Journal of Mathematical Biology (2019) 78:1459–1484
https://doi.org/10.1007/s00285-018-1316-9

Mathematical Biology

Quantitative flux coupling analysis

Mojtaba Tefagh[1] · **Stephen P. Boyd**[1]

Received: 27 February 2018 / Revised: 2 December 2018 / Published online: 10 December 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract
Flux coupling analysis (FCA) aims to describe the functional dependencies among reactions in a metabolic network. Currently studied coupling relations are qualitative in the sense that they identify pairs of reactions for which the activity of one reaction necessitates the activity of the other one, but without giving any numerical bounds relating the possible activity rates. The potential applications of FCA are heavily investigated, however apart from some trivial cases there is no clue of what bottleneck in the metabolic network causes each dependency. In this article, we introduce a quantitative approach to the same flux coupling problem named *quantitative flux coupling analysis* (QFCA). It generalizes the current concepts as we show that all the qualitative information provided by FCA is readily available in the quantitative flux coupling equations of QFCA, without the need for any additional analysis. Moreover, we design a simple algorithm to efficiently identify these flux coupling equations which scales up to the genome-scale metabolic networks with thousands of reactions and metabolites in an effective way. Furthermore, this framework enables us to quantify the "strength" of the flux coupling relations. We also provide different biologically meaningful interpretations, including one which gives an intuitive certificate of precisely which metabolites in the network enforce each flux coupling relation. Eventually, we conclude by suggesting the probable application of QFCA to the metabolic gap-filling problem, which we only begin to address here and is left for future research to further investigate.

Keywords Systems biology · Metabolic network analysis · Flux coupling analysis · FCA · QFCA · Flux coupling equation

Mathematics Subject Classification 92C42 · 90C05 · 49N15

✉ Mojtaba Tefagh
mtefagh@stanford.edu

[1] Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA, USA

Author summary

Metabolic networks are genome-scale complex systems of metabolites and reactions among them. In order to model their dynamics, constraint-based approach specifies a set of local rules like the preservation of mass at every single metabolite. Consecutively, these basic constraints result in emergent constraints which may link two reactions in totally separated regions of the metabolic network. From an applied standpoint, these global relations are just as important as the local ones, but unlike them, they are not efficiently computable and even if proved by computation these computational certificates are not as insightful.

In this article, we introduce the concept of fictitious metabolites which are nothing but proxies for the ensembles of related metabolites and in this way, they resemble metabolites structurally. They mask all the micro-interactions within and only display their collective behavior which can relate two far away reactions linked together by a chain of metabolites. Coupling relations among distant reactions are induced by fictitious metabolites in the same way as they are induced among adjacent reactions by regular metabolites. Therefore, our biological intuition applies to their general certificates seamlessly. Moreover, all the well-known local techniques can be transferred to the next level to develop scalable tools which are not only faster but also more interpretable at the system level.

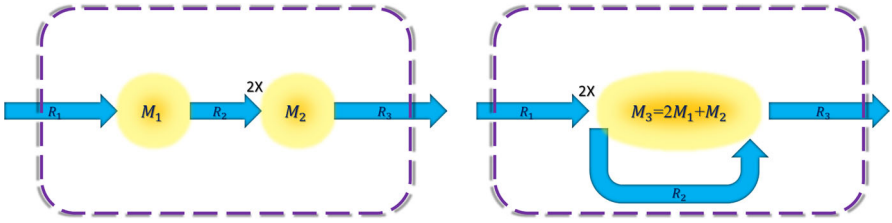
1 Introduction

Constraint-based reconstructions of (genome-scale) metabolic networks are models formally characterizing the metabolic activities of an organism. The relatively small set of data given by such models is then analyzed in order to predict different biological properties which are usually very hard to measure experimentally (Schilling et al. 1999a; Covert et al. 2001).

One instance of such *in silico* computational analysis is to investigate the effects of blocking a reaction on the rest of the reactions in the metabolic network. This has already many applications in systems biology and bioinformatics, for instance explaining the co-regulation of metabolic genes (Notebaart et al. 2008), or the identification of potential drug targets by understanding the immediate consequences of knockouts at the genome level (Haus et al. 2008).

However, because of the huge number of reactions the analysis of the dependencies among them is too intractable to be approached fully experimentally. An alternative approach though is to exploit the dependencies which are implied by the general constraints imposed on the biochemically legitimate flux distributions in the metabolic network of interest (Burgard et al. 2004).

The primary constraint assumed by most of the constraint-based models is the steady-state condition whereby we assume that the metabolites are balanced at an equilibrium (Varma and Palsson 1994). Solely, this constraint by itself may imply that the rate of some reactions cannot vary independently. As a simple example consider the metabolic network depicted in Fig. 1a. In this metabolic network the boundary reactions R_1 and R_2 import and export one unit of the internal metabolites M_1 and



(a) the original metabolic network **(b)** the transformed metabolic network

Fig. 1 Reactions coupled by metabolites

M_2, and the middle reaction consumes one unit of M_1 in order to produce two units of M_2. The mass balance of M_1 implies that R_1 and R_2 must be active at exactly the same rate. Likewise, the activity rate of R_3 must be twice the activity rate of R_2 in order to preserve the equilibrium at M_2.

On the other hand, consider a "fictitious metabolite" M_3 which is defined to be equal to $2M_1 + M_2$, namely each unit of M_1 counts as two units of M_3 and also each unit of M_2 counts as one unit of M_3. We express the stoichiometry of each reaction in terms of this new metabolite accordingly. In Fig. 1b, R_2 does not change the concentration of M_3 $(2(-1) + (2) = 0)$, R_1 imports two units $(2(1) + 0 = 2)$, and R_3 exports one unit of it $(2(0) + (-1) = -1)$. As a consequence, the mass balance of this fictitious metabolite which is conceived as a linear combination of some other metabolites, implies that the activity rate of R_3 equals two times the activity rate of R_1.

Although in this toy example there seems no need for certificates to verify the flux coupling relations, the key idea of introducing new fictitious metabolites as intuitive certificates extends naturally to the most general setting. The only remaining challenge is how to effectively identify the appropriate fictitious metabolites, for which we propose the QFCA algorithm as one of the possible solutions.

Outline. In Sect. 2, we review the constraint-based analysis of metabolic networks and formulate the flux coupling problem. In Sect. 3, we formally introduce the flux coupling certificates and derive an efficient algorithm to compute them. In Sect. 4, we discuss two of their biological interpretations, including fictitious metabolites as explained here, and the subjects of most interest which are left for future research. Finally in Sect. 5, we conclude by pointing out the major implications of this work.

2 Problem setup

2.1 Steady-state flux cone

Consider a given metabolic network consisting of m internal metabolites, $\mathcal{M} = \{M_i\}_{i=1}^m$, and n biochemical reactions, $\mathcal{R} = \{R_i\}_{i=1}^n$. In what continues, we will establish the mathematical framework for modelling the data we have and the constraints we assume.

The stoichiometric properties of the metabolic network of interest are encoded by an $m \times n$ stoichiometric matrix S. Each row of S corresponds to a unique internal metabolite in $\mathcal{M}$, and each column of it represents the stoichiometry of a reaction in $\mathcal{R}$. In the columns of S, consumed and produced metabolites are specified by negative and positive entries, respectively. In addition, the magnitude of each entry determines the relative rate of consumption or production. For instance, the stoichiometric matrix of the metabolic network depicted in Fig. 1a is

$$S = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 2 & -1 \end{bmatrix}.$$

In this setting, we think of any vector $v \in \mathbb{R}^n$ as a flux distribution whose entries, which we call flux coefficients, are the rates of the n reactions in the metabolic network. In addition, activity in the two possible directions for each reaction is shown by the sign of the corresponding flux coefficient. In other words, positive and negative flux coefficients correspond to forward and reverse directions, respectively.

In the *constraint-based reconstruction and analysis* (COBRA) of metabolic networks, the primary constraints imposed on a flux distribution to be feasible are mostly as follows (Schuster and Hilgetag 1994; Bonarius et al. 1997):

1. *The steady-state constraint* assumes that the metabolic network is in mass balance condition, which means that the concentration of each internal metabolite is almost constant throughout the time-scale of interest (Gunawardena 2014). We say a flux distribution $v \in \mathbb{R}^n$ is in steady-state condition, if

$$Sv = 0. \tag{1}$$

2. *The irreversibly constraint* states that each irreversible reaction is thermodynamically forced to proceed in only one fixed direction which is the forward direction, by convention. Let $\mathcal{I} \subseteq \mathcal{R}$ denote the subset of irreversible reactions. We say a flux distribution $v \in \mathbb{R}^n$ satisfies the irreversibility constraint, if for all $R_i \in \mathcal{I}$

$$v_i \geq 0. \tag{2}$$

A flux distribution $v \in \mathbb{R}^n$ is called feasible if it satisfies constraints (1) and (2). Since (1) is a homogeneous system of linear equations and (2) is a set of nonnegativity inequalities, the set of all the feasible flux distributions is a polyhedral convex cone, which we call the *steady-state flux cone* (Schilling et al. 1999b).

