Metabolic flux distributions

Daan de Groot[1], Wolfram Liebermeister[2], Maxime Mahout[3], Stefan Müller[4], David Ruckerbauer, Felipe Scott[5], and David Tourigny[6]

[1] Biozentrum and Swiss Institute of Bioinformatics, University of Basel, Switzerland; [2] Université Paris-Saclay, INRAE, MaIAGE, 78350 Jouy-en-Josas, France; [3] IRD, CNRS, University of Montpellier, France and Centre INRAE Occitanie-Montpellier, France; [4] Faculty of Mathematics, University of Vienna, Austria; [5] Green Technology Research Group, Facultad de Ingeniería y Ciencias Aplicadas, Universidad de Los Andes, Chile; [6] School of Mathematics, University of Birmingham, Birmingham UK

Abstract

A living cell can take up nutrients from its environment and chemically convert these substrates into products that it needs for its survival. The chemical conversion of these products is done by catalyzing so-called metabolic reactions. The whole of metabolic reactions that a cell can catalyze forms its metabolic network, and determines the metabolic versatility of the cell. In this chapter, we will investigate such metabolic networks and we will find that all metabolic capabilities of a cell can be captured in a mathematical space: the flux cone. Moreover, we will show that this flux cone can be decomposed into minimal metabolic modes, called elementary flux modes. To make this more precise: elementary flux modes are the minimal combinations of metabolic reactions out of which all other possible combinations of metabolic reactions that can be steadily catalyzed by a cell can be built. We show some applications of the analysis of a metabolic network through its elementary modes. Despite the benefits of elementary flux mode analysis, it cannot always be done because of its computational complexity, for those cases we describe alternative methods to explore the flux cone.

Keywords: constraint-based metabolic models, elementary flux modes, elementary conversion modes, minimal cut sets, flux sampling

Contributions: This chapter was originally drafted and written by Daan de Groot, David Tourigny and Felipe Scott. Stefan Müller worked substantially on the sections about elementary flux modes, David Ruckerbauer and Maxime Mahout added sections on Minimal Cut Sets and on the enumeration of EFMs under extra constraints. Wolfram Liebermeister wrote parts of the introduction. Exercises were added by Felipe Scott and Andreas Kremling. Feedback was provided by Diana Széliová, Samira van den Bogaard, Maarten Droste, Elad Noor, and Herbert Sauro.

To cite this chapter: D. de Groot, W. Liebermeister, M. Mahout, S. Müller, D. Ruckerbauer, F. Scott, and D. Tourigny. Metabolic flux distributions (Version July 2025). doi: 10.5281/zenodo.8156504. Chapter from: The Economic Cell Collective (2025). Economic Principles in Cell Biology. No commercial publisher | Online open access book | doi: 10.5281/zenodo.8156386

The authors are listed in alphabetical order.

doi: 10.5281/zenodo.8156386

Chapter overview

- The metabolic capabilities of an organism can be related to the individual chemical reactions it can catalyze
- Elementary flux modes are minimal metabolic strategies that together span all metabolic capabilities.
- When the analysis of elementary flux modes is prohibited by computational limits, alternatives could be used, such as elementary conversion modes, flux sampling and minimal cut sets.

4.1. Modeling metabolic fluxes in cells

In the previous chapters we have seen that cells can convert substances from their environment into building blocks for cell components: their metabolism allows cells to grow, reproduce, repair themselves, and produce compounds needed to resist environmental stresses. But how does a cell manage this in detail, and does it have alternative metabolic strategies in case one does not function properly?

The overall metabolic conversion, for example from nutrients and oxygen to all necessary cell components and carbon dioxide, that a cell can use to grow and reproduce is in fact the consequence of many smaller chemical reactions working in concert. All chemical reactions that a cell can catalyze by expressing its enzymes form a very versatile 'metabolic network', which enables a cell to survive and grow, even when the availability of nutrients in its environment changes. There are various (semi-)automatized methods available that can be used to reconstruct this metabolic network from an organism's genome sequence (for a review of the various methods, see [1]). In this chapter we will zoom in on this metabolic network and study the fluxes (reaction rates) through all individual reactions.

We call the combination of all reaction rates in a cell a 'metabolic flux distribution', and this flux distribution determines if and how a cell succeeds in taking up and converting the right nutrients to sustain itself. For a growing cell, we may ask: what will its flux distribution be, and how does this distribution change when its environment changes? Modeling metabolic fluxes allows us to answer specific questions, for instance about the change of a cell's metabolism after a gene is deleted: will it survive, and if so, will it take up different nutrients or produce different products? In contrast to the previous chapters, in the current and following chapters we are not satisfied with verbal descriptions, but seek predictive models that allow us to compute the state of a cell.

So how can we model metabolism in detail? Our main task is to describe and predict the uptake, conversion, and production of metabolites, as described by the metabolic fluxes. The rate at which a chemical reaction runs depends (through kinetics and thermodynamics) on metabolite concentrations and enzyme activities. Since enzymes are synthesized by the cell itself, the reaction rates are not only controlled by external nutrient supply, but also by gene expression. These dependencies make this a complicated field of study: the metabolic fluxes depend on the enzyme levels and metabolite concentrations, while the metabolite concentrations are again determined by the balance of fluxes through reactions that produce and consume the metabolites. In turn, enzyme levels are determined by gene expression, which is dependent on both external conditions and internal needs (e.g. the enzyme expression may change when different macromolecules need to be made in different phases of the cell cycle). To make matters even less transparent, most of the parameters (e.g. enzyme kinetic constants and details of enzyme regulation) are unknown.

For the moment, we therefore make some simplifying assumptions in order to obtain tractable models:

1. **Focus on small molecules** We focus on a subsystem of the cell, the metabolism of small molecules, which generates macromolecular precursors and energy carriers. All other processes (such as macromolecule synthesis) that happen "outside" our metabolic network are ignored.

2. **Ignore spatial structure** We largely ignore the spatial structure of cells: metabolite concentrations and reaction rates are assumed to be homogeneous across the cell. The exception to this rule occurs when there are cell com-

partments, in which case we describe the metabolites in both compartments as if they were separate compounds (e.g. cytosolic ATP vs mitochondrial ATP), which can be converted in each other through transport "reactions".

3. **Focus on fluxes as the only variables** Instead of considering metabolite concentrations, enzyme levels and metabolic fluxes together, we will only focus on metabolic fluxes. This has important consequences for the mathematical models that we will construct: many variables, and the corresponding equations, will be ignored. Additionally, fluxes cannot be computed through enzyme kinetics, so that we need to find other, non-mechanistic ways to compute the fluxes!

4. **Focus on steady-state metabolism** In a simplified picture of balanced growth (see the chapter on Balanced Growth), all metabolic processes are balanced: the rate at which material flows into the cell matches the rate at which it is converted, which again matches the production rate of macromolecule precursors. In addition, we assume that these fluxes are constant, such that the whole metabolic network is in a 'steady-state'. Taken together, we thus assume that the metabolic network can take up and produce external metabolites (e.g. extracellular metabolites and macromolecular precursors), but that all internal metabolites (inside the metabolic network) are mass-balanced, that is, for each of these metabolites, production and consumption cancel out.

5. **Describe precursor demand by a "biomass reaction"** We assume that cell growth (or: biomass production) requires a fixed set of macromolecule precursors in fixed proportions, corresponding to the average mixture of cell components that are necessary to make a cell. For metabolism, this means that the production of more macromolecule precursors only leads to more biomass production when the production of all precursors is scaled up proportionally. We formally express this by a hypothetical "biomass reaction" that consumes a mix of precursors and energy carriers in the predefined proportions. Hence, in the metabolic models we will describe the term "biomass" has a special meaning: while it usually means "the totality of compounds in a cell", here we use it for "the totality of compounds *outside* our metabolic model, which metabolism needs to produce".

6. **Ignore dilution of small molecules** When a cell doubles its size but does not produce a certain metabolite, the concentration of this metabolite will halve. This basic principle is called 'dilution by growth', and in principle affects all compounds in the cell. During balanced growth, the production of macromolecules that are produced but not degraded should balance dilution, i.e. the number of each macromolecule should double when the cell doubles its size. This requires the rate of precursor supply to match the dilution rate, and hence the cell's growth rate. Similarly, small molecules are diluted, but since these are also degraded by consuming reactions, the rate of dilution is usually negligible compared to the production and consumption by metabolic reactions. Therefore, the models below will usually ignore the dilution of such metabolites.

7. **Constrain solutions by modeling limited resources** Since each enzyme has a maximal catalytic rate (the k_{cat} value), a reaction flux will require a certain (minimal) amount of enzyme, which takes up cellular space; since cellular space is limited, fluxes cannot increase infinitely since there is always an upper bound on a weighted sum of reaction fluxes. This constraint implies compromises between different reaction fluxes: one flux can only be increased at the expense of others.

With these assumptions, we are converging on a mathematical model: we know which variables to describe (the metabolic fluxes in steady-state metabolism), which constraints to apply (the balance of production and consumption of all internal metabolites) and what main input information we need (the metabolic network, described by a list of chemical reaction equations). Importantly, the model will be able to describe compromise: for example, with a given carbon influx and assuming mass balance, the carbon atoms can either be used to generate energy or biomass; if one function increases, the other one goes down. To obtain realistic predictions, we may introduce additional constraints,

for example known flux directions or experimentally measured uptake rates. All this information will not suffice to predict metabolic fluxes precisely, but it allows us to narrow down the possible flux distributions. Importantly, all formulae in these models are linear, which makes them tractable even for very large model sizes (with thousands or even hundreds of thousands of variables).

Notably, all these assumptions depend only on the list of chemical reaction equations (the stoichiometry of the metabolic network), and nothing needs to be known about enzyme kinetics. So if the networks are already known, what do we gain from this kind of modeling? Even if a metabolic network structure is known reaction by reaction, this does not mean that we understand the network-wide behavior, i.e. which overall flux distributions are possible, and what overall flux distributions are useful for the cell. Our aim here is to make the step from structural information (about the network) to physiological insights about how the network can be used. We can learn, for example, how much biomass can be made from a certain amount of glucose, and whether an enzyme deletion is lethal because a certain precursor cannot be produced anymore.

Metabolic network structures (in the form of stoichiometric matrices) are approximately known for many microbial species, and to some extent for higher organisms. Together with the constraints outlined above, this network determines a range (or space) of possible flux distributions. In this chapter we will characterize this space of possible flux distributions according to our assumptions, and since we characterize fluxes entirely by constraints the models will be called "constraint-based models". We will get to know mathematical tools to characterize this space in a simple way: for instance, to describe all possibilities that a metabolic network provides we can use Elementary Flux Modes (EFMs).

In the next chapter, we will combine such constraint-based models with optimality principles: out of the space of possible flux distributions, specific "optimal" flux distributions will be selected because these are supposedly "most profitable", either for the cell or for metabolic engineering purposes. Some of the flux prediction methods that we will describe refer also to concentrations; for instance, metabolite concentrations play a role in thermodynamic constraints that exclude certain flux directions, and enzyme concentrations come into play in models that associate fluxes with an enzyme demand. However, in all cases, the connection between fluxes and concentrations is very simple, and real enzyme kinetics are ignored. In later chapters, we will then see how the models change when more and more of the complex details are added about metabolite concentrations, enzyme kinetics, and thermodynamics.

4.2. The flux cone

4.2.1. Mass-balance constraints

As described in the introduction, our models will be built on the metabolic network of all chemical reactions that an organism can catalyze. We can conveniently summarize all these chemical reactions as an $(m \times n)$-dimensional stoichiometric matrix $\mathbf{N}$ where each of the m rows corresponds to a metabolite and each of the n columns corresponds to a reaction. The entry $\mathbf{N}_{ij}$ is the coefficient of the i-th metabolite in the j-th chemical reaction. Then, we can gather all n net reaction rates in an n-dimensional *flux vector*: $\mathbf{v} = (v_1, \ldots, v_n)^T$. This is convenient because the multiplication $\mathbf{N}\ \mathbf{v}$ now captures the net production and consumption of all m metabolites at this flux distribution, and is therefore equal to the time derivative of the metabolite concentrations: $\mathrm{d}\mathbf{s}/\mathrm{d}t = \mathbf{N}\ \mathbf{v}$. Therefore, the steady-state assumption, combined with the assumption that dilution of metabolites due to growth is negligible, can be mathematically captured in a set of linear equations that we call the *mass-balance constraints* for $\mathbf{v}$,

$$\mathbf{N}\ \mathbf{v} = \mathbf{0}. \tag{4.1}$$

Since in a typical metabolic reaction network the number of metabolites is smaller than the number of reactions ($m < n$), the equations for $\mathbf{v}$ are under-determined. This means that there are infinitely many solutions, $\mathbf{v}$, that

satisfy the mass-balance constraints. The space of all such $\mathbf{v}$ is called the *nullspace* of $\mathbf{N}$.

4.2.2. Irreversibility constraints

In principle, all reactions in a metabolic reaction network are able to run in both directions, but in many practical examples certain thermodynamic arguments can be used to justify treating a subset of reactions as *irreversible*, meaning that in a given model they can run in only one direction. The choice of which reactions to assume irreversible depends on the experimental conditions and affects the results of the downstream constraint-based analysis.

Due to microscopic reversibility, the net reaction rate v_i (of reaction i) is the difference of the forward and backward reaction rates, that is, $v_i = v_i^{\rightarrow} - v_i^{\leftarrow}$ (with both $v_i^{\rightarrow} > 0$ and $v_i^{\leftarrow} > 0$ if all reactants are present), and v_i can be either positive, zero, or negative. As stated above, thermodynamics may determine the direction of certain reactions, that is, the sign of the net reaction rate. In this sense, if a reaction proceeds in the forward reaction, one adds the nonnegativity constraint $v_i \geq 0$. (Conversely, if a reaction proceeds in the backward reaction, one redefines the reaction by exchanging forward and backward and again adds $v_i \geq 0$.) For a compact notation, let $\mathcal{R}^{\rightarrow} \subseteq \{1, \ldots, n\}$ be the index set of the irreversible reactions (and $\mathcal{R}^{\rightleftharpoons} \subseteq \{1, \ldots, n\}$ be the reversible reactions). We require $\mathbf{v}^{\rightarrow} := \mathbf{v}_{\mathcal{R}^{\rightarrow}} \geq \mathbf{0}$, that is, $v_i \geq 0$ if $i \in \mathcal{R}^{\rightarrow}$.

