bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Finding MEMo: Minimum sets of elementary flux modes

Annika Röhl[1,2] and Alexander Bockmayr[1,3]

[1]Department of Mathematics and Computer Science, Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany
[2]email: annika.roehl@fu-berlin.de
[3]email: alexander.bockmayr@fu-berlin.de

July 16, 2019

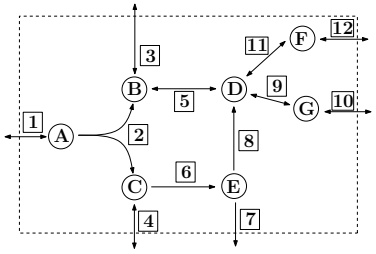
Abstract

Metabolic network reconstructions are widely used in computational systems biology for *in silico* studies of cellular metabolism. A common approach to analyse these models are *elementary flux modes* (EFMs), which correspond to minimal functional units in the network. Already for medium-sized networks, it is often impossible to compute the set of all EFMs, due to their huge number. From a practical point of view, this might also not be necessary because a subset of EFMs may already be sufficient to answer relevant biological questions. In this article, we study MEMos or minimum sets of EFMs that can generate all possible steady-state behaviours of a metabolic network. The number of EFMs in a MEMo may be by several orders of magnitude smaller than the total number of EFMs. Using MEMos, we can compute generating sets of EFMs in metabolic networks where the whole set of EFMs is too large to be enumerated.

1 Introduction

Genome-scale metabolic network reconstructions have been widely used in computational systems biology to develop *in silico* models of cellular metabolism [7, 8, 10, 30, 51]. Typically, a *metabolic network* $\mathcal{N} = (\mathfrak{M}, \mathfrak{R}, S, \mathrm{Irr})$ is given by a set $\mathfrak{M}$ of (internal) metabolites, a set $\mathfrak{R}$ of reactions, a stoichiometric matrix $S \in \mathbb{R}^{\mathfrak{M}\times\mathfrak{R}}$, and a set $\mathrm{Irr} \subseteq \mathfrak{R}$ of irreversible reactions, see Fig. 1 for a small example. The set of reversible reactions is $\mathrm{Rev} = \mathfrak{R} \setminus \mathrm{Irr}$. We always assume

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



$$\begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \end{pmatrix}$$

The corresponding stoichiometric matrix.

Figure 1 Example metabolic network with set of (internal) metabolites $\mathfrak{M} = \{A, B, \ldots, G\}$, and set of reactions $\mathfrak{R} = \{1, 2, \ldots, 11, 12\}$, where $\mathrm{Irr} = \{2, 6, 7, 8\}$ is the set of irreversible reactions.

that the metabolic network is in steady-state, i.e., we analyse the (steady-state) *flux cone*

$$\Gamma_{\mathcal{N}} = \left\{ v \in \mathbb{R}^{\mathfrak{R}} \mid Sv = 0,\ v_{\mathrm{Irr}} \geq 0 \right\}.$$

The elements $v \in \Gamma_{\mathcal{N}}$ are called (feasible) *flux vectors* and can be interpreted as steady-state flux distributions over the network $\mathcal{N}$.

The *support* of a flux vector $v \in \Gamma_{\mathcal{N}}$ is the set $\mathrm{supp}(v) = \{i \in \mathfrak{R} \mid v_i \neq 0\}$ of active reactions in v. An *elementary flux mode* (EFM) [43] is a flux vector $v \in \Gamma_{\mathcal{N}}, v \neq 0$ with inclusion-minimal support. One can show [44] that up to scaling an EFM is uniquely determined by its support. Therefore, two EFMs $e, e' \in \Gamma_{\mathcal{N}}$ are considered to be the same if $\mathrm{supp}(e) = \mathrm{supp}(e')$. We denote by $\mathcal{E}_{\mathcal{N}}$ the finite set of EFMs in the metabolic network $\mathcal{N}$. An EFM $e \in \mathcal{E}_{\mathcal{N}}$ is called *reversible* if $e, -e \in \Gamma_{\mathcal{N}}$, otherwise e is called *irreversible*. We denote the set of reversible resp. irreversible EFMs by $\mathcal{E}_{\mathcal{N}}^{\mathrm{Rev}}$ resp. $\mathcal{E}_{\mathcal{N}}^{\mathrm{Irr}}$.

Example 1.1 (Elementary flux modes) *In the network from Fig. 1 there are 18 EFMs with the following sets of active reactions:* $\{1, 2, 3, 4\}, \{1, 2, 3, 5, 6, 8\}, \{1, 2, 3, 6, 7\}, \{1, 2, 3, 6, 8, 9, 10\}, \{1, 2, 3, 6, 8, 11, 12\}, \{1, 2, 4, 5, 9, 10\}, \{1, 2, 4, 5, 11, 12\}, \{1, 2, 5, 6, 7, 9, 10\}, \{1, 2, 5, 6, 7, 11, 12\}, \{1, 2, 5, 6, 8, 9, 10\}, \{1, 2, 5, 6, 8, 11, 12\}, \{3, 4, 5, 6, 8\}, \{3, 5, 9, 10\}, \{3, 5, 11, 12\}, \{4, 6, 7\}, \{4, 6, 8, 9, 10\}, \{4, 6, 8, 11, 12\}, \{9, 10, 11, 12\}$. *The sets* $\{3, 5, 9, 10\}, \{3, 5, 11, 12\}, \{9, 10, 11, 12\}$ *are the supports of reversible EFMs, which can carry flux in both the forward and the backward direction.*

EFMs are a popular approach to analyse metabolic networks because every steady-state behaviour of the network can be represented with help of the EFMs [42, 43]. Formally speaking, the set $\mathcal{E}_{\mathcal{N}}$ is a *generating set* or *conic basis* of the flux cone $\Gamma_{\mathcal{N}}$. This means that every flux vector $v \in \Gamma_{\mathcal{N}}$ can be represented as a *conical combination* $v = \sum_{e \in \mathcal{E}_{\mathcal{N}}} \lambda_e e$, for some $\lambda_e \geq 0$.

While the EFMs are of great theoretical and practical interest, already for medium-sized metabolic networks it is often not possible to enumerate all of them, since their number grows exponentially with the size of the network. Thus, over the last 20 years, various methods have been developed to compute some or all EFMs in a given metabolic network, see e.g. [6, 9, 12, 13, 32, 33, 35, 48, 52].

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In this paper, we study inclusion-minimal sets of EFMs that generate the full flux cone $\Gamma_{\mathcal{N}}$. We study their mathematical properties and develop an algorithm to compute them. We start with a formal definition.

Definition 1.2 (MEMo) *Let $\mathcal{N} = (\mathfrak{M}, \mathfrak{R}, S, \mathrm{Irr})$ be a metabolic network with flux cone $\Gamma_{\mathcal{N}}$ and set of elementary flux modes $\mathcal{E}_{\mathcal{N}}$. A **M**inimum set of **E**lementary **Mo**eds or* MEMo *is an inclusion-minimal set $\mathcal{U} \subseteq \mathcal{E}_{\mathcal{N}}$ such that every $v \in \Gamma_{\mathcal{N}}$ can be represented as a linear combination*

$$v = \sum_{e \in \mathcal{U} \cap \mathcal{E}_{\mathcal{N}}^{\mathrm{Rev}}} \lambda_e e + \sum_{f \in \mathcal{U} \cap \mathcal{E}_{\mathcal{N}}^{\mathrm{Irr}}} \lambda_f f,$$

for some $\lambda_e, \lambda_f \in \mathbb{R}$, with $\lambda_f \geq 0$, for all irreversible $f \in \mathcal{U} \cap \mathcal{E}_{\mathcal{N}}^{\mathrm{Irr}}$.

A MEMo $\mathcal{U}$ can be used to represent every other flux vector as a conical combination, in particular all EFMs not belonging to $\mathcal{U}$. The minimum number of EFMs needed to represent the whole flux cone is in general much smaller than the number of all EFMs. Therefore, it may be possible to compute a MEMo in a reasonable amount of time even for large networks where the set of all EFMs can not be enumerated.

1.1 Intuition

We begin by providing some intuition on how to compute MEMos using the example in Fig. 1. Mathematically speaking, the flux cone $\Gamma_{\mathcal{N}} = \left\{v \in \mathbb{R}^{\mathfrak{R}} \mid Sv = 0,\ v_{\mathrm{Irr}} \geq 0\right\}$ is a *finitely generated* polyhedral cone. This means that there exist finite sets $\mathcal{G} \subset \Gamma_{\mathcal{N}}$ such that every flux vector $v \in \Gamma_{\mathcal{N}}$ can be obtained as a conical combination $v = \sum_{g \in \mathcal{G}} \lambda_g g, \lambda_g \geq 0$ of the elements of $\mathcal{G}$. If the polyhedral cone is *pointed*, i.e., it does not contain any line $\{\lambda v \mid \lambda \in \mathbb{R}\}$, for some $v \in \Gamma_{\mathcal{N}} \setminus \{0\}$, there exists a unique minimal set of generators. The elements of this set are the *extreme rays* (ERs) of the cone $\Gamma_{\mathcal{N}}$, see Fig. 2 for illustration. In general, flux cones of metabolic networks are *non-pointed* because they contain *reversible flux vectors* $v \neq 0$ for which both $v, -v \in \Gamma_{\mathcal{N}}$. Non-pointed polyhedral cones do not have a unique minimal set of generators, see Fig. 3. The set of reversible flux vectors defines the *lineality space* $\Lambda_{\mathcal{N}} = \{v \in \mathbb{R}^{\mathfrak{R}} \mid Sv = 0,\ v_{\mathrm{Irr}} = 0\}$. Following [26], a reversible reaction is called *fully reversible* if it can carry flux in a reversible flux vector. We denote the set of all fully reversible reactions in $\mathcal{N}$ by Frev. The cone $\Gamma_{\mathcal{N}}$ is pointed if and only if $\Lambda_{\mathcal{N}} = \{0\}$.

Example 1.3 (Reversible flux vectors and fully reversible reactions) *The network in Fig. 1 contains flux vectors that consist of reversible reactions only. For example, $v^{\intercal} = (0, 0, -1, 0, 1, 0, 0, 0, 1, 1, 0, 0)$ is a vector with support $\{3, 5, 9, 10\}$ and can operate in both directions. Thus, the reactions 3,5,9, and 10 are fully reversible and $v^{\intercal}$ is a reversible flux vector. In addition, reactions 11 and 12 of the network in Fig. 1 are fully reversible as well. On the other hand, reaction 4 is reversible, but not fully reversible: Whenever reaction 4 is active at least one irreversible reaction is active as well.*

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

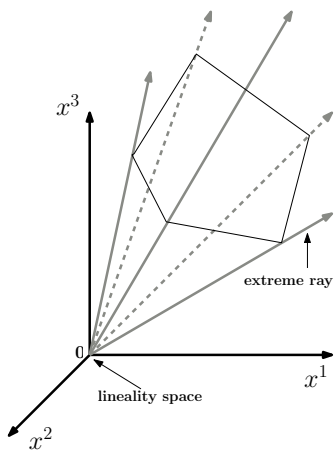


Figure 2 A pointed polyhedral cone in 3-dimensional space with 5 extreme rays. The lineality space is $\{0\}$.

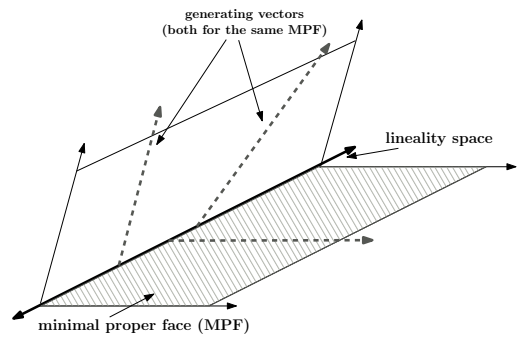


Figure 3 A non-pointed polyhedral cone in 3-dimensional space with a lineality space of dimension 1 and 2 minimal proper faces (MPFs) of dimension 2. The dashed vectors in one of the MPFs illustrate that the set of generating vectors is not unique: one of the two vectors can be used as well as any other vector in the MPF. To represent the whole polyhedral cone, one vector of each MPF is needed and one basis vector of the lineality space.

