Qualitative and Thermodynamic Flux Coupling Analysis

Arne C. Reimers [1,3,4,*] Yaron A. B. Goldstein [1,2] Alexander Bockmayr [1,2]

March 27, 2014

Abstract

Flux coupling analysis (FCA) has become a useful tool for aiding metabolic reconstructions and guiding genetic manipulations. Originally, it was introduced for constraint-based models of metabolic networks that are based on the steady-state assumption. Recently, we have shown that the steady-state assumption can be replaced by a much weaker lattice-theoretic property related to the supports of metabolic fluxes. In this paper, we further extend our approach and delevelop an efficient algorithm for general qualitative flux coupling analysis (QFCA). We illustrate our method by thermodynamic flux coupling analysis (tFCA), which allows studying steady-state metabolic models with loop-law thermodynamic constraints. These models do not satisfy the lattice-theoretic properties required in our previous work. For a selection of genome-scale metabolic network reconstructions, we discuss both theoretically and practically, how thermodynamic constraints strengthen the coupling results that can be obtained with classical FCA.

1 Introduction

Constraint-based modeling has become a widely used approach for the analysis of genome-scale reconstructions of metabolic networks [16, 4]. Given a set $\mathcal{M}$ of *metabolites* and a set $\mathcal{R}$ of *reactions*, the metabolic network is modeled by its stoichiometric matrix $S \in \mathbb{R}^{\mathcal{M}\times\mathcal{R}}$ and a set of irreversible reactions $\mathrm{Irr} \subseteq \mathcal{R}$. Based on this description, constraints are used to characterize the space of possible metabolic behaviors. The classical starting point of constraint-based modeling is the *steady-state* (or mass balance) constraint $Sv = 0$. It states that every metabolite has to be produced at the same rate as it is consumed. The set

$$C = \{v \in \mathbb{R}^{\mathcal{R}} \mid Sv = 0, v_{\mathrm{Irr}} \geq 0\}$$

is called the *steady-state flux cone*. It contains all flux distributions $v \in \mathbb{R}^{\mathcal{R}}$ satisfying the stoichiometric and irreversibility constraints.

A prominent example of constraint-based methods is *flux balance analysis* (FBA) [26]. Here, linear programming is used to predict how efficiently an organism can realize a certain biological objective, e.g., how much biomass can be produced out of a limited amount of nutrients. FBA can be used to predict the impact of gene or reaction knockouts in an organism. However, typically not all reactions in a metabolic network can carry flux independently from each other. In other words, by knocking out one reaction, one may implicitly disable flux through other

*e-mail: arne.c.reimers@gmail.com
[1] FB Mathematik und Informatik, Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany
[2] DFG-Research Center Matheon, Berlin, Germany
[3] Berlin Mathematical School (BMS), Berlin, Germany,
[2] International Max Planck Research School for Computational Biology and Scientific Computing, Berlin, Germany

reactions. Detecting this kind of dependencies is important for the identification of knock-out targets, because some reactions may be easier to manipulate than others [12]. In addition, dependency information can also be used to check the consistency of metabolic network reconstructions [5] or to find co-regulated reactions [20].

Flux coupling analysis (FCA) [6] was introduced to comprehensively analyze these kinds of dependencies between reactions. A reaction $s \in \mathcal{R}$ is called *blocked* if $v_s = 0$, for all $v \in C$, otherwise s is called *unblocked*. Burgard et al. [6] defined three types of coupling relations for a given pair of unblocked reactions (r, s):

- s is *directionally coupled* to r, written $s \to r$, if for all $v \in C$, $v_s \neq 0$ implies $v_r \neq 0$. Equivalently, we could require that for all $v \in C$, $v_r = 0$ implies $v_s = 0$, which is denoted by $r \xrightarrow{=0} s$ in [15] or [14].
- r and s are *partially coupled*, written $r \leftrightarrow s$, if both $r \to s$ and $s \to r$ hold.
- r and s are *fully coupled*, if there exists $\lambda \neq 0$ with $v_r = \lambda v_s$, for all $v \in C$.
- r and s are *uncoupled*, if neither $r \to s$ nor $s \to r$ holds.

Due to a number of theoretical and algorithmic improvements [15, 9, 14], it is now possible to perform FCA on large genome-scale metabolic network reconstructions in a few minutes of computation time on a standard desktop computer.

As can be seen from the definition, directional coupling does not depend on the precise amount of flux through a given reaction. It only matters whether there is a flux different from zero or not. In this sense, FCA is a *qualitative* method. As we will see, FCA can also be applied to more general qualitative models of metabolic networks such as those introduced in [24, 7]. These models do not use the steady-state assumption because it turns out to be too strong for certain applications.

