

METHODOLOGY ARTICLE

Open Access

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

Abdelhalim Larhlmi<sup>1,2\*†</sup>, Laszlo David<sup>3,4,5\*†</sup>, Joachim Selbig<sup>1,2</sup> and Alexander Bockmayr<sup>3,4</sup>

## 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]:

$$C = \{v \in \mathbb{R}^n \mid Sv = 0, v_i \geq 0, \text{ for all } i \in Irr\}, \quad (1)$$

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,  $Irr \subseteq \{1, \dots, n\}$  denotes the set of *irreversible* reactions in the network, and  $Rev = \{1, \dots, n\} \setminus Irr$  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: Larhlmi@mpimp-golm.mpg.de; Laszlo.David@fu-berlin.de

†Equal contributors

<sup>1</sup>Department of Bioinformatics, Institute for Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Str. 24-25, D-14476 Potsdam, Germany

<sup>3</sup>FB Mathematik und Informatik, Freie Universität Berlin, Arnimallee 6, D-14195 Berlin, Germany

Full list of author information is available at the end of the article

[30]. 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  $i \in \{1, \dots, n\}$  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  $\xrightarrow{=0}$ ,  $\leftrightarrow$ ,  $\sim$  and  $\mapsto$  are defined in the following way:*

- $i \xrightarrow{=0} j$  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  $v_j = 0$ .
- $i \sim j$   
if there exists  $\lambda \neq 0$  such that for all  $v \in C$ ,  $v_j = \lambda v_i$ .
- $i \mapsto j$  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  $i \sim j$  (resp.  $i \xrightarrow{=0} j$ ,  $i \xrightarrow{=0} j$ ) holds. Otherwise,  $i$  and  $j$  are uncoupled.

Note that  $i \sim j$  (resp.  $i \xrightarrow{=0} j$ ) is equivalent to  $j \sim i$  (resp.  $j \xrightarrow{=0} i$ ). In addition,  $i \sim j$  implies  $i \leftrightarrow j$ , which in turn is equivalent to  $(i \xrightarrow{=0} j$  and  $j \xrightarrow{=0} i)$ .

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

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

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

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

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  $Blk$ ,  $Irev$ ,  $Prev$  and  $Frev$ . 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  $|Rev|$  (resp.  $2|Rev|$ ). 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  $i, j \in Prev$  or  $i, j \in Frev$ ,  $i \xrightarrow{=0} j, i \xleftarrow{=0} j$  and  $i \sim j$  are equivalent. More importantly, only the stoichiometric constraints determine whether  $i \xrightarrow{=0} j$  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

$$\max\{v_j : Sv = 0, v_i = 0, 0 \leq v_j \leq 1\}, \quad (2)$$

and let  $v^*$  be an optimal solution. According to Proposition 6.13 in [34],  $i \xrightarrow{=0} j$  if and only if  $v_j^* = 0$ .

#### Flux coupling between (pseudo-)irreversible reactions

If  $i \in Irev$  and  $j \in Prev$ , we only have to check whether  $i \xrightarrow{=0} j$ . The other coupling relationships cannot occur. Let  $v^1$  resp.  $v^2$  be optimal solutions of the two LP problems

$$\begin{aligned} \max\{v_j : Sv = 0, v_k \geq 0, k \in Irr, v_i = 0, v_j \leq 1\}, \\ \min\{v_j : Sv = 0, v_k \geq 0, k \in Irr, v_i = 0, v_j \geq -1\}. \end{aligned} \quad (3)$$

Then  $i \xrightarrow{=0} j$  if and only if  $v_j^1 = v_j^2 = 0$ .

#### Flux coupling between irreversible reactions

If  $i, j \in Irev$ , 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

$$\begin{aligned} L_{ij} = \min\{v_i : Sv = 0, v_j = 1, v_k \geq 0, k \in Irr\}, \\ U_{ij} = \max\{v_i : Sv = 0, v_j = 1, v_k \geq 0, k \in Irr\}. \end{aligned} \quad (4)$$