As mentioned before, the set $\mathcal{E}_{\mathcal{N}}$ of EFMs is a finite generating set of the flux cone $\Gamma_{\mathcal{N}}$. However, $\mathcal{E}_{\mathcal{N}}$ need not be minimal. A well-known method to compute the set of all EFMs consists in reducing the lineality space $\Lambda_{\mathcal{N}}$ to $\{0\}$ by splitting all reversible reactions in $\mathcal{N}$, leading to a new network $\widetilde{\mathcal{N}}$. Splitting a reversible reaction $i \in \mathrm{Rev}$ means that reaction i is replaced by two reactions i^+ and i^-, where i^+ represents the forward and i^- the backward direction of i.

Example 1.4 (Splitting all reversible reactions) *In Fig. 4 all reversible reactions of the network in Fig. 1 are split. After splitting, there exists no reversible flux vector anymore. For example, a flux vector of the original network with support* $\{3, 5, 9, 10\}$ *has now either the support* $\{3^-, 5^+, 9^+, 10^+\}$ *or* $\{3^+, 5^-, 9^-, 10^-\}$ *(the reverse direction).*

The modified flux cone $\Gamma_{\widetilde{\mathcal{N}}}$, where all reversible reactions have been split, is pointed. Thus the extreme rays of $\Gamma_{\widetilde{\mathcal{N}}}$ form a unique minimal set of generators. It can be shown that these extreme rays are in a 1-1 correspondence with the EFMs of the original network $\mathcal{N}$ [5]. However, the set of $\mathcal{E}_{\mathcal{N}}$ of EFMs need not be minimal for $\Gamma_{\mathcal{N}}$.

Example 1.5 (Not all EFMs are needed) *For the network in Fig. 1 the set of reactions* $\{1, 2, 3, 4\}$ *is the support of the EFM* $v^\intercal = (1, 1, 1, 1, 0, 0, 0, 0, 0, 0)$. *The reactions* $\{3, 5, 11, 12\}$ *are the support of an EFM as well* $w^\intercal = (0,\ 0,\ -1,\ 0,\ 1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 1,\ 1)$. *Note that reaction 3 has a positive flux of 1 in* v *and a negative flux of -1 in* w. *The sum of* v *and* w *generates another EFM,* $u^\intercal = (1,$

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

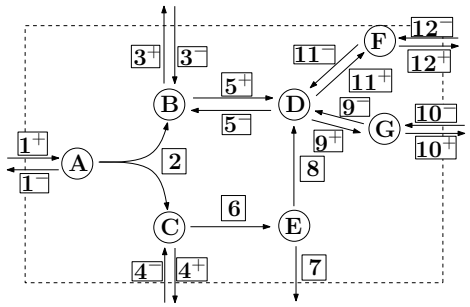


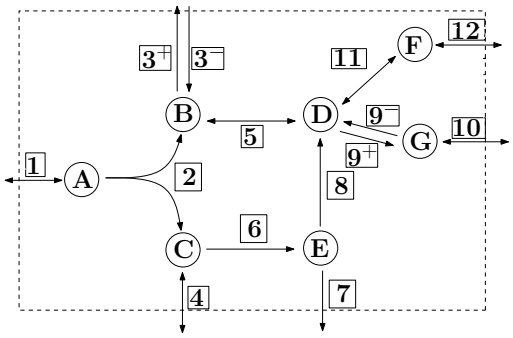
Figure 4 Same metabolic network as in Fig. 1, but with all reversible reactions split. For example, reaction 3 is split into 3^- and 3^+.

1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 1), *consisting of the reactions* $\{1, 2, 4, 5, 11, 12\}$. *Thus, the EFM with the support* $\{1, 2, 4, 5, 11, 12\}$ *is redundant and not required in a minimum generating set of EFMs for the network in Fig. 1.*

In this paper, we present a method for finding a minimum number of reversible reactions such that after splitting these reactions no feasible reversible flux vector exists in the modified network $\widetilde{\mathcal{N}}$. Therefore, the modified flux cone $\Gamma_{\widetilde{\mathcal{N}}}$ is pointed. We will show that the ERs of $\Gamma_{\widetilde{\mathcal{N}}}$ correspond to a MEMo of the original network $\mathcal{N}$, c.f. Def. 1.2.

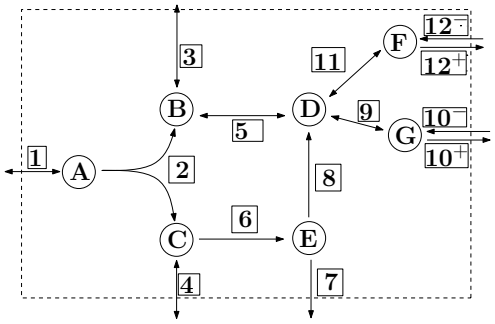
Example 1.6 (Splitting a minimum set of reversible reactions) *For the network in Fig. 1, it is sufficient to split only two reversible reactions, for example* 3 *and 9, see Fig. 5, to obtain a modified flux cone* $\Gamma_{\widetilde{\mathcal{N}}}$ *which is pointed. There are 5 ERs of* $\Gamma_{\widetilde{\mathcal{N}}}$*, which correspond to 5 EFMs in* $\mathcal{N}$*. Thus, only 5 instead of 18 EFMs are needed to generate* $\Gamma_{\mathcal{N}}$*. The minimum set of reversible reactions to split is not unique. Instead of splitting* 3 *and* 9 *as in Fig. 5 one could also split the reactions* 10 *and* 12*, as in Fig. 6. Again, we obtain a MEMo containing 5 different EFMs, which however are different from the 5 EFMs in the MEMo obtained by splitting the reactions* 3 *and* 9*. Finally, not any set of two reversible reactions can be split to obtain a pointed cone. For example, splitting reactions* 11 *and* 12 *will not eliminate all reversible vectors, as shown in Fig. 7.*

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



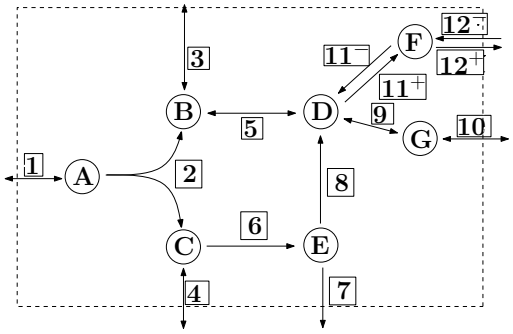
The corresponding flux cone is pointed, with 7 extreme rays with the following supports: $\{1, 2, 4, 5, 11, 12\}, \{3^+, 5, 11, 12\}, \{3^-, 5, 11, 12\}, \{4, 6, 7\}, \{4, 6, 8, 11, 12\}, \{9^+, 10, 11, 12\}, \{9^-, 10, 11, 12\}$. After recombination, because of the 2 reversible EFMs, this results in a MEMo if size 7-2=5, where the contained EFMs have the following supports: $\{1, 2, 4, 5, 11, 12\}, \{3, 5, 11, 12\}, \{4, 6, 7\}, \{4, 6, 8, 11, 12\}, \{9, 10, 11, 12\}$.

Figure 5 The fully reversible reactions 3 and 9 are split, leading to a pointed flux cone for the new network. The transformed network does not contain any reversible flux vectors.



The corresponding flux cone is pointed, with 7 extreme rays with the following supports: $\{1, 2, 3, 4\}, \{3, 4, 5, 6, 8\}, \{3, 5, 9, 10^+\}, \{3, 5, 9, 10^-\}, \{3, 5, 11, 12^+\}, \{3, 5, 11, 12^-\}, \{4, 6, 7\}$. After recombination, because of the 2 reversible EFMs, this results in a MEMo if size 7-2=5, where the contained EFMs have the following supports: $\{1, 2, 3, 4\}, \{3, 4, 5, 6, 8\}, \{3, 5, 9, 10\}, \{3, 5, 11, 12\}, \{4, 6, 7\}$.

Figure 6 The fully reversible reactions 10 and 12 are split, leading to a pointed flux cone for the new network. The transformed network does not contain any reversible flux vectors.



The transformed network still contains reversible flux vectors. Thus, the corresponding cone is not pointed. *Support of the reversible flux vector:* $\{3, 5, 9, 10\}$

Figure 7 The fully reversible reactions 11 and 12 are split.

1.2 Contributions

The contributions of this paper are the following:

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

- We prove that whenever a (sub-)set of reversible reactions is split and the corresponding augmented flux cone is pointed, the extreme rays of the pointed augmented flux cone correspond to EFMs of the original network (Thm. 3.3).
- We introduce a method and provide an implementation for finding a minimum number of reversible reactions such that after splitting these reactions the resulting augmented flux cone is pointed. Additionally, we prove the correctness of the method (Thm. 4.2).
- We prove that the extreme rays of the pointed augmented flux cone (after splitting a minimum set of reversible reactions) correspond to a MEMo, needed to fully describe the underlying metabolic network (Thm. 5.5).
- We apply this method to several metabolic networks and the results show that the number of EFMs in a MEMo is by several orders of magnitude smaller than the number of all EFMs (Sect. 6).

The paper is organised as follows: We start in Sect. 2 with basic facts about polyhedral cones and their connection to metabolic networks. In Sect. 3, we focus on pointed flux cones and formally introduce splitting of reversible reactions. We prove that if splitting results in a pointed polyhedral cone, then the extreme rays of the augmented flux cone correspond to EFMs of the original metabolic network. In Sect. 4 we show how to find a minimum set of reversible reactions to split such that the resulting augmented flux cone is pointed. In Sect. 5 we prove that the extreme rays of the pointed augmented flux cone (after splitting a minimum set of reversible reactions) correspond to a MEMo. Finally, in Sect. 6 we discuss and provide computational results for various genome-scale metabolic networks that were obtained by the software that we implemented. Related and future work are discussed in Sect. 7 and Sect. 8.

A preliminary version of this article appeared as a conference paper in [37].

2 Polyhedral cones

After providing the intuition underlying MEMos, we now formally introduce the method. For this, we start with some mathematical background on polyhedral cones, see [41] for more details.

2.1 Notation

For a vector $x \in \mathbb{R}^n$ we denote by x_i the i-th element of x. For a set H of indices, x_H is the subvector of x corresponding to these indices. We use superscripts to refer to different vectors. For example, $\{x^1, \ldots, x^t\}$ denotes a set of t vectors and we use a set of indices I here as well: $x^I = \{x^i \mid i \in I\}$. If all elements of a vector should be greater or equal to zero we write $x \geq 0$ instead of $x_i \geq 0$ for all i. With $\hat{e}^i \in \mathbb{R}^n$ we denote the i-th unit vector. We denote by $A_{i,*}$ the i-th row and by $A_{*,j}$ the j-th column of the matrix A. With $A_{H,*}$ resp. $A_{*,H}$,

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

we refer to a set of rows, resp. to a set of columns of A. We denote the rank of a matrix $A \in \mathbb{R}^{m\times n}$ with $\rho(A)$, where $\rho(A)$ is the maximum number of linearly independent columns in A. The horizontal resp. vertical concatenation of matrices A, B is denoted by $(A|B)$ resp. $\begin{pmatrix} A \\ B \end{pmatrix}$. Finally $\mathbf{0}_{m,n}$ are zero matrices of the size $m \times n$.

2.2 Polyhedral cones

A subset $\Gamma \subseteq \mathbb{R}^n$ is called a *cone* if any *conical combination* of two elements $x, y \in \Gamma$ belongs to Γ again, i.e., $\lambda x + \mu y \in \Gamma$, for any non-negative $\lambda, \mu \in \mathbb{R}_{\geq 0}$. An element $x \in \Gamma \setminus \{0\}$ is called a *ray* of Γ. A cone Γ is called *polyhedral* if there exists a matrix $A \in \mathbb{R}^{m\times n}$ such that

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

A cone Γ is *finitely generated* if there exists a finite set of generators $\{g^1, \ldots, g^t\} \subseteq \Gamma$ such that every element $x \in \Gamma$ can be written in the form $x = \sum_{k=1}^{t} \lambda_k g^k$, for some non-negative $\lambda_k \in \mathbb{R}_{\geq 0}$. By a classical theorem of Farkas-Minkowski-Weyl, a cone is polyhedral if and only if it is finitely generated. The *dimension* $\dim(\Gamma)$ of a polyhedral cone Γ is the maximum number of affinely independent points of Γ minus 1. For a polyhedral cone $\Gamma = \{x \in \mathbb{R}^n \mid Ax \geq 0\}$, the linear subspace

$$\Lambda = \{x \in \mathbb{R}^n \mid Ax = 0\} = \Gamma \cap (-\Gamma)$$

is called the *lineality space* of Γ. An inequality $a^\intercal x \geq 0,\ a \in \mathbb{R}^n \setminus \{0\}$ is *valid* for Γ if $\Gamma \subseteq \{x \in \mathbb{R}^n \mid a^\intercal x \geq 0\}$. A set $F \subseteq \Gamma$ is called a *face* of Γ, if there exist a valid inequality $a^\intercal x \geq 0$ for Γ such that $F = \Gamma \cap \{x \in \mathbb{R}^n \mid a^\intercal x = 0\}$. F is called a *facet* if $\dim(F) = \dim(\Gamma) - 1$. A face F is called a *minimal proper face* (MPF) if $\dim(F) = t + 1$, where t is the dimension of the lineality space Λ, see Fig. 3 for illustration.