To the end of this paper, we will assume for notational convenience that the irreversible reactions are indexed as $\mathcal{I} = \{R_1, R_2, \ldots, R_k\}$ for some k, unless stated otherwise. Subsequently, we can denote the steady-state flux cone by $\mathcal{C} = \{v = (v_I, v_R) \mid Sv = 0,\ v_I \geq 0\}$, where $v_I \in \mathbb{R}^k$ and $v_R \in \mathbb{R}^{n-k}$.

2.2 Blocked reactions and flux coupling relations

For any $R_i \in \mathcal{R}$, we say R_i is a blocked reaction if for all the feasible flux distributions $v \in \mathcal{C}$ there is zero flux through R_i, namely $v_i = 0$. Otherwise, we call R_i an unblocked

reaction. We call a metabolic network consistent if it contains no blocked reactions (Schuster and Hilgetag 1994).

After this preliminary definition, we are ready to define the three kinds of possible flux coupling relations between two unblocked reactions. By inspection, each one is a stronger condition than the previous one.

Definition 1 (Burgard et al. 2004) Let $R_i, R_j \in \mathcal{R}$ be an arbitrary pair of unblocked reactions.

Directional coupling: R_i is *directionally coupled* to R_j, denoted by $R_i \longrightarrow R_j$, if for all feasible flux distributions $v_i \neq 0$ implies $v_j \neq 0$.

Partial coupling: R_i is *partially coupled* to R_j, denoted by $R_i \longleftrightarrow R_j$, if for all feasible flux distributions $v_i \neq 0$ implies $v_j \neq 0$ and vice versa.

Full coupling: R_i is *fully coupled* to R_j, denoted by $R_i \Longleftrightarrow R_j$, if there exists a constant $c \neq 0$ such that for all feasible flux distributions

$$v_i = c v_j. \tag{3}$$

Note that we have excluded the blocked reactions from the coupling definition because otherwise, any blocked reaction would be directionally coupled to every other reaction. Additionally, no unblocked reaction could be directionally coupled to any blocked reaction. Consequently, there is no nontrivial information that flux coupling relations can tell about the blocked reactions, and therefore, we omitted them from the definition of flux coupling.

Despite the fact that we have excepted blocked reaction from the flux coupling definition, there is an intrinsic relationship between the two concepts. One equivalent definition of directional coupling is that a reaction R_i is directionally coupled to a reaction R_j if R_i is unblocked originally but will become blocked upon the removal of R_j from the metabolic network. Therefore, identifying all the reactions which are directionally coupled to R_j is essentially the same as finding the newly blocked reactions after removing R_j from the metabolic network.

In fact, one reason for the existence of blocked reactions in the first place is that metabolic network reconstructions are incomplete models of their real-world counterparts, and without the missing reactions many pathways can never carry nonzero fluxes. For instance, we can be missing some reactions while detecting some other ones which are directionally coupled to them. Then, the latter would be blocked.

2.3 Flux coupling equations

In this paper we call (3) a *full coupling equation* (FCE). We begin by observing that any unblocked reaction R_i is fully coupled to itself by $v_i = 1 v_i$ as the corresponding FCE.

Another observation is that, if there is a metabolite which is produced by exactly one reaction and consumed by exactly one other reaction, these reactions are fully coupled and the steady-state condition for this metabolite provides the corresponding

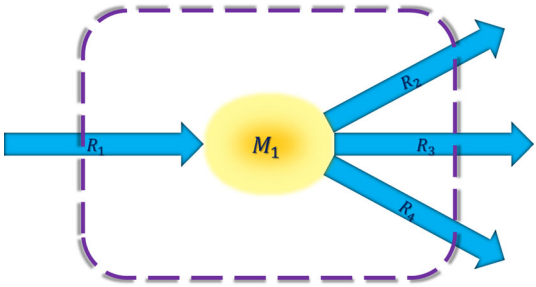


Fig. 2 For $t = 2, 3, 4$, $R_t \longrightarrow R_1$ can be inferred from the DCE corresponding to M_1

FCE. In Fig. 1a, equilibrium at M_1 proves $v_1 = 1v_2$, and equilibrium at M_2 proves $v_2 = \frac{1}{2}v_3$. In Fig. 1b, equilibrium at M_3 proves $v_1 = \frac{1}{2}v_3$.

Unlike full coupling relations, directional and partial coupling relations are qualitative in the sense that there is no equation in analogy to FCE which quantizes them. This motivates the search for at least some special cases where similar to the previous observation, the steady-state condition of specific metabolites provide us with flux coupling equations.

As illustrated in Fig. 2, suppose that there exists a metabolite which is either produced only by R_j and consumed only by $R_{i_1}, R_{i_2}, \ldots, R_{i_l}$, or consumed only by R_j and produced only by $R_{i_1}, R_{i_2}, \ldots, R_{i_l}$. Either way, from the mass balance of this metabolite we have that

$$v_j = c_{i_1} v_{i_1} + c_{i_2} v_{i_2} + \cdots + c_{i_l} v_{i_l}, \tag{DCE}$$

holds for some $c_{i_1}, c_{i_2}, \ldots, c_{i_l} > 0$. From this equation, it is immediate that if $R_{i_1}, R_{i_2}, \ldots, R_{i_l} \in \mathcal{I}$, then $R_{i_t} \longrightarrow R_j$ for all $t = 1, 2, \ldots, l$, because $v_{i_t} \neq 0$ implies $v_j \neq 0$. This is a generalization of FCE relating the rates of a fully coupled pair of reactions, and hence we call it *directional coupling equation* (DCE).

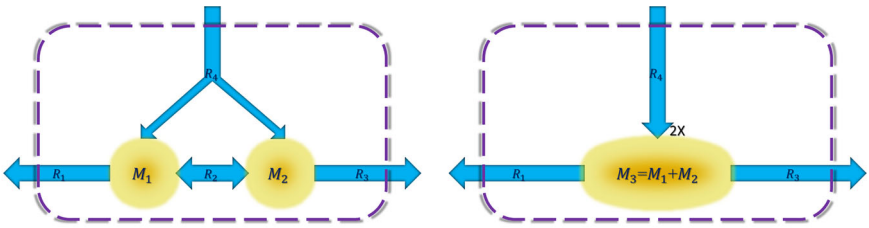
Moreover, the coefficients c_{i_t} indicate the contribution of each R_{i_t} in the rate of R_j and the greater this multiplier, the stronger the impact of R_{i_t} on R_j. In this sense, DCE coefficients can compare and quantify the strength of the directional coupling relations, just like the role of constant c in (3), for the case of full coupling relations.

Suppose that besides the previous DCE, we have another equation of the form

$$v_j = c'_{i_1} v_{i_1} + c'_{i_2} v_{i_2} + \cdots + c'_{i_l} v_{i_l} + c'_{i_{l+1}} v_{i_{l+1}}, \tag{EDCE}$$

where $c'_{i_{l+1}} \neq 0$. If $v_{i_{l+1}} \neq 0$, then either $v_j \neq 0$ or one of $v_{i_t} \neq 0$ holds, which in turn implies $v_j \neq 0$. As a result, if $R_{i_1}, R_{i_2}, \ldots, R_{i_l} \in \mathcal{I}$, then $R_{i_t} \longrightarrow R_j$ for all $t = 1, 2, \ldots, l+1$. Accordingly, we call it an *extended directional coupling equation* (EDCE).

Unlike FCE and DCE, just a single EDCE cannot prove any flux coupling relation on its own, because it is too general and $R_{i_{l+1}}$ can be either reversible or irreversible. However, this extra flexibility allows us to prove the directional coupling relations $R_{i_{l+1}} \longrightarrow R_j$, where $R_{i_{l+1}} \notin \mathcal{I}$. In such a case,



(a) The two-sided arrow indicates R_2's reversibility. **(b)** Drawing R_2 is no longer necessary.

Fig. 3 $R_2 \longrightarrow R_4$ can be inferred from the EDCEs corresponding to M_1 and M_2

$$\left(1+\frac{1}{c}\right)v_j = \left(c_{i_1}+\frac{c'_{i_1}}{c}\right)v_{i_1} + \left(c_{i_2}+\frac{c'_{i_2}}{c}\right)v_{i_2} + \cdots + \left(c_{i_l}+\frac{c'_{i_{l+1}}}{c}\right)v_{i_l} + \frac{c'_{i_{l+1}}}{c}v_{i_{l+1}},$$

as $c \longrightarrow \infty$ and $c \longrightarrow -\infty$ generates two series of DCEs proving $v_j \neq 0$ whenever $v_{i_{l+1}}$ is either positive (forward direction) or negative (reverse direction).

2.4 Fictitious metabolites

The steady-state constraint (1) is a homogeneous system of linear equations whose rows are conceived as local mass balance conditions for single metabolites. However, any arbitrary linear combination of these linear equations is again a valid constraint on the rate of reactions which are not necessarily close to each other in the metabolic network. Note that if such linear combinations are added to the rows of the stoichiometric matrix S, then the rank of S is constant and its null space does not change at all; hence the steady-state flux cone $\mathcal{C}$ remains the same too. Therefore, they can be interpreted as fictitious metabolites, even though we never actually add them to the metabolic network in practice.

