F2C2: a fast tool for the computation of flux coupling in genome-scale metabolic networks

Abdelhalim Larhlimi¹,²*,†, Laszlo David³,⁴,⁵*,†, Joachim Selbig¹,² and Alexander Bockmayr³,⁴

Abstract
Background: Flux coupling analysis (FCA) has become a useful tool in the constraint-based analysis of genome-scale metabolic networks. FCA allows detecting dependencies between reaction fluxes of metabolic networks at steady-state. On the one hand, this can help in the curation of reconstructed metabolic networks by verifying whether the coupling between reactions is in agreement with the experimental findings. On the other hand, FCA can aid in defining intervention strategies to knock out target reactions.

Results: We present a new method F2C2 for FCA, which is orders of magnitude faster than previous approaches. As a consequence, FCA of genome-scale metabolic networks can now be performed in a routine manner.

Conclusions: We propose F2C2 as a fast tool for the computation of flux coupling in genome-scale metabolic networks. F2C2 is freely available for non-commercial use at https://sourceforge.net/projects/f2c2/files/.

Background
The huge amount of genomic, transcriptomic and related data has allowed for a fast reconstruction of an increasing number of genome-scale metabolic networks, e.g. [1-7]. In the absence of detailed kinetic information, constraint-based modeling and analysis has recently attracted ample interest due to its ability to analyze genome-scale metabolic networks using very few information [8-10]. Constraint-based analysis is based on the application of a series of constraints that govern the operation of a metabolic network at steady state. This includes the stoichiometric and thermodynamic constraints, which limit the range of possible behaviors of the metabolic network, corresponding to different metabolic phenotypes. Applying these constraints leads to the definition of the solution space, called the steady-state flux cone [11]:

$$C = \{ v \in \mathbb{R}^n \mid Sv = 0, \quad v_i \geq 0, \quad \text{for all } i \in \text{Irr} \}$$

(1)

where $S$ is the $m \times n$ stoichiometric matrix of the network, with $m$ internal metabolites (rows) and $n$ reactions (columns), and the vector $v \in \mathbb{R}^n$ gives a flux distribution. Furthermore, $\text{Irr} \subseteq \{1, \ldots, n\}$ denotes the set of irreversible reactions in the network, and $\text{Rev} = \{1, \ldots, n\} \setminus \text{Irr}$ denotes the set of reversible reactions.

The flux cone contains the full range of achievable behaviors of the metabolic network at steady state. Various approaches have been proposed either to search for single optimal behaviors using optimization-based methods [12-16] or to assess the whole capabilities of a metabolic network by means of network-based pathway analysis [11,17-20].

Flux coupling analysis (FCA) is concerned with describing dependencies between reactions [21]. The stoichiometric and thermodynamic constraints not only determine all possible steady-state flux distributions over a network, they also induce coupling relations between the reactions. For instance, some reactions may be unable to carry flux under steady-state conditions. If a non-zero flux through a reaction in steady-state implies a non-zero flux through another reaction, then the two reactions are said to be coupled (see Def. 2 for a formal definition). FCA has been used for exploring various biological questions such as network evolution [22-24], gene essentiality [22], gene regulation [25-27], analysis of experimentally measured fluxes [28,29], or implications of the structure of the human metabolic network for disease co-occurrences.

*Correspondence: Larhlimi@mpimp-golm.mpg.de; Laszlo.David@fu-berlin.de
†Equal contributors
¹Department of Bioinformatics, Institute for Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Str. 24-25, D-14476 Potsdam, Germany
³FB Mathematik und Informatik, Freie Universität Berlin, Arndtstraße 2, D-14476 Berlin, Germany
Full list of author information is available at the end of the article.

© 2012 Larhlimi et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Having a time efficient implementation of FCA is important in such studies.

After introducing the main existing algorithms for flux coupling analysis, we propose in this paper a new algorithm which significantly speeds up the calculation of flux coupling. Our algorithm is based on two main principles. First, we reduce the stoichiometric model as much as possible when parsing the stoichiometric matrix. Second, we use inference rules to minimize the number of linear programming problems that have to be solved. We prove the efficiency of our algorithm by successfully competing with the most recent approach [31]. We show that FCA can now be quickly performed even for very large genome-scale metabolic networks.

**Approaches for flux coupling analysis**

Several algorithms were developed to calculate flux coupling between reactions. For a comparison among the existing approaches, the reader may refer to [31,32]. In the following, we focus on flux coupling methods based on solving a sequence of linear programming (LP) problems. These methods have proved to be significantly faster than other algorithms.

**Definitions**

We give a short overview of the important concepts we will use throughout this paper. First, we formally define blocked reactions in a metabolic network.

**Definition 1 (Blocked reaction).** Given the steady-state flux cone $C$, let $i \in \{1, \ldots, n\}$ be a reaction. If $v_i = 0$, for all $v \in C$, reaction $i$ is called blocked, otherwise $i$ is unblocked.

In the following, we assume that the flux cone is not trivial, i.e., not all reactions are blocked. Next, we define the (un)coupling relationships between reactions.

