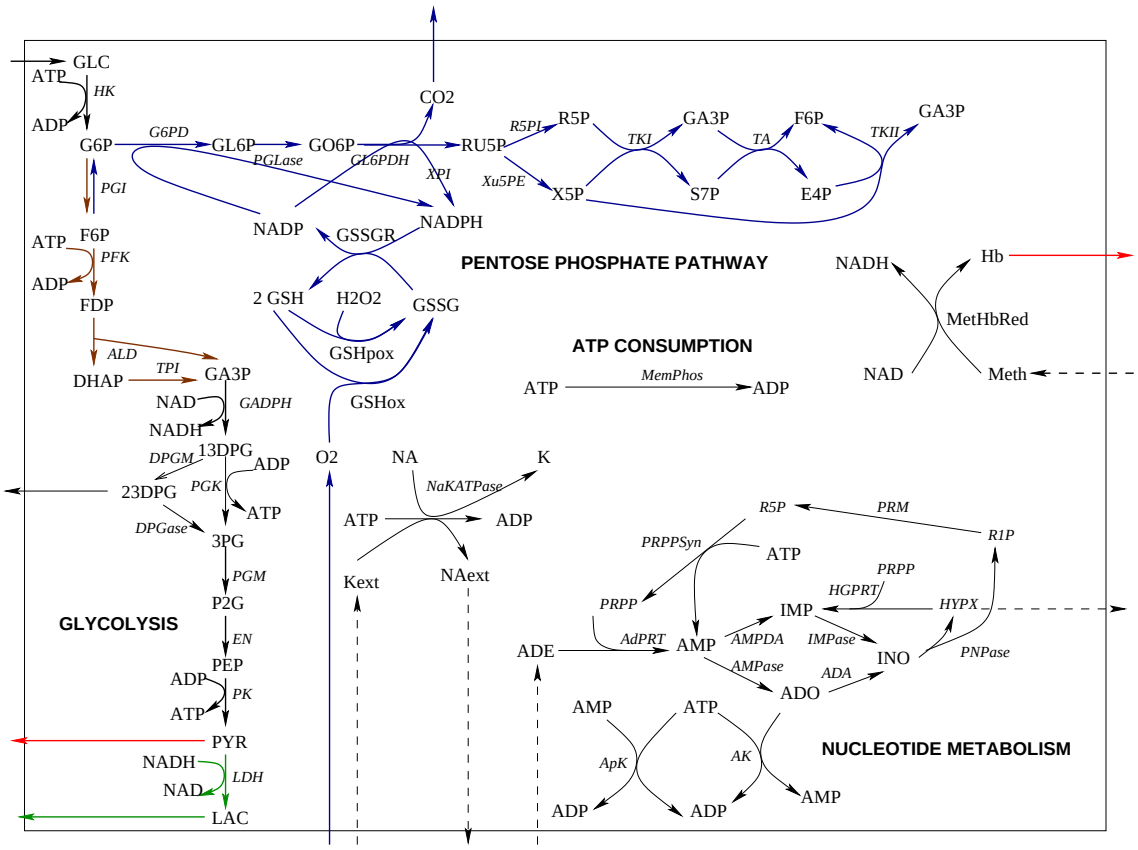


Figure 4.4: The red blood cell metabolism adapted from [16]. The bold reactions HK, GAPDH, PGM, EN and PK participate in all steady-state flux distributions. Except reaction PGI, each reversible reaction is operating only in one direction. Glucose (GLC) can be metabolized either by using reaction PGI in the forward direction together with reactions PFK, ALD and TPI (brown color) or by using reaction PGI in the backward direction together with reactions from PPP (blue color). The end product of each MMB is either pyruvate or lactate.

cleotides [16]. One of the uses of NADPH in the red blood cell is to prevent oxidative stress by reducing GSSG to GSH through reaction GSSGR. The reduced glutathione (GSH) is required to remove hydrogen peroxide (H_2O_2) through reaction GSHox [9]. The MMBs16-32 are similar to the former MMBs1-16. The only difference is that instead of using reactions TPI, ALD and PFK to prevent the toxic accumulation of DHAP, these MMBs utilize reactions G6PDH, GSSGR and GSHox to remove H_2O_2 and protect the cell against oxidative stress. This task requires oxygen (O_2) and NADPH. While the former is taken up into the cell, the latter is generated by only PPP . Accordingly, all the MMBs16-32 involve reactions from PPP.

In addition to the 32 elementary modes given in Tab. 4.4, which represent the MMBs of the red blood cell metabolic network, there are 16 additional elementary modes. These are obtained by combining MMBs that involve reaction PGI in opposite directions. Interestingly, Fig. 4.5 shows that each of these additional elementary modes

is a combination of two MMBs that fulfill the following conditions:

1. Both MMBs have the same metabolic function,
2. Both MMBs have the same end product,
3. Exactly one of these MMBs involves reactions from PPP.

We conclude that all metabolic functions of elementary modes are carried out by minimal metabolic behaviors. Note that except reaction PGI, each reversible reaction is operating only in one direction shown in Fig. 4.4. Accordingly, except reaction PGI, all (pseudo-) irreversible reactions defining an MMB D^1 will participate in an additional EM e if the latter is obtained by combining D^1 with another MMB D^2. The set of active reactions of e is then the union of those of D^1 and D^2. We should also mention that although only reaction PGI is able to work in both directions, the number of elementary modes is larger than that of minimal metabolic behaviors.

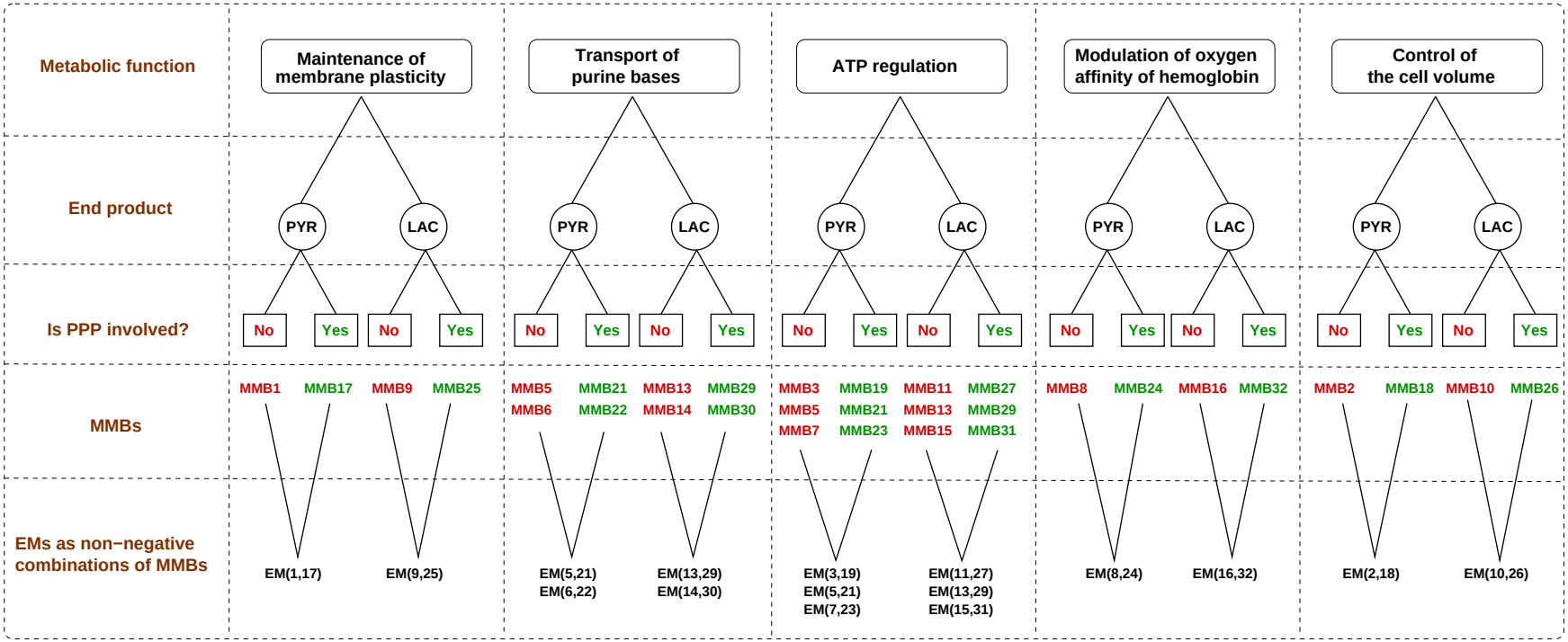


Figure 4.5: The main metabolic functions of the red blood cell metabolic network. Minimal metabolic behaviors (MMBs) are classified with respect to three characteristics: their functions, their end products and whether they involve reactions from PPP. The additional 16 elementary modes (EMs) are given as combinations of MMBs. EM(i, j) means that this mode is a non-negative combination of MMBi and MMBj.

CHAPTER

5 On Inner and Outer Descriptions of the Steady-State Flux Cone

In this chapter, we study the relationship between inner and outer descriptions of the flux cone. We first characterize the outcome of the network reconfiguration in terms of the outer description of the reconfigured cone. The reconfiguration leads to an increase in the size of the description and changes in the reversibility type of reactions. Then we give a generic procedure to show how inner descriptions can be computed from the outer one. We use this procedure to explain why, for large-scale metabolic networks, the size of the inner descriptions may be several orders of magnitude larger than that of the outer description. The main results of this chapter are published in [62].

5.1 Outer Description of the Reconfigured Flux Cone

In this section, we analyze the impact of reconfiguring the metabolic network. The effects include an increase in the size of the outer description of the reconfigured cone and changes in the reversibility type of reactions. Here, we define the *size* of an outer description of a flux cone as the sum of the number of its minimal proper faces and the dimension of its lineality space.

Let SR $\subseteq Rev$ be the set of split reactions. The network reconfiguration can be seen as an iterative procedure that consists of |SR| iterations, each splitting some reversible reaction. As will be shown, each iteration increases the description of the flux cone depending on the reversibility type of the split reaction. The increase is significant when the split reaction is pseudo-irreversible. Note that there are at most t iterations where the split reaction can be fully reversible, with $t = \dim(\text{lin.space}(C))$.

In the following, we consider the case of splitting one reaction, which is denoted by j. The reconfigured flux cone C', which contains all possible steady-state flux distributions in the reconfigured network, is given by

$$C' = \{(v, w) \in \mathbb{R}^{n+1} \mid Sv = w \cdot S_{*j},\ v_i \geq 0, \text{ for all } i \in Irr,\ v_j \geq 0,\ w \geq 0\}. \quad (5.1)$$

According to equation (5.1), splitting reaction j increases the number of variables and constraints by 1 and 2, respectively. Indeed, the reconfigured network contains one more reaction denoted by $n + 1$. The set of irreversible reactions in the reconfigured

network is $Irr' = Irr \cup \{j, n+1\}$. Accordingly, the lineality space of the reconfigured flux cone C' is given by

$$\text{lin.space}(C') = \{(v,0) \in \mathbb{R}^{n+1} \mid Sv = 0,\ v_i = 0, \text{ for all } i \in Irr,\ v_j = 0\},$$

or equivalently,

$$\text{lin.space}(C') = \{(v,0) \in \mathbb{R}^{n+1} \mid v \in \text{lin.space}(C \cap \{v \in \mathbb{R}^n \mid v_j = 0\})\}. \quad (5.2)$$

Since $\text{lin.space}(C \cap \{v \in \mathbb{R}^n \mid v_j = 0\}) = \text{lin.space}(C) \cap \{v \in \mathbb{R}^n \mid v_j = 0\}$, it follows from equation (5.2) that

$$\dim(\text{lin.space}(C')) = \dim(\text{lin.space}(C) \cap \{v \in \mathbb{R}^n \mid v_j = 0\}).$$

Lemma 5.1. *If $j \in Prev_0$ is pseudo-irreversible, then $\dim(\text{lin.space}(C')) = \dim(\text{lin.space}(C))$. If $j \in Frev$ is fully reversible, then $\dim(\text{lin.space}(C')) = \dim(\text{lin.space}(C)) - 1$.*

Proof. Suppose $j \in Prev_0$ is pseudo-irreversible. Then, $b_j = 0$ for each vector $b \in \text{lin.space}(C)$. Hence, $\text{lin.space}(C) \subseteq \{v \in \mathbb{R}^n \mid v_j = 0\}$ and so $\dim(\text{lin.space}(C')) = \dim(\text{lin.space}(C))$. Now suppose $j \in Frev$ is fully reversible. There exists then $b \in \text{lin.space}(C)$ such that $b_j \neq 0$ and so $\text{lin.space}(C) \nsubseteq \{v \in \mathbb{R}^n \mid v_j = 0\}$. Therefore, $\dim(\text{lin.space}(C')) = \dim(\text{lin.space}(C)) - 1$. □

In the following, we will characterize the minimal proper faces of the reconfigured flux cone C'. We first consider the case of a minimal proper face G' with $v'_j = v'_{n+1} = 0$ for all $v' \in G'$.

Lemma 5.2. *Let $G' \subseteq C'$ such that $v'_j = v'_{n+1} = 0$ for all $v' \in G'$. Then, the following are equivalent:*

- *G' is a minimal proper face of C'.*
- *There exists a minimal proper face G of $C \cap \{v \in \mathbb{R}^n \mid v_j = 0\}$ such that $G' = \{(v,0) \in \mathbb{R}^{n+1} \mid v \in G\}$.*

If this is the case, G and G' have the same characteristic set.

Proof. "⇒": Suppose G' is a minimal proper face of C' and D' is its characteristic set. Since $v'_j = v'_{n+1} = 0$ for all $v' \in G'$, we get $D' \subseteq Irr$. Let $G = \{v \in C \mid v_j = 0,\ v_i = 0, \text{ for all } i \in Irr \setminus D'\}$. We have $G' = \{(v,0) \in \mathbb{R}^{n+1} \mid v \in G\}$ and so $\dim(G) = \dim(G')$. Since G' is a minimal proper face of C' and $\dim(\text{lin.space}(C')) = \dim(\text{lin.space}(C \cap \{v \in \mathbb{R}^n \mid v_j = 0\}))$, we get $\dim(G) = \dim(\text{lin.space}(C \cap \{v \in \mathbb{R}^n \mid v_j = 0\})) + 1$ and so the claim follows.

"⇐": Immediate. □

We now will study the minimal proper faces of the reconfigured flux cone C', depending on the reversibility type of the split reaction.

5.1.1 Splitting a Fully Reversible Reaction

If $j \in \mathit{Frev}$ is fully reversible, there exists a flux distribution in the reconfigured network that involves either reaction j or $n+1$ and no other irreversible reactions. Accordingly, as will be stated in the following proposition, reactions j and $n+1$ define two trivial minimal proper faces of C' given by

$$\begin{aligned} G^j &= \{(v,0) \in \mathbb{R}^{n+1} \mid Sv = 0,\ v_i = 0, \text{ for all } i \in \mathit{Irr},\ v_j \geq 0\}, \\ G^{n+1} &= \{(v,w) \in \mathbb{R}^{n+1} \mid Sv = w \cdot S_{*j},\ v_i = 0, \text{ for all } i \in \mathit{Irr},\ v_j = 0,\ w \geq 0\}. \end{aligned}$$

Proposition 5.3. *If $j \in \mathit{Frev}$ is fully reversible, then G^j and G^{n+1} are two minimal proper faces of C' whose characteristic sets are $D^j = \{j\}$ and $D^{n+1} = \{n+1\}$, respectively.*

Proof. Suppose $j \in \mathit{Frev}$ is fully reversible. Then there exists $b \in \mathrm{lin.space}(C)$ such that $b_j > 0$. Let $I_j = \mathit{Irr} \cup \{n+1\}$. We have $G^j = \{v' \in C' \mid v'_j \geq 0,\ v'_i = 0, \text{ for all } i \in I_j\}$ and $\mathrm{lin.space}(C') = \{v' \in C' \mid v'_j = 0,\ v'_i = 0, \text{ for all } i \in I_j\}$. In addition, we have $(b,0) \in G^j \setminus \mathrm{lin.space}(C')$ and so $\dim(G^j) = \dim(\mathrm{lin.space}(C')) + 1$. Therefore, G^j is a minimal proper face characterized by reaction j. Similarly, let $I_{n+1} = \mathit{Irr} \cup \{j\}$. We have $G^{n+1} = \{v' \in C' \mid v'_{n+1} \geq 0,\ v'_i = 0, \text{ for all } i \in I_{n+1}\}$ and $\mathrm{lin.space}(C') = \{v' \in C' \mid v'_{n+1} = 0,\ v'_i = 0, \text{ for all } i \in I_{n+1}\}$. Define $u \in \mathbb{R}^{n+1}$ by $u_j = 0$, $u_{n+1} = b_j$ and $u_i = -b_i$ for all $i \in \{1,\ldots,n\} \setminus \{j\}$. We have $u \in G^{n+1} \setminus \mathrm{lin.space}(C')$ and so $\dim(G^{n+1}) = \dim(\mathrm{lin.space}(C')) + 1$. Accordingly, G^{n+1} is a minimal proper face characterized by reaction $n+1$. Since for each $v' \in G^j$ (resp. $v' \in G^{n+1}$), $v'_i = 0$ for all $i \in \mathit{Irr}' \setminus \{j\}$ (resp. $i \in \mathit{Irr}' \setminus \{n+1\}$), the claim follows. □

Next we are interested in non-trivial minimal proper faces of C'. Here, we get the following result.

Proposition 5.4. *Let $G' \subseteq C'$ such that $G' \neq G^j$ and $G' \neq G^{n+1}$. If $j \in \mathit{Frev}$ is fully reversible, then the following are equivalent:*

- *G' is a minimal proper face of C'.*
- *There exists a minimal proper face G of C such that $G' = \{(v,0) \in \mathbb{R}^{n+1} \mid v \in G \cap \{v \in \mathbb{R}^n \mid v_j = 0\}\}$.*

Proof. "⇒": According to Lemma 5.2, there exists a minimal proper face G'' of $C \cap \{v \in \mathbb{R}^n \mid v_j = 0\}$ such that $G' = \{(v,0) \in \mathbb{R}^{n+1} \mid v \in G''\}$. Let D be the characteristic set of G'' and let $G = \{v \in C \mid v_i = 0, \text{ for all } i \in \mathit{Irr} \setminus D\}$. We have $G'' = G \cap \{v \in \mathbb{R}^n \mid v_j = 0\}$. Let $G^0 \subseteq G$ be a minimal proper face of C and $D^0 \subseteq D$ its characteristic set. Since $j \in \mathit{Frev}$, there exists $g \in G^0 \setminus \mathrm{lin.space}(C)$ such that $g_j = 0$. Therefore, $g \in G^0 \cap \{v \in \mathbb{R}^n \mid v_j = 0\} \setminus \mathrm{lin.space}(C)$. Suppose there exists $k \in D \setminus D^0$. Then $v_k = 0$ for all $v \in G^0$ and $G^0 \subseteq G \cap \{v \in \mathbb{R}^n \mid v_k = 0\}$. Since G'' is a minimal proper face of $C \cap \{v \in \mathbb{R}^n \mid v_j = 0\}$, we have $G'' \cap \{v \in \mathbb{R}^n \mid v_k = 0\} = \mathrm{lin.space}(C \cap \{v \in \mathbb{R}^n \mid v_j = 0\})$. It follows that

$G^0 \cap \{v \in \mathbb{R}^n \mid v_j = 0\} \subseteq \text{lin.space}(C \cap \{v \in \mathbb{R}^n \mid v_j = 0\})$, in contradiction to $g \in G^0 \cap \{v \in \mathbb{R}^n \mid v_j = 0\} \setminus \text{lin.space}(C)$. We conclude that $D^0 = D$ and so G is a minimal proper face of C.

"$\Leftarrow$": Immediate. □

In summary, if reaction $j \in Frev$ is fully reversible, the minimal proper faces of C' are G^j, G^{n+1} and those which are in a 1-1 correspondence with the minimal proper faces of C. The dimension of the lineality space of C' decreases by one. Accordingly, the size of the flux cone description increases by one after splitting a fully reversible reaction.

5.1.2 Splitting a Pseudo-irreversible Reaction

If $j \in Prev_0$ is pseudo-irreversible, there is no flux distribution in the reconfigured network that involves reaction j (resp. $n+1$) and no other irreversible reactions. The following proposition shows that both (and only) reactions j and $n+1$ characterize a trivial minimal proper face of C' given by

$$G^c = \{(v, w) \in \mathbb{R}^{n+1} \mid Sv = w \cdot S_{*j},\ v_i = 0, \text{ for all } i \in Irr,\ v_j \geq 0,\ w \geq 0\}.$$

The minimal proper face G^c contains all the (2-cycle) flux distributions in the reconfigured network that involve only the forward and backward reactions j and $n+1$.

Proposition 5.5. *If $j \in Prev_0$ is pseudo-irreversible, then G^c is a minimal proper face of C' whose characteristic set is $D^c = \{j, n+1\}$.*

Proof. We have $G^c = \{v' \in C' \mid v'_j \geq 0,\ v'_i = 0, \text{ for all } i \in Irr\}$ and $\text{lin.space}(C') = \{v' \in C' \mid v'_j = 0,\ v'_i = 0, \text{ for all } i \in Irr\}$. Let $u \in \mathbb{R}^{n+1}$ with $u_j = u_{n+1} = 1$ and $u_i = 0$ for all $i \in \{1, \ldots, n\} \setminus \{j\}$. We have $u \in G^c \setminus \text{lin.space}(C')$ and so $\dim(G^c) = \dim(\text{lin.space}(C')) + 1$. Therefore, G^c is a minimal proper face characterized by reaction j. Since $u_{n+1} \neq 0$ and $u_i = 0$ for all $i \in Irr' \setminus \{j, n+1\}$, $D^c = \{j, n+1\}$ is the characteristic set of G^c. □

Let $G^1, \ldots, G^s$ be the minimal proper faces of C and $D^1, \ldots, D^s$ their characteristic sets, respectively. Starting from [33; 58] and using that reaction j is pseudo-irreversible, we partition the set $J = \{G^1, \ldots, G^s\}$ of minimal proper faces of C into three parts:

$$\begin{aligned} J^0 &= \{G \in J \mid v_j = 0 \text{ for all } v \in G\}, \\ J^+ &= \{G \in J \mid v_j > 0 \text{ for all } v \in G \setminus \text{lin.space}(C)\}, \\ J^- &= \{G \in J \mid v_j < 0 \text{ for all } v \in G \setminus \text{lin.space}(C)\}. \end{aligned}$$

From each of the sets J^0, J^+, J^- we will obtain different minimal proper faces of C'. We start by characterizing minimal proper faces G' with $v'_j = v'_{n+1} = 0$ for all $v' \in G'$. As will be stated in the next proposition, in addition to minimal proper faces

$G \in J^0$, some minimal proper faces of C' are obtained by combining particular pairs $(G^k, G^l) \in J^+ \times J^-$. The set Φ of these pairs is given by

$$\Phi = \{(G^k, G^l) \in J^+ \times J^- \mid D^i \nsubseteq D^k \cup D^l \text{ for all } i \in \{1, \ldots, s\} \setminus \{k, l\}\}. \quad (5.3)$$

Actually, $(G^k, G^l) \in \Phi$ means that the minimal proper faces G^k and G^l are adjacent in the flux cone C (see Definition 2.9 in Chap. 2 for a definition of adjacent minimal proper faces). Accordingly, each pair $(G^k, G^l) \in \Phi$ defines a minimal proper face of C'

$$\zeta(G^k, G^l) = \{v \in C \mid v_j = 0,\ v_i = 0, \text{ for all } i \in Irr \setminus D^k \cup D^l\}.$$

Finally, the set Adj of minimal proper faces of C' that are obtained by combining the pairs $(G^k, G^l) \in \Phi$ is given by

$$Adj = \{\zeta(G^k, G^l) \mid (G^k, G^l) \in \Phi\}. \quad (5.4)$$

Proposition 5.6. *Let $G' \subseteq C'$ such that $v'_j = v'_{n+1} = 0$ for all $v' \in G'$. If $j \in Prev_0$ is pseudo-irreversible, then the following are equivalent:*

- *G' is a minimal proper face of C'.*
- *There exists $G \in J^0 \cup Adj$ such that $G' = \{(v, 0) \in \mathbb{R}^{n+1} \mid v \in G\}$.*

Proof. According to Lemma 5.2, G' is a minimal proper face of C' if and only if there is a minimal proper face G of $C \cap \{v \in \mathbb{R}^n \mid v_j = 0\}$ such that $G' = \{(v, 0) \in \mathbb{R}^{n+1} \mid v \in G\}$. We show that G is a minimal proper face of $C \cap \{v \in \mathbb{R}^n \mid v_j = 0\}$ if and only if $G \in J^0 \cup Adj$. Since $j \in Prev_0$, we have $\text{lin.space}(C \cap \{v \in \mathbb{R}^n \mid v_j = 0\}) = \text{lin.space}(C)$.

"⇒": Let $G = \{v \in C \cap \{v \in \mathbb{R}^n \mid v_j = 0\} \mid v_i = 0, \text{ for all } i \in Irr \setminus D\}$. There exists $g \in G \setminus \text{lin.space}(C)$ such that $D = \{i \in Irr \mid g_i \neq 0\}$. Let $g^i \in G^i \setminus \text{lin.space}(C)$ for $i = 1, \ldots, s$. Since $g \in C$, g can be written in the form $g = \sum_{i=1}^{s} \alpha_i g^i + b$, for some $\alpha_i \geq 0$ and $b \in \text{lin.space}(C)$. Since $g \notin \text{lin.space}(C)$, there exists $k \in \{1, \ldots, s\}$ such that $\alpha_k \neq 0$. Accordingly, $D^k \subseteq D$. We have the following cases:

1. $G^k \in J^0$: Since $D^k \subseteq D$ and $G^k \subseteq C \cap \{v \in \mathbb{R}^n \mid v_j = 0\}$, we get $G^k \subseteq G$. Since G is a minimal proper face of $C \cap \{v \in \mathbb{R}^n \mid v_j = 0\}$ and $\text{lin.space}(C \cap \{v \in \mathbb{R}^n \mid v_j = 0\}) \subsetneq G^k$, we get $G^k = G$ and $G \in J^0$.

2. $G^k \in J^+$: Suppose $\alpha_i \neq 0$ implies $G^i \in J^+$ for all $i = 1, \ldots, s$. We get $g_j = \alpha_k g^k_j + \sum_{i \neq k} \alpha_i g^i_j > 0$, contradicting $g_j = 0$. Then there exists $l \in \{1, \ldots, s\}$ such that $\alpha_l \neq 0$ and $G^l \in J^-$. It follows that $D^l \subseteq D$ and $D^k \cup D^l \subseteq D$. Let $g' = g^l - (g^l_j / g^k_j) \cdot g^k$ and $G' = \{v \in C \mid v_j = 0,\ v_i = 0, \text{ for all } i \in Irr \setminus (D^k \cup D^l)\}$. We have $g' \in G' \setminus \text{lin.space}(C \cap \{v \in \mathbb{R}^n \mid v_j = 0\})$ and $G' \subseteq G$. Since G is a minimal proper face of $C \cap \{v \in \mathbb{R}^n \mid v_j = 0\}$, we get $G' = G$ and $D = D^k \cup D^l$. Suppose there exists $i \in \{1, \ldots, s\}$ such that $D^i \subseteq D^k \cup D^l$. If $G^i \in J^+$ (resp. $G^i \in J^-$), we prove in a similar way that $D^i \cup D^l = D^k \cup D^l$ (resp. $D^k \cup D^i = D^k \cup D^l$) and so $D^i = D^k$ (resp. $D^i = D^l$). It follows that $(G^k, G^l) \in \Phi$, $G = \zeta(G^k, G^l)$ and $G \in Adj$.

3. $G^k \in J^-$: The proof is similar to that of the case above.

"$\Leftarrow$": We can easily see that if $G \in J^0$, then G is a minimal proper face of $C \cap \{v \in \mathbb{R}^n \mid v_j = 0\}$. Suppose $G = \zeta(G^k, G^l)$ for some $(G^k, G^l) \in \Phi$. Let $G' \subseteq G$ be a minimal proper face of $C \cap \{v \in \mathbb{R}^n \mid v_j = 0\}$ and $D' \subseteq D^k \cup D^l$ its characteristic set. Accordingly, $G' \in J^0 \cup Adj$. Suppose $G' \in J^0$. It follows from $D' \subseteq D^k \cup D^l$ and $(G^k, G^l) \in \Phi$ that $D' = D^k$ or $D' = D^l$, contradicting $v_j = 0$ for all $v \in G'$. We conclude that $G' = \zeta(G^{k'}, G^{l'})$ for some $(G^{k'}, G^{l'}) \in \Phi$ and $D' = D^{k'} \cup D^{l'}$. Since $D' \subseteq D^k \cup D^l$, we get $D^{k'} \subseteq D^k \cup D^l$ and $D^{l'} \subseteq D^k \cup D^l$. Therefore, $D^{k'} = D^k$ and $D^{l'} = D^l$. We get $G' = G$ and so G is a minimal proper face of $C \cap \{v \in \mathbb{R}^n \mid v_j = 0\}$. □

Next, we characterize non-trivial minimal proper faces G' of C' with $v'_j > 0$ for all $v' \in G' \setminus \text{lin.space}(C')$.

Proposition 5.7. *Let $G' \subseteq C'$ such that $G' \neq G^c$. If $j \in Prev_0$ is pseudo-irreversible, then the following are equivalent:*

- *G' is a minimal proper face of C' such that $v'_j > 0$ for all $v' \in G' \setminus \text{lin.space}(C')$.*
- *There exists $G \in J^+$ such that $G' = \{(v, 0) \in \mathbb{R}^{n+1} \mid v \in G\}$.*

Proof. Suppose $j \in Prev_0$. Then, $\dim(\text{lin.space}(C')) = \dim(\text{lin.space}(C))$.

"$\Rightarrow$": Suppose G' is a minimal proper face of C' such that $v'_j > 0$ for all $v' \in G' \setminus \text{lin.space}(C')$ and let D' be its characteristic set. Since $G' \neq G^c$ and $j \in D'$, we have $n+1 \notin D'$ and $D' \setminus \{j\} \subseteq Irr$. Let $(g, 0) \in G' \setminus \text{lin.space}(C')$, $D = D' \setminus \{j\}$ and $G = \{v \in C \mid v_i = 0, \text{ for all } i \in Irr \setminus D\}$. We have $g \in G \setminus \text{lin.space}(C)$ and $g_j > 0$. Suppose there exists $v \in G \setminus \text{lin.space}(C)$ such that $v_j \leq 0$ and let $w = v - (v_j/g_j) \cdot g$. We have $(w, 0) \in G' \setminus \text{lin.space}(C')$ and $w_j = 0$, in contradiction to $v'_j > 0$ for all $v' \in G' \setminus \text{lin.space}(C')$. We conclude that $v_j > 0$ for all $v \in G \setminus \text{lin.space}(C)$ and $G' = \{(v, 0) \in \mathbb{R}^{n+1} \mid v \in G\}$. Accordingly, $\dim(G) = \dim(G')$. Since G' is a minimal proper face of C', we have $\dim(G') = \dim(\text{lin.space}(C \cap \{v \in \mathbb{R}^n \mid v_j = 0\})) + 1 = \dim(\text{lin.space}(C)) + 1$ and so $G \in J^+$.