Definition 2 Suppose that we are given a metabolic network with the stoichiometric matrix S. We call $\lambda \in \mathbb{R}^n$ a fictitious metabolite if there exists $\nu \in \mathbb{R}^m$ such that $\lambda = S^T\nu$.

At this point, DCE and EDCE are inspired by the mass balance of a specific metabolite. However, we showed that they can be used to prove directional coupling and consequently partial coupling relations. Recall that a fictitious metabolite provided an FCE in Fig. 1b. Similarly, M_3 is a fictitious metabolite providing the DCE $v_4 = \frac{1}{2}v_1 + \frac{1}{2}v_3$ in Fig. 3b. Furthermore in Fig. 3a, M_1 and M_2 provide the EDCEs $v_4 = v_1 + v_2$ and $v_4 = v_3 - v_2$, respectively.

Earlier in Sect. 1, we briefly mentioned the idea of proving full coupling relations by building fictitious metabolites whose corresponding FCEs verify the desired relations. The next step is to follow the same intuition and push this idea even further to include DCE and EDCE as well. The following theorem summarizes the desired result.

Theorem 2.1 (Fictitious metabolites) *Suppose that in a given metabolic network specified by S and $\mathcal{I}$, there are no irreversible blocked reactions. Then for any $\lambda \in \mathbb{R}^n$, λ is a fictitious metabolite if and only if*

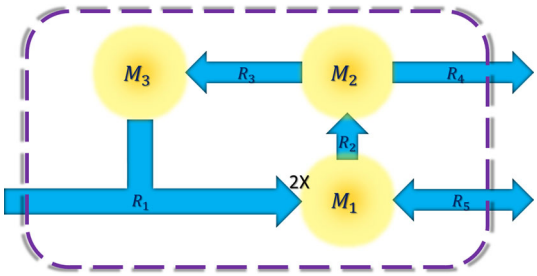


Fig. 4 M_1 and $M_1 + M_3$ provide EDCEs, M_2 and $M_2 + M_3$ provide DCEs, and M_3 provides an FCE

$$\lambda^T v = 0, \quad \forall v \in \mathcal{C}.$$

Remark Note that $\lambda^T v = 0$ is an arbitrary linear equation over the flux coefficients. As a special case, when λ represents the coefficients of any FCE, DCE, or EDCE, we will call the equation $\lambda = S^T \nu$ the fictitious metabolite form of the flux coupling equation λ. Later in Sect. 4, we will come back to this point and elaborate on how the fact that λ is written as a linear combination of the rows of S motivates the name of fictitious metabolite.

The proof of this theorem follows from the following lemma.

Lemma 2.2 *Suppose that in a given metabolic network specified by S and $\mathcal{I}$, there are no irreversible blocked reactions. Then for any $\lambda \in \mathbb{R}^n$,*

$$\lambda^T v = 0, \quad \forall v \in \mathcal{C} \Leftrightarrow \lambda^T u = 0, \quad \forall u \in \ker(S), \tag{4}$$

where $\ker(S)$ *is the set of all vectors $u \in \mathbb{R}^n$ such that $Su = 0$.*

Remark If there are no irreversible blocked reactions, then by (4) the affine hull of the steady-state flux cone is the null space of the stoichiometric matrix. As an immediate corollary, the affine dimension of $\mathcal{C}$ equals the nullity of S.

For the sake of example, consider the metabolic network depicted in Fig. 4. Similar to Figs. 1 and 2, M_3 provides the FCE $v_1 = v_3$ coupling R_1 to R_3, and M_2 provides the DCE $v_2 = v_3 + v_4$ coupling R_3 and R_4 to R_2. From these two equations, we get another DCE $v_2 = v_1 + v_4$ which is also given by merging M_2 and M_3 similar to Fig. 1b. From M_1 we have $2v_1 = v_2 + v_5$ must hold, which we can rewrite as the EDCE $v_2 = 2v_1 - v_5$ coupling R_5 to R_2. Also by substituting v_1 with v_3 from the former equations, we get another EDCE $v_2 = v_1 + v_3 - v_5$ which is also given by merging M_1 and M_3 into a fictitious metabolite.

2.5 Problem statement and prior works

In the sequel, the problem is given the stoichiometric matrix S and the subset of irreversible reactions $\mathcal{I}$ of a metabolic network, how to

1. identify all the blocked reactions in order to exclude them from FCA,

2. find all the coupled pairs of reactions and specify the type of coupling relation for each pair,
3. and ultimately for each coupling relation give the appropriate FCE, DCE, or EDCE in order to prove and quantize it.

We start by deriving a naive method for the first two items which are referred to collectively as FCA.

A simple observation is that we can determine whether $R_i \in \mathcal{R}$ is blocked or not by solving at most two *linear programs* (LP). For $R_i \in \mathcal{I}$, the following LP

$$\begin{aligned} &\text{maximize } v_i \\ &\text{subject to } v \in \mathcal{C} \\ &\qquad\qquad v_i \leq 1, \end{aligned} \tag{5}$$

has optimal value 0 if and only if R_i is blocked. The only other possible optimal value for (5) is 1 in which case R_i is unblocked. For $R_i \notin \mathcal{I}$, both (5) and

$$\begin{aligned} &\text{minimize } v_i \\ &\text{subject to } v \in \mathcal{C} \\ &\qquad\qquad v_i \geq -1, \end{aligned} \tag{6}$$

have optimal values 0 if and only if R_i is blocked. The only other possible optimal value for (6) is -1 in which case R_i is unblocked.

From this observation, a simple solution to our problem is to first identify the blocked reactions by (5) and (6). Then in analogy to (5), we determine whether $R_i \longrightarrow R_j$ holds or not by solving the LP to maximize the flux through R_i after blocking R_j, *i.e.*,

$$\begin{aligned} &\text{maximize } v_i \\ &\text{subject to } v \in \mathcal{C} \\ &\qquad\qquad v_j = 0 \\ &\qquad\qquad v_i \leq 1. \end{aligned} \tag{7}$$

For $R_i \in \mathcal{I}$, $R_i \longrightarrow R_j$ if and only if the optimal value of (7) is zero. However in analogy to (6), for $R_i \notin \mathcal{I}$ we should also solve the LP for maximizing flux through the other possible direction of R_i, *i.e.*,

$$\begin{aligned} &\text{minimize } v_i \\ &\text{subject to } v \in \mathcal{C} \\ &\qquad\qquad v_j = 0 \\ &\qquad\qquad v_i \geq -1. \end{aligned}$$

Then $R_i \longrightarrow R_j$ if and only if the optimal values of both objectives are zero.

Let n_i, and n_r denote the number of irreversible and reversible reactions, respectively. This naive method first identifies all the blocked reactions by $n_i + 2n_r$ LPs which can run in parallel. Then it solves $n_i((n_i - 1) + 2n_r) + n_r(n_i + 2(n_r - 1))$ additional LPs in order to find all the directional couplings which in turn determines

all the partial couplings as well. Again one can solve these LPs in parallel. However, to the best of our knowledge, this is not exploited yet in any existing algorithm.

Eventually, it only remains to find all the fully coupled pairs of reactions. We already know that full coupling is a stronger condition than partial coupling, hence it is enough to check if it holds only for the partially coupled reactions. For each pair of $R_i \longleftrightarrow R_j$, they are fully coupled to each other if and only if the optimal values of

$$\begin{aligned} &\text{maximize } v_i \\ &\text{subject to } v \in \mathcal{C} \\ &\qquad\qquad v_j = 1, \end{aligned}$$

and

$$\begin{aligned} &\text{minimize } v_i \\ &\text{subject to } v \in \mathcal{C} \\ &\qquad\qquad v_j = 1, \end{aligned}$$

are equal, and if so then it also equals the constant c in (3).

In summary, if there are n_p partially coupled pairs of reactions, this naive method requires to solve

$$n(n_i + 2n_r) + 2n_p \tag{8}$$

LPs in total in order to conduct FCA. An algorithm named *feasibility-based flux coupling analysis* (FFCA) (David et al. 2011) takes a similar approach as the basis for the generic case. However, it substantially reduces the number of required LPs to solve from (8) for the naive method.

This additional efficiency is achieved by exploiting theorems such as the following proposition which study the possible reversibility types for a coupled pair of reactions and so are called *reversibility-type prunings*. We refer the interested reader to (Larhlimi and Bockmayr 2006) to read more about them.

Proposition 1 (Larhlimi and Bockmayr 2006) *Suppose that R_i and R_j are unblocked reactions. Moreover, R_j is reversible and unblocked in both directions, i.e., there exist feasible flux distributions $u, v \in \mathcal{C}$ such that $u_j > 0$ and $v_j < 0$. Then $R_i \longrightarrow R_j$ if and only if $R_i \Longleftrightarrow R_j$.*

Note that we require not only $R_j \notin \mathcal{I}$, but also both directions of R_j to be unblocked. If this is not the case, *i.e.*, R_j is not explicitly constrained to be but is effectively irreversible, we can detect it by solving (5) and (6). Without loss of generality, suppose that R_j can only happen at positive rates and is blocked in the reverse direction. Otherwise, we multiply the corresponding column of S by -1. By adding such reactions to $\mathcal{I}$, we will assume to the end of this paper that if $R_j \notin \mathcal{I}$ is unblocked, then there exist feasible flux distributions $u, v \in \mathcal{C}$ such that $u_j > 0$ and $v_j < 0$. This issue is treated more extensively in the "Appendix".