**Definition 2 (Coupling relations).** Let $i,j$ be two unblocked reactions. The (un)coupling relationships $\rightarrow, \leftrightarrow, \Leftarrow$ and $\subseteq$ are defined in the following way:

- $i \rightarrow j$ if for all $v \in C, v_i = 0$ implies $v_j = 0$.
- $i \leftrightarrow j$ if for all $v \in C, v_i = 0$ is equivalent to $v_j = 0$.
- $i \Leftarrow j$ if there exists $\lambda \neq 0$ such that for all $v \in C, v_i = \lambda v_j$.
- $i \subseteq j$ if there exists $v \in C$ such that $v_i = 0$ and $v_j \neq 0$.

Reactions $i$ and $j$ are fully (resp. partially, directionally) coupled if the relation $i \Leftarrow j$ (resp. $i \leftrightarrow j$, $i \rightarrow j$) holds. Otherwise, $i$ and $j$ are uncoupled.

As shown in [33] the reversibility type of reactions is a key concept in flux coupling analysis.

**Definition 3 (Reversibility types).** A reversible reaction $i \in \text{Rev}$ is called fully reversible if there exists a flux vector $v \in C$ such that $v_i \neq 0$ and $v_i = 0$ for all $j \in \text{Irr}$. Otherwise, reaction $i$ is called pseudo-irreversible.

Using the reversibility type of reactions, we define the following reaction sets which will be used to determine the cases where flux coupling can occur between reactions.

- $\text{Frev} = \{i \mid i$ is fully reversible$\}$,
- $\text{Prev} = \{i \mid i$ is pseudo-irreversible and there exist $v^+, v^- \in C$ such that $v^+_i > 0, v^-_i < 0\}$,
- $\text{Irev} = \{i \mid i \notin \text{Frev} \cup \text{Prev} \text{ and } v_i \neq 0 \text{ for some } v \in C\}$,
- $\text{Blk} = \{i \mid i$ is blocked$\}$.

Note that the above reaction sets are disjoint and their union is equal to the set of all reactions, i.e., $\text{Blk} \cup \text{Irev} \cup \text{Prev} \cup \text{Frev} = \{1, \ldots, n\}$.

As the classification of reactions according to their reversibility types has been treated elsewhere [31,33,34], we focus on calculating the flux coupling between reactions given the reaction sets $\text{Blk}$, $\text{Irev}$, $\text{Prev}$ and $\text{Frev}$. First, we briefly review the main existing flux coupling methods based on linear programming. Afterwards, we present our new method to speed up the flux coupling analysis.

**Flux coupling finder**

The Flux Coupling Finder (FCF) algorithm [21] determines blocked reactions as well as coupled reactions by solving a sequence of linear programming (LP) problems. The FCF algorithm requires that each reversible reaction is split into two irreversible reactions, one forward and one backward. This may hamper the application of FCF to determine flux coupling in large genome-scale metabolic networks. Indeed, splitting reversible reactions results in an increase in the number of variables (resp. constraints) by $|\text{Rev}|$ (resp. $2|\text{Rev}|$). Since the FCF algorithm solves four LP problems to identify the maximum and minimum flux ratios for every pair of reactions, the total number of LP problems that have to be solved is very large. Furthermore, the FCF algorithm does not compute directly coupling relationships between reactions. A post-processing step is needed to deduce couplings between reactions (in the original network) from those between reaction directions (in the reconfigured network). More importantly, the FCF algorithm explores exhaustively all possible reaction pairs. This leads to a very big number of LP problems that have to be solved. This strategy may not scale well for genome-scale models of complex microorganisms which involve a large number of reactions.
Reversibility-based flux coupling analysis

To cope with the drawbacks of the FCF algorithm, we used the reversibility type of reactions to develop an improved version (WRP-FCF) of the original FCF method [31,34]. When looking for coupled reactions, WRP-FCF applies linear programming only in those cases where coupling relationships can occur [33]. Namely, given two unblocked reactions \(i\) and \(j\), we have exactly three cases where a flux coupling is possible between \(i\) and \(j\):

Flux coupling between reversible reactions

If \(i, j \in \text{Prev} \lor i, j \in \text{Irev}\), \(i \rightarrow j\), \(i \leftrightarrow j\) and \(i \leftarrow j\) are equivalent. More importantly, only the stoichiometric constraints determine whether \(i \rightarrow j\) holds, independently of the thermodynamic constraints.

Assume that blocked reactions have been identified beforehand and their respective columns in the stoichiometric matrix have been removed. Consider the following LP problem

\[
\max(v_j : Sv = 0, \quad v_i = 0, \quad 0 \leq v_j \leq 1),
\]
and let \(v^*\) be an optimal solution. According to Proposition 6.13 in [34], \(i \rightarrow j\) if and only if \(v^*_j = 0\).