4.2.3. The flux cone

Mass balance and irreversibility constraints together define the flux cone

$$\mathcal{C} = \{\mathbf{v} \mid \mathbf{N}\mathbf{v} = \mathbf{0},\ \mathbf{v}^{\rightarrow} \geq \mathbf{0}\}\,. \tag{4.2}$$

Elements of the flux cone are called *flux modes*. The flux cone $\mathcal{C}$ is called an s-cone (subspace cone) in Müller and Regensburger (2016) [2], since it arises from a linear subspace and nonnegativity constraints.

To provide a concrete example, we consider the simple representation of central carbon metabolism presented in Figure 4.1. In this example there are four external metabolites, $G_{\text{ex}}, O, P_1, P_2$ and two internal metabolites: G and P. In our model we only require mass-balance for internal metabolites, such that the steady-state constraint can be written as

$$\mathbf{N}\mathbf{v} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 2 & -1 & -1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix} = \mathbf{0}, \tag{4.3}$$

where each column thus corresponds to one of the four (reversible or irreversible) reactions, and where the rows correspond to G and P respectively. The entry 1 in the first row of the first column thus corresponds to the import of one glucose molecule G. The mass-balance equations

$$v_1 - v_2 = 0, \quad 2\,v_2 - v_3 - v_4 = 0, \tag{4.4a}$$

and the non-negativity conditions

$$v_1, v_2, v_3 \geq 0, \tag{4.4b}$$

induced by the irreversible reactions 1, 2, 3, define the flux cone $\mathcal{C}$ as the space of all flux vectors $\mathbf{v}$ that satisfy all of these constraints simultaneously.

4.3. Elementary flux modes

Equation (4.2) gives a mathematical definition of the flux cone (via equations and inequalities). Here, we will provide an equivalent characterization of this space (via generators, see Math box 4.A). Note that definition (4.2) makes it

Σ Math box 4.A Generators of a polyhedral cone

For every polyhedral cone, and hence for every subspace cone such as the flux cone $\mathcal{C}$, there exists a finite, minimal set of *generators* (minimal in the sense that no proper subset forms a generating set). In particular, for the flux cone $\mathcal{C}$, there exists a finite set $\{\boldsymbol{f}^{(1)}, \ldots, \boldsymbol{f}^{(\ell)}\}$ of n-dimensional vectors such that

$$\mathcal{C} = \left\{ \mathbf{v} \mid \mathbf{v} = \sum_{k=1}^{\ell} \lambda_k \, \boldsymbol{f}^{(k)} \text{ with } \lambda_k \geq 0 \right\},$$

that is, any flux vector $\mathbf{v}$ in the flux cone $\mathcal{C}$ can be expressed as a conical (non-negative) linear combination of generators $\{\boldsymbol{f}^{(1)}, \ldots, \boldsymbol{f}^{(\ell)}\}$.
Remark. The generators $\boldsymbol{f}^{(k)}$ can be multiplied with scalars, that is, any $\lambda \, \boldsymbol{f}^{(k)}$ with $\lambda > 0$ could replace $\boldsymbol{f}^{(k)}$ in the set of generators.
Remark. For a general polyhedral cone, there is no *unique* minimal generating set. However, there is a unique minimal set of *conformal* generators; see [2, Section 3.4]. For subspace cones (such as the flux cone), these are the support-minimal vectors (EFMs); see Theorem 1 below.

easy to check if a given steady-state flux distribution $\mathbf{v}$ lies in $\mathcal{C}$. However, it is not clear how to generate the flux cone. As a set of generators, we will introduce "minimal" flux distributions, called *elementary flux modes* (EFMs), that can be combined to generate all possible flux distributions in $\mathcal{C}$. These EFMs generate the flux cone, similar to how basis vectors generate a linear subspace (but with non-negative coefficients).

In order to define EFMs formally, we introduce the *support* of a vector $\mathbf{v}$ as the index set $\operatorname{supp}(\mathbf{v}) = \{i \mid v_i \neq 0\}$, that is, the support of a flux vector is the set of reactions that have a nonzero rate.

Definition 1. *A nonzero vector $\mathbf{v} \in \mathcal{C}$ is an EFM if it is* support-minimal, *that is, if $\operatorname{supp}(\mathbf{v}') \subseteq \operatorname{supp}(\mathbf{v})$ for any nonzero vector $\mathbf{v}' \in \mathcal{C}$ implies $\operatorname{supp}(\mathbf{v}') = \operatorname{supp}(\mathbf{v})$.*

Remark. If $\mathbf{v}$ is an EFM and $\operatorname{supp}(\mathbf{v}') = \operatorname{supp}(\mathbf{v})$, then further $\mathbf{v}' = \lambda \, \mathbf{v}$ for some scalar λ.

Definition 1 states that $\mathbf{v}$ is an EFM if there is no nonzero flux vector in the flux cone that uses only a strict subset of the reactions that are active in $\mathbf{v}$. This also means that if any of the flux-carrying reactions in an EFM is deleted, the flux through the remaining reactions must violate the mass-balance constraints and can therefore not occur in steady-state metabolism; the EFMs are thus minimal in the sense that they cannot be reduced further.

To illustrate the concept of EFMs, we return to the simple representation of central carbon metabolism presented in Figure 4.1 with the stoichiometric matrix

$$\mathbf{N} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 2 & -1 & -1 \end{pmatrix} \tag{4.5}$$

and the flux vector $\mathbf{v} = (v_1, v_2, v_3, v_4)^\top$, where $v_1, v_2, v_3 \geq 0$. As it turns out, the set of EFMs is given by

$$\boldsymbol{f}^{(1)} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 2 \end{pmatrix}, \quad \boldsymbol{f}^{(2)} = \begin{pmatrix} 1 \\ 1 \\ 2 \\ 0 \end{pmatrix}, \quad \boldsymbol{f}^{(3)} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}, \tag{4.6}$$

and these are depicted in Figure 4.1 (B). From our understanding of central carbon metabolism, we see that these three EFMs represent the "minimal" metabolic pathways of ($\boldsymbol{f}^{(1)}$) glycolytic fermentation, ($\boldsymbol{f}^{(2)}$) oxidative metabolism of glucose, and ($\boldsymbol{f}^{(3)}$) oxidative metabolism of the fermentation product.

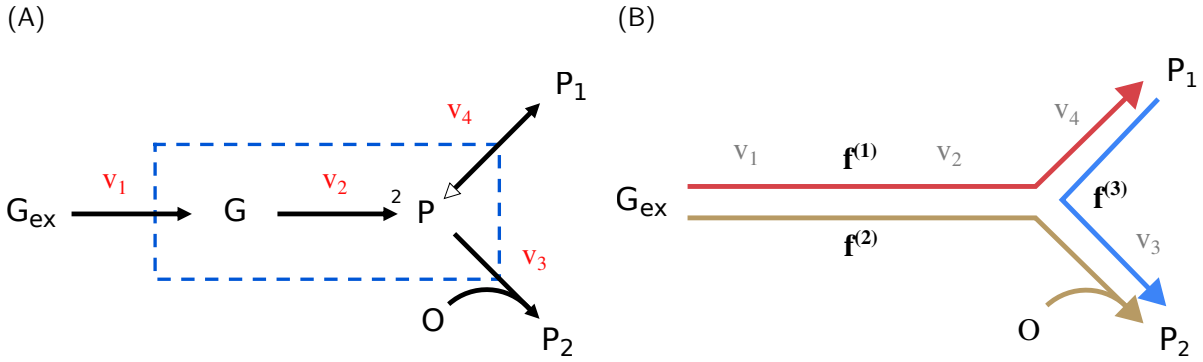


Figure 4.1: A simple representation of central carbon metabolism as a metabolic network. (A) Extracellular glucose, G_{ex}, is imported into the cell via reaction 1, and intracellular glucose, G, is converted to pyruvate, P, via reaction 2, having stochiometric coefficients of two pyruvate molecules for one glucose molecule. Pyruvate is then either converted to a fermentation product, P_1, via reaction 4 or, in the presence of oxygen, O, converted to an oxidative phosphorylation (OXPHOS) terminal product P_2 via reaction 3. The fermentation product P_1 can also be converted back to pyruvate via the backward reaction of 4. (B) EFMs $\boldsymbol{f}^{(1)}, \boldsymbol{f}^{(2)}, \boldsymbol{f}^{(3)}$. From our understanding of central carbon metabolism, $\boldsymbol{f}^{(1)}$ represents glycolytic fermentation, $\boldsymbol{f}^{(2)}$ the oxidative metabolism of glucose, and $\boldsymbol{f}^{(3)}$ the oxidative metabolism of the fermentation product.

In Math box 4.A, we have characterized a polyhedral cone (the flux cone $\mathcal{C}$) in terms of its generators (the EFMs). In our toy carbon metabolism network, this means that any flux vector $\mathbf{v}$ can be viewed as a conical combination of these three minimal metabolic pathways. This interpretation remains true for any metabolic reaction network: *elementary flux modes represent the minimal metabolic pathways through the metabolic reaction network at steady state*.

In Math box 4.A, we have also mentioned that EFMs need not form the *unique* minimal set of generators, but they form the unique minimal set of *conformal* generators. We first motivate conformality by thermodynamic arguments and then provide a formal definition. For every reaction, Gibbs free energy determines its direction, and hence for every flux, it determines its sign $(-, 0, +)$. Now, since any flux vector is the conical combination of EFMs, the signs of the flux vector determine the signs of the EFMs. In particular, if a certain flux component is zero, then this flux component is zero in all EFMs. (Zero flux cannot arise from a cancellation of positive and negative fluxes.) If a certain flux component is nonzero, then this flux component has the same sign or is zero in all EFMs. (Zero flux can arise from a zero enzyme concentration and hence is thermodynamically sound.)

The above argument can be formalized as follows: For a vector $\mathbf{v} \in \mathbb{R}^n$, we obtain the sign vector $\operatorname{sign}(\mathbf{v}) \in \{-, 0, +\}^n$ by applying the sign function componentwise, that is, $\operatorname{sign}(\mathbf{v})_i = \operatorname{sign}(v_i)$. In order to capture "conformal signs", we define the partial order $0 < -$ and $0 < +$ on $\{-, 0, +\}$, which implies the inequalities $0 \leq 0$ (zero flux conforms to zero flux), $+, 0 \leq +$ (positive or zero flux conforms to positive flux), and $-, 0 \leq -$ (negative or zero flux conforms to negative flux). The partial order on $\{-, 0, +\}$ induces a partial order on $\{-, 0, +\}^n$: for two sign vectors $\sigma, \tau \in \{-, 0, +\}^n$, we write $\sigma \leq \tau$ if the inequality holds componentwise, and we say that σ conforms to τ. If $\sigma \leq \tau$ (and τ_i is given), then $\sigma_i = \tau_i$ or $\sigma_i = 0$. To summarize, if σ conforms to τ, then it has the same entries or some more zeros.

Now, we can refine the characterization of a flux cone in terms of generators, as given in Math Box 4.A. Indeed, we have the following conformal sum theorem, see [2, Theorem 3].

Theorem 1. *Let $\mathcal{C}$ be the flux cone and $\{\boldsymbol{f}^{(1)}, \ldots, \boldsymbol{f}^{(\ell)}\}$ be the set of EFMs. Then,*

$$\mathcal{C} = \left\{ \mathbf{v} \mid \mathbf{v} = \sum_{k=1}^{\ell} \lambda_k \boldsymbol{f}^{(k)} \text{ with } \lambda_k \geq 0 \text{ and } \operatorname{sign}(\boldsymbol{f}^{(k)}) \leq \operatorname{sign}(\mathbf{v}) \right\}.$$

That is, any flux vector $\mathbf{v}$ in the flux cone $\mathcal{C}$ can be expressed as a conformal sum of EFMs $\{\boldsymbol{f}^{(1)}, \ldots, \boldsymbol{f}^{(\ell)}\}$.

Again, we illustrate the theoretical concepts in the simple representation of central carbon metabolism. The flux distribution $\mathbf{v} = (1,1,1,1)^T$ lies in the flux cone, cf. (4.4), and hence can be written as a conical linear combination of EFMs (in a non-unique way):

$$\begin{aligned}\mathbf{v} = \begin{pmatrix}1\\1\\1\\1\end{pmatrix} &= \boldsymbol{f}^{(1)} + \boldsymbol{f}^{(3)} = \begin{pmatrix}1\\1\\0\\2\end{pmatrix} + \begin{pmatrix}0\\0\\1\\-1\end{pmatrix} \\ &= \frac{1}{2}\,\boldsymbol{f}^{(1)} + \frac{1}{2}\,\boldsymbol{f}^{(2)} = \begin{pmatrix}\frac{1}{2}\\\frac{1}{2}\\0\\1\end{pmatrix} + \begin{pmatrix}\frac{1}{2}\\\frac{1}{2}\\1\\0\end{pmatrix}.\end{aligned} \tag{4.7}$$

Note that the first sum is not conformal: The fourth component of $\mathbf{v}$ is positive, whereas the corresponding component of $\boldsymbol{f}^{(3)}$ is negative. That is, the contributing EFMs have different signs in the net reaction rates of the fourth reaction, which leads to cancellation and is not meaningful thermodynamically. (Gibbs free energy determines reaction directions, see Section 4.4.3.) Still, the second sum is conformal: no cancellation occurs, and the decomposition is thermodynamically meaningful. Theorem 1 states that a decomposition as a conformal sum is always possible.

