Computing Elementary Flux Modes Involving a Set of Target Reactions

Laszlo David and Alexander Bockmayr

Abstract—Elementary flux mode (EM) computation is an important tool in the constraint-based analysis of genome-scale metabolic networks. Due to the combinatorial complexity of these networks, as well as the advances in the level of detail to which they can be reconstructed, an exhaustive enumeration of all EMs is often not practical. Therefore, in recent years interest has shifted towards searching EMs with specific properties. We present a novel method that allows computing EMs containing a given set of target reactions. This generalizes previous algorithms where the set of target reactions consists of a single reaction. In the one-reaction case, our method compares favorably to the previous approaches. In addition, we present several applications of our algorithm for computing EMs containing two target reactions in genome-scale metabolic networks. A software tool implementing the algorithms described in this paper is available at https://sourceforge.net/projects/caefm.

Index Terms—Metabolic networks, constraint-based analysis, elementary flux modes, mixed-integer linear programming

1 Introduction

Elementary (flux) modes (EMs) [1], [2], [3], [4], [5] are an important concept for the structural analysis of metabolic networks, with many practical applications (see e.g. [6] and references therein). As a consequence, the development of methods for the computation of EMs has become an active research area over the past years [7], [8], [9], [10], [11], [12], [13]. The computational complexity of enumerating all EMs is not known [14]. However, there exist several algorithms and software packages for an exhaustive enumeration in a given metabolic network [7], [8], [9], [10]. While these methods work very well for small networks, due to the possibly exponential number of EMs, they may fail for medium or large genome-scale networks. With the ever increasing size of genome-scale metabolic network reconstructions, EM analysis nowadays can often be used only under additional assumptions (e.g., modifying the system boundary of the network, blocking a large number of uptake reactions etc.). These extra assumptions may have the bad side effect of changing the structure of the network, sometimes introducing artificial pathways [15].

One way to deal with genome-scale networks is to define a subset of interest $\mathcal{R}_I$ of the full reaction set $\mathcal{R}$, without altering the network topology. Kaleta et al. [16] look for sets of reactions in $\mathcal{R}_I$, called flux patterns, which indicate the existence of an EM having those reactions in its support. They enumerate a basis for this set, the elementary flux patterns. Urbanczik and Wagner [17] and Marashi et al. [15] project the steady-state flux cone onto the subspace defined by $\mathcal{R}_I$, and enumerate the partial EMs for this subspace.

Given the difficulty of computing and analyzing the full set of all EMs, recent research has focused on finding a special subset of EMs [11], [12], [13]. De Figueiredo et al. [11] describe a mixed-integer programming method to enumerate the k shortest EMs ($k \geq 1$). Their method has been extended to find shortest EMs involving one reaction of choice and also to enumerate a minimal generating set of EMs [12]. In the last paper, the authors note that their method cannot be applied to find elementary modes (EMs) involving two predefined reactions.

The problem of finding an EM involving two or more given reactions was also considered by [14]. They give a more general formulation, in that they not only want their EM to involve a certain set of target reactions $T \subset \mathcal{R}$ (with $|T| = t > 0$), but also to avoid another set of reactions $F \subset \mathcal{R}$ (with $F \cap T = \emptyset$). In other words, the goal is to find an EM e, with $e_i \neq 0$ for $i \in T$ and $e_i = 0$ for $i \in F$. The authors show that defining a set of reactions F to be avoided does not add to the difficulty of the problem, and in fact reactions belonging to F can simply be removed from the network. Acuña et al. [18] study the complexity and prove that the decision problem, whether an EM involving two or more target reactions exists, is not solvable in polynomial time, unless $P = NP$.

To the best of our knowledge, no algorithm for this problem has been published so far. Here, we develop a mixed-integer programming approach for the more general problem of computing k elementary modes ($k \geq 1$) involving a given set of target reactions T, for $|T| \geq 2$. Computational experiments show that the method can be applied even to large genome-scale networks.

2 Preliminaries

We consider a metabolic network N with a set of metabolites $\mathcal{M}$ and a set of reactions $\mathcal{R}$. Formally, the network N is

- L. David is with the Department of Mathematics and Computer Science, Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany, the DFG-Research Center Matheon, and the Berlin Mathematical School (BMS). E-mail: Laszlo.David@fu-berlin.de.
- A. Bockmayr is with the Department of Mathematics and Computer Science, Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany and the DFG-Research Center Matheon. E-mail: Alexander.Bockmayr@fu-berlin.de.

Manuscript received 15 Dec. 2013; revised 13 May 2014; accepted 9 July 2014. Date of publication 28 July 2014; date of current version 4 Dec. 2014.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TCBB.2014.2343964

defined as $N = (\mathcal{M}, \mathcal{R}, S, Irr)$, with a stoichiometric matrix $S \in \mathbb{R}^{\mathcal{M}\times\mathcal{R}}$, and a set of irreversible reactions $Irr \subseteq \mathcal{R}$. For $I \subseteq \mathcal{M}$ and $J \subseteq \mathcal{R}$, we denote by S_{IJ} the submatrix of S defined by the rows from I and the columns from J. Any subset of reactions $\mathcal{R}' \subseteq \mathcal{R}$ defines a metabolic *subnetwork* $N' = (\mathcal{M}, \mathcal{R}', S', Irr')$, with $S' = S_{N'} = S_{\mathcal{M}\mathcal{R}'} \in \mathbb{R}^{\mathcal{M}\times\mathcal{R}'}$ and $Irr' = Irr \cap \mathcal{R}'$.

The stoichiometric constraints $Sv = 0$ (mass balance) and the thermodynamic constraints $v_{Irr} \geq 0$ define the *steady-state flux cone* $C = \{v \in \mathbb{R}^{\mathcal{R}} \mid Sv = 0, v_{Irr} \geq 0\}$. Any vector $v \in C\backslash\{0\}$ is called a *flux mode*. The *support* of v is the set $supp(v) = \{j \in \mathcal{R} \mid v_j \neq 0\}$ of active reactions in v. A reaction that is not active in any flux mode $v \in C$ is called *blocked*. A flux mode $v \in C \setminus \{0\}$ with minimal support (with respect to $\subseteq$) is called an *elementary mode*. The goal of this paper is to study the following *problem* ($\mathscr{P}$):

> Given a metabolic network $N = (\mathcal{M}, \mathcal{R}, S, Irr)$, a set of target reactions $T \subseteq \mathcal{R}$ and $k \geq 1$, compute a set E of EMs in N, $|E| = k$, such that $supp(e) \supseteq T$, for all $e \in E$.

Flux coupling analysis [19] defines four binary relations for pairs of unblocked reactions $i, j \in \mathcal{R}$. We say that i is *directionally coupled* to j and write $i \rightarrow j$, if $v_i \neq 0$ implies $v_j \neq 0$, for all $v \in C$. The reactions i and j are *partially coupled* if both $i \rightarrow j$ and $j \rightarrow i$ hold. Two partially coupled reactions i and j are *fully coupled* if there is $\lambda \neq 0$ such that $v_i = \lambda v_j$, for all $v \in C$. If neither $i \rightarrow j$ nor $j \rightarrow i$, then the reactions i and j are *uncoupled*. Following [20], we call two uncoupled reactions i and j *mutually exclusive* if they are never active together in the same EM, otherwise they are called *sometimes coupled*.