"$\Leftarrow$": Let $G \in J^+$ such that $G' = \{(v, 0) \in \mathbb{R}^{n+1} \mid v \in G\}$. Since $\dim(G') = \dim(G)$ and $\dim(G) = \dim(\text{lin.space}(C)) + 1 = \dim(\text{lin.space}(C \cap \{v \in \mathbb{R}^n \mid v_j = 0\})) + 1$, we conclude that G' is a minimal proper face of C'. Since $v_j > 0$ for all $v \in G \setminus \text{lin.space}(C)$, it follows that $v'_j > 0$ for all $v' \in G' \setminus \text{lin.space}(C')$. □

Finally, we characterize non-trivial minimal proper faces $G' \neq G^c$ of C' with $v'_{n+1} > 0$ for all $v' \in G' \setminus \text{lin.space}(C')$. In such a case, $v'_j = 0$ for all $v' \in G'$ and the characteristic set of G' is $D \cup \{n+1\}$ for some $D \subseteq Irr$.

Proposition 5.8. *Let $D \subseteq Irr$ be a set of irreversible reactions. If $j \in Prev_0$ is pseudo-irreversible, then the following are equivalent:*

- *There exists a minimal proper face G' of C' whose characteristic set is $D \cup \{n+1\}$.*

– *There exists $G \in J^-$ whose characteristic set is D.*

Proof. Suppose $j \in Prev_0$. Then, $\dim(\text{lin.space}(C')) = \dim(\text{lin.space}(C))$.

"⇒": Suppose G' is a minimal proper face of C' whose characteristic set is $D \cup \{n+1\}$. Let $g' \in G' \setminus \text{lin.space}(C')$ and $g \in \mathbb{R}^n$ such that $g_i = g'_i$ for all $i \in \{1, \ldots, n\} \setminus \{j\}$ and $g_j = -g'_{n+1}$. Let $G = \{v \in C \mid v_i = 0, \text{ for all } i \in Irr \setminus D\}$. We have $g \in G \setminus \text{lin.space}(C)$ and $g_j < 0$. Suppose there exists $v \in G \setminus \text{lin.space}(C)$ such that $v_j \geq 0$ and let $w = v - (v_j/g_j) \cdot g$. We have $(w, 0) \in G' \setminus \text{lin.space}(C')$, in contradiction to $v'_{n+1} > 0$ for all $v' \in G' \setminus \text{lin.space}(C')$. We conclude that $v_j < 0$ for all $v \in G \setminus \text{lin.space}(C)$. To show $G \in J^-$, let $F \subseteq G$ be a minimal proper face of C and $D' \subseteq D$ its characteristic set. Let $f \in F \setminus \text{lin.space}(C)$ and $f' \in \mathbb{R}^{n+1}$ with $f'_i = f_i$ for all $i \in \{1, \ldots, n\} \setminus \{j\}$, $f'_j = 0$ and $f'_{n+1} = -f_j > 0$. Since $f' \in C'$ and $\{i \in Irr' \mid f'_i > 0\} = D \cup \{n+1\}$, we have $f' \in G' \setminus \text{lin.space}(C')$. Suppose there exists $k \in D \setminus D'$. Then $v_k = 0$ for all $v \in F$ and $F \subseteq G \cap \{v \in \mathbb{R}^n \mid v_k = 0\}$. Accordingly, $f \in G \cap \{v \in \mathbb{R}^n \mid v_k = 0\} \setminus \text{lin.space}(C)$ and $f' \in G' \cap \{v \in \mathbb{R}^n \mid v_k = 0\} \setminus \text{lin.space}(C')$. Since G' is a minimal proper face of C' and $k \in D$, $G' \cap \{v \in \mathbb{R}^n \mid v_k = 0\} = \text{lin.space}(C')$, contradicting $f' \in G' \cap \{v \in \mathbb{R}^n \mid v_k = 0\} \setminus \text{lin.space}(C')$. We conclude that $D' = D$, $F = G$ and so the claim follows.

"⇐": Let $G \in J^-$ such that D is its characteristic set. Let $G' = \{(v, w) \in \mathbb{R}^{n+1} \mid Sv = w \cdot S_{*j},\ v_i = 0, \text{ for all } i \in Irr \setminus D,\ v_i \geq 0, \text{ for all } i \in D,\ v_j = 0,\ w \geq 0\}$. Let $F' \subseteq G'$ be a minimal proper face of C' and $D' \subseteq D \cup \{n+1\}$ its characteristic set. Suppose $n+1 \notin D'$. Since $j \notin D'$, by Proposition 5.6, there exists $F \in J^0 \cup Adj$ such that $F' = \{(v, 0) \in \mathbb{R}^{n+1} \mid v \in F\}$. The characteristic set of F is D'. Then either $D' = D^i$ with $G^i \in J^0$ or $D' = D^k \cup D^l$ with $(G^l, G^k) \in \Phi$. Since $D' \subseteq D$, both cases are contradicting $G \in J^-$. We conclude that $n+1 \in D'$. Since $j \notin D'$, $F' \neq G^c$ and its characteristic set is $(D' \setminus \{n+1\}) \cup \{n+1\}$. There exists then $K \in J^-$ whose characteristic set is $D' \setminus \{n+1\}$. Since $D' \setminus \{n+1\} \subseteq D$ and both G and K are minimal proper faces of C, it follows that $K = G$, $D' = D \cup \{n+1\}$ and $F' = G'$. We conclude that G' is a minimal proper face of C' whose characteristic set is $D \cup \{n+1\}$. □

To summarize, a non-trivial minimal proper face G' of C' is given either by

$$G' = \{(v, 0) \in \mathbb{R}^{n+1} \mid v \in G\}, \text{ for some } G \in J^0 \cup J^+ \cup Adj,$$

or by

$$G' = \{v' \in C' \mid v'_i = 0 \text{ for all } i \in (Irr \cup \{j\}) \setminus D^k\}, \text{ for some } G^k \in J^-.$$

Since $\dim(\text{lin.space}(C')) = \dim(\text{lin.space}(C))$, it follows that the size of the flux cone description increases by $|Adj| + 1$ after splitting a pseudo-irreversible reaction. Note that the set Adj can be quite large (cf. Sect. 5.2).

5.1.3 Changes in the Reversibility Type of Reactions

Another consequence of the network reconfiguration is the change in the reversibility type of reactions. Indeed, possibly many fully reversible reactions in the original

network may become pseudo-irreversible in the reconfigured network. Let $Frev'$ and $Prev'_0$ be the sets of fully and pseudo-irreversible reversible reactions in the reconfigured network, respectively, i.e.,

$$Frev' = \{i \in Rev \setminus \{j\} \mid b'_i \neq 0, \text{ for some } b' \in \mathrm{lin.space}(C')\},$$
$$Prev'_0 = Rev \setminus (Frev'_0 \cup \{j\}).$$

Since $\mathrm{lin.space}(C') = \{(v,0) \in \mathbb{R}^{n+1} \mid v \in \mathrm{lin.space}(C),\ v_j = 0\}$, we have $Frev' \subseteq Frev \setminus \{j\}$ and $Prev_0 \setminus \{j\} \subseteq Prev'_0$. Let Δ be the set of fully reversible reactions of the original network which become pseudo-irreversible in the reconfigured network, i.e.,

$$\Delta = Frev \setminus (Frev' \cup \{j\}).$$

We can easily see that $\Delta = \{i \in Frev \setminus \{j\} \mid b_i = 0 \text{ for each } b \in \mathrm{lin.space}(C) \cap \{v \in \mathbb{R}^n \mid v_j = 0\}\}$. The following proposition further characterizes the set Δ using a basis of the lineality space of C.

Proposition 5.9. *Let $B = (b^1, \ldots, b^t)$ be a basis of the lineality space $\mathrm{lin.space}(C)$. Then,*

$$\Delta = \{i \in Frev \setminus \{j\} \mid \textit{there exists } \lambda \neq 0 \textit{ such that } b_i^k = \lambda b_j^k \textit{ for all } k = 1, \ldots, t\}.$$

Proof. Let $\Omega = \{i \in Frev \setminus \{j\} \mid \text{there exists } \lambda \neq 0 \text{ such that } b_i^k = \lambda b_j^k \text{ for all } k = 1, \ldots, t\}$.. Then $\Omega \subseteq \Delta$. To show the reverse inclusion, suppose $i \in \Delta$. Since $i \in Frev$, there exists $b \in B$ such that $b_i \neq 0$. Since $i \in \Delta$, we have $b_j \neq 0$. Let $b' \in B$ and let $w = b' - (b'_j/b_j) \cdot b$. We have $w \in \mathrm{lin.space}(C) \cap \{v \in \mathbb{R}^n \mid v_j = 0\}$ and $w_i = b'_i - (b_i/b_j)b'_j$. Since $i \in \Delta$, we get $w_i = 0$ and so $b'_i/b'_j = b_i/b_j \stackrel{\text{def}}{=} \lambda \neq 0$, independently from b'. □

Corollary 5.10. *If $j \in Prev_0$ is pseudo-irreversible, then $Frev' = Frev$ and $Prev'_0 = Prev_0 \setminus \{j\}$.*

Proof. Suppose $j \in Prev_0$. Then, $b_j^k = 0$ for all $k = 1, \ldots, t$. Consider $i \in Frev \setminus \{j\}$. There exists $b \in B$ such that $b_i \neq 0$. Since $b_j = 0$, it follows that $i \notin \Delta$. Therefore, $\Delta = \emptyset$ and the claim follows. □

5.2 From Outer to Inner Descriptions

In this section, we give a generic procedure to show how inner descriptions can be computed from the outer one. We use this procedure to explain why, for large-scale metabolic networks, the size of the inner descriptions may be several orders of magnitude larger than that of the outer description.

The results in Sect. 5.1 allow for obtaining an outer description of the reconfigured flux cone after splitting one reversible reaction. Now we are seeking for an inner description of the reconfigured flux cone after splitting a set $\mathrm{SR} = \{j_1, \ldots, j_p\}$ of reversible reactions. We propose an iterative procedure that splits, in each iteration k,

a reversible reaction, and obtains a minimal generating set of the reconfigured flux cone using the following scheme. Let (R^0, B^0) be a minimal generating set of the original flux cone. The set (R^0, B^0) can be computed using an existing software for polyhedral computations such as `cdd` [33]. For $1 \leq k \leq p$, let j_k be the reversible reaction to be split in iteration k and let Irr^{k-1}, $Prev_0^{k-1}$ and $Frev^{k-1}$ be the set of irreversible, pseudo-irreversible and fully reversible reactions after splitting reactions $j_1, \ldots, j_{k-1}$, respectively. Set $Irr^0 := \{1, \ldots, n\} \setminus Rev$, $Prev_0^0 := \{i \in Rev \mid b_i = 0 \text{ for all } b \in B^0\}$ and $Frev^0 := Rev \setminus Prev_0^0$. Iteration k comprises two basic steps. First, we deduce a minimal generating set (B^k, R^k) from (B^{k-1}, R^{k-1}) based on the results given in Sect. 5.1. This step is straightforward if $j_k \in Frev^{k-1}$. In such a case, the inner description of the reconfigured flux cone increases by one. However, if $j_k \in Prev_0^{k-1}$, in addition to the generators we can directly deduce from (B^{k-1}, R^{k-1}), the inner description of the reconfigured flux cone includes a subset $\Psi \subseteq R^{k-1} \times R^{k-1}$ that contains possibly many generators. In this case, the increase in the inner description is equal to $|\Psi| + 1$. In the second step, we update the reversibility type of reactions using Proposition 5.9. The deduction procedure terminates in iteration p and an inner description of the reconfigured flux cone is (B^p, R^p). For a more detailed description, see Algorithm 1.

Algorithm 1 Procedure for deducing an inner description from an outer description of the flux cone.

Input:
- Set of reversible reactions $Rev \subseteq \{1, \ldots, n\}$;
- Set of reversible reactions to be split SR $= \{j_1, \ldots, j_p\}$;
- Set of minimal proper faces of the flux cone J;
- Lineality space of the flux cone L.

Output:
- Minimal generating set GenSet of the reconfigured flux cone.

Initialization: $R^0 := \{g \in G \setminus L \mid G \in J\}$, $B^0 = (b^1, \ldots, b^t)$ a vector basis of L, $Prev_0^0 := \{i \in Rev \mid b_i = 0 \text{ for all } b \in B^0\}$, $Frev^0 := Rev \setminus Prev_0^0$, $Irr^0 := \{1, \ldots, n\} \setminus Rev$.

for all $k \in \{1, \ldots, p\}$ **do**
  if $j_k \in Frev^{k-1}$ **then**
    /* Deduce a minimal generating set of minimal proper faces */
    Choose $u \in B^{k-1}$ such that $u_{j_k} > 0$,
    add$((u, 0), R^k)$;
    Define $w \in \mathbb{R}^{n+k}$ by $w_i' := -u_i$ for all $i \in \{1, \ldots, n+k-1\} \setminus \{j_k\}$, $w_{j_k} := 0$ and $w_{n+k} := u_{j_k}$,
    add(w, R^k);
    for all $g \in R^{k-1}$ **do**
      add$((g - (g_{j_k}/u_{j_k}) \cdot u, 0), R^k)$;
    end for

Continued from previous page

/* Deduce a minimal generating set of the lineality space */
for all $b \in B^{k-1} \setminus \{u\}$ **do**
add($(b - (b_{j_k}/u_{j_k}) \cdot u, 0), B^k$);
end for

/* Update the reversibility type of reactions */
$\Delta := \{i \in Frev^{k-1} \setminus \{j_k\} \mid$ there exists $\lambda \neq 0$ such that $b_i = \lambda b_{j_k}$ for all $b \in B^{k-1}\}$,
$Frev^k := Frev^{k-1} \setminus (\Delta \cup \{j_k\})$, $Prev_0^k := Prev_0^{k-1} \cup \Delta$.
else
/* Deduce a minimal generating set of minimal proper faces */
Define $w \in \mathbb{R}^{n+k}$ by $w_i' := 0$ for all $i \in \{1, \ldots, n+k-1\} \setminus \{j_k\}$,
$w_{j_k} := w_{n+k} := 1$,
add(w, R^k),
$P := \{g \in R^{k-1} \mid g_{j_k} > 0\}$, $N := \{g \in R^{k-1} \mid g_{j_k} < 0\}$, $Z := \{g \in R^{k-1} \mid g_{j_k} = 0\}$.
for all $g \in P \cup Z$ **do**
add($(g, 0), R^k$).
end for
for all $g \in N$ **do**
Define $g' \in \mathbb{R}^{n+k}$ by $g_i' := g_i$ for all $i \in \{1, \ldots, n+k-1\} \setminus \{j_k\}$,
$g_{j_k}' := 0$ and $g_{n+k}' := -g_{j_k}$,
add(g', R^k).
end for
$\Psi := \{(g^1, g^2) \in P \times N \mid \{i \in Irr^{k-1} \mid g_i > 0\} \not\subseteq \{i \in Irr^{k-1} \mid g_i^1 + g_i^2 > 0\}$ for all $g \in R^{k-1} \setminus \{g^1, g^2\}\}$,
for all $(g^1, g^2) \in \Psi$ **do**
add($(g^2 - (g_{j_k}^2/g_{j_k}^1) \cdot g^1, 0), R^k$).
end for

/* Deduce a minimal generating set of the lineality space */
for all $b \in B^{k-1}$ **do**
add($(b, 0), B^k$).
end for

/* Update the reversibility type of reactions */
$Frev^k := Frev^{k-1}$, $Prev_0^k := Prev_0^{k-1} \setminus \{j_k\}$.
end if
$Irr^k := Irr^{k-1} \cup \{j_k, n+k\}$.
end for
GenSet := (B^p, R^p).

Note that, for the reconfigured flux cone to be pointed, we must have $p \geq t$, where t is the dimension of the lineality space of the original flux cone, i.e., $t = \dim(\text{lin.space}(C))$. This is typically the case for the extreme pathway and extremal current approaches. In such a case, we have

$$\eta \stackrel{\text{def}}{=} |\{k \in \{1, \ldots, p\} \mid j_k \in Prev_0^{k-1}\}| = p - t.$$

Accordingly, the above procedure contains η iterations where the increase in the inner description of the flux cone is significant. This explains why, for large-scale metabolic networks, the size of the inner descriptions may be several orders of magnitude larger than that of the outer description.

Metabolic network	Network size				Outer description size		Inner description size		
	Met	Irr	Rev-Int	Rev-Ext	RMS	MMB	EM	EP	EC
Chloroplast stroma [80]	19	9	12	3	0	11	15	27	30
Human red blood cell [121]	38	18	17	15	1	48	3557	127	3590
Saccharomyces cerevesiae [17]	48	30	17	0	0	657	8726	8743	8743
Escherichia coli [48]	90	83	27	1	0	3560	507632	?	?
Purple bacteria [48]	77	61	24	3	2	12	393524	?	?

Table 5.1: Metabolic networks, with the number of internal metabolites (Met), the number of irreversible (Irr) reactions, the number of reversible internal (Rev-Int) and external (Rev-ext) reactions, the number of minimal metabolic behaviors (MMB), the dimension of the reversible metabolic space (RMS), the number of elementary modes (EM), extreme pathways (EP), and extremal currents (EC). "?" indicates that the existing implementation of `cdd` has failed in the computation of the inner description. This is not the case for the computation of the outer description, showing that the network reconfiguration renders more complex the constraint system that defines the reconfigured flux cone. Except the 2-cycles corresponding to the split reactions, the set of EPs corresponds to a subset of the set of EMs, which is equivalent to the set of ECs.

Tab. 5.1 shows the sizes of the inner and outer descriptions of the flux cone of some typical metabolic networks. The computation of the extreme pathways, extremal currents, the minimal metabolic behaviors and the reversible metabolic space was done using the software `cdd` [33]. For computing the elementary flux modes, we used METATOOL [117]. We can see that the size of the outer description, given as the sum of the number of MMBs and dim(RMS), is typically much smaller than the number of elementary flux modes, extreme pathways and extremal currents. This observation holds even if the flux cone is pointed. In such a case, the MMBs correspond to the set of extreme rays of the flux cone. The extreme pathways and extremal currents are extreme for the only reason that the split reversible reactions have been decomposed into forward and backward reactions. In the initial cone, these extreme rays are conically dependent and their numbers are much larger than the number of MMBs.

CHAPTER

6 A New Approach For Flux Coupling Analysis

The stoichiometric and thermodynamic constraints not only determine all possible flux distributions over a metabolic network at steady state. They also induce different dependencies between the reactions. For example, some reactions are incapable of carrying flux under steady-state conditions. Furthermore, in several metabolic networks, a zero flux through one reaction implies a zero flux through other reactions. Since constraining the flux through some reaction to be equal to zero corresponds to the deletion of the corresponding gene, such dependencies also link the metabolic to the gene regulatory network. Accordingly, flux coupling analysis [11; 14; 69], which seeks for the elucidation of blocked and coupled reactions, helps to better understand metabolic interactions within cellular networks.

In this chapter, we present a new approach for flux coupling analysis, based on our constrained-based approach introduced in Chap. 4. Using our reaction classification, we study mathematical dependencies between coupling relationships and the reversibility type of the reactions. We show that coupling relationships can only hold between certain reaction types. These results not only allow for improving an existing algorithm, but also lead to a new algorithm for flux coupling analysis. Parts of this chapter have been published in [59].

6.1 Definitions

Under stoichiometric and thermodynamic constraints, some reactions are unable of carrying any flux, i.e., their fluxes are always equal to zero. Such reactions, which are called *blocked reactions* [14] or *strictly detailed balanced reactions* [101], are often not relevant and may be caused by omission and/or errors in the model reconstruction process. According to [14], for the *E. coli* metabolic network [28], 14% of the 740 reactions are blocked, whereas 29% of the 1173 reactions in the *S. cerevisiae* metabolic network [32] are unable of carrying any flux.

In the following, we formally define blocked reactions using the steady-state flux cone. Remember that, given a stoichiometric matrix S and a set of irreversible reactions Irr, the steady-state flux cone is given by

$$C = \{v \in \mathbb{R}^n \mid Sv = 0,\ v_i \geq 0, \text{for all } i \in Irr\}.$$

Definition 6.1 (Blocked reactions)**.** *Given the flux cone C, the set*

$$Blk = \{i \in \{1, \ldots, n\} \mid v_i = 0 \text{ for all } v \in C\}$$

is called the set of blocked reactions. The remaining reactions are called unblocked.

In our flux coupling analysis, we start by identifying blocked reactions. Afterwards, we determine dependencies between unblocked reactions using Definition 6.2. But first, we check whether all reactions in the network are blocked, or equivalently, we check whether the flux cone is trivial. This is the case if the two following statements hold:

1. The flux cone C is equal to its lineality space, i.e., $C = \text{lin.space}(C)$, or equivalently, all irreversible reactions are blocked.

2. The lineality space $\text{lin.space}(C)$ is trivial, i.e., $\text{lin.space}(C) = \{0\}$.

To check condition (1), consider the following LP problem

$$\max \{ \sum_{j \in Irr} v_j : Sv = 0,\ 0 \leq v_j \leq 1 \text{ for all } j \in Irr\}. \tag{6.1}$$

Let v^* be an optimal solution of the above LP problem. All irreversible reactions are blocked if and only if $\sum_{j \in Irr} v_j^* = 0$. If this is the case, the flux cone C has no minimal proper face, and so is equal to its lineality space, i.e., $C = \text{lin.space}(C)$.

Now, to check condition (2), let $I \in \mathbb{R}^{n \times n}$ be the identity matrix and L be the matrix given by

$$L = \begin{pmatrix} S \\ I_{Irr*} \end{pmatrix}.$$

Since $\text{lin.space}(C)$ is the null space of the matrix L, i.e., $\text{lin.space}(C) = kern(L)$, $\text{lin.space}(C) = \{0\}$ if the matrix L has full rank, i.e., $\text{rank}(L) = n$.

In the rest of this chapter, we assume that the flux cone is not trivial, and so some reactions are unblocked.

Definition 6.2. *Let i, j be two unblocked reactions. The coupling relationships $\overset{=0}{\rightarrow}, \overset{=0}{\leftrightarrow}, \backsim^{\lambda}$ are defined in the following way:*

- $i \overset{=0}{\rightarrow} j$ *if for all* $v \in C$, $v_i = 0$ *implies* $v_j = 0$.

- $i \overset{=0}{\leftrightarrow} j$ *if for all* $v \in C$, $v_i = 0$ *is equivalent to* $v_j = 0$.

- $i \backsim^{\lambda} j$ *if there exists* $\lambda \in \mathbb{R}$ *such that for all* $v \in C, v_j = \lambda v_i$.

Reactions i and j are coupled if at least one of the relations $i \overset{=0}{\rightarrow} j$, $i \overset{=0}{\leftrightarrow} j$ or $i \backsim^{\lambda} j$ holds. Otherwise, i and j are uncoupled.

Note that $i \backsim^{\lambda} j$ (resp. $i \overset{=0}{\leftrightarrow} j$) is equivalent to $j \backsim^{\lambda} i$ (resp. $j \overset{=0}{\leftrightarrow} i$). Hence, when looking for reaction pairs (i, j) fulfilling $i \backsim^{\lambda} j$ or $i \overset{=0}{\leftrightarrow} j$, we should restrict ourselves to reaction pairs (i, j) with $i < j$. In addition, $i \backsim^{\lambda} j$ implies $i \overset{=0}{\leftrightarrow} j$, which in turn is equivalent to $(i \overset{=0}{\rightarrow} j$ and $j \overset{=0}{\rightarrow} i)$. Therefore, we need to check whether $i \overset{=0}{\leftrightarrow} j$ holds only if $i \backsim^{\lambda} j$ does not hold. Similarly, we need to check whether $i \overset{=0}{\rightarrow} j$ holds only if $i \backsim^{\lambda} j$ and $j \overset{=0}{\leftrightarrow} i$ do not hold.

Using the notations from Definitions 6.1 and 6.2, we formally define the problem of *flux coupling analysis (FCA)* as follows:

Flux Coupling Analysis (FCA)

Given: $S \in \mathbb{R}^{m \times n}$ stoichiometric matrix,
$Irr \subseteq \{1, \ldots, n\}$ set of irreversible reactions.
Find: $\text{Blk} = \{i \mid i \text{ is blocked}\}$,
$A = \{(i, j) \mid i \backsim^{\lambda} j, 1 \leq i < j \leq n, i, j \notin \text{Blk}\}$,
$B = \{(i, j) \mid i \overset{=0}{\leftrightarrow} j, 1 \leq i < j \leq n, i, j \notin \text{Blk}, (i, j) \notin A\}$,
$C = \{(i, j) \mid i \overset{=0}{\rightarrow} j, (i, j) \notin \text{Blk}^2, (i, j) \notin A \cup B, (j, i) \notin A \cup B\}$.

6.2 The FCF Algorithm

The *Flux Coupling Finder (FCF)* algorithm [14] has been developed in an attempt to identify blocked reactions as well as coupled reactions in metabolic networks. This algorithm requires the solution of a sequence of linear programming (LP) problems. In contrast to the FCA formulation given in the preceding section, the FCF algorithm requires that each reversible reaction is split into a forward and a backward reaction, which both are constrained to be irreversible. Fig. 6.1 shows a simple hypothetical network before and after reconfiguration. The reversible reaction 2 is split into a forward and a backward reaction 2^+ and 2^-, which both are irreversible.

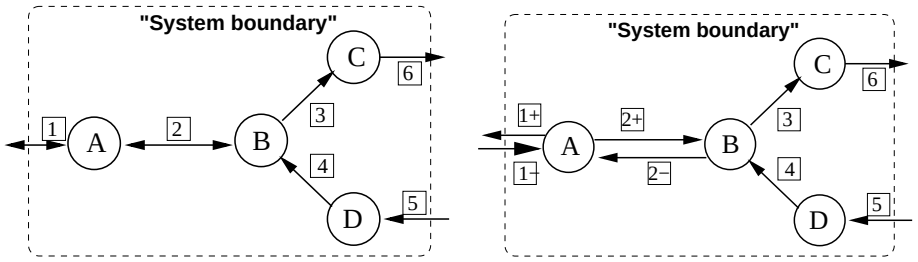
Let $Rev = \{i_1, \ldots, i_{|Rev|}\}$. For convenience, the stoichiometric matrix $S' \in \mathbb{R}^{m*(n+|Rev|)}$ of the reconfigured network can be written as follows:

$$\begin{aligned} S'_{*j} &= S_{*j} && \text{for all } j \in \{1, \ldots, n\}, \\ S'_{*(n+j)} &= -S_{*i_j} && \text{for all } j \in \{1, \ldots, |Rev|\}. \end{aligned}$$

All reactions in the reconfigured network are irreversible. Given $Irr' = \{1, \ldots, n + |Rev|\}$, the reconfigured flux cone C^{rec}, which contains all possible steady-state flux distributions in the reconfigured network, is given by

$$C^{rec} = \{d \in \mathbb{R}^{n+|Rev|} \mid S'd = 0,\ d_i \geq 0, \text{for all } i \in Irr'\}. \tag{6.2}$$

Actually, each reaction in the reconfigured network corresponds either to an irreversible reaction or to a possible direction of a reversible reaction in the original network. In both cases, a reaction in the reconfigured network corresponds to a direction of some reaction in the original network. Therefore, it is suitable to call a reaction



(a) **Network example** (b) **Reconfigured network**

Figure 6.1: The reversible reaction 2 is split into a forward and backward reaction 2^+ and 2^-. According to the FCF algorithm, a zero flux through direction 3 (resp. 4) implies a zero flux through direction 2^+ (resp. 2^-), i.e., a negative (resp. positive) flux through the reversible reaction 2. However, neither a zero flux through reaction 3 nor a zero flux through reaction 4 implies a zero flux through reaction 2.

in the reconfigured network a *reaction direction*, or shortly a *direction*. Accordingly, the FCF algorithm allows for identifying blocked directions as well as dependencies between directions.

The FCF algorithm decides whether a direction is blocked by maximizing the flux through that direction under the constraints defining the reconfigured cone C^{rec}. Indeed, for each direction $j \in \{1, \ldots, n + |Rev|\}$, the FCF algorithm uses the following LP problem

$$\max \{d_j : S'd = 0,\ d_i \geq 0, \text{for all } i \in Irr',\ d_i \leq d_i^{max} \text{ for all } i \in \text{BR}\}, \qquad (6.3)$$

where BR is the set of boundary reactions and for each $i \in$ BR, d_i^{max} is an upper bound on the flux in direction i. Because of these bounds, the above LP is finite. If the maximum value of the LP is zero, then direction j is blocked, i.e., $d_j = 0$ for all $d \in C^{rec}$.

We can easily see that, for each irreversible reaction $j \in Irr$, the following statements are equivalent:

1. $v_j = 0$ for all $v \in C$,
2. $d_j = 0$ for all $d \in C^{rec}$.

Accordingly, an irreversible reaction $j \in Irr$ is blocked if and only if direction j is blocked. Therefore, the set of irreversible reactions that are blocked can be identified by solving the LP problem (6.3) once for every irreversible reaction. Similarly, for each reversible reaction $i_j \in Rev$, the following statements are equivalent:

1. $v_{i_j} = 0$ for all $v \in C$,
2. $d_{i_j} = d_{n+j} = 0$ for all $d \in C^{rec}$.

Accordingly, a reversible reaction $i_j \in Rev$ is blocked if and only if directions i_j and $n + j$ are blocked. Hence, reversible reactions that are blocked can be identified by

solving the LP problem (6.3) twice for every reversible reaction. We conclude that the FCF algorithm requires the solution of $|Irr| + 2|Rev|$ LP problems to identify blocked reactions.