Flux coupling between (pseudo-)irreversible reactions

If \(i \in \text{Irev}\) and \(j \in \text{Prev}\), we only have to check whether \(i \rightarrow j\). The other coupling relationships cannot occur. Let \(v^1\) resp. \(v^2\) be optimal solutions of the two LP problems

\[
\max(v_j : Sv = 0, v_k \geq 0, k \in \text{Irr}, \quad v_i = 0, \quad v_j \leq 1),
\quad \min(v_j : Sv = 0, v_k \geq 0, k \in \text{Irr}, \quad v_i = 0, \quad v_j \geq -1).
\]

Then \(i \rightarrow j\) if and only if \(v^1_j = v^2_j = 0\).

Flux coupling between irreversible reactions

If \(i, j \in \text{Irev}\), in analogy with the FCF algorithm, we determine upper and lower bounds \(U_{ij}\) and \(L_{ij}\) such that \(0 \leq L_{ij}v_j \leq v_i \leq U_{ij}v_j\) for all \(v \in C\) by solving the LP problems

\[
L_{ij} = \min \{v_j : Sv = 0, \quad v_j = 1, \quad v_k \geq 0, k \in \text{Irr}\},
U_{ij} = \max \{v_j : Sv = 0, \quad v_j = 1, \quad v_k \geq 0, k \in \text{Irr}\}. (4)
\]

Comparison of \(L_{ij}\) and \(U_{ij}\) allows us to determine whether reactions \(i\) and \(j\) are coupled using the following rules:

- \(i \rightarrow 0 \rightarrow j\) (resp. \(j \rightarrow 0 \rightarrow i\)) if and only if \(L_{ij} \neq 0\) (resp. \(U_{ij} \neq +\infty\)),
- \(j \leftarrow i\) if and only if \(L_{ij} = 0\), \(U_{ij} \neq 0\) and \(L_{ij} = U_{ij}\).

The improved version WRP-FCF does not make an exhaustive computation for all pairs of reactions and does not require a reconfiguration of the metabolic network. Accordingly, not only is the number of LP problems used by WRP-FCF smaller than the number solved by FCF, but also the LP problems used by WRP-FCF are simpler than those employed by FCF. For mathematical proofs, the reader may refer to [31,34].

Feasibility-based flux coupling analysis

Linear programming can be solved by enumerating a set of adjacent feasible solutions such that at each solution the objective function improves or is unchanged. Accordingly, searching for a feasible solution can be faster than computing an optimal solution [35]. Based on this observation, [31] proposed the FFCA approach for flux coupling analysis transforming the optimization-based LP problems used in the WRP-FCF method into feasibility-based LP problems. The authors compared FFCA with other available flux coupling algorithms, and showed that FFCA is slightly faster than WRP-FCF.

Methods

Before using linear programming to calculate flux coupling between reactions, we preprocess the metabolic network of MetNet to remove the blocks of irreversible reactions and the merging of the pairs of equivalent reactions (1, 2), (8, 9) and (11, 12).

![Exemplary metabolic network MetNet before and after preprocessing](image-url)

Figure 1: Exemplary metabolic network MetNet before and after preprocessing. (a) MetNet consists of eight metabolites (A, ..., H), and thirteen reactions (1, ..., 13), whereof six reactions are irreversible. Metabolites are depicted as nodes while reactions are depicted as arrows. Reversible reactions are indicated by double arrowheads. (b) MetNet after preprocessing where dead-end metabolites and blocked reactions were removed and fully coupled reactions were merged iteratively. This resulted in the removal of the blocked reaction 13 and the merging of the pairs of equivalent reactions (1, 2), (8, 9) and (11, 12).
network in order to i) reduce the number of variables and constraints of the subsequent LP problems and ii) classify reactions according to their reversibility type. The network reduction is mainly based on the removal of trivially blocked reactions and the merging of the stoichiometric columns corresponding to trivially coupled reactions [36-39]. For this, one can use the kernel of the stoichiometric matrix. Alternatively, we can apply the following reduction rules which require a simple parsing of the stoichiometric columns corresponding to trivially coupled reactions [36-39]. For this, one can use the kernel of the stoichiometric matrix and are not time consuming. This strategy allows avoiding potential numerical instabilities related to the computation of a basis of the kernel.

Preprocessing
Certain metabolites, called dead-end metabolites [37], are produced (resp. consumed) by some reactions without being consumed (resp. produced) by other reactions. This concept is illustrated in Figure 1(a) where the dead-end metabolite $H$ is produced by reaction 13 without being consumed by any of the remaining reactions.

As stated below, reactions which are consuming or producing dead-end metabolites are blocked.

**Observation 1** (Dead-End Metabolite). Let $k \in \{1, \ldots, m\}$ be a metabolite. Then, the following holds:

- If there exists a reaction $i$ such that $S_{ki} \neq 0$ and $S_{kj} = 0$ for all $j \neq i$, then reaction $i$ is blocked.