For the rest of this paper, we assume that all reactions are irreversible, i.e., $Irr = \mathcal{R}$, and that none of the reactions is blocked. These assumptions do not limit the applicability of our methods. Reversible reactions can be split into a forward and backward reaction. When $|T| > 1$ and $l \in \{1, \ldots, |T|\}$ reactions in T are reversible, then the original problem can be reduced to 2^l subproblems by considering every combination of forward and backward arcs for the reversible reactions in T. Each of these subproblems can be solved independently of the others.

In general, the splitting operation may induce a number of artificial EMs in the form of two-cycles. However, these two-cycles in most cases do not increase the set of solutions for problem ($\mathscr{P}$). In fact, the only case where such two-cycles satisfy the conditions of problem ($\mathscr{P}$) is for $|T| = 1$. In this case, there is exactly one artificial EM that needs to be filtered out from the final solution set. For $|T| > 1$ no pairs of split reactions will be part of the same EM.

All blocked reactions can be identified in polynomial time, by solving a linear number of linear programs. Afterwards, they can be removed from the network without altering the underlying flux cone.

3 Methods

In the following, we divide the general problem ($\mathscr{P}$) into three subproblems and discuss them individually: the one-reaction case, the two-reaction case and the general t-reaction case (where $t = |T| > 2$). The underlying details vary in each case, but there is a general concept followed by all three methods. In every case, we aim to incrementally find an alternating sequence N^1, e^1, N^2, $e^2, \ldots, N^k$, e^k, of subnetworks and EMs of the input network N with the following properties for all $i \in \{1, \ldots, k\}$:

1) The target reactions $r_1, \ldots, r_t$ are part of every subnetwork N^i.
2) In every subnetwork N^i, reaction r_1 is directionally coupled to the reactions $r_2, \ldots, r_t$.
3) No subnetwork N^i has as flux mode any of the EMs e^l for $l < i$.
4) e^i is an EM in N^i involving r_1.

Clearly, the main difficulty in finding such a sequence of subnetworks is in imposing condition (2). In turn, once a new subnetwork $N^i = (\mathcal{M}, \mathcal{R}^i, S^i)$ has been found, a corresponding EM e^i can be computed by solving the linear program LP(N^i) [14].

$$\begin{aligned} \text{LP}(N^i):\ \min\ & 0 \\ \text{s.t.}\ & S^i v = 0, \\ & v_{r_1} \geq 1, \\ & v_j \geq 0, \quad \forall j \in \mathcal{R}^i. \end{aligned}$$

Using a Simplex-based method, we can compute a basic feasible solution e^i of LP(N^i), which corresponds to a vertex of the truncated flux cone of N^i (resp. an extreme ray of the flux cone of N^i), and thus defines an EM in N^i [7], which involves r_1. Due to the conservation property of EMs (see e.g. [20, Lemma 1]), e^i is also an EM in N. The set $E^i := \{e^1, \ldots, e^i\}$, $i \geq 1$, contains EMs of the original network N that involve every target reaction. Terminating the search after k EMs have been found provides a solution to problem $\mathscr{P}$.

3.1 The One-Reaction Problem

If the set of target reactions consists of only one reaction, condition (2) is trivially satisfied for every subnetwork satisfying condition (1). Assuming $N^1, e^1, N^2, e^2, \ldots N^i, e^i$ have already been computed, we can determine a new subnetwork N^{i+1} by solving the following mixed-integer linear program (MILP1).

$$\begin{aligned} &\text{MILP1}(E^i): \\ &\quad \min\ 0 \\ &\quad \text{s.t.}\ Sv = 0, \\ &\quad v_{r_1} \geq 1, \\ &\quad a_l \leq v_l \leq M a_l, && \forall l \in \mathcal{R}, \\ &\quad \sum_{l \in supp(e^u)} a_l \leq |supp(e^u)| - 1, && \forall u \in [1..i], \\ &\quad v_l \geq 0, && \forall l \in \mathcal{R}, \\ &\quad a_l \in \{0, 1\}, && \forall l \in \mathcal{R}. \end{aligned}$$

The constraints in (MILP1) are the same as for the computation of the so-called shortest EMs in [11]. There are two groups of variables: v represents the steady-state flux values of the reactions, while the 0-1 vector a models the support

TABLE 1
Algorithm 1 for the One-Reaction Case

Step	Action
0.	Initialize $i := 1, E := \emptyset$.
1.	Try to find a feasible solution (v', a') of MILP1(E).
2.	If MILP1(E) is infeasible, then STOP.
3.	Otherwise, use (v', a') to derive subnetwork N^i.
4.	Find a basic feasible solution e^i of LP(N^i).
5.	Let $E := E \cup \{e^i\}$ and $i := i + 1$.
6.	If $i > k$ then STOP.
7.	Go to Step 1.

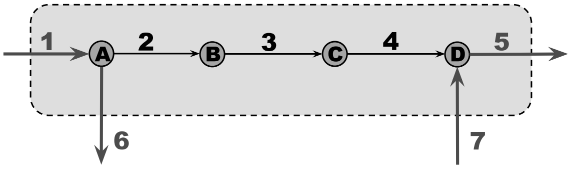

Fig. 1. Example network. Thick arrows represent the optimal solution of (MILP2).

of v (i.e., $v_l > 0 \Leftrightarrow a_l = 1$). The variables v_l and a_l are linked by the 3rd constraint, using a suitably large constant $M > 0$. An important difference to [11] is the objective function (see also Section 4.2). We do not try to find the smallest set of reactions satisfying the constraints of (MILP1). Instead computing any feasible solution is sufficient. This turns out to be enough to derive a subnetwork that satisfies conditions (1-3). Indeed, given a feasible solution (v', a') of (MILP1), define N^{i+1} by the set of reactions $\mathcal{R}^{i+1} := \{l \in \mathcal{R} \mid a'_l = 1\}$. This subnetwork N^{i+1} clearly satisfies conditions (1) and (2), while the fourth constraint in (MILP1), the so-called *no-good constraints*, guarantees condition (3). By solving LP (N^{i+1}), we obtain an EM e^{i+1}.

Table 1 summarizes the method for the one-reaction case. The two conditional exit points of the algorithm are Steps 2 and 6. If the algorithm terminates at Step 2, Proposition 1 assures all EMs will be found. In contrast, if the exit occurs at Step 6, we enumerate k EMs.

Proposition 1. *For any EM $e \notin E$ and sufficiently large $M > 0$, there is a flux mode v' in N such that $(v', supp(e))$ is a feasible solution for MILP1(E).*

A formal proof of Proposition 1 is omitted here. It can be easily obtained by suitably scaling the vector e.