To decide whether two directions $i, j \in \{1, \ldots, n + |Rev|\}$ are coupled amounts to determining the upper and lower bounds R_{max} and R_{min} such that $0 \leq R_{min} d_j \leq d_i \leq R_{max} d_j$ for all $d \in C^{rec} \cap \{d \in \mathbb{R}^{n+|Rev|} \mid d_i \leq d_i^{max} \text{ for all } i \in \text{BR}\}$. The authors of [14] showed that R_{max} and R_{min} are the optimal values for maximizing and minimizing the following LP problems

$$\max \{d_i : S'd = 0,\ d_j = 1,\ d_k \geq 0, \text{for all } k \in Irr',\ d_l \leq d_i^{max} t \text{ for all } l \in \text{BR},\ t \geq 0\},$$
$$\min \{d_i : S'd = 0,\ d_j = 1,\ d_k \geq 0, \text{for all } k \in Irr',\ d_l \leq d_i^{max} t \text{ for all } l \in \text{BR},\ t \geq 0\},$$

respectively. Comparison of R_{max} and R_{min} allows the FCF algorithm to determine whether directions i and j are coupled using the following rules:

- $d_i = 0$ implies $d_j = 0$ for all $d \in C^{rec}$ if and only if $R_{min} \neq 0$,
- $d_j = 0$ implies $d_i = 0$ for all $d \in C^{rec}$ if and only if $R_{max} \neq +\infty$,
- $d_j = \lambda d_i$ for all $d \in C^{rec}$ if and only if $R_{min} = R_{max} = \lambda \neq 0$.

Accordingly, the FCF algorithm needs the solution of at most $(|Irr| + 2|Rev|) \cdot (|Irr| + 2|Rev| - 1)$ LP problems to find all directions that are coupled. Altogether, this algorithm requires the solution of $(|Irr| + 2|Rev|)^2$ LP problems to identify blocked and coupled directions.

While the FCF algorithm has proved successful in computing blocked and coupled reactions in some genome-scale metabolic networks [14], this algorithm may be hampered by the reconfiguration of the metabolic network. This reconfiguration implies that the number of variables (resp. constraints) increases by $|Rev|$ (resp. $2|Rev|$). Since the FCF algorithm uses linear programming (LP) to identify the maximum and minimum flux ratios for every pair of directions, a very big number of LP problems has to be solved. In addition, as another consequence of the network reconfiguration, the FCF algorithm does not compute directly coupling relationships between reactions. For instance, if we consider again the hypothetical network given in Fig. 6.1, according to the FCF algorithm, a zero flux through direction 3 (resp. 4) implies a zero flux through direction 2^+ (resp. 2^-), i.e., a negative (resp. positive) flux through the reversible reaction 2. However, neither a zero flux through reaction 3 nor a zero flux through reaction 4 implies a zero flux through reaction 2. Therefore, a post-processing step is needed to deduce couplings between reactions (in the original network) from those between directions (in the reconfigured network). In addition, the FCF algorithm explores exhaustively all possible reaction pairs. This leads to a very big number of LP problems that have to be solved. This strategy may not scale well for genome-scale models of complex microorganisms which involve a large number of reactions. In the next section, we show that coupling relationships depend on the reversibility type of reactions. For instance, irreversible and pseudo-irreversible reactions cannot be coupled with fully reversible reactions. Hence, when looking for coupled reactions,

we do not have to explore exhaustively all possible reaction pairs. We can improve the FCF procedure significantly by applying linear programming only in those cases where coupling relationships can occur.

6.3 Flux Coupling Analysis Based on the Reversibility Type of Reactions

For the rest of this chapter, we assume that $G^k, k = 1, \ldots, s$, are the minimal proper faces of the steady-state flux cone C, represented by vectors $g^k \in G^k \setminus \mathrm{lin.space}(C)$, and that $b^l, l = 1, \ldots, t$, is a vector basis of $\mathrm{lin.space}(C)$. Accordingly, for all $v \in C$, there exist $\alpha_k, \beta_l \in \mathbb{R}, \alpha_k \geq 0$ such that

$$v = \sum_{k=1}^{s} \alpha_k g^k + \sum_{l=1}^{t} \beta_l b^l. \tag{6.4}$$

In the following, we show that the reversibility type is an important key to elucidate reaction couplings. Remember that a reversible reaction $j \in Rev$ is called pseudo-irreversible if $v_j = 0$, for all $v \in \mathrm{lin.space}(C)$. A reversible reaction that is not pseudo-irreversible is called fully reversible. Moreover, given a minimal proper face G of the flux cone C and a reaction $j \in \{1, \ldots, n\}$, we have the following properties (see Chap. 4 for more details):

- If $j \in Irr$ is irreversible, then $v_j > 0$, for all $v \in G \setminus \mathrm{lin.space}(C)$, or $v_j = 0$, for all $v \in G$. Furthermore, $v_j = 0$, for all $v \in \mathrm{lin.space}(C)$.
- If $j \in Rev$ is pseudo-irreversible, then the flux v_j through j has a unique sign in $G \setminus \mathrm{lin.space}(C)$, i.e., either $v_j > 0$, for all $v \in G \setminus \mathrm{lin.space}(C)$, or $v_j = 0$, for all $v \in G \setminus \mathrm{lin.space}(C)$, or $v_j < 0$, for all $v \in G \setminus \mathrm{lin.space}(C)$. For all $v \in \mathrm{lin.space}(C)$, we have again $v_j = 0$.
- If $j \in Rev$ is fully reversible, there exists $v \in \mathrm{lin.space}(C)$ such that $v_j \neq 0$. We can then find pathways $v^+, v^-, v^0 \in G \setminus \mathrm{lin.space}(C)$ with $v_j^+ > 0$, $v_j^- < 0$ and $v_j^0 = 0$.

Based on the properties given above, we define the following decomposition of the reaction set $\{1, \ldots, n\}$, which reflects that pseudo-irreversible reactions taking the same direction in all minimal proper faces behave like irreversible reactions.

- $Irev = Irr \cup \{i \mid i$ is pseudo-irreversible and $v_i \geq 0,$ for all $v \in C$ or $v_i \leq 0,$ for all $v \in C\}$,
- $Prev = \{i \mid i$ is pseudo-irreversible and there exist $v^+, v^- \in C$ such that $v_i^+ > 0, v_i^- < 0\}$,
- $Frev = \{i \mid i$ is fully reversible$\}$.

We should mention that the above sets are disjoint and their union is equal to the set of all reactions, i.e., $Irev \cup Prev \cup Frev = \{1, \ldots, n\}$. Furthermore, for the set Blk of blocked reactions, we have Blk $\cap$ $Frev$ = Blk $\cap$ $Prev = \emptyset$, and so Blk $\subseteq Irev$.

Example 6.3. *Consider the hypothetical network depicted in Fig. 6.2. It consists of thirteen metabolites $(A, \ldots, O)$, and nineteen reactions $(1, \ldots, 19)$. The steady-state flux cone is defined by $C = \{v \in \mathbb{R}^{19} \mid Sv = 0, v_i \geq 0$ for all $i \in Irr\}$, with the stoichiometric matrix S and the set of irreversible reactions $Irr = \{2, 3, 4, 5, 6, 7, 8\}$. Fig. 6.2 shows four flux vectors*

$$\begin{aligned}
g^1 &= (2,\ 2,\ 1,\ 0,\ 0,\ 0,\ 0,\ 0,\quad 2,\quad 2,\quad 1,\quad 1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0),\\
g^2 &= (0,\ 0,\ 1,\ 2,\ 0,\ 0,\ 0,\ 0,\quad 0,\quad 0,\ -1,\ -1,\ 2,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0),\\
g^3 &= (0,\ 0,\ 0,\ 0,\ 1,\ 1,\ 1,\ 0,\ -1,\ -1,\quad 0,\quad 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0),\\
g^4 &= (0,\ 0,\ 0,\ 0,\ 1,\ 1,\ 0,\ 1,\ -1,\ -1,\quad 0,\quad 0,\ 0,\ 0,\ 1,\ 1,\ 0,\ 0,\ 0),
\end{aligned}$$

representing the four minimal proper faces $G^k, k = 1, 2, 3, 4$ of the network. The lineality space $\mathrm{lin.space}(C) = \{v \in C \mid v_i = 0, i \in Irr\}$ *has dimension 2. It can be generated by the pathways*

$$\begin{aligned}
b^1 &= (0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ -1,\ 1,\ 1,\ 1,\ 0,\ 0,\ 0),\\
b^2 &= (0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ -1,\ 1,\ 0,\ 0,\ 1,\ 1,\ 0).
\end{aligned}$$

An arbitrary flux vector $v \in C$ can be written in the form combination $v = \sum_{k=1}^{4} \alpha_k g^k + \sum_{l=1}^{2} \beta_l b^l$, for some $\alpha_k \geq 0$ and $\beta_1, \beta_2 \in \mathbb{R}$.

Reaction 1 is pseudo-irreversible and operates in the forward direction in all minimal proper faces, and so $1 \in Prev$. For this network, we have the following decomposition of the reaction set:

$$\begin{aligned}
Irev &= \{1, 2, 3, 4, 5, 6, 7, 8, 19\},\\
Prev &= \{9, 10, 11, 12\},\\
Frev &= \{13, 14, 15, 16, 17, 18\}.
\end{aligned}$$

First, we characterize blocked reactions using generators of the cone. The following proposition follows directly from equation (6.4).

Proposition 6.4. *For any reaction $i \in \{1, \ldots, n\}$, the following are equivalent:*

1. *The reaction i is blocked.*
2. *$g_i^k = 0$, for all $k = 1, \ldots, s$, and $b_i^l = 0$, for all $l = 1, \ldots, t$.*

Example 6.5. *Reaction 19 is blocked since flux balance around metabolite O implies that $v_{19} = 0$ for all $v \in C$.*

The next results shows that the relations $i \overset{=0}{\rightarrow} j$, $i \overset{=0}{\leftrightarrow} j$, $i \backsim^{\lambda} j$ cannot hold for arbitrary pairs of reactions.

Theorem 6.6. *Let i, j be two unblocked reactions such that at least one of the relations $i \overset{=0}{\rightarrow} j$, $i \overset{=0}{\leftrightarrow} j$ or $i \backsim^{\lambda} j$ is satisfied. Then either (a) or (b) holds:*

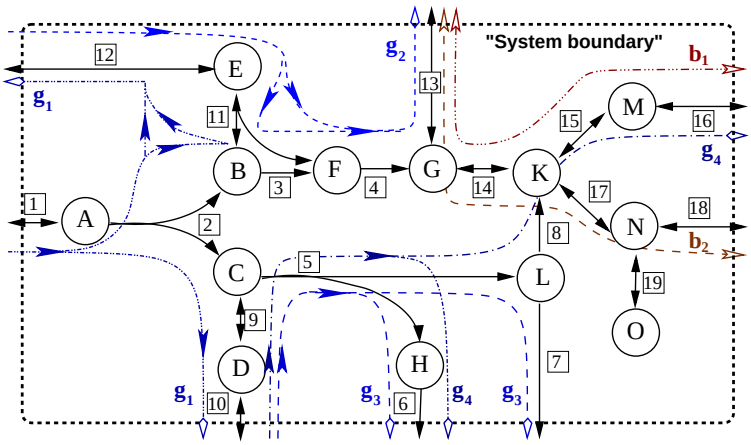


Figure 6.2: Network example with representative pathways

(a) *i and j are both (pseudo-) irreversible: $i, j \in Irev \cup Prev$.*

(b) *i and j are both fully reversible: $i, j \in Frev$.*

Proof. First suppose $i \in Irev \cup Prev$ and $j \in Frev$. Since $j \in Frev$, there exists $v \in \mathrm{lin.space}(C)$ such that $v_j \neq 0$. Since $i \in Irev \cup Prev$, we have $v_i = 0$, so $i \overset{=0}{\nrightarrow} j$.

Now suppose that $i \in Frev$ and $j \in Irev \cup Prev$. Since j is unblocked, there exists $w \in C$ such that $w_j \neq 0$. Since i is fully reversible, there exists $b \in \mathrm{lin.space}(C)$ such that $b_i \neq 0$. Define $v = w - (w_i/b_i) \cdot b$. It follows $v \in C$, $v_i = 0$ and $v_j = w_j \neq 0$, which implies $i \overset{=0}{\nrightarrow} j$. □

In the following, we study the coupling relationships for the different types of reactions. We first consider the case $i \in Prev$.

Proposition 6.7. *Suppose i, j are unblocked, $i \in Prev$ and $j \in Irev \cup Prev$. Then the following are equivalent:*

1. $i \overset{=0}{\rightarrow} j$.
2. $i \overset{=0}{\leftrightarrow} j$.
3. $i \backsim^{\lambda} j$.
4. $g_j^k = \lambda g_i^k$, *for all* $k = 1, \ldots, s$.

In each of these cases, $j \in Prev$.

Proof. (3) ⇒ (2) ⇒ (1) is immediate.

(1) ⇒ (4): Let $K^+ = \{k \mid g_i^k > 0\}$ and $K^- = \{k \mid g_i^k < 0\}$. Since $i \in Prev$, there exist $v^+, v^- \in C$ with $v_i^+ > 0$ and $v_i^- < 0$. If $K^+ = \emptyset$ (resp. $K^- = \emptyset$), we would have $v_i \leq 0$ (resp. $v_i \geq 0$), for all $v \in C$, which is a contradiction. So both K^+ and K^- must be non-empty. Let $p \in K^+$ and $q \in K^-$. Define $w = g_i^p \cdot g^q - g_i^q \cdot g^p$. Then $w \in C$

and $w_i = 0$. Since $i \overset{=0}{\to} j$, we get $w_j = g_i^p g_j^q - g_i^q g_j^p = 0$, or $g_j^p/g_i^p = g_j^q/g_i^q \overset{\text{def}}{=} \lambda$, independently of the choice of p and q. We conclude $g_j^k = \lambda g_i^k$, for all $k \in K^+ \cup K^-$. Since $i \overset{=0}{\to} j$, this holds for all $k = 1, \ldots, s$.

(4) $\Rightarrow$ (3): For all $v \in C$, there exists $b \in \text{lin.space}(C)$ and $\alpha_k \geq 0$ such that $v = \sum_{k=1}^s \alpha_k g^k + b$. Since $i \in Prev$ and $j \in Irev \cup Prev$, we have $b_i = b_j = 0$. It follows that $v_j = \sum_{k=1}^s \alpha_k g_j^k = \sum_{k=1}^s \alpha_k \lambda g_i^k = \lambda v_i$. □

Next, we characterize the case $i \in Frev$.

Proposition 6.8. *Suppose i, j are unblocked and $i \in Frev$ is fully reversible. Then the following are equivalent:*

1. $i \overset{=0}{\to} j$.
2. $i \overset{=0}{\leftrightarrow} j$.
3. $i \backsim^{\lambda} j$.
4. $g_j^k = \lambda g_i^k$, *for all* $k = 1, \ldots, s$, *and* $b_j^l = \lambda b_i^l$, *for all* $l = 1, \ldots, t$.

In each of these cases, $j \in Frev$.

Proof. (3) $\Rightarrow$ (2) $\Rightarrow$ (1) and (3) $\Leftrightarrow$ (4) are immediate.

To prove (1) $\Rightarrow$ (3), we suppose $i \overset{=0}{\to} j$. Since i is fully reversible, there exists $b \in \text{lin.space}(C)$, with $b_i \neq 0$. Given $v \in C$, define $w = v - (v_i/b_i) \cdot b$. Then $w \in C$ and $w_i = 0$. Since $i \overset{=0}{\to} j$, we get $w_j = v_j - (v_i/b_i)b_j = v_j - (b_j/b_i)v_i = 0$. Defining $\lambda = b_j/b_i$, this shows $v_j = \lambda v_i$, for all $v \in C$. □

Finally, we have to consider $i \in Irev$.

Proposition 6.9. *Suppose i, j are unblocked, $i \in Irev$ and $j \in Irev \cup Prev$. Then the following are equivalent:*

1. $i \overset{=0}{\to} j$ *holds in the flux cone* C.
2. $i \overset{=0}{\to} j$ *holds in all minimal proper faces* $G^k, k = 1, \ldots, s$.
3. $g_i^k = 0$ *implies* $g_j^k = 0$, *for all* $k = 1, \ldots, s$.

If also $j \overset{=0}{\leftrightarrow} i$ or $i \backsim^{\lambda} j$, then $j \in Irev$.

Proof. (1) $\Rightarrow$ (2) $\Rightarrow$ (3) is obvious, so we have to prove only (3) $\Rightarrow$ (1). For all $v \in C$, there exist $b \in \text{lin.space}(C)$ and $\alpha_k \geq 0$ such that $v = \sum_{k=1}^s \alpha_k g^k + b$. Since $i \in Irev$ and $j \in Irev \cup Prev$, we get $b_i = b_j = 0$. By the definition of $Irev$, either $g_i^k \geq 0$, for all $k = 1, \ldots, s$, or $g_i^k \leq 0$, for all $k = 1, \ldots, s$. Suppose $v_i = \sum_{k=1}^s \alpha_k g_i^k = 0$. It follows $g_i^k = 0$, for $k = 1, \ldots, s$. Using (3), we obtain $g_j^k = 0$ for $k = 1, \ldots, s$, and so $v_j = \sum_{k=1}^s \alpha_k g_j^k = 0$.

Under the hypotheses of Prop. 6.9, suppose $i \overset{=0}{\leftrightarrow} j$. Clearly, $j \overset{=0}{\to} i$. If $j \in Prev$, then by Prop. 6.7, $i \in Prev$, which is a contradiction. So $j \in Irev$. Similarly, if $i \backsim^{\lambda} j$, then $i \overset{=0}{\leftrightarrow} j$, and again $j \in Irev$. □

i/j	$Irev$			$Prev$			$Frev$		
	$\overset{=0}{\longrightarrow}$	$\overset{=0}{\longleftrightarrow}$	$\backsim^{\lambda}$	$\overset{=0}{\longrightarrow}$	$\overset{=0}{\longleftrightarrow}$	$\backsim^{\lambda}$	$\overset{=0}{\longrightarrow}$	$\overset{=0}{\longleftrightarrow}$	$\backsim^{\lambda}$
$Irev$	Prop.6.9	Cor.6.10	Cor.6.10	Prop.6.9					
$Prev$				Prop.6.7	Prop.6.7	Prop.6.7			
$Frev$							Prop.6.8	Prop.6.8	Prop.6.8

Table 6.1: Reaction coupling cases

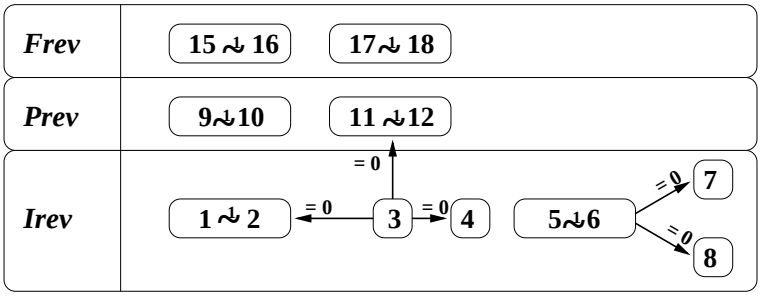


Figure 6.3: Coupled reactions in the network depicted in Fig. 6.2

Corollary 6.10. *Suppose i, j are unblocked and $i, j \in Irev$. Then we have:*

a) *$i \overset{=0}{\leftrightarrow} j$ iff $g_i^k = 0$ is equivalent to $g_j^k = 0$, for all $k = 1, \ldots, s$.*

b) *$i \backsim^{\lambda} j$ iff $g_j^k = \lambda g_i^k$, for all $k = 1, \ldots, s$.*

Proof. a) Suppose $g_j^k = 0$ is equivalent to $g_i^k = 0$, for all $k = 1, \ldots, s$. Then $g_i^k = 0$ implies $g_j^k = 0$, and vice versa, for all $k = 1, \ldots, s$. Since $i, j \in Irev$, we may apply Prop. 6.9 and get $i \overset{=0}{\rightarrow} j$ and $j \overset{=0}{\rightarrow} i$. So $i \overset{=0}{\leftrightarrow} j$.
b) Suppose $g_j^k = \lambda g_i^k$, for all $k = 1, \ldots, s$. For all $v \in C$, there exist $b \in \mathrm{lin.space}(C)$ and $\alpha_k \geq 0$ such that $v = \sum_{k=1}^{s} \alpha_k g^k + b$. Since $i, j \in Irev$, we have $b_i = b_j = 0$. It follows that $v_j = \sum_{k=1}^{s} \alpha_k g_j^k = \sum_{k=1}^{s} \alpha_k \lambda g_i^k = \lambda v_i$. □

Tab. 6.1 summarizes the different possible coupling relationships. Note that $i \overset{=0}{\rightarrow} j$, $i \overset{=0}{\leftrightarrow} j$ and $i \backsim^{\lambda} j$ are equivalent for $i, j \in Prev$ or $i, j \in Frev$.

Example 6.11. *Fig. 6.3 shows all coupled reactions in the network from Fig. 6.2. We see that many reactions depend on reaction* 3*. A zero flux for this reaction implies a zero flux for the reactions* $1, 2, 4, 11$ *and* 12*. Thus, reaction* 3 *plays a crucial role in the network. Reaction* 19 *is blocked, because it is involved neither in the definition of the minimal proper faces nor in the definition of the lineality space.*

6.4 An Improvement in the FCF Algorithm

It follows from the preceding section that coupling relationships can only hold between certain reaction types. According to Tab. 6.1, to detect coupling relationships, we do

not have to explore exhaustively all possible reaction pairs. In the following, we show that we can improve the FCF procedure significantly by applying linear programming only in those cases where coupling relationships can occur. All the possible cases are given in Tab. 6.1, where an empty entry indicates that the corresponding coupling relationship is not possible. Accordingly, the number of LP problems that have to be solved in this improved version of the FCF algorithm is much smaller than that of the original FCF algorithm. Moreover, this improved version does not require a reconfiguration of the metabolic network. Hence, no increase in the number of variables and constraints occurs and no post-processing procedure is needed.

Before looking for blocked and coupled reactions, we need to classify reactions according to their reversibility types. This task can easily be done if we are given a minimal set of generating vectors of the flux cone. Alternatively, we can classify reactions using linear programming.

6.4.1 Reaction Classification

In the following, we use the same linear programs for identifying the reversibility type of reactions, i.e., the sets $Irev, Prev$ and $Frev$ as well as the set Blk of blocked reactions.

As we did in Sect. 6.1, we first check whether the lineality space is trivial, i.e., $\mathrm{lin.space}(C) = \{0\}$. If this is the case, $Frev = \emptyset$. Otherwise, given a reversible reaction $j \in Rev$, consider the following LP problem

$$\max \{v_j : Sv = 0,\ v_k = 0 \text{ for all } k \in Irr,\ v_j \leq 1\}. \tag{6.5}$$

Since the above LP problem is feasible and bounded in the direction of the objective function, this LP problem has an optimal solution, say v^*. Let $F = \{i \in Rev \mid v_i^* \neq 0\}$. Reaction j is fully reversible if and only if $j \in F$. Moreover, for each $i \in F$, there exists a vector $b \in \mathrm{lin.space}(C)$ such that $b_i \neq 0$, namely $b = v^*$. Accordingly, all reactions in F are fully reversible, i.e., $F \subseteq Frev$.

Here, we are interested in finding all fully reversible reactions. We propose an iterative procedure that obtains, in each iteration k, a subset of fully reversible reactions using the following scheme. For $k \geq 1$, let R^k be the set of reversible reactions that are candidate to be chosen in iteration k to perform the LP problem (6.5), and let F^{k-1} be the set of all fully reversible reactions identified in iterations i for $i = 1, \ldots, k-1$. Set $R^1 = Rev$ and $F^0 = \emptyset$. The basic steps of iteration k are as follows:

1. Choose a reaction $j \in R^k$,
2. Solve the LP problem (6.5) and let v^* its optimal solution,
3. Let $F = \{i \in Rev \mid v_i^* \neq 0\}$. Set $F^k = F^{k-1} \cup F$ and set $R^{k+1} = R^k \backslash (\{j\} \cup F)$.

The above procedure terminates in iteration p if no reversible reaction $j \in R^{p+1}$ exists to perform the LP problem (6.5), i.e., $R^{p+1} = \emptyset$. If this is the case, we get $Frev = F^p$.

Once we have identified all fully reversible reactions, we need to determine the set of pseudo-irreversible reactions that take the same direction for all steady-state flux distributions. We first check whether the flux cone is equal to its lineality space, i.e., $C = \text{lin.space}(C)$, or equivalently, if all irreversible reactions are blocked. We proposed an LP problem in Sect. 6.1 to check whether this is the case. If all irreversible reactions are blocked, Blk $= Irev = \{1, \ldots, n\} \setminus Frev$ and $Prev = \emptyset$. Otherwise, given a pseudo-irreversible reaction $j \in Rev \setminus Frev$, consider the following two LP problems

$$\max \{v_j : Sv = 0,\ v_k \geq 0 \text{ for all } k \in Irr,\ v_j \leq 1\}, \tag{6.6}$$

$$\min \{v_j : Sv = 0,\ v_k \geq 0 \text{ for all } k \in Irr,\ v_j \geq -1\}. \tag{6.7}$$

Let v^+ and v^- be optimal solutions of the LP problems (6.6) and (6.6), respectively, and let P and N be the sets given by

$$P = \{i \in Rev \setminus Frev \mid v_i^+ > 0 \text{ or } v_i^- > 0\},$$

$$N = \{i \in Rev \setminus Frev \mid v_i^+ < 0 \text{ or } v_i^- < 0\}.$$

Then either (1), (2) or (3) holds:

1. $j \notin P \cup N$ and so, j is a blocked reaction, i.e., $j \in$ Blk,
2. $j \in P \cup N \setminus (P \cap N)$, then j takes the same direction for all steady-state flux distributions, and so $j \in Irev$,
3. $j \in P \cap N$, and so $j \in Prev$.

Moreover, we have $(P \cup N) \cap \text{Blk} = \emptyset$ and $P \cap N \subseteq Prev$.

In analogy with our procedure for identifying fully reversible reactions, we propose an iterative procedure that obtains the sets $Prev$ and $Irev$, and determines the Blk$\cap Rev$ of all reversible reactions that are blocked. For $k \geq 1$, let R^k be the set of reversible reactions that are candidate to be chosen in iteration k to perform the LP problems (6.6) and (6.7), and let P^{k-1} (resp. N^{k-1}) be the set of all reversible reactions i, for which a vector $v \in C$, with $v_i > 0$ (resp. $v_i < 0$), has been found in the previous iterations i for $i = 1, \ldots, k-1$. Set $R^1 = Rev \setminus Frev$ and $P^0 = N^0 = \emptyset$. The basic steps of iteration k are as follows:

1. Choose a reaction $j \in R^k$,
2. If $j \notin P^{k-1}$, then solve the LP problem (6.6) and let v^+ its optimal solution, otherwise set $v^+ = 0$.
3. If $j \notin N^{k-1}$, then solve the LP problem (6.7) and let v^- its optimal solution, otherwise set $v^- = 0$.
4. Let $P = \{i \in Rev \setminus Frev \mid v_i^+ > 0 \text{ or } v_i^- > 0\}$ and $N = \{i \in Rev \setminus Frev \mid v_i^+ < 0 \text{ or } v_i^- < 0\}$. Set $P^k = P^{k-1} \cup P$, set $N^k = N^{k-1} \cup N$, and set $R^{k+1} = R^k \setminus (\{j\} \cup (P \cap N))$.

The above procedure terminates in iteration q if no reversible reaction $j \in R^{q+1}$ exists to perform the LP problems (6.6) and (6.7), i.e., $R^{q+1} = \emptyset$. If this is the case, we have the following results:

- $Prev = P^q \cap N^q$,
- $Irev = Irr \cup ((P^q \cup N^q) \setminus (P^q \cap N^q))$,
- Blk $\cap\, Rev = Rev \setminus (P^q \cup N^q)$.

Together, the above procedures allow for identifying the reversibility type of the reactions ($Irev$, $Prev$ and $Frev$) as well as the set of reversible reactions that are blocked (Blk $\cap\, Rev$). In the following, we determine the set Blk $\cap\, Irr$ of irreversible reactions which are blocked.

Given an irreversible reaction $j \in Irr$, consider the following LP problem

$$\max \{v_j : Sv = 0,\ v_k \geq 0 \text{ for all } k \in Irr,\ v_j \leq 1\}. \tag{6.8}$$

Let v^* be an optimal solution of the above LP problem and let $I = \{i \in Irr \mid v_i^* > 0\}$. Reaction j is blocked if and only if $j \notin I$. Moreover, for each $i \in I$, there exists a vector $c \in C$ such that $c_i > 0$, namely $c = v^*$. Accordingly, we have $I \subseteq Irr \setminus$ Blk. Here again, we seek to find all irreversible reactions which are blocked. To do so, we propose the following iterative procedure. For each iteration $k \geq 1$, let R^k be the set of irreversible reactions that are candidate to be chosen in iteration k to perform the LP problem (6.8), and let I^{k-1} be the set of all irreversible reactions identified in iterations i for $i = 1, \ldots, k-1$. Set $R^1 = Irr$ and $I^0 = \emptyset$. The basic steps of iteration k are as follows:

1. Choose a reaction $j \in R^k$,
2. Solve the LP problem (6.8) and let v^* its optimal solution,
3. Let $I = \{i \in Irr \mid v_i^* > 0\}$. Set $I^k = I^{k-1} \cup I$ and set $R^{k+1} = R^k \setminus (\{j\} \cup I)$.

The above procedure terminates, say in iteration r, when no irreversible reaction $j \in R^{r+1}$ exists to perform the LP problem (6.8), i.e., $R^{r+1} = \emptyset$. If this is the case, we get Blk $\cap\, Irr = Irr \setminus I^r$.

Altogether, for identifying blocked reactions and classifying reactions according to their reversibility types, we need to solve at most $2|Rev| + |Irev| + |Prev| - |Frev|$ LP problems.

6.4.2 Reaction Couplings

In the following, we are interested in finding coupled reactions. Since coupling relationships are defined only for unblocked reactions, we assume that blocked reactions have been identified beforehand using the iterative procedures given in the preceding subsection, and that all their respective columns in the stoichiometric matrix have been

removed. In addition, since each reaction $j \in Irev$ is operating in the same direction, in what follows we assume that all reactions in $Irev$ are operating in the forward direction without loss of generality. Hence, by abuse of notation, all reactions in $Irev$ will be called irreversible in the current subsection.

According to Tab. 6.1, when looking for coupled reactions, we do not have to explore exhaustively all possible reaction pairs.

Couplings between reversible reactions

Given two reversible reactions i, j with $i, j \in Prev$ or $i, j \in Frev$, $i \overset{=0}{\rightarrow} j, i \overset{=0}{\leftrightarrow} j$ and $i \backsim^{\lambda} j$ are equivalent. In the following, we propose an LP problem to check whether $i \overset{=0}{\rightarrow} j$ holds. First, we need the following lemma.