On the one hand, we introduced EFMs as the support-minimal vectors of the flux cone, corresponding to minimal metabolic subnetworks. On the other hand, EFMs form the (unique minimal) set of (conformal) generators of the flux cone. Indeed, the beautiful thing about EFMs is that they have several equivalent (but complementary) definitions, see Math box 4.B for examples and proofs.

Viewing EFMs as minimal metabolic subnetworks enables us to interpret an EFM in terms of its biological function; an EFM can be seen as a metabolic strategy that a cell can use to obtain steady-state metabolism, and which it can combine with other strategies to reach its purpose. The interpretation as conformal generators allows us to write an arbitrary flux vector $\mathbf{v} \in \mathcal{C}$ as a combination of EFMs in a thermodynamically meaningful way, see Theorem 1. This also means that we can learn something about all flux vectors $\mathbf{v}$ by learning something about all EFMs. For example, if we know that there is no EFM that produces compound Y without using reaction r, this immediately implies that there is no flux vector at all that can do this, and that reaction r is thus essential for the production of Y.

Finally, after reaction splitting, as described in Section 4.3.2, the flux cone is contained in the non-negative orthant and hence is pointed. Then, EFMs agree with the extreme vectors and can be computed via algorithms based on the double-description method, as discussed in Section 4.4.4.

So far, we did not consider a limit on the amount of flux that a particular EFM may carry, since $\lambda\,\boldsymbol{f}^{(k)}$ is an EFM for any $\lambda > 0$ and any EFM $\boldsymbol{f}^{(k)}$, and consequently the absolute value of any flux vector $\mathbf{v}$ in $\mathcal{C}$ is unbounded. In Section 4.4, we will see that this is not necessarily true when additional constraints are introduced.

4.3.1. Practical relevance of EFMs

EFMs represent the full set of possible metabolic capacities of an organism, which can therefore make EFM analysis a useful tool for biology. To this end, application of EFM analysis to bioengineering has been proposed to guide the genetic manipulation of microorganisms to perform desirable properties such as synthesis of a bio-compound or efficient production of a recombinant protein (e.g. [3, 4]). From a more theoretical point of view, EFMs have also been used in attempts to quantify cellular robustness [5], in particular regarding robustness under genetic perturbations [6]. The relevance of elementary flux mode analysis to cellular robustness stems from the fact that there is rarely a unique conical combination of elementary flux modes for any given flux vector, which implies there are multiple

Σ Math box 4.B Equivalent Definitions of Elementary Flux Modes (EFMs)

In the main text, we have introduced EFMs as the support minimal vectors of the flux cone, see Definition 1. In fact, EFMs can be defined as the support-minimal, support-wise non-decomposable, sign-minimal, sign-wise non-decomposable, and conformally non-decomposable vectors of the flux cone; cf. [2]. Here, we consider the latter definition for three reasons: (i) it matches Theorem 1 on the decomposition of flux distribtions into conformal sums of EFMs, (ii) it also applies to general polyhedral cones (not just s-cones such as the flux cone) and even to polyhedra and polytopes, and (iii) it establishes a link to the case when the flux cone is contained in the negative orthant. (In the latter case, the cone is pointed and generated by the extreme vectors.)

Definition 2. *A nonzero vector* $\mathbf{v} \in \mathcal{C}$ *is* conformally non-decomposable *if* $\mathbf{v} = \mathbf{v}^1 + \mathbf{v}^2$ *for any nonzero vectors* $\mathbf{v}^1, \mathbf{v}^2 \in \mathcal{C}$ *with* $\operatorname{sign}(\mathbf{v}^1), \operatorname{sign}(\mathbf{v}^2) \leq \operatorname{sign}(\mathbf{v})$ *implies* $\mathbf{v}^1 \sim \mathbf{v}^2$ *(that is,* $\mathbf{v}^1 = \lambda\,\mathbf{v}^2$*).*

As stated above, EFMs can be defined as the conformally non-decomposable vectors of the flux cone. Indeed, we have the following equivalence.

Proposition 1. *A nonzero vector* $\mathbf{v} \in \mathcal{C}$ *is conformally non-decomposable if and only if it is support-minimal.*

Proof. Assume that $\mathbf{v} \in \mathcal{C}$ is conformally decomposable, that is, $\mathbf{v} = \mathbf{v}^1 + \mathbf{v}^2$ for nonzero $\mathbf{v}^1, \mathbf{v}^2 \in \mathcal{C}$ with $\operatorname{sign}(\mathbf{v}^1), \operatorname{sign}(\mathbf{v}^2) \leq \operatorname{sign}(\mathbf{v})$ and $\mathbf{v}^1 \not\sim \mathbf{v}^2$. Then also $\mathbf{v}^1 \not\sim \mathbf{v}$, and there exists a largest $\lambda > 0$ such that the nonzero vector $\mathbf{v}' = \mathbf{v} - \lambda\,\mathbf{v}^1$ fulfills $\operatorname{sign}(\mathbf{v}') \leq \operatorname{sign}(\mathbf{v})$. For this λ, $\mathbf{v}' \in \mathcal{C}$ (that is, $\mathbf{N}\,\mathbf{v}' = \mathbf{0}$ and $\mathbf{v}'^{\rightarrow} \geq \mathbf{0}$) and $\operatorname{sign}(\mathbf{v}') < \operatorname{sign}(\mathbf{v})$ (in particular, $v'_i = 0$ and $v_i \neq 0$ for some i). Hence, $\operatorname{supp}(\mathbf{v}') \subset \operatorname{supp}(\mathbf{v})$, that is, $\mathbf{v}$ is not support-minimal.

Conversely, assume that $\mathbf{v} \in \mathcal{C}$ is not support-minimal, that is, $\operatorname{supp}(\mathbf{v}') \subset \operatorname{supp}(\mathbf{v})$ for a nonzero $\mathbf{v}' \in \mathcal{C}$. Then, there exists a largest $\lambda > 0$ such that the nonzero vectors $\mathbf{v}^1 = \frac{1}{2}\mathbf{v} + \lambda\mathbf{v}'$ and $\mathbf{v}^2 = \frac{1}{2}\mathbf{v} - \lambda\mathbf{v}'$ fulfill $\operatorname{sign}(\mathbf{v}^1), \operatorname{sign}(\mathbf{v}^2) \leq \operatorname{sign}(\mathbf{v})$. For this λ, either $\operatorname{sign}(\mathbf{v}^1) < \operatorname{sign}(\mathbf{v})$ or $\operatorname{sign}(\mathbf{v}^2) < \operatorname{sign}(\mathbf{v})$; in any case, $\mathbf{v}^1, \mathbf{v}^2 \in \mathcal{C}$ and $\mathbf{v}^1 \not\sim \mathbf{v}^2$. Clearly, $\mathbf{v} = \mathbf{v}^1 + \mathbf{v}^2$, that is, $\mathbf{v}$ is conformally decomposable. □

Conformally non-decomposable vectors are closely related to extreme (or non-decomposable) vectors.

Definition 3. *A nonzero vector* $\mathbf{v} \in \mathcal{C}$ *is* extreme *if* $\mathbf{v} = \mathbf{v}^1 + \mathbf{v}^2$ *for any nonzero vectors* $\mathbf{v}^1, \mathbf{v}^2 \in \mathcal{C}$ *implies* $\mathbf{v}^1 \sim \mathbf{v}^2$.

If the flux cone is contained in the non-negative orthant (in particular, after reaction splitting, as described in Section 4.3.2) and hence is pointed, EFMs can be defined as the extreme vectors.

Proposition 2. *Let* $\mathcal{C} \subseteq \mathbb{R}^n_{\geq}$. *A nonzero vector* $\mathbf{v} \in \mathcal{C}$ *is extreme if and only if it is conformally non-decomposable.*

Proof. If $\mathbf{v}, \mathbf{v}^1, \mathbf{v}^2 \in \mathcal{C} \subseteq \mathbb{R}^n_{\geq}$, then $\mathbf{v} = \mathbf{v}^1 + \mathbf{v}^2$ implies $\operatorname{sign}(\mathbf{v}^1), \operatorname{sign}(\mathbf{v}^2) \leq \operatorname{sign}(\mathbf{v})$, and Definitions 2 and 3 agree. □

combinations of minimal metabolic pathways to achieve the same desired effect. This redundancy can be interpreted as a measure for the metabolic robustness of an organism, in terms of preserving essential metabolic functionalities under loss of a gene, for example.

There have also been several ways that EFM analysis has been incorporated into analysis of multi-omics data. For example, on the basis of transcriptomic profiling of microorganisms, metabolic pathways associated with elementary flux modes have been scored according to their probability of carrying flux [7]. The principle here is that, although levels of RNA often serve as a poor proxy for flux through the reaction associated with that particular enzyme's gene, by creating a gene set associated with an entire EFM there might be a better chance of concretely assessing whether the metabolic pathway as a whole is likely to carry flux. The study [7] suggested that the integration of EFM analysis with gene expression data enabled the identification of certain metabolic pathways activated during stress conditions, and that the organization of elementary flux mode utilization in *Saccharomyces cerevisiae* involves a disparate combination of highly specialized and multi-tasking roles. Beyond transcriptomic profiling, isotope tracing experiments in principle provide a much more direct insight into quantifying metabolic flux. To interpret isotope

tracing data, an extension of the concept of an EFM was introduced in [8].

4.3.2. Reaction splitting for EFM computation

The computation of EFMs via the double description (DD) method as well as the solution of linear programs (LPs) via the simplex algorithm assume that the flux cone is given in certain standard forms. (Note, however, that the computation of EFMs via lexicographic reverse search (lrs) does not involve such an assumption.)

Recall that the flux cone is given by the mass-balance constraints $\mathbf{N}\,\mathbf{v} = \mathbf{0}$ and the irreversibility constraints $\mathbf{v}^{\rightarrow} \geq \mathbf{0}$, whereas standard forms are given by $\mathbf{A}\,\mathbf{v} = \mathbf{0}$ (for DD) or $\mathbf{A}\,\mathbf{v} \geq \mathbf{0}$ and $\mathbf{v} \geq \mathbf{0}$ (for LP) with a matrix $\mathbf{A}$ of appropriate dimensions. To bring the flux cone into standard form, we will split reversible reactions into irreversible forward and backward reactions. First, we order reactions such that

$$\mathbf{N}\,\mathbf{v} = \begin{pmatrix}\mathbf{N}^{\rightarrow} & \mathbf{N}^{\rightleftharpoons}\end{pmatrix}\begin{pmatrix}\mathbf{v}^{\rightarrow}\\ \mathbf{v}^{\rightleftharpoons}\end{pmatrix}, \tag{4.8}$$

where the superscripts $\rightarrow$ and $\rightleftharpoons$ refer to the irreversible and reversible reactions, $\mathcal{R}^{\rightarrow}$ and $\mathcal{R}^{\rightleftharpoons}$, respectively. Next, for every reversible reaction $i \in \mathcal{R}^{\rightleftharpoons}$ with net reaction rate v_i, we define a forward reaction with "rate" $w_i^{\rightharpoonup} \geq 0$ and a backward reaction with "rate" $w_i^{\leftharpoondown} \geq 0$ such that $v_i = w_i^{\rightharpoonup} - w_i^{\leftharpoondown}$. (In vector form, $\mathbf{v}^{\rightleftharpoons} = \mathbf{w}^{\rightharpoonup} - \mathbf{w}^{\leftharpoondown}$.) Note that the "rates" $w_i^{\rightharpoonup}, w_i^{\leftharpoondown}$ do not denote the (microscopic) forward and backward reaction rates $v_i^{\rightharpoonup}, v_i^{\leftharpoondown}$ that determine the net reaction rate $v_i = v_i^{\rightharpoonup} - v_i^{\leftharpoondown}$. They are auxiliary quantities, and only their difference $w_i^{\rightharpoonup} - w_i^{\leftharpoondown} = v_i$ has a biochemical meaning (and is the subject of constraint-based metabolic modeling). Further, for every irreversible reaction $i \in \mathcal{R}^{\rightarrow}$, we write $v_i = w_i^{\rightarrow}$ to obtain a uniform notation. (In vector form, $\mathbf{v}^{\rightarrow} = \mathbf{w}^{\rightarrow}$.) Now,

$$\mathbf{N}\,\mathbf{v} = \begin{pmatrix}\mathbf{N}^{\rightarrow} & \mathbf{N}^{\rightleftharpoons} & -\mathbf{N}^{\rightleftharpoons}\end{pmatrix}\begin{pmatrix}\mathbf{w}^{\rightarrow}\\ \mathbf{w}^{\rightharpoonup}\\ \mathbf{w}^{\leftharpoondown}\end{pmatrix}. \tag{4.9}$$

By introducing the augmented stoichiometric matrix $\overline{\mathbf{N}}$ and the corresponding non-negative flux vector $\mathbf{w}$, we can write

$$\mathbf{N}\,\mathbf{v} = \overline{\mathbf{N}}\,\mathbf{w} \quad \text{with} \quad \overline{\mathbf{N}} = \begin{pmatrix}\mathbf{N}^{\rightarrow} & \mathbf{N}^{\rightleftharpoons} & -\mathbf{N}^{\rightleftharpoons}\end{pmatrix}, \quad \mathbf{w} = \begin{pmatrix}\mathbf{w}^{\rightarrow}\\ \mathbf{w}^{\rightharpoonup}\\ \mathbf{w}^{\leftharpoondown}\end{pmatrix}. \tag{4.10}$$

As a consequence, the augmented flux cone is given by

$$\bar{\mathcal{C}} = \{\mathbf{w} \mid \overline{\mathbf{N}}\,\mathbf{w} = \mathbf{0},\, \mathbf{w} \geq \mathbf{0}\} \tag{4.11}$$

or, in LP standard form, by

$$\bar{\mathcal{C}} = \{\mathbf{w} \mid \mathbf{A}\,\mathbf{w} \geq \mathbf{0},\, \mathbf{w} \geq \mathbf{0}\} \quad \text{with} \quad \mathbf{A} = \begin{pmatrix}\overline{\mathbf{N}}\\ -\overline{\mathbf{N}}\end{pmatrix}, \tag{4.12}$$

after writing equations as non-strict inequalities.

