Flux Modules in Metabolic Networks

Arne C. Müller [1,2,4,*], Alexander Bockmayr [1,3]

Abstract

The huge number of *elementary flux modes* (EFM) in genome-scale metabolic networks makes analysis based on elementary flux modes intrinsically difficult. However, it has been shown that the elementary flux modes with optimal yield often contain highly redundant information. The set of optimal-yield elementary flux modes can be compressed using modules. Up to now, this compression was only possible by first enumerating the whole set of all optimal-yield elementary flux modes.

We present a direct method for computing modules of the thermodynamically constrained optimal flux space of a metabolic network. This method can be used to decompose the set of optimal-yield elementary flux modes in a modular way and to speed up their computation. In addition, it provides a new form of coupling information that is not obtained by classical flux coupling analysis. We illustrate our approach on a set of model organisms.

Keywords: Metabolic Network, Flux Balance Analysis, Elementary Flux Modes, Modules

1 Introduction

Constraint-based methods have provided a very successful means to analyze genome-scale metabolic network reconstructions (Papin *et al.*, 2004; Price *et al.*, 2004). Instead of predicting explicit flux distributions, constraints are used to characterize biologically feasible phenomena. For metabolic networks, the most typical constraint is the steady-state assumption $Sv = 0$, where S denotes the stoichiometric matrix and v the flux vector. It is based on the assumption that no internal metabolite is allowed to accumulate or degenerate. Additional constraints may be thermodynamic irreversibilities $v_{\text{Irrev}} \geq 0$ (here Irrev denotes the set of irreversible reactions), bounds on nutrient uptake, etc.

To understand the *feasible flux space*, i.e., the set of all flux vectors satisfying the given constraints, various methods have been proposed. *Elementary modes* (Schuster and Hilgetag, 1994; Schuster *et al.*, 2000), *extreme pathways* (Schilling *et al.*, 2000), and *minimal metabolic behaviors* (Larhlimi and Bockmayr, 2009) have been developed to describe all pathways of the network. Unfortunately, the number of elementary flux modes, extreme pathways or minimal metabolic behaviors usually grows exponentially with the size of the network. Therefore, it has been impossible so far to perform this kind of analysis for large genome-scale models. Also, even small networks pose the problem that so many pathways are generated that understandable results can only be obtained using additional post-processing steps. Hence, other methods have been developed that just study specific characteristics of the network. *Flux variability analysis* (FVA) (Burgard *et al.*, 2001; Mahadevan and Schilling, 2003) is used to study the range of flux rates in which reactions can operate. *Flux coupling analysis* (FCA) (Burgard *et al.*, 2004; Larhlimi *et al.*, 2012) and correlation analysis (Papin and Palsson, 2004; Sarıyar *et al.*, 2006) study dependencies between fluxes through pairs of reactions.

*to whom correspondence should be addressed

[1] Department of Mathematics and Computer Science, Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany

[2] International Max Planck Research School for Computational Biology and Scientific Computing (IMPRS-CBSC), Max Planck Institute for Molecular Genetics, Ihnestr 63-73, D-14195 Berlin, Germany

[3] DFG-Research Center Matheon, Berlin, Germany

[4] Berlin Mathematical School (BMS), Berlin, Germany

Another possibility is to reduce the feasible flux space by an optimality criterion. Optimization-based methods like *flux balance analysis* (FBA) (Varma and Palsson, 1994; Mahadevan and Schilling, 2003; Price *et al.,* 2004; Terzer *et al.,* 2009; Durot *et al.,* 2009; Schuster *et al.,* 2007; Teusink *et al.,* 2009; Orth *et al.,* 2010) formulate queries on the model and can give quantitative results. A typical objective function asks for the maximal possible biomass production yield (Feist and Palsson, 2010), i.e., we consider the optimisation problem

$$\text{opt} := \max\{v_{\text{Biomass}} : Sv = 0,\ \ell \le v \le u\},$$

where ℓ and u denote lower and upper flux bounds. This can efficiently be solved with linear programming. A deeper analysis, however, requires the study of all the flux vectors realizing the optimal objective value. Usually, we do not get a single optimal flux vector (Khannapho *et al.*, 2008), but again a set of optimal solutions, which we call the *optimal flux space*:

$$P_{opt} := \{v : Sv = 0,\ \ell \le v \le u,\ v_{\text{Biomass}} = \text{opt}\}.$$

This space is mathematically very similar to the feasible flux space. Hence, adaptations of the above methods can also be used to study the optimal flux space.

Kelk *et al.* (2012) analyzed the optimal flux space using the theory of polyhedra (see e.g. Grünbaum (2003)). After removing linealities and rays (which are caused by thermodynamically infeasible cycles), they obtain a polytope. They observe that this polytope can be decomposed into *modules*. Their decomposition has the nice property that the modules can be analyzed independently and together they provide a comprehensive understanding of the whole polytope. In other words, the modules allow a losslessly compressed description of the extreme points of the polytope of optimal fluxes.

The contributions of this paper are the following:

- We introduce a mathematical framework for studying such modules. In particular, we give a simple mathematical definition of a module. We prove that it is sufficient to analyze the modules independently from each other to understand the optimal flux space of the whole network.
- We derive an alternative method to determine these modules, which does not require the removal of linealities and rays. An important difference to (Kelk *et al.*, 2012) is that we do not need to compute all optimal-yield elementary flux modes in our method. Since the number of elementary flux modes grows exponentially with the size of the network, this result is major algorithmic improvement.
- We present a novel analysis method for metabolic networks, which we call *blocking graphs*. These graphs contain information such as "*if flux through reaction r is fixed to its maximal value, then reaction s cannot carry any flux*". We show that this gives useful additional information that complements standard flux coupling analysis.
- We show how the modules can be used to compute the set of optimal-yield elementary flux modes.

The paper is organised as follows. We start in Sect. 2 with an intuitive description of our method. In Sect. 3, we develop the necessary mathematical theory. The corresponding computational methods are derived and presented in Sect. 4. In Sect. 5 we apply these methods on different genome-scale networks and discuss the results. Finally, in Sect. 5.2, we compare our method with the method of Kelk *et al.* (2012).

2 Approach

When studying the minimal and maximal possible flux through reactions in the optimal flux space, one usually observes that many reactions do not show any flux variability at all. Hence, these reactions must be part of any optimal flux vector. In the following, we are interested in those reactions that show flux variability in the optimal flux space. Here, the organism has abilities to adjust to environment changes or knockouts without changing its biomass production rate.

As an example, let us consider the network shown in Fig. 1. Assume that the restriction to optimal fluxes enforces a flux of 1 through reaction r_1. As one can easily see, this implies that reactions r_{13}, r_{14}, r_{15} also have a fixed flux. We now want to find out how the remaining (unfixed) reactions group into modules.

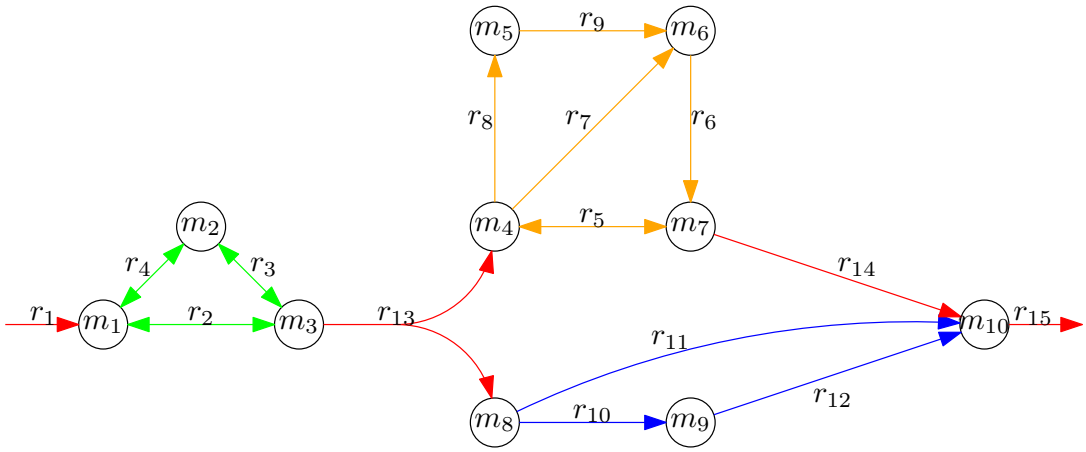

Figure 1: All stoichiometric coefficients in this example are 1. Assume flux through reaction r_1 is fixed to 1. Then flux through reactions $(r_1, r_{13}, r_{14}, r_{15})$ is fixed and we get the three modules (r_2, r_3, r_4), $(r_5, r_6, r_7, r_8, r_9)$, and (r_{10}, r_{11}, r_{12}).

Informally speaking, we consider a set of reactions to be a *module* if they behave together like one reaction with a fixed flux. This means that the rate (or interface flux) of metabolites that are produced and consumed by the reactions in the module must be constant for all optimal flux vectors. For example the reactions r_{10}, r_{11}, r_{12} form a module, since those reactions together always take up 1 of metabolite m_8 and always produce 1 of m_{10}. The interface flux of this module is $(0,0,0,0,0,0,0,-1,0,1)$.

The algorithm for finding modules follows the following idea: Suppose we fix flux through reaction r_{11} to its maximal value (in this case $v_{r_{11}} = 1$). It follows that there is no flux possible anymore through reactions r_{10}, r_{12}. The other reactions, however, are not affected by this choice at all. This is a hint that reactions r_{10}, r_{11}, r_{12} belong to the same module, while the other reactions are independent.

However, this approach requires that for each reaction we must be able to fix the flux to its maximal value. This is not possible if we just do standard flux variability analysis (FVA). Let us consider reaction r_9. Since this reaction is part of an internal cycle (r_9, r_6, r_5, r_8), we can send unbounded flux through r_9. Hence, we cannot set flux through this reaction to its maximal flux rate and observe the effects. To cope with this problem, we use *thermodynamically constrained flux variability analysis* (tFVA) (Schellenberger *et al.*, 2011; Müller and Bockmayr, 2013). Since thermodynamic constraints prohibit fluxes along internal cycles, these reactions obtain a finite maximal flux rate. We can now use this flux rate to perform our analysis.

3 Mathematical Theory

In this section, we develop the mathematical theory underlying our approach. The corresponding analysis methods will be described in Sect. 4.

3.1 Preliminaries

A metabolic network $\mathcal{N} = (\mathcal{M}, \mathcal{R}, \mathcal{I}, S, b, \ell, u)$ is given by a finite set of *metabolites* $\mathcal{M}$, a finite set of *reactions* $\mathcal{R}$, a subset $\mathcal{I} \subseteq \mathcal{R}$ of *internal reactions*, the *stoichiometric matrix* $S \in \mathbb{R}^{\mathcal{M}\times\mathcal{R}}$, and vectors $b \in \mathbb{R}^{\mathcal{M}}$, $\ell, u \in \mathbb{R}^{\mathcal{R}}_{\pm\infty}$, where $\mathbb{R}_{\pm\infty} := \mathbb{R} \cup \{-\infty, +\infty\}$. Reactions in $\mathcal{E} := \mathcal{R} \setminus \mathcal{I}$ are *exchange reactions* that can supply the network with inflow or outflow from the environment. A *steady-state flux* is a vector $v \in \mathbb{R}^{\mathcal{R}}$ that satisfies $Sv = 0$. More generally, we consider flux vectors $v \in \mathbb{R}^{\mathcal{R}}$ satisfying $Sv = b$, for some right-hand side $b \in \mathbb{R}^{\mathcal{M}}$. In addition we will use *lower* and *upper bounds* $\ell, u \in \mathbb{R}^{\mathcal{R}}_{\pm\infty}$, for which we require $\ell \le v \le u$. By abuse of notation, $v_r \le \infty$ resp. $-\infty \le v_r$ means that v_r is unbounded from above resp. below. We use subscripts to index flux through reactions, i.e., v_r denotes flux through reaction r.