A polyhedral cone Γ is called *pointed* if $\Lambda = \{0\}$. This means that whenever $x \in \Gamma, x \neq 0$, we have $-x \notin \Gamma$. In other words, Γ does not contain any line $\{\lambda x \mid \lambda \in \mathbb{R}\}$, for $x \neq 0$. By basic linear algebra, the cone Γ is pointed if and only if the matrix A has full column rank, i.e., $\rho(A) = n$. Two rays $x, x' \in \Gamma$ are considered identical if $x' = \lambda x$ for some $\lambda > 0$. A ray $x \in \Gamma \setminus \{0\}$ is called an *extreme ray* (ER) of Γ, if there exist no two linearly independent rays $x^1, x^2 \in \Gamma$ such that $x = x^1 + x^2$. If Γ is pointed, the minimal proper faces have dimension one. They form a unique minimum set of generators $\{g^1, \ldots, g^l\} \subseteq \Gamma$ which correspond to the ERs of Γ, see Fig. 2 for illustration.

For a more algebraic characterisation of the ERs of a pointed polyhedral cone we can use the *zero set* $Z(x) = \{i \mid x_i = 0\}$, which denotes the set of indices of entries of a given vector x which are zero. According to [12, 19], a ray x of a pointed polyhedral cone $\Gamma = \{x \in \mathbb{R}^n \mid Ax \geq 0\}$ is an ER if and only if $\rho(A_{Z(Ax),*}) = n - 1$, where $\rho(A) = n$.

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The set of feasible flux vectors in a metabolic network $\mathcal{N} = (\mathfrak{M}, \mathfrak{R}, S, \mathrm{Irr})$ defines a polyhedral cone

$$\Gamma_{\mathcal{N}} = \left\{ v \in \mathbb{R}^{\mathfrak{R}} \mid Sv = 0,\ v_{\mathrm{Irr}} \geq 0 \right\},$$

which is called the *flux cone*. As before, we denote by $\mathrm{supp}(v) = \{i \in \mathfrak{R} \mid v_i \neq 0\}$ the *support of* $v \in \Gamma_{\mathcal{N}}$ and by $\mathrm{supp}_{\mathrm{Irr}}(v) = \{i \in \mathrm{Irr} \mid v_i \neq 0\}$ the *support of* v *in* Irr. An *elementary flux mode* in $\mathcal{N}$ is a flux vector $e \in \Gamma_{\mathcal{N}} \setminus \{0\}$ with inclusion-minimal support. According to the *rank test* [19, 52], a flux vector $v \in \Gamma_{\mathcal{N}}$ is an EFM if and only if $\rho(S_{*,\mathrm{supp}(v)}) = |\mathrm{supp}(v)| - 1$, where $|\mathrm{supp}(v)|$ is the number of non-zero entries of v.

[26] introduced a notion similar to the EFMs, using the set of irreversible reactions Irr. A *minimal metabolic behavior* is an inclusion-minimal non-empty set $D \subseteq \mathrm{Irr}$ for which there exists $v \in \Gamma_{\mathcal{N}}$ such that $\mathrm{supp}_{\mathrm{Irr}}(v) = D$.

Example 2.1 (Minimal metabolic behaviours) *The network in Fig. 1 has three MMBs, namely* $\{2\}, \{6, 7\}$, *and* $\{6, 8\}$. *The minimal metabolic behaviour* $\{6, 8\}$ *is present in 8 EFMs with the supports* $\{1, 2, 3, 5, 6, 8\}, \{1, 2, 3, 6, 8, 9, 10\}, \{1, 2, 3, 6, 8, 11, 12\}, \{1, 2, 5, 6, 8, 9, 10\}, \{1, 2, 5, 6, 8, 11, 12\}, \{3, 4, 5, 6, 8\}, \{4, 6, 8, 9, 10\}, \{4, 6, 8, 11, 12\}$. *The last 3 EFMs, for which* $\mathrm{supp}_{\mathrm{Irr}}(\cdot) = \{6, 8\}$, *are all contained in the minimal proper face associated with the MMB* $\{6, 8\}$, *cf. [26]. The remaining 5 EFMs, for which* $\mathrm{supp}_{\mathrm{Irr}}(\cdot) = \{2, 6, 8\}$, *lie in the interior of the cone* $\Gamma_{\mathcal{N}}$.

The following proposition summarises what is known about the relationship between extreme rays, EFMs, and MMBs.

Proposition 2.2 *Let* $\Gamma_{\mathcal{N}}$ *be the flux cone of a metabolic network* $\mathcal{N}$.

1. *If all reactions are irreversible, then* $\Gamma_{\mathcal{N}}$ *is pointed and the extreme rays of* $\Gamma_{\mathcal{N}}$ *are exactly the EFMs of* $\mathcal{N}$ *[13, 43].*

2. *If there are reversible reactions and* $\Gamma_{\mathcal{N}}$ *is pointed, then the extreme rays of* $\Gamma_{\mathcal{N}}$ *form a subset of the set of EFMs in* $\mathcal{N}$. *However, not every EFM needs to be an extreme ray [19].*
   The extreme rays are exactly the vectors in $\Gamma_{\mathcal{N}}$ *with inclusion-minimal support in* Irr.

3. *If* $\Gamma_{\mathcal{N}}$ *is not pointed, then there is a 1-1 correspondence between the MMBs of* $\mathcal{N}$ *and the minimal proper faces of* $\Gamma_{\mathcal{N}}$ *[26].*

Proof: We have only to prove the last part of 2). By [26] (or using 3)) there is a 1-1 correspondence between the minimal proper faces of a flux cone and the MMBs. All vectors in a minimal proper face F have the same support in Irr, which is exactly the MMB corresponding to F. If a polyhedral cone is pointed, the minimal proper faces are the extreme rays. □

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In general, a metabolic network may contain reversible reactions together with reversible flux vectors $v \neq 0$, for which $v, -v \in \Gamma_{\mathcal{N}}$, and the flux cone is non-pointed. For example, in Fig. 1, there is a reversible flux vector with the support $\{3, 5, 9, 10\}$. If there exist reversible flux vectors, a minimal generating set of the flux cone is not unique anymore. It may consist of an arbitrary flux vector for each minimal proper face of the flux cone, together with a vector space basis of the lineality space, see [41, Sect. 8.8]. Computing a minimal generating set by one of the standard software tools, usually does not deliver a set of EFMs. In order to obtain a minimal generating set consisting of EFMs, one can transform the flux cone into a pointed cone. In the next section, we explain how this can be done.

3 Splitting reversible reactions

In many cases it is desirable to have a cone Γ which is pointed. If $x \geq 0$ for all $x \in \Gamma$, then the resulting cone will be pointed. In general, however, a variable $x_i \in \mathbb{R}$ can take negative values and the constraint $x_i \geq 0$ does not hold. To overcome this problem, a well-known method also used in linear programming is to *split* variables. Splitting a variable $x_i \in \mathbb{R}$ means replacing it by two non-negative variables $\widetilde{x}_i, \widetilde{x}_{n+1} \geq 0$, such that $x_i = \widetilde{x}_i - \widetilde{x}_{n+1}$. Note that this will change the structure of the cone and increase the dimension of the underlying vector space by 1. To describe this transformation formally and for several variables, we use a map $\pi_I : \mathbb{R}^n \to \mathbb{R}^{n+|I|}$, where $I = \{i_1, \ldots, i_s\}$ denotes the set of variables $x_{i_k}, k \in \{1, \ldots s\}$ to be split. For $x \in \mathbb{R}^n$ we get $\pi_I(x) = \widetilde{x}$ with $\widetilde{x}_j = x_j$, for all $j \in \{1, \ldots, n\} \setminus I$, and for each $i_k \in I$:

$$\begin{array}{llll} \widetilde{x}_{i_k} = x_{i_k} & \text{and} & \widetilde{x}_{n+k} = 0 & \text{if } x_{i_k} \geq 0, \\ \widetilde{x}_{i_k} = 0 & \text{and} & \widetilde{x}_{n+k} = -x_{i_k} & \text{if } x_{i_k} < 0. \end{array}$$

Using the map π_I, a polyhedral cone $\Gamma = \{x \in \mathbb{R}^n \mid Ax \geq 0\}$, with $A \in \mathbb{R}^{m \times n}$, is mapped to the *augmented polyhedral cone*

$$\widetilde{\Gamma}^I \quad = \quad \{\widetilde{x} = (x, y) \in \mathbb{R}^n \times \mathbb{R}^{|I|} \mid Ax - A_{*,I}\, y \geq 0,\ x_I \geq 0,\ y \geq 0\}.$$

The inverse transformation $\pi_I^r : \widetilde{\Gamma}^I \to \Gamma$ maps each vector $\widetilde{x} \in \widetilde{\Gamma}^I$ to $x = \pi_I^r(\widetilde{x})$ such that $x_j = \widetilde{x}_j$, for all $j \in \{1, \ldots, n\} \setminus I$, and $x_{i_k} = \widetilde{x}_{i_k} - \widetilde{x}_{n+k}$, for all $i_k \in I, k \in \{1, \ldots, s\}$. If we apply π_I^r, we say that we *recombine* the variables that were split before.

We now turn to the lineality space of $\widetilde{\Gamma}^I$. Generalising a result from [25], we get:

Proposition 3.1 *Let $\Gamma \subseteq \mathbb{R}^n$ be a polyhedral cone with lineality space Λ. For a set of variables $I \subseteq \{1, \ldots, n\}$, the lineality space $\widetilde{\Lambda}^I$ of the augmented polyhedral cone $\widetilde{\Gamma}^I$ is given by*

$$\widetilde{\Lambda}^I = \left\{ \begin{pmatrix} x \\ 0 \end{pmatrix} \in \mathbb{R}^{n+|I|} \mid x \in \Lambda \text{ and } x_I = 0 \right\}.$$

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Proof: For the cone $\Gamma = \{x \in \mathbb{R}^n \mid Ax \geq 0\}$ the lineality space is $\Lambda = \{x \in \mathbb{R}^n \mid Ax = 0\}$. Splitting the variables in I delivers the cone $\widetilde{\Gamma}^I = \{ \begin{pmatrix} x \\ w \end{pmatrix} \in \mathbb{R}^{n+|I|} \mid (A \mid -A_{*,I}) \begin{pmatrix} x \\ w \end{pmatrix} \geq 0,\ x_I \geq 0,\ w \geq 0\}$ For the lineality space we get:

$$\begin{aligned}
\widetilde{\Lambda}^I &= \left\{ \begin{pmatrix} x \\ w \end{pmatrix} \in \mathbb{R}^{n+|I|} \mid (A \mid -A_{*,I}) \begin{pmatrix} x \\ w \end{pmatrix} = 0,\ x_I = 0,\ w = 0 \right\} \\
&= \left\{ \begin{pmatrix} x \\ 0 \end{pmatrix} \in \mathbb{R}^{n+|I|} \mid (A \mid -A_{*,I}) \begin{pmatrix} x \\ 0 \end{pmatrix} = 0,\ x_I = 0 \right\} \\
&= \left\{ \begin{pmatrix} x \\ 0 \end{pmatrix} \in \mathbb{R}^{n+|I|} \mid Ax = 0,\ x_I = 0 \right\} \\
&= \left\{ \begin{pmatrix} x \\ 0 \end{pmatrix} \in \mathbb{R}^{n+|I|} \mid x \in \Lambda \text{ and } x_I = 0 \right\}
\end{aligned}$$

□

Corollary 3.2 *$\widetilde{\Gamma}^I$ is pointed if and only if $\rho(A_{*,\{1,\ldots,n\}\setminus I}) = n - |I|$.*

In the case of a metabolic network $\mathcal{N}$, the variables corresponding to the irreversible reactions can take only non-negative values. In order to obtain a pointed cone, one may split all reversible reactions into two irreversible ones, see Fig. 4 and Example 1.4. This leads to the pointed augmented flux cone $\widetilde{\Gamma}^{\text{Rev}}_{\mathcal{N}}$. The uniquely determined extreme rays of $\widetilde{\Gamma}^{\text{Rev}}_{\mathcal{N}}$ are called *extreme currents* [5]. It has been shown in [13] that after recombination and up to so-called 2-cycles, the extreme currents correspond exactly to the EFMs of the metabolic network $\mathcal{N}$. For each split reaction $i_k \in \text{Rev}$, the corresponding *2-cycle* is the extreme ray $\widetilde{v} \in \widetilde{\Gamma}^{\text{Rev}}_{\mathcal{N}}$ with $\widetilde{v}_{i_k} = \widetilde{v}_{n+k}$, and $v_i = 0$ otherwise. These cycles do not have a biological meaning and can be eliminated (they become zero after applying π^r_{Rev}).