We note that it is possible to initialize $N^1 := N$, thus avoiding the need to solve the very first mixed-integer program. The computationally hard part is Step 1 of the algorithm, while the other steps can be done in polynomial time.

3.2 The Two-Reaction Problem

A natural idea to find a shortest EM involving a pair of reactions $\{r_1, r_2\} \subseteq \mathcal{R}$ would be to extend the previous method by forcing both r_1 and r_2 to be active, while minimizing the total number of active reactions:

$$
\begin{array}{llcll}
(\text{MILP2}) : \min & \displaystyle\sum_{l \in \mathcal{R}} a_l & & & \\
\text{s.t.} & Sv & = & 0, & \\
 & v_{r_1} & \geq & 1, & \\
 & v_{r_2} & \geq & 1, & \\
 & a_l \leq v_l & \leq & M a_l, & \forall l \in \mathcal{R}, \\
 & v_l & \geq & 0, & \forall l \in \mathcal{R}, \\
 & a_l & \in & \{0, 1\}, & \forall l \in \mathcal{R}.
\end{array}
$$

(MILP2) indeed finds a shortest flux mode containing r_1 and r_2. However, it may fail to produce an EM. As illustrated by Fig. 1, if we apply (MILP2) for reactions 1 and 5, a non-elementary flux mode will be found, involving the reactions 1, 6, 7, 5. However, an optimal solution (v^*, a^*) of (MILP2) still has interesting properties, which will turn out to be useful for refining our method. These are described in the following propositions.

Proposition 2. *Let $v^* = \sum_{i=1}^{s} \lambda_i e^i$, with $\lambda_i > 0$, be any decomposition of an optimal solution (v^*, a^*) of (MILP2) into s EMs in N. Then for all $i \in [1..s]$, $e^i_{r_1} > 0$ or $e^i_{r_2} > 0$. There exists $i \in [1..s]$ with $e^i_{r_1} > 0$ and $e^i_{r_2} > 0$, if and only if $s = 1$.*

Proof. Suppose there exists $i \in [1..s]$ such that $e^i_{r_1} = 0$ and $e^i_{r_2} = 0$. Let $p := \min\{v^*_j / e^i_j \mid j \in supp(e^i)\}$ and $v' := v^* - pe^i$. Then there exist $\lambda > 0$ and $a' \in \{0,1\}^{\mathcal{R}}$ such that $(\lambda v', a')$ is a feasible solution of (MILP2), with $\sum_{l \in \mathcal{R}} a'_l < \sum_{l \in \mathcal{R}} a^*_l$, in contradiction to the optimality of v^*.

From $s = 1$, we get $v^* = \lambda_1 e^1, \lambda_1 > 0$, and therefore $e^i_{r_1} > 0$ and $e^i_{r_2} > 0$. Conversely, suppose there exists $i \in [1..s]$ such that $e^i_{r_1} > 0$ and $e^i_{r_2} > 0$. Up to scaling, e^i is a feasible solution of (MILP2) with $supp(e^i) \subseteq supp(v^*)$. From the optimality of v^*, we get $supp(e^i) = supp(v^*)$, which implies $s = 1$. □

Proposition 2 shows that any EM participating in a decomposition of v^* must contain at least one of the two target reactions. The only case when an EM contains both r_1 and r_2 is when the optimal solution v^* itself is an EM.

The result can be formulated in a stronger form, extending it to the whole subnetwork $N^* = (\mathcal{M}, \mathcal{R}^*, S_{\mathcal{M}\mathcal{R}^*})$ defined by v^* (with $\mathcal{R}^* = supp(v^*)$). Every EM e in this subnetwork will satisfy $e_{r_1} > 0$ or $e_{r_2} > 0$. If v^* is itself an EM, then also the subnetwork N^* will have only one EM, namely v^*. The following corollary summarizes our previous results.

Corollary 3. *In the subnetwork N^* defined by v^*, the reactions r_1 and r_2 are either fully coupled or mutually exclusive.*

Our next proposition shows that any flux mode using the reactions r_1 and r_2 can be scaled by a positive factor so that it becomes a feasible solution of (MILP2).

Proposition 4. *Let $v \in C$ with $v_{r_1} > 0$ and $v_{r_2} > 0$. Then for sufficiently large $M > 0$, there exists a feasible solution (v', a') of (MILP2), such that $v' = \lambda v$, for some $\lambda > 0$.*

Proof. Let $\lambda := 1/\min\{v_i \mid i \in supp(v)\}$ and $M := \max\{v_i / v_j \mid i, j \in supp(v)\}$. Now consider $v' := \lambda v$ and let $a'_l := 1$, for all $l \in supp(v)$, and $a'_l := 0$, for all $l \in \mathcal{R} \setminus supp(v)$. For all $l \in supp(v)$, $v'_l = v_l / \min\{v_i \mid i \in \ supp(v)\} \geq 1$. Thus, $v'_{r_1} \geq 1$ and $v'_{r_2} \geq 1$. By definition of a', for all $l \in supp(v)$, $a'_l \ = \ 1 \leq v'_l = v_l / \min\{v_i \mid i \in supp(v)\} \leq \max\{v_i / v_j \mid i, j \in$

$supp(v)\} = M = Ma'_l$. Since $v' \in C$ also holds, we conclude that (v', a') is a feasible solution of (MILP2). □

In general, when decomposing an arbitrary steady-state flux vector into EMs, the number of participating EMs ranges from only one to many. Moreover, a decomposition does not necessarily have to be unique. It turns out that in our special case, a decomposition is much more constrained. In fact, no decomposition can contain more than two EMs. As direct corollary of this result, we will also get the uniqueness of the decomposition.

Proposition 5. *Let (v^*, a^*) be an optimal solution of (MILP2) and let $v^* = \sum_{i=1}^{s} \lambda_i e^i$, with $\lambda_i > 0$, for all $i \in [1..s]$, be a decomposition of v^* into s EMs with pairwise different support. Then $s \leq 2$.*

Proof. Assume $s > 2$. From Proposition 2 it follows that there exist $i, j \in [1..s]$ such that e^i and e^j both contain r_1 or both contain r_2. Without loss of generality, we assume that e^i and e^j both contain r_1. Then, since $s \neq 1$, both e^i and e^j do not contain r_2. Let $\mathcal{R}_{diff} := supp(e^j) \setminus supp(e^i)$. Since e^i is an EM, we have $\mathcal{R}_{diff} \neq \emptyset$. Let $p := \min\{v^*_r/e^j_r \,|\, r \in \mathcal{R}_{diff}\}$. The vector $v' := v^* - pe^j$ satisfies $supp(v') \subsetneq supp(v^*)$. However, it need not be feasible for (MILP2) because it may violate some constraint $v_r \geq 0$. Let $V := \{r \in \mathcal{R} \,|\; v'_r < 0\} \subseteq supp(e^j) \cap supp(e^i)$. If $V \neq \emptyset$, define $q := \min\ \{v'_r/e^i_r \,|\; r \in V\} < 0$, otherwise $q := 0$. Then $v'' := v' + (1 - q)e^i \geq 0, v''_{r_1} > 0, v''_{r_2} > 0$, and we still have $supp(v'') \subsetneq supp(v^*)$. By Proposition 4, this implies the existence of a feasible solution for (MILP2) with a smaller objective function value than v^*, which is a contradiction. □