Goldstein et al. [11] generalized FCA to constraint-based models that do not have to satisfy the steady-state assumption, but instead fulfill certain lattice-theoretic properties related to the supports of the metabolic fluxes. Given a flux vector $v \in \mathbb{R}^{\mathcal{R}}$, its *support* is defined by

$$\operatorname{supp}(v) := \{i \in \mathcal{R} \mid v_i \neq 0\}.$$

By working with $\operatorname{supp}(v)$ instead of v, flux coupling analysis can be performed in a qualitative way. While lattices are a very general concept [8], we have to consider here only finite lattices L with reaction sets $a \subseteq \mathcal{R}$ as elements. Let $2^{\mathcal{R}}$ denote the power set of $\mathcal{R}$. A family of reaction sets $L \subseteq 2^{\mathcal{R}}$ defines a *lattice* if $\emptyset \in L$ and if L is union-closed, i.e., $a, b \in L$ implies $a \cup b \in L$. Note that each reaction set $a \in L$ can be naturally interpreted as a metabolic pathway consisting exactly of the reactions $r \in a$. For standard FCA, the *flux lattice* L^C is defined by the supports of the steady-state flux vectors,

$$L^C := \{\operatorname{supp}(v) \mid Sv = 0, v_{\mathrm{Irr}} \geq 0\}.$$

Given an arbitrary lattice $L \subseteq 2^{\mathcal{R}}$, we generalize the notion of flux coupling in the following way [11]. A reaction $s \in \mathcal{R}$ is *blocked in* L if $s \notin a$, for all $a \in L$, otherwise s is *unblocked in* L. Given two reactions r and s that are unblocked in L, we say that s is *directionally coupled in* L to r if and only if

$$\forall a \in L : s \in a \Rightarrow r \in a. \tag{1}$$

In the case of $L = L^C$, this corresponds exactly to standard FCA as introduced before.

The generalization of FCA to lattices already allows analyzing a wide range of metabolic network models. For example, we may use lattices to perform FCA on constraint-based models

with lower and upper bounds on the flux rates [11]. However, there are interesting cases where the lattice assumption is still too strong. Consider the network in Fig. 1. Here, reaction 1 is not directionally coupled to reaction 3, because $v = (0, 0, 1, 1)$ is a steady-state flux. However, this flux is an internal circulation, which violates the second law of thermodynamics. As observed by Beard et al. [2], a thermodynamically feasible flux must not contain any internal circulation. If we add thermodynamic constraints, reaction 3 becomes fully coupled to 1 and thus, we get a stronger result than without these constraints.

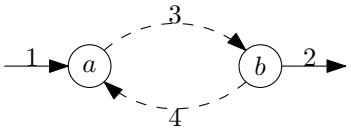

Figure 1: Here, all reactions are irreversible. Without thermodynamic constraints, reaction 1 is not directionally coupled to reaction 3, since reactions 3 and 4 form an internal circuit (dashed arrows). With thermodynamic constraints, reactions 1 and 3 are fully coupled.

The key property of lattices is that metabolic pathways can be combined by taking the union of the supports. However, if the pathways have to satisfy thermodynamic constraints, this is not always possible. For example, in Fig. 2, the pathways $\{r, b, s, d, e\}, \{r, a, s, c, e\}$ are thermodynamically feasible, but the combination $\{r, a, b, s, c, d, e\}$ is not thermodynamically feasible, since it contains the internal circuit $\{r, a, d\}$.

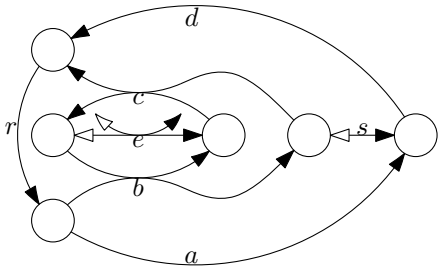

Figure 2: The only exchange reaction in this network is e and the white arrow-heads indicate reversible reactions. If r carries thermodynamically feasible flux, then s must also carry flux. But, s can carry both positive or negative flux.

In this paper, we show how to overcome this problem. We develop an efficient algorithm for general *qualitative flux coupling analysis* (QFCA), where we do not require the lattice-theoretic axioms. As a concrete instance of our framework, we present *thermodynamically constrained flux coupling analysis* (tFCA) and apply it to a number of genome-scale metabolic network reconstructions.

2 Qualitative FCA

The definition of directional coupling in (1) could also be applied to the family $T \subseteq 2^{\mathcal{R}}$ of supports of thermodynamically feasible fluxes (see Sect. 3 for a formal definition). However, as shown in Sect. 1, T need not be a lattice, and thus we cannot directly use the results of [11].

2.1 QFCA for $P \subseteq 2^{\mathcal{R}}$

In this section, we present a qualitative form of FCA that works not only for the supports $T \subseteq L^C$ of the thermodynamically feasible fluxes, but for any $\emptyset \neq P \subseteq 2^{\mathcal{R}}$. First we define

$$s \text{ blocked in } P :\Leftrightarrow \forall a \in P : s \notin a, \text{ and}$$
$$r \xrightarrow{=0} s \text{ in } P :\Leftrightarrow \forall a \in P : r \notin a \Rightarrow s \notin a.$$

Next we introduce the set of *irreducible* elements in P

$$\mathcal{J}(P) := \left\{ b \in P \setminus \{\emptyset\} \;\middle|\; \forall A \subseteq P : b = \bigcup_{a \in A} a \Rightarrow b \in A \right\},$$

and the set $\mathcal{M}(P)$ of (non-trivial) *minimal* elements in P

$$\mathcal{M}(P) := \{ e \in P \setminus \{\emptyset\} \mid \forall a \in P : a \subsetneq e \Rightarrow a = \emptyset \}.$$

Note that $\mathcal{M}(P) \subseteq \mathcal{J}(P)$, but in general not $\mathcal{M}(P) = \mathcal{J}(P)$. The *closure* of P is defined by

$$\langle P \rangle = \left\{ \bigcup_{a \in A} a \;\middle|\; A \subseteq P \right\}.$$