A subsequent work, *fast flux coupling calculator* (F2C2) (Larhlimi et al. 2012) also preprocesses the trivial cases like Fig. 2, along with many others, where it is obvious from M_1 that upon removal of R_1, all the reactions R_2, R_3, and R_4 would become blocked and hence $R_t \longrightarrow R_1$ for $t = 2, 3, 4$.

The purpose of this study is to generalize the latter approach to the most general setting in order to prove that like the role of FCE in the case of fully coupled reactions, DCE and EDCE have the same role for verifying the other flux coupling relations. This is in sharp contrast to all the previous works which provide no straightforward certificate to verify each flux coupling relation without solving any LPs.

3 Methods

3.1 Quantitative flux coupling analysis

In this subsection, our primary goal is to develop an algorithm finding all the blocked reactions and the pairs of reactions which are directionally, partially, or fully coupled by solving only a linear number of LPs in the worst case. To the best of our knowledge, this is the first algorithm which is guaranteed to achieve this linear bound in the number of reactions.

We begin by the identification of blocked reactions. Since for each unblocked $R_i \in \mathcal{R}$, there exists a feasible flux distribution v^i with nonzero flux through R_i, $v^i_i \neq 0$, we can form a weighted sum of these v^i's,

$$v^\star = \sum_i c_i v^i,$$

in order to get a feasible flux distribution $v^\star$ with nonzero flux coefficients for all the unblocked reactions. Furthermore, we can assume that all the nonzero entries of $v^\star$ are greater than one in absolute value by multiplying c_i's with a large enough scalar.

Indeed, we should form this linear combination with appropriate coefficients c_i's in a way that for any $R_j \notin \mathcal{I}$, different flux coefficients v^i_j do not cancel away to $v^\star_j = 0$. However, this is always possible as we can scale each feasible flux distribution v^i such that the entries of different $c_i v^i$'s have different orders of magnitude so they never add up to zero, $\sum_i c_i v^i_j \neq 0$.

Let $\mathcal{B}_I$ and $\mathcal{B}_R$ denote the set of all the irreversible and reversible blocked reactions, respectively. The critical observation here is that if we restrict attention to the identification of $\mathcal{B}_I$, it is enough to solve the following optimization problem

$$\begin{aligned} &\text{maximize } \mathbf{1}^T \min(v_I, \mathbf{1}) \\ &\text{subject to } v \in \mathcal{C}, \end{aligned} \tag{9}$$

where $v \in \mathbb{R}^n$. The previous two paragraphs constructs a solution for this optimization problem, hence the optimal objective value is achieved and equals $k - |\mathcal{B}_I|$.

We can reformulate this optimization problem as the following equivalent LP (see OnePrune in Dreyfuss et al. 2013 or LP-7 in Vlassis et al. 2014)

$$\begin{aligned} &\text{maximize } \mathbf{1}^T u \\ &\text{subject to } Sv = 0 \\ &\qquad\qquad v_I \geq u \\ &\qquad\qquad \mathbf{1} \geq u \geq 0, \end{aligned} \tag{10}$$

where $u \in \mathbb{R}^k$. This LP is always both feasible (*e.g.*, $u = 0$, $v = 0$ is feasible) and bounded ($\mathbf{1}^T u \leq k$). Any optimal $v^\star$ for (10) is also optimal for (9) and vice versa. Since $u^\star = \min(v_I^\star, \mathbf{1})$, the optimal $u^\star$ has zero and one entries corresponding to the blocked and unblocked irreversible reactions, respectively.

Now that we can identify $\mathcal{B}_I$ by solving one LP, we assume that all the reactions in $\mathcal{I}$ are unblocked or otherwise remove them from the metabolic network. As a result, $v_i^\star > 0$ for all $R_i \in \mathcal{I}$.

Suppose that $R_i \notin \mathcal{I}$ is blocked or equivalently there does not exist $v \in \mathcal{C}$ such that $v_i \neq 0$. We claim that $Sx = 0$ has no solutions for which $x_i \neq 0$. Otherwise, we choose a large enough $v^\star$ such that $v = x + v^\star \in \mathcal{C}$. However, this cannot be true because v_i and $v_i^\star$ are both zero as R_i is blocked and hence x_i must be zero too.

Consequently, $R_i \notin \mathcal{I}$ is unblocked if and only if there exists $x \in \mathbb{R}^n$ such that $Sx = 0$ and $x_i \neq 0$, or equivalently $Sx = 0$ and $x_i = 1$. However, this latter condition can be expressed as the system of linear equations

$$\begin{cases} Sx = 0 \\ e_i^T x = 1, \end{cases} \tag{11}$$

where e_i is the ith element of the standard basis of $\mathbb{R}^n$.

In a nutshell, we can determine $\mathcal{B}_R$ by only solving linear equations. Combining this with our previous result, we can identify all the blocked reactions by solving only one LP and n_r systems of linear equations of the form (11).

Moreover, we can find all the flux coupling relations by repeatedly using this procedure on metabolic networks with a reaction removed, in order to find all the reactions directionally coupled to it. Thereafter, partial couplings follow from directional couplings trivially.

Finally, we try to find all the fully coupled pairs of reactions without solving any optimization problem and any LP in particular. We have seen that full couplings are deduced from FCEs. Thus, we can check whether each $R_i \iff R_j$ is true or not by searching for a $\lambda \in \mathbb{R}^n$ such that $\lambda_l = 0 \Leftrightarrow l \neq i, j$, and $\lambda^T v = 0$ for all $v \in \mathcal{C}$. If there exists such a λ, then it contains the corresponding FCE embedded in itself which can be retrieved by

$$v_i = -\frac{\lambda_j}{\lambda_i} v_j.$$

Without loss of generality, assume that $\lambda_i = 1$ and thus $\lambda_j = -c$. By applying Theorem 2.1, there exists $\nu \in \mathbb{R}^m$ such that $\lambda = S^T \nu$. Therefore, we can solve for ν by

$$S^T \nu = e_i - c e_j. \tag{12}$$

Nevertheless, there is a subtlety here that if we have not determined whether $R_i \iff R_j$ or not yet, we do not know the value of c in advance. In the next Sect. 3.2, we will revisit this obstacle in greater detail.

This derives the promised FCA algorithm which solves at most $1 + n_i$ LPs, which is linear in n. Later in Sect. 3.3, we will find out that even the same number of LPs is also enough to infer all the corresponding DCEs and EDCEs too.

3.2 Positive and negative certificates

In the previous subsection, we designed an efficient algorithm to carry out FCA which is the first two tasks of Sect. 2.5, but the third one still remains. Fortunately, standard results from the theory of Lagrange duality provide dual variables as optimality certificates for any optimization problem solved. And by the way we defined (9), its dual problem turns out to be (13) which is the key to computing the desired flux coupling equations. For more details on the derivation of this dual problem and a brief overview of the theory of Lagrange duality, we refer the interested reader to the "Appendix" of this article.

To recap the results of Sect. 1, recall that if $v^\star$ and $\lambda^\star$ are a primal-dual optimal pair for the problems (9) and

$$\begin{aligned} &\text{maximize } \mathbf{1}^T \min(\lambda, \mathbf{1}) \\ &\text{subject to } S^T v = \lambda \\ &\qquad\qquad \lambda_R = 0 \\ &\qquad\qquad \lambda_I \geq 0, \end{aligned} \tag{13}$$

then exactly one of $v_i^\star$ and $\lambda_i^\star$ equals zero for each $R_i \in \mathcal{I}$. We think of $\lambda_i \neq 0$ as a *positive certificate* verifying that R_i is blocked and $v_i \neq 0$ as a *negative certificate* verifying that R_i is unblocked. Last but not least, there is no additional computational cost for getting the optimal dual variables since most LP solvers compute both $v^\star$ and $\lambda^\star$ jointly when solving either (9) or (13).

Up to here, we can determine $\mathcal{B}_I$ by just one LP of the form either (9) finding negative certificates or (13) finding positive certificates. In order to determine $\mathcal{B}_R$, we continue in the same manner by again defining negative and positive certificates in analogy. We already know that any solution to (11) proves $R_i \notin \mathcal{I}$ is unblocked, hence we call it a negative certificate.

On the other hand, assume that $R_i \notin \mathcal{I}$ is blocked thus there does not exists $v \in \mathcal{C}$ such that $v_i \neq 0$. Since $v_i = 0$ for all $v \in \mathcal{C}$, if $\lambda \in \mathbb{R}^n$ is any vector of all zeros except for its ith entry $\lambda_i \neq 0$, then

$$\lambda^T v = 0, \quad \forall v \in \mathcal{C}.$$

However by Theorem 2.1, we know that this implies $\lambda = S^T \nu$ for some $\nu \in \mathbb{R}^m$.