Corollary 6. *Any decomposition of v^* into EMs is unique.*

Proof. The result is trivial if v^* is itself an EM. Thus we only have to consider the case $s = 2$. Assume by contradiction that $v^* = \lambda_1 e^1 + \lambda_1 e^2 = \mu_1 e^3 + \mu_2 e^4$, where at least three of e^1, e^2, e^3, e^4 have pairwise different support. Clearly, then $v^* = (\lambda_1 e^1 + \lambda_1 e^2 + \mu_1 e^3 + \mu_2 e^4)/2$, which contradicts the result of Proposition 5. □

Of special interest is Corollary 3, which asserts that in the subnetwork N^* defined by v^*, r_1 and r_2 are either fully coupled or mutually exclusive. Thus, to make the optimal solution of (MILP2) an EM, it is enough to add constraints that exclude the second case. These additional constraints must forbid the existence of an EM in N^* involving exactly one of r_1 and r_2. This can be achieved by requiring that r_1 should be directionally coupled to r_2 in N^* (or alternatively that r_2 is directionally coupled to r_1). To formulate this mathematically, we delete r_2 from N^* and require that r_1 is blocked in the resulting subnetwork N'. More formally, if $S' := S_{*,supp(v^*)\setminus\{r_2\}}$ is the stoichiometric matrix of N', then the following system should be infeasible:

$$\begin{aligned} S'z &= 0, \\ z_{r_1} &= 1, \\ z &\geq 0. \end{aligned}$$

By applying Farkas' Lemma (see e.g. [21]), this infeasibility requirement can be turned into a feasibility condition in the dual space. Let $y \in \mathbb{R}^m$, $x \in \mathbb{R}$. Then the following system in y and x should be feasible:

$$\begin{aligned} (S')^T y + u^{r_1} x &\geq 0, \\ x &\leq -1. \end{aligned}$$

Here u^r is the rth unit vector (with an entry 1 for component r, and 0 otherwise) and $.^T$ denotes transposition of a matrix.

This formulation inherently uses S', the stoichiometric matrix of the new subnetwork. Naturally, information about it is not derivable independently. Thus the first constraint set needs to be reformulated to dynamically adjust itself according to the current solution (v, a) of (MILP2). This leads to the following constraints:

$$\begin{aligned} S^T y + u^{r_1} x &\geq -M(\mathbf{1} - a + u^{r_2}), \\ x &\leq -1, \end{aligned} \qquad \text{DirC}(r_1, r_2)$$

where $\mathbf{1}$ denotes a vector all whose components are 1.

By using a large enough constant M, the first constraint becomes trivially satisfiable for inactive reactions ($a_l = 0$) and for r_2, where the right-hand side simplifies to $-M$. In contrast, for active reactions ($a_l = 1$) different from r_2, the right-hand side sums up to 0, thus effectively activating the constraint. The inequalities $\text{DirC}(r_1, r_2)$ are called *directional coupling constraints for r_1 implying r_2*. Extending (MILP2) with $\text{DirC}(r_1, r_2)$, allows computing a shortest EM through r_1 and r_2. The following Proposition 7 summarizes our construction. It guarantees that by adding the constraints $\text{DirC}(r_1, r_2)$, any feasible solution of (MILP2) defines a subnetwork in which condition (2) is satisfied.

Proposition 7. *Let (v', a') be any feasible solution of (MILP2) augmented with the directional coupling constraints $DirC(r_1, r_2)$. Then in the subnetwork N^* defined by $supp(v')$, reaction r_1 is directionally coupled to reaction r_2.*

Proof. Let (v', a') be a feasible solution of (MILP2) augmented with $\text{DirC}(r_1, r_2)$. Assume by contradiction that in the subnetwork N^* defined by $supp(v')$, r_1 is not directionally coupled to r_2. Then there exists a flux mode w in N^* with $w_{r_1} > 0$ and $w_{r_2} = 0$. Since (v', a') is feasible, there exist $y' \in \mathbb{R}^m$, $x' \in \mathbb{R}$ such that

$$\begin{aligned} S^T y' + u^{r_1} x' &\geq -M(\mathbf{1} - a' + u^{r_2}), \\ x' &\leq -1. \end{aligned}$$

Let N' be the subnetwork obtained from N^* by deleting reaction r_2, with corresponding stoichiometric matrix S'. Removing the inequality corresponding to r_2, we get the feasible system

$$\begin{aligned} (S')^T y' + u^{r_1} x' &\geq 0, \\ x' &\leq -1. \end{aligned}$$

Applying the Farkas' Lemma, we can now derive that the system

$$\begin{aligned} S'z &= 0, \\ z_{r_1} &= 1, \\ z &\geq 0, \end{aligned}$$

TABLE 2
Algorithm 2 for the Two-Reaction Case

Step	Action
0.	Initialize $i := 1, E := \emptyset$.
1.	Try to find a feasible solution (v', a') of MILP3(E).
2.	If MILP3(E) is infeasible, then STOP.
3.	From (v', a') derive subnetwork N^i.
4.	Find a basic feasible solution e^i of LP(N^i).
5.	Let $E := E \cup \{e^i\}$ and $i := i + 1$.
6.	If $i > k$ then STOP.
7.	Go to Step 1.

is infeasible. This is a contradiction to the existence of w. We conclude that in the subnetwork N^* defined by $supp(v')$, r_1 is directionally coupled to r_2. □

By iteratively adding no-good constraints corresponding to already found EMs, we are able to enumerate any number of EMs in an increasing order of length. The resulting mixed-integer linear program can be expected to work for smaller-scale network models, but due to the difficulty of proving optimality in mixed-integer linear programs, the algorithm will most likely turn impractical for larger models. The reason for the bottle-neck is clearly that in every iteration we aim to find the shortest EM not yet discovered. Similar to the one-reaction case, we next trade the optimality condition on the length for an easier to solve program. The final form of our method is given in the mixed-integer linear program (MILP3). Table 2 summarizes the algorithm in the two-reaction case.

$$\begin{array}{rrcll}
\text{MILP3}(E^i)\text{: min} & 0 & & & \\
\text{s.t.} & Sv & = & 0, & \\
& v_{r_1} & \geq & 1, & \\
& v_{r_2} & \geq & 1, & \\
& a_l \leq v_l & \leq & M_0\ a_l, & \forall l \in \mathcal{R}, \\
& S^T y + u^{r_1} x & \geq & M_1(a - \mathbf{1} - u^{r_2}), & \\
& -x & \geq & 1, & \\
& \sum_{l \in supp(e^q)} a_l & \leq & |supp(e^q)| - 1, & \forall q < i, \\
& v_l & \geq & 0, & \forall l \in \mathcal{R}, \\
& a_l & \in & \{0, 1\}, & \forall l \in \mathcal{R}, \\
& x, y_m & \in & \mathbb{R}, & \forall m \in \mathcal{M}.
\end{array}$$

3.3 The General t-Reaction Case

Although this problem seems to be much harder at first sight, it turns out that the previous results provide all the ingredients necessary to tackle this general case. We propose two strategies for building a mixed-integer linear program that can be used in a similar fashion to (MILP3).