Splitting all reversible reactions may highly increase the number of variables and thus the dimension of the vector space where the augmented flux cone lives in. Already for medium-sized networks, the number of extreme rays of $\widetilde{\Gamma}^{\text{Rev}}_{\mathcal{N}}$, which corresponds to the number of EFMs (up to the 2-cycles) of $\mathcal{N}$, will be huge and therefore computing the whole set may not be feasible or desirable. Already in [40], it has been shown that splitting only the *internal* reversible reactions delivers an augmented flux cone which is pointed (assuming that there is at most one exchange reaction per internal metabolite). After recombination, the extreme rays of this cone are called *extreme pathways* (EPs). However, the number of EPs can be very large and usually the set of all EPs cannot be computed for genome-scale metabolic networks, see Sect. 7.1 for further discussion.

Extending the results in [13], we show next that if splitting a subset $I \subseteq \text{Rev}$ of reversible reactions results in a pointed augmented flux cone $\widetilde{\Gamma}^I_{\mathcal{N}}$ then the extreme rays of $\widetilde{\Gamma}^I_{\mathcal{N}}$ after reconfiguration define a subset of EFMs in the metabolic network $\mathcal{N}$.

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Theorem 3.3 *Let $\Gamma_{\mathcal{N}} = \left\{v \in \mathbb{R}^{\mathfrak{R}} \mid Sv = 0,\ v_{\mathrm{Irr}} \geq 0\right\}$ be the flux cone of a metabolic network $\mathcal{N}$. Let $I \subseteq \mathrm{Rev}$ be a set of reversible reactions such that the augmented cone $\widetilde{\Gamma}^I_{\mathcal{N}}$ obtained by splitting the reactions in I is pointed. With exception of the 2-cycles, the extreme rays of $\widetilde{\Gamma}^I_{\mathcal{N}}$ after recombination are elementary flux modes in $\mathcal{N}$.*

Proof: By splitting the reactions in I we get the metabolic network $\widetilde{\mathcal{N}} = (\mathfrak{M}, \widetilde{\mathfrak{R}}, \widetilde{S}, \widetilde{\mathrm{Irr}})$ with $\widetilde{\mathfrak{R}} = \mathfrak{R} \cup I_{\mathrm{split}}$, $\widetilde{S} = (S \mid -S_{*,I})$ and $\widetilde{\mathrm{Irr}} = \mathrm{Irr} \cup I \cup I_{\mathrm{split}}$. Here, I_{split} is the set of additional irreversible reactions obtained by splitting the reactions in I. The corresponding flux cone is

$$\Gamma_{\widetilde{\mathcal{N}}} = \widetilde{\Gamma}^I_{\mathcal{N}} = \{\begin{pmatrix} v \\ w \end{pmatrix} \in \mathbb{R}^{|\mathfrak{R}|+|I|} \mid \widetilde{S}\begin{pmatrix} v \\ w \end{pmatrix} = 0,\ v_{\mathrm{Irr}} \geq 0,\ v_I \geq 0,\ w \geq 0\}.$$

Let $\widetilde{e}$ being an ER of $\Gamma_{\widetilde{\mathcal{N}}}$ and $e = \pi^r_I(\widetilde{e})$ the corresponding flux vector in $\Gamma_{\mathcal{N}}$. By Prop. 2.2, the ERs of the pointed cone $\Gamma_{\widetilde{\mathcal{N}}}$ are EFMs in $\widetilde{\mathcal{N}}$. By the rank test for EFMs, this implies $\rho(\widetilde{S}_{*,\mathrm{supp}(\widetilde{e})}) = |\mathrm{supp}(\widetilde{e})| - 1$.

Suppose first that for all $i_k \in I$ we have $\widetilde{e}_{i_k} = 0$ or $\widetilde{e}_{n+k} = 0$, i.e., there is no non-zero flux through both directions of a former reversible reaction. Then $|\mathrm{supp}(\widetilde{e})| = |\mathrm{supp}(e)|$ and $\rho(\widetilde{S}_{*,\mathrm{supp}(\widetilde{e})}) = \rho(S_{*,\mathrm{supp}(e)})$ (by the definition of $\widetilde{S} = (S \mid -S_{*,I})$). It follows that $\rho(S_{*,\mathrm{supp}(e)}) = |\mathrm{supp}(e)| - 1$. Using again the rank test, this shows that e is an EFM in $\Gamma_{\mathcal{N}}$.

Suppose now that there exists $i_k \in I$ with $\widetilde{e}_{i_k} \neq 0$ and $\widetilde{e}_{n+k} \neq 0$. Since $\widetilde{e}$ is an EFM of $\widetilde{\mathcal{N}}$, it follows $\widetilde{e}_{i_k} = \widetilde{e}_{n+i}$ and $\widetilde{e}_j = 0$, for all $j \in \widetilde{\mathfrak{R}} \setminus \{i_k, n+k\}$. Hence, $\widetilde{e}$ is a 2-cycle, which we excluded from our considerations. □

Thus, whenever the augmented flux cone $\widetilde{\Gamma}^I_{\mathcal{N}}$ is pointed, the recombined extreme rays of $\widetilde{\Gamma}^I_{\mathcal{N}}$ correspond to a subset of EFMs in the original metabolic network $\mathcal{N}$. The remaining EFMs of $\mathcal{N}$ correspond to rays inside of $\widetilde{\Gamma}^I_{\mathcal{N}}$ and can be obtained from the extreme rays by conical combinations. In the following, we develop a method for determining a minimum set of reversible reactions that have to be split in order to obtain a pointed cone.

4 Splitting a minimum number of reversible reactions

We show in this section that there exist minimum sets of reversible reactions such that, after splitting these reactions, the augmented flux cone becomes pointed. These sets always contain t fully reversible reactions, where t is the dimension of the lineality space. We note that splitting one reaction reduces the dimension of the lineality space by at most 1 [25]. Therefore at least t reversible reactions have to be split in order to obtain a pointed cone.

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$B^{\text{rcef}} = \begin{array}{c} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ * & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ * & * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ * & * & * & * & 0 & 0 \end{pmatrix} \end{array} \begin{array}{l} \leftarrow j_1 \\ \\ \leftarrow j_2 \\ \leftarrow j_3 \\ \vdots \\ \\ \leftarrow j_r \\ \\ \end{array}$$

Figure 8 B^{rcef} is a matrix in reduced column echelon form, where $*$ are appropriate values in $\mathbb{R}$.

Definition 4.1 (minFrev) *Let $\Gamma_{\mathcal{N}}$ be the flux cone of a metabolic network $\mathcal{N}$. Let t be the dimension of the lineality space $\Lambda_{\mathcal{N}}$. A* minimum set of fully reversible reactions *is a set* minFrev *of t fully reversible reactions such that the flux cone $\widetilde{\Gamma}_{\mathcal{N}}^{\text{minFrev}}$ obtained by splitting the reactions in* minFrev *is pointed. We denote by* $\text{minFrev}_{\text{split}}$ *the set of the t additional irreversible reactions obtained by splitting. Thus*

$$\widetilde{\Gamma}_{\mathcal{N}}^{\text{minFrev}} = \{\widetilde{v} \in \mathbb{R}^{|\mathfrak{R}|+t} \mid (S \mid -S_{*,\text{minFrev}})\widetilde{v} = 0, \widetilde{v}_{\text{Irr} \cup \text{minFrev} \cup \text{minFrev}_{\text{split}}} \geq 0\}.$$

The next theorem states that for any metabolic network $\mathcal{N}$ there exists a minimum set of fully reversible reactions minFrev in the sense of Def. 4.1. To find such a set, we use a well-known concept from matrix theory [14, p.85]. A matrix $B^{\text{rcef}} \in \mathbb{R}^{m \times n}$ is in *reduced column echelon form* if the following properties hold (see Fig. 8):

1. The first non-zero element in column k is a 1 in row j_k, for $k = 1, 2, \ldots, r$ (this 1 is called a *pivot*).
2. $1 \leq j_1 < j_2 < \cdots < j_r \leq m$ (i.e., for each change in columns from left to right, the pivot appears in a lower row).
3. For $k = 1, \ldots, r$, the pivot in column k is the only non-zero element in row j_k.
4. Each of the last $n - r$ columns consists entirely of zeros.

Theorem 4.2 *Let $\Gamma_{\mathcal{N}}$ be the flux cone of a metabolic network $\mathcal{N}$. If t is the dimension of the lineality space $\Lambda_{\mathcal{N}}$, then there exists a set* minFrev *of t fully reversible reactions such that the cone $\widetilde{\Gamma}_{\mathcal{N}}^{\text{minFrev}}$ obtained by splitting the reactions in* minFrev *is pointed.*

Proof: The idea of the proof is the following. As a vector space, the lineality space $\Lambda_{\mathcal{N}}$ has a basis consisting of t linearly independent vectors. In order to obtain a pointed flux cone, we intuitively have to destroy $\Lambda_{\mathcal{N}}$, i.e., all vectors in the basis. We do this by splitting one reaction in each vector of the basis. To

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

make sure that each of these reactions is present in exactly one basis vector, we make use of the reduced column echelon form, cf. Fig. 8.

We may assume that $\Gamma_{\mathcal{N}}$ is not pointed, hence $\Lambda_{\mathcal{N}} \neq \{0\}$. Let $B \in \mathbb{R}^{|\mathfrak{R}| \times t}$ be a matrix whose columns $(\check{b}^1, \ldots, \check{b}^t)$ form a basis of the lineality space $\Lambda_{\mathcal{N}}$, where $t \geq 1$ is the dimension of $\Lambda_{\mathcal{N}}$. By applying elementary column operations, we can obtain the reduced column echelon form B^{rcef} of B, see Fig. 8, which is uniquely determined by B [14, p.85]. The columns $(b^1, \ldots, b^t)$ of B^{rcef} define again a basis of $\Lambda_{\mathcal{N}}$. The row indices $j_1, \ldots, j_r$ for the pivot 1's in B^{rcef} are exactly the indices of the fully reversible reactions we are looking for. To see this, define $\text{minFrev} = \{j_1, \ldots, j_r\}$. From $t = \rho(B) = \rho(B^{\text{rcef}}) = r$, we get $r = t$. Since $b^k_{j_k} = 1 \neq 0$, for $k = 1, \ldots, r = t$, we have $\text{minFrev} \subseteq \text{Frev}$.

After splitting the reactions in minFrev we get the augmented flux cone $\widetilde{\Gamma}^{\text{minFrev}}_{\mathcal{N}}$, which by Prop. 3.1 has the lineality space

$$\widetilde{\Lambda}^{\text{minFrev}}_{\mathcal{N}} = \left\{ \begin{pmatrix} v \\ 0 \end{pmatrix} \in \mathbb{R}^{|\mathfrak{R}|+|\text{minFrev}|} \mid v \in \Lambda_{\mathcal{N}} \text{ and } v_{\text{minFrev}} = 0 \right\}.$$

Since $(b^1, \ldots, b^t)$ defines a basis of $\Lambda_{\mathcal{N}}$, any $v \in \Lambda_{\mathcal{N}}$ can be written as $v = \sum_{k=1}^{t} \lambda_k b^k$, for some $\lambda_k \in \mathbb{R}$. The matrix B^{rcef} is in reduced column-echelon form. This means that for each $j \in \text{minFrev}$ there is exactly one b^{k_j} with $b^{k_j}_j = 1$, and for all $k \neq k_j$, it holds $b^k_j = 0$. Since for $\begin{pmatrix} v \\ 0 \end{pmatrix} \in \widetilde{\Lambda}^{\text{minFrev}}_{\mathcal{N}}$ we have $v_j = \sum_{k=1}^{t} \lambda_k b^k_j = \lambda_{k_j} b^{k_j}_j = \lambda_{k_j} = 0$, for all $j \in \text{minFrev}$, it follows that $\lambda_k = 0$, for all $k \in \{1, \ldots, t\}$. This implies $v = 0$ and thus $\widetilde{\Lambda}^{\text{minFrev}}_{\mathcal{N}} = \{0\}$, which proves that $\widetilde{\Gamma}^{\text{minFrev}}_{\mathcal{N}}$ is pointed. □

Based on the previous proof, Algorithm 1 summarizes how to find a minimum set minFrev such that $\widetilde{\Gamma}^{\text{minFrev}}_{\mathcal{N}}$ is pointed. First we compute a basis of the lineality space $\Lambda_{\mathcal{N}}$ of $\Gamma_{\mathcal{N}}$. Let B be the matrix whose columns are the vectors of this basis. Next we transform B in the reduced column echelon form (e.g. in `MATLAB` using `rref`). The row indices of the pivots then correspond to the indices of the reactions that form the set minFrev.