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

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

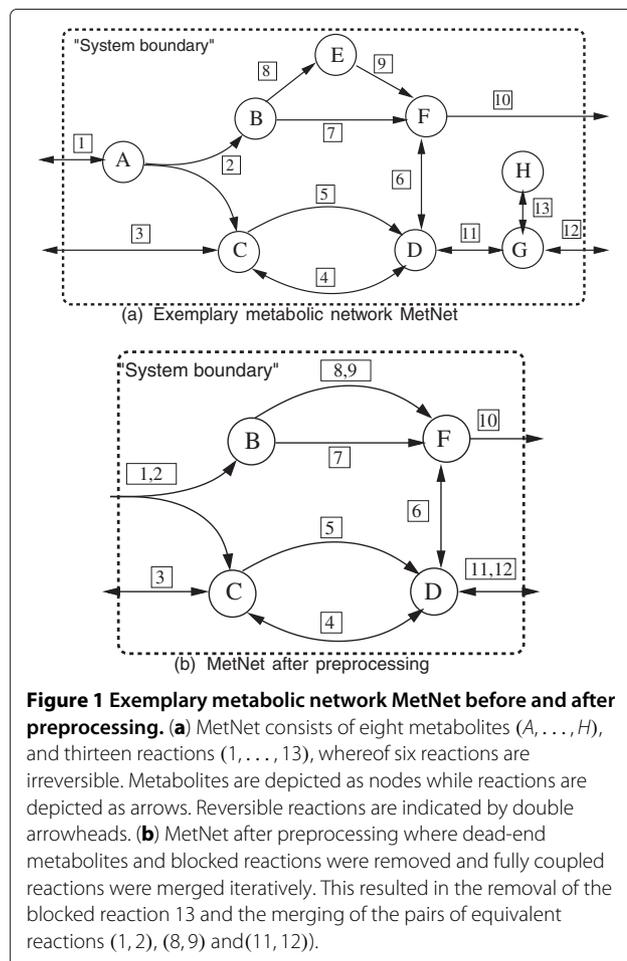
Accordingly, not only is the number of LP problems used by WRP-FCF smaller than the number solved by FCF, but also the LP problems used by WRP-FCF are simpler than those employed by FCF. For mathematical proofs, the reader may refer to [31,34].

### 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



**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).

**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.

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 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  $k \in \{1, \dots, m\}$  be a metabolite. Then, the following hold:*

- 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 2 Transitivity inferred flux (un)coupling**

Known flux coupling	$i \rightsquigarrow j$	$i \overset{=0}{\leftrightarrow} j$	$i \overset{=0}{\rightarrow} j$	$j \overset{=0}{\rightarrow} i$
$k \rightsquigarrow i$	$k \rightsquigarrow j$	$i \overset{=0}{\leftrightarrow} j$	$k \overset{=0}{\rightarrow} j$	$j \overset{=0}{\rightarrow} k$
		$k \rightsquigarrow j$		
$k \overset{=0}{\leftrightarrow} i$	$k \overset{=0}{\leftrightarrow} j$	$k \overset{=0}{\leftrightarrow} j$	$k \overset{=0}{\rightarrow} j$	$j \overset{=0}{\rightarrow} k$
$k \overset{=0}{\rightarrow} i$	$k \overset{=0}{\rightarrow} j$	$k \overset{=0}{\rightarrow} j$	$k \overset{=0}{\rightarrow} j$	
$k \mapsto i$	$k \mapsto j$	$k \mapsto j$		$k \mapsto j$
$i \mapsto k$	$j \mapsto k$	$j \mapsto k$	$j \mapsto k$	

- If there exists a set of reactions  $I \subseteq Irr$  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  $k \in \{1, \dots, m\}$ ,  $S_{ki} \neq 0$ ,  $S_{kj} \neq 0$  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  $B = \{s + 1, \dots, n\}$ . Let  $S'$  be the submatrix of  $S$  formed by the rows  $S_{k*}$  with  $k \notin D$  and the columns  $S_{*l}$  with  $l \notin B$ . Let  $Irr' = Irr \setminus B$ . Then, for all pairs of reactions  $i \notin B$  and  $j \notin B$ ,*

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

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

**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 $kern(S_{*Rev})$ 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  $S'_{*p} = S_{*p}$  for all  $p \neq k, l$  and  $S'_{*k} = S_{*k} + \lambda S_{*l}$ . Let  $Irr' = (Irr \cup \{k\}) \setminus \{l\}$  if  $l \in Irr$ , and  $Irr' = Irr$  otherwise. Then, for all pairs of reactions  $i \neq l$  and  $j \neq l$ ,*

- $i \stackrel{=0}{\rightsquigarrow} j$  if and only if  $v'_i = 0$  implies  $v'_j = 0$ , for all  $v' \in \mathbb{R}^{n-1}$  such that  $S'v' = 0$  and  $v'_p \geq 0$  for all  $p \in Irr'$ .
- $i \rightsquigarrow j$  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  $p \in Irr'$ .