We will also be interested in the flux through a set of reactions $A \subseteq \mathcal{R}$. Hence, we write v_A to denote the components of v corresponding to the reactions in A and we use S_A to denote the stoichiometric matrix that only contains the columns corresponding to the reactions in A.

A set of vectors $P \subseteq \mathbb{R}^{\mathcal{R}}$ is called a *flux space*. For $A \subseteq \mathcal{R}$, we define the projection $\mathrm{pr}_A(P) := \{v_A : v \in P\} \subseteq \mathbb{R}^A$. For disjoint sets $A, B \subseteq \mathcal{R}$ and $X \subseteq \mathbb{R}^A, Y \subseteq \mathbb{R}^B$ we define $X \times Y := \{v \in \mathbb{R}^{A \dot\cup B} : v_A \in X, v_B \in Y\}$. Here, $\dot\cup$ denotes the union of disjoint sets. Finally, we use $\subseteq$ to denote subsets with equality, and $\subset$ for subsets without equality.

3.2 Thermodynamically feasible fluxes

Following Beard *et al.* (2004), we introduce the notion of a thermodynamically feasible flux.

Definition 1 (Thermodynamically Feasible Flux) A flux vector $v \in \mathbb{R}^{\mathcal{R}}$ is *thermodynamically feasible* (thermo. feasible) if there exists a vector $\mu \in \mathbb{R}^{\mathcal{M}}$ s.t.

$$\mu S_r v_r < 0 \text{ or } v_r = 0, \qquad \text{for every internal reaction } r \in \mathcal{I}.$$

Note that other authors, e.g. Fleming *et al.* (2012), use a slightly stronger definition where $v_r = 0$ is not always allowed. As shown in (Beard *et al.*, 2004; Noor *et al.*, 2012; Müller, 2012), Def. 1 can be equivalently restated as follows:

Theorem 1 (Looplaw) *A flux vector $v \in \mathbb{R}^{\mathcal{R}}$ is thermodynamically feasible if and only if there exists no $w \in \mathbb{R}^{\mathcal{I}} \setminus \{0\}$ with*

$$\begin{aligned} S_{\mathcal{I}} w &= 0 \\ w_r &\ge 0 \qquad \forall r \in \mathcal{I} \text{ with } v_r \ge 0, \\ w_r &\le 0 \qquad \forall r \in \mathcal{I} \text{ with } v_r \le 0. \end{aligned}$$

From this theorem, we observe the importance of steady-state fluxes that use only internal reactions.

Definition 2 (Internal Cycle) We call $w \in \mathbb{R}^{\mathcal{I}} \setminus \{0\}$ an *internal cycle* if it satisfies

$$S_{\mathcal{I}} w = 0, \quad \ell_{\mathcal{I}} \le w \le u_{\mathcal{I}}.$$

The *set of reactions contained in internal cycles* is defined as

$$\mathcal{C} := \{r \in \mathcal{I} : w_r \ne 0, \text{ for some internal cycle } w\}.$$

We say that an internal cycle w is *contained in a flux vector* $v \in \mathbb{R}^{\mathcal{R}}$ if

$$\operatorname{sign}(w) \subseteq \operatorname{sign}(v),$$

where for vectors $x \in \mathbb{R}^A$ and sets X^+, X^-, Y^+, Y^-, we define

$$\operatorname{sign}(x) := (\{r \in A : x_r > 0\}, \{r \in A : x_r < 0\}),$$
$$(X^+, X^-) \subseteq (Y^+, Y^-) :\Leftrightarrow X^+ \subseteq Y^+ \wedge X^- \subseteq Y^-.$$

Using this notion, we can say that a flux vector is thermodynamically feasible if and only if it does not contain an internal cycle.

3.3 Decomposition of Flux Spaces into Modules

The main goal of this paper is the decomposition of flux spaces into modules. We first propose a definition of module for general flux spaces $P \subseteq \mathbb{R}^{\mathcal{R}}$.

Definition 3 (Module) A set of reactions $A \subseteq \mathcal{R}$ is called a *module w.r.t. a flux space* $P \subseteq \mathbb{R}^{\mathcal{R}}$, or shortly A is called a *P-module*, if there exists a vector $d \in \mathbb{R}^{\mathcal{M}}$ with $S_A v_A = d$ for all flux vectors $v \in P$. The vector d is called the *right-hand side* of the module A. Since d operates as the interface of the module to the rest of the network, we refer to d also as the *interface flux* of A. □

We observe that if and only if $P \neq \emptyset$ then the interface flux of each P-module is well defined. Hence, all of the following theorems will require $P \neq \emptyset$. If $P \subseteq \{v \in \mathbb{R}^{\mathcal{R}} : Sv = b\}$, then $\mathcal{R}$ is itself a P-module. We also observe that, given two disjoint P-modules A and B, their union is again a P-module:

Lemma 1 *Let $P \subseteq \mathbb{R}^{\mathcal{R}}$ be an arbitrary flux space and let A and B be disjoint P-modules with right-hand sides d^A, d^B respectively. Then $A \dot\cup B$ is also a P-module with right-hand side $d^A + d^B$.* □

PROOF For every $v \in P$, we have $S_A v_A = d^A$ and $S_B v_B = d^B$, which implies $S_{A\dot\cup B} v_{A\dot\cup B} = S_A v_A + S_B v_B = d^A + d^B$. Hence, $A \dot\cup B$ is a P-module with right-hand side $d^A + d^B$. ■

We now focus on flux spaces consisting of steady-state fluxes and the more restricted flux spaces, where the fluxes are also thermodynamically feasible.

Definition 4 (Thermodynamically constrained flux space) Given a metabolic network $\mathcal{N} = (\mathcal{M}, \mathcal{R}, \mathcal{I}, S, b, \ell, u)$, the *thermodynamically constrained flux space* T is defined as

$$T := \{v \in \mathbb{R}^{\mathcal{R}} : Sv = b,\ \ell \leq v \leq u,\ v \text{ thermo. feasible}\}. \tag{1}$$

If $T \neq \emptyset$, we similarly define for T-modules $A \subseteq \mathcal{R}$ with right-hand side d

$$T^A := \left\{v \in \mathbb{R}^A :\ \ S_A v = d,\ \ell_A \leq v \leq u_A, v \text{ thermo. feasible}\ \right\}, \tag{2}$$

and in addition

$$\tilde{T}^A := \operatorname{pr}_A(T). \tag{3}$$

Definition 5 (Steady-state flux space) Given a metabolic network $\mathcal{N} = (\mathcal{M}, \mathcal{R}, \mathcal{I}, S, b, \ell, u)$, the *steady-state flux space* F is defined as

$$F := \{v \in \mathbb{R}^{\mathcal{R}} : Sv = b,\ \ell \leq v \leq u\}. \tag{4}$$

For F-modules (if $F \neq \emptyset$) and T-modules (if $T \neq \emptyset$) $A \subseteq \mathcal{R}$ with right-hand side d we define

$$F^A := \{v \in \mathbb{R}^A : S_A v = d,\ \ell_A \leq v \leq u_A\}. \tag{5}$$

Observe that F^A is also well defined if A is both a F-module and a T-module (if $T \neq \emptyset$), because $T \subseteq F$ and thus the right-hand sides must coincide. We will later see (Application of Cor. 1 with $\mathcal{I} = \emptyset$) that for F-modules A we have $F^A = \mathrm{pr}_A(F)$. While $\tilde{T}^A \subseteq T^A$ holds for any T-module A, we do not always have $\tilde{T}^A = T^A$ (see however Cor. 1 and the supplementary material for an example).

In what follows, we will use P to denote general flux spaces, while T and F always denote the flux spaces defined above.

Before we study the decomposition of the flux space T, we first investigate the decomposition of the flux space of a union of two T-modules.

Lemma 2 *Assume $T \neq \emptyset$. Then for any disjoint T-modules A and B we have*

$$\tilde{T}^{A\dot{\cup}B} \subseteq \tilde{T}^A \times T^B \subseteq T^A \times T^B \subseteq F^{A\dot{\cup}B}.$$

If $\ell_{\mathcal{C}} \leq 0 \leq u_{\mathcal{C}}$, then $\tilde{T}^{A\dot{\cup}B} = T^A \times T^B$. □

PROOF By the definition of module, we have $\tilde{T}^{A\dot{\cup}B} \subseteq \tilde{T}^A \times \tilde{T}^B$ and $\tilde{T}^A \subseteq T^A, \tilde{T}^B \subseteq T^B$. Hence, we get $\tilde{T}^{A\dot{\cup}B} \subseteq \tilde{T}^A \times \tilde{T}^B \subseteq \tilde{T}^A \times T^B \subseteq T^A \times T^B$. Let d^A, d^B denote the right-hand sides of the T-modules A and B respectively. By Lemma 1, it follows that $A\dot{\cup}B$ is a T-module with right-hand side $d^A + d^B$.

Let $x \in T^A, y \in T^B$ be fixed but arbitrary. Define $v' \in \mathbb{R}^{A\dot{\cup}B}$ by $v'_A := x$ and $v'_B := y$. Since A and B are T-modules, we have $S_{A\dot{\cup}B}v' = d^A + d^B$. Since $\ell_A \leq x \leq u_A$ and $\ell_B \leq y \leq u_B$, it follows that $v' \in F^{A\dot{\cup}B}$. Hence, $T^A \times T^B \subseteq F^{A\dot{\cup}B}$.

Now we continue with the case $\ell_{\mathcal{C}} \leq 0$ and $u_{\mathcal{C}} \geq 0$. Since $T \neq \emptyset$, there exists $w' \in T$. Define $w \in \mathbb{R}^{\mathcal{R}}$ by $w_{A\dot{\cup}B} := v'$ and $w_{\mathcal{R}\setminus(A\dot{\cup}B)} := w'_{\mathcal{R}\setminus(A\dot{\cup}B)}$. Since $A\dot{\cup}B$ is a T-module, it follows that $w \in F$, as defined in (4). Let v be the flux vector obtained from w by subtracting all contained internal cycles using Alg. 1 from Müller and Bockmayr (2013). Since $\ell_{\mathcal{C}} \leq 0 \leq u_{\mathcal{C}}$, it follows by Thm. 2 in Müller and Bockmayr (2013) that $v \in T$ and $\mathrm{sign}(w - v) \subseteq \mathrm{sign}(w)$.