In the *cascade* method, we extend (MILP2) with DirC (r_1, r_2), DirC$(r_2, r_3), \ldots,$ DirC(r_{t-1}, r_t). Based on Proposition 7 and the transitivity of directional coupling [22], in any feasible solution of this new MILP, reaction r_1 will imply reactions $r_2, \ldots, r_t$, thus satisfying condition (2). Similarly, in the *hub* method, DirC(r_1, r_2), DirC$(r_1, r_3), \ldots,$ DirC(r_1, r_t) are added to (MILP2).

Alternative coupling strategies can be thought of. Indeed, by constructing any spanning tree on the vertices r_1, $r_2, \ldots r_t$, with r_1 being the root of the tree, and taking the union of the directionality constraints corresponding to each edge, we create the conditions for a subnetwork where every desired reaction is directionally coupled to r_1. At this time, it is unclear whether there is a practical advantage in choosing one of these methods compared to the others. In all cases, the total number of constraints (and variables) is the same and grows linearly with t.

3.4 Flux Uncoupling

The computational complexity of the problem to decide for a pair of uncoupled reactions whether they are sometimes coupled or mutually exclusive [20] is NP-complete [18]. However, this problem can be seen as a special case of Section 3.2. Indeed, rather than enumerating some (or all) EMs involving these two reactions, we are merely interested in the existence of any. For this purpose, it is enough to execute Step 1 of Algorithm 2. If MILP3($\emptyset$) is feasible (resp. infeasible), we can conclude that our two reactions are sometimes coupled (resp. mutually exclusive).

While this works for any pair of reactions, the above solution might not be optimal if our goal is to find all uncoupling relations. For a given pair of reactions, rather than stopping after Step 1 of Algorithm 2, one could continue executing Steps 2-7 and potentially compute an EM, if one exists. This EM can then be used to decide not only about the uncoupling relation for the current pair of reactions, but for other pairs as well. More specifically, once an EM e has been computed for two reactions i and j, this EM can also be used to show that other pairs of reactions in $supp(e)$ are sometimes coupled. This may greatly reduce the number of pairs for which solving an MILP is necessary. Moreover, by performing FCA [22], [23] as a heuristic presolving step, many uncoupling pairs can be deduced without having to solve an MILP.

3.5 Choosing Big-M Values

One issue that we have not addressed so far is the importance of choosing right values for M_0 and M_1. From the theoretical side, we are assured about the existence of M_0 and M_1 values for which the algorithm behaves as intended. However, when implementing these algorithms, it becomes crucial to choose correct values. On one hand, we are inclined to select large constants to guarantee the correctness of the solutions. If we select constants that are not large enough, we risk cutting off feasible solutions corresponding to EMs we may interested in. On the other hand, the larger these values are, the less numerically stable the MILPs become. In the following, we analyze how setting M_0 and M_1 affects the output of the algorithms.

We note that every feasible solution (v, a) of (MILP3) satisfies $v_i \in [1, M_0]$ for $a_i = 1$. Thus, M_0 represents the largest ratio v_i/v_j that can occur for non-zero fluxes v_i and v_j.

M_1 appears only in the directional coupling constraints. For any flux vector that contains r_1 but not r_2, the directional coupling constraints will be infeasible, regardless of the choice of M_1. Thus the only bad case occurs when an EM containing both r_1 and r_2 that should be feasible is rendered infeasible by the directional coupling constraints. This happens if M_1 is chosen not large enough, such that some of the constraints that should be trivially satisfied become unsatisfiable. Let v' be an EM for which this case occurs, i.e., $v'_{r_1} \geq 0$, $v'_{r_2} \geq 0$, with infeasible $DirC(r_1, r_2)$ constraints. Let $a \in \{0,1\}^n$ be a vector corresponding to the support of v'. We investigate the following LP-relaxation of (MILP3).

$$\begin{aligned} &\min \quad 0 \\ S^T y + u^{r_1} x \quad &\geq \quad M_1(a - \mathbf{1} - u^{r_2}), \\ -x \quad &\geq \quad 1. \end{aligned}$$

Its corresponding dual LP reads

$$\begin{aligned} \max \quad w - M_1 \sum_{i \in \{i \,|\, a_i = 0\} \cup \{r_2\}} v_i& \\ Sv \quad &= \quad 0, \\ v_{r_1} - w \quad &= \quad 0, \\ v, w \quad &\geq \quad 0. \end{aligned}$$

Note that the null vector 0 is always dual feasible. Since in our assumption the primal is infeasible, the dual must be unbounded. We conclude that the primal is infeasible because $w = v_{r_1} > M_1 \sum_{i \in \{i \,|\, a_i = 0\} \cup \{r_2\}} v_i$. In order to avoid this problem, we recommend over-approximating M_1 with $M_1 \geq v_{r_1}/v_i, i \in \mathcal{R} \setminus \{r_1\}$.

4 Results and Discussion

The above described methods have been implemented in Matlab. The MILP solver of choice was Gurobi 5.0. In the following, we present several use-case scenarios that have been performed on different real-world networks. These tests were aimed at validating the correctness of the methods, and also to motivate their existence, by applying them on networks where an exhaustive EM enumeration would fail. Furthermore, we ran benchmarking tests to measure the running time of the algorithms and other statistical properties. All computations were performed using a single Intel T2600 (2.16 GHz) processor on a 32-bit Windows 7 system, with a maximum memory of 640 MB allocated to Matlab.

Checking the correctness of the resulting flux vectors is easy. A simple rank test as in [24] can prove the EM property, while checking for non-zero entries in the target reactions assures that we are indeed using them. Due to the nature of the algorithms, the EM property never gets violated. Indeed, because the last step of every method presented involves solving an LP to optimality using a Simplex-based method, the result will be a vertex of the truncated flux cone.

4.1 Use-Case Scenario 1

In [25], the authors compare graph-based pathway enumeration with EM analysis in the context of discovering pathways producing G6P (KEGG entry C00668) from AcCoA (KEGG entry C00024). The underlying network was based on the human Krebs-cycle with two possible configurations, one of which was able to display the required phenotype, while the other was not. Answering this type of questions is a perfect use-case scenario for our method. Indeed, we ran our algorithms, trying to find EMs containing the required reactions, and in both configurations we were able to replicate the answers presented by the authors.