Lemma 6.12. *Let $i, j \in Rev$ be two reversible reactions, with $i, j \in Prev$ or $i, j \in Frev$. If $i \overset{=0}{\rightarrow} j$, then for each vector $v \in kern(S)$, with $v_i = 0$, there exists a vector $u \in kern(S)$ with $u_i = 0$, $u_j = v_j$ and*

$$\{k \in Irr \mid u_k < 0\} \subsetneq \{k \in Irr \mid v_k < 0\}.$$

Proof. Let $v \in kern(S)$, with $v_i = 0$, and let $l \in \{k \in Irr \mid v_k < 0\}$. Since $l \in Irev$ and $i \in Prev \cup Frev$, $i \overset{=0}{\rightarrow} l$ does not hold and there exists a vector $v' \in C$ such that $v'_i = 0$ and $v'_l > 0$. Moreover, since $i \overset{=0}{\rightarrow} j$ and $v'_i = 0$, we get $v'_j = 0$. Consider $u = v - (v_l/v'_l) \cdot v'$. We have $Su = 0$, $u_i = 0$, $u_j = v_j$ and for each $k \in Irr$, we have $u_k = v_k - (v_l/v'_l)v'_k \geq v_k$. Therefore, for each $k \in Irr$, $u_k < 0$ implies $v_k < 0$, and so $\{k \in Irr \mid u_k < 0\} \subseteq \{k \in Irr \mid v_k < 0\}$. Since $u_l = 0$ and $v_l < 0$, $\{k \in Irr \mid u_k < 0\} \subsetneq \{k \in Irr \mid v_k < 0\}$, and the claim follows. □

The following proposition states that, given two reversible reactions $i, j \in Rev$, with $i, j \in Prev$ or $i, j \in Frev$, only the stoichiometric constraints determine whether $i \overset{=0}{\rightarrow} j$ holds, independently of the thermodynamic constraints.

Proposition 6.13. *Let $i, j \in Rev$ be two reversible reactions, with $i, j \in Prev$ or $i, j \in Frev$. The following are equivalent:*

1. $i \overset{=0}{\rightarrow} j$.
2. *For each $v \in kern(S)$, $v_i = 0$ implies $v_j = 0$.*
3. *For each $v \in kern(S) \cap \{v \in \mathbb{R}^n \mid v_j \geq 0\}$, $v_i = 0$ implies $v_j = 0$.*

Proof. (2) ⇒ (1), (2) ⇒ (3) and (3) ⇒ (2) are immediate. So we have to prove only (1) ⇒ (2). Let $v \in kern(S)$, with $v_i = 0$. According to Lemma 6.12, there exist $u^0, \ldots, u^p \in kern(S)$ such that $u^0 = v$ and for all $l = 1, \ldots, p$, we have $u^l_i = 0$, $u^l_j = u^{l-1}_j$, $\{k \in Irr \mid u^l_k < 0\} \subsetneq \{k \in Irr \mid u^{l-1}_k < 0\}$, and $\{k \in Irr \mid u^p_k < 0\} = \emptyset$. Since $\{k \in Irr \mid u^p_k < 0\} = \emptyset$, we have $u^p_k \geq 0$ for all $k \in Irr$, and so $u^p \in C$. In addition, since $i \overset{=0}{\rightarrow} j$, $u^p \in C$ and $u^p_i = 0$, we get $u^p_j = 0$. Therefore, we have $u^l_j = 0$ for $l = 1, \ldots, p$. Hence, we have $v_j = 0$, and the claim follows. □

Given two reversible reactions i, j with $i, j \in Prev$ or $i, j \in Frev$, let $C^{ij} = \{v \in \mathbb{R}^n \mid Sv = 0,\ v_i = 0,\ v_j \geq 0\}$. According to Proposition 6.13, $i \overset{=0}{\rightarrow} j$ holds if and only if reaction j is blocked in the cone C^{ij}, i.e., $v_j = 0$ for all $v \in C^{ij}$. To check whether this is the case, consider the following LP problem

$$\max \{v_j : Sv = 0,\ v_i = 0,\ 0 \leq v_j \leq 1\}, \tag{6.9}$$

and let v^* be its optimal solution. Therefore, $i \overset{=0}{\rightarrow} j$ if and only if $v_j^* = 0$.

Couplings between reversible and irreversible reactions

Given two reactions $i \in Irev$ and $j \in Prev$, we only have to check whether $i \overset{=0}{\rightarrow} j$. The other coupling relationships cannot occur. Let $C^{ij} = \{v \in \mathbb{R}^n \mid Sv = 0,\ v_k \geq 0,\ \text{for all } k \in Irr,\ v_i = 0\}$. We have $i \overset{=0}{\rightarrow} j$ if and only if reaction j is blocked in the cone C^{ij}, i.e., $v_j = 0$ for all $v \in C^{ij}$. Accordingly, if v^1 and v^2 are optimal solutions of the following LP problems

$$\begin{aligned} &\max \{v_j : Sv = 0,\ v_k \geq 0 \text{ for all } k \in Irr,\ v_i = 0,\ v_j \leq 1\}, \\ &\min \{v_j : Sv = 0,\ v_k \geq 0 \text{ for all } k \in Irr,\ v_i = 0,\ v_j \geq -1\}, \end{aligned} \tag{6.10}$$

then $i \overset{=0}{\rightarrow} j$ if and only if $v_j^1 = v_j^2 = 0$.

Couplings between irreversible reactions

In analogy with the FCF algorithm, given two reactions $i, j \in Irev$, we determine the upper and lower bounds R_{max} and R_{min} such that $0 \leq R_{min} v_j \leq v_i \leq R_{max} v_j$ for all $v \in C$. R_{max} and R_{min} are the optimal values for maximizing and minimizing the following LP problems

$$\begin{aligned} &\max \{v_i : Sv = 0,\ v_j = 1,\ v_k \geq 0, \text{for all } k \in Irr\}, \\ &\min \{v_i : Sv = 0,\ v_j = 1,\ v_k \geq 0, \text{for all } k \in Irr\}, \end{aligned}$$

respectively. Comparison of R_{max} and R_{min} allows the FCF algorithm to determine whether reactions i and j are coupled using the following rules:

- $i \overset{=0}{\rightarrow} j$ if and only if $R_{min} \neq 0$,
- $j \overset{=0}{\rightarrow} i$ if and only if $R_{max} \neq +\infty$,
- $j \curvearrowright^{\lambda} i$ if and only if $R_{min} = R_{max} = \lambda \neq 0$.

Altogether, the number of LP problems that have to be solved for identifying coupled reactions is at most

$$\frac{|Prev|(|Prev| - 1)}{2} + \frac{|Frev|(|Frev| - 1)}{2} + 2|Irev||Prev| + |Irev|(|Irev| - 1).$$

The next proposition compares the total number of the LP problems that have to be solved in the original version of the FCF algorithm with that of our improved version of the FCF algorithm.

Proposition 6.14. *Let $\mathscr{N}^{original}$ (resp. $\mathscr{N}^{improved}$) the number of the LP problems that have to be solved in the original (resp. improved) version of the FCF algorithm. Then,*

$$\begin{aligned}\mathscr{N}^{original} - \mathscr{N}^{improved} &= |Rev|^2 + 2|Frev|(|Irev| + |Prev| + 1) + \\ &\quad \tfrac{|Prev|}{2}(|Prev| - 1) + \tfrac{|Frev|}{2}(|Frev| - 1) + \\ &\quad 2|Rev|(n-1) \geq 0.\end{aligned}$$

Proof. We have $\mathscr{N}^{\text{original}} = (|Irr| + 2|Rev|)^2$. Since $n = |Irr| + |Rev| = |Irev| + |Prev| + |Frev|$, we get $\mathscr{N}^{\text{original}} = (|Irev| + |Prev| + |Frev| + |Rev|)^2$. Moreover, we have $\mathscr{N}^{\text{improved}} = 2|Rev| + |Irev| + |Prev| - |Frev| + \frac{|Prev|(|Prev|-1)}{2} + \frac{|Frev|(|Frev|-1)}{2} + 2|Irev||Prev| + |Irev|(|Irev| - 1)$. Then, the claim is straightforward. □

The proposition above states that the number of the LP problems that have to be solved in the original version of the FCF algorithm is bigger than that of our improved version. Note that both numbers are equal if no reaction is reversible, i.e., $|Rev| = 0$. In such a case, the reconfigured network is identical to the original one. However, in the presence of reversible reactions, we have $\mathscr{N}^{\text{original}} > \mathscr{N}^{\text{improved}}$. Moreover, since our version does not require a reconfiguration of the metabolic network, the number of variables and constraints in our LP problems is smaller than that of variables and constraints in the original FCF algorithm. In addition, no post-processing procedure is needed in our improved version.

6.5 A New Algorithm For Flux Coupling Analysis

The results in Sect. 6.3 also suggest a new algorithm to identify blocked and coupled reactions. This method does not require any reconfiguration of the metabolic network. It is only based on the reversibility type of the reactions and a minimum set of generators of the flux cone. The basic steps of this new method are shown in Algorithm 2. First, we compute a set of generators of the flux cone C using existing software for polyhedral computations. Second, we classify the reactions according to their reversibility type. This classification allows us to determine whether a coupling between two reactions is possible.

Both our new algorithm and the FCF algorithm have been implemented in the Java language. The FCF procedure was realized using CPLEX 9.0 (a solver for linear and integer programming problems) accessed via Java. To compute a set of generating vectors of the steady-state flux cone, our algorithm uses the software `cdd` [33], which is a C++ implementation of the Double Description method of Motzkin et al. for general convex polyhedra in $\mathbb{R}^n$.

To compare the two approaches, we computed blocked and coupled reactions for some genome-scale networks. The computations were performed on a Linux server with an AMD Athlon Processor 1.6 GHz and 2 GB RAM. We present computation times for models of the human red blood cell [121], the human cardiac mitochondria [116], the central carbon metabolism of *E. coli* [50; 106], the *E. coli K-12* (iJR904

Metabolic network	Blk	Irev	Prev	Frev	MMB	FCMMB	FCF
Red Blood Cell	0	31	14	6	2.32	0.26	110.65
Central metabolism of *E. coli*	0	92	18	0	214.49	2.55	477.14
Human cardiac mitochondria	121	83	3	9	1262.65	0.34	13426.91
Helicobacter pylori	346	128	15	39	13551.44	0.43	318374.15
E. coli K-12	435	480	49	110	261306.15	5.32	≥ 1 week

Table 6.2: Metabolic systems, with the number of blocked reactions (*Blk*), the size of the sets $Irev, Prev, Frev$, the running time (in seconds) of computing a set of generators (*MMB*), reaction coupling using this set (*FCMMB*), and reaction coupling using the FCF procedure (*FCF*).

GSM/GPR) [86], and the *H. pylori* (iIT341 GSM/GPR) [111]. We refer to [14] for a discussion of the biological aspects of flux coupling analysis.

Tab. 6.2 summarizes our computational results. It shows that flux coupling analysis can be done extremely fast if a set of generators of the flux cone is available. Computing such a set is the most time-consuming part in our algorithm. However, it should be noted that this step has an interest in its own. We obtain similar information as by computing the elementary flux modes or extreme pathways of the network. The overall running time of the new algorithm is still significantly faster than the original FCF method.

Algorithm 2 Procedure for identifying blocked and coupled reactions.

Input: • Sets $Irev, Prev, Frev \subseteq \{1, \ldots, n\}$,
• For each minimal proper face $G^k, k = 1, \ldots, s,$ a generating vector $g^k \in G^k \setminus \text{lin.space}(C)$,
• A vector basis $b^1, \ldots, b^t$ of $\text{lin.space}(C)$.

Output: • Blocked reactions: Blk $= \{i \mid i \text{ is blocked}\}$,
• Coupled reactions: $A = \{(i,j) \mid i \rightharpoonup^{\lambda} j, 1 \leq i < j \leq n\}$,
$B = \{(i,j) \mid i \overset{=0}{\leftrightarrow} j, 1 \leq i < j \leq n, (i,j) \notin A\}$,
$C = \{(i,j) \mid i \overset{=0}{\rightarrow} j, (i,j) \notin A \cup B, (j,i) \notin A \cup B\}$.

Initialization: Blk $:= \emptyset,\ A := \emptyset,\ B := \emptyset,\ C := \emptyset$.

/* Blocked reactions */
for all $i \in \{1, \ldots, n\}$ **do** ▷ Proposition 6.4
  if $(b_i^l = 0, \forall l = 1, \ldots, t)$ and $(g_i^k = 0, \forall k = 1, \ldots, s)$ **then**
    add$(i,$ Blk$)$;
  end if
end for

Continued from previous page

/* Coupled reactions */
$Irev := Irev \setminus$ Blk;
for all $i, j \in Prev$ with $i < j$ **do** ▷ Proposition 6.7
  if $\exists \lambda \in \mathbb{R}$ such that $g_j^k = \lambda g_i^k, \forall k = 1, \ldots, s$ **then**
    add($(i, j), A$);
  end if
end for
for all $i, j \in Frev$ with $i < j$ **do** ▷ Proposition 6.8
  if $\exists \lambda \in \mathbb{R}$ such that $b_j^l = \lambda b_i^l, \forall l = 1, \ldots, t,$ and $g_j^k = \lambda g_i^k, \forall k = 1, \ldots, s$ **then**
    add($(i, j), A$);
  end if
end for
for all $i \in Irev, j \in Irev \cup Prev$ **do** ▷ Proposition 6.9
  if $g_i^k \neq 0$ or $g_j^k = 0, \forall k = 1, \ldots, s$ **then**
    add($(i, j), C$);
  end if
end for
for all $(i, j) \in C$ with $i, j \in Irev$ and $i < j$ **do** ▷ Corollary 6.10
  if $(j, i) \in C$ **then**
    remove($(i, j), C$), remove($(j, i), C$));
    if $\exists \lambda \in \mathbb{R}$ such that $g_j^k = \lambda g_i^k, \forall k = 1, \ldots, s$ **then**
      add($(i, j), A$);
    else
      add($(i, j), B$);
    end if
  end if
end for

CHAPTER 7

Minimal Direction Cuts in Metabolic Networks

7.1 Introduction

In the previous chapter, we introduced flux coupling analysis which allows for identifying dependencies between reactions. This analysis not only can help us to enhance our understanding of biological systems, but also can be used to determine a reaction that is critical for the survival of a certain pathogen. Such a reaction could be a suitable target to be repressed by network perturbations. A recent approach [46; 47] has introduced minimal cut sets to identify such perturbations. A *minimal cut set* (MCS) is a minimal set of reactions whose removal represses a given target reaction. However, this approach suffers from two major drawbacks that are limiting in practice. First, using MCSs becomes inadequate when we are interested in inhibiting only one direction of a reversible target reaction. Second, and more importantly, the computation of MCSs is currently based on the conservation property of elementary modes (EMs) and the principle that each MCS is a minimal hitting set for all EMs involving the target reaction. This computation is hampered by the combinatorial explosion of the number of EMs and hence becomes impractical for genome-scale networks.

To overcome these limitations, we introduce the concept of *minimal direction cuts* in metabolic networks [61]. A *minimal forward* (resp. *backward*) *direction cut (MFC)* (resp. *MBC*) is a minimal set of reactions whose removal prevents the target reaction from carrying a flux in the forward (resp. backward) direction. If the target reaction is irreversible, MFCs are identical to MCSs and the empty set is the unique MBC. If, however, the target reaction is reversible, each MCS is the union of an MFC and an MBC. More importantly, the computation of MFCs and MBCs does not require that of EMs. Instead it can be based on the Farkas lemma for equality and inequality constraints. Finally, MCSs can be directly calculated from MFCs and MBCs.

This chapter highlights the key results presented in [61] and is organized as follows. In Sect. 7.2, we give a first presentation of our approach. In Sect. 7.3, we formally characterize MFCs and MBCs using an extended Farkas lemma. In Sect. 7.4, we propose an iterative algorithm to identify MFCs and MBCs in a metabolic network. Finally, some computational results are given in Sect. 7.5.

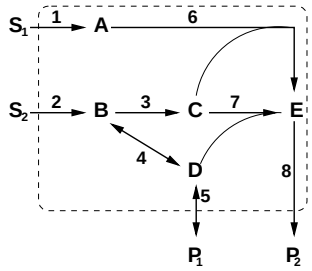


Figure 7.1: Hypothetical network from [46]

7.2 Minimal Forward and Backward Direction Cuts

For the rest of the chapter, let $\tau \in \{1, \ldots, n\}$ be the target reaction. A *forward* (resp. *backward*) *direction cut* for reaction τ is a set of reactions whose removal from the network implies that all possible steady-state flux distributions over the network do not involve reaction τ in the forward (resp. backward) direction. Here, the removal of a set of reactions is mathematically expressed by constraining the fluxes through all these reactions to zero.

Definition 7.1. *Let C be the flux cone defined in equation (3.5).*

- *A forward direction cut (FC) for reaction τ is a set of reactions $M \subseteq \{1, \ldots, n\}$, such that for any $v \in C$ with $v_i = 0$ for all $i \in M,\ v_\tau \leq 0$.*
- *A backward direction cut (BC) for reaction τ is a set of reactions $M \subseteq \{1, \ldots, n\}$, such that for any $v \in C$ with $v_i = 0$ for all $i \in M,\ v_\tau \geq 0$.*

A forward (resp. backward) direction cut M is minimal, denoted MFC (resp. MBC), if there is no forward (resp. backward) direction cut $M' \subsetneq M$ strictly contained in M.

Example 7.2. *Consider the hypothetical network [46] depicted in Fig. 7.1. It consists of five internal metabolites $(A, \ldots, E)$, and eight reactions $(1, \ldots, 8)$, whereof reactions 4 and 5 are reversible. Let reaction 5 be the target reaction. Fig. 7.2 (resp. Fig. 7.3) displays MFCs and MBCs (resp. MCSs) for reaction 5. For instance, the forward direction cut MFC2 does not inhibit the backward direction of reaction 5. MCS2 = $\{2, 3\}$ is a minimal cut set to repress reaction 5. However, this set is not minimal to repress the production of metabolite P_1 since this is possible by removing only reaction 2. Indeed, MCS2 is the union of the minimal direction cuts MFC2 and MBC2.*

The meaningful direction cuts are those which are minimal, since the trivial solution, which consists of removing all reactions, always is a direction cut for τ. Obviously, if there is no flux vector $v \in C$ with $v_\tau < 0$ (resp. $v_\tau > 0$), then the empty set is the unique MBC (resp. MFC) for τ. In particular, if τ is an irreversible reaction, then the empty set is the unique MBC for τ and MFCs for τ are identical to minimal cut sets. On the other hand, if τ is reversible, then the computation of MFCs is equivalent to that of MBCs in the sense that an algorithm for computing MFCs could also be used

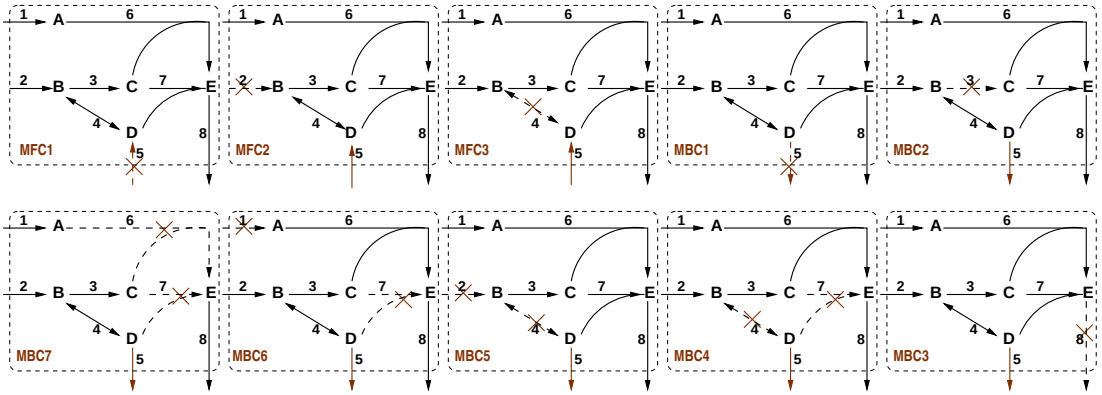


Figure 7.2: Minimal forward and backward direction cuts for the target reaction 5.

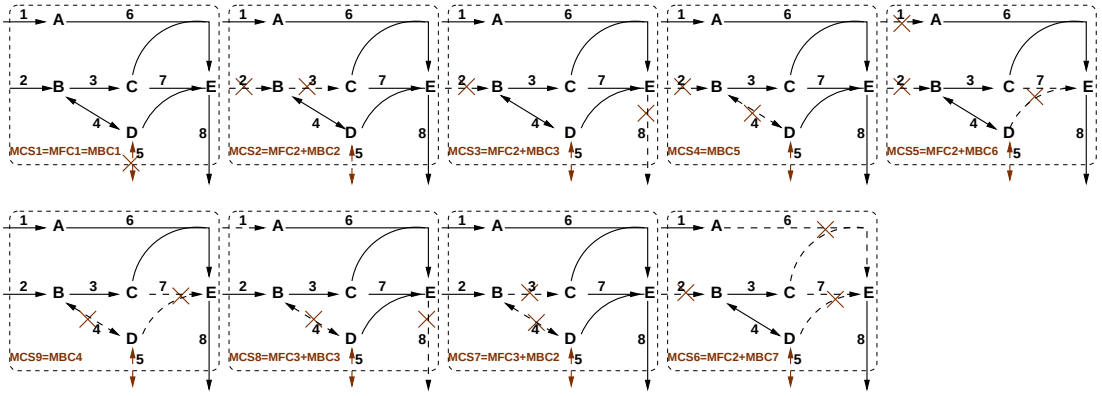


Figure 7.3: Minimal cut sets for the target reaction 5.

to calculate MBCs. To show this, let S' be the $m \times n$ matrix whose columns are defined by

$$S'_{*j} = S_{*j} \text{ for all } j \in \{1, \ldots, n\} \setminus \{\tau\} \text{ and } S'_{*\tau} = -S_{*\tau},$$

and let

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

Lemma 7.3. *Let $M \subseteq \{1, \ldots, n\} \setminus \{\tau\}$ be a set of reactions. If τ is reversible, i.e., $\tau \in Rev$, then the following are equivalent:*

1. *For any $v \in C$ with $v_i = 0$ for all $i \in M$, $v_\tau \geq 0$.*
2. *For any $v' \in C'$ with $v'_i = 0$ for all $i \in M$, $v'_\tau \leq 0$.*

Proof. Let $v, v' \in \mathbb{R}^n$ such that $v'_\tau = -v_\tau$ and $v'_i = v_i$ for all $i \in \{1, \ldots, n\} \setminus \{\tau\}$. Since $\tau \in Rev$, $v \in C$ and $v_\tau \geq 0$ if and only if $v' \in C'$ and $v'_\tau \leq 0$. □

Given a set $M \subseteq \{1, \ldots, n\} \setminus \{\tau\}$ of reactions, it follows from Lemma 7.3 that M is an MBC if and only if M is an MFC with respect to the stoichiometric matrix S'. This is why we focus in the following on characterizing MFCs. The same results hold also for MBCs. Note that $\{\tau\}$ is a trivial MFC for τ. Therefore, each MFC $M \neq \{\tau\}$,

called *non-trivial* MFC for τ, does not include τ, i.e., $\tau \notin M$. In the following, we are interested in finding these non-trivial MFCs.

In analogy with minimal cut sets, MFCs could be computed using the conservation property of elementary modes (EMs) (see Lemma 3.9 in Chap. 3). Remember that according to this property, whenever the flux through some reaction is constrained to zero, the EMs that remain are those which do not contain that reaction. Accordingly, each MFC is a minimal hitting set of all EMs involving the target reaction in the forward direction. A set $M \subseteq \{1, \ldots, n\}$ is a *hitting set* of a collection C of EMs if M meets every EM in the collection C, i.e., $M \cap \{i \mid e_i \neq 0\} \neq \emptyset$ for each $e \in C$. In view of the combinatorial explosion of EMs, computing MFCs using this strategy does not scale well for genome-scale metabolic networks. In the following, we instead use the Farkas lemma for equality and inequality constraints to further characterize MFCs.

7.3 Characterizing MFCs Using the Farkas Lemma

In this section we give a formal characterization of minimal forward direction cuts in a metabolic network. We use mainly the Farkas lemma, which states that the unsolvability of a system of constraints can be established by finding a solution for a corresponding dual system (see Sect. 2.3 in Chap. 2 for more details).

Let $M \subseteq \{1, \ldots, n\}$ be an FC and let $(e_1, \ldots, e_n)$ be the canonical basis of $\mathbb{R}^n$. In the following, I_M (resp. I_{Irr}) denotes the matrix whose rows are the vectors e_j with $j \in M$ (resp. $j \in Irr$). From Definition 7.1, we get that

$$\text{there is no } v \in \mathbb{R}^n \text{ such that } Sv = 0,\ I_M v = 0,\ I_{Irr} v \leq 0 \text{ and } v_\tau \leq -1. \tag{7.1}$$

Let A, B and C be the matrices defined by

$$A = \begin{pmatrix} S \\ -I_M \end{pmatrix}, B = -I_{Irr} \text{ and } C = e_\tau,$$

and let $x = y = 0$ and $z = -1$. By (7.1), there exists no solution vector $v \in \mathbb{R}^n$ for $Av = x,\ Bv \geq y,\ Cv \leq z$. According to the Farkas lemma 2.15 in Chap. 2, there exist row vectors $\alpha \in \mathbb{R}^m$, $\mu \in \mathbb{R}^{|M|}$, $\beta \in \mathbb{R}^{|Irr|}$ and $\gamma \in \mathbb{R}$, such that $\gamma C = (\alpha, \mu)A + \beta B,\ \beta \geq 0,\ \gamma \geq 0$ and $\gamma z < 0$. This leads to the following proposition which characterizes FCs using the Farkas lemma.

Proposition 7.4. *Let $M \subseteq \{1, \ldots, n\}$ be a set of reactions. Then, the following are equivalent:*

1. *M is an FC for τ.*
2. *There exist $\alpha_i, \beta_j, \mu_k, \gamma \in \mathbb{R}$ with $\beta_j \geq 0$ and $\gamma \geq 1$, such that*

$$\sum_{k \in M} \mu_k e_k + \gamma e_\tau = \sum_{i=1}^{m} \alpha_i S_{i*} - \sum_{j \in Irr} \beta_j e_j. \tag{7.2}$$

Let D be the cone defined by

$$D = \{\sum_{i=1}^{m} \alpha_i S_{i*} - \sum_{j \in Irr} \beta_j e_j \mid \beta_j \geq 0 \text{ for all } j \in Irr\}. \tag{7.3}$$

The following corollary describes FCs by means of the support of vectors in D.

Corollary 7.5. *Let D be the cone (7.3) and let $M \subseteq \{1, \ldots, n\} \setminus \{\tau\}$ be a set of reactions. Then, the following are equivalent:*

1. *M is an FC for τ.*
2. *There exists a vector $w \in D$, such that $w_\tau \geq 1$ and $Supp(w) \setminus \{\tau\} \subseteq M$.*

Proof. (1) $\Rightarrow$ (2): By Proposition 7.4, there exist $\alpha_i, \beta_j, \mu_k, \gamma \in \mathbb{R}$ with $\beta_j \geq 0$, and $\gamma \geq 1$ fulfilling equation (7.2). Let $w = \sum_{k \in M} \mu_k e_k + \gamma e_\tau$. On the one hand, since $\tau \notin M$, we get $w_\tau = \gamma \geq 1$ and $Supp(w) \setminus \{\tau\} = \{k \in M \mid \mu_k \neq 0\} \subseteq M$. On the other hand, we have $w = \sum_{i=1}^{m} \alpha_i S_{i*} - \sum_{j \in Irr} \beta_j e_j$, $w \in D$ and so the claim follows.

(1) $\Leftarrow$ (2): Since $w \in D$, there exist $\alpha_i, \beta_j \in \mathbb{R}$ with $\beta_j \geq 0$, such that $w = \sum_{i=1}^{m} \alpha_i S_{i*} - \sum_{j \in Irr} \beta_j e_j$. Moreover, since $\tau \notin M$ and $Supp(w) \setminus \{\tau\} \subseteq M$, we can write $w = \sum_{k \in M} w_k e_k + w_\tau e_\tau$ with $w_k = 0$ for each $k \in M \setminus Supp(w)$. We then get

$$\sum_{k \in M} w_k e_k + w_\tau e_\tau = \sum_{i=1}^{m} \alpha_i S_{i*} - \sum_{j \in Irr} \beta_j e_j.$$

By Proposition 7.4, M is an FC for τ.

□

Note that the minimality of MFCs is quite similar to that of elementary modes (EMs). An EM corresponds to a flux vector $e \in C$ involving a minimum set of reactions [99], i.e., $Supp(e)$ is minimal. In the same way, an MFC corresponds to a vector $w \in D$ such that $w_\tau \geq 1$ and $Supp(w)$ is minimal. We will call such a vector *simple in D with respect to τ and $Supp$*. More generally, we define:

Definition 7.6. *For $Q \subseteq \mathbb{R}^n$ let $\varphi : Q \to 2^{\{1,\ldots,n\}}$ be the function that maps each vector $s \in Q$ to a subset $\varphi(s) \subseteq \{1, \ldots, n\}$. A vector $s \in Q$ is simple in Q with respect to τ and φ if the following conditions hold:*

1. *$s_\tau \geq 1$,*
2. *There is no vector $s' \in Q$ such that $s'_\tau \geq 1$ and $\varphi(s') \subsetneq \varphi(s)$.*

If this is the case, we also say that $simp_{(Q,\tau,\varphi)}(s)$ holds.

The following theorem shows that MFCs are in a 1-1 correspondence with the support of vectors simple in D with respect to τ and $Supp$.

Theorem 7.7. *Let D be the cone (7.3) and let $M \subseteq \{1, \ldots, n\} \setminus \{\tau\}$ be a set of reactions. Then, the following are equivalent:*

1. *M is an MFC for τ.*

2. *There exists a vector w simple in D with respect to τ and $Supp$, such that $Supp(w) \setminus \{\tau\} = M$.*

Proof. (1) $\Rightarrow$ (2): By Corollary 7.5, there exists $w \in D$ such that $w_\tau \geq 1$ and $Supp(w) \setminus \{\tau\} \subseteq M$. Let $N = Supp(w) \setminus \{\tau\}$. By Corollary 7.5, N is an FC for τ. Since N is contained in M and M is minimal, we get $M = N = Supp(w) \setminus \{\tau\}$. Suppose that $simp_{(D,\tau,Supp)}(w)$ does not hold. Since $w_\tau \geq 1$, there exists a vector $w' \in D$ such that $w'_\tau \geq 1$ and $Supp(w') \subsetneq Supp(w)$. Let $N' = Supp(w') \setminus \{\tau\}$. By Corollary 7.5, N' is an FC for τ. Moreover, $N' = Supp(w') \setminus \{\tau\} \subsetneq Supp(w) \setminus \{\tau\} = M$, which is a contradiction.