Since A is a T-module, we have $S_A v_A = d^A$. From $S_A x - S_A v_A = d^A - d^A = 0$, we get $S_A(x - v_A) = 0$. If $x - v_A \neq 0$, it follows that x would have contained an internal cycle (note that $\mathrm{sign}(x - v_A) \subseteq \mathrm{sign}(x)$). This is a contradiction and hence, $v_A = x$. By the same argument, we can show $v_B = y$. Since $v \in T$, we obtain $v' = v_{A\dot{\cup}B} \in \tilde{T}^{A\dot{\cup}B}$ and since we can do this for every $x \in T^A$ and $y \in T^B$, we get $\tilde{T}^{A\dot{\cup}B} \supseteq T^A \times T^B$. Therefore, $\tilde{T}^{A\dot{\cup}B} = T^A \times T^B$. ■

Corollary 1 *If $T \neq \emptyset$ with $\ell_{\mathcal{C}} \leq 0 \leq u_{\mathcal{C}}$ and A is a T-module, then $T^A = \tilde{T}^A = pr_A(T)$.* □

PROOF Let $B := \mathcal{R} \setminus A$. Since A is a T-module and $T \subseteq \{v \in \mathbb{R}^{\mathcal{R}} : Sv = b\}$, B is a T-module, too. With Lemma 2, it follows $\tilde{T}^{A\dot{\cup}B} = T = T^A \times T^B$ and by projection $T^A = pr_A(T) = \tilde{T}^A$. ■

Using Lemma 2, we can now show by induction that from a partition of the reaction set $\mathcal{R}$ into a set of T-modules, we can get a decomposition of the thermodynamically constrained flux space T.

Theorem 2 (Product Space from Modules) *Assume $T \neq \emptyset$. If $\mathcal{X} = \{A_1, \ldots, A_n\}$ is a partition of $\mathcal{R}$ into T-modules, then*

$$T \subseteq \prod_{A \in \mathcal{X}} T^A \subseteq F.$$

If $\ell_{\mathcal{C}} \leq 0 \leq u_{\mathcal{C}}$, then

$$T = \prod_{A \in \mathcal{X}} T^A.$$

PROOF Define $B_1 := A_1$ and $B_i := A_i \cup B_{i-1}$, $i = 2,\ldots,n$. It follows from the definition of module and Lemma 1 that A_i and B_i, $i = 1,\ldots,n$, are T-modules. By Lemma 2 it follows that $\tilde{T}^{B_j} \subseteq \tilde{T}^{B_{j-1}} \times T^{A_j}$, for all $j = 2,\ldots,n$. We already observed that $\tilde{T}^{B_1} \subseteq T^{B_1} = T^{A_1}$, hence it follows by induction that $\tilde{T}^{B_j} \subseteq \prod_{i=1}^{j} T^{A_i}$. Since $B_n = \mathcal{R}$ we obtain that $T \subseteq \prod_{i=1}^{n} T^{A_i}$.

To prove $\prod_{i=1}^{n} T^{A_i} \subseteq F$, let $v^i \in T^{A_i}$ be arbitrary but fixed. Since $T \neq \emptyset$, it follows that there exists a $w \in T$. Let d^i denote the right-hand side of T-module A_i. We get $S_{A_i} w_{A_i} = d^i$, for all $i = 1,\ldots,n$, which implies $b = Sw = \sum_{i=1}^{n} S_{A_i} w_{A_i} = \sum_{i=1}^{n} d^i$. Now define $v \in \mathbb{R}^{\mathcal{R}}$ with $v_{A_i} := v^i$, for all $i = 1,\ldots,n$. It follows that $Sv = \sum_{i=1}^{n} S_{A_i} v^i = \sum_{i=1}^{n} d^i = b$. Clearly, v also satisfies $\ell \leq v \leq u$. We conclude $v \in F$ and thus $\prod_{i=1}^{n} T^{A_i} \subseteq F$.

If in addition, we have $\ell_{\mathcal{C}} \leq 0 \leq u_{\mathcal{C}}$, then by Lemma 2 and Cor. 1 we also get the equalities $\tilde{T}^{B_j} = T^{B_j} = T^{B_{j-1}} \times T^{A_j}$ and hence, $T = \prod_{i=1}^{n} T^{A_i}$. ■

3.4 Unique Decomposition

Next we study the existence and uniqueness of the decomposition of a network into minimal modules.

Definition 6 (Minimal Module) Let $P \subseteq \mathbb{R}^{\mathcal{R}}$ be a flux space. A P-module $A \subseteq \mathcal{R}$ is *minimal* if there exists no P-module B s.t. $B \subset A$. □

The following result shows that we can always decompose a network into minimal modules.

Proposition 1 (Existence) *Let P be a flux space. Every P-module A can be partitioned into minimal P-modules, i.e., there exist minimal P-modules $A_1,\ldots,A_k$ s.t. $A = A_1 \dot{\cup} A_2 \dot{\cup} \ldots \dot{\cup} A_k$.* □

PROOF Assume the proposition is false. Then there exists a non-minimal P-module $A \subset \mathcal{R}$ that cannot be partitioned into smaller P-modules and a minimal P-module $B \subset A$. By definition there exist $a, b \in \mathbb{R}^{\mathcal{M}}$ with $S_A v_A = a$ and $S_B v_B = b$. It follows $a - b = S_A v_A - S_B v_B = S_{A\setminus B} v_{A\setminus B} + S_B v_B - S_B v_B = S_{A\setminus B} v_{A\setminus B}$, which implies that $C := A \setminus B$ is also a P-module. Thus $C \dot{\cup} B$ is a partition of A, which contradicts the assumption. ■

Note that this proposition holds for arbitrary flux spaces P. The following lemma holds for every flux space that satisfies the steady-state assumption. In particular, it holds for the thermodynamically constrained flux space T.

Lemma 3 (Modules form Product Space) *Let $P \subseteq \{v \in \mathbb{R}^{\mathcal{R}} : Sv = b\}$. Assume $P = \prod_{i=1}^{n} P_i$ with $P_i \subseteq \mathbb{R}^{A_i}$, where $A_i \subseteq \mathcal{R}$. Then for every $i = 1,\ldots,n$, there exists a vector $b^i \in \mathbb{R}^{\mathcal{M}}$ s.t. $S_{A_i} v_{A_i} = b^i$, for all $v \in P$, i.e., A_i is a P-module.* □

PROOF Assume the lemma is false. Then there exist $i \in \{1,\ldots,n\}$ and $v, w \in P$ s.t. $S_{A_i} v_{A_i} \neq S_{A_i} w_{A_i}$. Define $w' \in \mathbb{R}^{\mathcal{R}}$ by $w'_{A_i} = w_{A_i}, w'_{\mathcal{R}\setminus A_i} = v_{\mathcal{R}\setminus A_i}$. Since $P = \prod_{i=1}^{n} A_i$, it follows that $w' \in P$. Since $S_{A_i} w_{A_i} \neq S_{A_i} v_{A_i}$, we get $Sw' = S_{A_i} w_{A_i} + S_{\mathcal{R}\setminus A_i} v_{\mathcal{R}\setminus A_i} = S_{A_i} w_{A_i} + b - S_{A_i} v_{A_i} \neq b$. Thus $w' \notin P$, which is a contradiction. ■

To prove uniqueness of the decomposition into minimal T-modules (Thm. 3), we use Thm. 2 from the previous section. Given a partition of the reaction set $\mathcal{R}$ into T-modules A, the thermodynamically constrained flux space T can be written as the product of the flux spaces T^A. Assuming that there exist two different partitions with minimal T-modules, we show that we can write T as a product of smaller factors (Lemma. 4). We then go in the reverse direction and show with Lemma 3 that from this we can obtain smaller T-modules, contradicting the minimality.

Lemma 4 *Let X, I be sets. Let $P \subseteq X^I$. Let $P = \prod_{i=1}^{n} A_i$ and $P = \prod_{i=1}^{m} B_i$, with $A_i \subseteq X^{a_i}$ and $B_i \subseteq X^{b_i}$, where $a_i \subseteq I, i = 1,\ldots,n$, and $b_i \subseteq I, i = 1,\ldots,m$. Then $P = \prod_{i=1}^{n} \prod_{j=1}^{m} C_{ij}$ with $C_{ij} = \mathrm{pr}_{a_i \cap b_j} P$.* □

PROOF For every $j = 1,\ldots,m$ holds $B_j = \mathrm{pr}_{b_j}(P) = \mathrm{pr}_{b_j}(\prod_{i=1}^n A_i) = \prod_{i=1}^n \mathrm{pr}_{a_i \cap b_j} A_i = \prod_{i=1}^n C_{ij}$. Since $P = \prod_{j=1}^m B_j = \prod_{j=1}^m \prod_{i=1}^n C_{ij} = \prod_{i=1}^n \prod_{j=1}^m C_{ij}$, the claim follows. ■

Theorem 3 (Uniqueness) *Assume $T \neq \emptyset$, where $\ell_{\mathcal{C}} \leq 0 \leq u_{\mathcal{C}}$. Then the partition of $\mathcal{R}$ into minimal T-modules exists and is unique.* □

PROOF Since $\mathcal{R}$ is a T-module, it follows by Prop. 1 that there always exists a partition into minimal T-modules. Assume there exist two partitions $\mathcal{X} \neq \mathcal{Y}$ of $\mathcal{R}$ into minimal T-modules. By Thm. 2 and Cor. 1, we can write

$$\prod_{x \in \mathcal{X}} \mathrm{pr}_x(T) = \prod_{x \in \mathcal{X}} T^x = T = \prod_{y \in \mathcal{Y}} T^y = \prod_{y \in \mathcal{Y}} \mathrm{pr}_y(T).$$

By Lemma 4 it follows that there exists a partition $\mathcal{Z}$ of $\mathcal{R}$ which is finer than $\mathcal{X}, \mathcal{Y}$, i.e., every $z \in \mathcal{Z}$ is contained in some $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. The partition $\mathcal{Z}$ also satisfies

$$T = \prod_{z \in \mathcal{Z}} \mathrm{pr}_z(T).$$

It follows by Lemma 3 that every $z \in \mathcal{Z}$ is also a T-module. Since $\mathcal{X} \neq \mathcal{Y}$, there exists at least one T-module of $\mathcal{Z}$ that is strictly contained in a T-module of $\mathcal{X}$. This contradicts the minimality of the T-modules in $\mathcal{X}$. ■

3.5 Elementary Flux Modes

An important consequence of the decomposition into T-modules is that we can describe the set of elementary flux modes in a more compact form. This result follows basically directly from the product form of the flux space (Thm. 2). We define the set of *elementary flux modes of a flux space P* as the flux modes with minimal sign-support:

$$\mathrm{EFM}(P) := \{v \in P : \mathrm{sign}(w) \not\subset \mathrm{sign}(v)\ \forall w \in P \setminus \{0\}\}.$$

From this, we can derive the following relationships between the elementary flux modes of F^A and T^A:

Proposition 2 *Assume $T \neq \emptyset$. Let A be a T-module with right-hand side $d \neq 0$. Let*

$$P := \{(v,x) \in \mathbb{R}^A \times \mathbb{R}^+ : S_A v - dx = 0, \ell_A \leq v \leq u_A\}.$$

Then

a) $\mathrm{EFM}(T^A) \subseteq \mathrm{EFM}(F^A)$

b) $\{v \in \mathbb{R}^A : (v,1) \in \mathrm{EFM}(P)\} \subseteq \mathrm{EFM}(T^A)$

c) If $\ell_{A \cap \mathcal{C}} \leq 0 \leq u_{A \cap \mathcal{C}}$, then $\mathrm{EFM}(T^A) = \mathrm{EFM}(F^A)$.