It is important to mention that the network used in the previous study is quite small (20 metabolites and 26 reactions). EM enumeration tools work very well for models of this size. Thus, we performed the same analysis for a genome-scale reconstruction of the human metabolic model [26]. In [27] a similar analysis has been performed with the use of elementary flux patterns. A crucial difference between analyzing the small and the genome-scale models, beside their size, is that in the latter network, the two metabolites we were interested in are internal to the model. This is important for two reasons. First, the number of adjacent reactions to each metabolite can be more than one, making the selection of a pair of reactions involving G6P and AcCoA non-unique. Indeed, after eliminating the blocked reactions from the network, we still had 63 viable pairs of reactions.

Secondly, and more importantly, for two internal reactions r_1 and r_2, one cannot easily find out whether reaction r_1 is a predecessor of r_2 or vice-versa. In the case of boundary reactions, it is very natural to claim that in an EM, uptake reactions precede outgoing reactions. We could say that the products of the outgoing reactions are obtained from the substrates of the uptake reactions. A similar statement for internal reactions is not trivial. Deciding whether one internal reaction precedes another one is an unsolved problem. While our method can still identify EMs with internal reactions, the interpretation of these EMs has to be done with care.

For these reasons, two artificial transport reactions were added to the network, one importing AcCoA and the other secreting G6P. The biological interpretation of our method applied to this modified system would be the following. ”Assuming we can inject AcCoA into the network and have a method to secret G6P. Can the underlying network convert AcCoA into G6P?”. Allowing one hour of running time, we could identify six EMs with the desired property, which can be found in the supplementary material, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TCBB.2014.2343964. The number of reactions taking part in these EMs ranged from 53 to 267.

In a similar way, a possible usage of the tool would be to enumerate EMs that produce biomass, allowing the system to grow, but at the same time also produce one or more byproducts (e.g. biofuels or toxic compounds). Such a set of EMs gives us insight into the growth-coupled production capabilities of a microorganism. Alternatively, these EMs could also be used as an input for other methods, such as constrained cut set computation [28].

4.2 Use-Case Scenario 2

While the main novelty of the paper is Algorithm 2, Algorithm 1 provides an alternative method for computing EMs through a single given reaction. Compared to methods like [11], our algorithm improves running time at the cost of computing not necessarily the shortest EMs. While shorter

TABLE 3
Comparison between Shortest EMs, Shortest GFMs and Algorithm 1

Method	NoR	LI	AHD
shortest EM*	54	25-26	12.79
shortest GFMs*	132	25-37	16.21
Algorithm 1	272	25-57	26.082

(NoR)—total number of reactions involved in any of the computed modes. (LI)—length interval of the modes. (AHD)—Average Hamming distance. '' indicates numbers taken from [12].*

EMs are more likely to represent biological phenotypes that occur in practice, it is unclear if only the shortest ones should be of interest. Rezola et al. [12] introduced the concept of *generating flux modes* (GFMs) and compared the 100 shortest EMs with the 100 shortest GFMs producing lysine in the *E. coli* iAF1260 model [29]. The authors argued that even though GMFs are longer and take more time to compute, they present a more varied description of the total solution set. We extended their experiment by also considering Algorithm 1. We enumerated 100 EMs producing lysine and compared their statistical properties with the other two methods. The results are summarized in Table 3.

It becomes clear from Table 3 that Algorithm 1 computes a broader range of pathways that can be used for lysine production. While it still finds some representatives of the shortest EMs, it also detects more advanced EMs. Fig. 2 presents the length distribution of the EMs obtained by Algorithm 1.

Even if we extend our interest beyond the set of shortest EMs, one might still want to avoid unreasonably long ones (i.e., those involving more than L reactions, for some $L \geq 1$). This can easily be achieved by adding the constraint $\sum_{l \in \mathcal{R}} a_l \leq L$ to the MILP formulation.

In our next experiment, we compared Algorithm 1 with the method of [11] for computing shortest EMs. As input we used the metabolic network of *S. cerevisiae iND750* [30] with ethanol (R_ETOH) being the target reaction, required to be active in all EMs. For various M values and choosing the flux variables to be either integer or continuous, 20 EMs were computed in each case. Table 4 summarizes the results.

TABLE 4
Comparing Algorithm 1 with the Computation of Shortest EMs (Running Time in Secs)

		Integer variables		Continuous variables	
	Method	Length	Time	Length	Time
M = 10	Shortest	6-10	1719 s	6-10	2074 s
	Algorithm 1	6-15	16 s	6-23	15 s
M = 100	Shortest	6-10	8158 s	6-10	3421 s
	Algorithm 1	6-21	21 s	6-31	18 s
M = 1000	Shortest	6-10	14362 s	6-10	7780 s
	Algorithm 1	6-31	16 s	6-50	29 s

Table 4 highlights the main differences in the two methods. Algorithm 1 finds EMs much quicker at the expense of individual EMs' length. As expected, in almost every case, the higher the constant M was chosen, the more expensive the computation becomes. De Figueiredo et al. [11] remark that using continuous flux variables, the MILPs might be more time consuming to solve. However, when computing the shortest EMs in this particular network, we do not consistently observe this behavior. Indeed, while for $M = 10$, choosing the flux variables as continuous proved to be slightly more expensive, in all other cases continuous variables led to shorter running times.

In every iteration of Algorithm 1, the MILPs to be solved become more complex, since they contain additional constraints. Over time, these constraints are expected to slow down the solving of (MILP2). Hence, the rate at which new EMs are computed is expected to decrease, the more EMs we enumerate. To study this effect, we enumerated 1,000 EMs in the network of *S. cerevisiae iND750* [30], with ethanol (R_ETOH) being the target reaction. The flux variables were chosen as continuous, while a value for M of 1,000 was used. The total running time was 4,350 seconds. For reference, in the same time frame we could compute only 16 EMs with the method presented by [11]. We measured the time t_i required for the computation of the ith EM, for all $i \in \{1, \ldots, 1{,}000\}$. In order to study how the lengths of consecutive EMs vary, we also measured the length l_i of the ith EM, for all $i \in \{1, \ldots, 1{,}000\}$. In Fig. 3, we display the evolu-

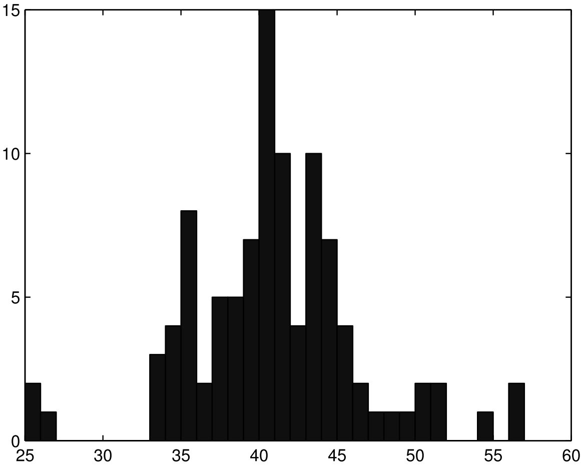

Fig. 2. Histogram of the length of EMs found by Algorithm 1.