To sum up, $R_i \notin \mathcal{I}$ is blocked if and only if there exists $\nu \in \mathbb{R}^m$ such that $S^T \nu = \lambda$, and $\lambda_i \neq 0$ but the rest of the entries of λ are zero. Similar to what we had before, we call such λ a positive certificate that R_i is blocked. As opposed to the previous case of irreversible reactions, this time positive certificates can be found by just solving the system of linear equations

$$S^T x = e_i. \tag{14}$$

As a simple observation, (12) reduces to (14) after the removal of the jth column of S and the jth entry of e_i. For an index set A, let X^A and x^A denote the corresponding columns of the matrix X and the corresponding entries of the vector x. Moreover, let $X^{(A)}$ and $x^{(A)}$ denote the result of removing X^A and x^A from X and x, respectively. By this notation, in order to solve the system of linear equations (12) for both v and c, it is enough to solve

$$(S^{(j)})^T x = e_i^{(j)},$$

and then $v = x$ is a solution of (12) for $c = S^{j\,T} x$. The converse is also true that any solution of (12) is a solution of the above equation. This simple trick overcomes the previously mentioned obstacle of solving (12) when we do not know the value of c beforehand.

To conclude this subsection, recall from Sect. 3.1 that $R_i \iff R_j$ if and only if there exists $v \in \mathbb{R}^m$ such that $S^T v = \lambda$, where λ is the corresponding FCE. In this case, we have already shown that λ can be found from (12) if we only know the value of v. And here we showed how to determine if such a v exists. Accordingly, we call this fictitious metabolite form of an FCE a positive certificate for $R_i \iff R_j$. This is the first kind of QFCA positive certificates, but there remain two more which we will examine in the next subsection.

3.3 Directional coupling equations

We have already proved the "if" parts of the following theorem in Sect. 2.3, and we will give a constructive proof of the "only if" parts in this subsection.

Theorem 3.1 (Directional coupling equations) *Suppose that in a given metabolic network specified by S and $\mathcal{I}$, there are no irreversible blocked reactions. Let R_j be an arbitrary unblocked reaction, and $\mathcal{D}_j$ denote the set of all the irreversible reactions which are directionally coupled to R_j excluding itself. Then, $\mathcal{D}_j$ is nonempty if and only if there exists a DCE such that for all $v \in \mathcal{C}$,*

$$v_j = \sum_{d:R_d \in \mathcal{D}_j} c_d v_d, \tag{15}$$

for some fixed $c_d > 0$.

Furthermore, for any unblocked $R_i \notin \mathcal{I}$, $R_i \longrightarrow R_j$ if and only if there exists an EDCE such that for all $v \in \mathcal{C}$,

$$v_j = \sum_{d:R_d \in \mathcal{D}_j} c'_d v_d + c'_i v_i, \tag{16}$$

where $c'_i \neq 0$.

Remark Comparing (3) and (15), directional coupling can be thought of as full coupling between a set of reactions $\mathcal{D}_j$ and a single one R_j. One aspect of this is the following weak converse for directional coupling relations. If R_j is implied by $\mathcal{D}_j$ and no other irreversible reaction, then the activity of R_j implies that at least one reaction in the set $\mathcal{D}_j$ must be active too. The proof is immediate by (15). This converse gives another way of thinking about directional coupling relations as the weak variant of the stronger condition of full coupling relations.

Back to the problem of computing the coupled reactions to an arbitrary reaction R_j, recall that Proposition 1 states if $R_i \longrightarrow R_j$ and $R_j \notin \mathcal{I}$, then $R_i \Longleftrightarrow R_j$. In addition, we have already resolved the case of full coupling, hence from now on we assume that $R_j \in \mathcal{I}$.

Let $\mathcal{D}_j$ denote the set of all the directionally coupled irreversible reactions to R_j excluding itself. We have already argued that all we need in order to determine $\mathcal{D}_j$ is finding $\mathcal{B}_I$ after imposing the additional constraint $v_j = 0$. Analogues to (9), the zero entries of any solution to

$$\begin{aligned} &\text{maximize } \mathbf{1}^T \min(v_I, \mathbf{1}) \\ &\text{subject to } v \in \mathcal{C} \\ &\qquad\qquad v_j = 0, \end{aligned}$$

determine $\mathcal{B}_I$. Therefore, we call such v a negative certificate.

Applying the results of Sect. 3.2, we can also determine $\mathcal{D}_j$ for $R_j \in \mathcal{I}$ by solving the following optimization problem

$$\begin{aligned} &\text{maximize } \mathbf{1}^T \min(\lambda^{(j)}, \mathbf{1}) \\ &\text{subject to } S^T v = \lambda \\ &\qquad\qquad \lambda_i = 0, \quad i \notin \mathcal{I} \\ &\qquad\qquad \lambda_i \geq 0, \quad i \in \mathcal{I} \setminus \{j\}, \end{aligned} \tag{17}$$

where $v \in \mathbb{R}^m, \lambda \in \mathbb{R}^n$. Intuitively, this optimization problem is equivalent to (13) after the removal of the jth column of S, and the jth entry of λ. In other words, this is the problem of finding $\mathcal{B}_I$ after the removal of R_j, hence $\mathcal{D}_j$ is precisely the set of positive entries of λ.

Suppose that $\mathcal{D}_1 = \{R_2, R_3, \ldots, R_l\}$ for some $2 \leq l \leq k$. From our earlier arguments, we have $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_l, 0, \ldots, 0)$ where $\lambda_2, \lambda_3, \ldots, \lambda_l > 0$. Therefore, for any feasible flux distribution $v \in \mathcal{C}$, we get

$$v_1 = -\frac{\lambda_2}{\lambda_1} v_2 - \frac{\lambda_3}{\lambda_1} v_3 - \cdots - \frac{\lambda_l}{\lambda_1} v_l.$$

Since $R_1 \in \mathcal{I}$, thus $\lambda_1 < 0$. We see that for any such λ, $\lambda v = 0$ is actually a DCE, hence this proves the first part of Theorem 3.1. Up to now, this is the second kind of QFCA positive certificates, corresponding to the directional coupling relations.

The only remaining case would be if $R_i \notin \mathcal{I}$ and $R_j \in \mathcal{I}$. By definition, $R_i \longrightarrow R_j$ if and only if after the removal of R_j, R_i becomes blocked. Thus, we can resolve this case by finding $\mathcal{B}_R$ after the removal of R_j and the irreversible reactions which imply it, namely $\mathcal{D}_j$. Hence, the negative certificates can found by solving (11) after substituting S and e_i by $S^{(\mathcal{D}_j \cup \{R_j\})}$ and $e_i^{(\mathcal{D}_j \cup \{R_j\})}$, respectively.

To finish the proof of Theorem 3.1, the third and final kind of positive certificates that the QFCA algorithm computes is the fictitious metabolite form of an EDCE. Therefore, it is supposed to be of the form $\lambda = S^T v$ such that $\lambda_l = 0$ for all $l \notin \mathcal{D}_j \cup \{i, j\}$, and $l_i \neq 0$. Like the previous cases, we solve for this fictitious metabolite form of an EDCE by the following analogue of (14)

Table 1 A bird's eye view of QFCA

	Positive certificates		Negative certificates		A
$\mathcal{B}_R$			$S^{(A)}u = 0$		$\emptyset$
EDCE	$\left(S^{(A)}\right)^T x = e_i^{(A)}$		${e_i^{(A)}}^T u = 1$		$\mathcal{D}_j \cup \{R_j\}$
FCE					$\{j\}$
$\mathcal{B}_I$	maximize	$\mathbf{1}^T \min(\lambda^{(A)}, \mathbf{1})$	maximize	$\mathbf{1}^T \min(v_I, \mathbf{1})$	$\emptyset$
DCE	subject to	$S^T \nu = \lambda$	subject to	$v \in \mathcal{C}$	$\{j\}$
		$\lambda_i = 0, \quad i \notin \mathcal{I}$		$v_A = 0$	
		$\lambda_i \geq 0, \quad i \in \mathcal{I} \setminus A$			

$$\left(S^{(\mathcal{D}_j \cup \{R_j\})}\right)^T x = e_i^{(\mathcal{D}_j \cup \{R_j\})}.$$

In the end, by the definition of partial coupling $R_i \longleftrightarrow R_j$ if and only if both $R_i \longrightarrow R_j$ and $R_j \longrightarrow R_i$. Therefore, this part reduces to the previous parts and we determine a partial coupling relation by either a pair of positive certificates to prove or a single negative certificate to disprove it. For an overview, see Table 1.

As a final remark, one can use the same negative certificates for the full coupling relations too, because if such a negative certificate exists the pair of interest cannot be fully coupled. However, only in the case that the reactions are partially but not fully coupled, though there does not exist any such negative certificate, yet they are not fully coupled to each other. We waive these elusive instances of partially but not fully coupled reactions as they are rare and awkward (Marashi and Tefagh 2014). Furthermore, there are always positive certificates which can be used instead of negative certificates, particularly for the full coupling which is itself defined by a positive certificate, namely FCE.

4 Discussion

One possible biological interpretation of the positive certificates is from viewing the dual variable ν as the vector of chemical potentials for different metabolites in $\mathcal{M}$. In this perspective, the dual variable λ is the vector of potential differences between the products and the reactants for different reactions in $\mathcal{R}$.