It is easy to see that $\langle P \rangle$ is the smallest lattice that contains P. We say P is a *generator* of $\langle P \rangle$. We observe that for any lattice L, $\mathcal{J}(L)$ is the unique minimal generator, thus $L = \langle \mathcal{J}(L) \rangle$. It follows $\langle \mathcal{J}(P) \rangle = \langle P \rangle$ and $\mathcal{J}(P) = \mathcal{J}(\langle P \rangle)$. Now we can state our first result:

Theorem 1. *Consider any $\emptyset \neq P \subseteq 2^{\mathcal{R}}$ and suppose $B \subseteq 2^{\mathcal{R}}$ such that $\mathcal{J}(P) \subseteq B \subseteq \langle P \rangle$. For a reaction $s \in \mathcal{R}$, the following are equivalent:*

(a) *s blocked in P,*

(b) *s blocked in $\langle P \rangle$,*

(c) *s blocked in B,*

(d) *s blocked in $\langle B \rangle$.*

For two reactions r, s that are unblocked in P, the following are equivalent:

(a) *$r \xrightarrow{=0} s$ in P,*

(b) *$r \xrightarrow{=0} s$ in $\langle P \rangle$,*

(c) *$r \xrightarrow{=0} s$ in B,*

(d) *$r \xrightarrow{=0} s$ in $\langle B \rangle$.*

Proof. Since $\mathcal{J}(P) \subseteq B \subseteq \langle P \rangle$, we have $\langle P \rangle = \langle B \rangle$. Thus, it is sufficient to prove (c) $\Leftrightarrow$ (d).

$\Rightarrow$: Assume s is unblocked (resp. s is not directionally coupled to r) in $\langle B \rangle$. By definition, there exists $a \in \langle B \rangle$ such that $s \in a$ (resp. $s \in a \not\ni r$). Since $\mathcal{J}(\langle B \rangle)$ is a generator of $\langle B \rangle$, there exists $b \in \mathcal{J}(\langle B \rangle)$ with $s \in b$ (resp. $s \in b \not\ni r$). Since $b \in \langle B \rangle$ and by definition of $\langle B \rangle$, there exists $A \subseteq B \subseteq \langle B \rangle$ with $b = \bigcup_{a \in A} a$. Since b is irreducible, it follows that $b \in A$ and thus $b \in B$. This proves that s is unblocked (resp. s is not directionally coupled to r) in B.

$\Leftarrow$: Assume s is unblocked (resp. s is not directionally coupled to r) in B. Then there exists $a \in B$ such that $s \in a$ (resp. $s \in a \not\ni r$). Since $B \subseteq \langle B \rangle$, we have $a \in \langle B \rangle$. It follows that s is unblocked (resp. s is not directionally coupled to r) in $\langle B \rangle$. $\square$

2.2 Algorithm for QFCA in P

In [11], we presented a generic algorithm for finding blocked reactions and determining flux coupling pairs, which works for arbitrary lattices $L \subseteq 2^{\mathcal{R}}$. In order to use this algorithm in a particular lattice L, we only have to provide a method `test(s)` (resp. `test(r,s)`) that returns a lattice element $a \in L$ with $s \in a$ (resp. $s \in a \not\ni r$), if such elements exists, and $\emptyset$ otherwise. Here we show that a similar result holds for flux coupling analysis in P. Given $B \subseteq 2^{\mathcal{R}}$ with $\mathcal{J}(P) \subseteq B \subseteq \langle P \rangle$, we only have to define a method $\texttt{testBlocked}_B(s)$ (resp. $\texttt{testCoupled}_B(r,s)$) that finds $a \in B, s \in a$ (resp. $s \in a \not\ni r$), if such an element exists, and returns $\emptyset$ otherwise.

Theorem 2. *Assume $\emptyset \neq P \subseteq L$ and $\mathcal{J}(P) \subseteq B \subseteq \langle P \rangle$. Let $\texttt{testBlocked}_B : \mathcal{R} \to 2^{\mathcal{R}}$ be a function satisfying*

$$\texttt{testBlocked}_B(s) = \begin{cases} a, & \text{if } \exists a \in B \ : \ s \in a, \\ \emptyset, & \text{otherwise.} \end{cases}$$

Let $\texttt{testCoupled}_B : \mathcal{R} \times \mathcal{R} \to 2^{\mathcal{R}}$ be a function satisfying

$$\texttt{testCoupled}_B(r,s) = \begin{cases} a, & \text{if } \exists a \in B \ : \ s \in a \not\ni r, \\ \emptyset, & \text{otherwise.} \end{cases}$$

Then a reaction s is blocked in P if and only if $\texttt{testBlocked}_B(s) = \emptyset$. Two unblocked reactions r, s are directionally coupled in P, i.e., $r \xrightarrow{=0} s$ in P holds, if and only if $\texttt{testCoupled}_B(r,s) = \emptyset$.