d) If $\ell_r \in \{-\infty, 0\}$ and $u_r \in \{0, \infty\}$ for all $r \in A$, then $\{v \in \mathbb{R}^A : (v,1) \in \mathrm{EFM}(P)\} = \mathrm{EFM}(T^A)$. □

PROOF We show all statements separately:

a) Let $v \in \mathrm{EFM}(T^A)$. Assume $v \notin \mathrm{EFM}(F^A)$. Since $T^A \subseteq F^A$, it follows that there exists $w \in F^A$ with $\mathrm{sign}(w) \subset \mathrm{sign}(v)$ and w is thermodynamically infeasible. By Thm. 1, there exists an internal cycle $c \in \mathbb{R}^A$ with $\mathrm{sign}(c) \subseteq \mathrm{sign}(w)$. It follows that $\mathrm{sign}(c) \subseteq \mathrm{sign}(v)$, contradicting $v \in T^A$.

b) Let $(v,1) \in \mathrm{EFM}(P)$. Assume $v \notin \mathrm{EFM}(T^A)$. Since $T^A \subseteq \mathrm{pr}_A(P)$, it follows that $v \notin T^A$. By Thm. 1, there exists an internal cycle c with $\mathrm{sign}(c) \subseteq \mathrm{sign}(v)$. It follows that $\mathrm{sign}(c) \subset \mathrm{sign}(\{v,1\})$. This contradicts $(v,1) \in \mathrm{EFM}(P)$.

c) Assume $\ell_{A\cap\mathcal{C}} \leq 0 \leq u_{A\cap\mathcal{C}}$, and suppose there exists $v \in \mathrm{EFM}(F^A) \setminus \mathrm{EFM}(T^A)$. We conclude that $v \notin T^A$, hence v contains an internal cycle c. Since $\ell_{A\cap\mathcal{C}} \leq 0 \leq u_{A\cap\mathcal{C}}$, we can subtract all internal cycles using Alg. 1 from Müller and Bockmayr (2013). By Thm. 2 in Müller and Bockmayr (2013), we obtain $v' \in F^A$ with $\mathrm{sign}(v') \subset \mathrm{sign}(v)$, contradicting $v \in \mathrm{EFM}(F^A)$.

Thus, it follows that $\mathrm{EFM}(F^A) \subseteq \mathrm{EFM}(T^A)$. Together with a) we get that $\mathrm{EFM}(F^A) = \mathrm{EFM}(T^A)$.

d) We will now consider the case when $\ell_r \in \{-\infty, 0\}$ and $u_r \in \{0, \infty\}$ for all $r \in A$. Assume there exists $v \in \mathrm{EFM}(T^A)$ with $(v,1) \notin \mathrm{EFM}(P)$. Since $(v,1) \in P$, it follows that $(v,1)$ is not minimal. Hence, there exists a $(w,x) \in P$ with $\mathrm{sign}(w,x) \subseteq \mathrm{sign}(v,1)$.

If $x > 0$ we scale (w,x) to $(w',1)$. Since the bounds are only 0 or infinity, the flux bounds will also be satisfied by w'. It follows that $w' \in T^A$. Since $\mathrm{sign}(w') \subset \mathrm{sign}(v)$, this is a contradiction to the minimality of v. Therefore, we only need to consider the case where $x = 0$.

Let $v' = v - \alpha w$, where α is chosen as large as possible such that $\mathrm{sign}(v') \subseteq \mathrm{sign}(v)$. Since $\ell_A \leq 0 \leq u_A$, it follows that $v' \in P$ and $\mathrm{sign}(v') \subset \mathrm{sign}(v)$. Since $(w,0) \in P$, we have $Sw = 0$ and hence, $Sv' = d$. Since v was thermodynamically feasible, v' is also thermodynamically feasible, hence $v' \in T^A$, which contradicts the minimality of v. ∎

Theorem 4 *Let $b \neq 0$ and $T \neq \emptyset$. Let $\mathcal{X} = \{A_1, \ldots, A_n\}$ be a partition of $\mathcal{R}$ into T-modules, then*

$$\mathrm{EFM}(T) \subseteq 0^N \times \prod_{A\in\mathcal{X}:0\notin T^A} \mathrm{EFM}(T^A),$$
$$\mathrm{EFM}(T) \subseteq \mathrm{EFM}(F)$$

where $0^N \in \mathbb{R}^N$ with $0^N_r = 0$ for all $r \in N = \bigcup_{A\in\mathcal{X}:0\in T^A} A$.
If $T = \prod_{A\in\mathcal{X}} T^A$, then the first inclusion becomes an equality.
If $\ell_{\mathcal{C}} \leq 0 \leq u_{\mathcal{C}}$, then both inclusions become equalities. □

PROOF We start with the first inclusion. Let $v \in \mathrm{EFM}(T)$ be fixed but arbitrary. Assume there exists $A \in \mathcal{X}$ with $0 \in T^A$ and $v_A \neq 0$. We define $w \in \mathbb{R}^{\mathcal{R}}$ by $w_B = v_B$, for all $B \in \mathcal{X} \setminus \{A\}$, and $w_A = 0$. It follows by Thm. 2 that $w \in F$. By construction, we have $\mathrm{sign}(w) \subset \mathrm{sign}(v)$. Since v is thermodynamically feasible, it follows that w is thermodynamically feasible, hence $w \in T$. Since $b \neq 0$, it follows that $w \neq 0$, which is a contradiction to the minimality of v.

Assume there exists an $A \in \mathcal{X}$ with $0 \notin T^A$ s.t. $v_A \notin \mathrm{EFM}(T^A)$. Since $v_A \in T^A$ by Thm. 2, it follows that there exists $w^A \in T^A$ with $\mathrm{sign}(w^A) \subset \mathrm{sign}(v_A)$. We now define $w \in \mathbb{R}^{\mathcal{R}}$ by $w_B = v_B$, for all $B \in \mathcal{X} \setminus \{A\}$, and $w_A = w^A$. It follows by Thm. 2 that $w \in F$. By construction we have $\mathrm{sign}(w) \subset \mathrm{sign}(v)$. Since v is thermodynamically feasible, it follows that w is thermodynamically feasible, hence $w \in T$, which is a contradiction to the minimality of v. Therefore

$$\mathrm{EFM}(T) \subseteq 0^N \times \prod_{A\in\mathcal{X}:0\notin T^A} \mathrm{EFM}(T^A).$$

The second inclusion follows directly from Prop. 2 a) by choosing $A = \mathcal{R}$.

Now we consider the case where $T = \prod_{A\in\mathcal{X}} T^A$. Let $v^A \in \mathrm{EFM}(T^A)$ be a fixed but arbitrary elementary flux mode for each $A \in \mathcal{X}$ with $0 \notin T^A$. By assumption, it follows that $v \in \mathbb{R}^{\mathcal{R}}$ defined by $v_A = v^A$, for $A \in \mathcal{X}$ with $0 \notin T^A$, and $v_A = 0$, for $A \in \mathcal{X}$ with $0 \in T^A$, satisfies $v \in T$. Assume there exists a flux vector $w \in T$ with $\mathrm{sign}(w) \subset \mathrm{sign}(v)$. It follows that there exists a reaction $r \in \mathrm{supp}(v) \setminus \mathrm{supp}(w)$, where

supp denotes the support (non-zero entries). Since the T-modules form a partition of $\mathcal{R}$, it follows that there exists a T-module $A \in \mathcal{X}$ with $r \in A$ and $0 \notin T^A$. It follows that $\operatorname{sign}(w_A) \subset \operatorname{sign}(v^A)$. By Thm. 2 it follows that $w_A \in T^A$, a contradiction to the minimality of v^A. Hence, we conclude that if $T = \prod_{A\in\mathcal{X}} T^A$, then

$$\operatorname{EFM}(T) \supseteq 0^N \times \prod_{A\in\mathcal{X}:0\notin T^A} \operatorname{EFM}(T^A),$$

which implies

$$\operatorname{EFM}(T) = 0^N \times \prod_{A\in\mathcal{X}:0\notin T^A} \operatorname{EFM}(T^A).$$

If we have $\ell_\mathcal{C} \leq 0 \leq u_\mathcal{C}$ and $T \neq \emptyset$, it follows by Thm. 2 that $T = \prod_{A\in\mathcal{X}} T^A$ and thus we get the first inclusion. The second inclusion follows from Prop. 2 c) by choosing $A = \mathcal{R}$. ∎

4 Methods

The following methods operate on the thermodynamically constrained flux space, as introduced in Def. 4:

$$T := \{v \in \mathbb{R}^\mathcal{R} : Sv = b,\ \ell \leq v \leq u,\ v \text{ thermo. feasible}\}.$$

We recall that a flux vector v is considered thermodynamically (thermo.) feasible, if it does not contain any internal cycles (Def. 1, Thm. 1).

Note that usually the steady-state assumption is formulated as $Sv = 0$, i.e., by choosing $b = 0$. We also allow b to be different from 0. If $b_m > 0$ for a metabolite $m \in \mathcal{M}$, this means that there is a constant consumption with rate b_m of the metabolite m by some reaction that is not explicitly modeled. Similarly $b_m < 0$ means that there is a constant production with rate $-b_m$ of the metabolite m by some reaction that is not explicitly modeled. With these formulations the flux space of a T-module can be understood as the flux space of the subnetwork.

According to Def. 3, a T-module is a subset $A \subseteq \mathcal{R}$ of the reactions such that the interface flux $d \in \mathbb{R}^\mathcal{M}$ with the rest of the network is constant for each feasible flux vector $v \in T$. Formally, we have $S_A v_A = d$ for each $v \in T$. For the definition of the flux space of the T-module, we turned this property around and considered all flux vectors $v \in \mathbb{R}^A$ (through the module) that satisfy the interface flux condition (Def. 4):

$$T^A := \left\{v \in \mathbb{R}^A : S_A v = d,\ \ell_A \leq v \leq u_A,\ v \text{ thermo. feasible}\right\}$$

Observe that T^A is just a thermodynamically constrained flux space as T, except that we chose a different stoichiometric matrix and a different right-hand side vector (d instead of b).

For simplicity we will restrict the generality of the flux space T and assume in the following that the bounds ℓ and u on the fluxes satisfy $\ell_\mathcal{I} \leq 0$ and $u_\mathcal{I} \geq 0$, where $\mathcal{I}$ denotes the set of internal reactions. This means that we have no bound on an internal reaction that forces flux away from 0. This condition can be slightly relaxed, the reader is referred to Sect. 3 for details.

All the results also hold without thermodynamic constraints. To see this, observe that we impose no restrictions on which reactions must be considered internal reactions. Hence, it is also possible to (artificially) declare all reactions to be exchange reactions. In this case there cannot exist any internal cycles and hence, only the steady-state assumption and flux bounds are retained.

4.1 Formulation of the Optimal Flux Space

Usually, T-modules can only be found when analyzing a flux space subject to an optimization criterion, for example, when the fluxes that maximize biomass production are studied. Since our flux space T, as defined in (1), does not allow the addition of arbitrary linear constraints, we now discuss how the constraints that restrict the flux space to only optimal fluxes can be formulated. We point out that for many cases it will be possible to get a formulation in the form of (1), but in general this is not true to its full extent. We therefore consider a special and a more general case.