If there exists a potential vector ν for which all the reactions have nonnegative potential differences and some reactions have strictly positive ones, then this is a positive certificate that the latter reactions must be blocked. Extending the same argument for the flux coupling relations, if a potential difference vector λ is positive only for one reaction R_i, then the activity of any reaction with strictly negative potential difference implies the activity of R_i.

A thermodynamic constraint which has gained popularity in many recent constraint-based models is the loop-law (Beard et al. 2004), which itself is a special case of the second law of Thermodynamics. For the interested reader, we note that though our

approach can also be interpreted by the second law of Thermodynamics applied to the whole system instead of single reactions, this is different from the loop-law which asserts negative potential difference for every active reaction. Here, we do not intend to delve in greater depth and seek to explore a different direction.

Another interpretation for λ that we give in here is to think of it as a fictitious metabolite like the example depicted in Fig. 1b. Since by Theorem 2.1 we have $S^T \nu = \lambda$, the dual variable λ is a linear combination of metabolites. As a consequence, it can be treated as a metabolite itself. In particular, the steady-state constraint holds for λ because for any feasible flux distribution v, it holds that

$$\lambda^T v = \nu^T S v = \nu^T 0 = 0.$$

Considering the mass balance for this fictitious metabolite, if it is a dead-end whereby we mean it is either only produced or only consumed in the metabolic network, then all of the reactions producing or consuming it must be blocked.

On the other hand, if λ is nonpositive for all but one reaction R_i, it means that the only reaction producing this fictitious metabolite is R_i (see Fig. 2). Hence, the activity of any of the reactions consuming it implies the activity of R_i.

This point of view may be of biological interest as well. Here follows the fictitious metabolite computed by QFCA to find the reactions coupled to "ribose-5-phosphate isomerase" from E. coli core model (Orth et al. 2010).

$$\begin{aligned} M = {} & 4 \times 13dpg[c] + 2 \times 2pg[c] + 2 \times 3pg[c] \\ & + 4.8756 \times 6pgc[c] + 3.8756 \times 6pgl[c] + 2 \times actp[c] \\ & - 2 \times adp[c] - 4 \times amp[c] + 2 \times dhap[c] \\ & - 1.8756 \times e4p[c] + 2 \times f6p[c] + 4 \times fdp[c] \\ & + 2 \times g3p[c] + 2 \times g6p[c] + 2 \times pep[c] \\ & + 2 \times pi[c] + 1 \times pi[e] - 5.7513 \times r5p[c] \\ & + 5.8756 \times ru5p - D[c] - 1.8756 \times s7p[c] \\ & + 5.8756 \times xu5p - D[c] \end{aligned}$$

Looking up the full names of these metabolites, they are nearly all the metabolites in this network which contain phosphate.

- 3-Phospho-D-glyceroyl-phosphate
- D-Glycerate-2-phosphate
- 3-Phospho-D-glycerate
- 6-Phospho-D-gluconate
- 6-phospho-D-glucono-1-5-lactone
- Acetyl-phosphate
- ADP
- AMP
- Dihydroxyacetone-phosphate
- D-Erythrose-4-phosphate
- D-Fructose-6-phosphate

- D-Fructose-1-6-bisphosphate
- Glyceraldehyde-3-phosphate
- D-Glucose-6-phosphate
- Phosphoenolpyruvate
- Phosphate (pi[c])
- Phosphate (pi[e])
- alpha-D-Ribose-5-phosphate
- D-Ribulose-5-phosphate
- Sedoheptulose-7-phosphate
- D-Xylulose-5-phosphate

We have observed the same property for any other reaction in the instances we looked through, namely, the associated fictitious metabolite found by QFCA is a linear combination of the biologically relevant metabolites. In conclusion, the concept of fictitious metabolite introduced here can be a subject of future research, especially insightful for genome-scale metabolic networks where all the biological pathways cannot be annotated manually.

5 Applications of QFCA

5.1 A quantitative approach to FCA

QFCA redefines coupling relations in a more informative setting than FCA without losing any generality. The fact that the strength of all kinds of flux coupling relations can be measured numerically has not been investigated up to now. More specifically, (15) implies that

$$v_j \geq c_d v_d, \quad \forall R_d \in \mathcal{D}_j,$$

and (16) implies that

$$v_j \geq c'_i v_i.$$

Conversely, we claim that all such lower bounds can be derived from some DCEs or EDCEs.

Suppose that there exists a constant $c \neq 0$ such that

$$v_j \geq c v_i,$$

for all the feasible flux distributions $v \in \mathcal{C}$. Another way of putting this is to say that the optimal objective value for the following LP is zero.

$$\begin{aligned} &\text{minimize } v_j - c v_i \\ &\text{subject to } v \in \mathcal{C} \end{aligned}$$

This in turn is equivalent to the feasibility of the following dual problem.

$$\begin{aligned} &\text{maximize } 0 \\ &\text{subject to } S^T v + e_j - ce_i = \lambda \\ &\qquad\qquad \lambda_i = 0, \quad i \notin \mathcal{I} \\ &\qquad\qquad \lambda_i \geq 0, \quad i \in \mathcal{I} \end{aligned}$$

However if $v^\star$ and $\lambda^\star$ are dual feasible, then we have that

$$(1 - \lambda_j^\star)v_j = (c + \lambda_i^\star)v_i + \sum_{d \neq i,j} \lambda_d^\star v_d,$$

which is a DCE or EDCE that proves a stronger or at least equally strong inequality.

Historically, FCA itself was originally stated as flux ratio maximizations or minimizations (Burgard et al. 2004), which can also be done by DCEs and EDCEs as we have just shown. The only competing approach is to consider the following linear fractional program

$$\begin{aligned} &\text{minimize } \tfrac{v_j}{v_i} \\ &\text{subject to } v \in \mathcal{C}, \end{aligned}$$

and following Charnes–Cooper's transformation (Horst and Pardalos 2013), fix the rate of R_i to 1 for $R_i \in \mathcal{I}$ and to both $+1$ and -1 for $R_i \notin \mathcal{I}$ and then run *flux variability analysis* (FVA) (Fell and Small 1986; Savinell and Palsson 1992; Gudmundsson and Thiele 2010) to get lower and upper bounds on R_j.

Although these bounds provide quantitative information on the dependencies among reactions, this approach is complementary to what QFCA does. In the result of FVA, except for the fully coupled reactions where the lower and upper bounds meet at some specific value, we only get inequalities. However, QFCA gives equations like FCE, DCE, and EDCE which can also turn into inequalities. From this point of view, FVA is a complement to QFCA, not an alternative. Also it is worth mentioning, that conducting this procedure requires quadratically many LPs, namely $2(n-1)(n_i+2n_r)$, while QFCA is linear-time and only requires no more than $n+1$ LPs to be solved.

5.2 Sensitivity analysis and the metabolic gap-filling problem

Genome-scale metabolic network reconstructions are always susceptible to have many missing reactions creating a major challenge in the post-genomic era to pinpoint these missing biological components (Rolfsson et al. 2011). In (Marashi and Bockmayr 2011), the authors discuss the significant sensitivity of FCA results to this incomplete data. The main point is that we have no clue which blocked reactions can become unblocked after the addition of some missing reactions to the metabolic network. QFCA pinpoints the associated metabolites which cause any single flux coupling relation or blocked reaction as was discussed in Sect. 4. Therefore, even if some metabolites are suspect to involve missing reactions, we still know which flux coupling relations or blocked reactions are for sure true anyway.

More generally, suppose that we replace $\mathcal{I}$ by another close enough $\mathcal{I}'$, and S by S' where a few entries have changed, and even some columns are added. What is key in the fictitious metabolite form of the flux coupling equations is that if they satisfy the conditions for positive certificates, it does not matter how they have been discovered in the first place. We can pad zeros to ν corresponding to the added columns, and it might be the case that ${S'}^T \nu = \lambda$ still holds because ν is often sparse and it is only enough that the stoichiometry of the metabolites for which $\nu_i \neq 0$ is not changed. Then again, λ is often sparse, and it might satisfy the required irreversibility conditions since it is only enough that the irreversibility of the reactions for which $\lambda_i > 0$ is not changed for DCEs, and there is no requirement for FCEs and EDCEs.

To the best of our knowledge, all the other FCA methods rely on the optimality of the solutions to LPs like (7) to prove flux coupling relations in the generic case, in contrast to the easily verifiable certificates of QFCA. Consequently, their results are potentially vulnerable to any single missing reaction unless we double-check the optimality of the corresponding solutions. For QFCA the sanity-check, as explained in the previous paragraph, is much more computationally efficient since feasibility [*e.g.* for DCE, with respect to the constraint of (17)] is enough to test if positive certificates are valid. Depending on the confidence of our model, one can search for more robust positive certificates by sparse regularization.

In the metabolic gap-filling problem (Reed et al. 2006; Satish Kumar et al. 2007; Orth and Palsson 2010; Thiele et al. 2014), we search for a reaction which needs to be added so that at least one formerly blocked reaction can admit a nonzero flux. In our sensitivity analysis, we used the fact that for each flux coupling relation or blocked reaction, the corresponding QFCA certificates specify the liable metabolites. Turning this argument around, fictitious dead-end metabolites implicitly mark the responsible subnetworks for blocked reactions. In the metabolic gap filling problem, we can restrict our search to these subnetworks because any reaction outside them is provably irrelevant.