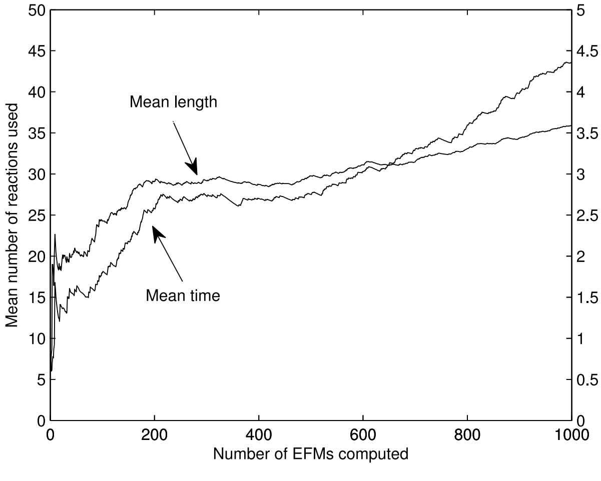

Fig. 3. Mean time and mean length of the EMs computed.

TABLE 5
Summary of the Correct Computational Results

Netw.	Reac.	Pairs	EM found	Mean length	No EM
ECC	90	8,010	7,691	24.36	176
HP	269	72,092	66,749	46.57	1,862

tion of the mean times $\bar{t}_i := (\sum_{j=1}^{i} t_j)/i$ together with the mean lengths $\bar{l}_i := (\sum_{j=1}^{i} l_j)/i$.

4.3 Statistical Analysis and Flux Uncoupling

The main bottleneck of the method are cases when EMs of the required type do not exist. While finding an EM if it exists seemed to work well in practice, proving their non-existence (i.e., showing (MILP3) to be infeasible) is rather hard and time-consuming. Intuitively, one can think of the MILPs search-tree. Depending on how many feasible solutions there are, the solver might find one without traversing the whole tree. For proving the non-existence, the whole search-tree must be traversed.

In the next experiment, we tried to compute an EM for every pair of reactions. A maximum time of 60 seconds was allocated for each pair. Reaching the timeout meant that we were unable to compute if this pair shared an EM or not. The test has been performed on two small to medium-sized real world networks, the central metabolism of *E. coli* (ECC) [31] and the *H. pylori* (HP) [32] genome-scale metabolic network. For the constants M_0 and M_1, the value of 10,000 was chosen, while the feasibility tolerance was set to 10^{-6}. The total running time in the case of ECC was approximately 3 hours, while the algorithm took close to 3 days in the case of HP.

Table 5 empirically sheds light on the nature of real-world networks. One can observe in these networks that most pairs of reactions share at least one EM, and the algorithm presented in this paper is able to find them.

From Table 6, it becomes clear that (although not many) there are cases where the algorithm does not produce a relevant result, either by not finishing before its time-out, or by producing an erroneous result. The latter cases are attributed to the incorrect choice of M_0 and M_1 and numerical imprecisions in the MILP solver.

Seeing the high probability for the existence of an EM containing two randomly chosen reactions, the question arises whether solving (MILP2) would suffice. If it correctly finds an EM in most cases, this would motivate using (MILP2) as a heuristic approach to the problem. Unless the two reactions are blocked, (MILP2) is always feasible, and the optimal solution is characterized by Corollary 3 which asserts that we either get an EM or a false positive. We performed the same experiment as before, and attempted to compute a flux vector for every pair of reactions. We decided about the EM property of the computed flux vectors by applying the rank test [7, Lemma 2]. For every instance, a maximum of 60 seconds was allowed. Table 7 summarizes these results.

TABLE 6
Summary of the Bad Instances

Network	Timed out		Numerical error	
	Absolute	Relative	Absolute	Relative
ECC	78	0.97%	65	0.81%
HP	2,078	2.88%	1,403	1.94%

TABLE 7
The Performance of (MILP2) on Metabolic Networks

Network	EM found	False positive	Timed out
ECC	5,212	2,686	202
HP	206	9,213	62,673

It turns out that in the case of the small *E. coli* network, (MILP2) performs reasonably well. In 67percent of the cases where EMs exist, it correctly finds one. However, for the medium-scale *H. pylori* network, solving (MILP2) to optimality almost never terminates in the allocated time. Based on this empirical evidence we conclude that (MILP2) may not be a viable approach for medium- to large-scale metabolic networks.

5 Conclusion

We have presented novel methods to compute EMs involving any number of predefined target reactions. These algorithms can also be used to distinguish between mutually exclusive and sometimes coupled reactions. From the application on genome-scale metabolic networks, we conclude that the methods work as intended and are fast enough for practical use. They should become a valuable asset for constraint-based analysis of metabolic networks.

A prototype implementation in Matlab is available for download at https://sourceforge.net/projects/caefm.

Acknowledgments

The authors would like to thank Sayed-Amir Marashi for constructive discussions on the experiments. The proof of Prop. 5 is based on ideas from Yaron Goldstein. They would also like to thank Luis de Figueiredo for providing us the network model from [25].

References

[1] S. Schuster and C. Hilgetag, “On elementary flux modes in biochemical reaction systems at steady state,” *J. Biol. Syst.*, vol. 2, pp. 165–182, 1994.

[2] S. Schuster, T. Dandekar, and A. Fell, “Detection of elementary flux modes in biochemical networks: A promising tool for pathway analysis and metabolic engineering,” *Trends Biotechnol.*, vol. 17, pp. 53–60, 1999.

[3] S. Schuster, D. A. Fell, and T. Dandekar, “A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks,” *Nature Biotechnol.*, vol. 18, pp. 326–332, 2000.

[4] S. Schuster, C. Hilgetag, J. H. Woods, and P. A. Fell, “Reaction routes in biochemical reaction systems: Algebraic properties, validated calculation procedure and example from nucleotide metabolism,” *J. Math. Biol.*, vol. 45, no. 2, pp. 153–181, 2002.

[5] A. J. Papin, J. Stelling, N. D. Price, S. Klamt, S. Schuster, and B. O. Palsson, “Comparison of network-based pathway analysis methods,” *TRENDS Biotechnol.*, vol. 22, no. 8, pp. 400–405, 2004.

[6] J. Zanghellini, D. E. Ruckerbauer, M. Hanscho, and C. Jungreuthmayer, "Elementary flux modes in a nutshell: Properties, calculation and applications," *Biotechnol. J.*, vol. 8, pp. 1009–1016, Jun. 2013.

[7] J. Gagneur and S. Klamt, "Computation of elementary modes: A unifying framework and the new binary approach," *BMC Bioinformatics*, vol. 5, p. 175, 2004.

[8] A. von Kamp and S. Schuster, "Metatool 5.0: Fast and flexible elementary modes analysis," *Bioinformatics*, vol. 22, pp. 1930–1931, 2006.

[9] M. Terzer and J. Stelling, "Large-scale computation of elementary flux modes with bit pattern trees," *Bioinformatics*, vol. 24, pp. 2229–2235, 2008.