(1) $\Leftarrow$ (2): By Corollary 7.5, M is an FC for τ. Suppose that M is not minimal. Then there exists a set N such that N is an FC for τ and $N \subsetneq M$. By Corollary 7.5, there exists $w' \in D$ such that $w'_\tau \geq 1$, $Supp(w') \setminus \{\tau\} \subseteq N$ and so $Supp(w') \subsetneq Supp(w)$, which is a contradiction. □

Suppose $w = \sum_{i=1}^{m} \alpha_i S_{i*} - \sum_{j \in Irr} \beta_j e_j$ is simple in D with respect to τ and $Supp$, and let $s = \sum_{i=1}^{m} \alpha_i S_{i*}$. By Theorem 7.7, the set $M = Supp(w) \setminus \{\tau\}$ is an MFC for τ. The following proposition shows that M is completely defined by the vector s.

Proposition 7.8. *Let D be the cone defined in equation (7.3). Let $w = \sum_{i=1}^{m} \alpha_i S_{i*} - \sum_{j \in Irr} \beta_j e_j$ be a simple vector in D with respect to τ and $Supp$ and let $s = \sum_{i=1}^{m} \alpha_i S_{i*}$. Then,*

$$Supp(w) \setminus \{\tau\} = \{i \in Irr \setminus \{\tau\} \mid s_i < 0\} \cup \{i \in Rev \setminus \{\tau\} \mid s_i \neq 0\}.$$

Proof. We have $\{i \in Irr \setminus \{\tau\} \mid s_i < 0\} \cup \{i \in Rev \setminus \{\tau\} \mid s_i \neq 0\} \subseteq Supp(w) \setminus \{\tau\}$. To show the reverse inclusion, suppose that $k \in (Supp(w) \setminus \{\tau\}) \cap Irr$, $\sum_{i=1}^{m} \alpha_i S_{ik} \geq 0$ and let

$$w' = \sum_{i=1}^{m} \alpha_i S_{i*} - \sum_{j \in Irr \setminus \{k\}} \beta_j e_j - (\sum_{i=1}^{m} \alpha_i S_{ik}) e_k.$$

We then have $w' \in D$. Since $w' = w - w_k e_k$, $w_k \neq 0$ and $w'_k = 0$, we get $Supp(w') \subsetneq Supp(w)$. Moreover, since $k \neq \tau$, we get $w'_\tau = w_\tau \geq 1$. Thus, there exists a vector $w' \in D$ such that $w'_\tau \geq 1$ and $Supp(w') \subsetneq Supp(w)$, which is a contradiction. □

The proposition above leads to an alternative characterization of the MFC M using the vector s lying in the vector subspace

$$E = \{\sum_{i=1}^{m} \alpha_i S_{i*} \mid \alpha_i \in \mathbb{R} \text{ for all } i = 1, \ldots, m\}. \tag{7.4}$$

Let $Supp' : E \rightarrow 2^{\{1,\ldots,n\}}$ be the function that maps each vector $u \in E$ to

$$Supp'(u) = \{i \in Irr \setminus \{\tau\} \mid u_i < 0\} \cup \{i \in Rev \setminus \{\tau\} \mid u_i \neq 0\}. \tag{7.5}$$

According to Proposition 7.8, we have $M = Supp'(s)$. Furthermore, as will be shown in the next lemma, s is simple in E with respect to τ and $Supp'$.

Lemma 7.9. *Let D be the cone (7.3). Let $w = \sum_{i=1}^{m} \alpha_i S_{i*} - \sum_{j \in Irr} \beta_j e_j$ be a vector in D and let $s = \sum_{i=1}^{m} \alpha_i S_{i*}$. Then, the following hold:*

1. *$Supp'(s) \subseteq Supp(w) \setminus \{\tau\}$.*
2. *If $simp_{(D,\tau,Supp)}(w)$ holds, then $simp_{(E,\tau,Supp')}(s)$ holds.*

Proof. 1. Let $k \in Supp'(s)$. We have then $k \neq \tau$. If $k \in Irr$, then $s_k < 0$, $w_k = s_k - \beta_k < 0$ and so $k \in Supp(w) \setminus \{\tau\}$. Otherwise, we get $w_k = s_k \neq 0$ and so $k \in Supp(w) \setminus \{\tau\}$. Thus, $Supp'(s) \subseteq Supp(w) \setminus \{\tau\}$.

2. Suppose that $simp_{(E,\tau,Supp')}(s)$ does not hold. Since $s_\tau \geq w_\tau \geq 1$, there exists a vector $s' \in E$ such that $s'_\tau \geq 1$ and $Supp'(s') \subsetneq Supp'(s)$. Let

$$w' = s' - \sum_{j \in Irr \setminus \{\tau\}, s'_j > 0} s'_j e_j.$$

We have $w' \in D$, $w'_\tau = s'_\tau \geq 1$ and $Supp(w') \setminus \{\tau\} = Supp'(s') \subsetneq Supp'(s)$. Moreover, since $simp_{(D,\tau,Supp)}(w)$ holds, it follows from Proposition 7.8 that $Supp'(s) = Supp(w) \setminus \{\tau\}$. Thus, there exists a vector $w' \in D$ with $w'_\tau \geq 1$ and $Supp(w') \subsetneq Supp(w)$, which is a contradiction.

□

The next theorem shows that each MFC corresponds to a simple vector in E with respect of τ and $Supp'$. MFCs can then be identified using simple vectors in E. This reduces the complexity of the MFC computation since E is generated by only m generators, while D is spanned by $m + |Irr|$ generators and defined by $|Irr|$ additional constraints.

Theorem 7.10. *Let E be the cone (7.4) and let $M \subseteq \{1, \ldots, n\} \setminus \{\tau\}$ be a set of reactions. Then, the following are equivalent:*

1. *M is an MFC for τ.*
2. *There exists a vector s simple in E with respect to τ and $Supp'$ such that $Supp'(s) = M$.*

Proof. (1) ⇒ (2): By Theorem 7.7, there exists a vector $w = \sum_{i=1}^{m} \alpha_i S_{i*} - \sum_{j \in Irr} \beta_j e_j$ simple in D with respect to τ and $Supp$, such that $M = Supp(w) \setminus \{\tau\}$. Let $s = \sum_{i=1}^{m} \alpha_i S_{i*}$. By Lemma 7.9, s is simple in E with respect to τ and $Supp'$. Moreover, by Corollary 7.8, $Supp'(s) = Supp(w) \setminus \{\tau\} = M$.

(2) ⇒ (1): Let

$$w = s - \sum_{j \in Irr \setminus \{\tau\}, s_j > 0} s_j e_j.$$

We have $w \in D$, $w_\tau = s_\tau \geq 1$ and $Supp(w) \setminus \{\tau\} = Supp'(s) = M$. Suppose that $simp_{(D,\tau,Supp)}(w)$ does not hold. Then, since $w_\tau \geq 1$, there exists a vector $w' = \sum_{i=1}^{m} \alpha'_i S_{i*} - \sum_{j \in Irr} \beta'_j e_j$ in D, such that $w'_\tau \geq 1$ and $Supp(w') \subsetneq Supp(w)$. Since

$Supp(w) \setminus \{\tau\} = Supp'(s)$ and $Supp(w') \subsetneq Supp(w)$, we get $Supp(w') \setminus \{\tau\} \subsetneq Supp'(s)$. Let $s' = \sum_{i=1}^{m} \alpha'_i S_{i*}$. We have then $s'_\tau \geq w'_\tau \geq 1$. Moreover, by Lemma 7.9, $Supp'(s') \subseteq Supp(w') \setminus \{\tau\}$ and so $Supp'(s') \subsetneq Supp'(s)$, which is a contradiction. Thus, w is simple in D with respect to τ and $Supp$ and so, following Theorem 7.7, M is an MFC for τ. □

The results above show that MFCs can be identified using vectors simple in E with respect of τ and $Supp'$. These results can also be used to determine MCSs. Indeed, an MCS can be seen as the union of an MFC and an MBC depending on the reversibility type of the target reaction. If the latter is irreversible, the empty set is the unique MBC and MFCs are identical to MCSs.

Corollary 7.11. *Let $M \subseteq \{1, \ldots, n\} \setminus \{\tau\}$ be a set of reactions. If reaction τ is irreversible, i.e., $\tau \in Irr$, then the following are equivalent:*

1. *M is an MCS for τ.*

2. *M is an MFC for τ.*

The following corollary shows that, if the target reaction is reversible, each MCS is the union of an MFC and an MBC.

Corollary 7.12. *Let $M \subseteq \{1, \ldots, n\} \setminus \{\tau\}$ be a set of reactions. If reaction τ is reversible, i.e., $\tau \in Rev$, and M is an MCS for τ, then there exist an MFC M_1 and an MBC M_2 for τ, such that $M = M_1 \cup M_2$.*

Proof. Suppose M is an MCS. Then, M is also an FC and an BC for τ. Thus, there exist two sets of reactions M_1 and M_2 such that M_1 is an MFC for τ, M_2 is an MBC for τ, $M_1 \subseteq M$ and $M_2 \subseteq M$. Then, we have $M_1 \cup M_2 \subseteq M$. Since $M_1 \cup M_2$ is also a cut set for τ and M is minimal, we get $M = M_1 \cup M_2$.

□

The mathematical results above suggest an algorithm to identify MFCs by means of vectors simple in E with respect of τ and $Supp'$. One possible approach to determine these simple vectors is the use of mixed-integer linear programming (MILP).

7.4 Algorithm for Finding MFCs

We have shown that each MFC corresponds to a simple vector in E with respect to τ and $Supp'$. To identify such a simple vector, we first introduce for each reaction $j \neq \tau$ a binary variable λ_j. By L and U we denote resp. some lower and upper bound for vectors $s \in E$, i.e., $L \leq s_j \leq U$ for all $j \in \{1, \ldots, n\} \setminus \{\tau\}$. A simple vector in E with respect to τ and $Supp'$ corresponds to an optimal solution of the following MILP:

$$
\begin{array}{ll}
\min \sum_{j \neq \tau} \lambda_j & \\
\text{subject to:} & s_\tau \geq 1, \\
& L\lambda_j \leq s_j \leq U\lambda_j, \quad \forall j \in Rev, \qquad \text{(a)} \\
& L\lambda_j \leq s_j, \quad \forall j \in Irr \setminus \{\tau\}, \qquad \text{(b)} \\
& (s_1, \ldots, s_n) \in E, \\
& \lambda_j \in \{0,1\}, \quad \forall j \in \{1, \ldots, n\} \setminus \{\tau\}.
\end{array}
$$

Conditions (a) and (b) ensure for all $j \in \{1, \ldots, n\} \setminus \{\tau\}$ that $j \in Supp'(s)$ implies $\lambda_j = 1$. Although a solution (s, λ) may exist with $\lambda_k = 1$ and $k \notin Supp'(s)$ for some $k \in \{1, \ldots, n\} \setminus \{\tau\}$, this solution is not optimal. Indeed, (s, λ') with $\lambda'_k = 0$ and $\lambda'_j = \lambda_j$ for all $j \neq k$ would be another solution with a smaller value of the objective function. For an optimal solution (s^*, λ^*), we have

$$\lambda^*_j = 1 \text{ if and only if } j \in Supp'(s^*). \tag{7.6}$$

Therefore, s^* is a simple vector in E with respect to τ and $Supp'$, and $M = Supp'(s^*) = \{j \mid \lambda^*_j = 1\}$ is an MFC.

Here, we are interested in finding all MFCs. We propose an iterative MILP that obtains, in each iteration i, an optimal solution (s^*, λ^*) and so an MFC $M^i = \{j \mid \lambda^*_j = 1\}$. We ensure the selection of a new MFC M^i by imposing the following constraint on the binary variables:

$$\sum_{j \in M^{i-1}} \lambda_j \leq |M^{i-1}| - 1, \tag{7.7}$$

where M^{i-1} is the MFC obtained in the previous iteration $i-1$. This condition signifies that at least one reaction $j \in M^{i-1}$ is not included in M^i, otherwise M^{i-1} would be contained in M^i, and since the latter is minimal, we would get $M^i = M^{i-1}$. The algorithm stops when no other optimal solution can be found.

Flux coupling analysis introduced in the previous chapter could be used to optimize our algorithm. This analysis allows identifying directional couplings between reactions. A reaction j is directionally coupled with some reaction i, written $i \overset{=0}{\rightarrow} j$, if a zero flux through i implies a zero flux through j. In this case, at most i or j is contained in an MFC. This is expressed by the following constraint:

$$\lambda_i + \lambda_j \leq 1 \quad \forall i, j \in \{1, \ldots, n\} \setminus \{\tau\} \text{ such that } i \overset{=0}{\rightarrow} j. \tag{7.8}$$

In addition, a reaction i such that $i \overset{=0}{\rightarrow} \tau$ is itself a trivial MFC. Therefore, the following constraint holds for non-trivial MFCs:

$$\lambda_i = 0 \quad \forall i \in \{1, \ldots, n\} \setminus \{\tau\} \text{ such that } i \overset{=0}{\rightarrow} \tau. \tag{7.9}$$

A typical example of coupled reactions are *enzyme subsets* [78]. Indeed, the fluxes through reactions in an enzyme subset are proportional to each other. Given the enzyme subsets of a metabolic network, we can obtain, in a pre-processing step of our

Algorithm 3 Procedure for computing minimal direction cuts.

Input: • Target reaction τ,
• Stoichiometric matrix S,
• Set of irreversible reactions $Irr \subseteq \{1, \ldots, n\}$,
• Maximal size of MFCs $MaxSize$.

Optional: • Coupled reactions
$\Phi = \{(i,j) \mid i \stackrel{=0}{\rightarrow} j \vee j \stackrel{=0}{\rightarrow} i, 1 \leq i < j \leq n, i \neq \tau, j \neq \tau\}$,
$\Psi = \{i \mid i \stackrel{=0}{\rightarrow} \tau, 1 \leq i \leq n, i \neq \tau\}$,
• Set $\Delta \subseteq \{1, \ldots, n\}$ of reactions that must not occur in MFCs,

Output: • Set $\mathcal{F}$ of minimal forward direction cuts (MFCs).

Initialization: $M := \emptyset$.

/* Trivial MFCs */
add($\{\tau\}, \mathcal{F}$);
for all $i \in \Psi$ **do**
  if $i \notin \Delta$ **then** add($\{i\}, \mathcal{F}$));
  end if
end for

repeat
  /* Find a vector simple in E with respect to τ and $Supp'$ */
  Solve the MILP problem

$$
\begin{array}{ll}
\min \sum_{j \neq \tau} \lambda_j & \\
\text{subject to:} & \sum_{i=1}^{m} \alpha_i S_{i\tau} \geq 1, \\
& L\lambda_j \leq \sum_{i=1}^{m} \alpha_i S_{ij} \leq U\lambda_j \quad \forall j \in Rev, \\
& L\lambda_j \leq \sum_{i=1}^{m} \alpha_i S_{ij} \quad \forall j \in Irr \setminus \{\tau\}, \\
& \lambda_i + \lambda_j \leq 1 \quad \forall (i,j) \in \Phi, \\
& \lambda_i = 0 \quad \forall i \in \Psi, \\
& \lambda_i = 0 \quad \forall i \in \Delta, \\
& \sum_{j \neq \tau} \lambda_j \leq \text{MaxSize} \\
& \lambda_j \in \{0,1\} \quad \forall j \in \{1, \ldots, n\} \setminus \{\tau\}, \\
& \alpha_i \in \mathbb{R} \quad \forall i \in \{1, \ldots, m\},
\end{array}
$$

  and let (α^*, λ^*) be its optimal solution.

  /* Retrieve the corresponding MFC */
  $M = \{j \mid \lambda_j^* = 1\}$;
  add(M, $\mathcal{F}$);

  /* Force the selection of a new MFC in the next iteration */
  add to the constraint system: $\sum_{j \in M} \lambda_j \leq |M| - 1$.
until No other solution is found

algorithm, a compressed network with the enzyme subsets taken as combined reactions. This compression, which is identical to that presented in [34], results in a smaller network, and so computing the MFCs for the compressed network is easier than computing those of the original network.

Our approach may be hampered by the huge number of possible minimal direction cuts. However, with our method, we can easily add new constraints to reduce the search space. For instance, from the practical viewpoint, we could require that minimal direction cuts contain only external reactions. These reactions reflect the experimental conditions the biological system is placed on and so could easily be controlled experimentally. On the other hand, while looking for minimal direction cuts for a target reaction, we should be careful not to perturb involuntarily other reactions that are important. As an example, consider again the network depicted in Fig. 7.1. Assume we wish to repress the consumption of metabolite P_1 and maintain the production of metabolite P_2. Removing reaction 3 inhibits the consumption of P_1. However, this perturbation forbids also the production of P_2. Therefore, reaction 3 should be excluded from minimal direction cuts. In general, we could define a set of reactions $\Delta \subseteq \{1, \ldots, n\}$ that should not occur in minimal direction cuts, and then impose the following additional condition on the binary variables corresponding to these reactions:

$$\lambda_i = 0 \quad \forall i \in \Delta. \tag{7.10}$$

Finally, we often are interested in removing a small number of reactions from a metabolic network. Enumerating all the MFCs is not always necessary. Our algorithm allows for defining the maximal size (MaxSize) of the MFCs to be computed by imposing the following constraint:

$$\sum_{j \neq \tau} \lambda_j \leq \text{MaxSize} \tag{7.11}$$

7.5 Computational Results for the Central Metabolism of E. coli

Our iterative algorithm, whose main steps are shown in Algorithm 3, has been implemented in the Java language. We used CPLEX 9.0 as our solver for linear and integer programming problems, accessed via Java. In order to check the capabilities of our algorithm, we computed the MFCs in the central metabolic network of *E. coli* with 'biomass reaction' as a target reaction. We considered four variants of the *E. coli* central metabolism corresponding to the growth on four different substrates (acetate, succinate, glycerol and glucose). In order to work on networks with different complexities, the authors of [106] have inserted in these variants several pseudo reactions. This results in an increase in the number of elementary modes. Recall that the difficulty of computing MCSs using the traditional algorithm grows with the number of elementary modes. On the other hand, the target reaction is irreversible, and so the MFCs and MCSs are identical. Therefore, working on these four variants also allows

Substrate	Acetate	Succinate	Glycerol	Glucose
♯reactions (n)	104	104	106	105
♯metabolites (m)	88	88	89	88
♯EMs (network complexity)	599	4250	11333	27100
♯MFCs	206	340	624	1249
Computation time	1min 28s	22m 20s	1h 14min	12h 10min

Table 7.1: Computation of MFCs for variants of the central metabolism of *E. coli*, with the number of reactions (n), the number of metabolites (m), the number of elementary modes (♯EMs) and the number of MFCs (♯MFCs). Contrary to MCSs, the computation of MFCs does not require that of elementary modes. For all computations, we defined six as the maximal size of the MFCs.

for comparing a direct computation of MCSs using our algorithm with a computation using elementary modes. Experiments for computing MCSs in the four variants of the *E. coli* metabolism using the traditional algorithm are described in [47].

Table 7.1 summarizes our computational results. The computations were performed on a Solaris server with a Sparc III Processor 900 MHz and 32 GB RAM. We did not use additional constraints (7.8), (7.9) and (7.10) that would allow for reducing the search space. The computation times of our current prototype implementation may be compared with those given in [47]. Note that the latter do not include the time needed for computing the elementary modes, which is known to be a complex combinatorial problem. In addition, our results show that the difficulty of computing MFCs grows with the number of elementary modes. This observation, which holds also for MCSs, reflects that the number of elementary modes is a measure of the complexity of the network analysis. Note however that, contrary to MCSs, the computation of MFCs does not require that of elementary modes.

CHAPTER

8 Constraint-Based Analysis of Gene Deletion in Metabolic Networks

The range of all possible behaviors, which is mathematically described by the steady-state flux cone, can be altered by gene deletion. In this chapter, we analyze the changes in the overall capabilities of a metabolic network caused by gene deletion. In particular, we show how to obtain in a constraint-based approach a description of the altered steady-state flux cone. The analysis is based on our refined classification of reactions (irreversible, pseudo-irreversible and fully reversible reactions). The work we present in this chapter has been partially published in [60].

8.1 Constraint-based Modeling of Gene Deletion

Recall that, in the context of metabolic network analysis, metabolic systems are assumed to operate at steady state such that for all internal metabolites the flux is balanced. In addition, the flux through each irreversible reaction must be non-negative. Fluxes through reversible reactions are not restricted with respect to their sign. The set of all possible flux distributions over the network at steady state defines the steady-state flux cone

$$C = \{v \in \mathbb{R}^n \mid Sv = 0,\ v_i \geq 0,\ \text{for all } i \in Irr\}. \tag{8.1}$$

Let $\tau \in \{1, \ldots, n\}$ be the target reaction associated with the deleted gene. To simulate gene deletion, we constrain the flux through reaction τ to zero. This leads to the *altered flux cone*

$$C_{\tau^0} = \{v \in \mathbb{R}^n \mid Sv = 0,\ v_\tau = 0,\ v_i \geq 0,\ \text{for all } i \in Irr\}. \tag{8.2}$$

The altered flux cone contains the full range of achievable behaviors of the altered metabolic network at steady state. Hence, it is of great interest to describe this cone in a mathematically and biologically meaningful way. This can be done using an existing description of the flux cone C. In what follows, we provide a minimal and unique description of the altered cone using minimal metabolic behaviors and the reversibility space of the original flux cone.

8.2 Characterizing Minimal Proper Faces of the Altered Flux Cone

In this section we mainly characterize the minimal proper faces of the altered flux cone C_{τ^0} defined in equation (8.2). Mathematically, this cone is also given by

$$C_{\tau^0} = C \cap \{v \in \mathbb{R}^n \mid v_\tau = 0\}.$$

Therefore, we may deduce an outer description of the altered cone C_{τ^0} from an outer description of the original cone C. First, we can easily see that the lineality space of the altered flux cone C_{τ^0} is given by

$$\text{lin.space}(C_{\tau^0}) = \text{lin.space}(C) \cap \{v \in \mathbb{R}^n \mid v_\tau = 0\}.$$

Obviously, if reaction τ is not fully reversible, i.e., $\tau \notin \mathit{Frev}$, we have $\text{lin.space}(C) \subseteq \{v \in \mathbb{R}^n \mid v_\tau = 0\}$ and so $\text{lin.space}(C_{\tau^0}) = \text{lin.space}(C)$.

In the following, we use the double description method (reviewed in Chap. 2) to determine the minimal proper faces of the altered cone C_{τ^0}. For this, we should mention that the altered flux cone C_{τ^0} is also given by

$$C_{\tau^0} = C \cap \{v \in \mathbb{R}^n \mid v_\tau \leq 0\} \cap \{v \in \mathbb{R}^n \mid -v_\tau \geq 0\}.$$

Let C_{τ^+} be the cone defined by $C_{\tau^+} = C \cap \{v \in \mathbb{R}^n \mid v_\tau \geq 0\}$. The altered cone C_{τ^0} is also given by $C_{\tau^0} = C_{\tau^+} \cap \{v \in \mathbb{R}^n \mid -v_\tau \geq 0\}$. Accordingly, using the double description method, we can deduce a description of C_{τ^0} from that of C_{τ^+}, which in turn can be computed from the description of C.

Let $J = \{G^1, \ldots, G^s\}$ be the set of minimal proper faces of C and $t = \dim(\text{lin.space}(C))$. Select for each $i = 1, \ldots, s$ a vector $g^i \in G^i \setminus \text{lin.space}(C)$, and let $R = (g^1, \ldots, g^s)$. Let $B = (b^1, \ldots, b^t)$ be a basis of $\text{lin.space}(C)$. Accordingly, (R, B) is a double description pair (DDP) of the flux cone C, i.e.,

$$C = \{x \in \mathbb{R}^n \mid x = R\lambda + B\mu \text{ for some } \lambda \in \mathbb{R}^s_{\geq 0} \text{ and } \mu \in \mathbb{R}^t\}.$$

Moreover, (R, B) is minimal, i.e., no proper submatrix of (R, B) can generate the cone C. In the following, we show how the description of the altered flux cone C_{τ^0} depends on the reversibility type of the target reaction τ. Remember that we defined the following decomposition of the reversible reaction set $\mathit{Rev} = \{1, \ldots, n\} \setminus \mathit{Irr}$:

- Prev_0: set of pseudo-irreversible reactions, i.e., $\mathit{Prev}_0 = \{i \in \mathit{Rev} \mid v_i = 0, \text{ for all } v \in \text{lin.space}(C)\}$,
- Frev: set of fully reversible reactions, i.e., $\mathit{Frev} = \mathit{Rev} \setminus \mathit{Prev}_0$.

We now will study the minimal proper faces of the altered flux cone C_{τ^0}, depending on the reversibility type of the target reaction τ.

8.2.1 Removing a Fully Reversible Reaction

If $\tau \in \mathit{Frev}$ is fully reversible, there exists a flux vector $b^k \in B$ such that $b^k_\tau \neq 0$, and so $\mathrm{lin.space}(C) \nsubseteq \{v \in \mathbb{R}^n \mid v_\tau = 0\}$. Assume that $b^k_\tau > 0$. Otherwise we can take the vector $-b^k$ since we also have $-b^k \in \mathrm{lin.space}(C)$. Moreover, for each $i = 1, \ldots, s$, there exists a vector $g^i \in G^i \setminus \mathrm{lin.space}(C)$ such that $g^i_\tau = 0$. Accordingly, assume that we choose the set R such that $g^i_\tau = 0$ for all $g^i \in R$.

Since $C_{\tau^+} = C \cap \{v \in \mathbb{R}^n \mid v_\tau \geq 0\}$ and $\mathrm{lin.space}(C) \nsubseteq \{v \in \mathbb{R}^n \mid v_\tau = 0\}$, according to the double description method, a minimal DDP (R_{τ^+}, B_{τ^+}) of C_{τ^+} is given by

- $B_{\tau^+} = (\mathfrak{b}^1, \ldots, \mathfrak{b}^{t-1})$ with $\mathfrak{b}^i = b^i - (b^i_\tau / b^k_\tau) \cdot b^k$ for all $b^i \in B \setminus \{b^k\}$,
- $R_{\tau^+} = R \cup \{b^k\}$.

Since $C_{\tau^0} = C_{\tau^+} \cap \{v \in \mathbb{R}^n \mid -v_\tau \geq 0\}$, $\mathrm{lin.space}(C_{\tau^+}) \subseteq \{v \in \mathbb{R}^n \mid v_\tau = 0\}$, $g^i_\tau = 0$ for all $g^i \in R$ and $b^k_\tau > 0$, following the double description method, a minimal DDP (R_{τ^0}, B_{τ^0}) of C_{τ^0} is given by

- $B_{\tau^0} = B_{\tau^+}$,
- $R_{\tau^0} = R$.

Based on the results above, the following proposition characterizes the minimal proper faces of the altered flux cone C_{τ^0}.

Proposition 8.1. *If $\tau \in \mathit{Frev}$ is fully reversible, the set J_{τ^0} of minimal proper faces of the altered cone C_{τ^0} is given by*

$$J_{\tau^0} = \{G^i \cap \{v \in \mathbb{R}^n \mid v_\tau = 0\} \mid G^i \in J\}.$$

Proof. Let G_{τ^0} be a minimal proper face of C_{τ^0} generated by a vector $g \in R_{\tau^0}$. Since $R_{\tau^0} = R$, we get $g = g^i$ for some $g^i \in R$. Accordingly, C_{τ^0} and G^i are given by

$$G_{\tau^0} = \{\alpha' \cdot g^i + b' \mid \alpha' \geq 0 \text{ and } b' \in \mathrm{lin.space}(C_{\tau^0})\},$$

$$G^i = \{\alpha \cdot g^i + b \mid \alpha \geq 0 \text{ and } b \in \mathrm{lin.space}(C)\}.$$

Since $\mathrm{lin.space}(C_{\tau^0}) \subseteq \mathrm{lin.space}(C)$, we have $G_{\tau^0} \subseteq G^i \cap \{v \in \mathbb{R}^n \mid v_\tau = 0\}$. To show the reverse inclusion, suppose $v \in G^i \cap \{v \in \mathbb{R}^n \mid v_\tau = 0\}$. There exist $\alpha \geq 0$ and $b \in \mathrm{lin.space}(C)\}$ such that $v = \alpha \cdot g^i + b$ and $v_\tau = \alpha g^i_\tau + b_\tau = 0$. Since $g^i_\tau = 0$, we get $b_\tau = 0$ and so $b \in \mathrm{lin.space}(C_{\tau^0})$. Therefore, $v \in G_{\tau^0}$. We conclude that

$$G_{\tau^0} = G^i \cap \{v \in \mathbb{R}^n \mid v_\tau = 0\}.$$

□

8.2.2 Removing a (Pseudo-) Irreversible Reaction

If $\tau \in Irr \cup Prev_0$ is (pseudo-) irreversible, we have $\mathrm{lin.space}(C) \subseteq \{v \in \mathbb{R}^n \mid v_\tau = 0\}$ and so

$$\mathrm{lin.space}(C_{\tau^0}) = \mathrm{lin.space}(C).$$

Consider the hyperplane $H = \{v \in \mathbb{R}^n \mid v_\tau = 0\}$ and let $H^+ = \{v \in \mathbb{R}^n \mid v_\tau > 0\}$ (resp. $H^- = \{v \in \mathbb{R}^n \mid v_\tau < 0\}$) be the positive (resp. negative) half-space supported by the hyperplane H. We partition the set R into three parts:

$$\begin{aligned} R^+ &= \{g^i \in R \mid g^i_\tau > 0\}, \\ R^0 &= \{g^i \in R \mid g^i_\tau = 0\}, \\ R^- &= \{g^i \in R \mid g^i_\tau < 0\}. \end{aligned}$$

We should mention that, since $\tau \in Irr \cup Prev_0$, for each $i = 1, \ldots, s$, $g^i_\tau > 0$ (resp. $g^i_\tau < 0$, $g^i_\tau = 0$) if and only if $G^i \setminus \mathrm{lin.space}(C) \subseteq H^+$ (resp. $G^i \setminus \mathrm{lin.space}(C) \subseteq H^-$, $G^i \setminus \mathrm{lin.space}(C) \subseteq H$). Accordingly, H partitions the set J of minimal proper faces of C into three parts:

$$\begin{aligned} J^+ &= \{G^i \in J \mid G^i \setminus \mathrm{lin.space}(C) \subseteq H^+\}, \\ J^0 &= \{G^i \in J \mid G^i \setminus \mathrm{lin.space}(C) \subseteq H\}, \\ J^- &= \{G^i \in J \mid G^i \setminus \mathrm{lin.space}(C) \subseteq H^-\}. \end{aligned}$$

Since $C_{\tau^+} = C \cap \{v \in \mathbb{R}^n \mid v_\tau \geq 0\}$ and $\mathrm{lin.space}(C) \subseteq \{v \in \mathbb{R}^n \mid v_\tau = 0\}$, following the double description method, a minimal DDP (R_{τ^+}, B_{τ^+}) of C_{τ^+} is given by

- $B_{\tau^+} = B$,
- $R_{\tau^+} = R^+ \cup R^0 \cup Adj$ with $Adj = \{g^k_\tau \cdot g^l - g^l_\tau \cdot g^k \mid G^k \in J^+,\ G^l \in J^-,\ G^k \text{ and } G^l \text{ are adjacent in } C\}$.