**Table 1** Main steps of the preprocessing procedure

| Step | Rule |
|------|------|
| 1.   | Identify dead-end metabolites and the corresponding blocked reactions. |
| 2.   | Apply Reduction Rule 1 to remove the rows (resp. columns) corresponding to dead-end metabolites (resp. blocked reactions) from the stoichiometric matrix. |
| 3.   | Apply the TFC rule to determine reactions which are proportional to each other and update their reversibility type. |
| 4.   | Apply Reduction Rule 2 to keep only one column for each set of reactions which are proportional to each other. |
| 5.   | Repeat Steps 1-4 until neither a dead-end metabolite nor a pair of fully coupled reactions can be identified. |
| 6.   | Update the reversibility type of each reaction and remove the columns corresponding to blocked reactions from the stoichiometric matrix. |
| 7.   | Use a basis of the kernel of the stoichiometric matrix to identify fully coupled reactions. This step is optional as the kernel computation may lead to numerical problems. |
| 8.   | Classify reactions according to their reversibility type. |

If there exists a set of reactions $I \subseteq \text{Irr}$ such that $S_{ki} > 0$ (resp. $S_{ki} < 0$) for all $i \in I$ and $S_{kj} = 0$ for all $j \notin I$, then all reactions $i \in I$ are blocked.

In each of these cases, $k$ is called a dead-end metabolite.

Certain metabolites are involved in exactly two reactions. For instance, in the network MetNet depicted in Figure 1(a), metabolite $E$ is produced/consumed only by reactions 8 and 9. The following observation states that the fluxes through reactions involving such metabolites are proportional to each other.

**Observation 2** (Trivial Full Coupling (TFC)). Let $i$ and $j$ be two reactions such that, for some metabolite $k \in \{1, \ldots, m\}$, $S_{ki} \neq 0$, $S_{kj} \neq 0$ and $S_{kl} = 0$ for all $l \neq i,j$. Then, reactions $i$ and $j$ are either blocked or fully coupled.

The identification of dead-end metabolites and their corresponding blocked reactions allows us to reduce the number of metabolites and reactions that matter for identifying coupled reactions. As stated in the following observation, the removal of the rows (resp. columns) in the stoichiometric matrix corresponding to dead-end metabolites (resp. blocked reactions) does not influence the flux coupling between reactions.

**Observation 3** (Reduction Rule 1). Let $D$ be a set of dead-end metabolites and let $B$ be a set of blocked reactions. For convenience, suppose $B = \{s + 1, \ldots, n\}$. Let $S'$ be the submatrix of $S$ formed by the rows $S_{ks}$ with $k \notin D$ and the columns $S_{sl}$ with $l \notin B$. Let $\text{Irr}' = \text{Irr} \setminus B$. Then, for all pairs of reactions $i \notin B$ and $j \notin B$,

- If $i \rightarrow j$ and only if $v'_i = 0$ implies $v'_j = 0$ for all $v' \in \mathbb{R}^n$ such that $S'v' = 0$ and $v'_p \geq 0$ for all $p \in \text{Irr}'$.
- If $i \not\rightarrow j$ and only if there exists $\lambda' \neq 0$ such that $v'_j = \lambda' v'_i$ for all $v' \in \mathbb{R}^n$ with $S'v' = 0$ and $v'_p \geq 0$ for all $p \in \text{Irr}'$.  

The next observation shows that two fully coupled reactions can be represented by only one column in the

**Table 2** Transitivity inferred flux (un)coupling

| Known flux coupling | $i \not\rightarrow j$ | $i \not\rightarrow j$ | $i \rightarrow j$ | $j \not\rightarrow k$ |
|---------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| $k \not\rightarrow i$ | $k \not\rightarrow j$ | $k \rightarrow j$ | $k \rightarrow j$ | $j \rightarrow k$ |
| $k \not\rightarrow j$ | $k \not\rightarrow j$ | $k \rightarrow j$ | $k \rightarrow j$ | $j \rightarrow k$ |
| $k \rightarrow j$ | $k \rightarrow j$ | $k \rightarrow j$ | $k \rightarrow j$ | $j \rightarrow k$ |
| $k \rightarrow j$ | $k \rightarrow j$ | $k \rightarrow j$ | $k \rightarrow j$ | $j \rightarrow k$ |
| $k \rightarrow j$ | $k \rightarrow j$ | $k \rightarrow j$ | $k \rightarrow j$ | $j \rightarrow k$ |
| $i \rightarrow j$ | $j \rightarrow k$ | $j \rightarrow k$ | $j \rightarrow k$ | $k \rightarrow j$ |
Irr are equivalent:

1. Apply the preprocessing procedure shown in Table 1.
2. Apply the feasibility rule using the feasible solutions obtained when solving the LP problems used in Step 1.
3. Apply the TDC and TUC rules to determine trivially (un)coupled reactions.
4. Identify fully coupled reversible reactions by solving the LP problems (2). This is not necessary if the kernel of the stoichiometric matrix is used in Step 1.
5. Determine the dimension $t$ of $\text{ker}(S_{\text{rev}})$ and remove $t$ independent fully reversible reactions and their coupled reactions. This step is optional since $t$ is often small.
6. Determine the flux coupling between (pseudo-)irreversible reactions by solving the LP problems (3) and (5).
7. For each LP problem solved in Step 6, use the inference rules in Table 2 in combination with the feasibility rule.