5.3 Implementation and runtime

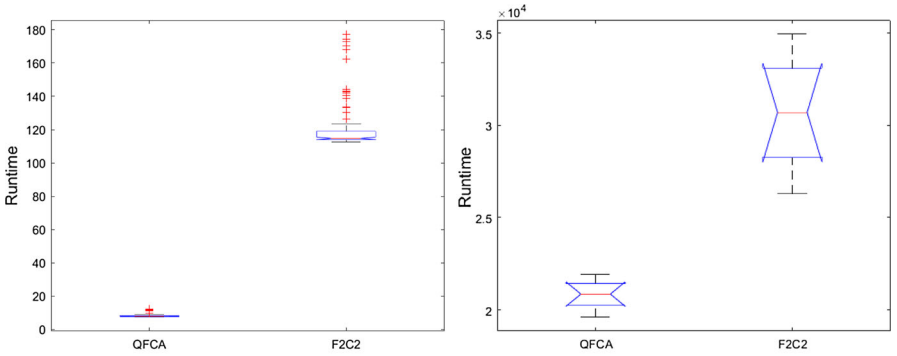
All the examples in this study were conducted by the accompanying MATLAB© package freely available for non-commercial use at GitHub[1]. This package implements QFCA as described in Procedure 1, using Gurobi™ optimizer or the built-in *linprog* of MATLAB© as the underlying LP solver. For convenience in comparing the results, we use the same format for the output as F2C2, *i.e.*, a 0-1 indicator vector b specifying the blocked reactions by ones, and an $n \times n$ flux coupling matrix A with the possible values 0, 1, 2, 3, 4 for its (i, j) entry indicating uncoupled, fully coupled, partially coupled, reaction i is directionally coupled to j, reaction j is directionally coupled to i, respectively.

From our earlier discussions in Sect. 3, QFCA flux coupling relations are computed by solving at most n LPs in total. If the initial metabolic network is not consistent, the preprocessing step associated with removing the blocked reactions requires one additional LP.

[1] https://mtefagh.github.io/qfca/.

Procedure 1 QFCA

Input: $\mathcal{M}, \mathcal{R}, S, \mathcal{I}$
Output: A, b
Identifying the blocked reactions and removing them from the metabolic network Sect. 3.2
Aggregating all the isozymes and removing the newly blocked reactions once more (Burgard et al. 2004)
Finding the fully coupled pairs of reactions and merging each pair into a single one (Larhlimi et al. 2012)
Computing the set of fully reversible reactions and the reversibility type pruning (David et al. 2011)
Identifying the directional and partial coupling relations by positive certificates Sect. 3.3



(a) YEASTNET v3.0 with 2292 reversible and 49 irreversible reactions

(b) Recon3D with 5238 reversible and 5362 irreversible reactions

Fig. 5 QFCA average runtime is 7% and 68% of F2C2 average runtime, respectively

Further speed-up is achieved by implementing two preprocessing techniques which prune some trivial cases by solving merely one LP, namely the reversibility type pruning as explained in FFCA (David et al. 2011), and the aggregation of isozymes as explained in *flux coupling finder* (FCF) (Burgard et al. 2004) (afterward, we compute the blocked reactions since the reaction pairs that become blocked after merging isozymes are fully coupled to each other). Also, we merge the fully coupled reactions after identifying them in the same way as F2C2 [see Figure 1 in (Larhlimi et al. 2012)].

Comparing the runtime of QFCA with F2C2 which is the state-of-the-art algorithm, we see substantial improvements in performance (solving 43 LPs and 9 systems of linear equations) when the metabolic network has mostly irreversible reactions, *e.g.*, version 3 of the YEASTNET (Herrgård et al. 2008) (see Fig. 5a), and comparable performance (solving 5458 LPs and 474 systems of linear equations) when the metabolic network has many reversible reactions, *e.g.*, Recon3D (Brunk et al. 2018) (see Fig. 5b). Note that both algorithms yield the same result, *i.e.*, all the directionally, partially, and fully coupled pairs of reactions.

6 Conclusions

The main significance of this work is that all the flux coupling relations, initially proposed by FCA, can be obtained by the more informative flux coupling equations (see Sect. 2.3). In most cases, we should compromise performance in order to get

more information, but in this case, it turns out that the more general approach is even more efficient from the computational viewpoint as demonstrated by both a theoretical worst-case analysis (see Sect. 3.1), and two real-world examples. On the whole, besides investigating the new concepts (see Sect. 4), we can reproduce many of the previously done research more precisely (see Sect. 5.1), more robustly (see Sect. 5.2), and more efficiently (see Sect. 5.3).

Appendix

Derivation of the dual problem for (10)

For the standard definitions from the theory of Lagrange duality used in this appendix, we refer the reader to the fifth chapter of (Boyd and Vandenberghe 2004).

By definition, the Lagrangian for the LP (10) is equal to

$$\mathcal{L}(u, v, \lambda_1, \lambda_2, \lambda_3, \nu) = -\mathbf{1}^T u + v^T S v + \lambda_1^T (u - v_I) + \lambda_2^T (u - \mathbf{1}) + \lambda_3^T (-u),$$

where $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}^k$, and $\nu \in \mathbb{R}^m$ are the Lagrange dual variables. The Lagrange dual function for this Lagrangian is defined as

$$g(\lambda_1, \lambda_2, \lambda_3, \nu) = \inf_{u \in \mathbb{R}^k, v \in \mathbb{R}^n} \mathcal{L}(u, v, \lambda_1, \lambda_2, \lambda_3, \nu).$$

Let $p^\star$ denote the optimal objective value and assume $\lambda_i \geq 0$ for $i = 1, 2, 3$. If $u^\star$ and $v^\star$ are optimal, then

$$\begin{aligned}
g(\lambda_1, \lambda_2, \lambda_3, \nu) &= \inf_{u \in \mathbb{R}^k, v \in \mathbb{R}^n} \mathcal{L}(u, v, \lambda_1, \lambda_2, \lambda_3, \nu) \\
&\leq \mathcal{L}\big(u^\star, v^\star, \lambda_1, \lambda_2, \lambda_3, \nu\big) \\
&= -\mathbf{1}^T u^\star + v^{\star T} S v^\star + \lambda_1^T \big(u^\star - v_I^\star\big) + \lambda_2^T (u^\star - \mathbf{1}) + \lambda_3^T (-u^\star) \\
&= -\mathbf{1}^T u^\star + \lambda_1^T \big(u^\star - v_I^\star\big) + \lambda_2^T (u^\star - \mathbf{1}) + \lambda_3^T (-u^\star) \\
&\leq -\mathbf{1}^T u^\star \\
&= -p^\star.
\end{aligned}$$

Therefore, for any nonnegative $\lambda_1, \lambda_2, \lambda_3$, and any ν, the negative Lagrange dual function yields an upper bound on $p^\star$. In order to find the tightest bound, we should solve the following Lagrange dual problem

$$\begin{aligned}
&\text{maximize } \mathbf{1}^T \lambda_2 \\
&\text{subject to } S^T \nu = \begin{bmatrix} \lambda_1 \\ 0 \end{bmatrix} \\
&\qquad\qquad \lambda_1 + \lambda_2 \geq \mathbf{1} \\
&\qquad\qquad \lambda_1 \geq 0 \\
&\qquad\qquad \lambda_2 \geq 0,
\end{aligned}$$

over the Lagrange dual variables. The proof follows from rewriting the Lagrangian as

$$\mathcal{L}(u, v, \lambda_1, \lambda_2, \lambda_3, \nu) = \left(S^T \nu - \begin{bmatrix} \lambda_1 \\ 0 \end{bmatrix}\right)^T v - \lambda_2^T \mathbf{1} + (-\mathbf{1} + \lambda_1 + \lambda_2 - \lambda_3)^T u.$$

Again, the dual LP is always both feasible (*e.g.*, $\lambda_1 = \lambda_2 = 0$, $\nu = 0$ is feasible) and bounded ($\mathbf{1}^T \lambda_2 \leq k$).

For this primal-dual pair of LPs strong duality holds which means that the gap between the dual optimal objective and the primal optimal objective $p^\star$ is zero. In other words, the bound given by the optimal dual variables is sharp.

It is easily seen that

$$\lambda_2^\star = \max(\mathbf{1} - \lambda_1^\star, 0),$$

has either zero or one entries just like the optimal $u^\star$. However, from zero duality gap

$$\mathbf{1}^T u^\star = \mathbf{1}^T \lambda_2^\star,$$

hence $\lambda_2^\star$ and $u^\star$ have the same number of ones. Also by complementary slackness, $\lambda_2^\star$ is zero wherever $u_i^\star \neq 1$. Altogether, $\lambda_2^\star$ and $u^\star$ are 0-1 vectors with the same sparsity pattern, thus they are equal.