Proof. We show only the last part. By Thm. 1, we know that $r \xrightarrow{=0} s$ in P holds if and only if $r \xrightarrow{=0} s$ in B. From the definition of $\texttt{testCoupled}_B$, we get $\texttt{testCoupled}_B(r,s) = \emptyset$ if and only if $r \xrightarrow{=0} s$ in B. □

Corollary 1. *The algorithm introduced in [11] implemented by using the functions $\texttt{testBlocked}_B$ and $\texttt{testCoupled}_B$ performs FCA for P.*

3 Thermodynamic Constraints

Now we apply the framework from Sect. 2 to develop an algorithm for thermodynamic FCA (tFCA). We will work with the relaxed form of thermodynamic constraints introduced by Beard et al. [2], who use the following formulation:

$$Sv = 0, \quad \text{steady-state} \tag{2}$$

$$v_{\text{Irr}} \geq 0, \quad \text{irreversible reactions} \tag{3}$$

$$\Delta\mu_i v_i < 0 \vee v_i = 0, \ \forall i \in \mathcal{I}, \quad \text{thermodyn. constr.} \tag{4}$$

$$\Delta\mu = \mu^T S_{\mathcal{I}}, \quad \text{potential differences} \tag{5}$$

$$v \in \mathbb{R}^{\mathcal{R}}, \mu \in \mathbb{R}^{\mathcal{M}},$$

where $\mathcal{I}$ denotes the set of internal reactions (i.e., reactions that are not exchange reactions), $S_{\mathcal{I}}$ the submatrix of the stoichiometric matrix S corresponding to the reactions in $\mathcal{I}$, and μ the chemical potential of each metabolite (Gibbs free energy of formation). Given a metabolic network $\mathcal{N} = (\mathcal{M}, \mathcal{R}, \mathcal{I}, S, \text{Irr})$, a flux vector $v \in \mathbb{R}^{\mathcal{R}}$ is called *thermodynamically feasible* if there exists a vector $\mu \in \mathbb{R}^{\mathcal{M}}$ such that (2), (3), (4), and (5) are satisfied.

By multiplying μ^T from the left side with $S_\mathcal{I}$ ($\cdot^T$ denotes transposition), the potential differences for internal reactions are obtained. This is equivalent to the often found formulation $K\Delta\mu = 0$, where K is the null-space matrix of $S_\mathcal{I}$. The motivation behind (4) is that normally a chemical reaction carries flux if and only if it reduces Gibbs free energy [3, 1, 21]. Since many reactions are catalyzed by enzymes, however, it can happen that an enzyme, for example because of regulatory control, is not present; and hence, the corresponding reaction is effectively not possible and carries essentially no flux, even if there is a negative potential difference. Thus, zero flux is always allowed, too. Note that other formulations of thermodynamic constraints (for example in [10]) do not have this property, .

As it can easily be seen from the definition, thermodynamic feasibility depends only on the sign of the fluxes. This is the idea underlying the use of oriented matroids as suggested by Beard et al. [2]. Let $\operatorname{sign}(v) \in \{-, 0, +\}^\mathcal{R}$ denote the vector of signs of $v \in \mathbb{R}^\mathcal{R}$. For example, for $v = (3, 5, 0, -1, 2)$ we have $\operatorname{sign}(v) = (+, +, 0, -, +)$. Given a sign vector $A \in \{-, 0, +\}^\mathcal{R}$, we will also write $A = (A^+, A^-)$ with

$$A^+ = \{r \in \mathcal{R} : A_r = +\}, \;\; A^- = \{r \in \mathcal{R} : A_r = -\}.$$

Since we want to talk about fluxes contained in other fluxes, we define the following inclusion relation:

$$(A^+, A^-) \subseteq (B^+, B^-) \text{ if and only if } A^+ \subseteq B^+ \text{ and } A^- \subseteq B^-.$$

Note that if all reactions are irreversible (and thus all sign vectors of feasible fluxes are non-negative), then this subset relation is equivalent to the ordinary subset relation on the support of the fluxes.

It has been shown [2, 19, 18] that a flux v is thermodynamically feasible if and only if v does not contain an *internal circulation*, i.e., a flux vector $w \in \mathbb{R}^\mathcal{I} \setminus \{0\}$ with $S_\mathcal{I} w = 0$ and $w_{\mathcal{I}\cap\mathrm{Irr}} \geq 0$:

Theorem 3. *Given a metabolic network $\mathcal{N}$ = ($\mathcal{M}$, $\mathcal{R}$, $\mathcal{I}$, S, Irr), a flux vector $v \in \mathbb{R}^\mathcal{R}$ satisfying* (2) *and* (3) *is thermodynamically feasible if and only if there is no $w \in \mathbb{R}^\mathcal{I} \setminus \{0\}$ with* $\operatorname{sign}(w) \subseteq \operatorname{sign}(v_\mathcal{I})$ *and $S_\mathcal{I} w = 0$.*

Using this characterization we define the *space T of thermodynamically feasible fluxes* as follows:

$$T := \left\{ \operatorname{supp}(v) : \begin{array}{c} Sv = 0, v_\mathrm{Irr} \geq 0, \\ \nexists w \neq 0 : S_\mathcal{I} w = 0, \operatorname{sign}(w) \subseteq \operatorname{sign}(v) \end{array} \right\}$$

4 Thermodynamic FCA

For our implementation of thermodynamically constrained FCA, we do not operate directly on the space T of thermodynamically feasible fluxes. Instead, we make use of Thm. 1 and work on the space

$$B := \left\{ \operatorname{supp}(v) : \begin{array}{c} Sv = 0, v_\mathrm{Irr} \geq 0, \\ \nexists w \neq 0 : S_\mathcal{I} w = 0, w_{\mathcal{I}\cap\mathrm{Irr}} \geq 0, \\ \operatorname{supp}(w) \subseteq \operatorname{supp}(v) \end{array} \right\}.$$

Since $\operatorname{sign}(w) \subseteq \operatorname{sign}(v)$ implies $\operatorname{supp}(w) \subseteq \operatorname{supp}(v)$, we get immediately $B \subseteq T$. However, B can be strictly smaller than T. For example, we may have flux through parallel, reversible reactions in T. This is not allowed in B, because parallel, reversible reactions together form an internal circuit.