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 Met-Net 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  $kern(S) = \{v \in \mathbb{R}^n \mid Sv = 0\}$  of the stoichiometric matrix after the removal of all blocked reactions.

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

- $i \rightsquigarrow j$ .

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

*Proof.*  $\Leftarrow$  Immediate.

$\Rightarrow$  Since  $i \rightsquigarrow j$ , there exists  $\lambda \neq 0$  such that  $v_i = \lambda v_j$  for all  $v \in C$ . Assume by contradiction that there is  $v \in kern(S)$  such that  $v_i \neq \lambda v_j$  and let  $L = \{l \in Irr \mid v_l < 0\}$ . 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,  $w \in kern(S)$  and  $w_l > 0$  for all  $l \in Irr$ , thus  $w \in C$ . However,  $w_i \neq \lambda w_j$ , contradicting  $i \rightsquigarrow j$ . The required statement follows.  $\square$

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 (#met.) and reactions (#reac.) before and after applying the preprocessing steps**

Network name	Original size		Prepr. size	
	#met.	#reac.	#met.	#reac.
<i>M. barkeri</i> , iAF692	628	690	149	221
<i>S. cerevisiae</i> , iND750	1061	1266	248	446
<i>M. tuberculosis</i> , iNJ661	826	1025	240	418
<i>E. coli</i> , iJR904	761	1075	269	565
<i>E. coli</i> , iAF1260	1668	2382	604	1272
<i>E. coli</i> , iJO1366	1805	2582	651	1376
<i>H. sapiens</i> , Recon1	2766	3742	725	1573

**Table 5 Performance comparison between the FFCA and F2C2 algorithms in terms of the number of LP problems solved (#LPs) and their total running times (TRT)**

Network	FFCA		F2C2		
	#LPs	TRT	#LPs	TRT	Prepr. RT
<i>M. barkeri</i> , iAF692	301975	59 m 40 s	774	7 s	5 s
<i>S. cerevisiae</i> , iND750	472629	1 h 50 m 17s	1280	21 s	15 s
<i>M. tuberculosis</i> , iNJ661	556504	3 h 5 m 36 s	1506	22 s	16 s
<i>E. coli</i> , iJR904	655437	2 h 40 m 33 s	1580	26 s	18 s
<i>E. coli</i> , iAF1260	4256786	4 d 31 m 26 s	3309	2 m 47 s	2 m
<i>E. coli</i> , iJO1366	4877262	4 d 5 h 30 m 46 s	3955	3 m 55 s	2 m 45 s
<i>H. sapiens</i> , Recon1	4566304	4 d 18 h 3 m 37 s	3903	5 m 20 s	4 m 9 s

For the F2C2 algorithm, TRT includes the time (Prepr. RT) required for the preprocessing step. Computation times are given in days (d), hours (h), minutes (m) and seconds (s).

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  $i, j \in Irev$  and  $k, l \in Prev \cup Frev$  be four reactions. Then, the following holds:*

- If  $S_{*i} = \alpha S_{*j}$  for some  $\alpha > 0$ , then  $i \mapsto p$  and  $j \mapsto p$  for all  $p \notin Blk$ .
- If  $S_{*i} = \alpha S_{*j}$  for some  $\alpha < 0$ , then  $p \mapsto i$  (resp.  $p \mapsto j$ ) for all  $p \notin Blk \cup \{j\}$  (resp.  $p \notin Blk \cup \{i\}$ ).
- If  $S_{*i} = \alpha S_{*k}$  for some  $\alpha \neq 0$ , then  $i \mapsto p$  and  $p \mapsto i$  for all  $p \notin Blk$  and  $q \mapsto k$  for all  $q \notin Blk \cup \{i\}$ .
- If  $S_{*k} = \alpha S_{*l}$  for some  $\alpha \neq 0$ , then  $k \mapsto p$  and  $p \mapsto k$  for all  $p \notin Blk \cup \{l\}$  and  $l \mapsto q$  and  $q \mapsto l$  for all  $q \notin Blk \cup \{k\}$ .