Obviously, $\bar{\mathcal{C}}$ is contained in the non-negative orthant and hence pointed. As an important consequence, EFMs can be defined as the extreme vectors of the flux cone, see the 'Math box', and be computed by algorithms based on the DD method.

For examples of pointed polyhedral cones (in the non-negative orthant), see Figure 4.2. Note that the cone in Figure 4.2 (A) is not an s-cone and hence not a flux cone. In particular, its generators/extreme vectors lie in the interior of the non-negative orthant. On the other hand, the cone in Figure 4.2 (B) is an s-cone. Its generators/support-

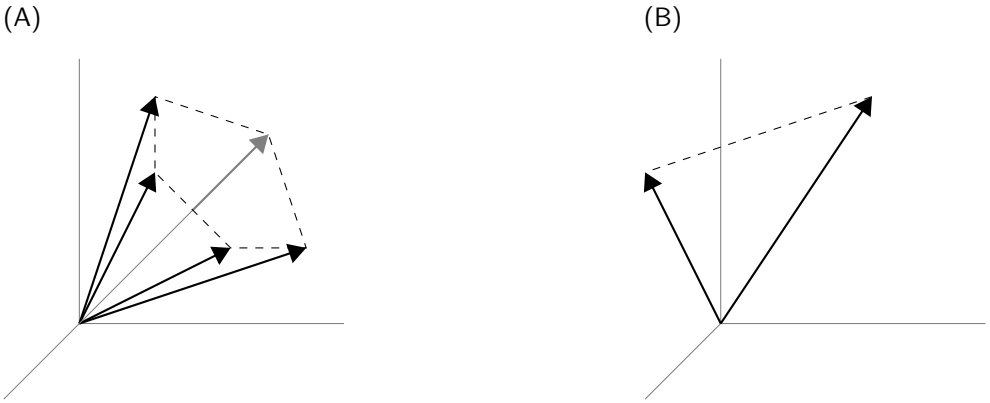


Figure 4.2: Pointed polyhedral cones. (A) A pointed polyhedral cone that is not a flux cone; all its generators lie in the interior of the non-negative orthant. (B) A pointed polyhedral cone that is a flux cone; its generators arise from the intersection of a subspace with the boundaries of the non-negative orthant.

minimal vectors/EFMs arise from the intersection of a subspace (the nullspace of the stoichiometric matrix) with the boundaries of the non-negative orthant.

Again, we return to the simple representation of central carbon metabolism presented in Figure 4.1. After reaction splitting, the mass-balance constraint can be written as

$$\overline{\mathbf{N}}\,\mathbf{w} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 2 & -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} w_1^{\rightarrow} \\ w_2^{\rightarrow} \\ w_3^{\rightarrow} \\ w_4^{\rightarrow} \\ w_4^{\leftarrow} \end{pmatrix} = \mathbf{0}. \tag{4.13}$$

In particular, the reversible fourth reaction has been split into irreversible forward and backward reactions with reaction vectors $\binom{0}{-1}$, $\binom{0}{1}$ and "rates" $w_4^{\rightarrow}$, $w_4^{\leftarrow}$, see Equation (4.5). Now, algorithms based on the DD method can be applied to the mass-balance and irreversibility constraints in standard form, $\overline{\mathbf{N}}\,\mathbf{w} = \mathbf{0}$ and $\mathbf{w} \geq \mathbf{0}$. As it turns out, the set of EFMs (support-minimal vectors) is given by

$$\boldsymbol{g}^{(1)} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 2 \\ 0 \end{pmatrix}, \quad \boldsymbol{g}^{(2)} = \begin{pmatrix} 1 \\ 1 \\ 2 \\ 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{g}^{(3)} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \quad \text{and} \quad \boldsymbol{g}^{(4)} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}. \tag{4.14}$$

EFMs $\boldsymbol{g}^{(1)}, \boldsymbol{g}^{(2)}, \boldsymbol{g}^{(3)}$ correspond to EFMs $\boldsymbol{f}^{(1)}, \boldsymbol{f}^{(2)}, \boldsymbol{f}^{(3)}$ before reaction splitting, see Equation (4.6). Just recall $v_4 = w_4^{\rightarrow} - w_4^{\leftarrow}$. However, EFM $\boldsymbol{g}^{(4)}$ corresponds to zero flux. More specifically, it represents the fourth reaction having equal forward and backward "rates" and hence zero net reaction rate. Such EFMs are artifacts of reaction splitting and need to be discarded when translating the EFMs of the augmented flux cone back to the EFMs of the original flux cone.

4.4. Extra constraints and flux polyhedra

4.4.1. Inhomogeneous linear flux constraints

We have so far been working exclusively with mass-conservation and irreversiblity constraints, which are captured entirely by the stochiometric matrix where each row is associated with a metabolite concentration at steady state. We also saw that these considerations alone result in a flux cone that is by definition unbounded, meaning that a flux vector in this space is allowed to take on any absolute value (i.e. multiplying a flux vector in the flux cone by an arbitrarily large positive number again returns a flux vector in the flux cone). However, there are physical constraints limiting the magnitude of flux vectors, especially on the values of flux through exchange reactions that may depend on concentrations of extracellular substrates, numbers of transporter molecules in the membrane, or for which we might have direct experimental measurements. Typically, such bounds on flux values are imposed using inequality constraints of the form $v_i^{\mathsf{lb}} \leq v_i \leq v_i^{\mathsf{ub}}$ where v_i^{lb} and v_i^{ub} are lower and upper bounds, respectively, for the flux through the ith reaction. When reactions have been decomposed into forward and reverse directions, both upper and lower bounds are non-negative where the latter is usually zero.

In the example from Figure 4.1, one may impose an upper bound on the flux value v_1, suggesting that there is a maximal rate at which the cell or organism can import glucose from the extracellular environment. In this case the total set of constraints on the flux vector $\mathbf{v}$ take the form

$$\mathbf{N}\,\mathbf{v} = \mathbf{0}, \quad \mathbf{v}^{\rightarrow} \geq \mathbf{0}, \quad v_1 \leq v_1^{\mathsf{ub}}, \tag{4.15}$$

where v_1^{ub} is the maximal glucose uptake rate. It is important to note that the new constraint is of a different kind than the mass-balance and irreversibility constraints: the right-hand side of the constraint is nonzero. Constraints that involve a nonzero are called *inhomogeneous constraints*. We can write these constraints in matrix form as

$$\mathbf{G}\,\mathbf{v} \geq \mathbf{h}, \tag{4.16}$$

where in this particular case

$$\mathbf{G} = \begin{pmatrix} -1 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{h} = \begin{pmatrix} -v_1^{\mathsf{ub}} \end{pmatrix}. \tag{4.17}$$

In general, the matrix $\mathbf{G}$ will have ℓ rows corresponding to ℓ inhomogeneous linear constraints of the form

$$\sum_i G_{ji}\, v_i \leq h_j, \quad j = 1, \ldots \ell. \tag{4.18}$$

That is, for constraint j, there are n entries G_{ji} $(i = 1, \ldots, n)$ of the matrix $\mathbf{G}$ and the component h_j of the ℓ-dimensional vector $\mathbf{h}$. Many constraints can be written in this general form. For example, after reaction splitting, one may impose a bound on the total flux that a cell can catalyze, by setting all entries (in the corresponding row of $\mathbf{G}$) to 1.

Altogether, the constraints on $\mathbf{v}$ define a *flux polyhedron* that is necessarily contained within the flux cone given by the homogeneous constraints $\mathbf{N}\,\mathbf{v} = \mathbf{0}$ and $\mathbf{v}^{\rightarrow} \geq \mathbf{0}$. The additional inhomogeneous constraints serve to further restrict the cone such that various (if not all) dimensions become bounded, thus bounding the total magnitude of the flux vector $\mathbf{v}$.

4.4.2. From the flux polyhedron to the EFM weight polyhedron

Via the conical sum $\mathbf{v} = \sum_{k=1}^{\ell} \lambda_k\, \boldsymbol{f}^{(k)}$, constraints on the fluxes $\mathbf{v}$ define constraints on the EFM weights $\boldsymbol{\lambda}$ and hence a corresponding *EFM weight polyhedron*. Whereas elements $\mathbf{v}$ of the flux polyhedron have entries v_i for every reaction $i \in \mathcal{R}^{\rightleftharpoons}$, elements $\boldsymbol{\lambda}$ of the EFM weight polyhedron have entries λ_k for every EFM $\boldsymbol{f}^{(k)}$, $k = 1, \ldots, \ell$ (and

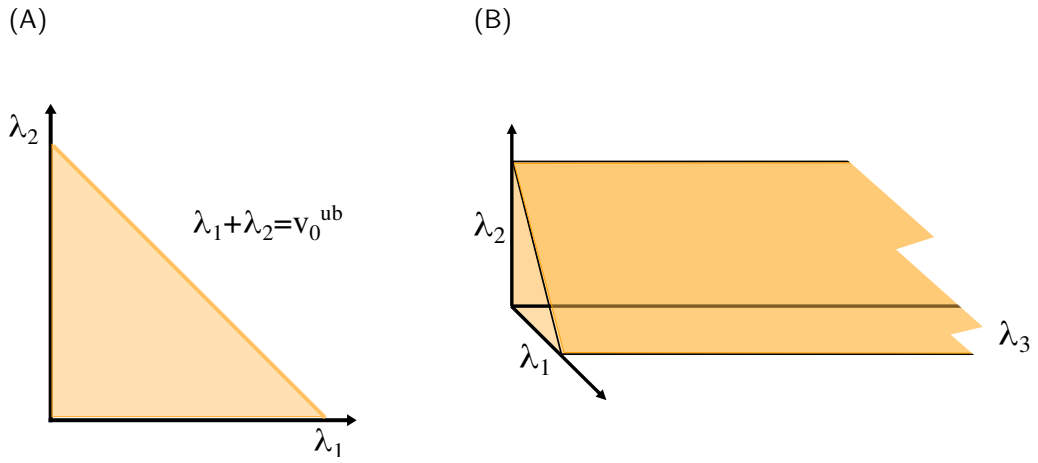


Figure 4.3: Feasible regions in the space of EFM weights - (A) Possible combinations of EFM weights λ_1 and λ_2, given by the inequality $\lambda_1 + \lambda_2 \leq v_1^{\mathsf{ub}}$ (and $\lambda_1, \lambda_2 \geq 0$). (B) Geometry of the EFM weight polyhedron (blue) representing any flux vector that satisfies the mass-balance, irreversibility, and maximal glucose uptake rate constraints. While bounded in λ_1, λ_2, it is unbounded in λ_3.

hence can be very high-dimensional).

In the example from Figure 4.1, let $\lambda_1, \lambda_2, \lambda_3 \geq 0$ be the weights of the corresponding EFMs in the (conformal) sum $\mathbf{v} = \sum_{i=1}^{3} \lambda_i \boldsymbol{f}^{(i)}$. Bounding the extracellular glucose uptake rate puts an upper bound on the weights of EFMs $\boldsymbol{f}^{(1)}, \boldsymbol{f}^{(2)}$ (involving the glucose uptake reaction),

$$\lambda_1 + \lambda_2 \leq v_1^{\mathsf{ub}}, \tag{4.19}$$

see also Figure 4.1 (B). However, the weight of EFM $\boldsymbol{f}^{(3)}$ (associated with uptake and oxidation of the fermentation product) can remain unbounded.

For this simple example, it is quite straightforward to interpret the geometric consequences of the maximal glucose uptake rate. Any flux vector $\mathbf{v}$ in the resulting flux polyhedron now corresponds to a point (λ_1, λ_2) in the (projected) EFM weight polyhedron depicted in Figure 4.3 (A). However, the weight λ_3 remains free, and the (full) EFM weight polyhedron is depicted in Figure 4.3 (B). In terms of the flux polyhedron, the maximal glucose uptake has restricted the flux cone along v_1, v_2 while leaving v_3, v_4 unbounded.

In order to obtain a bounded flux polyhedron (a flux *polytope*), we impose an upper bound on the uptake rate of the fermentation product, that is, $-v_4 \leq v_4^{\mathsf{ub}}$. In terms of the EFM weights, we obtain the bound $-2\lambda_1 + \lambda_3 \leq v_4^{\mathsf{ub}}$. Since conformal sums are sufficient to generate the flux cone, this simplifies to $\lambda_3 \leq v_4^{\mathsf{ub}}$. Altogether, the EFM weight polyhedron is given by

$$\lambda_1, \lambda_2, \lambda_3 \geq 0, \quad \lambda_1 + \lambda_2 \leq v_1^{\mathsf{ub}}, \quad \lambda_3 \leq v_4^{\mathsf{ub}}. \tag{4.20}$$

Indeed, all EFM weights and hence all fluxes are bounded.

More general constraints, for larger metabolic reaction networks will be more difficult to interpret and visualize in such simple geometric terms. Quite quickly the combinatorial complexity associated with combinations of multiple constraints and EFMs will become unmanageable. The intuitive treatment of inhomogeneous linear constraints is partially assisted using the concept of elementary flux vectors on which we will add a section in a later version of this book, but both geometrically and biologically these objects are nowhere near as easy to interpret as their EFM counterparts. We shall see that alternative computational methods for exploring flux space therefore become imperative.