---

Table 3 The main steps of the F2C2 algorithm

| Step | Rule |
|------|------|
| 1.   | Apply the preprocessing procedure shown in Table 1. |
| 2.   | Apply the feasibility rule using the feasible solutions obtained when solving the LP problems used in Step 1. |
| 3.   | Apply the TDC and TUC rules to determine trivially (un)coupled reactions. |
| 4.   | Identify fully coupled reversible reactions by solving the LP problems (2). This is not necessary if the kernel of the stoichiometric matrix is used in Step 1. |
| 5.   | Determine the dimension $t$ of $\text{ker}(S_{\text{rev}})$ and remove $t$ independent fully reversible reactions and their coupled reactions. This step is optional since $t$ is often small. |
| 6.   | Determine the flux coupling between (pseudo-)irreversible reactions by solving the LP problems (3) and (5). |
| 7.   | For each LP problem solved in Step 6, use the inference rules in Table 2 in combination with the feasibility rule. |

---

stoichiometric matrix, without altering the flux coupling between reactions.

**Observation 4 (Reduction Rule 2).** Let $k, l$ be two reactions such that for all $v \in C$, $v_l = \lambda v_k$ for some $\lambda \neq 0$. For convenience, suppose $l = n$ and $\lambda > 0$. Let $S'$ be the $m \times (n - 1)$ matrix defined by $S'_{ij} = S_{ik} - S_{lj}$ for all $p \neq k, l$ and $S'_{ik} = S_{ik} + S_{lj}$. Let $\text{Irr}' = (\text{Irr} \cup \{k\}) \setminus \{l\}$ if $l \in \text{Irr}$, and $\text{Irr}' = \text{Irr}$ otherwise. Then, for all pairs of reactions $i \neq l$ and $j \neq l$,

- $i \sim 0$ if and only if $v_j' = 0$ implies $v_j = 0$, for all $v' \in \mathbb{R}^{n-1}$ such that $S'v' = 0$ and $v_{p}' \geq 0$ for all $p \in \text{Irr}'$.
- $i \sim j$ if and only if there exists $\lambda' \neq 0$ such that $v_j' = \lambda' v_k'$, for all $v' \in \mathbb{R}^{n-1}$ with $S'v' = 0$ and $v_p' \geq 0$ for all $p \in \text{Irr}'$.

Note that when applying the reduction rules of Observations 3 and 4, further metabolites and reactions may fulfill the conditions of Observations 1 and 2. Accordingly, we apply these reduction rules in an iterative way. As an illustration, the reduction of the network MetNet depicted in Figure 1(a) involves two iterations. In the first one, metabolite $H$ and reaction 13 are removed, the pairs of reactions (1, 2) and (8, 9) are merged and metabolites $A$ and $E$ are removed. In the second iteration, the equivalent reactions (11, 12) are merged and metabolite $G$ is removed. The preprocessed network depicted in Figure 1(b) contains only four metabolites and nine reactions.

Certain fully coupled reactions could not be identified using Observation 2. The following lemma proves that all fully coupled reaction pairs can be deduced from the kernel $\text{ker}(S) = \{v \in \mathbb{R}^n \mid Sv = 0\}$ of the stoichiometric matrix after the removal of all blocked reactions.

**Lemma 1.** Let $(S, \text{Irr})$ be a metabolic network with $n$ unblocked reactions. For a pair of reactions $(i, j)$ the following are equivalent:

- $i \sim j$.
- there exists $\lambda \in \mathbb{R} \setminus \{0\}$ such that $v_i = \lambda v_j$, for all $v \in \text{ker}(S)$.

**Proof.** $\Rightarrow$ Immediate.

$\Rightarrow$ Since $i \sim j$, there exists $\lambda \neq 0$ such that $v_i = \lambda v_j$ for all $v \in C$. Assume by contradiction that there is $v \in \text{ker}(S)$ such that $v_i \neq \lambda v_j$ and let $L = \{l \in \text{Irr} \mid v_l < 0\}$. Since every reaction is unblocked, for every $l \in L$ there exists $g_l \in C$ with $g_l(l) = 1$. Let $w = v - \sum_{l \in L} v_l g_l$. Clearly, $w \in \text{ker}(S)$ and $w_i > 0$ for all $l \in \text{Irr}$, thus $w \in C$. However, $w_i \neq \lambda w_j$, contradicting $i \sim j$. The required statement follows.

In analogy with the WRP-FCF and FFCA approaches, we identify the reversibility type of reactions in order to apply linear programming only in those cases where coupling relationships can occur [33]. Here, we use the procedure for reaction classification described in [31,34]. Note that applying the above reduction rules beforehand reduces the number of variables and constraints in the LP problems used for the classification of reactions.