4.1.1 Special Case:

We assume that our objective is to maximize flux through a given reaction $t \in \mathcal{R}$. Usually t is the biomass reaction. Formally, we solve

$$\text{opt} := \max\{v_t : Sv = 0,\ \ell \le v \le u,\ v \text{ thermo. feasible}\}.$$

The optimal flux space can then be formulated as

$$T_{\text{opt}} := \left\{v \in \mathbb{R}^{\mathcal{R}} : Sv = 0,\ \ell \le v \le u,\ \text{opt} \le v_t \le \text{opt},\ v \text{ thermo. feasible}\right\}.$$

Note that we assume that $\ell_{\mathcal{I}} \le 0 \le u_{\mathcal{I}}$. If t is the biomass reaction, it is an exchange reaction and thus, this will not be an issue. However, if t is an internal reaction (like the `ATPase` reaction), this condition will be violated. We refer the reader to Sect. 3 for more details to this case.

If t is an exchange reaction, we can also eliminate it and write with $\mathcal{R}' := \mathcal{R} \setminus \{t\}$

$$T'_{\text{opt}} := \left\{v \in \mathbb{R}^{\mathcal{R}'} : S_{\mathcal{R}'}v = -S_t \cdot \text{opt},\ \ell_{\mathcal{R}'} \le v \le u_{\mathcal{R}'}, v \text{ thermo. feasible}\right\}.$$

This is a useful transformation if the bounds ℓ and u just encode irreversibilities, like in elementary flux mode analysis. It follows that we then obtain a network, where ℓ and u still only encode irreversibilities.

4.1.2 General Case:

We assume that we are given an arbitrary cost function c and we solve

$$\text{opt} = \max\{cv : Sv = 0,\ \ell \le v \le u,\ v \text{ thermodynamically feasible}\}$$

The optimal flux space can then be formulated as

$$P_{\text{opt}} := \left\{v \in \mathbb{R}^{\mathcal{R}} : Sv = 0,\ \ell \le v \le u,\ cv = \text{opt}, v \text{ thermo. feasible}\right\}.$$

However, our results do not allow the addition of arbitrary linear constraints like $cv = \text{opt}$. Hence, we will have to slightly modify the metabolic network to account for the additional constraint. This can be done by introducing a new metabolite with right-hand side opt and modifying the reactions r with $c_r \neq 0$ to produce this new metabolite with stoichiometric coefficient c_r:

$$T''_{\text{opt}} := \left\{v \in \mathbb{R}^{\mathcal{R}} : \begin{pmatrix} S \\ c \end{pmatrix} v = \begin{pmatrix} 0 \\ \text{opt} \end{pmatrix},\ \ell \le v \le u, v \text{ thermo. feasible}\right\}$$

Ignoring thermodynamic feasibility, this corresponds to adding the linear constraint $cv = \text{opt}$. However, since we modify the reactions of the network, we also modify the structure of internal circuits. Hence, thermodynamic feasibility properties will in general only be retained if all reactions r with $c_r \neq 0$ are exchange reactions.

4.2 Computing Minimal Modules

The central result of this paper is the following direct method for computing the decomposition of $\mathcal{R}$ into minimal (w.r.t. set inclusion) T-modules. We have already shown that this decomposition is always well defined (Thm. 3). The algorithm for computing T-modules is based on the insight that the flux space T can be written as the Cartesian product of the flux spaces for the T-modules (Thm. 2). From this, we can derive that if we fix the flux value of one reaction to a fixed value, this will have no influence on the flux variability of a reaction in a different T-module:

Corollary 2 *Let $P := \{v : Sv = b, \ell \leq v \leq u\}$ or $P := T$ with $\ell_{\mathcal{I}} \leq 0 \leq u_{\mathcal{I}}$. Assume $P \neq 0$. Let $\mathcal{R}$ be partitioned into P-modules. Let A be a P-module and $r \in A, s \in \mathcal{R} \setminus A$. Let x be a feasible flux rate for r, i.e., $x \in \mathrm{pr}_r(P)$. Then*

$$\max\{cv_s : v \in P, v_r = x\} = \max\{cv_s : v \in P\} \qquad \text{for all } c \in \mathbb{R}.$$

PROOF Since x is a feasible flux rate for r, there exists a flux vector $w \in P$ with $w_r = x$. Let v be a flux vector maximizing $\max\{cv_s : v \in P\}$. By Thm. 2 it follows that $v' \in \mathbb{R}^{\mathcal{R}}$, with $v'_A = w_A$ and $v'_{\mathcal{R}\setminus A} = v_{\mathcal{R}\setminus A}$, satisfies $v' \in P$. Since $s \notin A$, it follows that $cv_s = cv'_s$, showing $\max\{cv_s : v \in P, v_r = x\} \geq \max\{cv_s : v \in P\}$. The other direction follows immediately, since addition of constraints can never increase the objective value. ∎

We can now use this result to put together our method for computing minimal T-modules. Since we want to see dependencies between reactions in the same minimal T-module, we choose extreme values for x (minimal and maximal flux). This will likely cause big effects on the variability of reactions in the same minimal T-module, but by Cor. 2 we will see no effect on the variability in other reactions.

To compute candidate sets of T-modules, we compute a graph $G = (\mathcal{R}, E)$ using Alg. 1. For the algorithm to work, we assume that there exists no pathway that can carry unbounded flux with thermodynamic constraints (this is only possible if nutrient/energy uptake is unbounded).

Algorithm 1 Computation of candidate sets for T-modules.

1. Compute thermodynamic flux variability $v_r \in [v_r^{\min}, v_r^{\max}]$ for each reaction r in the network. Define $V := \{r : v_r^{\min} < v_r^{\max}\}$.
2. Each reaction $r \notin V$ forms a T-module by itself.
3. For each reaction $r \in V$ do the following
   (a) Fix r to its maximal/minimal flux rate (which exists because of thermodynamic feasibility)
   (b) Compute thermodynamic flux variability $v_s \in [v_s^{\min,r}, v_s^{\max,r}]$ for each reaction $s \in V$.
   (c) If $v_s^{\min,r} > v_s^{\min}$ or $v_s^{\max,r} < v_s^{\max}$, then we say that s is influenced by r and add the edge (r,s) to E.
   (d) If $v_s^{\min,r} > v_s^{\min} \geq 0$ or $v_s^{\max,r} < v_s^{\max} \leq 0$, we say that r forces flux through s.
   (e) If $v_s^{\min,r} = 0 = v_s^{\max,r}$, we say that r blocks flux through s.
4. Compute the connected components $\mathcal{X} = \{A_1, \ldots, A_n\}$ of G.

To run *thermodynamically constrained flux variability analysis* (tFVA), we use the fast-tFVA tool (Müller and Bockmayr, 2013). With it, we were able to run this algorithm on genome-scale networks like *E. coli* iAF1260 and *S. cerevisiae* iND750.

By Cor. 2 it follows that every connected component of G is a subset of a minimal T-module. However, in practice usually all connected components are indeed T-modules. To check if a subset $A \subset \mathcal{R}$ is indeed a T-module, we run Alg. 2 with $P = T$. This algorithm returns YES if and only if A is a P-module, because we individually minimize and maximize each component (metabolite) of the right-hand side of the candidate P-module. If and only if A is a P-module, d is a fixed vector, i.e., for each metabolite the minimum and maximum must be the same. If there were a flux vector $v \in P$ with different right-hand side, then also the maximum or minimum would be different.

Algorithm 2 Checks if a candidate set A is indeed a P-module. This algorithm works not only for the flux space $P = T$, but for arbitrary flux spaces P.

Input: $A \subseteq \mathcal{R}$
$M := \{m \in \mathcal{M} : \exists r \in A : S_{mr} \neq 0\}$
for $m \in M$ **do**
  $d_{\min} := \min\{S_{mA} v_A : v \in P\}$
  $d_{\max} := \max\{S_{mA} v_A : v \in P\}$
  if $d_{\min} \neq d_{\max}$ **then**
    return NO
  end if
end for
return YES

In practice, it rarely happens that the connected components are not T-modules. If, however, connected components are detected that are not T-modules, these have to be combined manually to form T-modules. This is an easy task if only two connected components A and B are not T-modules. Since the T-modules partition the set of all reactions, it follows that the union of A and B forms a minimal T-module. In general, however, there are exponentially many combinations possible. This would then require an additional algorithm, but since we did not encounter this in our test cases, we did not implement an algorithm for it.

4.3 Blocking Graph

The blocking graph visualizes reactions that are on alternative pathways. We define it as the directed graph $\mathcal{B} = (\mathcal{R}, E)$, where

$$E := \{(r,s) : (v_r = v_r^{\max} \to v_s = 0) \vee (v_r = v_r^{\min} \to v_s = 0) \ \ \forall v \in T\}.$$

Observe that these arcs are computed in step 3e of the Alg 1. It follows that no reactions of different T-modules will be connected in $\mathcal{B}$.

Hence, we will usually only look at the subgraph of the blocking interaction graph consisting of the nodes of a T-module. We will call this subgraph the *blocking graph* of the T-module. An example can be seen in Fig. 2.

4.4 Modular Decomposition of Optimal-Yield Elementary Flux Modes

An important application of the decomposition into T-modules is that we can write the set of elementary flux modes into a more compact form. This result also follows from the product form of the flux space (Thm. 2).

In this subsection, we assume that the flux spaces only have irreversibilities as flux bounds (i.e. $\ell \in \{-\infty, 0\}^{\mathcal{R}}$ and $u \in \{0, \infty\}^{\mathcal{R}}$), as it is common in elementary flux mode analysis. However, since we are particularly interested in looking at the optimal flux space, we allow the constraint $Sv = b$ with $b \neq 0$ in the definition of the flux space. We refer the reader to Sect. 4.1 on how to obtain such a formulation.

We point out that the original definition by Schuster and Hilgetag (1994) only defines elementary flux modes for flux cones with $Sv = 0$. Hence, we will use a slightly extended version of elementary flux modes. For a given arbitrary flux space $P \subseteq \mathbb{R}^{\mathcal{R}}$, we defined (Sect. 3.5) the set of elementary flux modes of P as the flux modes with minimal sign-support:

$$\mathrm{EFM}(P) = \{v \in P : \mathrm{sign}(w) \not\subset \mathrm{sign}(v)\ \forall w \in P \setminus \{0\}\},$$

where sign is defined as in Sect. 3.1. We observe that this definition, if applied to the flux cone, is equivalent to the definition given by Schuster and Hilgetag (1994).

First of all, we observe that in this sense the set of thermodynamically feasible elementary flux modes is not very different from the set of elementary flux modes of the steady-state flux space F itself. More formally we showed that

$$\begin{aligned} \mathrm{EFM}(T) &= 0^N \times \prod_{A \in \mathcal{X} : 0 \notin T^A} \mathrm{EFM}(T^A) = \mathrm{EFM}(F) && \text{(by Prop. 2)} \\ \mathrm{EFM}(T^A) &= \mathrm{EFM}(F^A) = \{v \in \mathbb{R}^A : (v,1) \in \mathrm{EFM}(P)\} && \text{(by Thm. 4),} \end{aligned}$$

where for a T-module A with right-hand side d, we have

$$\begin{aligned} F &:= \{v \in \mathbb{R}^{\mathcal{R}} : Sv = b,\ \ell \le v \le u\}, \\ F^A &:= \{v \in \mathbb{R}^A : S_A v = d, \ell_A \le v \le u_A\}, \\ P &:= \{(v,x) \in \mathbb{R}^A \times \mathbb{R}_{\ge 0} : S_A v - dx = 0, \ell_A \le v \le u_A\}. \end{aligned}$$

We conclude that we can easily combine the elementary flux modes of the T-modules to obtain all elementary flux modes of the network and although our definition of elementary flux mode is slightly different from the definition given in Schuster and Hilgetag (1994), the same algorithms can be applied to compute the elementary flux modes of the T-modules. We only need to enumerate the elementary flux modes (v,x) of P and filter out those which have $x > 0$.