As a final remark, we clarify once more that the general form of constraints (4.16) is by no means restricted to sums on the left hand side that involve just a single reaction and can of course include constraints on weighted sums of flux values for different reactions. These weighted sums are often associated with particular biological interpretations: in the example from Figure 4.1, we might want to restrict our search of flux space to those flux vectors $\mathbf{v}$ that produce adenosine triphosphate (ATP) at a rate of at least v^{ATP}. Although a more elaborate model would of course include ATP as one of the metabolites, in this example we can use our biological understanding of central carbon metabolism to see that ATP is produced in reactions v_2 and v_3. A lower bound on ATP production would thus be a lower bound on a combination of v_2 and v_3 with coefficients determined by stoichiometry (depending on the organism under investigation). We could write such a constraint as

$$\alpha_1 v_1 + \alpha_3 v_3 \geq v^{\mathrm{ATP}} \tag{4.21}$$

with appropriate coefficients α_1, α_3. Such a constraint forms an additional row of the matrix $\mathbf{G}$ and we leave it as an exercise for the reader to explore how this affects the geometry of the flux polytope for various values of the coefficients, minimal ATP production rate and maximal glucose and fermentation product uptake rates. Particular combinations of constraints will be impossible to satisfy simultaneously (i.e. when the minimal rate of ATP production is impossible to achieve under the given bounds on glucose and fermentation product uptake rates), resulting in a flux polytope that is empty. In such cases the set of constraints on $\mathbf{v}$ are called *infeasible*.

4.4.3. Thermodynamic constraints

In Chapter 3 in [9] the basic concepts of chemical thermodynamics were introduced, in particular, the Gibbs free energy of a metabolic reaction was defined in terms of the concentrations of its products and substrates. For a metabolic reaction network with stochiometric matrix $\mathbf{N}$, the vector of Gibbs free energies (one for each reaction in the network) $\mathbf{\Delta_r G'}$ can be written in matrix form as

$$\mathbf{\Delta_r G'} = \mathbf{\Delta_r G'^{o}} + RT \cdot \mathbf{N}^{\top} \ln(\mathbf{s}) \tag{4.22}$$

where R is the gas constant, T the temperature and $\mathbf{s}$ the vector of metabolite concentrations at steady state. The components of the vector $\mathbf{\Delta_r G'^{o}}$ are the changes in standard Gibbs free energy for each corresponding reaction. Typically, these values are not known precisely for reactions in the network, but can be estimated or approximated from experimental data using methods beyond the scope of this chapter. Similarly, although it is often difficult to accurately measure all metabolite concentrations, in principle the vector $\mathbf{s}$ can be obtained experimentally. However, in practice experimental data on $\mathbf{s}$ and $\mathbf{\Delta_r G'^{o}}$ are almost never available. Various methods have therefore been developed to combine estimation of $\mathbf{\Delta_r G'^{o}}$ (sometimes with partial measurements of $\mathbf{s}$) with advanced computational techniques that allow simultaneous optimization (see next chapter) or sampling (see below) of $\mathbf{v}$ and $\mathbf{s}$ (or equivalently: $\mathbf{\Delta_r G'}$).

The second law of thermodynamics applied to chemical reaction networks can be summarized by saying that every component of the metabolic flux vector $\mathbf{v}$ must satisfy the condition

$$\mathrm{sign}(v_i) = -\mathrm{sign}(\Delta_{\mathrm{r}} G'_i) \tag{4.23}$$

where v_i and $\Delta_{\mathrm{r}} G'_i$ are the ith components of $\mathbf{v}$ and $\mathbf{\Delta_r G'}$, respectively, and $\mathrm{sign}(x)$ denotes the sign of a variable x, and $\mathrm{sign}(0) = 0$. It is important to point out that this notation is different to that used previously, where we had assumed all v_i to be non-negative by decomposing each reaction into irreversible forward and backward reactions. Returning to this reversible notation simplifies the inclusion of thermodynamic constraints into constraint-based models and also their interpretation. According to the second law, a reaction can only proceed in a direction where the change in Gibbs free energy is negative. Thus, to be consistent with mass-balance *and* the second law of thermodynamics, a flux vector $\mathbf{v}$ must simultaneously satisfy both (4.1) *and* (4.23), with $\mathbf{\Delta_r G'}$ defined in (4.22).

The consequence of these additional constraints on the geometry of the space of metabolic flux distributions is to exclude quadrants incompatible with the signs of $\mathbf{\Delta_r G'}$. Equivalently, imposing the second law of thermodynamics on metabolic flux distributions removes regions of the space that are associated with combinations of thermodynamically-infeasible reaction directionalities.

The resulting space of feasible flux vectors is almost always non-convex, which means more advanced computational methods are required to explore it efficiently. The intuitive reason for this is that imposing thermodynamic constraints on top of the mass-balance constraint is usually done in terms of Boolean variables, which breaks the linearity of the problem that we had and exploited so far. Relating this to the EFMs that were discussed previously, it for example becomes clear that any EFM representing an internal cycle –not including any exchange reactions– will never be thermodynamically feasible. Thus, thermodynamic constraints also reduce the set of EFMs that are possible in a metabolic network. Interestingly, it turns out that any thermodynamically-feasible metabolic flux vector can be expressed solely in terms of thermodynamically-feasible EFMs [10], but the converse statement is not true: a linear combination of thermodynamically-feasible elementary flux mode does not necessarily satisfy the thermodynamic constraints. This shows how the workable properties of convex spaces break down as the mathematical models become more complex, in this case by accounting for thermodynamics.

4.4.4. Computational challenges for EFM analysis

Enumerating EFMs for large networks can be computationally challenging if not impossible. In principle, EFMs can be found by removing one reaction at a time and solving the resulting mass-balance constraint problem until it is no longer possible to remove a reaction and still obtain a flux vector that satisfies the steady state conditions. However, the equivalence of EFMs and extreme vectors of the flux cone (after reaction splitting) described in Section 4.3.2 enables the use of algorithms that are specialized in the efficient enumeration of extreme rays of polyhedral cones, such as the double description method [11]. Various tools have been developed for elementary flux mode enumeration based on this algorithm (e.g. EFMTOOL [12] or MetaTool [13]). However, when the size of the metabolic reaction network grows, the number of EFMs scales disproportionately, leading to a combinatorial explosion that effectively makes enumeration impossible for genome-scale networks containing several thousands of reactions [14]. Currently, EFM analysis is therefore restricted to medium-scale reconstructions containing on the order of several hundreds of reactions, and results in the identification of several hundred million EFMs (e.g. enumeration based on the *Escherichia coli* core model results in approximately 272 million EFMs).

Approaches to reduce the complexity of dealing with so many EFMs even for metabolic reaction networks of modest size have also been proposed. These include invoking transcriptional regulatory constraints to eliminate most of the EFMs to be considered in downstream analysis. Imposing additional constraints based on thermodynamic conditions similarly reduces the set of EFMs considerably. A problem with these approaches is evidently that they do still depend on an initial calculation of all EFMs, and so do not solve the problem of enumeration complexity. A rigorous study of the complexity of EFM mode enumeration was performed by Acuña and colleagues [15]. They showed that the decision problem if there exists an EFM containing two specific reactions is NP-complete whilst the complexity of enumerating all EFMs remains open.

Later in this chapter we will explore some alternatives to EFM enumeration that reduce the difficulty of enumeration, cf. Section 4.5.

4.4.5. Reducing combinatorics of EFMs computation

In order to reduce the combinatorics of EFM computation to a feasible order, the search space may be limited to the biologically relevant EFMs only. This can be done by considering additional biological constraints before, during, or after the computation of EFMs. One way to restrict the search space is to remove all 'irrelevant' reactions in a metabolic network, that is,

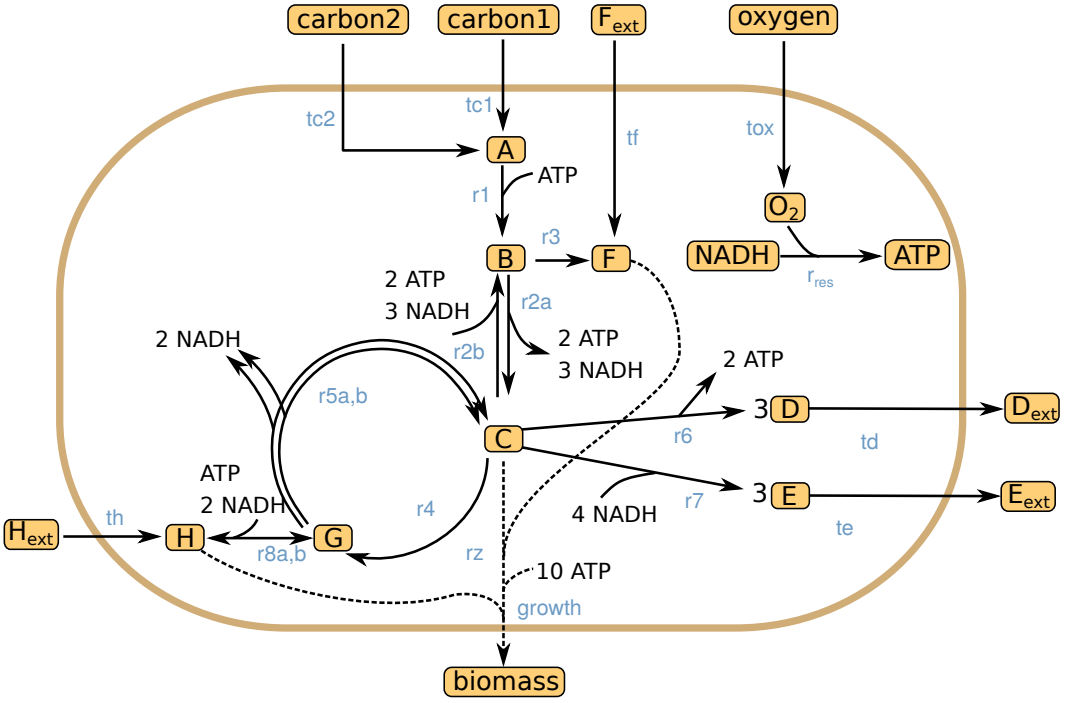


Figure 4.4: Metabolic model from [16]. Transcriptional regulation shown in Figure 4.5.

- reactions that are not essential for the cell (not part of the core metabolism),
- reactions that are not performed for *chemo-physical*, *kinetic*, or *thermodynamic* reasons,
- reactions that are too expensive in terms of enzymatic resource allocation,
- reactions that transport metabolites which are not present in the growth medium ('*environmental regulation*'),
- reactions that are catalyzed by enzymes whose expression is inhibited by *transcriptional regulation*.

The purpose of incorporating biological constraints, from the perspective of a modeler, is to reduce the number of pathways the biologist needs to analyze. Additionally, the computation of EFMs becomes much more efficient because fewer solutions need to be computed.

Below we are going to illustrate the last two types of constraints: environmental and transcriptional regulation. Both types can be expressed using Boolean constraints. A Boolean constraint is a Boolean function $f : \mathbb{B}^k \to \mathbb{B}$, where $\mathbb{B} = \{0, 1\}$, which takes in k Boolean inputs $z \in \mathbb{B}^k$ and produces a Boolean output $b \in \mathbb{B}$ such that $b = f(z)$. In our case, Boolean functions determine whether reactions are allowed or not in EFMs based on biological conditions. To this end, reactions are associated with a Boolean indicator. The value of this indicator (either 0 or 1) determines whether that reaction can participate in an EFM.

The following relationship, for a set of reactions R with corresponding fluxes v and indicators z, determines how Boolean regulation affects the presence of reactions in EFMs:

$$\forall r \in R : \ (z_r = 0) \implies (v_r = 0)$$

As an example, we consider the following small metabolic model from [17, 16], which involves transcriptional and environmental regulation. The network contains 18 reactions and 18 metabolites (10 internal and 8 external) and has 80 EFMs. For an illustration, see Figure 4.4. The formulae describing reaction stoichiometries and regulation rules are shown in Figure 4.5. This model makes for a good basis for studying the effect of Boolean constraints on

(A) Variables

Internal metabolites: $M = \{\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}, \mathrm{E}, \mathrm{F}, \mathrm{G}, \mathrm{H}, \mathrm{O_2}, \mathrm{ATP}, \mathrm{NADH}\}$

External metabolites: $E = \{\mathrm{carbon1}, \mathrm{carbon2}, \mathrm{D_{ext}}, \mathrm{E_{ext}}, \mathrm{F_{ext}}, \mathrm{H_{ext}}, \mathrm{oxygen}, \mathrm{biomass}\}$

Boolean indicators for growth medium: $G = \{m_{\mathrm{carbon1}}, m_{\mathrm{carbon2}}, m_{\mathrm{H}}, m_{\mathrm{F}}, m_{\mathrm{oxygen}}\}$

Reactions: $R = \{\mathrm{r1}, \mathrm{r2a}, \mathrm{r2b}, \mathrm{r3}, \mathrm{r4}, \mathrm{r5a}, \mathrm{r5b}, \mathrm{r6}, \mathrm{r7}, \mathrm{r8a}, \mathrm{r8b}, \mathrm{rres}, \mathrm{tc1}, \mathrm{tc2}, \mathrm{td}, \mathrm{te}, \mathrm{tf}, \mathrm{th}, \mathrm{tox}, \mathrm{growth}\}$

Flux vector: $v = \{v_{\mathrm{r1}}, v_{\mathrm{r2a}}, v_{\mathrm{r2b}}, \ldots\}$

Boolean indicators for reactions: $z = \{z_{\mathrm{r1}}, z_{\mathrm{r2a}}, z_{\mathrm{r2b}}, \ldots\}$