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

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  $l \in Frev \cup Prev$ . Let  $P = \{i \mid S_{ki} > 0\}$  and  $N = \{j \mid S_{kj} < 0\}$ . If  $card(P) = 1$  (resp.  $card(N) = 1$ ), then  $i \xrightarrow{=0} j$  (resp.  $j \xrightarrow{=0} i$ ) for all  $(i, j) \in P \times N$ .*

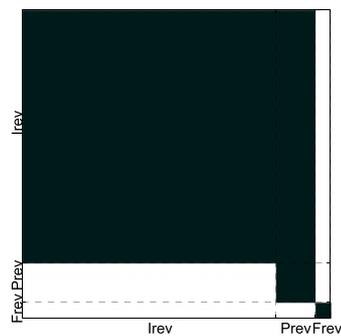
Since directional flux coupling is a transitive relation, the flux (un)coupling between many reactions can simply be deduced from dependencies between reactions whose flux (un)coupling has been determined beforehand. This allows us to significantly reduce the total number of LP problems to be solved. Examples of such inferred flux (un)couplings are given in Figure 1(b). According to the TDC rule, we have  $(1, 2) \xrightarrow{=0} (8, 9)$ . By solving the LP problems (4), we obtain  $10 \mapsto (8, 9)$ . We can easily conclude that  $10 \mapsto (1, 2)$ .

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

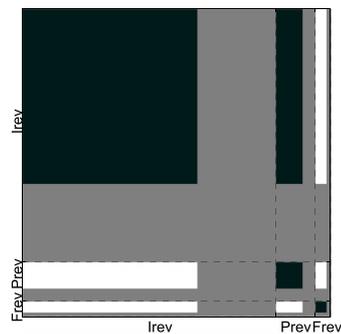
For some pairs of reactions, we need to solve at least one LP problem. The optimal solution not only determines the flux coupling between the considered pair of reactions, but also allows one to determine many other uncoupled reactions.

**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  $i \mapsto j$  for all  $(i, j) \in I \times J$ .*

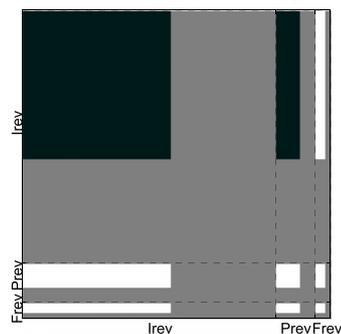
In general, most irreversible reactions are uncoupled to each other. Accordingly, the LP problems (4) used to determine coupled irreversible reactions are often unbounded. This limits the use of the feasibility rule, which requires the calculation of a feasible flux vector. In order to optimally use the feasibility rule, instead of solving the LP problems (4) to decide whether two irreversible



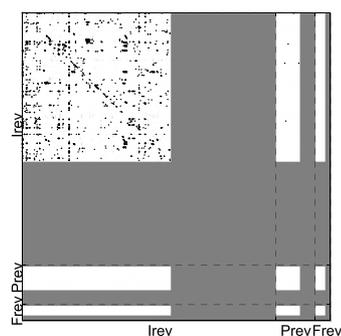
(a) FFCA algorithm



(b) FFCA with the preprocessing without using the kernel



(c) FFCA with the preprocessing using the kernel



(d) F2C2 algorithm

**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 *Irev*, *Prev* and *Frev* 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  $i, j \in Irev$  are coupled, we propose to solve the bounded LP problems

$$\begin{aligned} L_{ij} &= \min\{v_i : Sv = 0, v_j = 1, v_k \geq 0, k \in Irr\}, \\ L_{ji} &= \min\{v_j : Sv = 0, v_i = 1, v_k \geq 0, k \in Irr\}, \end{aligned} \quad (5)$$

and to use the following scheme:

- $i \xrightarrow{=0} j$  (resp.  $j \xrightarrow{=0} i$ ) 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  $k \in Frev$  be a fully reversible reaction. For convenience, suppose  $k = n$ . Let  $S'$  be the  $m \times (n - 1)$  matrix defined by  $S'_{*p} = S_{*p}$  for all  $p \neq k$  and let  $Irr' = Irr$ . Then, for all pairs of reactions  $i \notin Frev$  and  $j \notin Frev$ ,