Algorithm 1: Finding a minimum set of fully reversible reactions

Input : Stoichiometric matrix S
Set of irreversible reactions Irr
Output: Minimum set of fully reversible reactions minFrev

1 $B \leftarrow \text{basis}(\Lambda_{\mathcal{N}})$;
`/* columns of` B `are vectors of the basis of the lineality space` $\Lambda_{\mathcal{N}}$ `*/`
2 $B^{\text{rcef}} \leftarrow \text{ReducedColumnEchelonForm}(B)$;
3 minFrev $\leftarrow$ row indices of the pivots of B^{rcef};
4 **return** minFrev;

We note that the set minFrev computed by Algorithm 1 does not depend on

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the particular basis B of $\Lambda_{\mathcal{N}}$ because two matrices B, B' with the same column space have the same reduced column-echelon form [14].

Example 4.3 (Compute minFrev**)** *The flux cone for the network in Fig. 1 has a lineality space of dimension 2. Therefore any set* minFrev *contains two reactions. The lineality space can be described by the matrix*

$$B^{\intercal} = \begin{array}{c} \begin{array}{cccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \end{array} \\ \begin{pmatrix} 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & -1 & -1 \end{pmatrix} \end{array}.$$

We can see that only the fully reversible reactions $3, 5, 9, 10, 11$ *and 12 are active in the lineality space. Furthermore, B is already in reduced column echelon form and thus $B = B^{\mathrm{rcef}}$. The pivot elements are 3 and 9, therefore* $\mathrm{minFrev} = \{3, 9\}$. *These are the reactions which are split in Example 1.6. Splitting reactions 3 and 9 delivers an augmented flux cone which is pointed. This can be seen from the modified network in Fig. 5, since there exists no feasible flux vector that is reversible.*

5 MEMo: Minimum set of Elementary Modes

The main result of this section is that the extreme rays of an augmented flux cone $\widetilde{\Gamma}_{\mathcal{N}}^{\mathrm{minFrev}}$, where minFrev is a minimum set of fully reversible reactions, after recombination form a Minimum set of Elementary Modes or MEMo, cf. Def. 1.2. We start by using Prop. 2.2 to prove the following result.

Proposition 5.1 *Let* minFrev *be a minimum set of fully reversible reactions in a metabolic network $\mathcal{N}$ such that $\widetilde{\Gamma}_{\mathcal{N}}^{\mathrm{minFrev}}$ is pointed. Then the set of extreme rays of the augmented cone $\widetilde{\Gamma}_{\mathcal{N}}^{\mathrm{minFrev}}$ does not contain any 2-cycle.*

Proof: By the definition of the lineality space $\Lambda_{\mathcal{N}}$ and the set Frev, we have $\Lambda_{\mathcal{N}} = \{v \in \mathbb{R}^{\mathfrak{R}} \mid Sv = 0, v_{\mathrm{Irr}} = 0\} = \{v \in \mathbb{R}^{\mathfrak{R}} \mid S_{*,\mathrm{Rev}}\, v_{\mathrm{Rev}} = 0, v_{\mathrm{Irr}} = 0\} = \{v \in \mathbb{R}^{\mathfrak{R}} \mid S_{*,\mathrm{Frev}}\, v_{\mathrm{Frev}} = 0, v_{\mathrm{Irr}\cup(\mathrm{Rev}\setminus\mathrm{Frev})} = 0\}$. From $\dim(\Lambda_{\mathcal{N}}) = t$, we get $\rho(S_{*,\mathrm{Frev}}) = |\mathrm{Frev}| - t$.

By Prop. 3.1, it follows that $\widetilde{\Lambda}_{\mathcal{N}}^{\mathrm{minFrev}} = \{(v \mid 0)^{\intercal} \in \mathbb{R}^{|\mathfrak{R}|+|\mathrm{minFrev}|} \mid S_{*,\mathrm{Frev}\setminus\mathrm{minFrev}}\, v_{\mathrm{Frev}\setminus\mathrm{minFrev}} = 0, v_{\mathrm{Irr}\cup(\mathrm{Rev}\setminus\mathrm{Frev})\cup\mathrm{minFrev}} = 0\}$. Since $\widetilde{\Gamma}_{\mathcal{N}}^{\mathrm{minFrev}}$ is pointed, this implies $\rho(S_{*,\mathrm{Frev}\setminus\mathrm{minFrev}}) = |\mathrm{Frev}| - |\,\mathrm{minFrev}\,| = |\mathrm{Frev}| - t = \rho(S_{*,\mathrm{Frev}})$. For each $i_k \in \mathrm{minFrev}$, the column S_{*,i_k} is therefore linearly dependent on the columns of $S_{*,\mathrm{Frev}\setminus\mathrm{minFrev}}$. Thus, there exists a flux vector $v \in \Gamma_{\mathcal{N}} \setminus \{0\}$ with $v_{i_k} = 1$ and $\mathrm{supp}(v) \subseteq (\mathrm{Frev} \setminus \mathrm{minFrev}) \cup \{i_k\}$. It follows that $\pi_{\mathrm{minFrev}}(v) = \widetilde{v} \in \widetilde{\Gamma}_{\mathcal{N}}^{\mathrm{minFrev}}$ and $\mathrm{supp}_{\widetilde{Irr}}(\widetilde{v}) = \{i_k\}$, with $\widetilde{Irr} = \mathrm{Irr} \cup \mathrm{minFrev} \cup \mathrm{minFrev}_{\mathrm{split}}$. This implies that $\widetilde{v}$ has minimal support w.r.t irreversible reactions in $\widetilde{\Gamma}_{\mathcal{N}}^{\mathrm{minFrev}}$. According to Prop. 2.2, $\widetilde{v}$ is an ER of $\widetilde{\Gamma}_{\mathcal{N}}^{\mathrm{minFrev}}$.

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Suppose there exists an ER $\widetilde{w} \in \widetilde{\Gamma}_{\mathcal{N}}^{\text{minFrev}}$ with $\widetilde{w}_{i_k} \neq 0 \neq \widetilde{w}_{n+k}$. This would imply that $\text{supp}_{\widetilde{Irr}}(\widetilde{v}) \subsetneq \text{supp}_{\widetilde{Irr}}(\widetilde{w})$, which is a contradiction. It follows that the set of ERs of $\widetilde{\Gamma}_{\mathcal{N}}^{\text{minFrev}}$ does not contain any 2-cycle. □

From the proof of Prop. 5.1 we can deduce the following corollaries:

Corollary 5.2 *Let $\widetilde{e}$ be an extreme ray in $\widetilde{\Gamma}_{\mathcal{N}}^{\text{minFrev}}$. Then there is at most one reaction $j \in \text{minFrev} \cup \text{minFrev}_{\text{split}}$ with $\widetilde{e}_j \neq 0$. If $\widetilde{e}_j \neq 0$, for some $j \in \text{minFrev} \cup \text{minFrev}_{\text{split}}$, then $\widetilde{e}_i = 0$, for all $i \in \text{Irr}$. Moreover, $\pi^r_{\text{minFrev}}(\widetilde{e})$ is a reversible EFM of the original network $\mathcal{N}$.*

Corollary 5.3 *For each $i_k \in \text{minFrev}$ there exist exactly two extreme rays $\widetilde{e}, \widetilde{f} \in \widetilde{\Gamma}_{\mathcal{N}}^{\text{minFrev}}$ such that $\widetilde{e}_{i_k} \neq 0$ and $\widetilde{f}_{n+k} \neq 0$. After recombination, both $\widetilde{e}$ and $\widetilde{f}$ define a single reversible EFM e in $\Gamma_{\mathcal{N}}$ with $\text{supp}(e) = \text{supp}(\pi^r_{\text{minFrev}}(\widetilde{e})) = \text{supp}(\pi^r_{\text{minFrev}}(\widetilde{f}))$. Altogether, the t reactions in minFrev define $2t$ extreme rays in $\widetilde{\Gamma}_{\mathcal{N}}^{\text{minFrev}}$ and t linearly independent reversible EFMs in $\mathcal{N}$.*

The following proposition relates the supports of the extreme rays of $\widetilde{\Gamma}_{\mathcal{N}}^{\text{minFrev}}$ to the minimal metabolic behaviors (MMBs) of $\mathcal{N}$.

Proposition 5.4 *Let minFrev be a minimum set of fully reversible reactions in a metabolic network $\mathcal{N}$ such that $\widetilde{\Gamma}_{\mathcal{N}}^{\text{minFrev}}$ is pointed. The non-empty supports $\text{supp}_{\text{Irr}}(\widetilde{e})$ of the extreme rays $\widetilde{e}$ of $\widetilde{\Gamma}_{\mathcal{N}}^{\text{minFrev}}$ are exactly the minimal metabolic behaviors of $\mathcal{N}$.*

Proof: Let D be an MMB in $\mathcal{N}$ and let $v \in \Gamma_{\mathcal{N}}$ such that $\text{supp}_{\text{Irr}}(v) = D$. By Lemma 2 in [18], there is a decomposition $v = e + y$, with $e, y \in \Gamma_{\mathcal{N}}, e_{\text{minFrev}} = 0$ and $y_{\text{Irr}} = 0$. For $\widetilde{e} = \pi_{\text{minFrev}}(e)$ and $\widetilde{Irr} = \text{Irr} \cup \text{minFrev} \cup \text{minFrev}_{\text{split}}$ it follows that $\text{supp}_{\widetilde{Irr}}(\widetilde{e}) = \text{supp}_{\text{Irr}}(\widetilde{e}) = \text{supp}_{\text{Irr}}(e) = \text{supp}_{\text{Irr}}(v) = D$. Suppose $\widetilde{e}$ is not an extreme ray in $\widetilde{\Gamma}_{\mathcal{N}}^{\text{minFrev}}$. By Prop. 2.2, there exists $\widetilde{w} \in \widetilde{\Gamma}_{\mathcal{N}}^{\text{minFrev}}$ with $\emptyset \neq \text{supp}_{\widetilde{Irr}}(\widetilde{w}) \subsetneq \text{supp}_{\widetilde{Irr}}(\widetilde{e})$. For $w = \pi^r_{\text{minFrev}}(\widetilde{w}) \in \Gamma_{\mathcal{N}}$ this implies $\text{supp}_{\text{Irr}}(w) = \text{supp}_{\widetilde{Irr}}(\widetilde{w}) \subsetneq D$, in contradiction to D being an MMB in Γ_N.

Conversely, let $\widetilde{e}$ be an extreme ray in $\widetilde{\Gamma}_{\mathcal{N}}^{\text{minFrev}}$ with $\text{supp}_{\text{Irr}}(\widetilde{e}) \neq \emptyset$. By Cor. 5.2, we have $\text{supp}_{\widetilde{\text{Irr}}}(\widetilde{e}) = \text{supp}_{\text{Irr}}(\widetilde{e})$. Assume $D = \text{supp}_{\text{Irr}}(\pi^r_{\text{minFrev}}(\widetilde{e})) = \text{supp}_{\text{Irr}}(\widetilde{e})$ is not an MMB in $\mathcal{N}$. Then there exists $v \in \Gamma_{\mathcal{N}}$ with $\emptyset \neq \text{supp}_{\text{Irr}}(v) \subsetneq D$. As before, we consider a decomposition $v = u + y$, with $u, y \in \Gamma_{\mathcal{N}}, u_{\text{minFrev}} = 0$ and $y_{\text{Irr}} = 0$. For $\widetilde{u} = \pi_{\text{minFrev}}(u)$ we get $\text{supp}_{\widetilde{Irr}}(\widetilde{u}) = \text{supp}_{\text{Irr}}(u) = \text{supp}_{\text{Irr}}(v) \subsetneq D = \text{supp}_{\widetilde{\text{Irr}}}(\widetilde{e})$, in contradiction to $\widetilde{e}$ being an extreme ray in $\widetilde{\Gamma}_{\mathcal{N}}^{\text{minFrev}}$. □

We prove now one of our main theorems, stating that for any minimum set of fully reversible reactions minFrev, the extreme rays of $\widetilde{\Gamma}_{\mathcal{N}}^{\text{minFrev}}$ after recombination define a MEMo of $\mathcal{N}$.