(B) Stoichiometry

Internal reactions

$$\mathrm{r1} : \mathrm{A} + \mathrm{ATP} \rightarrow \mathrm{B}$$
$$\mathrm{r2a}, \mathrm{r2b} : \mathrm{B} \leftrightarrow 2\ \mathrm{ATP} + 2\ \mathrm{NADH} + \mathrm{C}$$
$$\mathrm{r3} : \mathrm{B} \rightarrow \mathrm{F}$$
$$\mathrm{r4} : \mathrm{C} \rightarrow \mathrm{G}$$
$$\mathrm{r5a} : \mathrm{G} \rightarrow 0.8\ \mathrm{C} + 2\ \mathrm{NADH}$$
$$\mathrm{r5b} : \mathrm{G} \rightarrow 0.8\ \mathrm{C} + 2\ \mathrm{NADH}$$
$$\mathrm{r6} : \mathrm{C} \rightarrow 2\ \mathrm{ATP} + 3\ \mathrm{D}$$
$$\mathrm{r7} : \mathrm{C} + 4\ \mathrm{NADH} \rightarrow 3\ \mathrm{E}$$
$$\mathrm{r8a}, \mathrm{r8b} : \mathrm{G} + \mathrm{ATP} + 2\ \mathrm{NADH} \leftrightarrow \mathrm{H}$$
$$\mathrm{rres} : \mathrm{NADH} + \mathrm{O_2} \rightarrow \mathrm{ATP}$$

Transport reactions

$$\mathrm{tc1} : \mathrm{carbon1} \rightarrow \mathrm{A}$$
$$\mathrm{tc2} : \mathrm{carbon2} \rightarrow \mathrm{A}$$
$$\mathrm{td} : \mathrm{D} \rightarrow \mathrm{D_{ext}}$$
$$\mathrm{te} : \mathrm{E} \rightarrow \mathrm{E_{ext}}$$
$$\mathrm{tf} : \mathrm{F_{ext}} \rightarrow \mathrm{F}$$
$$\mathrm{th} : \mathrm{H_{ext}} \rightarrow \mathrm{H}$$
$$\mathrm{tox} : \mathrm{oxygen} \rightarrow \mathrm{O_2}$$
$$\mathrm{growth} : \mathrm{C} + \mathrm{F} + \mathrm{H} + 10\ \mathrm{ATP} \rightarrow \mathrm{biomass}$$

(C) Regulation

Transcriptional

$$z_{\mathrm{r2b}} \implies \neg z_{\mathrm{r2a}}$$
$$\neg m_{\mathrm{oxygen}} \implies \neg z_{\mathrm{r5a}}$$
$$m_{\mathrm{oxygen}} \implies \neg z_{\mathrm{r5b}}$$
$$z_{\mathrm{r2b}} \implies \neg z_{\mathrm{r7}}$$
$$m_{\mathrm{H}} \implies \neg z_{\mathrm{r8a}}$$
$$\neg m_{\mathrm{oxygen}} \implies \neg z_{\mathrm{rres}}$$
$$m_{\mathrm{carbon1}} \implies \neg z_{\mathrm{tc2}}$$

Environmental

$$\neg m_{\mathrm{carbon2}} \implies \neg z_{\mathrm{tc2}}$$
$$\neg m_{\mathrm{carbon1}} \implies \neg z_{\mathrm{tc1}}$$
$$\neg m_{\mathrm{H}} \implies \neg z_{\mathrm{th}}$$
$$\neg m_{\mathrm{F}} \implies \neg z_{\mathrm{tf}}$$
$$\neg m_{\mathrm{oxygen}} \implies \neg z_{\mathrm{tox}}$$

Figure 4.5: Formulae for the metabolic model from Figure 4.4. Stoichiometry is given for reactions and metabolites; simple arrow or double arrow represent reversibility, for instance reaction $\mathrm{r1}$ consumes one A and one ATP to produce one B. The names $\mathrm{r2a}, \mathrm{r2b}$ and $\mathrm{r8a}, \mathrm{r8b}$ denote the forward and backward directions of the respective reactions, while $\mathrm{r5a}, \mathrm{r5b}$ represent isozymes. Boolean inputs for the Boolean functions can be either growth medium metabolites or reactions.

a small scale: out of 80 EFMs, only 26 are consistent with the regulation in the most permissive growth medium – and even fewer are found when the growth medium gets restricted [16].

As mentioned above, we distinguish two types of Boolean functions. First, *environmental regulation* applies to uptake transporters and is automatically constructed from the defined growth medium. For example, the oxygen transport reaction can only be active if external oxygen is present in the growth medium ($\neg m_{\mathrm{oxygen}} \implies \neg z_{\mathrm{tox}}$). Second, *transcriptional regulation* is reconstructed from a literature review and curated by the modeler. For instance, $\mathrm{r7}$ is regulated by the level of metabolite B in the cell, its enzyme cannot be expressed at the same time as B is being produced by $\mathrm{r2}$ ($z_{\mathrm{r2b}} \implies \neg z_{\mathrm{r7}}$). Some individual constraints mimic the behavior of *E. coli*: the activation of

respiration reaction $\mathrm{r5a}$, $\mathrm{r5b}$, rres depends on the presence of oxygen (motivated by the transcriptional factors ArcA and FNR); $\mathrm{tc2}$ is deactivated when faced with $\mathrm{carbon1}$, mimicking the behavior of glucose catabolite repression by CRP. Ultimately, these transcriptional and environmental constraints serve to filter out EFMs. For instance, the elementary mode $\{\mathrm{r2b}, \mathrm{r3}, \mathrm{r4}, \mathrm{r5b}, \mathrm{r8b}, \mathrm{rres}, \mathrm{th}, \mathrm{tox}, \mathrm{growth}\}$ is not consistent with regulation. Indeed, we have: $z_{\mathrm{r5b}} \implies \neg m_{\mathrm{oxygen}}$ and $z_{\mathrm{tox}} \implies m_{\mathrm{oxygen}}$, a contradiction.

Regulation Boolean constraints could be incorporated into the EFM computation by the method *regEFMTool*, as well as in the tools *SMTTool* and *aspefm* [18, 19, 20]. These constraints lend themselves naturally to logical encoding, making logic programming such as Answer Set Programming (ASP) well suited to this type of problem. Unlike traditional double description methods, which struggle with the combinatorial explosion of EFMs by the number of reactions and do not inherently handle regulatory constraints, ASP allows for an intuitive representation of Boolean constraints and efficient pruning of infeasible solutions early in the computation. In the simplest cases reactions that cannot respect the regulation constraints are directly deactivated in pre-processing.

Adding *environmental regulation* and restricting the analysis to a limited growth medium is crucial for reducing the computational load of the analysis. The software *regEFMTool* from Jungreuthmayer *et al.* was tested on Orth, Fleming and Palsson's *E. coli* core model [18, 21], a central carbon metabolic model of 95 reactions containing a complete *transcriptional regulation network*. The analysis was performed with all uptake reactions allowed. The total number of EFMs was reduced from 226.3 million to 2 million EFMs after post-processing.

Using *aspefm*, Mahout *et al* applied *environmental regulation*, *transcriptional regulation*, as well as *thermodynamic constraints* in order to further reduce that set to a subset of only 10^3 EFMs for post-processing analysis of optimal uptake rates [20, 22]. In general, we therefore recommend to routinely incorporate basic regulation constraints checking in order to drastically reduce the complexity of search of EFMs on metabolic models. This is particularly true for genome-scale models, which number of reactions reach thousands and number of EFMs reach billions. Ideally the procedure should be done in pre-processing, coupled with network compression.

Instead of inactivating reactions, one might be interested by computing all EFMs containing a specific reaction, such as the biomass, or several reactions, *e.g.* biomass synthesis and ATP maintenance. This is not a good idea to try to incorporate these constraints directly into the computation as such a constraint adds an hyperplane on the solution space, changing the resulting solutions [23, 24]. As a result, these kind of candidate constraints are best left for post-processing.

4.5. Alternative methods for flux space exploration

As we described above, exploration of all possible flux distributions using EFMs can become very complex for larger models. A genome-scale model, which comprises all metabolic reactions that an organism can catalyze, typically contains thousands of reactions, which prohibits the enumeration of EFMs. At the moment, it is unclear whether, even if we would have an enormously fast computer that could compute all EFMs, the number of EFMs would not be so large that we cannot store the EFMs anywhere, nor analyze it in any meaningful way. Here we discuss several alternatives for exploring the metabolic capabilities of a cell that try to avoid the combinatorial complexity that hinders EFM analysis.

4.5.1. Elementary conversion modes

If we are interested in the metabolic capabilities of an organism, is it always necessary to know all possible flux vectors? For example, what if we want to lab-culture an organism of which we have a reconstructed metabolic network, but no idea what nutrients it needs to grow. Then we only need to know from what combinations of nutrients it can make all its cell components. Or, what if we want to model the possible cross-feeding interactions between several microbial species? Then we are mostly interested in what each of them can consume and produce,

and not really in how they do that. Elementary conversion modes (ECMs), introduced in 2005 by Urbanczik and Wagner [25], capture all possible overall conversions from nutrients to products that an organism can catalyze, while ignoring which individual reactions are used for this.

ECMs focus on the net results of metabolism, i.e. on the uptake and production of compounds external to the metabolic network, such as sugars, nitrogen sources, fermentation products but also 'biomass'. To get information about these compounds we need to extend our metabolic network by including the external compounds as rows in the stoichiometry matrix; this is in general easy to do since we already had exchange reactions (reactions where an external compound was imported or exported) so we only have to find the stoichiometric coefficient in which the external compound was involved in these reactions. Let us denote the original stoichiometry matrix by $\mathbf{N}_{\text{int}}$ and the submatrix that we add by $\mathbf{N}_{\text{ext}}$; together they form $\mathbf{N}_{\text{tot}}$. We can then define the *conversion cone*:

$$\mathcal{C} = \left\{ \frac{\mathrm{d}\mathbf{s}}{\mathrm{d}t} = \mathbf{N}_{\text{ext}}\mathbf{v} \mid \mathbf{N}_{\text{int}}\mathbf{v} = \mathbf{0}, \mathbf{v} \geq \mathbf{0} \right\}. \tag{4.24}$$

If we look carefully at this definition we can see that the flux vectors $\mathbf{v}$ need to satisfy exactly the same constraints as in the flux cone (Eq. (4.2)). The only difference between flux and conversion cones is that we are either interested in the fluxes themselves, or rather in the conversions that they induce: $\mathrm{d}\mathbf{s}/\mathrm{d}t = \mathbf{N}_{\text{ext}}\mathbf{v}$.

Definition 4. *The set of* ECMs *is the minimal set of conversions* $\{ecm^1, \dots ecm^\ell\}$ *(where* ecm^i_k *is the amount of metabolite* k *produced in the* i*th elementary conversion mode), such that*

1. *all conversions* $\mathrm{d}\mathbf{s}/\mathrm{d}t \in \mathcal{C}$ *can be written as a positive sum of these elementary conversion modes:* $\mathrm{d}\mathbf{s}/\mathrm{d}t = \sum_i \lambda_i ecm^i$*, with* $\lambda_i \geq 0$*,*
2. *without the production of any metabolite being canceled in that sum, i.e. for all metabolites* k *we either have for all* $\lambda_i > 0$ *that* $ecm^i_k \geq 0$ *or for all* $\lambda_i > 0$ *that* $ecm^i_k \leq 0$*.*

We will explain both parts of this definition below, but let us first remark that the definition is in fact perfectly analogous to the definition of EFMs: EFMs are the *elementary vectors* (or precisely: conformally non-decomposable vectors) of the flux cone, and ECMs of the conversion cone. The reason that the definition of ECMs has an additional requirement (2.) is just that the analogous requirement was automatically satisfied for EFMs because we assumed all reactions to be irreversible.

In Figure 4.6A we show a small metabolic network with external metabolites A, B and BM, and internal metabolites C, D and E. We can find 9 EFMs in this network: one that goes from A to B, four that produce BM starting from A and four that produce BM from B. We get four EFMs to go from A to BM because there are two ways of going from C to D and again two for converting D into E. This makes clear that having a number of modules of alternative reactions can quickly give rise to large numbers of EFMs, even though the overall conversion from nutrients to products remains the same. In contrast, we will explain that we only get three ECMs.

In Figure 4.6B we see the conversion cone in gray. Note that this cone does not live in flux space, but rather in the space of external metabolite changes, or conversions. We recognize that the cone can be spanned by two extreme rays, which correspond to converting A into B (blue) and to using $2B$ to produce BM (yellow), so these rays correspond to elementary conversion modes following the first part of Definition 4. Now why do we have a third ECM, when the blue and yellow one already span the whole conversion cone? Indeed, the third vector in Figure 4.6B can be obtained by summing the yellow vector and two times the blue vector: $2(-1,1,0) + (0,-2,1) = (-2,0,1)$. However, note that the production of metabolite B would cancel in this sum, which is not allowed according to the second part of Definition 4. The reason that this second part of the definition is important, is that the elementary conversion modes are intended to capture all metabolic capabilities of an organism, so taking only the first two modes would not be enough: we also want to account for the possibility of making BM from A even if we decide that the elementary conversion mode from B to BM is not possible in the current environment, for example because B is

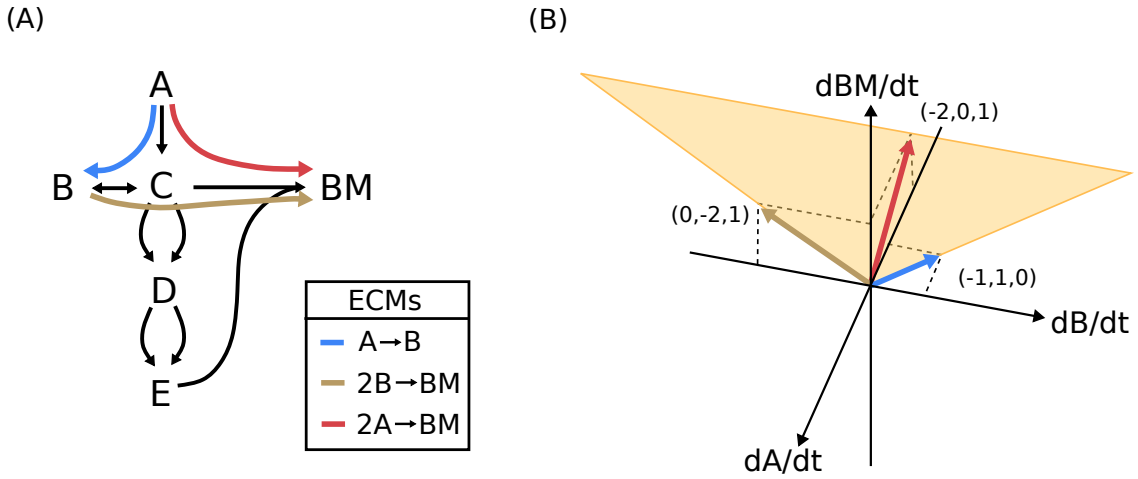


Figure 4.6: Elementary conversion modes – (A) Small toy network with three ECMs shown in blue, yellow and red. Note that the red mode can be decomposed as a positive combination of the blue and yellow elementary conversion modes, but that would cancel the production of B so this is not allowed. (B) The conversion cone is shown in gray, and the blue and yellow arrow correspond to the blue and yellow ECMs are the extreme rays. The red ECM needs to be added because it is on the intersection with the $\mathrm{d}B/\mathrm{d}t = 0$-plane.

not present as a nutrient in the medium.