Recall that two minimal proper faces of C are adjacent if they are contained in one face of C of dimension $t + 2$.

Since $C_{\tau^0} = C_{\tau^+} \cap \{v \in \mathbb{R}^n \mid -v_\tau \geq 0\}$ and $\mathrm{lin.space}(C_{\tau^+}) \subseteq \{v \in \mathbb{R}^n \mid v_\tau = 0\}$, according to the double description method, a minimal DDP (R_{τ^0}, B_{τ^0}) of C_{τ^0} is given by

- $B_{\tau^0} = B_{\tau^+} = B$,
- $R_{\tau^0} = R^0 \cup Adj$.

In the following, we characterize the minimal proper faces of the altered flux cone C_{τ^0}.

Definition 8.2. *For each pair of minimal proper faces $G^k \in J^+$ and $G^l \in J^-$, let*

$$comb(G^k, G^l) = \{g^k_\tau \cdot g^l - g^l_\tau \cdot g^k \mid g^k \in G^k \text{ and } g^l \in G^l\}.$$

Note that, given two minimal proper faces $G^k \in J^+$ and $G^l \in J^-$, we have $comb(G^k, G^l) \subseteq C \cap \{v \in \mathbb{R}^n \mid v_\tau = 0\}$ and so $comb(G^k, G^l) \subseteq C_{\tau^0}$.

Proposition 8.3. *If $\tau \in Irr \cup Prev_0$ is (pseudo-) irreversible, the set J_{τ^0} of minimal proper faces of the altered cone C_{τ^0} is given by*

$$J_{\tau^0} = J^0 \cup \{comb(G^k, G^l) \mid G^k \in J^+,\ G^l \in J^-,\ G^k \text{ and } G^l \text{ are adjacent in } C\}.$$

Proof. Let G_{τ^0} be a minimal proper face of C_{τ^0} generated by a vector $g \in R_{\tau^0}$. Since $R_{\tau^0} = R^0 \cup Adj$, either $g \in R^0$ or $g \in Adj$.

– If $g \in R^0$, we have $g = g^i$ for some $g^i \in R^0$. Accordingly, C_{τ^0} and G^i are given by

$$G_{\tau^0} = \{\alpha' \cdot g^i + b' \mid \alpha' \geq 0 \text{ and } b' \in \text{lin.space}(C_{\tau^0})\},$$

$$G^i = \{\alpha \cdot g^i + b \mid \alpha \geq 0 \text{ and } b \in \text{lin.space}(C)\}.$$

It follows from $\text{lin.space}(C_{\tau^0}) = \text{lin.space}(C)$ that $G_{\tau^0} = G^i$ and so $G_{\tau^0} \in J^0$.

– If $g \in Adj$, there exist $G^k \in J^+$ and $G^l \in J^-$ that are adjacent and $g = g^k_\tau \cdot g^l - g^l_\tau \cdot g^k$. Accordingly, C_{τ^0} is given by

$$G_{\tau^0} = \{\alpha \cdot (g^k_\tau \cdot g^l - g^l_\tau \cdot g^k) + b \mid \alpha \geq 0 \text{ and } b \in \text{lin.space}(C)\}.$$

We can easily see that $comb(G^k, G^l) \subseteq G_{\tau^0}$. To show the reverse inclusion, suppose $v \in G_{\tau^0} \setminus \text{lin.space}(C)$. There exist $\alpha > 0$ and $b \in \text{lin.space}(C)$ such that $v = \alpha \cdot (g^k_\tau \cdot g^l - g^l_\tau \cdot g^k) + b$. Let $g^1 = g^k - (1/\alpha) \cdot b$ and $g^2 = \alpha \cdot g^l$. We have $g^1 \in G^k$, $g^2 \in G^l$ and $v = g^1_\tau \cdot g^2 - g^2_\tau \cdot g^1$. Therefore, $v \in comb(G^k, G^l)$, and so $G_{\tau^0} \subseteq comb(G^k, G^l)$. We conclude that $G_{\tau^0} = comb(G^k, G^l)$ and the claim follows.

□

If $J^+ = \emptyset$ or $J^- = \emptyset$, which is particularly the case if $\tau \in Irr$ is irreversible, we get $J_{\tau^0} = J^0$. Accordingly, the set of minimal proper faces of C_{τ^0} is the set of minimal proper faces G of C for which there exists no vector $v \in G$ such that $v_\tau \neq 0$.

In the following, we assume that $J^+ \neq \emptyset$ and $J^- \neq \emptyset$. In addition to minimal proper faces $G \in J^0$, J_{τ^0} contains the minimal proper faces $comb(G^k, G^l)$ where $G^k \in J^+$ and $G^l \in J^-$ are adjacent. In the following, we further characterize these adjacent faces.

Lemma 8.4. *Let $G^k \in J^+$ and $G^l \in J^-$ be two minimal proper faces of C and let D^k and D^l their characteristic sets, respectively. Let $G = \{v \in C \mid v_i = 0$ for all $i \in Irr \setminus (D^k \cup D^l)\}$. The following are equivalent:*

1. *G^k and G^l are adjacent.*
2. *G is of dimension $t + 2$.*

Proof. (1) $\Rightarrow$ (2): Since G is a face of C that contains two different faces G^k and G^l, G is not a minimal proper face of C, and so $\dim(G) \geq t+2$. On the other hand, G^k and G^l are adjacent and so are contained in a face F of C such that $\dim(F) = t+2$. Let $F = \{v \in C \mid v_i = 0 \text{ for all } i \in Irr \setminus D\}$ for some $D \subseteq Irr$. Since $G^k \subseteq F$ and $G^l \subseteq F$, we get $D^k \subseteq D$ and $D^l \subseteq D$ and so $D^k \cup D^l \subseteq D$. Therefore, $G \subseteq F$, $\dim(G) \leq t+2$ and so the claim follows.

(2) $\Rightarrow$ (1): G^k and G^l are contained in a face G of dimension $t+2$ and so are adjacent. $\square$

As we pointed out before, given two minimal proper faces $G^k \in J^+$ and $G^l \in J^-$ which are adjacent, $comb(G^k, G^l)$ is a minimal proper face of C_{τ^0}. While this face is defined by flux vectors from G^k and G^l (see Definition 8.2), it is also interesting to determine the irreversible reactions which characterize this face. The following proposition states that the characteristic set of $comb(G^k, G^l)$ is the set $D^k \cup D^l$.

Proposition 8.5. *Let $G^k \in J^+$ and $G^l \in J^-$ be two minimal proper faces of C and let D^k and D^l their characteristic sets, respectively. Let $G = \{v \in C \mid v_i = 0$ for all $i \in Irr \setminus (D^k \cup D^l)\}$. If $\tau \in Prev_0$ is pseudo-irreversible and the faces G^k and G^l are adjacent, then*

$$comb(G^k, G^l) = G \cap \{v \in \mathbb{R}^n \mid v_\tau = 0\},$$

and so the characteristic set of $comb(G^k, G^l)$ is the set $D^k \cup D^l$.

Proof. Let $G' = G \cap \{v \in \mathbb{R}^n \mid v_\tau = 0\}$. Consider $g \in comb(G^k, G^l)$. We have $g_\tau = 0$ and $g_i = 0$ for all $i \in Irr \setminus (D^k \cup D^l)$. Therefore, $g \in G'$ and so $comb(G^k, G^l) \subseteq G'$. We now will show the reverse inclusion. Since $\tau \in Prev_0$, $\dim(\text{lin.space}(C_{\tau^0})) = \dim(\text{lin.space}(C)) = t$. Since G^k and G^l are adjacent, $comb(G^k, G^l)$ is a minimal proper face of C_{τ^0}. Therefore, $\dim(comb(G^k, G^l)) = t+1$ and so $\dim(G') \geq t+1$. In addition, according to Lemma 8.4, G is a face of C of dimension $t+2$. Since $G^k \subseteq G$ and $G^l \in J^+$, we get $G \nsubseteq \{v \in \mathbb{R}^n \mid v_\tau = 0\}$. It follows that $\dim(G') \leq \dim(G) - 1$ and so $\dim(G') \leq t+1$. We can now conclude that G' is a face of C_{τ^0} of dimension $t+1$ and so is a minimal proper face of C_{τ^0}. Since $comb(G^k, G^l) \subseteq G'$ and both $comb(G^k, G^l)$ and G' are minimal proper faces of C_{τ^0}, the claim follows. $\square$

The following proposition states that the pair $(G^k, G^l) \in J^+ \times J^-$ defining a minimal proper face $comb(G^k, G^l)$ of C_{τ^0} is unique.

Proposition 8.6. *Let $G^k \in J^+$ and $G^l \in J^-$ be two minimal proper faces of C and let D^k and D^l their characteristic sets, respectively. Let $G = \{v \in C \mid v_i = 0$ for all $i \in Irr \setminus (D^k \cup D^l)\}$. If $\tau \in Prev_0$ is pseudo-irreversible and the faces G^k and G^l are adjacent, then G contains exactly two minimal proper faces of C, i.e., G^k and G^l.*

Proof. Suppose there exists $G^p \in J$ such that $G^p \subseteq G$. By Lemma 8.4, since G^k and G^l are adjacent, we have $\dim(G) = t+2$. Moreover, it follows from $\dim(G^k) = t+1$, $G^k \subseteq G$ and $g^l \notin G^k$ that $(b^1, \ldots, b^t, g^k, g^l)$ is a basis of $\text{lin}(G)$. Accordingly, since $g^p \in \text{lin}(G)$, there exist $\alpha_k, \alpha_l \in \mathbb{R}$ and $b \in \text{lin.space}(C)$ such that $g^p = \alpha_k \cdot g^k + \alpha_l \cdot g^l + b$. Since $D^k \nsubseteq D^l$ and $D^l \nsubseteq D^k$, there exist $i \in D^k \setminus D^l$ and $j \in D^l \setminus D^k$. We

have $g_i^p = \alpha_k g_i^k \geq 0$ and $g_j^p = \alpha_l g_j^l \geq 0$. Therefore, $\alpha_k \geq 0$ and $\alpha_l \geq 0$. Moreover, since $g^p \notin \text{lin.space}(C)$, $\alpha_k \neq 0$ or $\alpha_l \neq 0$, or equivalently, $D^k \subseteq D^p$ or $D^l \subseteq D^p$. Since $G^k, G^l, G^p \in J$, we conclude that $G^p = G^k$ or $G^p = G^l$. □

8.3 Minimal Metabolic Behaviors of the Altered Flux Cone

Minimal metabolic behaviors of the altered flux cone C_{τ^0} are in a 1-1 correspondence with minimal proper faces of C_{τ^0}. Using the results from the preceding section, this observation leads to the following results.

8.3.1 Removing a Fully Reversible Reaction

If $\tau \in Frev$ is fully reversible, the unique effect of removing reaction τ is the reduction of the dimension of the reversible metabolic space RMS_{τ^0}, i.e.,

$$\dim(RMS_{\tau^0}) = \dim(RMS) - 1.$$

The MMBs of C_{τ^0} are the same as those of C.

8.3.2 Removing an Irreversible Reaction

If $\tau \in Irr$ is irreversible, the reversible metabolic space does not change, i.e.,

$$RMS_{\tau^0} = RMS.$$

In analogy with elementary modes, we can state the following *conservation property* for minimal metabolic behaviors: if the flux through an irreversible reaction is constrained to zero, the set of MMBs of the altered flux cone C_{τ^0} is the set of all MMBs of the flux cone C which do not involve this reaction. Accordingly, if $\tau \in Irr$ is irreversible, the conservation property of minimal metabolic behaviors guarantees that the MMBs of C_{τ^0} are exactly the MMBs D of C for which $j \notin D$.

8.3.3 Removing a Pseudo-irreversible Reaction

If $\tau \in Prev_0$ is pseudo-irreversible, the reversible metabolic space does not change, i.e.,

$$RMS_{\tau^0} = RMS.$$

Let M^0 (resp. M^+, M^-) be the set of characteristic sets of a minimal proper face $G \in J^0$ (resp. $G \in J^+, G \in J^-$), and let M_{τ^0} be the set of MMBs of C_{τ^0}. If $M^+ = \emptyset$ or $M^- = \emptyset$, then $M_{\tau^0} = M^0$. Accordingly, the set of MMBs of C_{τ^0} is the set of MMBs of C whose auxiliary sets do not contain τ. Remember that if G a is minimal proper face of C characterized by an MMB D, the auxiliary set of D is the set of pseudo-irreversible reactions $j \in Prev_0$ for which there exists $v \in G$ such that $v_j \neq 0$.

In the following, we assume that $M^+ \neq \emptyset$ and $M^- \neq \emptyset$. In addition to the MMBs $D \in M^0$, M_{τ^0} contains those which are combinations of the MMBs of C. The next proposition defines an algebraic characterization of these MMBs.

Proposition 8.7. *Let $\tau \in Prev_0$ be a pseudo-irreversible reaction and let $t = \dim(\text{lin.space}(C))$. Given two MMBs $D^k \in M^+$, $D^l \in M^-$ of the flux cone C, the following are equivalent*

1. *$D^k \cup D^l$ is an MMB of the altered cone C_{τ^0}.*
2. $\text{rank}(S_{*D^k \cup D^l \cup Rev}) = |D^k \cup D^l| + |Rev| - t - 2$.

Proof. According to Propositions 8.3 and 8.5 and Lemma 8.4, $D^k \cup D^l$ is an MMB of the altered cone C_{τ^0} if and only if $G = \{v \in C \mid v_i = 0 \text{ for all } i \in Irr \setminus (D^k \cup D^l)\}$ is a face of C such that

$$\dim(\text{lin}(G)) = t + 2. \tag{8.3}$$

Let $I \in \mathbb{R}^{n \times n}$ be the identity matrix and let $\overline{D} = Irr \setminus \{D^k \cup D^l\}$. Since $\text{lin}(G)$ is the null space of the matrix

$$\begin{pmatrix} S \\ I_{\overline{D}*} \end{pmatrix},$$

(8.3) is equivalent to

$$\text{rank}(\begin{pmatrix} S \\ I_{\overline{D}*} \end{pmatrix}) = n - t - 2. \tag{8.4}$$

Using some row operations, we get the following equation

$$\text{rank}(\begin{pmatrix} S \\ I_{\overline{D}*} \end{pmatrix}) = |\overline{D}| + \text{rank}(S_{*D^k \cup D^l \cup Rev}). \tag{8.5}$$

We also know that

$$n = |Rev| + |Irr|. \tag{8.6}$$

Combining equations (8.4), (8.5) and (8.6), we obtain

$$\text{rank}(S_{*D^k \cup D^l \cup Rev}) = |D^k \cup D^l| + |Rev| - t - 2. \tag{8.7}$$

□

We can use statement (8.7) to define an upper bound on the cardinality of all pairs of MMBs $D^k \in M^+$ and $D^l \in M^-$ for which $D = D^k \cup D^l$ can be an MMB of C_{τ^0}. Indeed, we know that

$$\text{rank}(S_{*D^k \cup D^l \cup Rev}) \leq \text{rank}(S) \tag{8.8}$$

holds for all MMBs $D^k \in M^+$ and $D^l \in M^-$ of C. Accordingly, (8.7) and (8.8) lead to the following corollary.

Corollary 8.8. *Let $D^k \in M^+$, $D^l \in M^-$ two MMBs of the flux cone C. The set $D^k \cup D^l$ can be an MMB of the altered flux cone C only if*

$$|D^k \cup D^l| \leq \text{rank}(S) - |Rev| + t + 2. \tag{8.9}$$

In the following proposition, we provide a combinatorial characterization of the MMBs of C that can be combined to generate the MMBs of C_{τ^0}. First we need the following lemma.

Lemma 8.9. *For each MMBs $D^k \in M^+$, $D^l \in M^-$ of the flux cone C, the set $D^k \cup D^l$ is a metabolic behavior (MB) of the altered cone C_{τ^0}.*

Proof. Let $D^k \in M^+$, $D^l \in M^-$ be two MMBs of C. We have $D^k = \{i \in Irr \mid g_i^k \neq 0\},\ D^l = \{i \in Irr \mid g_i^l \neq 0\},\ g_\tau^k > 0$ and $g_\tau^l < 0$. Let $g = g_\tau^k \cdot g^l - g_\tau^l \cdot g^k$. We have $g \in C,\ g_\tau = 0$ and so $g \in C_{\tau^0}$. Moreover, $D^k \cup D^l = \{i \in Irr \mid g_i \neq 0\}$. Therefore, the set $D^k \cup D^l$ is an MB of the altered cone C_{τ^0}. □

Proposition 8.10. *Given two MMBs $D^k \in M^+$, $D^l \in M^-$ of the flux cone C, the following are equivalent*

1. *$D^k \cup D^l$ is an MMB of the altered cone C_{τ^0}.*
2. *For each MMB D^i of C, $D^i \subseteq D^k \cup D^l$ implies that $D^i = D^k$ or $D^i = D^l$.*

Proof. (1) ⇒ (2): Let D^i be an MMB of C such that $D^i \subseteq D^k \cup D^l$. Let $G^i, G^k, G^l \in J$ the minimal proper faces of C whose characteristic sets are D^i, D^k, D^l, respectively, and let $G = \{v \in C \mid v_i = 0 \text{ for all } i \in Irr \setminus (D^k \cup D^l)\}$. It follows from $D^i \subseteq D^k \cup D^l$ that $G^i \subseteq G$. Since $D^k \cup D^l$ is an MMB of the altered cone C_{τ^0}, G^k and G^l are adjacent in C. By Proposition 8.6, we get $G^i = G^k$ or $G^i = G^l$, or equivalently, $D^i = D^k$ or $D^i = D^l$.

(2) ⇒ (1): According to Lemma 8.9, $D^k \cup D^l$ is an MB of C_{τ^0}. Suppose there exists an MMB D of C_{τ^0} such that $D \subseteq D^k \cup D^l$. If $D \in M^0$, then D is an MMB of C and so $D = D^k$ or $D = D^l$. This is a contradiction since $D^k \notin M^0$ and $D^l \notin M^0$. Therefore, $D \notin M^0$ and so there exist $D^p \in M^+$ and $D^q \in M^-$ such that $D = D^p \cup D^q$. Accordingly, $D^p \subseteq D$ and $D^q \subseteq D$. Since $D \subseteq D^k \cup D^l$, we get $D^p \subseteq D^k \cup D^l,\ D^q \subseteq D^k \cup D^l$, and so $D^p = D^l$ and $D^q = D^l$. Hence, $D \subseteq D^k \cup D^l$ implies $D = D^k \cup D^l$ and so $D^k \cup D^l$ is an MMB of C_{τ^0}. □

Example 8.11. *In the metabolic network from Fig. 8.1, $\dim(RMS) = 2$. The MMBs and the corresponding minimal proper faces of the flux cone C are as follows:*

$$D^1 = \{2\}, \quad D^2 = \{6, 7\}, \quad D^3 = \{6, 8\},$$
$$G^k = \{v \in C \mid v_i = 0, i \in Irr \setminus D^k\}, \quad k = 1, 2, 3.$$

Consider the following cases:

1. *$\tau = 3$: $\tau \in Frev$ and so $\dim(RMS_{\tau^0}) = \dim(RMS) - 1 = 1$. The MMBs and the corresponding minimal proper faces of the altered cone C_{τ^0} are as follows:*

$$D_{\tau^0}^k = D^k, \quad k = 1, 2, 3,$$
$$G_{\tau^0}^k = G^k \cap \{v \in \mathbb{R}^n \mid v_\tau = 0\}, \quad k = 1, 2, 3.$$

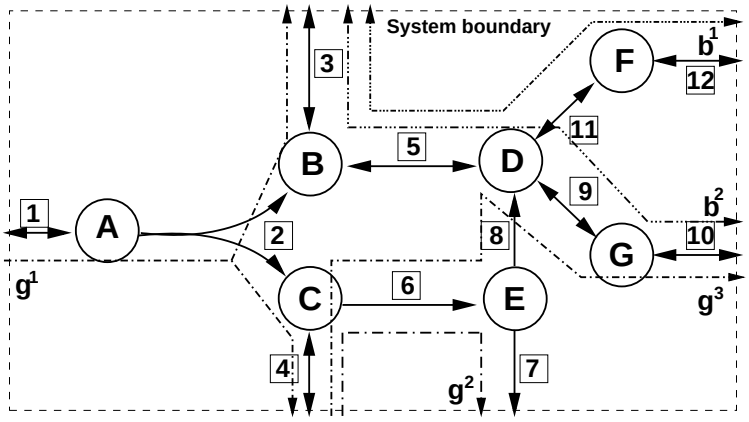


Figure 8.1: Representative pathways in ILLUSNET

2. *$\tau = 8$: $\tau \in Irr$ and so $RMS_{\tau^0} = RMS$. Moreover, $J^0 = \{D^1, D^2\}$. Therefore, the MMBs and the corresponding minimal proper faces of the altered cone C_{τ^0} are as follows:*

$$D^k_{\tau^0} = D^k, \quad k = 1, 2,$$
$$G^k_{\tau^0} = G^k, \quad k = 1, 2.$$

3. *$\tau = 1$: $\tau \in Prev_0$ and so $RMS_{\tau^0} = RMS$. Moreover, $J^0 = \{D^2, D^3\}$ and $J^- = \emptyset$. Therefore, the MMBs and the corresponding minimal proper faces of the altered cone C_{τ^0} are as follows:*

$$D^{k-1}_{\tau^0} = D^k, \quad k = 2, 3,$$
$$G^{k-1}_{\tau^0} = G^k, \quad k = 2, 3.$$

4. *$\tau = 4$: $\tau \in Prev_0$ and so $RMS_{\tau^0} = RMS$. Moreover, $J^0 = \emptyset$, $J^+ = \{D^1\}$ and $J^- = \{D^2, D^3\}$. Since $D^1 \in J^+$, $D^2, D^3 \in J^-$, $D^3 \not\subseteq D^1 \cup D^2$ and $D^2 \not\subseteq D^1 \cup D^3$, the MMBs and the corresponding minimal proper faces of the altered cone C_{τ^0} are as follows:*

$$D^1_{\tau^0} = D^1 \cup D^2, \quad D^2_{\tau^0} = D^1 \cup D^3,$$
$$G^k_{\tau^0} = \{v \in C \mid v_i = 0, i \in Irr \setminus D^k_{\tau^0}\} \cap \{v \in \mathbb{R}^n \mid v_\tau = 0\}, \quad k = 1, 2.$$

The results above can be extended to predict the effect on the flux cone when constraining the reversibility of some reaction. If a reversible reaction ι is constrained to operate in the forward (resp. backward) direction only, the altered flux cone will be $C_{\iota^+} = C \cap \{v \in \mathbb{R}^n \mid v_\iota \geq 0\}$ (resp. $C_{\iota^-} = C \cap \{v \in \mathbb{R}^n \mid v_\iota \leq 0\}$). Again, the description of C_{ι^+} and C_{ι^-} can be deduced from that of the flux cone C depending on the reversibility type of reaction ι. Indeed, if $\iota \in Prev_0$ is pseudo-irreversible, the MMBs of C_{ι^+} (resp. C_{ι^-}) are the MMBs of the altered cone $C_{\iota^0} = C \cap \{v \in \mathbb{R}^n \mid v_\iota = 0\}$, together with the MMBs $D \in M^+$ (resp. $D \in M^-$) of the flux cone C. The reversible metabolic space does not change, i.e., $RMS_{\iota^+} = RMS_{\iota^-} = RMS$. On the other hand, if $\iota \in Frev$ is fully reversible, the MMBs of C_{ι^+} and C_{ι^-} are the MMBs of C, together with a new MMB $D = \{\iota\}$ and $\dim(RMS_{\iota^+}) = \dim(RMS_{\iota^-}) = \dim(RMS) - 1$.

8.4 Reaction Reversibility in the Altered Network

We conclude this chapter by studying the changes in the reversibility type of reactions following a gene deletion. The set of irreversible (resp. reversible) reactions Irr' (resp. Rev') in the altered network is the same as those of the original network, i.e., $Irr' = Irr$ (resp. $Rev' = Rev$). Accordingly, we restrict ourselves to study the changes in the reversibility type of reversible reactions.

Possibly many fully reversible reactions in the original network may become pseudo-irreversible in the altered network. Let $Prev'_0$ and $Frev'$ be the sets of pseudo-irreversible and fully reversible reactions in the altered network, respectively, i.e.,

$$\begin{aligned} &Frev' = \{i \in Rev \mid b'_i \neq 0, \text{ for some } b' \in \mathrm{lin.space}(C_{\tau^0})\}, \\ &Prev'_0 = Rev \setminus Frev'. \end{aligned}$$

Since $\mathrm{lin.space}(C_{\tau^0}) \subseteq \mathrm{lin.space}(C)$, we have $Frev' \subseteq Frev \setminus \{\tau\}$ and $Prev_0 \subseteq Prev'_0$. In the following, based on the description of the altered flux cone given in Sect. 8.2, we further characterize the relationship between $Frev'$ (resp. $Prev'_0$) and $Frev$ (resp. $Prev_0$).

Let Δ be the set of fully reversible reactions of the original network which become pseudo-irreversible in the altered network, i.e.,

$$\Delta = Frev \setminus Frev' = Prev'_0 \setminus Prev_0.$$

We can easily see that $\Delta = \{j \in Frev \mid b_\tau = 0 \text{ implies } b_j = 0, \text{ for all } b \in \mathrm{lin.space}(C)\}$. The following proposition further characterizes the set Δ using a basis of the lineality space of C.

Proposition 8.12. *Let $B = (b^1, \ldots, b^t)$ be a basis of the lineality space $\mathrm{lin.space}(C)$. Then,*

$$\Delta = \{j \in Frev \mid \textit{there exists } \lambda \neq 0 \textit{ such that } b^i_j = \lambda b^i_\tau \textit{ for all } i = 1, \ldots, t\}.$$

Proof. Let $\Omega = \{j \in Frev \mid \text{there exists } \lambda \neq 0 \text{ such that } b^i_j = \lambda b^i_\tau \text{ for all } i = 1, \ldots, t\}$. Then $\Omega \subseteq \Delta$. To show the reverse inclusion, suppose $j \in \Delta$. Since $j \in Frev$, there exists $b \in B$ such that $b_j \neq 0$. Since $j \in \Delta$, we have $b_\tau \neq 0$. Let $b' \in B$ and let $w = b' - (b'_\tau/b_\tau) \cdot b$. We have $w \in \mathrm{lin.space}(C) \cap \{v \in \mathbb{R}^n \mid v_\tau = 0\}$ and $w_j = b'_j - (b_j/b_\tau)b'_\tau$. Since $j \in \Delta$, we get $w_j = 0$ and so $b'_j/b'_\tau = b_j/b_\tau \overset{\text{def}}{=} \lambda \neq 0$, independently from b'. □

Corollary 8.13. *If $\tau \in Irr \cup Prev_0$ is (pseudo-) irreversible, then $Frev' = Frev$ and $Prev'_0 = Prev_0$.*

Proof. Suppose $\tau \in Irr \cup Prev_0$. Then, $b^k_\tau = 0$ for all $k = 1, \ldots, t$. Consider $j \in Frev$. There exists $b \in B$ such that $b_j \neq 0$. Since $b_\tau = 0$, it follows that $j \notin \Delta$. Therefore, $\Delta = \emptyset$ and so the claim follows. □

Example 8.14. *Consider again the metabolic network from Fig. 8.1. The sets* $Prev_0$ *and* $Frev$ *of pseudo-irreversible and fully reversible reactions in this network are respectively:*

$$Prev_0 = \{1, 4\},$$
$$Frev = \{3, 5, 9, 10, 11, 12\}.$$

If $\tau = 2$*, we have* $\Delta = \emptyset$ *and so*

$$Prev'_0 = Prev_0,$$
$$Frev' = Frev.$$

If $\tau = 3$*, we have* $\Delta = \{3, 5\}$ *and so*

$$Prev'_0 = \{1, 3, 4, 5\},$$
$$Frev' = \{9, 10, 11, 12\}.$$

CHAPTER

9 Control-effective Flux Analysis

The importance of single reactions for the overall metabolic network performance can be assessed using knockout mutations. It has been suggested that a reaction is crucial if its removal from the network prevents certain critical metabolic functions [2; 15; 82]. Such an important reaction can be identified using flux coupling analysis or gene deletion analysis discussed in Chap. 6 and Chap. 8, respectively. Alternatively, the essentiality of some reaction could correlate with how this reaction participates in flexible and efficient operations of the metabolic network. The *flexibility* of the latter can be defined as its capability to adapt to different environmental conditions. An *efficient* operation corresponds to a flux distribution which carries out an optimal outcome, such as maximal growth rate, while using a minimum investment, i.e., the sum of all absolute fluxes. In this chapter, we discuss *control-effective flux (CEF) analysis* [106], which has proved promising in assessing a metabolism. The CEFs, which are directly computed from elementary modes, indicate the importance of each reaction for the overall metabolic network. After discussing the main advantages of using elementary modes in CEF analysis, we consider the use of a minimal generating set of the flux cone in such an analysis. To compare both approaches, we compute the CEFs for the *red blood cell* and *S. cerevesiae* metabolisms.

9.1 Definitions

In the following, we generalize the concept of *efficiency* which has initially been defined for elementary modes [106]. In a first step, we define a set of irreversible reactions, $Mfun \subseteq Irr$, as the basis for the main metabolic functions (e.g., growth, ATP maintenance). In addition, we assume that the operation of each reaction $\tau \in Mfun$ requires a non-zero flux through a certain irreversible reaction $\iota \in Irr$. For instance, reaction ι may correspond to the total substrate uptake.