Theorem 5.5 *Let minFrev be a minimum set of fully reversible reactions in a metabolic network $\mathcal{N}$ such that $\widetilde{\Gamma}_{\mathcal{N}}^{\text{minFrev}}$ is pointed. After recombination, the*

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

extreme rays of the pointed cone $\widetilde{\Gamma}_{\mathcal{N}}^{\mathrm{minFrev}}$ *define a minimum set of EFMs or MEMo in* $\mathcal{N}$*. Any set of EFMs of smaller cardinality cannot be a generating set for the flux cone* $\Gamma_{\mathcal{N}}$*.*

Proof: According to [41, Sect. 8.8], a minimum set of generating vectors for $\Gamma_{\mathcal{N}}$ always consists of one vector in each minimal proper face of $\Gamma_{\mathcal{N}}$ and a vector space basis of $\Lambda_{\mathcal{N}}$. According to Cor. 5.2, there are two types of extreme rays $\widetilde{e} \in \widetilde{\Gamma}_{\mathcal{N}}^{\mathrm{minFrev}}$:

Case 1: $\mathrm{supp}_{\mathrm{Irr}}(\widetilde{e}) \neq \emptyset$ and $\widetilde{e}_{\mathrm{minFrev} \cup \mathrm{minFrev}_{\mathrm{split}}} = 0$. Let $e = \pi^r_{\mathrm{minFrev}}(\widetilde{e})$ be the recombined vector. By Prop. 5.4, $\mathrm{supp}_{\mathrm{Irr}}(e) = \mathrm{supp}_{\mathrm{Irr}}(\widetilde{e})$ is an MMB D in $\mathcal{N}$. By [26] this implies that e belongs to the minimal proper face defined by D. Since $\mathrm{supp}_{\mathrm{Irr}}(e) \neq 0$, we also have $e \notin \Lambda_{\mathcal{N}}$. By Prop. 2.2, the extreme ray $\widetilde{e}$ is an EFM in $\widetilde{\Gamma}_{\mathcal{N}}^{\mathrm{minFrev}}$. Thus, $\mathrm{supp}(\widetilde{e})$ is inclusion-minimal in $\widetilde{\Gamma}_{\mathcal{N}}^{\mathrm{minFrev}}$. Since $\widetilde{e}_{\mathrm{minFrev} \cup \mathrm{minFrev}_{\mathrm{split}}} = 0$, we have $\mathrm{supp}(e) = \mathrm{supp}(\widetilde{e})$, which implies that e is an EFM in $\Gamma_{\mathcal{N}}$ (otherwise, there would exist $v \in \Gamma_{\mathcal{N}} \setminus \{0\}$ with $\mathrm{supp}(v) \subsetneq \mathrm{supp}(e)$, and for $\widetilde{v} = \mathrm{supp}(\pi_{\mathrm{minFrev}}(v))$ we would get $\mathrm{supp}(\widetilde{v}) = \mathrm{supp}(v) \subsetneq \mathrm{supp}(e) = \mathrm{supp}(\widetilde{e})$, which is a contradiction).

Case 2: $\widetilde{e}_j \neq 0$ for some $j \in \mathrm{minFrev} \cup \mathrm{minFrev}_{\mathrm{split}}$ and $\widetilde{e}_{\mathrm{Irr}} = 0$. According to Cor. 5.2, all ERs with this property lead to t reversible EFMs in $\mathcal{N}$, which after recombination generate $\Lambda_{\mathcal{N}}$.

In conclusion, after recombination, the ERs with the first property represent the MPFs of $\Gamma_{\mathcal{N}}$ and the ERs with the second property form a basis of $\Lambda_{\mathcal{N}}$. Altogether, the recombined ERs of $\widetilde{\Gamma}_{\mathcal{N}}^{\mathrm{minFrev}}$ form a minimum set of EFMs generating $\Gamma_{\mathcal{N}}$, and therefore define a MEMo. □

Corollary 5.6 *For a given metabolic network* $\mathcal{N}$ *the number of EFMs in a MEMo is always* $s + t$*, where* s *is the number of minimal proper faces in* $\Gamma_{\mathcal{N}}$ *and* t *is the dimension of* $\Lambda_{\mathcal{N}}$*.*

Alg. 2 describes how to compute a MEMo given a minimum set of fully reversible reactions minFrev, which can be obtained by Alg. 1. First, the reactions in minFrev are split in order to get the pointed augmented flux cone $\widetilde{\Gamma}_{\mathcal{N}}^{\mathrm{minFrev}}$. Next, we compute the ERs of $\widetilde{\Gamma}_{\mathcal{N}}^{\mathrm{minFrev}}$ using an existing tool based on the double description method [12], such as `polco` [50] or `cdd` [11].

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Algorithm 2: Finding MEMo

Input : Stoichiometric matrix S
Set of irreversible reactions Irr
A minimum set of fully reversible reactions minFrev

Output: MEMo

/* A minimum set of EFMs, describing the underlying metabolic network */

1 $\widetilde{S} \leftarrow (S| - S_{*,\text{minFrev}})$;

/* $\widetilde{S}$ consists of the columns of S and the negative of the columns corresponding to the reactions of minFrev */

2 $I \leftarrow (\text{Irr} \cup \text{minFrev} \cup \text{minFrev}_{\text{split}})$;

/* I contains the indices of the irreversible reactions of the pointed augmented cone $\widetilde{\Gamma}_{\mathcal{N}}^{\text{minFrev}}$, including the split reactions of the set minFrev and the corresponding additional reactions in $\text{minFrev}_{\text{split}}$ */

3 ER $\leftarrow$ extreme rays of $\widetilde{\Gamma}_{\mathcal{N}}^{\text{minFrev}} = \{\widetilde{v} \in \mathbb{R}^{|\mathfrak{R}|+t} \mid \widetilde{S}\widetilde{v} = 0,\ \widetilde{v}_I \geq 0\}$;

/* use polco to compute the ERs of $\widetilde{\Gamma}_{\mathcal{N}}^{\text{minFrev}}$ */

4 MEMo $\leftarrow$ recombination(ER);

/* apply π^r_{minFrev} to the ERs of $\widetilde{\Gamma}_{\mathcal{N}}^{\text{minFrev}}$ */

5 **return** MEMo;

6 Computational results and discussion

In this section we first give some computational results, followed by a discussion and potential applications for MEMos.

6.1 Computational results

We implemented our method in `MATLAB` using `polco` [50] to compute the extreme rays of a pointed cone. Our software is available at `https://sourceforge.net/projects/findingmemo`. In the following, we present computational results for various metabolic network reconstructions taken from `BiGG` Models [22], `KEGG` [21] (together with `KEGGtranslator` [55]), and the `BioModels` Database [27]. Tab. 1 summarises the main characteristics of the metabolic networks studied. In Tab. 2, we study the size of the MEMos compared to the number of extreme pathways (EPs) and the number of EFMs, which were computed with `polco` [50] resp. `efmtool` [49]. In Fig. 9, we compare the number of flux vectors in a MEMo to the number of EPs and EFMs for those networks for which we were able to compute the EPs and EFMs. Since the number of EFMs of a MEMo is much smaller than the number of all EFMs, it is faster to compute them. A table with running times for all computations can be found in the Appendix.

In all networks that we considered the number of fully reversible reactions and the number of internal reversible reactions is greater than or equal to the

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

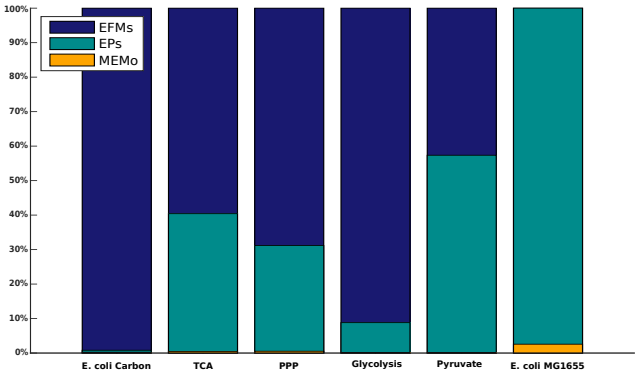


Figure 9 Comparison of the number of flux vectors of the three different generating sets we discussed in this article. We set the number of all EFMs to 100% and compare the vectors of the other sets accordingly. The bars overlay and the largest number is always on the top. Hence, if there is no blue bar for the EFMs *Escherichia coli* MG1655 the number of EPs and the number of EFMs are the same. Note that there is always a yellow bar indicating the size of the MEMo even it is hard to see. This shows how small the number of flux vectors in a MEMo is, compared to the whole set of EFMs.

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1 Characteristics of different metabolic networks used for the comparison in Tab. 2 **network**: name of the network. **rxns**: number of unblocked reactions. **mets**: number of metabolites. **rev**: number of unblocked reversible reactions. **intrev**: number of unblocked internal reversible reactions. **frev**: number of unblocked fully reversible reactions. **t**: dimension of the lineality space of the flux cone.

network	rxns	mets	rev	intrev	frev	t
Escherichia coli carbon metabolism [4]	34	18	34	18	32	15
Citrate cycle (TCA) [21]	35	22	31	25	30	12
Pentose phosphate pathway [21, 55]	57	34	31	18	27	8
Glycolysis / Gluconeogenesis [21, 55]	61	32	42	28	33	13
Pyruvate metabolism [21, 55]	81	28	34	29	24	12
Escherichia coli MG1655 [22]	87	68	34	28	0	0
Rhizobium etli iOR363 [27, 34]	194	371	71	60	0	0
Buchnera iSM197 [27, 29]	261	255	17	11	0	0
Mycoplasma pneumoniae iJW145 [27, 54]	306	266	73	47	0	0
Blattabacterium cuenoti iCG238 [15, 27]	350	364	73	60	5	2
Mycoplasma genitalium iPS189 [27, 47]	351	346	107	26	3	1
Mannheimia succiniciproducens iSH335 [16, 27]	374	316	58	49	3	1
Blattabacterium iCG230 [15, 27]	412	358	71	59	5	2
Helicobacter pylori 26695 [39]	452	396	94	82	31	6
Human Erythrocyte iAB-RBC-283 [2, 27]	469	342	123	68	10	2
Helicobacter pylori iCS291 [27, 39]	493	396	94	67	31	6
Clostridium thermocellum iSR432 [27, 36]	581	564	95	95	32	20

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 2 Size of MEMos, number of EPs and number of EFMs for different networks. **t**: dimension of the lineality space of the flux cone. **rxns**: number of unblocked reactions of the network. **network**: name of the network. **MEMo**: size of a minimum set of elementary modes. **EPs**: number of extreme pathways. **EFMs**: number of elementary flux modes. **N/A**: The programs used (`polco` [50] and `efmtool` [49]) run out of memory.

network	rxns	t	MEMo	EPs	EFMs
Escherichia coli carbon metabolism [4]	34	15	15	52	6,421
Citrate cycle (TCA) [21]	35	12	16	1,564	3,870
Pentose phosphate pathway [21]	57	8	27	1,607	5,155
Glycolysis / Gluconeogenesis [21]	61	13	29	1,716	19,464
Pyruvate metabolism [21]	81	12	49	27,361	47,708
Escherichia coli MG1655 [22]	87	0	2,572	100,274	100,274
Rhizobium etli iOR363 [27, 34]	194	0	6,147	N/A	N/A
Buchnera iSM197 [27, 29]	261	0	245	1,863	N/A
Mycoplasma pneumoniae iJW145 [27, 54]	306	0	1,593,880	N/A	N/A
Blattabacterium cuenoti iCG238 [15]	350	2	376	1,863	N/A
Mycoplasma genitalium iPS189 [27, 47]	351	1	190,471	N/A	N/A
Mannheimia succiniciproducens iSH335 [16, 27]	374	1	2,064,760	N/A	N/A
Blattabacterium iCG230 [15, 27]	412	2	84	N/A	N/A
Helicobacter pylori 26695 [39]	452	6	150,138	N/A	N/A
Human Erythrocyte iAB-RBC-283 [2, 27]	469	2	6,179	N/A	N/A
Helicobacter pylori iCS291 [27, 39]	493	6	150,138	N/A	N/A
Clostridium thermocellum iSR432 [27, 36]	581	20	1,039,267	N/A	N/A

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

dimension t of the lineality space $\Lambda_{\mathcal{N}}$. While there always exist internal reversible reactions, it can happen that there are no fully reversible reactions. In this case, $t = 0$ and the original flux cone is already pointed, i.e., there is no need to split any reaction before computing a MEMo. This holds for the networks *Escherichia coli* MG1655 [22], *Rhizobium etli* iOR363 [27,34], *Buchnera* iSM197 [27,29], and *Mycoplasma pneumoniae* iJW145 [27,54]. For all these networks, there exist internal reversible reactions which are split when computing the extreme pathways (EPs). This is why the number of EPs is larger than the number of flux vectors in a MEMo. For example, instead of 2,572 EFMs in a MEMo for the network *Escherichia coli* MG1655 [22], there exist 100,274 EPs. For this network, the number of EFMs and EPs is the same, although not all reversible reactions are internal.