Because many EFMs result in the same overall conversion, the exploration of metabolic capabilities can now be done in larger networks, at the cost of ignoring information about which reactions are used [26]. This way of thinking can be pushed even further: what if one is not interested in the conversions between all nutrients and products, but only between a subset of these? In that case, one would want to compute the ECMs only between the external metabolites of the most interest. This can be done with a small trick. Say that we are not interested in the production of external metabolite X. Before we start the enumeration algorithm we add a virtual reaction to the network that consumes and produces X from nothing, i.e. we add $X \rightleftarrows \emptyset$, and then we change X from an external metabolite to an internal metabolite. Consequently, it now has to satisfy the mass-balance constraint (which can always be done trivially using the added virtual reaction), and will thus never show up in the computed elementary conversions. In this way it was possible to compute all ECMs between glucose, oxygen and biomass for a real genome-scale network of *E. coli*.

4.5.2. Flux sampling

In addition to the computational complexity of EFM enumeration for large metabolic networks, these objects are not necessarily related to experimentally-derived flux measurements. This is because when a vector of experimentally-measured flux values $\mathbf{v}$ would be decomposed into EFMs, this generally does not give a unique solutions because it can be done in many ways. Flux sampling methods can be employed to solve both the computational and interpretability problems simultaneously, exploring the set of flux vectors (i.e. directly measurable in principle) by computationally sampling from the flux space. The goal of flux sampling in general terms is to produce a sequence of flux vectors that satisfy the steady state constraints until enough samples have been generated to provide an approximate representation of the entire flux space. The flux polyhedra defined by mass-balance and additional inhomogenous linear constraints are convex, and therefore uniform sampling of these flux spaces can be achieved using variants of an algorithm developed for convex analysis called the coordinate hit-and-run (CHR) algorithm [27]. Briefly, the most basic implementation of the CHR algorithm generates a Markov chain of flux vectors by starting in a random position within the flux polytope, picking a direction at random (uniform), and moving a random distance (uniform) in that direction from the current point. The resulting point is returned as a flux vector instance and the process repeats from there. It has been proven that the CHR algorithm converges to a stationary distribution of the Markov chain that is a uniform distribution in the flux space. Alternatives to uniform sampling (i.e. alternative

distributions across the flux polytope) can also be achieved using variants of the CHR algorithm.

As highlighted previously in Section 4.4.3, mass-balance and inhomogeneous linear constraints alone often do not contain enough information to sufficiently reduce the space of biologically-feasible flux vectors. For example, thermodynamic constraints on flux vectors are important for ruling out a large proportion of the sampled flux vectors as infeasible, but this may disproportionately dominate the resulting sampling distributions. Unfortunately, for mathematical reasons too deep to go into here, simply removing these infeasible flux distributions post-sampling will not result in a uniform distribution over the thermodynamically-feasible portion of flux space. In fact, this relevant subset of flux space cannot be defined explicitly, and is usually neither convex nor connected meaning that no Markov chain methods exist for sampling. As an alternative, a recent method [28] has been developed to combine thermodynamic constraints, physiological observations and estimated thermodynamic parameters, with mass-balance and inhomogeneous linear constraints to provide a probabilistic thermodynamic analysis of metabolic reaction networks. Advances such as these will almost certainly aid a more complete characterization of flux space as data and methods become available.

4.5.3. Minimal cut sets

A minimal cut set (MCS) is a set of reactions that, when disabled, disables a set of modes, which in turn can represent a biological function, such as the secretion of a side product. This enables the prediction of gene deletion targets, given that the genes coding for the involved reactions are known. A cut set is minimal if the removal of one or more reactions from the set leads to at least one of the targeted modes not being disabled.

In order to avoid also disabling desired functionalities, such as product secretion and growth, the concept of constrained minimal cut sets (cMCSs) has been developed. cMCSs enable targeting a set of modes while at the same time making sure that some elements of another set of modes will remain active.

Motivation for (constrained) Minimal Cut Sets The concept of MCSs was introduced by Klamt and Gilles in 2004 [29] and subsequently generalized and improved [30, 31, 32]. As briefly outlined above, the idea is to define a set of EFMs which should be disabled, for example because they generate an unwanted side product or because they don't generate the product of interest with a sufficiently high yield. Since EFMs are minimal, removing a single reaction will disable it. A cut set is a set of reactions of which at least one is active in each of the EFMs in the targeted group. Thus, disabling the reactions contained in the cut set will disable all of the targeted EFMs, and each cut set therefore represents the prediction of a set of gene deletions. Since it would be pointless to remove reactions which only target EFMs that were already targeted by other reactions, cut sets are required to be minimal. This means that removing a single reaction from the cut set would lead to one or more of the targeted EFMs to survive the intervention and also that adding a single reaction to the cut set would have no additional effect on the set of target EFMs.

The pitfall when using MCSs is that while they guarantee the elimination of the targeted EFMs, all other EFMs may be affected as well. This means that modes with desired phenotypes, such as high growth and/or high product yield, may become impossible. Therefore, cMCSs were developed [33]. In this extension of the concept of MCSs it is now possible to additionally define a set of EFMs which are desired, i.e. which can not be disabled by the cMCSs. This is usually implemented by the requirement that at least a specified minimum number of EFMs of the desired set need to remain active. Summarizing, cMCSs are sets of reactions which guarantee that (i) the full set of target EFMs is disabled and (ii) a certain minimum of desired EFMs has to remain unaffected. The drawback, with both MCSs and cMCSs, is that the target (and desired) EFMs need to be defined. This is generally achieved by defining cut-offs in terms of product yield and growth, which is, however, ultimately arbitrary.

Calculation of (constrained) Minimal Cut Sets Since minimal cut sets in a metabolic network are EFMs in a dual network [34], methods used for calculating EFMs can be used to calculate MCSs. Among other approaches [35] one based on binary integer programming has been developed [36, 37]. While it requires that the EFMs are

calculated before it can be applied, the advantage is that the algorithm is very intuitive. After having calculated the modes, each is represented as a binary vector which is zero for reactions with zero flux and one otherwise. The EFMs are then divided into either targeted or desired. A binary vector, corresponding to the cMCSs being calculated is introduced. It will have a one if the corresponding reaction remains active and zero if the reaction is disabled. The first requirement is that cMCS needs to disable all target modes and thus the vector must have zero elements such that each target EFM must have at least one corresponding non-zero element. The second requirement is that at least a defined minimum of desired modes must remain active. This is achieved by introducing a second binary vector. This vector has an element for each EFM and is calculated so that it has a zero when the mode is disabled by the cMCS and one otherwise. By adding the constraint that the number of ones in this vector must at least equal the previously defined minimum, the second requirement is met. Maximizing the vector corresponding to the cMCS yields the first solution. The next solution can be found by adding constraints to make sure that the current one is excluded.

4.6. Concluding remarks

In this chapter we studied how the individual reactions that an organism can catalyze together give rise to the overall conversion of nutrients into cell components and secretion products. For that, we studied the cell's metabolism under a number of simplifying assumptions, most notably, we model metabolism in steady-state. Given this steady-state constraint, we explained how all feasible flux distributions form a space of a specific type: a pointed polyhedral cone. By exploring this 'flux cone' we can chart the metabolic capabilities of an organism.

We have seen that an exhaustive charting of these metabolic capabilities is the computation of all *elementary flux modes*: minimal subnetworks that can individually give rise to steady-state flux distributions, and that may be interpreted as minimal metabolic strategies. An especially important use of EFM analysis can be found in the prediction of the effect of gene knockouts: when all EFMs that produce compound Y use reaction r, then the organism cannot make this compound when the gene is knocked out that codes for the enzyme that catalyzes r. And conversely, sometimes gene knockouts can be found such that the cell cannot grow anymore without producing a certain compound of interest. Clearly, these analyses can be very useful for the design of organisms in bio-industry.

On the other hand, we also saw that for large models the computation of all EFMs becomes impossible. There are simply too many of these minimal subnetworks. We presented several alternatives. One could use *elementary conversion modes* if one still desires an exhaustive list of the metabolic capabilities of the cell. The ECMs are easier to enumerate because one can choose to focus only on all possible conversions between (a subset of) the nutrients and products, instead of requiring all information about which reactions are used to get these conversions. For the design of gene knockouts specifically, *minimal cut sets* may be used. Finally, we discussed that the flux cone can be sampled randomly to characterize the flux cone, if this characterization does not need to be exhaustive.

In many cases we have additional information that determines that part of the flux cone is infeasible. For example, some metabolic fluxes may have been measured so that these reaction rates can be fixed to their observed value. In other cases, one may want to use thermodynamic properties to prohibit reactions from occurring that would violate the second law of thermodynamics. These additional constraints can be imposed on top of the mass-balance constraint to further bound the space of feasible flux distributions; each correctly-imposed constraint narrows down the space of feasible fluxes, and thus increases our knowledge of the metabolic state of the cell.

All explorations of the space of feasible flux distributions show one unavoidable conclusion: the metabolic network is incredibly flexible. Even when several constraints are imposed, a genome-scale metabolic model will allow for an almost incomprehensible number of modes in which the metabolic network can function. Consequently, to predict the metabolic state of a cell in more detail we need to make an additional assumption. In the following chapter, we will study what predictions we can make when we assume that the metabolic state is optimized to perform a certain function.

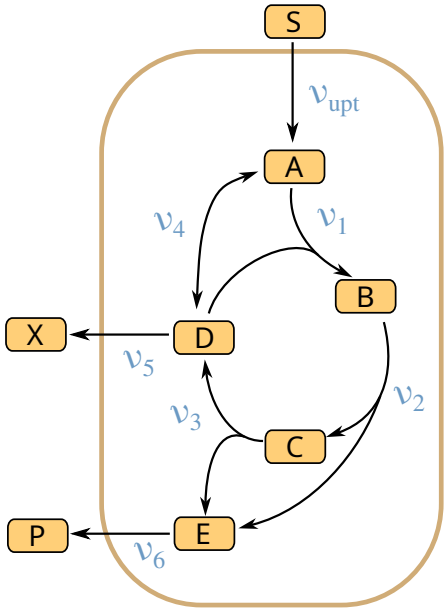


Figure 4.7: *Spirallus insilicus* network, adapted from [38]

Recommended readings

Elementary flux modes and their applications are introduced in an intuitive way in: J. Zanghellini, D. E. Ruckerbauer, M. Hanscho, C. Jungreuthmayer (2013). Elementary flux modes in a nutshell: Properties, calculation and applications. Biotechnology Journal 8 (9), 1009. doi: 10.1002/biot.201200269

Elementary Flux Vectors were introduced as an analog of Elementary Flux Modes in the case that the flux mode is further bound by at least one inhomogeneous constraint. A nice review of these EFVs is can be found in: S. Klamt, G. Regensburger, M. P. Gerstl, C. Jungreuthmayer, S. Schuster, R. Mahadevan, J. Zanghellini, and S. Müller (2017). From elementary flux modes to elementary flux vectors: Metabolic pathway analysis with arbitrary linear flux constraints. PLoS Computational Biology, 13(4):e1005409, doi: 10.1371/journal.pcbi.1005409.

Problems

Computer exercises for this chapter can be found on the book website.

Problem 4.1 A small metabolic network (1)

Spirallus insilicus, a completely fictional organism [39], is characterized by the metabolic network depicted in Figure 4.7 X, S and P represent the biomass, one substrate and one product, while metabolites A to E denote intracellular metabolites. One directional arrows indicate irreversible reactions (all but v_4)

(a) How many intracellular metabolites, intracellular reactions and transport reactions are involved in the model?
(b) Obtain the stoichiometric matrix ($\mathbf{N}$) and the vector of fluxes. How many elements are in the product $\mathbf{N}\ \mathbf{v}$ and what do they represent?
(c) Is the matrix $\mathbf{N}$ of full rank? How many fluxes should be specified to have a unique solution?
(d) Transform the set of constraints so that they define a pointed cone. Determine the number of variables (fluxes) and constraints.

Problem 4.2 A small metabolic network (2)

Consider the following small metabolic network:

$$\mathrm{S}_e \xrightarrow{v_0} \mathrm{S}_c$$
$$\mathrm{S}_c \xrightarrow{v_1} \mathrm{P}_c$$
$$\mathrm{P}_c \xrightarrow{v_2} \mathrm{C}_c$$
$$\mathrm{P}_c \xrightarrow{v_3} \mathrm{D}_c$$
$$\mathrm{P}_c + 2\,\mathrm{C}_c \xrightarrow{v_4} \mathrm{X}$$

Metabolites with a c subscript are located in the cytosol (intracellular) while e stands for extracellular and X represent biomass. All fluxes are positive.