Theoretically, this can be done with any kind of elementary flux mode enumeration tool, like `metatool` (von Kamp and Schuster, 2006), or `EFMtool` (Terzer and Stelling, 2008). We point out that d (the stoichiometry of the artificial reaction with flux x) is the result of an optimization step (if we analyze the optimal flux space). Hence, d is very likely to be not exact and have non-integer coefficients which are also not approximable by nice fractions. This turned out to be a significant problem for `EFMtool`, which sometimes did not find any elementary flux modes with $x > 0$, although there clearly existed flux modes with $x > 0$. Hence, we used `metatool`, which did not have these problems if the presolver was turned off. If the T-modules are very small, also a MILP based approach (de Figueiredo *et al.*, 2009) works very well.

5 Results and Discussion

Using the presented method, we were able to compute the modules of the optimal flux space of many genome-scale networks like *E. coli* iAF1260, *S. cerevisiae* iND750, or *M. tuberculosis* iNJ661. Comprehensive lists of the detected modules for each of the analyzed networks can be found in the supplementary material. There we also list the optimal elementary flux modes of *E. coli* iJR904 grown on *threonine* resp. *arginine*. With the help of the modules, it was possible to write down all optimal elementary flux modes on a single page.

Next to the modules, we also get interaction relationships between the reactions inside of modules (see Sect. 4.3 and steps 3c, 3d, 3e of Alg. 1). Let us consider the module consisting of the reactions L-alanine transaminase (`ALATA_L`), valine transaminase (`VALTA`), and valine-pyruvate aminotransferase (`VPAMT`). This is depicted in Fig. 2a and the blocking graph is shown in Fig. 2b. By studying Fig. 2a,

we might think that VALTA and VPAMT together form an alternative route to ALATA_L. However, a look at Fig. 2b reveals that if VALTA carries maximal flux, then VPAMT does not carry any flux and vice versa. The blocking interaction graph actually shows us that VPAMT and ALATA_L together form an alternative route to VALTA. Furthermore, we can derive that ALATA_L is also important for other pathways, since even maximal or minimal flux through VALTA cannot force flux through ALATA_L to zero.

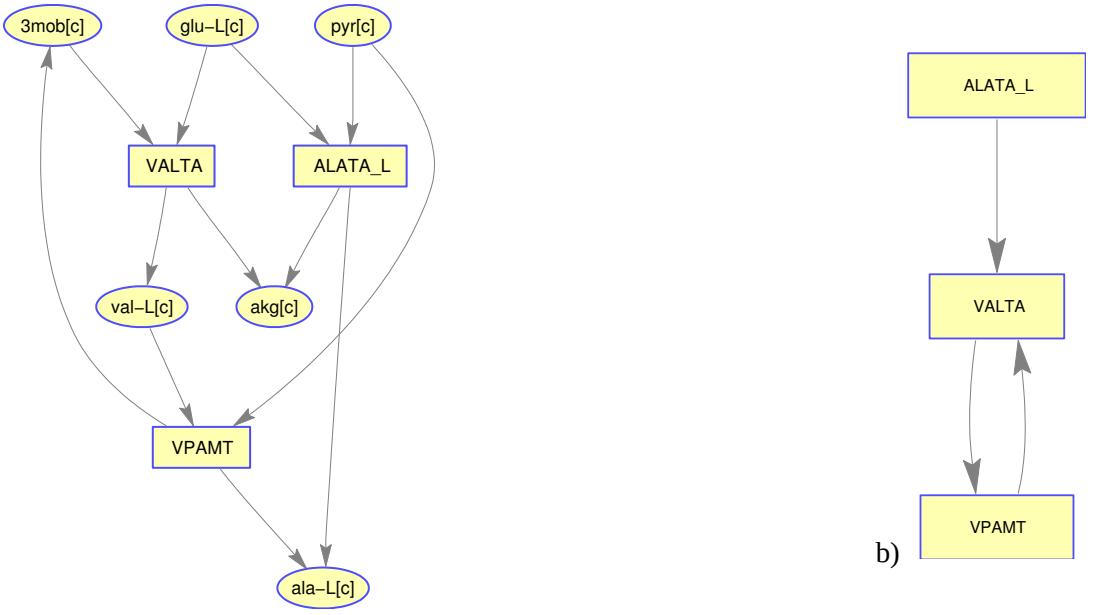

Figure 2: Graphical representations of a module of *E. coli* iAF1260.
a) The subnetwork of the module consisting of ALATA_L, VPAMT, and VALTA. The participating metabolites are drawn in ellipses. Stoichiometric coefficients are not shown.
b) The blocking graph of the module. An arc from reaction r_1 to r_2 means that if r_1 has minimal or maximal flux, then no flux is possible through r_2. In this example it means that VPAMT and VALTA block each other when carrying optimal flux. However, VALTA does not block ALATA_L when carrying optimal flux. This shows that ALATA_L is also necessary for other pathways and crucial for optimal growth.

The blocking graph may also give us information about which reactions may be subject to regulatory control in order to obtain a specific effect. For example, if we consider the module of *S. aureus* iSB619 shown in Fig. 3, we see that regulatory control on LDH_D or LDH_L will potentially influence what kind of lactate is produced.

5.1 Sensibility to Growth Conditions in *E. coli*

When we analyzed *E. coli* iAF1260 grown on *glucose*, we discovered instead of the biggest module found by Kelk *et al.* (2012) three smaller modules, seen in Figs. 4, 5, 6, which mostly contain the same reactions. It turns out that the difference was actually not caused by the different analysis methods, but actually by slight modifications of the metabolic network. Kelk *et al.* (2012) used an uptake flux of at most 12.7777mmol/gDW/h (mmol per gram dry weight per hour) for *glucose*, while we used an uptake flux of 8mmol/gDW/h for *glucose* as originally given in the model. All other bounds on the network were essentially the same (they additionally allow uptake of *Cob(I)alamin*, which however is blocked in the network).

A careful analysis of the network revealed that fluxes do not scale linearly with the uptake of *glucose* as assumed in Kelk *et al.* (2012). This is caused by two reactions with small absolute flux bounds: The network requires a flux through *ATP maintenance* of 8.39mmol/gDW/h and maximal *oxygen* uptake of

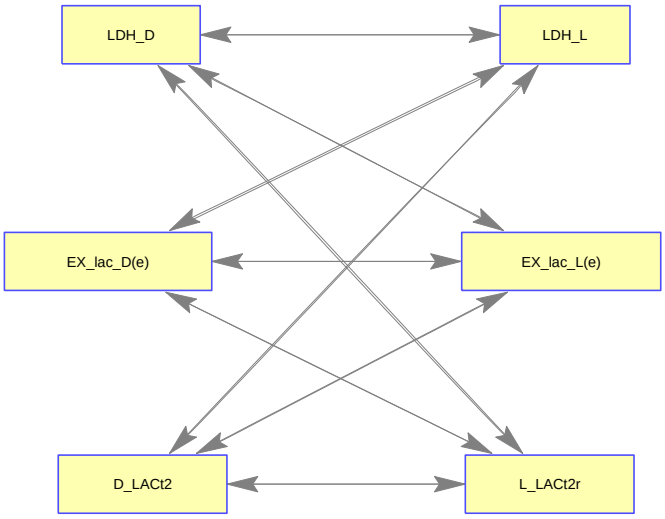

Figure 3: The blocking interaction graph shows clearly that (`LDH_D`, `D_LACt2`, `EX_lac_D(e)`) forms an alternative pathway to (`LDH_L`, `D_LACt2r`, `EX_lac_L(e)`).

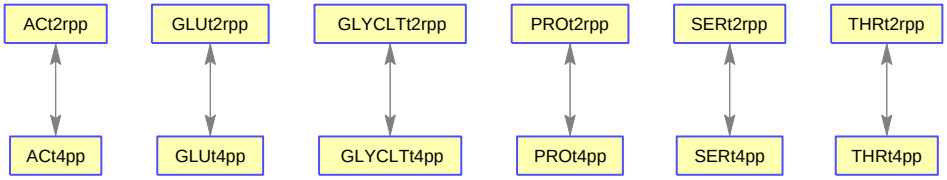

Figure 4: An arrow is drawn between two reactions in this module of *E. coli* iAF1260 (grown on *glucose*) if the reactions do not have a blocking interaction. This module is part of the biggest module found for this network in Kelk *et al.* (2012). This figure raises the assumption that we have 6 alternative pathways that realize the same function and that each of the pathways is realized by the two reactions connected by an edge. Indeed, this is the case: This module transports sodium from the periplasm to the cytosol in exchange to hydrogen.

18.5mmol/gDW/h. These values are fixed and do not scale with the *glucose* uptake.

If *E. coli* iAF1260 is allowed to have only an uptake flux of 8mmol/gDW/h for *glucose*, it will not consume all the *oxygen* to achieve optimal growth. However, with an uptake flux greater than 12mmol/gDW/h for *glucose*, it will require all the supplied oxygen to grow optimally. It follows that in this case the structure of the optimal flux space of *E. coli* iAF1260 will also change structurally. Consequently the optimal flux space gets partitioned into different modules.

To understand this structural change, we also analyzed anaerobic growth of *E. coli* iAF1260 under *glucose*. Interestingly, the modules shown in Figs. 4, 6 also existed under anaerobic growth conditions. This was unexpected since in the aerobic growth case with limited oxygen supply as studied by Kelk *et al.* (2012), these modules do not exist. Instead of the module in Fig. 5, we found a module consisting only of a subset of the reactions, as shown in Fig. 7.

A comparison of the modules in Fig. 5 (aerobic) and Fig. 7 revealed that the former is transforming *succinate* into *fumerate*, while the latter is doing the reverse transformation (see Tab. 1). Hence, the modules just look similar but actually perform a different metabolic function. This also explains why

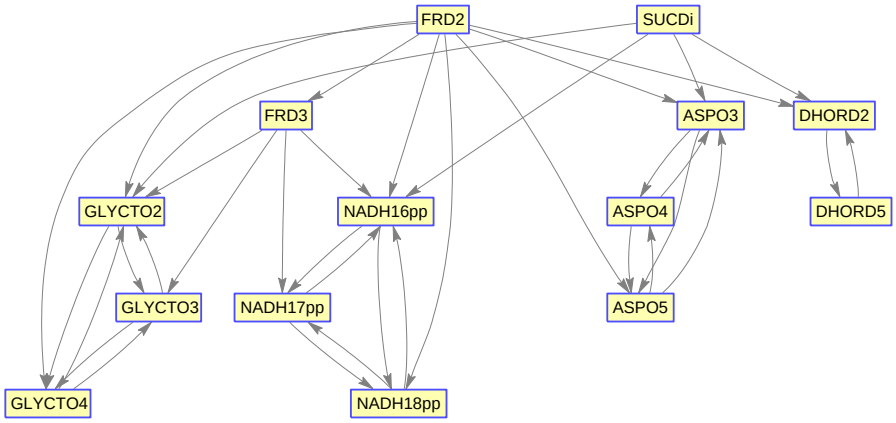

Figure 5: Blocking interaction graph of a module of *E. coli* iAF1260 (grown on *glucose*, aerobic). This module is part of the biggest module found for this network in Kelk *et al.* (2012). Here, we see that the this module has 4 submodules which interact by the reactions `FRD2`, `FRD3` and `SUCDi`.

under limited oxygen supply these modules do not appear.