Based on the results above, we propose to apply the preprocessing procedure given in Table 1 before identifying coupled reactions using linear programming. We show

---

Table 4 Genome-scale metabolic networks with the number of metabolites ($\text{met.}$) and reactions ($\text{reac.}$) before and after applying the preprocessing steps

| Network name     | Original size | Prepr. size |
|------------------|---------------|-------------|
| $M$. barkeri, IAF692 | 628 | 149 | 221 |
| $S$. cerevisiae, IN750 | 1061 | 248 | 446 |
| $M$. tuberculosis, IN661 | 826 | 240 | 418 |
| $E$. coli, iJR904 | 761 | 269 | 565 |
| $E$. coli, iAF1260 | 1668 | 604 | 1272 |
| $E$. coli, iJO1366 | 1805 | 651 | 1376 |
| $H$. sapiens, Recon1 | 2766 | 725 | 1573 |

---
later that the preprocessing step turns out to be crucial for obtaining an efficient flux coupling algorithm.

**Algorithmic improvements**

In certain metabolic networks, the conversion of a set of substrates into a set of products can be made by different reactions having the same stoichiometry. A simple example of such reactions are isoenzymes which make the same conversion of substrates into products. This concept is illustrated in Figure 1(a) where both reactions 4 and 5 convert C into D in the same way. This holds also for reaction 7 and the merged equivalent reactions (8, 9) in Figure 1(b) showing that the network preprocessing may simplify the identification of such alternative conversions. The flux coupling of such reactions is trivial using the following lemma.

**Lemma 2** (Trivial Uncoupling (TUC)). Let \( i, j \in \text{Irev and } k, l \in \text{Prev \cup Frev be four reactions. Then, the following holds:} \)

- If \( S_{ai} = \alpha S_{aj} \) for some \( \alpha > 0 \), then \( i \rightarrow p \) and \( j \rightarrow p \) for all \( p \notin \text{Blk} \).
- If \( S_{ai} = \alpha S_{aj} \) for some \( \alpha < 0 \), then \( p \rightarrow i \) (resp. \( p \rightarrow j \)) for all \( p \notin \text{Blk} \cup \{j\} \) (resp. \( p \rightarrow \text{Blk} \cup \{i\} \)).
- If \( S_{ai} = \alpha S_{sk} \) for some \( \alpha \neq 0 \), then \( i \rightarrow p \) and \( p \rightarrow i \) for all \( p \notin \text{Blk} \) and \( q \rightarrow k \) for all \( q \notin \text{Blk} \cup \{i\} \).
- If \( S_{ki} = \alpha S_{sk} \) for some \( \alpha \neq 0 \), then \( k \rightarrow p \) and \( p \rightarrow k \) for all \( p \notin \text{Blk} \cup \{l\} \) and \( l \rightarrow q \) and \( q \rightarrow l \) for all \( q \notin \text{Blk} \cup \{k\} \).

**Proof.** The proofs of the four statements are similar. We only consider the first one. Suppose \( S_{ai} = \alpha S_{aj} \) for some \( \alpha > 0 \) and let us prove \( i \rightarrow p \). Let \( p \notin \text{Blk} \). There exists \( v \in C \) such that \( v_p = 0 \). Let \( w \in \mathbb{R}^n \) such that \( w_i = 0 \), \( w_j = v_i + v_j \) and \( w_q = v_q \) for all \( q \neq i, j \). We have \( w \in C \), \( w_i = 0 \), \( w_p = 0 \) and so the claim follows.

The next observation states that metabolites which are involved only in irreversible reactions and consumed or produced by exactly one reaction define trivial directional couplings between these reactions.

**Observation 5** (Trivial Directional Coupling (TDC)). Let \( k \) be some metabolite such that \( S_{kl} = 0 \) for all \( l \in \text{Frev \cup Prev} \). Let \( P = \{i \mid S_{ki} > 0\} \) and \( N = \{j \mid S_{ij} < 0\} \). If \( \text{card}(P) = 1 \) (resp. \( \text{card}(N) = 1 \)), then \( i \rightarrow 0 \) (resp. \( j \rightarrow 0 \)) for all \( (i, j) \in P \times N \).

Since directional flux coupling is a transitive relation, the flux (un)coupling between many reactions can simply be deduced from dependencies between reactions whose flux (un)coupling has been determined beforehand. This allows us to significantly reduce the total number of LP problems to be solved. Examples of such inferred flux (un)couplings are given in Figure 1(b). According to the TDC rule, we have \((1, 2) \rightarrow (8, 9) \). By solving the LP problems (4), we obtain \( 10 \leftrightarrow (8, 9) \). We can easily conclude that \( 10 \leftrightarrow (1, 2) \).

Table 2 shows the inferred flux (un)coupling relations we can deduce from known relationships between reactions.

For some pairs of reactions, we need to solve at least one LP problem. The optimal solution not only determines the flux coupling between the considered pair of reactions, but also allows one to determine many other uncoupled reactions.

**Observation 6** (Feasibility Rule). Let \( v \in C \) be a steady state flux vector and let \( I = \{i \mid v_i = 0\} \) and \( J = \{j \mid v_j \neq 0\} \). Then \( i \rightarrow j \) for all \( (i, j) \in I \times J \).