To apply Thm. 1, we have to show that the irreducible elements of T are contained in B. First, we note that the irreducible elements of T form a subset of the *elementary modes* defined as

$$E := \mathcal{M}\left(L^C\right) = \left\{e \in L^C \mid \forall a \in L^C \ : \ a \subsetneq e \Rightarrow a = \emptyset\right\}.$$

This is an immediate consequence of the next lemma.

Lemma 4.1. *For every $v \in T$, there exist elementary modes $e_1, \ldots, e_k \in E \cap T$ with $v = \bigcup_{i=1}^{k} e_i$.*

Proof. This lemma follows directly from Lemma 4 in [18]. Alternatively, the notion of feasibility classifier as introduced by Terzer in [25] and used in the supplementary material of [13] leads to an easy proof. □

Corollary 2. $\mathcal{J}(T) \subseteq E = \mathcal{M}\left(L^C\right)$

Now we can state the desired result:

Proposition 1. $\mathcal{J}(T) \subseteq B$

Proof. By Cor. 2, every $e \in \mathcal{J}(T)$ is minimal in L^C. Thus, there is no $a \in L^C \setminus \{\emptyset\}$ with $a \subsetneq e$. Assume $e \notin B$. Then there exists $w \in \mathbb{R}^{\mathcal{I}} \setminus \{0\}$ with $S_{\mathcal{I}} w = 0, w_{\mathcal{I} \cap \mathrm{Irr}} \geq 0$, and $\mathrm{supp}(w) \subseteq e$. If $\mathrm{supp}(w) \subsetneq e$, then e is not minimal in L^C. If $\mathrm{supp}(w) = \mathrm{supp}(e)$, then e is the support of the internal circulation w, and it follows $e \notin T$. In both cases, we get a contradiction, hence $e \in B$. □

Altogether, we have shown $\mathcal{J}(T) \subseteq B \subseteq T \subseteq \langle T \rangle$. According to Thm. 1, we may perform FCA in T via FCA in B. Based on Thm. 2, we define methods $\texttt{testBlocked}_B(s)$ and $\texttt{testCoupled}_B(r, s)$, in which we solve the following mixed integer linear program (MILP) (for sufficiently large $M > 0$):

$$\begin{aligned}
& \min 0 \\
\text{s.t.} \quad & Sv = 0, && (6)\\
& v_{\mathrm{Irr}} \geq 0, && (7)\\
& -Ma_i \leq v_i \leq Ma_i, && (8)\\
& \sum_{i \in c} a_i \leq |c| - 1, \quad \forall c \in \mathcal{C} && (9)\\
& v_s = 2b - 1, \ v_s \in \{-1, 1\}, && (10)\\
& v_r = 0, && (11)\\
& a_i, b \in \{0, 1\} \quad \forall i \in \mathcal{R}.
\end{aligned}$$

Here, $\mathcal{C} := \mathcal{M}\left(\{\mathrm{supp}(w) \mid S_{\mathcal{I}} w = 0, w_{\mathcal{I} \cap \mathrm{Irr}} \geq 0\}\right)$ denotes the set of minimal supports of internal circulations, and $|c|$ is the cardinality of $c \in \mathcal{C}$. The idea of this MILP is the following. The variables a_i describe the support of the flux vector v. In (8), we require only that $a_i = 0$ implies $v_i = 0$, because the a_i appear only in the circuit constraints (9). Violated circuit constraints cannot become feasible by setting additional $a_i = 1$. The 0-1 variable b is used to force positive or negative flux through reaction s.

The functions $\texttt{testBlocked}_B(s)$ (resp. $\texttt{testCoupled}_B(r, s)$) can now be implemented by searching for a feasible solution v of the MILP (6) – (10) (resp. (6) – (11)). If this MILP is infeasible, we return the value $\emptyset$, otherwise the reaction set $a = \mathrm{supp}(v)$.

5 Implementation

The efficiency of the algorithm in [11] results from a search via nested intervals. The unblocked reactions in a lattice L are exactly those contained in the maximum element

$$\max := 1_L := \bigcup_{a \in L} a.$$

Similarly, the reactions *not* coupled to r in L are exactly those contained in the maximum element

$$\max_r := 1_{L_{\perp\{r\}}} := \bigcup_{a \in L_{\perp\{r\}}} a$$

of the lattice

$$L_{\perp\{r\}} = \{a \in L \mid r \notin a\}.$$

The algorithm in [11] determines max (resp. $\max_r$) via lower and upper bounds $\mathtt{lb} \subseteq \max \subseteq \mathtt{ub}$ (resp. $\mathtt{lb} \subseteq \max_r \subseteq \mathtt{ub}$). The lower bound $\mathtt{lb}$ is the union of known pathways ("witnesses"), which in the case of $\max_r$ must not contain r. The upper bound $\mathtt{ub}$ excludes reactions that are known to be blocked. Thus, only the remaining reactions $s \in \mathtt{ub} \setminus \mathtt{lb}$ have to be tested. Traditional FCA tests the feasibility of $\left\{v \in \mathbb{R}^{\mathcal{R}} \;\middle|\; Sv = 0, v_{\mathrm{Irr}} \geq 0, v_r = 0, |v_s| \geq 1\right\}$ via linear programming (LP). Supports of feasible solutions extend the lower bound (by at least adding s), while infeasibility leads to an update $\mathtt{ub} \leftarrow \mathtt{ub} \setminus \{s\}$.