Furthermore, we found small changes regarding the second largest module found by Kelk *et al.* (2012). Although it stays mostly the same, the reactions `ACKr`, `ACS`, `ADK1`, `PTAr`, `R15BPK`, `R1PK` leave and enter the module depending on the amount of *oxygen* supply. It is interesting to note that `ACKr` and `PTAr` are actually contained in the largest module from Kelk *et al.* (2012) (the module that decomposed into 3 smaller modules under high or no oxygen supply).

5.2 Evaluation on Genome-Scale Networks

Using fast-tFVA tool (Müller and Bockmayr, 2013) we were able to run this algorithm on many genome-scale networks. The results are given in Table. 2. In nearly all instances the algorithm directly computed the modules of the networks. Only in two cases (*E. coli* iJR094 grown on *threonine* resp. *tryptophan*), two connected components of the interaction graph were not modules. As already pointed out in Sect. 4.2, this is in accordance with our theoretical results. The example in Fig. 8 with the following flux polytope shows how this can happen:

$$P = \left\{ v \in \mathbb{R}^8 \;\middle|\; \begin{pmatrix} 1 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} v = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \; v \geq 0 \right\}.$$

For this reason, we always check at the end of Alg. 1 if the computed connected components are indeed modules (Alg. 2). It follows from Thm. 3 that this problem is not an intrinsic property of minimal modules. Instead, it is caused by how we detect interactions between reactions. We also remark that the addition of thermodynamic constraints can also not be its sole cause, since the example network does not involve thermodynamic constraints. We consider it likely that these effects may also happen to different

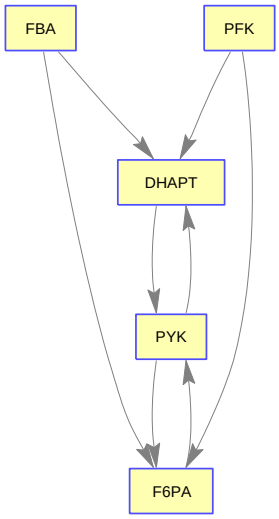

Figure 6: Blocking interaction graph of a module of *E. coli* iAF1260 (grown on *glucose*, aerobic). This module is part of the biggest module found for this network in Kelk *et al.* (2012). This diagram proposes the thesis that `F6PA` and `DHAPT` form an alternative pathway to `FBA`, `PFK`, and `PYK`. Indeed, both form two alternative pathways for transforming D-fructose 6-phosphate and phosphoenolpyruvate into dihydroxyacetone phosphate, pyruvate and glyceraldehyde 3-phosphate. The reason why the edges from `FBA` and `PFK` are not bidirectional is that these reactions are also used in other pathways.

module-finding algorithms that do not use thermodynamic constraints, like the method suggested by Kelk *et al.* (2012).

We remark that our new method only works well for the optimal flux space. Although theoretically it can also be applied to analyze the full flux space, it will most probably only compute one module. Fluxes in the optimal flux space usually have the following property: If v is a feasible flux vector, then αv with $0 < \alpha < 1$ is also a feasible flux vector. It follows that the interface flux of each module can only be 0. This implies that the network is simply the union of several networks that do not interact. For well curated models this is very unlikely to be the case.

6 Related Work

Our work was highly inspired by the work of Kelk *et al.* (2012). Both methods decompose the optimal flux space of a metabolic network into independent modules. However, there are some differences next to algorithmic issues that we want to point out.

Using the mathematical theory introduced in this paper, the idea of the method by Kelk *et al.* (2012) can be considered an application of Thm. 2, or more precisely of the following corollary:

Corollary 3 *Let $P := \{v : Sv = b, \ell \leq v \leq u\}$ or $P := T$ with $P \neq \emptyset$ and $\ell_{\mathcal{I}} \leq 0 \leq u_{\mathcal{I}}$. Let $\mathcal{X}$ be a partition of $\mathcal{R}$ into P-modules. For every $A \in \mathcal{X}$ let $\Omega_A = \mathrm{pr}_A(P)$ be the sample space of a fixed but arbitrary probability space. Let $B \in \mathcal{X}$, $r \in B, s \in \mathcal{R} \setminus B$ be arbitrary but fixed. Let*

$$E_1 = \{v \in P : v_r \in X\}, \qquad E_2 = \{v \in P : v_s \in Y\}$$

be events, where $X, Y \subseteq \mathbb{R}$. Then, E_1, E_2 are independent in the product probability space $P = \prod_{A \in \mathcal{X}} \Omega_A$. □

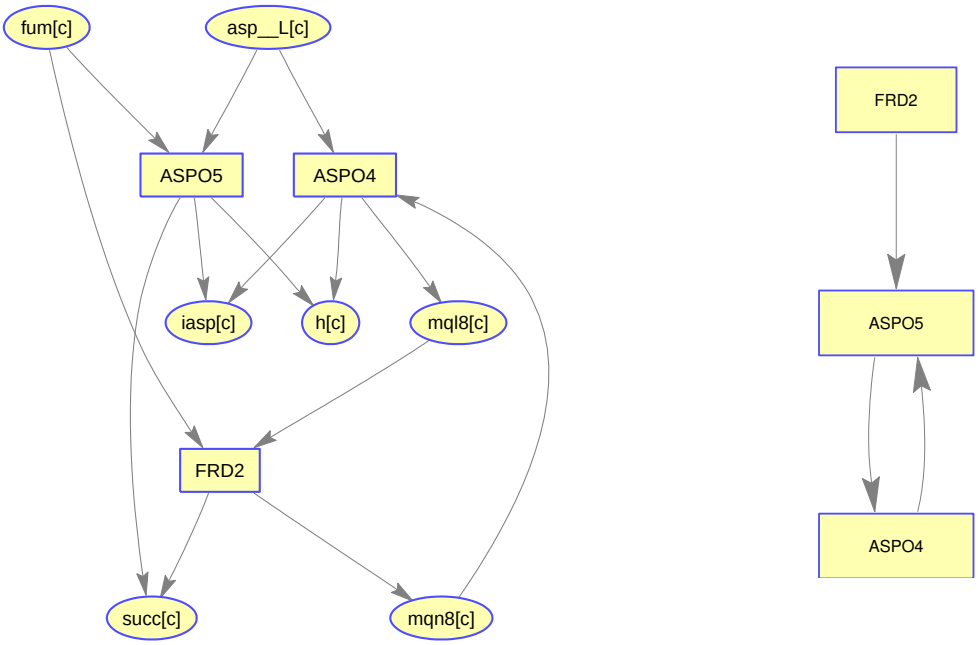

Figure 7: In anaerobic growth of *E.coli* iAF1260 under *glucose*, *fumarate reductase* (FRD2) and two kinds of *L-aspartate oxidase* (ASPO4, ASPO5) already form a module. Under aerobic growth, these are only part of a bigger module, which is shown in Fig. 5. This module essentially transforms *fumarate* and *L-aspartate* into *iminoaspartate* and *succinate*. FRD2 again is also used in a different pathway and additionally transforms small amounts of *menaquinol 8* into *menaquinone 8*.

PROOF By Thm. 2, we have $P = \prod_{A \in \mathcal{X}} \Omega_A$. The independence follows directly from the definition of product probability space. ■

Let us consider the following discrete probability measure for each module A (under the assumption $b \neq 0$ it follows that $\mathrm{EFM}(P)$ is finite):

$$Pr(v \in \mathrm{pr}_A(P)) := \frac{|\{w \in \mathrm{EFM}(P) : w_A = v\}|}{|\mathrm{EFM}(P)|},$$

where $P = \{v \in \mathbb{R}^{\mathcal{R}} : Sv = b,\ \ell \leq v \leq u\}$ and $b \neq 0$. It follows by Cor. 3 that the random variables $X_r, X_s : \mathbb{R}^{\mathcal{R}} \to \mathbb{R}$ with $X_r : v \mapsto v_r$, $X_s : v \mapsto v_s$ are independent if r and s belong to different modules. It follows that X_r and X_s can only be correlated if r and s belong to the same module.

By Thm. 4 it follows that the elementary flux modes are uniformly distributed in the product probability space P, i.e.,

$$Pr(v \in P) = \begin{cases} \frac{1}{|\mathrm{EFM}(P)|} & v \in \mathrm{EFM}(P) \\ 0 & \text{otherwise.} \end{cases}$$

Thus, X_r, X_s are exactly those flux variables that Kelk *et al.* (2012) used to compute flux correlations (if one ignores the fact that they modify the network first to get rid of linealities).

Note, that also our method can be considered an application of Cor. 3. The main difference between our method and Kelk *et al.* (2012) can hence be considered in the choice of probability measures and events that are used to detect interactions between reactions.

Metabolite	interface flux aerobic module	interface flux anaerobic module
L-Aspartate	-0.001678	-0.0005999
Iminoaspartate	0.001678	0.0005999
Fumarate	3.729	-0.08771
Succinate	-3.729	0.08771
H^+	-114.2	0.0005999
H^+ *(periplasm)*	85.67	0
Glyoxylate	0.0004929	0
Glycolate	-0.0004929	0
NAD	28.56	0
NADH	-28.56	0
(S)-Dihydroorotate	-0.2437	0
Orotate	0.2437	0
Ubiquinone-8	-32.53	0
Ubiquinol-8	32.53	0
Menaquinol 8	0	-0.08711
Menaquinone 8	0	0.08711

Table 1: Interface flux comparison of the Modules shown in Fig. 5 and Fig. 7. All metabolites are cytosolic except where stated otherwise. The flux units are in mmol/gDW/h (mmol per gram dry weight per hour). By definition of module these interface fluxes are constant for all optimal flux vectors under the corresponding growth condition.

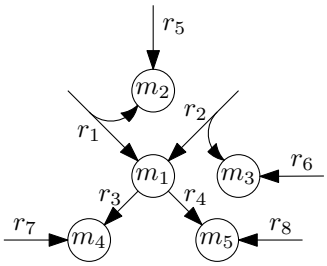

Figure 8: The proposed method computes that $\{r_1, r_5\}$, $\{r_2, r_6\}$, $\{r_3, r_7\}$, $\{r_4, r_8\}$ are minimal modules. However, this network only consists of exactly one minimal module.

Next to the fact that flux variability analysis is much easier to compute (even with thermodynamic constraints) than the enumeration of all elementary flux modes, an additional algorithmic difficulty is posed by internal cycles in the network. We solved the problem by using thermodynamic constraints. Kelk *et al.* (2012) handled the problem by first eliminating the cycles using a projection step. Hence, both methods do not solve the original problem directly, but slightly modified variants.