Ultimately, we can also rewrite the dual problem as (13) in analogy to the primal problem (9), by substituting

$$\lambda = \begin{bmatrix} \lambda_1 \\ 0 \end{bmatrix}.$$

Proofs of Sect. 2.4

Proof of Theorem 2.1 From the Lemma 2.2, we can assume the seemingly stronger but equivalent right hand side of (4), which by definition means that $\lambda \in \ker(S)^\perp$. From rank-nullity theorem we know that $\ker(S)^\perp = \text{range}(S^T)$. Thus, $\lambda \in \text{range}(S^T)$ which in turn implies the desired result. The converse also holds because in the reverse direction we have $\lambda = S^T \nu$ and hence, for any $v \in \mathcal{C}$,

$$\lambda^T v = \nu^T S v = \nu^T 0 = 0.$$

□

Proof of Lemma 2.2 ($\Leftarrow$) is immediate from $\mathcal{C} \subseteq \ker(S)$. For the other direction, suppose that the left-hand side is true.

The proof goes by contradiction. Assume to the contrary that there exists $u \in \ker(S)$ such that $\lambda^T u \neq 0$. Let

$$v = u + \sum_{i : R_i \in \mathcal{I}} |u_i| v^i,$$

where for any $R_i \in \mathcal{I}$, v^i is an arbitrary feasible flux distribution with its ith flux coefficient equal to one, namely $v_i^i = 1$. One can easily show that the feasibility constraints (1) and (2) hold for v, and hence $v \in \mathcal{C}$.

On the other hand,

$$\lambda^T v - \sum_{i:R_i \in \mathcal{I}} |u_i| \lambda^T v^i = \lambda^T u \neq 0,$$

by the way we constructed v. Therefore, at least one of the terms on the left hand side is nonzero. The proof is complete since this is in contradiction with the assumption that $\lambda^T v = 0$ for all $v \in \mathcal{C}$. □

Reversibility type correction

Earlier in Sect. 2.5, we have derived a naive method for reversibility type correction by solving $2n_r$ LPs. Recall that for $R_j \notin \mathcal{I}$, our task is to figure out if both its forward and reverse directions are unblocked, and as always we assume that all the blocked reactions are already removed.

Following the same fashion, this time we try to search for positive certificates proving that either the forward or reverse direction of R_j becomes blocked when R_j is added to $\mathcal{I}$ (for the reverse direction we should also replace S^j by $-S^j$). Applying (13) to the resulting modified metabolic network, either $\lambda^\star = 0$ or $\lambda_j^\star \neq 0$, otherwise the same $\lambda^\star$ can be considered as a positive certificate for the original metabolic network where $R_j \notin \mathcal{I}$ proving that some reactions other than R_j are blocked, which is in contradiction to the assumption that we have already removed all the blocked reactions.

If $\lambda_j^\star < 0$, then $\frac{\lambda^\star}{-\lambda_j^\star}$ is in fact a DCE which clearly shows $v_j > 0$ for all $v \in \mathcal{C}$. If $\lambda_j^\star > 0$, then $\frac{\lambda^\star}{\lambda_j^\star}$ shows $v_j < 0$ for all $v \in \mathcal{C}$ and becomes a DCE if we replace S^j by $-S^j$ permanently. If $\lambda_j^\star = 0$ and hence $\lambda^\star = 0$ for both directions, then R_j is truly reversible. As a consequence, we have just shown that in the modified metabolic network either R_j is not blocked in the selected direction or some other irreversible reactions should also become blocked which means that R_j is effectively irreversible because some irreversible reactions are directionally coupled to it.

In this way, DCEs can also be used for reversibility type correction with two major advantages over the naive method. The first one is that in order to derive DCEs with either $\lambda_j^\star < 0$ or $\lambda_j^\star > 0$, it is enough to solve (17) as there is no constraint on the sign of λ_j. This is a twofold decrease in the number of required LPs and brings the total number of them down to n_r. The second advantage is that we also get the $\mathcal{D}_j$ for free by the same LP.

References

Beard DA, Babson E, Curtis E, Qian H (2004) Thermodynamic constraints for biochemical networks. J Theor Biol 228(3):327–333

Bonarius HP, Schmid G, Tramper J (1997) Flux analysis of underdetermined metabolic networks: the quest for the missing constraints. Trends Biotechnol 15(8):308–314
Boyd S, Vandenberghe L (2004) Convex optimization. Cambridge University Press, Cambridge
Brunk E, Sahoo S, Zielinski DC, Altunkaya A, Dräger A, Mih N, Gatto F, Nilsson A, Gonzalez GAP, Aurich MK et al (2018) Recon3D enables a three-dimensional view of gene variation in human metabolism. Nat Biotechnol 36(3):272
Burgard AP, Nikolaev EV, Schilling CH, Maranas CD (2004) Flux coupling analysis of genome-scale metabolic network reconstructions. Genome Res 14(2):301–312
Covert MW, Schilling CH, Famili I, Edwards JS, Goryanin II, Selkov E, Palsson BØ (2001) Metabolic modeling of microbial strains in silico. Trends Biochem Sci 26(3):179–186
David L, Marashi S-A, Larhlimi A, Mieth B, Bockmayr A (2011) FFCA: a feasibility-based method for flux coupling analysis of metabolic networks. BMC Bioinform 12(1):236
Dreyfuss JM, Zucker JD, Hood HM, Ocasio LR, Sachs MS, Galagan JE (2013) Reconstruction and validation of a genome-scale metabolic model for the filamentous fungus neurospora crassa using farm. PLoS Comput Biol 9(7):e1003126
Fell DA, Small JR (1986) Fat synthesis in adipose tissue an examination of stoichiometric constraints. Biochem J 238(3):781–786
Gudmundsson S, Thiele I (2010) Computationally efficient flux variability analysis. BMC Bioinform 11(1):489
Gunawardena J (2014) Time-scale separation-michaelis and menten's old idea, still bearing fruit. FEBS J 281(2):473–488
Haus U-U, Klamt S, Stephen T (2008) Computing knock-out strategies in metabolic networks. J Comput Biol 15(3):259–268
Herrgård MJ, Swainston N, Dobson P, Dunn WB, Arga KY, Arvas M, Blüthgen N, Borger S, Costenoble R, Heinemann M et al (2008) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. Nat Biotechnol 26(10):1155–1160
Horst R, Pardalos PM (2013) Handbook of global optimization, vol 2. Springer, Berlin
Larhlimi A, Bockmayr A (2006) A new approach to flux coupling analysis of metabolic networks. In: International symposium on computational life science. Springer, pp 205–215
Larhlimi A, David L, Selbig J, Bockmayr A (2012) F2C2: a fast tool for the computation of flux coupling in genome-scale metabolic networks. BMC Bioinform 13(1):57
Marashi S-A, Bockmayr A (2011) Flux coupling analysis of metabolic networks is sensitive to missing reactions. Biosystems 103(1):57–66
Marashi S-A, Tefagh M (2014) A mathematical approach to emergent properties of metabolic networks: partial coupling relations, hyperarcs and flux ratios. J Theor Biol 355:185–193
Notebaart RA, Teusink B, Siezen RJ, Papp B (2008) Co-regulation of metabolic genes is better explained by flux coupling than by network distance. PLoS Comput Biol 4(1):e26
Orth JD, Palsson BØ (2010) Systematizing the generation of missing metabolic knowledge. Biotechnol Bioeng 107(3):403–412
Orth J, Fleming R, Palsson B (2010) Reconstruction and use of microbial metabolic networks: the core escherichia coli metabolic model as an educational guide. EcoSal Plus. https://doi.org/10.1128/ecosalplus.10.2.1
Reed JL, Patel TR, Chen KH, Joyce AR, Applebee MK, Herring CD, Bui OT, Knight EM, Fong SS, Palsson BO (2006) Systems approach to refining genome annotation. Proc Natl Acad Sci 103(46):17480–17484
Rolfsson O, Palsson BØ, Thiele I (2011) The human metabolic reconstruction Recon 1 directs hypotheses of novel human metabolic functions. BMC Syst Biol 5(1):155
Satish Kumar V, Dasika MS, Maranas CD (2007) Optimization based automated curation of metabolic reconstructions. BMC Bioinform 8(1):212
Savinell JM, Palsson BØ (1992) Network analysis of intermediary metabolism using linear optimization. I. development of mathematical formalism. J Theor Biol 154(4):421–454
Schilling CH, Edwards JS, Palsson BØ (1999a) Toward metabolic phenomics: analysis of genomic data using flux balances. Biotechnol Progress 15(3):288–295
Schilling CH, Schuster S, Palsson BØ, Heinrich R (1999b) Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. Biotechnol Progress 15(3):296–303
Schuster S, Hilgetag C (1994) On elementary flux modes in biochemical reaction systems at steady state. J Biol Syst 2(02):165–182

Thiele I, Vlassis N, Fleming RM (2014) fastGapFill: efficient gap filling in metabolic networks. Bioinformatics 30(17):2529–2531

Varma A, Palsson BØ (1994) Metabolic flux balancing: basic concepts, scientific and practical use. Nat Biotechnol 12:994

Vlassis N, Pacheco MP, Sauter T (2014) Fast reconstruction of compact context-specific metabolic network models. PLoS Comput Biol 10(1):e1003424