In general, most irreversible reactions are uncoupled to each other. Accordingly, the LP problems (4) used to determine coupled irreversible reactions are often unbounded. This limits the use of the feasibility rule, which requires the calculation of a feasible flux vector. In order to optimally use the feasibility rule, instead of solving the LP problems (4) to decide whether two irreversible
Figure 2 Visualization of the LP problems solved using different algorithms. (a) The FFCA algorithm, (b) the FFCA algorithm after applying the preprocessing procedure given in Table 1 without kernel computation (Step 7), (c) the FFCA algorithm after applying the preprocessing procedure and using the kernel of the stoichiometric matrix to identify fully coupled reactions and (d) the F2C2 algorithm given in Table 3. The dashed lines mark the boundary of the \( I_{rev} \), \( Prev \) and \( F_{rev} \) regions. Colors: Black - an LP problem is solved for the corresponding (ordered) pair of reactions; Gray - the corresponding LP problem is not solved due to one (or both) reactions being eliminated from the network (a preprocessing improvement); White - corresponding LP problem does not get solved due to an algorithmic improvement.

Reactions \( i, j \in I_{rev} \) are coupled, we propose to solve the bounded LP problems

\[
\begin{align*}
L_{ij} &= \min \{v_i : Sv = 0, v_j = 1, v_k \geq 0, k \in Irr\}, \\
L_{ji} &= \min \{v_j : Sv = 0, v_i = 1, v_k \geq 0, k \in Irr\},
\end{align*}
\]

and to use the following scheme:

- \( i \rightarrow j \) (resp. \( j \rightarrow i \)) if and only if \( L_{ij} \neq 0 \) (resp. \( L_{ji} \neq 0 \)),
- \( j \sim i \) if and only if \( L_{ij} \neq 0, L_{ji} \neq 0 \) and \( L_{ij} = 1/L_{ji} \).

The following observation states that removing a fully reversible reaction does not alter the flux coupling between (pseudo-)irreversible reactions.

**Observation 7.** Let \( k \in F_{rev} \) be a fully reversible reaction. For convenience, suppose \( k = n \). Let \( S' \) be the \( m \times (n - 1) \) matrix defined by \( S'_{sp} = S_{sp} \) for all \( p \neq k \) and let \( Irr' = Irr \).

Then, for all pairs of reactions \( i \notin F_{rev} \) and \( j \notin F_{rev} \),

- \( i \rightarrow j \) if and only if \( v'_i = 0 \) implies \( v'_j = 0 \), for all \( v' \in \mathbb{R}^{n-1} \) with \( S'v' = 0 \) and \( v'_p \geq 0 \) for all \( p \in Irr' \).
- \( i \sim j \) if and only if there exists \( \lambda' \neq 0 \) such that \( v'_j = \lambda'v'_i \), for all \( v' \in \mathbb{R}^{n-1} \) with \( S'v' = 0 \) and \( v'_p \geq 0 \) for all \( p \in Irr' \).

Let \( S_{aRev} \) be the submatrix defined by the columns in \( S \) corresponding to the reversible reactions and let \( t \) be the dimension of \( \text{ker}(S_{aRev}) \). Based on Observation 7, we can remove up to \( t \) independent fully reversible reactions without altering the flux coupling between (pseudo-)irreversible reactions. Since certain fully reversible reactions may change their reversibility type after the deletion of a fully reversible reaction, we first remove a randomly chosen reaction \( i \in F_{rev} \) together with the coupled reactions with \( i \). We calculate the impact of this deletion on the dimension of \( \text{ker}(S_{aRev}) \). If this dimension decreases by 1, the deletion is maintained; otherwise the removed reactions are restored to the network. This is repeated until \( t \) independent fully reversible reactions and their coupled reactions are removed. We assume that the flux coupling between fully reversible reactions is determined beforehand.
Based on the above results, we propose the Fast Flux Coupling Calculator (F2C2) to determine coupled reactions. The main steps of the F2C2 algorithm are given in Table 3.

**Results and discussion**

The F2C2 algorithm has been implemented within the MATLAB environment, using CLP (via the Mexclp [40] interface) as the underlying linear programming solver. For benchmarking, we analyzed the following genome-scale metabolic networks: *Escherichia coli*, iJR904 [1], *Saccharomyces cerevisiae*, iND750 [2], * Methanosarcina barkeri*, iAF692 [3], *Mycobacterium tuberculosis*, iNJ661 [4], *Escherichia coli*, iAF1260 [5], *Homo sapiens*, Recon1 [6] and *Escherichia coli*, iJO1366 [7]. For the numerically sensitive parts, a tolerance level of 1e-6 was set. All computations were performed using a single Intel Xeon 5160 (3.0 GHz) processor on a 64-bit Debian Linux 6.0 system.

As pointed out in the previous section, part of the performance gain of F2C2 over previous FCA algorithms stems from the fact that the preprocessing steps reduce the network size. This affects the running time on two levels: there are fewer reaction pairs and the LP problems to be solved have reduced size. The dramatic effect of the preprocessing steps on the network size is presented in Table 4.