In the case of Kelk *et al.* (2012), the actual modules may be obtainable by integrating information of the lineality space and the detected rays back into the computed modules. This is however not discussed in the original paper and it is not clear how it can be done in general. For example, the module {`ALATA_L`, `VALTA`, `VPAMT`}, which we found in all *E. coli* models (under various growth conditions) was detected by Kelk *et al.* (2012) as a ray and hence did not appear in the list of modules. It should be noted that not every ray is a module. For example, in all *E. coli* networks, {`ACCOAL`, `SUCOAS`, `PPCSCT`} was detected as a ray, but in only some of the networks we detected this also as a module. In particular, this is not recognizable by the results presented by Kelk *et al.* (2012), since there always exists a different ray that involves `ACCOAL`.

Model	no. reactions	no. modules	no. efms	run time
E. coli iJR904 on *Glucose*	1075	8	$2^6 \times 90 \times 96$	162*s*
E. coli iJR904 on *L-Threonine*	1075	12	$2^{10} \times 6 \times 90$	152*s*
E. coli iJR904 on *L-Arginine*	1075	11	$2^8 \times 6^2 \times 9$	139*s*
E. coli iJR904 on *Citrate*	1075	11	$2^8 \times 6 \times 15 \times 90$	138*s*
E. coli iJR904 on *Fumarate*	1075	11	$2^8 \times 6 \times 90$	144*s*
E. coli iJR904 on *L-glutamine*	1075	10	$2^7 \times 6 \times 90 \times 222$	145*s*
E. coli iJR904 on *Lactose*	1075	8	$2^6 \times 8 \times ?$	160*s*
E. coli iJR904 on *L-Malate*	1075	10	$2^7 \times 6 \times 15 \times 90$	135*s*
E. coli iJR904 on *L-Tryptophan*	1075	7	$2^5 \times 18 \times 96$	136*s*
E. coli iAF1260 on *Glucose*, aerobic	2382	9	$2^4 \times 3^2 \times 6 \times 54 \times 5184$	
E. coli iAF1260 on *Glucose*, anaerobic	2382	9	$2^6 \times 3 \times 6 \times 2592$	531*s*
E. coli iAF1260 on *Glucose*, limited oxygen	2382	6	$2^3 \times 3 \times 2592 \times ?$	976*s*
E. coli iAF1260 on *L-Threonine*, aerobic	2382	9	$2^4 \times 3 \times 8^2 \times 108 \times 4944$	6836*s*
H. pylori iIT341	554	5	$2^4 \times 18$	
M. barkeri iAF62	690	7	$2^4 \times 3 \times 28 \times 156$	941*s*
M. tuberculosis iNJ661	1025	10	$2^6 \times 3^3 \times ?$	1623*s*
S. aureus iSB619	743	10	$2^7 \times 4 \times 192$	
S. cerevisiae iND750	1266	8	$2^4 \times 5 \times 6 \times 80 \times ?$	

Table 2: For each of the analyzed networks the table shows the number of computed modules in the optimal flux space with respect to the specified growth condition. If no growth condition is specified, the default from the BiGG-database (Schellenberger *et al.*, 2011) was used. We also computed the number of optimal elementary flux modes through each module. Since every combination of elementary flux modes of the modules gives an optimal elementary flux mode of the whole network, we did not compute the product and simply stated the factors. We use '?' to denote that the we were not able to compute the elementary flux modes through a module, because there were too many elementary flux modes (≥ 300). We remark that we used the original networks from the BiGG-database, where no duplicate reactions were removed.

The modules of Kelk *et al.* (2012) have the property that it is possible to combine the vertices of the modules together to obtain the vertices of the whole projected flux space. However, this does not work for the original space, since this space may contain linealities. With our decomposition and definition of elementary flux mode, we actually can combine the elementary flux modes of the modules together to obtain the elementary flux modes of the optimal flux space.

The addition of thermodynamic constraints eliminates thermodynamically infeasible cycles and hence may also fix reactions that would otherwise be part of modules or connect different modules. Hence, our method may find smaller modules than the method by Kelk *et al.* (2012). However, thermodynamically constrained flux balance analysis is still an **NP**-hard problem (Müller and Bockmayr, 2013). There are many networks where even fast-tFVA needs more than half an hour of computation time. Since we need to run many iterations of tFVA, this may make the algorithm very slow. If we are only interested in the modules however, there is no need to compute all interactions between all reactions. Hence, we assume that there is a lot of potential that can be used to speed up the algorithm.

7 Conclusion

We introduced a mathematical theory for the analysis of modules in the optimal flux space of a metabolic network. We gave a nice and clean definition of module (invariance of interface fluxes) and derived properties on the decomposition of the network into modules and the structure of the flux space.

We showed that these modules can safely be analyzed independently from each other. In such an analysis each module can be considered a metabolic network on its own. We showed that if we have feasible fluxes for each module separately, these fluxes together also form a feasible flux in the whole network.

In addition, we also proved that we only need to compute elementary flux modes for each of the modules separately. These elementary flux modes can then be easily combined to describe the set of all elementary flux modes in the network. Hence, a full enumeration of elementary flux modes is not necessary anymore. We applied this on the optimal flux space of *E. coli* iJR904 on two different growth conditions. There, it gives us a description of the set of optimal elementary flux modes that can be even analyzed manually.

Using our mathematical framework, we derived a novel method for computing modules of the optimal flux space. We also showed how the method by Kelk *et al.* (2012) can be understood as an application of the theory that we introduced. However, it was only possible with the formal definition of module that we were able to mathematically prove the properties of the modules that they found.

Another concept introduced in this paper are blocking graphs. Blocking graphs are an alternative method to analyze the possible pathways of modules without enumerating all elementary flux modes. From this, we were able to read off alternative pathways directly.

We demonstrated our methods to compute and analyze the modules of several genome-scale networks. We showed that blocking graphs are a useful tool for the analysis of metabolic networks that complements flux coupling analysis. Also the definition of module itself proved to be very useful since the interface flux already gives a direct interpretation of the function of the module in the whole network.

Unfortunately, our methods only work well on the optimal flux space, since the full flux space is usually not decomposable into modules as we defined them (see Sect. 5.2). Suitable generalizations of the notion of module as defined in this paper, however, might be used to decompose also the whole flux space. This would give us the opportunity to also compute a compact description of the whole set of elementary flux modes directly. Thus, the problem of enumerating elementary flux modes may become tractable in this way even for genome-scale networks.

Finally we want to point out that our method relies on many runs of tFVA. We think that this is not really necessary and the algorithm can be sped up considerably. This and the extension to the whole flux space will be the focus of our future work.

Acknowledgements

We thank Leen Stougie, who presented this problem on the ENUMEX Summerschool (Enumeration Algorithms & Exact Methods For Exponential Problems in Computational Biology).

Funding: This work was funded by the Berlin Mathematical School in terms of a PhD stipend.

References

Beard, D. A., Babson, E., Curtis, E., and Qian, H. (2004). Thermodynamic constraints for biochemical networks. *Journal of Theoretical Biology*, **228**, 327–333.

Burgard, A. P., Vaidyaraman, S., and Maranas, C. D. (2001). Minimal reaction sets for escherichia coli metabolism under different growth requirements and uptake environments. *Biotechnology Progress*, **17**, 791–797.

Burgard, A. P., V., N. E., Schilling, C. H., and Maranas, C. D. (2004). Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Research*, **14**(2), 301312.

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

Durot, M., Bourguignon, P.-Y., and Schachter, V. (2009). Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS Microbiology Reviews*, **33**, 164–90.

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

Fleming, R. M., Maes, C. M., Saunders, M. A., Ye, Y., and Palsson, B. O. (2012). A variational principle for computing nonequilibrium fluxes and potentials in genome-scale biochemical networks. *Journal of Theoretical Biology*, **292**, 71–77.

Grünbaum, B. (2003). *Convex Polytopes*. Graduate Texts in Mathematics. Springer, 2 edition.

Kelk, S. M., Olivier, B. G., Stougie, L., and Bruggeman, F. J. (2012). Optimal flux spaces of genome-scale stoichiometric models are determined by a few subnetworks. *Scientific Reports*, **2**, 580.

Khannapho, C., Zhao, H., Bonde, B. L., Kierzek, A. M., Avignone-Rossa, C. A., and Bushell, M. E. (2008). Selection of objective function in genome scale flux balance analysis for process feed development in antibiotic production. *Metabolic Engineering*, **10**(5), 227–233.

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

Larhlimi, A., David, L., Selbig, J., and Bockmayr, A. (2012). F2c2: a fast tool for the computation of flux coupling in genome-scale metabolic networks. *BMC Bioinformatics*, **13**, 57.

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

Müller, A. (2012). *Thermodynamic Constraints in Metabolic Networks*. Master's thesis, Freie Universität Berlin, Fachbereich Mathematik und Informatik. `http://page.mi.fu-berlin.de/arnem/theses/master.pdf`.

Müller, A. and Bockmayr, A. (2013). Fast thermodynamically constrainted flux variability analysis. *Bioinformatics*, **29**(7), 903–909.

Noor, E., Lewis, N. E., and Milo, R. (2012). A proof for loop-law constraints in stoichiometric metabolic networks. *BMC Systems Biology*, **6**, 140.

Orth, J. D., Thiele, I., and Palsson, B. O. (2010). What is flux balance analysis. *Nature Biotechnology*, **28**, 245248.

Papin, A. J., Stelling, J., Price, N. D., Klamt, S., Schuster, S., and Palsson, B. O. (2004). Comparison of network-based pathway analysis methods. *TRENDS in Biotechnology*, **22**(8), 400–405.

Papin, Jason A. Reed, J. L. and Palsson, B. O. (2004). Hierarchical thinking in network biology: the unbiased modularization of biochemical networks. *TRENDS in Biochemical Sciences*, **29**(12), 641–647.

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

Sarıyar, B., Perk, S., Akman, U., and Hortaçsu, A. (2006). Monte carlo sampling and principal component analysis of flux distributions yield topological and modular information on metabolic networks. *Journal of Theoretical Biology*, **242**, 389400.

Schellenberger, J., Lewis, N. E., and Palsson, B. Ø. (2011). Elimination of thermodynamically infeasible loops in steady-state metabolic models. *Biophysical Journal*, **100**, 544–553.

Schilling, C. H., Letscher, D., and Palsson, B. O. (2000). Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function form a pathway-oriented perspective. *Journal of Theoretical Biology*, **203**, 229–248.

Schuster, S. and Hilgetag, C. (1994). On elementary flux modes in biochemical systems at steady state. *J. Biol. Systems*, **2**, 165–182.

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

Schuster, S., Pfeiffer, T., and Fell, D. A. (2007). Is maximization of molar yield in metabolic networks favoured by evolution? *Journal of Theoretical Biology*, **252**(3), 497–504.

Terzer, M. and Stelling, J. (2008). Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics*, **24**(19), 2229–2235.

Terzer, M., Maynard, N. D., Covert, M. W., and Stelling, J. (2009). Genome-scale metabolic networks. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, **1**(3), 285–297.

Teusink, B., Wiersma, A., Jacobs, L., Notebaart, R., and Smid, E. (2009). Understanding the adaptive growth strategy of *Lactobacillus plantarum* by in silico optimisation. *PLoS Computational Biology*, **5**(6), e1000410.

Varma, A. and Palsson, B. O. (1994). Metabolic flux balancing: Basic concepts, scientific and practical use. *Nature Biotechnology*, **12**, 994–998.

von Kamp, A. and Schuster, S. (2006). Metatool 5.0: fast and flexible elementary modes analysis. *Bioinformatics*, **22**(15), 1930–1931.