In principle, it would be possible to realize thermodynamic FCA just by replacing these feasibility tests with the functions $\mathtt{testBlocked}_B$ and $\mathtt{testCoupled}_B$ introduced in Sect. 4. However, when calling these functions, we have to solve an NP-hard problem due to the thermodynamic constraints [17]. Since solving the MILP is computationally hard, we decided to use our knowledge of the lattice structure to minimize the number of function calls. We introduce a relaxation for preprocessing, which is $\langle T \rangle \subseteq L^C$. For the search of unblocked reactions in T (note $1_T = 1_{\langle B \rangle}$), we start with $\mathtt{ub} = 1_{L^C}$. To find directionally coupled reactions in T (derived from $1_{\langle B \rangle_{\perp\{r\}}}$), we continue with $\mathtt{ub} = 1_T \cap 1_{L^C_{\perp\{r\}}}$, where $1_{L^C_{\perp\{r\}}}$ is the set of reactions not directionally coupled to r by traditional FCA.

Our software has been implemented in Java and alternates between traditional FCA and tFCA, using the results of FCA computations whenever possible to infer tFCA properties. To perform traditional FCA and the test $\mathtt{testCoupled}_B$, we use `Cplex` 12.5 for solving the LPs and MILPs. The internal circuits of the network are computed with a variant of the WW-algorithm [27] using the `efmtool` by Terzer et al. [25]. All the networks analyzed in this study have a low number of internal circuits, which made this approach feasible and easy to implement.

A prototype implementation of tFCA is available at `http://hoverboard.io/L4FC`.

6 Discussion

6.1 Theoretical Differences

6.1.1 Standard couplings are preserved

If $r \xrightarrow{=0} s$ in L^C, then $r \xrightarrow{=0} s$ in T, since an infeasible system cannot become feasible by adding constraints. It follows that if two reactions are directionally coupled in L^C, they are also directionally coupled in T.

6.1.2 New partial couplings

In Fig. 2, we see an example where $r \xrightarrow{=0} s$ in L^C, but r is not directionally coupled to s in L^C (and thus r is not partially coupled to s in L^C). However, $s \xrightarrow{=0} r$ in T, hence r is partially coupled to s in T. Further examples are the pairwise in L^C uncoupled reactions a, b, c, d. With thermodynamic constraints however, $a \leftrightarrow c$ in T and $b \leftrightarrow d$ in T. In particular, we can deduce $v_a = v_c$ and $v_b = v_d$ for every thermodynamically feasible flux vector.

6.1.3 New directional couplings

In Fig. 3, we see an example where reactions a, b are uncoupled in L^C (because of flux vectors with supports $\{a, c, d\}$ and $\{b, c\}$). But, b is directionally coupled to a in T, since flux through the circuit $\{b, c\}$ alone is thermodynamically infeasible.

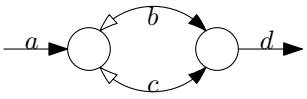

Figure 3: Example with uncoupled reactions that are thermodynamically coupled. The white arrowheads indicate that the reactions are reversible.

6.2 Practical Comparison

To evaluate our method in practice, we compared standard FCA and tFCA on several genome-scale metabolic network reconstructions from the BiGG-database [23]. The computational experiments were done with `Java Oracle JDK` 1.7.45 on a MacBook Air (2012), 1.8 GHz Intel Core i5, 4GB RAM, `Mac OS X` 10.8. For solving the linear programs and MILPs, we used `CPLEX` 12.5. The results are given in Tab. 1.

In all the considered networks, tFCA was able to detect additional blocked and coupled reactions. However, the impact of the additional thermodynamic constraints heavily depends on the network.

When we analyze all pairs of coupled reactions, there is a lot of redundant information. By transitivity, if we have couplings $a \xrightarrow{=0} b$, $b \xrightarrow{=0} c$, $a \xrightarrow{=0} c$, then the third can be inferred from the first and the second. In practice, this can lead to a quadratic blow-up of redundant couplings and hence the numbers of coupled pairs does not really reflect the gained information. In order to get a more adequate description, we computed a minimum set of couplings from which all other couplings can be deduced (in the case of FCA), and a minimum extension of this set from which all additional couplings in tFCA can be deduced. The details of this computation are described in [22].

In average, we found around $1\% - 6\%$ of additional couplings (in the minimal extension). In most cases, new directional couplings arised from previously uncoupled reactions. Only in one case, previously uncoupled reactions became partially coupled. We also observed that the reaction *cystathionine g-lyase* (`CYSTGL`) was part of new directional couplings in *M. tuberculosis*, while in *S. cerevisiae* the thermodynamic constraints blocked this reaction.

In the case of *H. pylori* iIT341, many new couplings are found because these reactions are part of the main biomass production pathway. If one of the reactions in the pathway is blocked, the whole pathway breaks down, which induces these couplings. However, if some reaction is part of an internal circuit, standard FCA allows such a reaction to be active via the circuit. This way standard FCA cannot detect that the reaction is part of the pathway. The reactions *homoserine O-trans-acetylase* (`HSERTA`), *O-Acetyl-L-homoserine succinate-lyase (adding cysteine)* (`METB1r`)

Model		blocked	couples	runtime
E. coli iAF1260	FCA	839	2101	36.35
	tFCA	848	2128	47.74
	extension	9	49	
S. cerevisiae iND750	FCA	635	885	7.73
	tFCA	640	935	12.43
	extension	5	58	
M. tuberculosis iNJ661	FCA	281	831	6.08
	tFCA	287	834	9.46
	extension	6	9	
S. aureus iSB619	FCA	278	544	3.02
	tFCA	279	546	3.95
	extension	1	3	
H. pylori iIT341	FCA	118	516	2.2
	tFCA	124	515	5.05
	extension	6	10	