[10] M. Terzer. (2009). Large scale methods to enumerate extreme rays and elementary modes. PhD dissertation, Swiss Federal Inst. Technol., Zurich, Switzerland [Online]. Available: http://e-collection.library.ethz.ch/eserv/eth:534/eth-534-02.pdf

[11] L. F. de Figueiredo, A. Podhorski, A. Rubio, C. Kaleta, J. E. Beasley, S. Schuster, and F. J. Planes. (2009). Computing the shortest elementary flux modes in genome-scale metabolic networks. *Bioinformatics* [Online]. *25(23)*, pp. 3158–3165. Available: http://bioinformatics.oxfordjournals.org/content/25/23/3158.abstract

[12] A. Rezola, L. F. de Figueiredo, M. Brock, J. Pey, A. Podhorski, C. Wittmann, S. Schuster, A. Bockmayr, and F. J. Planes. (2011). Exploring metabolic pathways in genome-scale networks via generating flux modes. *Bioinformatics* [Online]. *27(4)*, pp. 534–540. Available: http://bioinformatics.oxfordjournals.org/content/27/4/534.abstract

[13] A. Rezola, J. Pey, L. F. de Figueiredo, A. Podhorski, S. Schuster, A. Rubio, and F. J. Planes, "Selection of human tissue-specific elementary flux modes using gene expression data," *Bioinformatics*, vol. 29, no. 16, pp. 2009–2016, 2013.

[14] V. Acuña, F. Chierichetti, V. Lacroix, A. Marchetti-Spaccamela, M. Sagot, and L. Stougie, "Modes and cuts in metabolic networks: Complexity and algorithms," *BioSystems*, vol. 95, pp. 51–60, 2009.

[15] S.-A. Marashi, L. David, and A. Bockmayr. (2012, Jan.). Analysis of metabolic subnetworks by flux cone projection. *Algorithms Molecular Biol.: AMB* [Online]. *7(1)*, p. 17. Available: http://www.almob.org/content/7/1/17

[16] C. Kaleta, L. F. de Figueiredo, and S. Schuster, "Can the whole be less than the sum of its parts? Pathway analysis in genome-scale metabolic networks using elementary flux patterns," *Genome Res.*, vol. 19, pp. 1872–1883, 2009.

[17] R. Urbanczik and C. Wagner, "Functional stoichiometric analysis of metabolic networks," *Bioinformatics*, vol. 21, pp. 4176–4180, 2005.

[18] V. Acuña, A. Marchetti-Spaccamela, M. Sagot, and L. Stougie, "A note on the complexity of finding and enumerating elementary modes," *BioSystems*, vol. 99, pp. 210–214, 2010.

[19] A. P. Burgard, E. V. Nikolaev, C. H. Schilling, and C. D. Maranas, "Flux coupling analysis of genome-scale metabolic network reconstructions," *Genome Res.*, vol. 14, pp. 301–312, 2004.

[20] S.-A. Marashi and A. Bockmayr. (2011). Flux coupling analysis of metabolic networks is sensitive to missing reactions. *Biosystems* [Online]. *103(1)*, pp. 57–66. Available: http://www.sciencedirect.com/science/article/pii/S0303264710001711

[21] A. Schrijver, *Theory of Linear and Integer Programming*. New York, NY, USA: Wiley, 1998.

[22] A. Larhlimi, L. David, J. Selbig, and A. Bockmayr. (2012, Jan.). F2C2: A fast tool for the computation of flux coupling in genome-scale metabolic networks. *BMC Bioinformat.* [Online]. *13(1)*, p. 57. Available: http://www.biomedcentral.com/1471-2105/13/57

[23] L. David, S.-A. Marashi, A. Larhlimi, B. Mieth, and A. Bockmayr. (2011, Jan.). FFCA: A feasibility-based method for flux coupling analysis of metabolic networks. *BMC Bioinformatics* [Online]. *12 (1)*, p. 236. Available: http://www.biomedcentral.com/1471-2105/12/236

[24] A. V. Kamp and S. Schuster. (2006). Metatool 5.0: Fast and flexible elementary modes analysis. *Bioinformatics* [Online]. *22(15)*, pp. 1930–1931. Available: http://bioinformatics.oxfordjournals.org/content/22/15/1930.abstract

[25] L. F. de Figueiredo, S. Schuster, C. Kaleta, and D. A. Fell. (2008). Can sugars be produced from fatty acids? a test case for pathway analysis tools. *Bioinformatics* [Online]. *24(22)*, pp. 2615–2621. Available: http://bioinformatics.oxfordjournals.org/content/24/22/2615.abstract

[26] N. C. Duarte, S. A. Becker, N. Jamshidi, I. Thiele, M. L. Mo, T. D. Vo, R. Srivas, and B. Palsson. (2007). Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Nat. Academy Sci.* [Online]. *104(6)*, pp. 1777–1782. Available: http://www.pnas.org/content/104/6/1777.abstract

[27] C. Kaleta, L. F. de Figueiredo, S. Werner, R. Guthke, M. Ristow, and S. Schuster. (2011). In silico evidence for gluconeogenesis from fatty acids in humans. *PLoS Comput. Biol.* [Online]. *7(7)*. Available: http://dblp.uni-trier.de/db/journals/ploscb/ploscb7.html#KaletaFWGRS11

[28] O. Hädicke and S. Klamt. (2011). Computing complex metabolic intervention strategies using constrained minimal cut sets. *Metabolic Eng.* [Online]. *13(2)*, pp. 204–213. Available: http://www.sciencedirect.com/science/article/pii/S1096717610001084

[29] A. M. Feist, C. S. Henry, J. L. Reed, M. Krummenacker, A. R. Joyce, P. D. Karp, L. J. Broadbelt, V. Hatzimanikatis, and B. Ø. Palsson, "A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information," *Mol. Syst. Biol.*, vol. 3, no. 121, Jun. 2007.

[30] N. C. Duarte, M. J. Herrgård, and B. Ø. Palsson, "Reconstruction and validation of saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model," *Genome Res.*, vol. 14, no. 7, pp. 1298–1309, Jul. 2004.

[31] B. O. Palsson, *Systems Biology: Properties of Reconstructed Networks*. New York, NY, USA: Cambridge Univ. Press, 2006.

[32] I. Thiele, T. D. Vo, N. D. Price, and B. O. Palsson, "Expanded metabolic reconstruction of *Helicobacter pylori* (iIT341 GSM/GPR): An in silico genome-scale characterization of single- and double-deletion mutants," *J. Bacteriol.*, vol. 187, pp. 5818–5830, 2005.

Laszlo David received the MS degree in computer science in 2006 from the Babeş-Bolyai University, Cluj-Napoca, Romania. He is currently working toward the PhD degree at Freie Universität Berlin. His research interests include constraint-based modeling of metabolic networks and optimization.

Alexander Bockmayr is a full professor of Mathematics at Freie Universität Berlin, where he holds the chair for Mathematics in Life Sciences since 2004. His current research focuses on new approaches for mathematical and computational modeling of metabolic and regulatory networks, with special emphasis on constraint-based methods. His mathematical background lies in discrete mathematics and optimization.