(a) Represent the model as a reaction network (a sketch with metabolites and reactions)
(b) Obtain the stoichiometric matrix ($\mathbf{N}$) and list the variables of the metabolic model ($\mathbf{v}$)
(c) Show that there is no solution to the mass balance equation $\mathbf{N}\ \mathbf{v} = \mathbf{0}$ producing metabolite D. Identify why this is so and modify the model so the production of D is allowed ($v_3 > 0$)

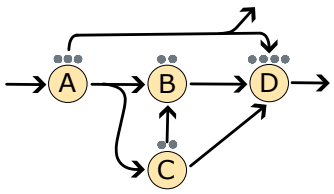
Problem 4.3 Elementary Flux Modes (1)

Assume reaction v_4 is irreversible from A to D in *Spirallus insilicus* (Problem 4.1). Calculate all the Elementary Flux Modes.

(a) By hand.
(b) Using a software of your choice (e.g. pypi.org/project/efmtool/)

Problem 4.4 Elementary Flux Modes (2)

Consider the following metabolic network



$$\mathbf{N} = \begin{pmatrix} 1 & -2 & 0 & 0 & 0 & 0 & -2 \\ 0 & 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & -1 & -2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & -1 & 1 \end{pmatrix}$$

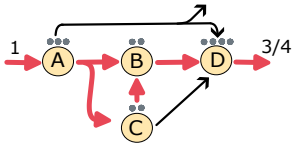
Please note that some stoichiometric coefficients in $\mathbf{N}$ are different from 1 (not shown in the graphics).

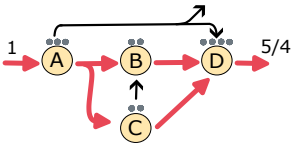
(a) In the network drawing, gray dots denote carbon atoms. Check that carbon atoms are conserved in all reactions. What's the carbon content of the byproduct (not shown) of the reaction from A to D?
(b) All metabolites are treated as internal, that is, they need to be mass-balanced. Find all EFMs (by pure reasoning or by using a software). Determine all EFMs in which all fluxes are in forward direction, i.e. along the conventional directions indicated by arrows.
(c) Which of the EFMs are thermodynamically realizable? Explain why.

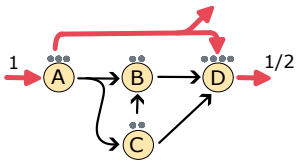
Solutions to problems

Problem 4.4 (Elementary Flux Modes (2))

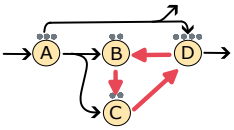
EFMs containing forward fluxes only:

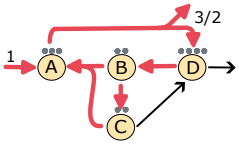


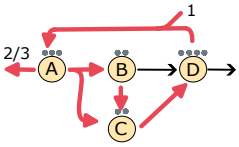


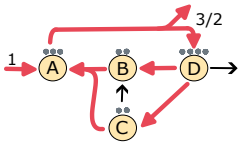


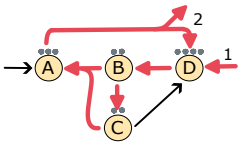
EFMs containing forward and backward fluxes:

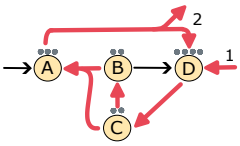


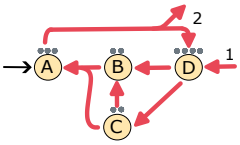


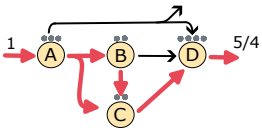












Bibliography

[1] Sebastián N Mendoza, Brett G Olivier, Douwe Molenaar, and Bas Teusink. A systematic assessment of current genome-scale metabolic reconstruction tools. *Genome Biology*, 20(6):158, 2019. doi: https://doi.org/10.1186/s13059-019-1769-1.

[2] Stefan Müller and Georg Regensburger. Elementary vectors and conformal sums in polyhedral geometry and their relevance for metabolic pathway analysis. *Front. Genet.*, 7(90):1–11, 2016. doi: 10.3389/fgene.2016.00090.

[3] Ross Carlson, Aaron Wlaschin, and Friedrich Srienc. Kinetic studies and biochemical pathway analysis of anaerobic poly-(R)-3-hydroxybutyric acid synthesis in Escherichia coli. *Appl Environ Microbiol*, 71(2):713–720, 2005. doi: https://doi.org/10.1128/AEM.71.2.713-720.2005.

[4] Guido Melzer, Manely Eslahpazir Esfandabadi, Ezequiel Franco-Lara, and Christoph Wittmann. Flux design: In silico design of cell factories based on correlation of pathway fluxes to desired properties. *BMC Systems Biology*, 3(1):120, 2009. doi: https://doi.org/10.1186/1752-0509-3-120.

[5] Jörg Stelling, Steffen Klamt, Katja Bettenbrock, Stefan Schuster, and Ernst Dieter Gilles. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420(6912):190–193, 2002. doi: https://doi.org/10.1038/nature01166.

[6] Jörn Behre, Thomas Wilhelm, Axel von Kamp, Eytan Ruppin, and Stefan Schuster. Structural robustness of metabolic networks with respect to multiple knockouts. *Journal of Theoretical Biology*, 252(3):433–441, 2008. doi: https://doi.org/0.1016/j.jtbi.2007.09.043.

[7] Jean-Marc Schwartz, Claire Gaugain, Jose C Nacher, Antoine de Daruvar, and Minoru Kanehisa. Observing metabolic functions at the genome scale. *Genome Biology*, 8(6):R123, 2008. doi: https://doi.org/10.1186/gb-2007-8-6-r123.

[8] J Pey and Francisco J Planes. Direct calculation of elementary flux modes satisfying several biological constraints in genome-scale metabolic networks. *Bioinformatics*, 30(15):2197–2203, 2014. doi: https://doi.org/10.1093/bioinformatics/btu193.

[9] The Economic Cell Collective, editor. *Economic Principles in Cell Biology*. Free online book, 2023. doi: 10.5281/zenodo.8156386.

[10] Matthias P Gerstl, Christian Jungreuthmayer, Stefan Müller, and Jürgen Zanghellini. Which sets of elementary flux modes form thermodynamically feasible flux distributions? *FEBS Journal*, 283(9):1782–1794, 2016. doi: https://doi.org/10.1111/febs.13702.

[11] K Fukuda and A Prodon. *Combinatorics and Computer Science*, chapter Double description method revisited. Springer, 1995.

[12] Marco Terzer and Jörg Stelling. Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics*, 24(19):2229–2235, 2008. doi: https://doi.org/10.1093/bioinformatics/btn401.

[13] T Pfeiffer, I Sánchez-Valdenebro, JC Nuño, F Montero, and S Schuster. METATOOL: for studying metabolic networks. *Bioinformatics*, 15(3):251–257, 1999. doi: https://doi.org/10.1093/bioinformatics/15.3.251.

[14] Steffen Klamt, Jörg Stelling, Martin Ginkel, and Ernst Dieter Gilles. FluxAnalyzer: exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps. *Bioinformatics*, 19(2):261–269, 2003. doi: https://doi.org/10.1093/bioinformatics/19.2.261.

[15] Vicente Acunña, Flavio Chierichetti, Vincent Lacroix, Alberto Marchetti-Spaccamela, Marie-France Sagot, and Leen Stougie. Modes and cuts in metabolic networks: complexity and algorithms. *Biosystems*, 95(1):51–60, 2009. doi: https://doi.org/10.1016/j.biosystems.2008.06.015.

[16] Markus W. Covert and Bernhard O. Palsson. Constraints-based models: Regulation of Gene Expression Reduces the Steady-state Solution Space. *Journal of Theoretical Biology*, 221(3):309–325, 2003. doi: 10.1006/jtbi.2003.3071.

[17] Markus W. Covert, Christophe H. Schilling, and Bernhard O. Palsson. Regulation of Gene Expression in Flux Balance Models of Metabolism. *Journal of Theoretical Biology*, 213(1):73–88, 2001. doi: 10.1006/jtbi.2001.2405.

[18] Christian Jungreuthmayer, David E. Ruckerbauer, and Jürgen Zanghellini. regEfmtool: Speeding up elementary flux mode calculation using transcriptional regulatory rules in the form of three-state logic. *Biosystems*, 113(1):37–39, 2013. ISSN 0303-2647. doi: 10.1016/j.biosystems.2013.04.002.

[19] Sabine Peres, Martin Morterol, and Laurent Simon. SAT-Based Metabolics Pathways Analysis without Compilation. In J.O. Dada P. Mendes and K. Smallbone, editors, *Lecture Note in Bioinformatics*, volume 8859, pages 20–31. Springer International Publishing, 2014. doi: 10.1007/978-3-319-12982-2_2.

[20] Maxime Mahout, Ross P. Carlson, and Sabine Peres. Answer Set Programming for Computing Constraints-Based Elementary Flux Modes: Application to Escherichia coli Core Metabolism. *Processes*, 8(12):1649, 2020. ISSN 2227-9717. doi: 10.3390/pr8121649.

[21] Jeffrey D. Orth, Ronan M. T. Fleming, and Bernhard Ø. Palsson. Reconstruction and use of microbial metabolic networks: The core escherichia coli metabolic model as an educational guide. *EcoSal*, 28:245–248, 2010. doi: 10.1128/ecosal.10.2.1.

[22] Emma Crisci, Maxime Mahout, and Sabine Peres. Computing Thermodynamically Consistent Elementary Flux Modes with Answer Set Programming. In Roberta Gori, Paolo Milazzo, and Mirco Tribastone, editors, *Computational Methods in Systems Biology*, pages 80–88. Springer Nature Switzerland, 2024. ISBN 978-3-031-71671-3. doi: 10.1007/978-3-031-71671-3_7.

[23] Jon Pey and Francisco J. Planes. Direct calculation of elementary flux modes satisfying several biological constraints in genome-scale metabolic networks. *Bioinformatics*, 30(15):2197–2203, 2014. ISSN 1367-4803. doi: 10.1093/bioinformatics/btu193.

[24] Martin Morterol, Philippe Dague, Sabine Peres, and Laurent Simon. Minimality of metabolic flux modes under boolean regulation constraints. In *Workshop on Constraint-Based Methods for Bioinformatics (WCB)*, 2016.

[25] R Urbanczik and C Wagner. Functional stoichiometric analysis of metabolic networks. *Bioinformatics*, 21(22):4176–4180, 2005. doi: https://doi.org/10.1093/bioinformatics/bti674.

[26] Tom J Clement, Erik B Baalhuis, Bas Teusink, Frank J Bruggeman, Robert Planqué, and Daan H de Groot. Unlocking elementary conversion modes: ecmtool unveils all capabilities of metabolic networks. *Patterns*, 2(1):100177, 2020. doi: https://doi.org/10.1016/j.patter.2020.100177.

[27] Nathan D Price, Jan Schellenberger, and Bernhard O Palsson. Uniform sampling of steady-state flux spaces: means to design experiments and to interpret enzymopathies. *Biophysical Journal*, 87(4):21722186, 2004. doi: https://doi.org/10.1529/biophysj.104.043000.

[28] Mattia G Gollub, Hans-Michael Kaltenbach, and Jörg Stelling. Probabilistic thermodynamic analysis of metabolic networks. *Bioinformatics*, 37(18):2938–2945, 2021. doi: https://doi.org/10.1093/bioinformatics/btab194.

[29] Steffen Klamt and Ernst Dieter Gilles. Minimal cut sets in biochemical reaction networks. *Bioinformatics*, 20 (2):226–234, 2004.

[30] Steffen Klamt. Generalized concept of minimal cut sets in biochemical networks. *Biosystems*, 83(2-3):233–247, 2006.

[31] Axel von Kamp and Steffen Klamt. Enumeration of smallest intervention strategies in genome-scale metabolic networks. *PLoS computational biology*, 10(1):e1003378, 2014.

[32] Steffen Klamt, Radhakrishnan Mahadevan, and Axel von Kamp. Speeding up the core algorithm for the dual calculation of minimal cut sets in large metabolic networks. *BMC bioinformatics*, 21(1):1–21, 2020.

[33] Oliver Hädicke and Steffen Klamt. Computing complex metabolic intervention strategies using constrained minimal cut sets. *Metabolic engineering*, 13(2):204–213, 2011.

[34] Kathrin Ballerstein, Axel von Kamp, Steffen Klamt, and Utz-Uwe Haus. Minimal cut sets in a metabolic network are elementary modes in a dual network. *Bioinformatics*, 28(3):381–387, 2012.

[35] David E Ruckerbauer, Christian Jungreuthmayer, and Jürgen Zanghellini. Predicting genetic engineering targets with elementary flux mode analysis: a review of four current methods. *New biotechnology*, 32(6):534–546, 2015.

[36] Christian Jungreuthmayer and Juergen Zanghellini. Designing optimal cell factories: integer programming couples elementary mode analysis with regulation. *BMC systems biology*, 6(1):1–12, 2012.

[37] Christian Jungreuthmayer, Govind Nair, Steffen Klamt, and Juergen Zanghellini. Comparison and improvement of algorithms for computing minimal cut sets. *BMC bioinformatics*, 14(1):1–12, 2013.

[38] Xiao Zhao, Stephan Noack, Wolfgang Wiechert, and Eric Von Lieres. Dynamic flux balance analysis with nonlinear objective function. *Journal of Mathematical Biology*, 75(6-7):1487–1515, 2017. doi: 10.1007/s00285-017-1127-4.

[39] Wolfgang Wiechert and Katharina Nöh. Isotopically non-stationary metabolic flux analysis: complex yet highly informative. *Current Opinion in Biotechnology*, 24(6):979–986, 2013. doi: https://doi.org/10.1016/j.copbio.2013.03.024.