For the networks *Mannheimia succiniciproducens* iSH335 [16,27], *Helicobacter pylori* 26695 [39], *Helicobacter pylori* iCS291 [27,39], and *Clostridium thermocellum* iSR432 [27,36] we were not able to compute the EPs because `polco` was running out of memory, while we were always able to compute a MEMo.

6.2 Discussion

When computing only a subset of EFMs and not the whole set, several issues arise. One drawback of computing a MEMo is that the resulting set of EFMs is not unique. This problem was already addressed by [44]. We prove in Sect. 5 that there exists a bijection between the irreversible EFMs in a MEMo and the MMBs of the network. Thus, the set of active irreversible reactions is the same for all flux vectors in all different MEMos, while the set of active reversible reactions is not, therefore giving rise to the non-uniqueness of MEMos.

The MEMos that can be obtained by Algor. 2 depend on the set minFrev of reactions that are split. As the following example shows, different minimum sets minFrev of fully reversible reactions may result in the same MEMo. Furthermore, there exist MEMos that cannot be obtained from any set minFrev.

Example 6.1 (Different MEMos) *For the network in Fig. 1 there exist three different reversible EFMs with the supports* $\{3,5,9,10\}, \{3,5,11,12\}, \{9,10,11,12\}$. *Each two of them form a basis of the lineality space* $\Lambda_{\mathcal{N}}$*, thus there are three bases. Furthermore, there exist three MMBs* $\{2\}, \{6,7\}, \{6,8\}$*, see Example 2.1, which correspond to three minimal proper faces of the flux cone* $\Gamma_{\mathcal{N}}$. *The minimal proper face for the MMB* $\{6,7\}$ *contains only one EFM with support* $\{4,6,7\}$. *The other two minimal proper faces each contain three EFMs, namely* $\{1,2,3,4\}, \{1,2,4,5,9,10\}, \{1,2,4,5,11,12\}$ *for the MMB* $\{2\}$ *and* $\{4,6,8,9,10\}, \{3,4,5,6,8\}, \{4,6,8,11,12\}$ *for the MMB* $\{6,8\}$. *The remaining eight EFMs lie in the interior of the cone. These EFMs contain more than one MMB, e.g.* $\{1,2,3,6,7\}$ *contains the MMBs* $\{1,2\}$ *and* $\{6,7\}$. *Therefore these EFMs do not lie in a minimal proper face but in the interior of the cone. Thus they are not contained in any MEMo. By choosing one basis for the lineality space and one EFM for each minimal proper face, we get in total* $3 \cdot 1 \cdot 3 \cdot 3 = 27$ *possible MEMos.*

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

By enumeration, we can determine that there exist 12 different minimum sets of fully reversible reactions minFrev *for the network in Fig. 1, which can be found in the Appendix. These 12 sets give rise to 5 different MEMos. For example, the two* minFrev *sets* $\{3, 9\}$ *and* $\{3, 10\}$ *result in the same MEMo. A MEMo that cannot be obtained by choosing a* minFrev *set and computing the extreme rays of the cone* $\widetilde{\Gamma}_{\mathcal{N}}^{\text{minFrev}}$ *is the set of EFMs with the supports* $\{1, 2, 3, 4\}, \{4, 6, 7\}, \{4, 6, 8, 9, 10\}, \{3, 5, 9, 10\}, \{3, 5, 11, 12\}$.

Alg. 1 computes always the same set minFrev for a given network. Therefore, the MEMo which Alg. 2 computes, will be always the same too, although there exist different MEMos for a given network. Unfortunately, Alg. 2 cannot be used to compute all MEMos of a network even if all sets of minFrev are given, see Example 6.1.

Another issue, besides non-uniqueness, when studying only a subset of EFMs is that not all EFMs of biological relevance may be contained in a MEMo, as pointed out in [44]. This is a valid argument for computing the set of all EFMs. However, for genome-scale metabolic networks this is often not feasible, or the number of EFMs is too large to identify relevant EFMs. In such cases, a MEMo is a good alternative for studying the underlying metabolic network because it is smaller but still represents the whole network. As described in the following subsection, several methods exist which can use a generating subset of EFMs or MEMo instead of the whole set.

6.3 Applications

A typical application for EFMs is the decomposition of a given flux vector into a set of EFMs [45, 46]. Since the set of all EFMs is often not computable, different methods exist which use only a subset of EFMs [3, 17]. [53] introduced a method which decomposes flux vectors to compute the *α-spectrum* of a given flux vector using a (generating) subset of EFMs. The α-spectrum defines which EFMs can and cannot be included in the reconstruction of a given flux vector and to what extent they individually contribute to the decomposition. So far, either the set of all EFMs is used or the set of EPs. As shown in Tab. 2, the number of EPs may be much larger than the number of flux vectors in a MEMo. Furthermore, there exist networks where it is possible to compute a MEMo but not the EPs. For these networks, the α-spectrum can be computed using a MEMo, whereas using the set of EPs or the set of all EFMs is not possible. Based on Wiback's work, [28] developed a method which computes the α-spectrum for a given flux vector including uncertainties. Their concept of *flux-spectrum* is used in scenarios where measurements are incomplete and/or uncertain. Since the method is based on the α-spectrum, again a generating subset or MEMo can be utilized instead of the whole set of EFMs.

Another well-known application of EFMs is the computation of *Minimal Cut Sets* (MCSs). MCSs are, like EFMs, minimal sets of reactions. Whenever the set of reactions of an MCS is removed from a metabolic network, certain predefined target behaviours are not possible anymore [24]. Given the whole set

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of EFMs, MCSs are inclusion-minimal hitting sets, i.e., sets of reactions such that each EFM contains at least one reaction of the hitting set [23]. Unfortunately, an arbitrary generating subset of EFMs cannot be used to compute MCSs. However, as shown in [38], one can use MMBs instead of EFMs to compute MCSs consisting only of irreversible reactions. Here, minimal hitting sets within the MMBs are enumerated, which correspond to *irreversible MCSs* (iMCSs). As shown in Sect. 5, there exists a bijection between the irreversible EFMs in a MEMo and the MMBs of the network. Given a MEMo, extracting the irreversible reactions from the EFMs in the MEMo provides the set of MMBs. In other words, when computing a MEMo we obtain at the same time the set of all MMBs. Therefore, the method presented here also gives a new way for computing MMBs, and based on these for computing iMCSs.

The EFMs in a MEMo can provide new insight into the importance of certain pathways or reactions in the network, especially since their irreversible part is unique. Indeed, investigating which reactions and EFMs are exchangeable and why, is an interesting task. Doing so, it would be possible to study the robustness of the underlying metabolic network. In addition, the EFMs of a MEMo could also be used to discover similarities to other networks by detecting subsets of EFMs that can be found in those other networks as well.

7 Related work

In the following, we discuss a number of papers that are closely related to our work.

7.1 Extreme pathways

In our approach, we compute a MEMo by splitting a minimum subset of reversible reactions. A closely related concept is the set of *extreme pathways* (EPs) [31, 40], which are computed by splitting all the internal reversible reactions. As shown in [31, 40], splitting only the *internal* reversible reactions always delivers an augmented flux cone which is pointed (given that there is only one exchange reaction per internal metabolite). This cone is unique and so are its extreme rays. After recombination, these extreme rays are called extreme pathways. The assumption that there exists only one reversible exchange reaction per metabolite is not true for all metabolic networks. Examples where this condition does not hold are the *Escherichia coli* carbon metabolism [4] or RECON1 [7, 22]. For computing a MEMo we do not need any condition on the network. However, the main advantage of our approach is that the number of EFMs in a MEMo is typically much smaller than the number of EPs, see Tab. 2.

7.2 Minimal metabolic behaviours

[26] introduce a unique minimum outer description of the flux cone $\Gamma_{\mathcal{N}}$ based on *minimal metabolic behaviours* (MMBs) and the lineality space $\Lambda_{\mathcal{N}}$. Each

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MMB corresponds to a minimal proper face of the flux cone or to set of flux vectors with the same minimum set of active irreversible reactions. Larhlimi and Bockmayr show that for each MMB there exists a corresponding EFM, but they do not consider reaction splitting and do not present a method for computing a minimal generating set consisting of EFMs only.

7.3 Minimal generating set

[18] introduce a method for finding a minimum generating set of flux vectors which corresponds to a set of EFMs and therefore is a MEMo. However, their approach is different from the one introduced here. We exploit splitting reversible reactions, while they use a method based on generating a basis of the null space of $\Gamma_{\mathcal{N}}$. Additionally, they only prove that the reversible generating flux vectors are EFMs, but do not consider the irreversible flux vectors. We prove that all generating flux vectors obtained by our approach are EFMs. [18] only mention that the minimum set of EFMs is not unique. Here we discuss this issue in a systematic way and show how to obtain different MEMo's by splitting different sets of reversible reactions. Furthermore, we provide a `MATLAB` tool to compute MEMos and illustrate the impact of our method by computational results for various genome-scale metabolic networks.

8 Conclusion and further work

In this paper we introduced the concept of a minimum set of elementary modes (MEMo) necessary to generate the flux cone of a metabolic network. We presented a method to compute these MEMos. We implemented our algorithm using `MATLAB` and showed that the size of MEMos often is by several orders of magnitude smaller than the number of EPs or EFMs. One drawback of the MEMos is that they are not unique because they depend on the set of fully reversible reactions which have to be split. However, one can show that the set of irreversible reactions involved in the MEMos is unique and that they will always be part of the MEMos independently of the reactions that are split. The biological relevance of the MEMos has to be further investigated. The need of having smaller sets of generating flux vectors has been addressed by various papers [1,9,18,20,35,40]. We expect that MEMos give rise to new opportunities for analysing large genome-scale metabolic networks.

Acknowledgements

We thank Alexandra and Arne Reimers for fruitful discussions and comments on this paper.

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

[1] Arabzadeh, M., Zamani, M.S., Sedighi, M., Marashi, S.A.: A graph-based approach to analyze flux-balanced pathways in metabolic networks. BioSystems (2018)

[2] Bordbar, A., Jamshidi, N., Palsson, B.Ø.: iAB-RBC-283: A proteomically derived knowledge-base of erythrocyte metabolism that can be used to simulate its physiological and patho-physiological states. BMC Systems Biology **5**, 110 (2011)

[3] Chan, S.H.J., Ji, P.: Decomposing flux distributions into elementary flux modes in genome-scale metabolic networks. Bioinformatics **27**(16), 2256–2262 (2011)

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

[5] Clarke, B.L.: Stoichiometric Network Analysis. Cell Biochemistry and Biophysics **12**(1), 237–253 (1988)

[6] David, L., Bockmayr, A.: Computing Elementary Flux Modes Involving a Set of Target Reactions. IEEE/ACM Transactions on Computational Biology and Bioinformatics **11**(6), 1099–1107 (2014)

[7] Duarte, N.C., Becker, S.A., Jamshidi, N., Thiele, I., Mo, M.L., Vo, T.D., Srivas, R., Palsson, B.Ø.: Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proceedings of the National Academy of Sciences **104**(6), 1777–1782 (2007)

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

[9] de Figueiredo, L.F., Podhorski, A., Rubio, A., Kaleta, C., Beasley, J.E., Schuster, S., Planes, F.J.: Computing the shortest elementary flux modes in genome-scale metabolic networks. Bioinformatics **25**(23), 3158–3165 (2009)

[10] Francke, C., Siezen, R.J., Teusink, B.: Reconstructing the metabolic network of a bacterium from its genome. Trends in Microbiology **13**(11), 550–558 (2005)

[11] Fukuda, K., Avis, D.: cdd (2001). https://www.inf.ethz.ch/personal/fukudak/cdd_home/cdd.html