The algorithmic improvements further reduce the number of LP problems to be solved. A direct performance comparison between the FFCA and F2C2 algorithms (including the running times and number of LP problems solved) is summarized in Table 5. In all cases, F2C2 outperformed FFCA by several orders of magnitude. As pointed out in the previous section, part of the performance gain of F2C2 over previous FCA algorithms stems from the fact that the preprocessing steps reduce the network size. This affects the running time on two levels: there are fewer reaction pairs and the LP problems to be solved have reduced size. The dramatic effect of the preprocessing steps on the network size is presented in Table 4.

The algorithmic improvements further reduce the number of LP problems to be solved. A direct performance comparison between the FFCA and F2C2 algorithms (including the running times and number of LP problems solved) is summarized in Table 5. In all cases, F2C2 outperformed FFCA by several orders of magnitude. In [31] it has been shown that FFCA is more efficient on genome-scale metabolic networks than other flux coupling algorithms.

For an intuitive, visual representation of the individual improvements’ impact on the algorithm's performance, a more in-depth analysis has been performed on the recent metabolic network of *E. coli*, iJO1366. Four different sets of improvements were cumulatively switched on, and the linear programs solved were plotted for each case (Figure 2). To better highlight the relevant differences, the following modifications were applied to the results. First, 249 (out of 2582) reactions identified as blocked were removed from the images. This is a common preprocessing step in most FCA algorithms. Second, the order of reactions was permuted such that the reactions in *Irev*, *Prev* and *Frev* are clustered together. Additionally, in each of these three clusters, the fully coupled reactions were moved towards the end of the segment.

Figure 2(a) plots the LP problems solved in the FFCA algorithm. Applying the simple preprocessing steps without using the kernel (Figure 2(b)), several reactions are found to be fully coupled with others, and as such can be merged together. When Lemma 1 is applied (Figure 2(c)), all fully coupled sets are detected. As a consequence the gray stripes get thicker and the LP problems corresponding to (*Prev*, *Prev*) and (*Frev*, *Frev*) pairs need not be solved anymore. The use of the algorithmic improvements (Figure 2(d)) filters the pairs in (*Irev*, *Irev*) and (*Irev*, *Frev*) blocks, greatly reducing the total number of LP problems solved.

**Conclusions**

We have presented the new flux coupling algorithm F2C2, which outperforms previous methods by orders of magnitude. Flux coupling analysis of genome-scale metabolic networks can now be performed in a routine manner. A software tool F2C2 is freely available for non-commercial use at https://sourceforge.net/projects/f2c2/files/.

**Abbreviations**

F2C2: Fast Flux Coupling Calculator; FCA: Flux Coupling Analysis; FFCA: Feasibility-based Flux Coupling Analysis; FCF: Flux Coupling Finder; LP: Linear Program; TDC: Trivial Directional Coupling; TFC: Trivial Full Coupling; TUC: Trivial Uncoupling.

**Competing interests**

The authors declare that they have no competing interests.

**Authors’ contributions**

The F2C2 algorithm was designed by all authors, based on ideas from AL and LD. The implementation and computational experiments were done by LD. The first draft of the manuscript was written by AL, and revised by LD and AB. The final document was approved by all authors.

**Author details**

1. Department of Bioinformatics, Institute for Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Str. 24-25, D-14476 Potsdam, Germany. 2. Max-Planck Institute for Molecular Plant Physiology, Am Mühlenberg 1, D-14476 Potsdam, Germany. 3. FB Mathematik und Informatik, Freie Universität Berlin, Arnimallee 6, D-14195 Berlin, Germany. 4. DFG-Research Center Matheon, Arnimallee 6, D-14195 Berlin, Germany. 5. Berlin Mathematical School (BMS), Arnimallee 6, D-14195 Berlin, Germany.

Received: 16 December 2011 Accepted: 23 February 2012 Published: 23 April 2012

**References**

1. Reed JL, Vo TD, Schilling CH, Palsson BØ: An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). Genome Biol 2003, 4:R54.
2. Duarte NC, Herrgard MJ, Palsson BØ: Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. Genome Res 2004, 14(7):1296–1309.
3. Feist AM, Scholes JCM, Palsson BØ, Brockman FJ, Ideker T: Modeling methanogenesis with a genome-scale metabolic reconstruction of *Methanosarcinaarkeri*. Mol Syst Biol 2006, 2:2006.0004.
4. Jamshidi N, Palsson BØ: Investigating the metabolic capabilities of *Mycobacterium tuberculosis* H37Rv using the in silico strain iNJ661 and proposing alternative drug targets. BMC Syst Biol 2007, 1:26.
5. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hazenmanikatis V, Palsson BØ: A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol 2007, 3:121.
6. Duarte NC, Becker SA, Jamshidi N, Theile I, Mo ML, Vo TD, Srinvas R, Palsson BØ: Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proc Natl Acad Sci U.S.A 2007, 104(8):1777–1782.