Table 1: Comparison of thermodynamic flux coupling results for different genome-scale networks.
blocked Number of blocked reactions in the network.
couples Minimal number of pairs of directionally coupled reactions $r \xrightarrow{=0} s$ from which all couplings can be induced by transitive closure.
runtime Total runtime without pre-processing (calculation of internal circuits for tFCA). Times are given in seconds.
FCA Results for traditional FCA (steady-state assumption).
tFCA Results for thermodynamical FCA (steady-state, no internal circuits).
extension Minimal number of changes necessary to extend the FCA coupling graph to the tFCA coupling graph.

and *O-succinylhomoserine lyase* (`SHSL1r`, `SHSL2r`) in *H. pylori* iIT341 provide an example for this effect.

A summary of the results for *H. pylori* iIT341 is given in Fig. 4. The coupling types are depicted in a set-diagram style:

- The set of reactions blocked without thermodynamic constraints (blocked by FCA) is contained in the box of reactions blocked with thermodynamic constraints (blocked by tFCA). The number of reactions that are only blocked due to thermodynamic constraints is indicated in the set difference. In the case of *H. pylori*, there are 6 such reactions.

- Since there are more reactions blocked with thermodynamic constraints than without, some of the coupled reactions that we found by standard FCA contain reactions that are blocked with thermodynamic constraints. This is why the set of reactions coupled by FCA intersects the set of thermodynamically blocked reactions and the set of thermodynamically coupled reactions. For both intersections, we report how many pairs of reactions fall into the respective category.

- In the set difference of the thermodynamically minus the normally coupled reactions, we included the number of coupling pairs that fell into this category.

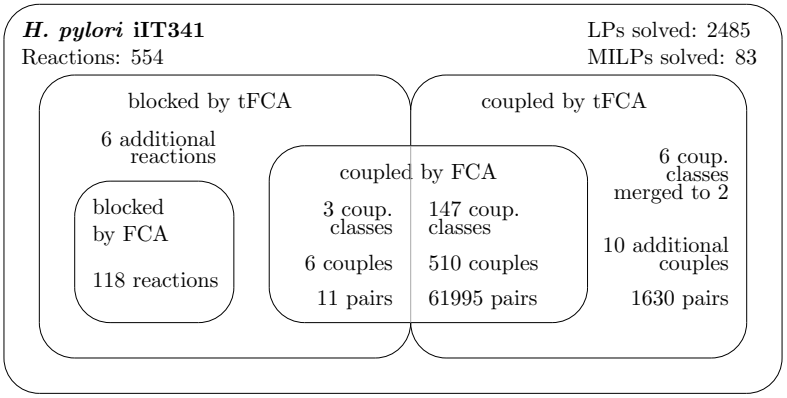

Figure 4: Results for *H. pylori* iIT341. We found that *homoserine O-trans-acetylase* (`HSERTA`), *O-Acetyl-L-homoserine succinate-lyase (adding cysteine)* (`METB1r`) and *O-succinylhomoserine lyase* (`SHSL1r`, `SHSL2r`) are not only necessary for biomass production (as computed by standard FCA) but cannot work without this function. Together with the biomass reaction and all its partially coupled reactions they form one of the new coupling classes. Furthermore, the under tFCA partially coupled (but not fully coupled) `SHSL2r` and `HSERTA` were originally (in FCA) uncoupled. "couples" refers to the minimal number of (additional) couplings from which all couplings can be inferred by transitive closure.

In Fig. 4, we see that the standard FCA couplings can be represented by 516 couplings, 6 of which are actually blocked by thermodynamic constraints. Thermodynamic constraints give additional information on 10 couplings. Thermodynamic constraints also merged 6 groups of partially coupled reactions to 2 groups.

Tab. 1 shows that the new tFCA algorithm runs only slightly slower than the FCA algorithm from [11]. This is achieved by the pre-processing step that first applies standard FCA, which already detects many couplings (see Sec. 6.1.1). In addition, witnesses are found that prove some reactions to be also thermodynamically unblocked, or uncoupled. Only for the few remaining cases where we cannot deduce any information from the previous step, we have to start the MILP solver.

7 Conclusion

We presented a refined version of FCA that finds more coupled reactions than standard FCA. Although thermodynamic constraints were used that are usually NP-hard, it was possible to also analyze genome-scale networks like *E. coli* iAF1260 in a few minutes. We observed that thermodynamic constraints do not only give additional blocked reactions but also additional coupled reactions. The concrete impact highly depends on the network that is analyzed.

We also observed in Thm. 1 that the presented approach is not only applicable to looplaw thermodynamic constraints but to any kind of restrictions of the flux space. Extensions of this method to also include concentration information or other constraints are straight-forward. Only the transformation in Sec. 4 cannot be applied directly, because it uses specific properties of loop-law thermodynamic constraints.

A prototype implementation of tFCA is available at `http://hoverboard.io/L4FC`.

Acknowledgments

We thank Alexandra Grigore for ideas on presenting the flux coupling results.

The PhD work of Arne Reimers and Yaron Goldstein was supported by the Berlin Mathematical School and the Gerhard C. Starck Stiftung.

References

[1] Robert A. Alberty. *Thermodynamics of Biochemcial Reactions*. Massachusetts Institute of Technology, Cambridge, MA, 2003.