[12] Fukuda, K., Prodon, A.: Double description method revisited. In: Combinatorics and computer science, pp. 91–111. Springer (1996)

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

[14] Gilbert, J., Gilbert, L.: Linear Algebra and Matrix Theory. Academic Press (2014)

[15] González-Domenech, C.M., Belda, E., Patiño-Navarrete, R., Moya, A., Peretó, J., Latorre, A.: Metabolic stasis in an ancient symbiosis: genome-scale metabolic networks from two *Blattabacterium cuenoti* strains, primary endosymbionts of cockroaches. BMC Microbiology **12**(1), S5 (2012)

[16] Hong, S.H., Kim, J.S., Lee, S.Y., In, Y.H., Choi, S.S., Rih, J.K., Kim, C.H., Jeong, H., Hur, C.G., Kim, J.J.: The genome sequence of the capnophilic rumen bacterium *Mannheimia succiniciproducens*. nature biotechnology **22**(10), 1275–1281 (2004)

[17] Ip, K., Colijn, C., Lun, D.S.: Analysis of complex metabolic behavior through pathway decomposition. BMC Systems Biology **5**(1), 91 (2011)

[18] Jevremovic, D., Boley, D.: Finding minimal generating set for metabolic network with reversible pathways. BioSystems **112**(1), 31–36 (2013)

[19] Jevremović, D., Trinh, C.T., Srienc, F., Boley, D.: On algebraic properties of extreme pathways in metabolic networks. J. Comput. Biol. **17**(2), 107–119 (2010)

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

[21] Kanehisa, M., Goto, S.: KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research **28**(1), 27–30 (2000)

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

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

[24] Klamt, S., Gilles, E.D.: Minimal cut sets in biochemical reaction networks. Bioinformatics **20**(2), 226–234 (2004)

[25] Larhlimi, A., Bockmayr, A.: On inner and outer descriptions of the steady-state flux cone of a metabolic network. In: International Conference on Computational Methods in Systems Biology, pp. 308–327. Springer (2008)

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[26] Larhlimi, A., Bockmayr, A.: A new constraint-based description of the steady-state flux cone of metabolic networks. Discrete Applied Mathematics **157**(10), 2257–2266 (2009)

[27] Li, C., Donizelli, M., Rodriguez, N., Dharuri, H., Endler, L., Chelliah, V., Li, L., He, E., Henry, A., Stefan, M.I., et al.: BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. BMC Systems Biology **4**, 92 (2010)

[28] Llaneras, F., Picó, J.: An interval approach for dealing with flux distributions and elementary modes activity patterns. Journal of Theoretical Biology **246**(2), 290–308 (2007)

[29] MacDonald, S., Thomas, G., Douglas, A.: Genetic and metabolic determinants of nutritional phenotype in an insect–bacterial symbiosis. Molecular Ecology **20**(10), 2073–2084 (2011)

[30] Notebaart, R.A., Van Enckevort, F.H.J., Francke, C., Siezen, R.J., Teusink, B.: Accelerating the reconstruction of genome-scale metabolic networks. BMC Bioinformatics **7**, 296 (2006)

[31] Papin, J.A., Stelling, J., Price, N.D., Klamt, S., Schuster, S., Palsson, B.Ø.: Comparison of network-based pathway analysis methods. Trends in Biotechnology **22**(8), 400–405 (2004)

[32] Pey, J., Planes, F.J.: Direct calculation of elementary flux modes satisfying several biological constraints in genome-scale metabolic networks. Bioinformatics pp. 2197–2209 (2014)

[33] Pfeiffer, T., Sánchez-Valdenebro, I., Nuño, J.C., Montero, F., Schuster, S.: METATOOL: for studying metabolic networks. Bioinformatics **15**(3), 251–257 (1999)

[34] Resendis-Antonio, O., Reed, J.L., Encarnación, S., Collado-Vides, J., Palsson, B.Ø.: Metabolic Reconstruction and Modeling of Nitrogen Fixation in *Rhizobium etli*. PLoS Computational Biology **3**(10), e192 (2007)

[35] Rezola, A., de Figueiredo, L.F., Brock, M., Pey, J., Podhorski, A., Wittmann, C., Schuster, S., Bockmayr, A., Planes, F.J.: Exploring metabolic pathways in genome-scale networks via generating flux modes. Bioinformatics **27**(4), 534–540 (2011)

[36] Roberts, S.B., Gowen, C.M., Brooks, J.P., Fong, S.S.: Genome-scale metabolic analysis of *Clostridium thermocellum* for bioethanol production. BMC Systems Biology **4**(31) (2010)

[37] Röhl, A., Bockmayr, A.: Reaction Splitting and Minimum Sets of Elementary Flux Modes. Proceedings of the Lyon Spring School on advances in Systems and Synthetic Biology **16** (2017)

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[38] Röhl, A., Riou, T., Bockmayr, A.: Computing irreversible minimal cut sets in genome-scale metabolic networks via flux cone projection. Bioinformatics (2018)

[39] Schilling, C.H., Covert, M.W., Famili, I., Church, G.M., Edwards, J.S., Palsson, B.Ø.: Genome-Scale Metabolic Model of *Helicobacter pylori* 26695. Journal of Bacteriology **184**(16), 4582–4593 (2002)

[40] Schilling, C.H., Letscher, D., Palsson, B.Ø.: Theory for the Systemic Definition of Metabolic Pathways and their use in Interpreting Metabolic Function from a Pathway-Oriented Perspective. Journal of Theoretical Biology **203**(3), 229–248 (2000)

[41] Schrijver, A.: Theory of linear and integer programming. Wiley, Chichester; New-York (1998)

[42] Schuster, S., Fell, D.A., Dandekar, T.: A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nature Biotechnology **18**(3), 326–332 (2000)

[43] Schuster, S., Hilgetag, C.: On elementary flux modes in biochemical reaction systems at steady state. Journal of Biological Systems **2**(02), 165–182 (1994)

[44] Schuster, S., Hilgetag, C., Woods, J.H., Fell, D.A.: Reaction routes in biochemical reaction systems: Algebraic properties, validated calculation procedure and example from nucleotide metabolism. Journal of Mathematical Biology **45**(2), 153–181 (2002)

[45] Schwartz, J.M., Kanehisa, M.: A quadratic programming approach for decomposing steady-state metabolic flux distributions onto elementary modes. Bioinformatics **21**(suppl_2), ii204–ii205 (2005)

[46] Schwartz, J.M., Kanehisa, M.: Quantitative elementary mode analysis of metabolic pathways: the example of yeast glycolysis. BMC Systems Biology **7**(1), 186 (2006)

[47] Suthers, P.F., Dasika, M.S., Kumar, V.S., Denisov, G., Glass, J.I., Maranas, C.D.: A Genome-Scale Metabolic Reconstruction of *Mycoplasma genitalium, i*PS189. PLoS Comput Biol **5**(2), e1000,285 (2009)

[48] Terzer, M.: Large scale methods to enumerate extreme rays and elementary modes. Ph.D. thesis, ETH Zurich (2009)

[49] Terzer, M.: efmtool (2017). http://www.csb.ethz.ch/tools/software/efmtool.html

[50] Terzer, M.: polco (2017). http://www.csb.ethz.ch/tools/software/polco.html

[51] Thiele, I., Palsson, B.Ø.: A protocol for generating a high-quality genome-scale metabolic reconstruction. Nature Protocols **5**(1), 93–121 (2010)

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[52] Urbanczik, R., Wagner, C.: An improved algorithm for stoichiometric network analysis: theory and applications. Bioinformatics **21**(7), 1203–1210 (2005)

[53] Wiback, S.J., Mahadevan, R., Palsson, B.Ø.: Reconstructing metabolic flux vectors from extreme pathways: defining the α-spectrum. Journal of Theoretical Biology **224**(3), 313–324 (2003)

[54] Wodke, J.A., Puchałka, J., Lluch-Senar, M., Marcos, J., Yus, E., Godinho, M., Gutiérrez-Gallego, R., Dos Santos, V.A.M., Serrano, L., Klipp, E., Maier, T.: Dissecting the energy metabolism in *Mycoplasma pneumoniae* through genome-scale metabolic modeling. Molecular Systems Biology **9**(1), 653 (2013)

[55] Wrzodek, C., Büchel, F., Ruff, M., Dräger, A., Zell, A.: Precise generation of systems biology models from KEGG pathways. BMC Systems Biology **7**(15) (2013)

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

9 Appendix

9.1 Example: all minFrevs and the corresponding MEMo

Table 3 All different sets of minFrev for the example network. The reactions of the given set minFrev are split and the extreme rays of the corresponding pointed flux cone are computed. After recombination the extreme rays always correspond to five different EFMs which form a MEMo. For different sets of minFrev the same set of EFMs can occur.
minFrev: a set of two reactions. The augmented flux cone, after splitting these two reactions, is pointed.
EFMi: the support of the i-th EFM of the MEMo.
MEMo: which MEMo is described by the five EFMs.
To point out which sets of minFrev result in the same MEMo we put a dividing line between different MEMos.

minFrev	EFM1	EFM2	EFM3	EFM4	EFM5	MEMo
3,9	1,2,4,5,11,12	3,5,11,12	4,6,7	4,6,8,11,12	9,10,11,12	I
3,10	1,2,4,5,11,12	3,5,11,12	4,6,7	4,6,8,11,12	9,10,11,12	I
3,11	1,2,4,5,9,10	3,5,9,10	4,6,7	4,6,8,9,10	9,10,11,12	II
3,12	1,2,4,5,9,10	3,5,9,10	4,6,7	4,6,8,9,10	9,10,11,12	II
5,9	1,2,3,4	3,5,11,12	4,6,7	4,6,8,11,12	9,10,11,12	III
5,10	1,2,3,4	3,5,11,12	4,6,7	4,6,8,11,12	9,10,11,12	III
5,11	1,2,3,4	3,5,9,10	4,6,7	4,6,8,9,10	9,10,11,12	IV
5,12	1,2,3,4	3,5,9,10	4,6,7	4,6,8,9,10	9,10,11,12	IV
9,11	1,2,3,4	3,4,5,6,8	3,5,9,10	3,5,11,12	4,6,7	V
9,12	1,2,3,4	3,4,5,6,8	3,5,9,10	3,5,11,12	4,6,7	V
10,11	1,2,3,4	3,4,5,6,8	3,5,9,10	3,5,11,12	4,6,7	V
10,12	1,2,3,4	3,4,5,6,8	3,5,9,10	3,5,11,12	4,6,7	V

bioRxiv preprint doi: https://doi.org/10.1101/705012; this version posted July 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

9.2 Time for computation

Table 4 Time to compute a MEMo, EPs and EFMs for different networks. **rxns**: number of unblocked reactions of the network. **network**: name of the network. **MEMo**: time to compute a minimum set of elementary modes. **EPs**: time to compute the extreme pathways. **EFMs**: time to compute the elementary flux modes. **N/A**: The programs used (`polco` [50] and `efmtool` [49]) run out of memory.

network	rxns	MEMo	EPs	EFMs
Escherichia coli carbon metabolism [4]	34	0.62 sec	6 sec	3 sec
Citrate cycle (TCA) [21]	35	0.34 sec	0.6 sec	0.3 sec
Pentose phosphate pathway [21]	57	0.5 sec	0.4 sec	0.5 sec
Glycolysis / Gluconeogenesis [21]	61	0.5 sec	0.3 sec	0.9 sec
Pyruvate metabolism [21]	81	0.5 sec	1.4 sec	2 sec
Escherichia coli MG1655 [22]	87	1.6 sec	17 sec	9 sec
Rhizobium etli iOR363 [27, 34]	194	5 sec	N/A	N/A
Buchnera iSM197 [27, 29]	261	2.5 sec	2 sec	N/A
Mycoplasma pneumoniae iJW145 [27, 54]	306	5 min	N/A	N/A
Blattabacterium cuenoti iCG238 [15]	350	10.6 sec	2 sec	N/A
Mycoplasma genitalium iPS189 [27, 47]	351	3 min	N/A	N/A
Mannheimia succiniciproducens iSH335 [16, 27]	374	3 min	N/A	N/A
Blattabacterium iCG230 [15, 27]	412	7.8 sec	N/A	N/A
Helicobacter pylori 26695 [39]	452	1 min	N/A	N/A
Human Erythrocyte iAB-RBC-283 [2, 27]	469	15 sec	N/A	N/A
Helicobacter pylori iCS291 [27, 39]	493	1 min 5 sec	N/A	N/A
Clostridium thermocellum iSR432 [27, 36]	581	6.5 min	N/A	N/A