Afterwards, for each reaction $\tau \in Mfun$, the *efficiency* w.r.t. reaction τ of a flux vector $v \in C \setminus \{0\}$, denoted by $Eff(v, \tau)$, is defined as the flux through reaction τ divided by the sum of all absolute fluxes through reactions participating in v,

$$Eff(v, \tau) = \frac{v_\tau}{\|v\|_1}, \tag{9.1}$$

where $\|v\|_1$ is the L_1 norm of the flux vector v, i.e., $\|v\|_1 = \sum_{i=1}^{n} |v_i|$. The efficiency

of v, denoted by $\mathit{Eff}(v)$, is defined as the sum of the efficiencies w.r.t. reactions in Mfun,

$$\mathit{Eff}(v) = \sum_{\tau \in \mathit{Mfun}} \mathit{Eff}(v, \tau).$$

According to equation (9.1), the efficiency $\mathit{Eff}(v, \tau)$ measures the flux through reaction τ in the flux distribution v while taking into consideration the investment needed to establish that flux distribution. Among flux vectors carrying the same fluxes through reaction τ, those with smaller investment are more efficient than the others. This is in agreement with the suggested optimization-based approach [39], which assumes that the optimal flux distribution is the one which has the minimal investment. In general, given two flux vectors $v^1, v^2 \in C$, if $\mathit{Eff}(v^1) \geq \mathit{Eff}(v^2)$, then we say v^1 is *at least as efficient as* v^2.

For the remaining of this chapter, we need to define *normalized flux vectors.*

Definition 9.1. *A flux vector $v \in C$ is called* normalized *(by the flux through reaction ι) if $v_\iota \in \{0, 1\}$.*

Now, let ϑ be the function that maps each finite set $E \subseteq C$ of normalized flux vectors to the *effectiveness vector* $w = \vartheta(E)$ given by

$$w_i = \sum_{\tau \in \mathit{Mfun}} \frac{1}{\sigma_\tau} \sum_{e \in E} \frac{\mathit{Eff}(e, \tau)}{\sum_{e' \in E} \mathit{Eff}(e', \tau)} \cdot |e_i| \quad \text{for all } i = 1, \ldots, n, \tag{9.2}$$

with $\sigma_\tau = \max\{e_\tau : e \in E\}$ for all $\tau \in \mathit{Mfun}$. For all $i \in \{1, \ldots, n\}$, w_i can be seen as the average of absolute fluxes through reaction i in all flux vectors $e \in E$. The latter are weighted by their relative efficiencies. If E is the set of elementary modes, for all $i \in \{1, \ldots, n\}$, the component $\vartheta(E)_i$ is identical to the *control-effective flux (CEF)* of reaction i [106]. It has been suggested in [106] that a reaction with a high CEF is crucial for the overall metabolic network performance. In general, given a subset $E \subseteq C$ of normalized flux vectors, the effectiveness vector $\vartheta(E)$ measures the importance of reactions for the flux vectors in E while taking into account the efficiencies of these flux vectors. Below we consider the extent to which the conclusions drawn from the effectiveness vector $\vartheta(E)$ about reaction essentiality can be generalized to the whole flux cone.

9.2 Effectiveness Sensitivity

Given a finite set $E \subseteq C \backslash \{0\}$, flux vectors $e \in E$ such that $\mathit{Eff}(e) = 0$, or equivalently, $e_\tau = 0$ for all $\tau \in \mathit{Mfun}$, do not influence the components of $\vartheta(E)$. This is particularly the case of flux vectors in $E \cap \mathrm{lin.space}(C)$.

Proposition 9.2. *Let $E \subseteq C \setminus \{0\}$ be a subset of normalized flux vectors and let $\mathcal{E} = \{e \in E \mid \mathit{Eff}(e) \neq 0\}$. Then, $\vartheta(E) = \vartheta(\mathcal{E})$.*

Proof. Immediate. □

In general, the effectiveness vectors $\vartheta(E)$ and $\vartheta(E')$ of two different subsets $E, E' \subseteq C \setminus \{0\}$ of normalized flux vectors can be different. The question arises on how to choose the set E such that $\vartheta(E)$ can be used to identify crucial reactions. One interesting possibility is to choose E as a generating set of the flux cone C. This strategy is analogous to using a generating set of C to elucidate the intrinsic properties that emerge from the whole metabolic network. For instance, if a coupling relationship between two reactions holds for the generators of the flux cone, the same property holds for the whole flux cone (see Chap. 6 for more details). Similarly, a reaction which is important for the generators of the flux cone can be expected to be crucial for the whole metabolic network. But again the question arises what generating set of the flux cone should be used to grasp the essentiality of reactions. This question is particularly interesting since there are infinitely many possible generating sets of the flux cone. In addition, as will be stated in Proposition 9.10, using two different generating sets of C can lead to different results.

In an attempt to choose a generating set E whose effectiveness vector $\vartheta(E)$ helps to evaluate the importance of reactions, one might wish that E covers all the most efficient operations of the metabolic network, i.e., for each non-zero flux vector $v \in C \setminus \{0\}$, there exists a generator $e \in E$ which is at least as efficient as v, i.e., $\mathit{Eff}(v) \leq \mathit{Eff}(e)$. There may exist two flux vectors $v^1, v^2 \in C$ with equal efficiencies, i.e., $\mathit{Eff}(v^1) = \mathit{Eff}(v^2)$, such that v^1 contains zero elements wherever v^2 does, and it includes at least one additional zero component, i.e., $\mathit{Supp}(v^1) \subsetneq \mathit{Supp}(v^2)$ (see Fig. 9.1 for an illustration). In such a case, v^1 could be more interesting to be included in the generating set E than v^2. Accordingly, one may requires that each generator $e \in E$ should be a simple flux vector, i.e., its set of active reactions is minimal.

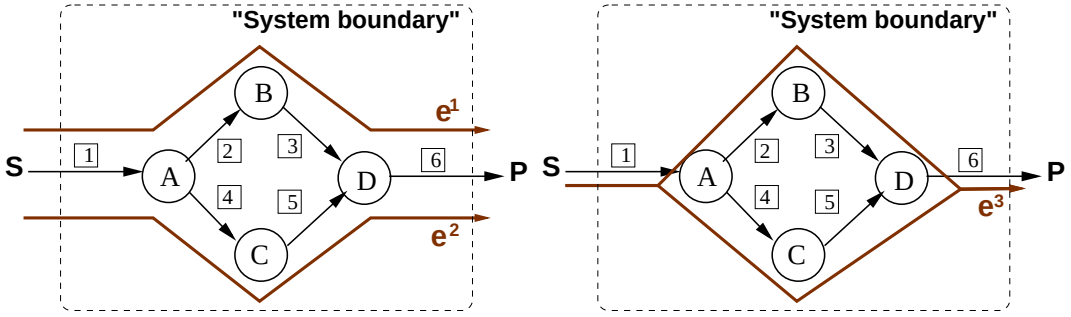


Figure 9.1: A hypothetic network contains two elementary modes e^1 and e^2. We assume that formation of product P, carried by reaction 6, is the main function of this network, i.e., $\mathit{Mfun} = \{6\}$. In addition, the operation of reaction 6 requires consumption of substrate S by reaction 1. Tab. 9.1 shows that the flux vector $e^3 = \frac{1}{2}(e^1 + e^2)$ is as efficient as both elementary modes, i.e., $\mathit{Eff}(e^3) = \mathit{Eff}(e^1) = \mathit{Eff}(e^2)$. However, e^3 is not a simple flux vector.

In the following, we show that finding a generating set E which fulfills both requirements above essentially amounts to assuming that E is a particular subset of ele-

Flux vector e^j	$\lvert e_i^j\rvert$ for $i = 1, \ldots, 6$						$\lVert e^j\rVert_1$	$\mathit{Eff}(e^j)$
	1	2	3	4	5	6		
e^1	1	1	1	0	0	1	4	$1/4$
e^2	1	0	0	1	1	1	4	$1/4$
e^3	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	4	$1/4$

Table 9.1: The three flux vectors e^1, e^2 and e^3 depicted in Fig. 9.1. and their respective efficiencies.

mentary modes. We first need to find out the connection between simplicity characterizing elementary modes and efficiency in metabolic networks. For this, we define the following relationship which takes account of these two properties.

Definition 9.3. *Given two non-zero flux vectors $v, u \in C \setminus \{0\}$, we write $v \trianglelefteq u$ if $\mathit{Supp}(u) \subseteq \mathit{Supp}(v)$ and $\mathit{Eff}(v) \leq \mathit{Eff}(u)$.*

For all flux vectors $v, u \in C \setminus \{0\}$, $v \trianglelefteq u$ means that u is at least as efficient as v even if the set of active reactions in v includes that of u. The next proposition shows that the efficiencies of v and u have the same sign.

Proposition 9.4. *Let $v, u \in C \setminus \{0\}$ be two non-zero flux vectors such that $v \trianglelefteq u$. Then, $\mathit{Eff}(v) \neq 0$ if and only if $\mathit{Eff}(u) \neq 0$.*

Proof. It follows from $v \trianglelefteq u$ that $\mathit{Supp}(u) \subseteq \mathit{Supp}(v)$ and $\mathit{Eff}(v) \leq \mathit{Eff}(u)$. Accordingly, $\mathit{Eff}(v) \neq 0$ implies $\mathit{Eff}(u) \neq 0$. Now, suppose $\mathit{Eff}(u) \neq 0$. Then, $\sum_{\tau \in \mathit{Mfun}} u_\tau \neq 0$. There exists $j \in \mathit{Mfun}$ such that $u_j \neq 0$ and so $j \in \mathit{Supp}(u)$. Since $\mathit{Supp}(u) \subseteq \mathit{Supp}(v)$, it follows that $j \in \mathit{Supp}(v)$. Since $\mathit{Mfun} \subseteq \mathit{Irr}$, we get $\sum_{\tau \in \mathit{Mfun}} v_\tau \neq 0$ and so $\mathit{Eff}(v) \neq 0$. □

By definition, each elementary mode is a simple steady-state flux vector, i.e., its support is minimal. This definition is equivalent to the following statement given in [113].

Proposition 9.5 ([113])**.** *A non-zero flux vector $e \in C$ is an elementary mode if and only if e lies on an extreme ray of some pointed cone obtained by intersecting the flux cone C with one of the 2^n orthants of $\mathbb{R}^n$.*

The next theorem states that elementary modes are the most efficient flux vectors in the steady-state flux cone.

Theorem 9.6. *Let $v \in C \setminus \{0\}$ be a non-zero flux vector. There exists an elementary mode e^* such that $v \trianglelefteq e^*$.*

Proof. Let $\mathcal{O}$ be an orthant of $\mathbb{R}^n$ such that $v \in \mathcal{O}$. Let $C' = C \cap \mathcal{O}$ and let E be the set of elementary modes of C that belong to C'. According to Proposition 9.5, C' is a

pointed cone and E is the set of its extreme rays. Since $v \in C'$, there exists $E' \subseteq E$ such that

$$v = \sum_{e \in E'} \lambda_e e, \text{ with } \lambda_e > 0 \text{ for all } e \in E'.$$

Let $\mu_e = \lambda_e \cdot \frac{\|e\|_1}{\|v\|_1}$ for all $e \in E'$. Then, $\mathit{Eff}(v) = \sum_{e \in E'} \mu_e \cdot \mathit{Eff}(e)$. Since $E' \subseteq \mathcal{O}$, it follows that $\mathit{Supp}(e) \subseteq \mathit{Supp}(v)$ for all $e \in E'$. In addition, we have $\|v\|_1 = \sum_{e \in E'} \lambda_e \cdot \|e\|_1$ and so $\sum_{e \in E'} \mu_e = 1$. Let $e^* \in E'$ such that $\mathit{Eff}(e) \leq \mathit{Eff}(e^*)$ for all $e \in E'$. We have $\mathit{Eff}(v) \leq (\sum_{e \in E} \mu_e) \cdot \mathit{Eff}(e^*) = \mathit{Eff}(e^*)$ and so the claim follows. □

From a biological point of view, the theorem above states that given a non-zero flux distribution $v \in C \setminus \{0\}$, the set of active reactions in v includes that of an elementary mode that has the same outcome through reactions $\tau \in \mathit{Mfun}$ and an investment which is smaller than that of v. Note that this result relies on the use of the L_1 norm in the definition of efficiency. The restriction of this norm on an orthant of $\mathbb{R}^n$ is linear.

Let E be the set of elementary modes and let $\mathcal{E} = \{e \in E \mid \mathit{Eff}(e) \neq 0\}$. According to Proposition 9.2, only elementary modes in $\mathcal{E}$ are relevant for the computation of CEFs. In the following, we show that $\mathcal{E}$ is the unique minimal set of simple flux vectors that covers all the most efficient operations of the metabolic network.

Definition 9.7. *An* efficient cover *of the flux cone C is a subset $U \subseteq C \setminus \{0\}$ such that for each non-zero flux vector $v \in C \setminus \{0\}$ with $\mathit{Eff}(v) \neq 0$, there exists a flux vector $u \in U$ such that $v \trianglelefteq u$. An efficient cover U is* minimal*, if there is no efficient cover $U' \subsetneq U$ strictly contained in U.*

Obviously, a trivial efficient cover is the flux cone C itself. In general, each subset $U' \subseteq C$ containing an efficient cover U of the flux cone C is itself an efficient cover. Here, we are interested in finding a minimal efficient cover of C whose effectiveness vector can be used to evaluate the importance of reactions. The following theorem not only states that the subset $\mathcal{E}$ of elementary modes fulfills this requirement, but also that $\mathcal{E}$ is the unique minimal efficient cover of C.

Theorem 9.8. *$\mathcal{E}$ is the unique minimal efficient cover of the flux cone C.*

Proof. Let $v \in C \setminus \{0\}$ be a flux vector such that $\mathit{Eff}(v) \neq 0$. According to Theorem 9.6, there exists an elementary mode e^* such that $v \trianglelefteq e^*$. With Proposition 9.4, it follows from $\mathit{Eff}(v) \neq 0$ that $\mathit{Eff}(e^*) \neq 0$ and so $e^* \in \mathcal{E}$. Therefore, $\mathcal{E}$ is an efficient cover of C. Now, let $\mathcal{F}$ be an efficient cover of C. Consider $e \in \mathcal{E}$. There exists a flux vector $e^* \in \mathcal{F}$ such that $e \trianglelefteq e^*$. Therefore, $\mathit{Supp}(e^*) \subseteq \mathit{Supp}(e)$. Since e is an elementary mode, we get $e = e^*$ or $e = -e^*$. Since $\mathit{Eff}(e) \neq 0$, it follows that $e \notin \mathrm{lin.space}(C)$ and so $e \neq -e^*$. We conclude that $e = e^*$, $e \in \mathcal{F}$ and so $\mathcal{E} \subseteq \mathcal{F}$. □

The set $\mathcal{E}$ can be determined by computing the set of all elementary modes and then selecting those whose efficiencies are not equal to zero. However, the computation of elementary modes is in general a hard computational task, which hampers the practical applicability of control-effective flux analysis. In the view of this limit, one might wish

to use another generating set of the flux cone C for the analysis of reaction importance. An interesting alternative is the use of a minimal set of generators of the flux cone, or equivalently minimal metabolic behaviors. An issue that may be encountered in this approach is that a minimal generating set need not be unique since the flux cone can be non-pointed. This limitation is not problematic for the following reasons. First, according to Proposition 9.2, generators of the lineality space $\mathrm{lin.space}(C)$ do not influence the components of the effectiveness vector since their efficiencies are equal to zero. These generators can be neglected in control-effective flux analysis. Second, given two normalized flux vectors g^1 and g^2 representing a minimal proper face G, both g^1 and g^2 carry the same fluxes through all (pseudo-) irreversible reactions. The only difference between g^1 and g^2 can be in their L_1 norm. We could choose the most efficient flux vector $g \in G \setminus \mathrm{lin.space}(C)$ to represent the minimal proper face G. The following proposition shows that the most efficient flux vector in a minimal proper face is the one with the smallest L_1 norm.

Proposition 9.9. *Let G be a minimal proper face of the flux cone C and let $g^1, g^2 \in G \setminus \mathrm{lin.space}(C)$ be two normalized flux vectors. Then,*

$$\mathit{Eff}(g^1) \cdot \|g^1\| = \mathit{Eff}(g^2) \cdot \|g^2\|$$

Proof. Either we have $g^1_\iota = g^2_\iota = 0$ and so $g^1_\tau = g^2_\tau = 0$, or $g^1_\iota = g^2_\iota = 1$ and so $g^1_\tau = g^2_\tau$. In both cases, $g^1_\tau = g^2_\tau$ and so the claim follows. □

There could be other interesting criterions to choose a generating set E such that its effectiveness vector $\vartheta(E)$ helps to grasp the importance of reactions. In any case, one should be able to compare results obtained by using different generating sets. The following proposition allows for such a comparison.

Proposition 9.10. *Let E be a set of normalized generators of the flux cone C and let $E' \subseteq C$ be a set of normalized flux vectors such that $E \subseteq E'$. Then, for each $i \in \{1, \ldots, n\}$*

$$\vartheta(E')_i - \vartheta(E)_i = \sum_{\tau \in \mathit{Mfun}} \sum_{e \in E' \setminus E} \frac{\mathit{Eff}(e,\tau)}{\sum_{e' \in E'} \mathit{Eff}(e',\tau)} \cdot \left(\frac{|e_i|}{\sigma_\tau} - \vartheta(E)_i \right), \tag{9.3}$$

with $\sigma_\tau = \max\{e_\tau : e \in E\}$ for all $\tau \in \mathit{Mfun}$.

Proof. For each $i \in \{1, \ldots, n\}$

$$\vartheta(E)_i = \sum_{\tau \in \mathit{Mfun}} \frac{1}{\sigma_\tau} \sum_{e \in E} \frac{\mathit{Eff}(e,\tau)}{\sum_{e' \in E} \mathit{Eff}(e',\tau)} \cdot |e_i|,$$

$$\vartheta(E')_i = \sum_{\tau \in \mathit{Mfun}} \frac{1}{\sigma'_\tau} \sum_{e \in E'} \frac{\mathit{Eff}(e,\tau)}{\sum_{e' \in E'} \mathit{Eff}(e',\tau)} \cdot |e_i|,$$

with $\sigma_\tau = \max\{e_\tau : e \in E\}$ and $\sigma'_\tau = \max\{e'_\tau : e' \in E'\}$ for all $\tau \in \mathit{Mfun}$. In what follows, we show that $\sigma'_\tau = \sigma_\tau$. Let $R = \{e \in E \mid e_\iota \neq 0\}$. We have $e_\iota = e_\tau = 0$ for

all $e \in E \setminus R$. Moreover, since $\iota \in Irr$, we have $E \cap \text{lin.space}(C) \subseteq E \setminus R$. Consider $e' \in E'$ and $\tau \in Mfun$. Since E is a generating set of the flux cone C, we have

$$e'_\iota = \sum_{e \in R} \lambda(e) e_\iota \text{ and } e'_\tau = \sum_{e \in R} \lambda(e) e_\tau \text{ for some } \lambda(e) \geq 0.$$

Since $e_\iota = 1$ and $e_\tau \leq \sigma_\tau$ for all $e \in R$, we get $e'_\iota = \sum_{e \in R} \lambda(e)$, $e'_\tau \leq \sigma_\tau \cdot \sum_{e \in R} \lambda(e)$ and so $e'_\tau \leq \sigma_\tau e'_\iota$. Since $e'_\iota \in \{0, 1\}$, $e'_\tau \leq \sigma_\tau$. Therefore, $\sigma'_\tau = \sigma_\tau$ and so the claim follows. □

Given two generating sets $E, E' \subseteq C$ such that $E \subseteq E'$, the differences in the effectiveness vectors $\vartheta(E)$ and $\vartheta(E')$ is given in equation (9.3). In some cases, we have $\vartheta(E') = \vartheta(E)$. This is particularly the case when $Eff(e) = 0$ for all $e \in E' \setminus E$. However, $\vartheta(E')$ and $\vartheta(E)$ can be different. For instance, if some reaction i is participating in no flux vector $e \in E' \setminus E$, i.e., $e_i = 0$ for all $e \in E' \setminus E$, then $\vartheta(E')_i < \vartheta(E)_i$. In general, given two reactions i and j, $\vartheta(E')_i - \vartheta(E)_i$ and $\vartheta(E')_j - \vartheta(E)_j$ can be different in sign and magnitude. If this is the case, using the effectiveness vectors $\vartheta(E)$ and $\vartheta(E')$ to evaluate the importance of reactions i and j would lead to conflicting conclusions. This shows that the results drawn from control-effective analysis are sensitive to the choice of the generating set of the flux cone.

9.3 Computational Results

In this section, control-effective fluxes obtained by means of elementary modes (resp. minimal metabolic behaviors) are called *EM-based* (resp. *MMB-based*) CEFs. We computed EM-based and MMB-based CEFs for the red blood cell metabolic network introduced in Sect. 4.5. For these computations, we consider reactions GSSGR, GSHpox, MemPhos, NaKATPase, MetHbRed and 23DPGdrain as the basis for the main functions of the metabolic network, i.e., $Mfun = \{$GSSGR, GSHpox, MemPhos, NaKATPase, MetHbRed, 23DPGdrain$\}$ [16]. Reaction ι corresponds to glucose uptake (GLK).

The flux cone of the red blood cell metabolic network contains 32 MMBs and 48 elementary modes (see Sect. 4.5). The 12 elementary modes that lie in the interior of the flux cone are obtained by combining two MMBs that involve reaction PGI in opposite directions. All these elementary modes have a zero flux through reaction PGI. As a result, the EM-based CEF of reaction PGI is lower than its MMB-based CEF. For the other reactions, the results of both approaches are qualitatively similar. Fig. 9.2, showing MMB-based and EM-based CEFs of all reactions, reveals that GSSGR has the highest control-effective flux, which can be explained by the importance of NADPH formation in the red blood metabolism [5].

As further demonstration we have applied control-effective analysis to determine critical reactions in the central carbon metabolism of *S. cerevesiae* [17]. Although this metabolism is relatively small (with 44 internal metabolites and 53 reactions), it contains 8726 (resp. 1309) elementary modes for growth on glucose (resp. ethanol).

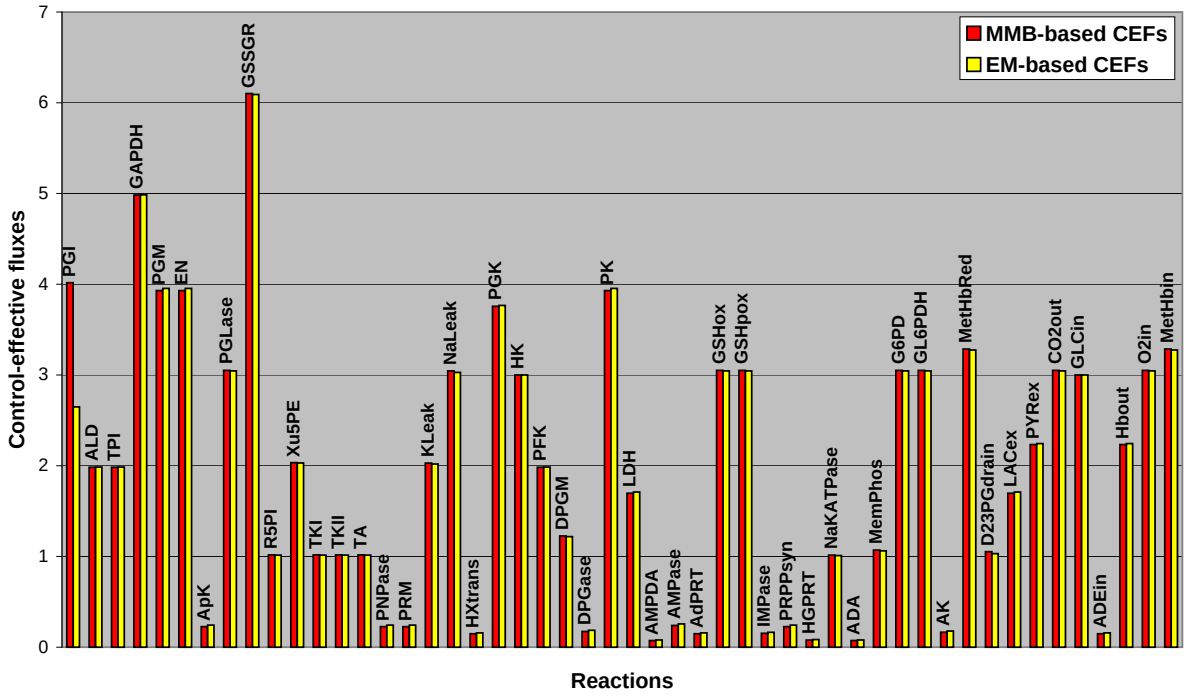


Figure 9.2: Comparison of MMB-based and EM-based control-effective fluxes for the red blood cell.

The corresponding flux cone is pointed and has 657 (resp. 224) minimal metabolic behaviors for growth on glucose (resp. ethanol). Fig. 9.3 and Fig. 9.4 show MMB-based and EM-based CEFs in different growth media (glucose and ethanol). Again, predictions of both approaches are qualitatively similar for the most of reactions. As expected, by using elementary modes, control-effective fluxes of reactions *MDH*1, *TPI*, *RPE*, *TKL*2, and *MDH*2 are underestimated for growth on glucose. All these reactions are reversible and are involved in forward and backward directions. Many elementary modes in the interior of the flux cone are combinations of other ones and decrease the EM-based CEF of these reactions. Interestingly reaction *TPI* is involved only in backward direction for growth on ethanol. In such a growth media, the EM-based CEF of reaction *TPI* is identical to its MMB-based CEF, while those of reactions *MDH*1, *RPE*, *TKL*2, and *MDH*2 are still different.

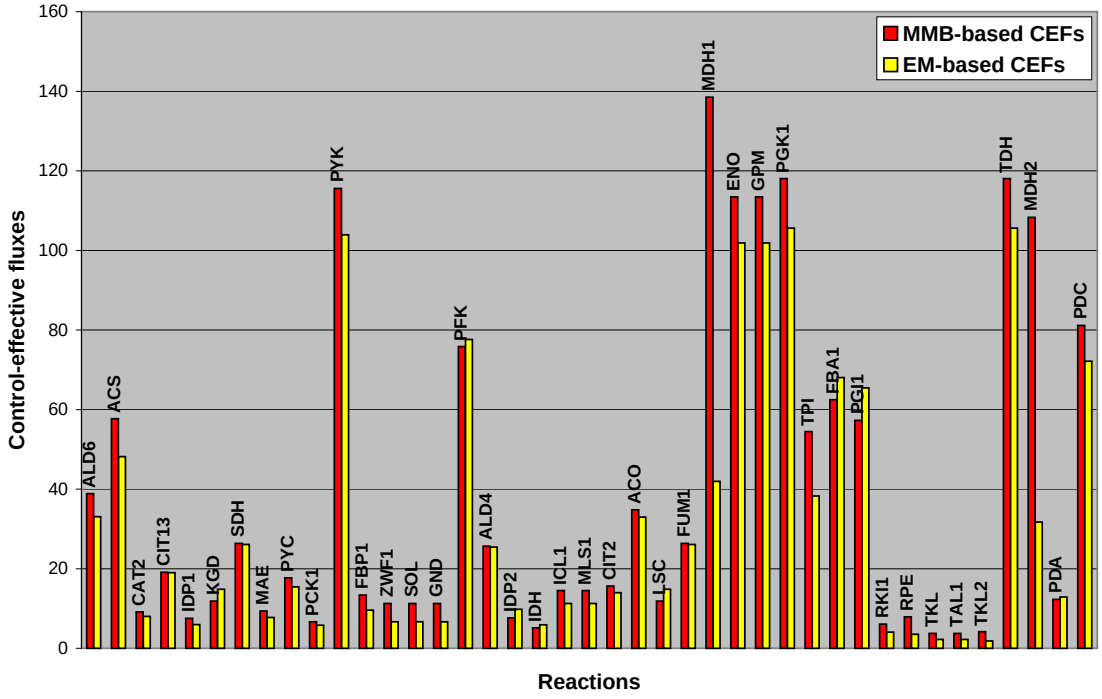


Figure 9.3: Comparison of MMB-based and EM-based control-effective fluxes for the yeast cells in glucose media.

Since the control-effective fluxes should correlate with transcript levels of metabolic genes [106], theoretical transcript ratios for growth on ethanol versus glucose were computed as ratios of EM-based CEFs and compared with experimental data from [25]. A good correlation ($R^2 = 0.60$) between theoretical and experimental transcript ratios has been found for 38 genes [17]. Similarly, we investigated the correlation between the same experimental data with ratios of MMB-based CEFs. Fig. 9.5 shows the same correlation strength ($R^2 = 0.60$), but for 41 genes.

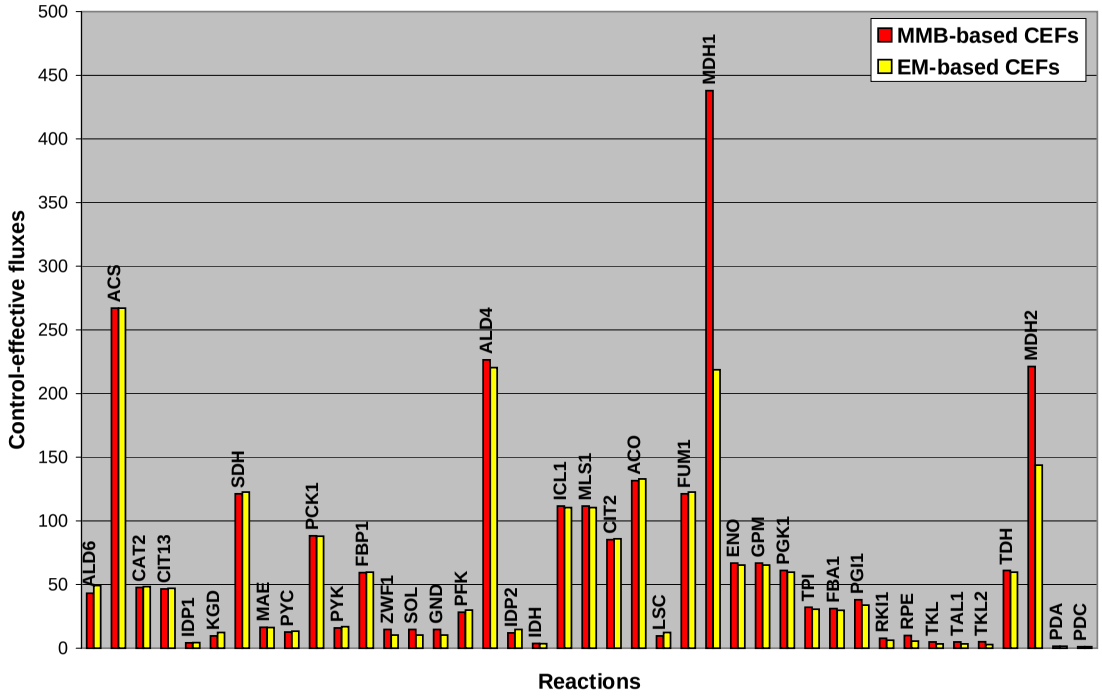


Figure 9.4: Comparison of MMB-based and EM-based control-effective fluxes for the yeast cells in ethanol media.