[2] Daniel A. Beard, Eric Babson, Edward Curtis, and Hong Qian. Thermodynamic constraints for biochemical networks. *Journal of Theoretical Biology*, 228:327–333, 2004.

[3] Daniel A. Beard, Shou dan Liang, and Hong Qian. Energy balance for analysis of complex metabolic networks. *Biophysical Journal*, 83:79–86, 2002.

[4] A. Bordbar, J.M. Monk, Z.A. King, and B. Palsson. Constraint-based models predict metabolic and associated cellular functions. *Nat Rev Genet*, 15(2):107–20, 2014.

[5] Jacob G. Bundy, Balzs Papp, Rebecca Harmston, Roy A. Browne, Edward M. Clayson, Nicola Burton, Richard J. Reece, Stephen G. Oliver, and Kevin M. Brindle. Evaluation of predicted network modules in yeast metabolism using nmr-based metabolite profiling. *Genome Research*, 17(4):510519, 2007.

[6] Anthony P. Burgard, Nikolaev Evgeni V., Christophe H. Schilling, and Costas D. Maranas. Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Research*, 14(2):301312, 2004.

[7] Pablo Carbonell, Davide Fichera, Shashi Pandit, and Jean-Loup Faulon. Enumerating metabolic pathways for the production of heterologous target chemicals in chassis organisms. *BMC Systems Biology*, 6(1):10, 2012.

[8] Brian A Davey. *Introduction to lattices and order*. Cambridge University Press, 2002.

[9] L. David, S. A. Marashi, A. Larhlimi, B. Mieth, and A. Bockmayr. FFCA: a feasibility-based method for flux coupling analysis of metabolic networks. *BMC Bioinformatics*, 12:236, 2011.

[10] Ronan M.T. Fleming, Christopher M. Maes, Michael A. Saunders, Yinyu Ye, and Bernhard Ø. Palsson. A variational principle for computing nonequilibrium fluxes and potentials in genome-scale biochemical networks. *Journal of Theoretical Biology*, 292:71–77, 2012.

[11] YaronA.B. Goldstein and Alexander Bockmayr. A lattice-theoretic framework for metabolic pathway analysis. In Ashutosh Gupta and ThomasA. Henzinger, editors, *Computational Methods in Systems Biology*, volume 8130 of *Lecture Notes in Computer Science*, pages 178–191. Springer Berlin Heidelberg, 2013.

[12] Utz-Uwe Haus, Steffen Klamt, and Tamon Stephen. Computing knock-out strategies in metabolic networks. *Journal of Computational Biology*, 15(3):259–268, 2008.

[13] Stefan J. Jol, Anne Kümmel, Marco Terzer, Jorg Stelling, and Matthias Heinemann. System-level insights into yeast metabolism by thermodynamic analysis of elementary flux modes. *PLoS Computational Biology*, 8(3), 2012.

[14] A. Larhlimi, L. David, J. Selbig, and A. Bockmayr. F2C2: a fast tool for the computation of flux coupling in gen ome-scale metabolic networks. *BMC Bioinformatics*, 13:57, 2012.

[15] Abdelhalim Larhlimi and Alexander Bockmayr. A new approach to flux coupling analysis of metabolic networks. *Computational Life Sciences II*, 4216:205–215, 2006.

[16] N. E. Lewis, H. Nagarajan, and B. Palsson. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat Rev Microbiol*, 10(4):291–305, 2012.

[17] A. C. Müller and A. Bockmayr. Fast thermodynamically constrained flux variability analysis. *Bioinformatics*, 29(7):903–909, 2013.

[18] Arne Müller. Thermodynamic constraints in metabolic networks. Master's thesis, Freie Universität Berlin, Fachbereich Mathematik und Informatik, February 2012.

[19] Elad Noor, Nathan E. Lewis, and Ron Milo. A proof for loop-law constraints in stoichiometric metabolic networks. *BMC Systems Biology*, 6:140, 2012.

[20] Richard A. Notebaart, Bas Teusink, Roland J. Siezen, and Balázs Papp. Co-regulation of metabolic genes is better explained by flux coupling than by network distance. *PLoS Computational Biology*, 4(1):e26, 2008.

[21] Hong Qian and Daniel A. Beard. Thermodynamics of stoichiometric biochemical networks in living systems far from equilibrium. *Biophysical Chemistry*, 114:213–220, 2005.

[22] A. C. Reimers, A. M. Reimers, and Y. A. B. Goldstein. Minimal equivalent subgraphs containing a given set of arcs. Technical Report #1053, Matheon, March 2014.

[23] Jan Schellenberger, Junyoung O. Park, Tom M. Conrad, and Bernhard Ø Palsson. BiGG: a biochemical genetic and genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics*, 11:213, 2010.

[24] Takehide Soh and Katsumi Inoue. Identifying necessary reactions in metabolic pathways by minimal model generation. In *ECAI*, pages 277–282, 2010.

[25] Marco Terzer. *Large Scale Methods to Enumerate Extreme Rays and Elementary Modes*. PhD thesis, Swiss Federal Institute of Technology, Zurich, 2009.

[26] Amit Varma and Bernhard Ø. Palsson. Metabolic flux balancing: Basic concepts, scientific and practical use. *Nature Biotechnology*, 12:994–998, 1994.

[27] Jeremiah Wright and Andreas Wagner. Exhaustive identification of steady state cycles in large stoichiometric networks. *BMC Systems Biology*, 2(61), 2008.