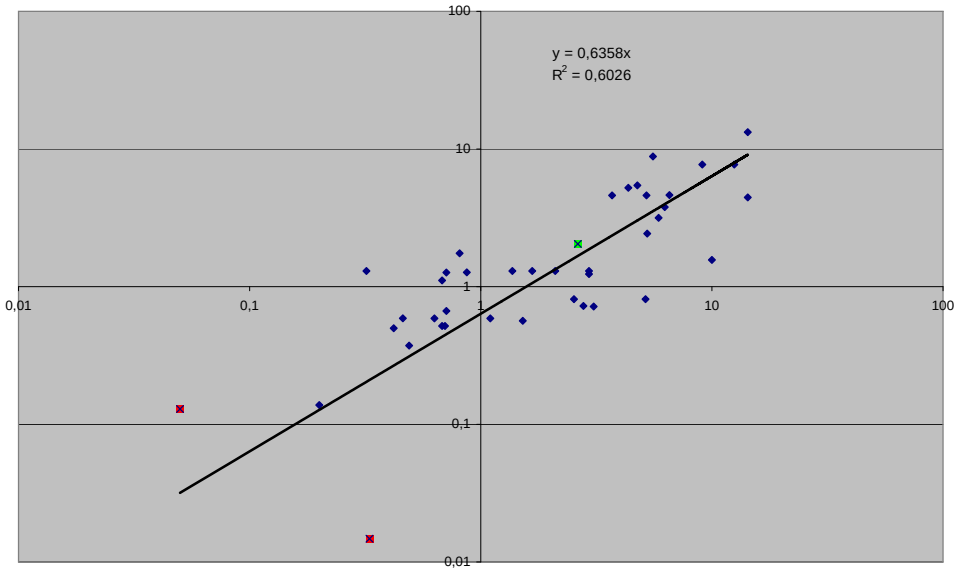


Figure 9.5: Comparison of theoretical and experimental data-based logarithmic ratios of gene expression levels for shift from glucose to ethanol. The experimental data are from [25]. Both X-axis and Y-axis are on logarithmic scale (X-axis: experimental data, Y-axis: MMB-based CEF ratios for shift from glucose to ethanol).

CHAPTER

10 Conclusions and Discussion

To conclude this dissertation on constraint-based analysis of metabolic networks, we summarize our main results, and we mention some interesting extensions of the presented work.

In order to achieve a system-level understanding of living systems, it is necessary to develop mathematical and computational methods guided by the concern of simplicity. For this reason, constraint-based modeling shows a wide applicability in the study of metabolic networks. In the scope of the presented work, a new constraint-based approach has been developed to describe the steady-state flux cone. The main difference with other recent constraint-based methods lies in the fact that our approach uses an outer description of the flux cone, based on sets of irreversible reactions. This is different from elementary mode and extreme pathway analysis, which both use an inner description, based on sets of generating vectors.

We have given several reasons to justify our new approach. First, it guarantees a certain compactness of the resulting description of the flux cone. Indeed, our description is not only unique, but also is minimal and satisfies a simplicity condition similar to the one that holds for inner description. The study of the relationship between inner and outer descriptions of the flux cone has allowed for explaining why, for large-scale metabolic networks, the size of the outer description is often significantly smaller than that of the inner descriptions. Nevertheless, since we still use a sparse matrix representation to store the resulting flux cone description, a more efficient representation of both types of descriptions is yet a challenging and interesting topic. A promising idea could be to transform the flux cone description into a logic function and then to represent this function in the form of a *binary decision diagram (BDD)*. This may reduce the space needed to store the large description of the flux cone. Another advantage of such a strategy is that BDD expressions can efficiently perform Boolean queries, allowing for identifying intrinsic properties of metabolic networks such as reaction dependencies.

Second, our approach suggests a refined classification of reactions according to their reversibility type (irreversible, pseudo-irreversible, and fully reversible). While the irreversible and pseudo-irreversible reactions completely characterize minimal metabolic behaviors, the fully reversible reactions define the reversible metabolic space, which may contain useful biological information. This information is no longer explicit if we replace each reversible reaction with two irreversible ones. Moreover, we have shown that the reversibility type provides a key to elucidate reaction dependen-

cies. Coupling relationships can only hold between reactions of a certain reversibility type. In particular, (pseudo-) irreversible reactions cannot be coupled with fully reversible reactions, and all reactions in an enzyme subset must have the same reversibility type. Using these concepts, we have improved an existing algorithm for identifying blocked and coupled reactions, and even devised a new algorithm for flux coupling analysis.

In many cases, a zero flux through one reaction implies a zero flux through several reactions. In other cases, many reactions must be constrained to have a zero flux for blocking a target reaction. A possible extension of flux coupling analysis is to study dependencies between subsets of reactions. More formally, given two subsets $M, N \subseteq \{1, \ldots, n\}$ of reactions, we say N is *dependent* of M, written $M \stackrel{=0}{\rightarrow} N$, if the removal of all reactions in M implies that all reactions in the set N cannot operate under steady-state conditions. More formally, $M \stackrel{=0}{\rightarrow} N$ if the following three conditions hold:

1. For all $v \in C$, $v_i = 0$ for all $i \in M$ implies $v_i = 0$ for all $i \in N$,
2. For all $M' \subsetneq M$, there exists $v \in C$ such that $v_i = 0$ for all $i \in M'$ and there exists $j \in N$ such that $v_j \neq 0$.
3. For all $N' \supsetneq N$, there exists $v \in C$ such that $v_i = 0$ for all $i \in M$ and there exists $j \in N'$ such that $v_j \neq 0$.

The second (resp. third) condition guarantees the minimality (resp. maximality) property of M (resp. N). If $N = \{i\}$ is a singleton set, M is identical to a minimal cut set for reaction i. In general, the identification of dependent reaction sets is more difficult than that of minimal cut sets, which is already known to be a hard computational task. The computation of minimal direction cuts is not trivial as well. Improving all these computations is another attractive perspective.

Our refined classification of reactions helps also for obtaining, in a constraint-based approach, a description of the altered steady-state flux cone. The latter contains the full range of achievable behaviors of the metabolic system after the removal of some reactions in the network. Certainly, the outer description of the flux cone may change considerably upon deletion of reactions. Nevertheless, this occurs only in case of the removal of pseudo-irreversible reactions which can operate in either direction under steady-state conditions. For the other reactions, the deduction of an outer description of the altered flux cone is straightforward.

Finally, we discussed control-effective flux (CEF) analysis, which has proved promising in assessing the importance of reactions. We formally justified why elementary modes are useful for CEF analysis. We also considered the use of a minimal generating set of the flux cone in such an analysis. It has been shown that the predictions about reaction importance using both strategies are similar for two metabolic systems, namely the *red blood cell* and *S. cerevesiae*. Overall, we anticipate that using a minimal generating set would enable a control-effective flux analysis of genome-scale metabolic networks. This, of course, still has to be proven.

Several challenges still remain in constraint-based analysis of metabolic networks. From a computational point of view, computing a description of the flux cone amounts to calculating a convex basis. This computation, which may be impractical for large-scale metabolic networks, is still a challenging task for the future. Further advancements in metabolic network modeling (e.g., dividing the network into modules, taking into account reaction dependencies) and in algorithm implementation (e.g., improving adjacency tests, parallelizing the double description method) may improve the existing tools.

In addition to the computational challenges, the biological interpretation of minimal metabolic behaviors and the reversible metabolic space deserves further attention. Our work suggests a modular approach to the study of metabolic networks. Each MMB could be seen as a module or a family of metabolic pathways sharing specific properties. The overall metabolic network can be understood as a combination of these different modules. We expect that more investigations of the biological implications of MMBs and RMS will further justify our approach.

Constraint-based metabolic network analysis is based on two main simplifying assumptions. First, metabolic networks are assumed to be at steady state. This can be considered as a limitation to this approach since no predictions about the dynamic behavior of the system can be made. However, the insight gained about the structure of metabolism may serve as a foundation for future studies whenever the understanding of the structure is of great interest. The second assumption concerns the reversibility of reactions. As we already mentioned in the beginning of this thesis, all metabolic reactions are thermodynamically reversible and can proceed in either direction depending on their Gibbs free energy differences. Since the computation of the latter is prohibitively expensive for large networks, reaction reversibility is still largely a matter of convention. In future studies, it is worth considering the reversibility of reactions as a parameter in the definition of the flux cone. A promising idea is to study the sensitivity of the flux cone definition to the reversibility of reactions. Another line of research could be to find out how reactions influence each other's reversibility using polyhedral theory.

The incorporation of regulatory constraints and kinetic information would expand the scope of our approach. Clearly the topic is not closed and constraint-based approaches will continue to provide useful modeling and computational tools for systems biology.

Curriculum Vitae

For reasons of data protection,
the curriculum vitae is not included in the online version

BIBLIOGRAPHY

[1] R. Albert and A.L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74(1):47–97, 2002.

[2] E. Almaas, Z.N. Oltvai, and A.L. Barabási. The activity reaction core and plasticity of metabolic networks. *PLoS Comput. Biol.*, 1(7):e68, 2005.

[3] D. Avis, K. Fukuda, and S. Picozzi. On canonical representations of convex polyhedra. In A. Cohen, X.S. Gao, and N. Takayama, editors, *Mathematical Software, Proceedings of the First International Congress*, pages 350–360. World Scientific, 2002.

[4] J.E. Bailey. Complex biology with no parameters. *Nat. Biotechnol.*, 19(6):503–504, 2001.

[5] J.W. Baynes and M. Dominiczak. *Medical Biochemistry*. Mosby, 1999.

[6] D.A. Beard and H. Hong. Thermodynamic-based computational profiling of cellular regulatory control in hepatocyte metabolism. *Am. J. Physiol. Endocrinol. Metab.*, 288(3):E633–44, 2005.

[7] J.E. Beasley and F.J. Planes. Recovering metabolic pathways via optimization. *Bioinformatics*, 23(1):92–98, 2007.

[8] A. Bockmayr and V. Weispfenning. Solving numerical constraints. In A. Robinson and A. Voronkov, editors, *Handbook of Automated Reasoning*, volume 1, chapter 12, pages 751–842. Elsevier, 2001.

[9] D. Bossi and B. Giardina. Red cell physiology. *Mol. Aspects Med.*, 17(2):117–128, 1996.

[10] C. Bro, B. Regenberg, J. Förster, and J. Nielsen. In silico aided metabolic engineering of Saccharomyces cerevisiae for improved bioethanol production. *Metab. Eng.*, 8(2):102–111, 2006.

[11] J.G. Bundy, B. Papp, R. Harmston, R.A. Browne, E.M. Clayson, N. Burton, R.J. Reece, S.G. Oliver, and K.M. Brindle. Evaluation of predicted network modules in yeast metabolism using NMR-based metabolite profiling. *Genome Res.*, 17(4):510–519, 2007.

[12] A.P. Burgard and C.D. Maranas. Probing the performance limits of the Escherichia coli metabolic network subject to gene additions or deletions. *Biotechnol. Bioeng.*, 74(5):364–75, 2001.

[13] A.P. Burgard and C.D. Maranas. Optimization-based framework for inferring and testing hypothesized metabolic objective functions. *Biotechnol. Bioeng.*, 82(6):670–677, 2003.

[14] A.P. Burgard, E.V. Nikolaev, C.H. Schilling, and C.D. Maranas. Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res.*, 14(2):301–312, 2004.

[15] A.P. Burgard, S. Vaidyaraman, and C.D. Maranas. Minimal reaction sets for Escherichia coli metabolism under different growth requirements and uptake environments. *Biotechnol. Prog.*, 17(5):791–797, 2001.

[16] T. Cakir, C.S. Tacer, and U.K. Ozergin. Metabolic pathway analysis of enzyme-deficient human red blood cells. *BioSystems*, 78(1-3):49–67, 2004.

[17] T. Cakir, D.S. Tekir, Z.I. Önsan, U.O. Kutlu, and J. Nielsen. Effect of carbon source perturbations on transcriptional regulation of metabolic fluxes in Saccharomyces cerevisiae. *BMC Syst. Biol.*, 18(1), 2007.

[18] R. Caspi, H. Foerster, C.A. Fulcher, R. Hopkinson, J. Ingraham, P. Kaipa, M. Krummenacker, S. Paley, J. Pick, S.Y. Rhee, C. Tissier, P. Zhang, and P.D. Karp. MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.*, 34(Database issue), 2006.

[19] V. Chvátal. *Linear Programming (Series of Books in the Mathematical Sciences).* W. H. Freeman, 1983.

[20] B.L. Clarke. Stability of complex reaction networks. In I. Prigogine and S.A. Rice, editors, *Advances in Chemical Physics*, volume 43, pages 1–216. John Wiley & Sons, 1980.

[21] B.L. Clarke. Complete set of steady states for the general stoichiometric dynamical system. *J. Chem. Phys.*, 75(10):4970–4979, 1981.

[22] A. Cornish-Bowden and M.L. Cardenas. Irreversible reactions in metabolic simulations: how reversible is irreversible? In *Animating the Cellular Map, Stellenbosch, South Africa*, pages 65–71. Stellenbosch University Press, 2000.

[23] M.W. Covert, I. Famili, and B.O. Palsson. Identifying constraints that govern cell behavior: a key to converting conceptual to computational models in biology? *Biotechnol. Bioeng.*, 84(7):763–772, 2003.

[24] T. Dandekar, F. Moldenhauer, S. Bulik, H. Bertram, and S. Schuster. A method for classifying metabolites in topological pathway analyses based on minimization of pathway number. *BioSystems*, 70(3):255–270, 2003.

[25] J.L. DeRisi, V.R. Iyer, and P.O. Brown. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, 278(5338):680–686, 1997.

[26] J.S. Edward and B.O. Palsson. How will bioinformatics influence metabolic engineering? *Biotechnol. Bioeng.*, 58(2-3):162–169, 1998.

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

[28] J.S. Edwards and B.O. Palsson. The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences of the United States of America*, 97(10):5528–5533, 2000.

[29] D.A. Fell. Metabolic control analysis: a survey of its theoretical and experimental development. *Biochem. J.*, 286 (Pt 2):313–330, 1992.

[30] D.A. Fell. The analysis of flux in substrate cycles. In S. Schuster, M. Rigoulet, R. Ouhabi, and J.P. Mazat, editors, *Modern Trends in Biothermokinetics*, volume 2, pages 97–101. Plenum Press, 1993.

[31] D.A. Fell. Systems properties of metabolic networks. In *Complex systems on Unifying themes in complex systems, Nashua, New Hampshire, USA*, pages 165–177. Perseus Books, 2000.

[32] J. Förster, I. Famili, P. Fu, B.O. Palsson, and J. Nielsen. Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network. *Genome Res.*, 13(2):244–253, 2003.

[33] K. Fukuda and A. Prodon. Double description method revisited. In *Combinatorics and Computer Science*, pages 91–111. Springer, LNCS 1120, 1995.

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

[35] D.T. Gillespie. Stochastic simulation of chemical kinetics. *Annu Rev Phys Chem*, 58:35–55, 2007.

[36] L.H. Hartwell, J.J. Hopfield, S. Leibler, and A.W. Murray. From molecular to modular cell biology. *Nature*, 402(6761 Suppl), 1999.

[37] M. Heiner, I. Koch, and K. Voss. Analysis and simulation of steady states in metabolic pathways with Petri nets. In *Workshop and Tutorial on Practical Use of Coloured Petri Nets and the CPN Tools, CPN'01, Aarhus University, Denmark*, pages 15–34, 2001.

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

[39] H.G. Holzhütter. The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *Eur. J. Biochem.*, 271(14):2905–2922, 2004.

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

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

[42] K.J. Kauffman, P. Prakash, and J.S. Edwards. Advances in flux balance analysis. *Curr. Opin. Biotechnol.*, 14(5):491–496, 2004.

[43] I.M. Keseler, J. Collado-Vides, S. Gama-Castro, J. Ingraham, S.M. Paley, I.T. Paulsen, M. Peralta-Gil, and P.D. Karp. EcoCyc: a comprehensive database resource for Escherichia coli. *Nucleic Acids Res.*, 33(Database-Issue):334–337, 2005.

[44] L. Khachiyan, E. Boros, B. Endre, K. Borys, K. Elbassioni, and V. Gurvich. Generating all vertices of a polyhedron is hard. In *ACM-SIAM symposium on Discrete algorithm VII, SODA '06, New York, NY, USA*, pages 758–765. ACM Press, 2006.

[45] H. Kitano. Systems biology: a brief overview. *Science*, 295(5560):1662–1664, 2002.

[46] S. Klamt. Generalized concept of minimal cut sets in biochemical networks. *Biosystems*, 83(2-3):233–247, 2006.

[47] S. Klamt and E.D. Gilles. Minimal cut sets in biochemical reaction networks. *Bioinformatics*, 20(2):226–234, 2003.

[48] S. Klamt, J. Saez-Rodriguez, and E.D. Gilles. Structural and functional analysis of cellular networks with cellnetanalyzer. *BMC Syst. Biol.*, 1:2, 2007.

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

[50] S. Klamt and J. Stelling. Combinatorial complexity of pathway analysis in metabolic networks. *Mol. Biol. Rep.*, 29(1-2):233–236, 2002.

[51] S. Klamt and J. Stelling. Two approaches for metabolic pathway analysis? *Trends Biotechnol.*, 21:64–69, 2003.

[52] G.W. Klau. On the relation between permutation-equivalent matrices and graph isomorphism. personal communication, 2007.

[53] D. Kondepudi. *Introduction to Modern Thermodynamics.* Wiley, 2008.

[54] J.O. Krömer, C. Wittmann, H. Schröder, and E. Heinzle. Metabolic pathway analysis for rational design of L-methionine production by Escherichia coli and Corynebacterium glutamicum. *Metab. Eng.*, 8(4):353–369, 2006.

[55] M. Krummenacker, S. Paley, L. Mueller, T. Yan, and P.D. Karp. Querying and computing with BioCyc databases. *Bioinformatics*, 21(16):3454–3455, 2005.

[56] V.S. Kumar, S.D. Madhukar, and C.D. Maranas. Optimization based automated curation of metabolic reconstructions. *BMC Bioinformatics*, 8:212, 2007.

[57] A. Kümmel, S. Panke, and M. Heinemann. Systematic assignment of thermodynamic constraints in metabolic network models. *BMC Bioinformatics*, 7:512, 2006.

[58] A. Larhlimi and A. Bockmayr. A new constraint-based description of the steady-state flux cone of metabolic networks. To appear in *Discrete Applied Mathematics*, 2008. Earlier version: Matheon Preprint Nr. 299. `http://page.mi.fu-berlin.de/~bockmayr/MMB.pdf`.

[59] A. Larhlimi and A. Bockmayr. A new approach to flux coupling analysis of metabolic networks. In *Computational Life Sciences II, CompLife'06, Cambridge, UK*, pages 205–215. Springer, LNBI 4216, 2006.

[60] A. Larhlimi and A. Bockmayr. Constraint-based analysis of gene deletion in a metabolic network. In *Workshop on Constraint based Methods for Bioinformatics, WCB'07, Porto, Portugak*, pages 48–55, 2007.

[61] A. Larhlimi and A. Bockmayr. Minimal direction cuts in metabolic networks. In *Computational Life Sciences III, CompLife'07, Utrecht, The Netherlands*, volume 940 of *American Institute of Physics Conference Series*, pages 73–86, 2007.

[62] A. Larhlimi and A. Bockmayr. On inner and outer descriptions of the steady-state flux cone of a metabolic network. In *Computational Methods in Systems Biology, CMSB'08, Rostock, Germany*, Lecture Notes in Bioinformatics. Springer, 2008. Accepted.

[63] J.M. Lee, E.P. Gianchandani, and J.A. Papin. Flux balance analysis in the era of metabolomics. *Brief. Bioinformatics*, 7(2):140–150, 2006.

[64] K. Lee, F. Berthiaume, G.N. Stephanopoulos, and M.L. Yarmush. Metabolic flux analysis: a powerful tool for monitoring tissue function. *Tissue Eng.*, 5(4):347–368, 1999.

[65] S. Lee, C. Phalakornkule, and M.M. Domach amd I.E. Grossmann. Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. *Comput. Chem. Eng.*, 24:711–716, 2000.

[66] J. Leiser and J.J. Blum. On the analysis of substrate cycles in large metabolic systems. *Cell Biophys.*, 11:123–138, 1987.

[67] M.V. Martinov, A.G. Plotnikov, V.M. Vitvitsky, and F.I. Ataullakhanov. Deficiencies of glycolytic enzymes as a possible cause of hemolytic anemia. *Biochim. Biophys. Acta*, 1474:75–87, 2000.

[68] P.J. Mulquiney and P.W. Kuchel. Model of 2,3-biphosphglycerate metabolism in the human erythrocyte based on detailed enzyme kinetic equations: equations and parameter refinement. *Biochem. J.*, 342:581–596, 1999.

[69] R.A. Notebaart, B. Teusink, R.J. Siezen, and B. Papp. Co-regulation of metabolic genes is better explained by flux coupling than by network distance. *PLoS Comput. Biol.*, 4(1), 2008.

[70] F. Nozicka, J. Guddat, H. Hollatz, and B. Bank. Theory of linear parametric optimization. *Akademie-Verlag*, 1974.

[71] I. Oancea. *Topological analysis of metabolic and regulatory networks by decomposition methods.* PhD thesis, Humboldt-Universität zu Berlin, 2003.

[72] B.O. Palsson. The challenges of in silico biology. *Nat. Biotechnol.*, 18(11):1147–1150, 2000.

[73] B.O. Palsson, N.D. Price, and J.A. Papin. Development of network-based pathway definitions: the need to analyze real metabolic networks. *Trends Biotechnol.*, 21(5):195–198, 2003.

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

[75] J.A. Papin, N.D. Price, S.J. Wiback, D.A. Fell, and B.O. Palsson. Metabolic pathways in the post-genome era. *Trends Biochem. Sci.*, 28(5):250–258, 2003.

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

[77] P.R. Patnaik. Microbial metabolism as an evolutionary response: the cybernetic approach to modeling. *Crit. Rev. Biotechnol.*, 21(3):155–175, 2001.

[78] T. Pfeiffer, I. Sánchez-Valdenebro, J.C. Nuno, F. Montero, and S. Schuster. Metatool: for studying metabolic networks. *Bioinformatics*, 15(3):251–257, 1999.

[79] S. Picozzi. *On canonical representation of convex polyhedra.* PhD thesis, Ecole Polytechnique Fédérale de Lausanne, 2005.

[80] M.G. Poolman, D.A. Fell, and C.A. Raines. Elementary modes analysis of photosynthate metabolism in the chloroplast stroma. *Eur. J. Biochem.*, 270(3):430–439, 2003.

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

[82] C. Pál, B. Papp, M.J. Lercher, P. Csermely, S.G. Oliver, and L.D. Hurst. Chance and necessity in the evolution of minimal metabolic networks. *Nature*, 440(7084):667–670, 2006.

[83] A. Raghunathan, N.D. Price, M.Y. Galperin, K.S. Makarova, S. Purvine, A.F. Picone, T. Cherny, T. Xie, T.J. Reilly, R. Munson, R.E. Tyler, B.J. Akerley, A.L. Smith, B.O. Palsson, and E. Kolker. In silico metabolic model and protein expression of Haemophilus influenzae strain rd kw20 in rich medium. *OMICS*, 8(1):25–41, 2004.

[84] R. Ramakrishna, J.S. Edwards, A. McCulloch, and B.O. Palsson. Flux-balance analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints. *Am. J. Physiol. Regul. Integr. Comp. Physiol.*, 280(3):R695–704, 2001.

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

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

[87] R.T. Rockafellar. The elementary vectors of a subspace of $\mathbb{R}^n$. In R.C. Bose and T.A. Dowling, editors, *Combinatorial Mathematics and its Applications*, pages 104–127. The Univ. of N. Carolina Press, 1969.

[88] R.T. Rockafellar. *Convex Analysis.* Princeton University press, 1970.

[89] C.H. Schilling, M.W. Covert, I. Famili, G.M. Church, J.S. Edwards, and B.O. Palsson. Genome-scale metabolic model of Helicobacter pylori 26695. *J. Bacteriol.*, 184(16):4582–4593, 2002.

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

[91] C.H. Schilling, S. Schuster, B.O. Palsson, and R. Heinrich. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnol. Prog.*, 15(3):296–303, 2000.

[92] A. Schrijver. *Theory of Linear and Integer Programming*. Wiley, 1986.

[93] R. Schuetz, L. Kuepfer, and U. Sauer. Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. *Mol. Syst. Biol.*, 3:119, 2007.

[94] R. Schuster and H.G. Holzhütter. Use of mathematical models for predicting the metabolic effect of large-scale enzyme activity alterations: application to enzyme deficiencies of red blood cells. *Eur. J. Biochem.*, 229:403–418, 1995.

[95] R. Schuster and S. Schuster. Refined algorithm and computer program for calculating all non-negative fluxes admissible in steady states of biochemical reaction systems with or without some flux rates fixed. *Comput. Appl. Biosci.*, 9(1):79–85, 1993.

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

[97] S. Schuster, A.D. Fell, and T. Dandekar. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.*, 18:326–332, 2000.

[98] S. Schuster and C. Hilgetag. On elementary flux modes in biochemical reaction systems at steady state. *J. Biol. Syst.*, 2(2):165–182, 1994.

[99] S. Schuster, C. Hilgetag, J.H. Woods, and D.A. Fell. Reaction routes in biochemical reaction systems: algebraic properties, validated calculation procedure and example from nucleotide metabolism. *J. Math. Biol.*, 45(2):153–181, 2002.

[100] S. Schuster, T. Pfeiffer, F. Moldenhauer, I. Koch, and T. Dandekar. Structural analysis of metabolic networks: elementary flux modes, analogy to Petri nets, and application to Mycoplasma pneumoniae. In *German Conference on Bioinformatics, GCB 2000, Heidelberg, Germany*, pages 115–120. Logos Verlag, 2000.

[101] S. Schuster and R. Schuster. Detecting strictly detailed balanced subnetworks in open chemical reaction networks. *J. Math. Chem.*, 6:17–40, 1991.

[102] J. Schwender, F. Goffman, J.B. Ohlrogge, and Y. Shachar-Hill. Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. *Nature*, 432(7018):779–782, 2004.

[103] D. Segrè, D. Vitkup, and G.M. Church. Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. U.S.A.*, 99(23):15112–15117, 2002.

[104] D. Segrè, J. Zucker, J. Katz, X. Lin, P. D'haeseleer, W.P. Rindone, P. Kharchenko, D.H. Nguyen, M.A. Wright, and G.M. Church. From annotated genomes to metabolic flux models and kinetic parameter fitting. *OMICS*, 7(3):301–316, 2003.

[105] T. Shlomi, O. Berkman, and E. Ruppin. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc. Natl. Acad. Sci. U.S.A.*, 102(21):7695–7700, 2005.

[106] J. Stelling, S. Klamt, K. Bettenbrock, S. Schuster, and E.D. Gilles. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420(6912):190–193, November 2002.

[107] G. Stephanopoulos. *Metabolic Engineering : Principles and Methodologies*. Academic Press, San Diego, 1998.

[108] R. Steuer, A.N. Nesi, A.R. Fernie, T. Gross, B. Blasiu, and J. Selbig. From structure to dynamics of metabolic pathways: application to the plant mitochondrial TCA cycle. *Bioinformatics*, 23:1378–1385, 2007.

[109] D.S. Tekir, T. Cakir, and U.O. Kutlu. Analysis of enzymopathies in the human red blood cells by constraint-based stoichiometric modeling approaches. *Comput. Biol. Chem.*, 30(5):327–338, 2006.

[110] M. Terzer and J. Stelling. Accelerating the computation of elementary modes using pattern trees. In P. Bucher and B.M.E. Moret, editors, *Workshop on Algorithms in Bioinformatics VI, WABI'06, Zurich, Switzerland*, volume 4175 of *Lecture Notes in Computer Science*, pages 333–343. Springer, 2006.

[111] I. Thiele, T.D. Vo, N.D. Price, and B.O. Palsson. Expanded metabolic reconstruction of Helicobacter pylori (iIT341 GSM/GPR): an in silico genome-scale characterization of single- and double-deletion mutants. *J. Bacteriol.*, 187(16):5818–5830, 2005.

[112] H.R. Tiwary. On the hardness of Minkowski addition and related operations. In *Symposium on Computational Geometry, SCG '07, New York, NY, USA*, pages 306–309. ACM, 2007.

[113] R. Urbanczik. Enumerating constrained elementary flux vectors of metabolic networks. *IET Syst. Biol.*, 1(5):274–279, 2007.

[114] R. Urbanczik and C. Wagner. An improved algorithm for stoichiometric network analysis: theory and applications. *Bioinformatics*, 21(7):1203–1210, 2005.

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

[116] T.D. Vo, H.J. Greenberg, and B.O. Palsson. Reconstruction and functional characterization of the human mitrochondrial metabolic network based on proteomic and biochemical data. *J. Biol. Chem.*, 279(38):39532–39540, 2004.

[117] A. von Kamp and S. Schuster. Metatool 5.0: fast and flexible elementary modes analysis. *Bioinformatics*, 22(15):1930–1931, 2006.

[118] A. Wagner and D.A. Fell. The small world inside large metabolic networks. *Proc. Biol. Sci.*, 268(1478):1803–1810, 2001.

[119] C. Wagner. Nullspace approach to determine the elementary modes of chemical reaction systems. *J. Phys. Chem. B*, 108(7):2425–2431, 2004.

[120] C. Wagner and R. Urbanczik. The geometry of the flux cone of a metabolic network. *Biophys. J.*, 89(6):3837–3845, 2005.

[121] S.J. Wiback and B.O. Palsson. Extreme pathway analysis of human red blood cell metabolism. *Biophys. J.*, 83(2):808–818, 2002.

[122] N. Wiener. *Cybernetics: or Control and Communication in Animal and the Machine*. MIT Press, Cambridge, MA, USA, 2000.

[123] T. Wilhelm, J. Behre, and S. Schuster. Analysis of structural robustness of metabolic networks. *Syst. Biol.*, 1(1):114–120, 2004.

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

[125] G.M. Ziegler. *Lectures on Polytopes*. Springer-Verlag, 1994.