- $i \xrightarrow{=0} j$  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  $S_{*Rev}$  be the submatrix defined by the columns in  $S$  corresponding to the reversible reactions and let  $t$  be the dimension of  $kern(S_{*Rev})$ . 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  $i \in Frev$  together with the coupled reactions with  $i$ . We calculate the impact of this deletion on the dimension of  $kern(S_{*Rev})$ . 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  $10e-6$  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. 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. Secondly, 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*, *Prev*) 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

<sup>1</sup>Department of Bioinformatics, Institute for Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Str. 24-25, D-14476 Potsdam, Germany. <sup>2</sup>Max-Planck Institute for Molecular Plant Physiology, Am Mühlenberg 1, D-14476 Potsdam, Germany. <sup>3</sup>FB Mathematik und Informatik, Freie Universität Berlin, Arnimallee 6, D-14195 Berlin, Germany. <sup>4</sup>DFG-Research Center Matheon, Arnimallee 6, D-14195 Berlin, Germany. <sup>5</sup>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):1298–1309.
3. Feist AM, Scholten JCM, Palsson BØ, Brockman FJ, Ideker T: **Modeling methanogenesis with a genome-scale metabolic reconstruction of *Methanosarcina barkeri***. *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, Hatzimanikatis 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, Thiele I, Mo ML, Vo TD, Srivas 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**(6):1777–1782.

7. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, Palsson BØ: **A comprehensive genome-scale reconstruction of Escherichia coli metabolism-2011.** *Mol Syst Biol* 2011, **7**:535.
8. Covert MW, Famili I, Palsson BØ: **Identifying constraints that govern cell behavior: a key to converting conceptual to computational models in biology?** *Biotechnol Bioeng* 2003, **84**(7):763–772.
9. Price ND, Reed JL, Palsson BØ: **Genome-scale models of microbial cells: evaluating the consequences of constraints.** *Nat Rev Microbiol* 2004, **2**(11):886–897.
10. Schellenberger J, Que R, Fleming R, Thiele I, Orth J, Feist A, Zielinski D, Bordbar A, Lewis N, Rahmanian S, Kang J, Hyduke D, Palsson B: **Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0.** *Nat Protoc* 2011, **6**:1290–1307.
11. Clarke BL: **Stability of complex reaction networks.** In *Adv. Chem. Phys. Volume 43*. Edited by Prigogine I, Rice SA. Hoboken: John Wiley & Sons, Inc.; 1980:1–216.
12. Edwards JS, Palsson BØ: **Metabolic flux balance analysis and the in silico analysis of Escherichia coli K-12 gene deletions.** *BMC Bioinf* 2000, **1**:1.
13. Segrè D, Vitkup D, Church G: **Analysis of optimality in natural and perturbed metabolic networks.** *Proc Natl Acad Sci USA* 2002, **99**(23):15112–15117.
14. Kauffman KJ, Prakash P, Edwards JS: **Advances in flux balance analysis.** *Curr Opin Biotechnol* 2004, **14**(5):491–496.
15. Shlomi T, Berkman O, Ruppin E: **Regulatory on/off minimization of metabolic flux changes after genetic perturbations.** *Proc Natl Acad Sci USA* 2005, **102**(21):7695–7700.
16. Lee JM, Gianchandani EP, Papin JA: **Flux balance analysis in the era of metabolomics.** *Brief Bioinf* 2006, **7**(2):140–150.
17. Schilling CH, Letscher D, Palsson BØ: **Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective.** *J Theor Biol* 2000, **203**(3):229–248.
18. Schuster S, Hilgetag C, Woods JH, Fell DA: **Reaction routes in biochemical reaction systems: algebraic properties, validated calculation procedure and example from nucleotide metabolism.** *J Math Biol* 2002, **45**(2):153–181.
19. Voss K, Heiner M, Koch I: **Steady state analysis of metabolic pathways using Petri nets.** *In Silico Biol* 2003, **3**(3):367–387.
20. Larhlimi A, Bockmayr A: **A new constraint-based description of the steady-state flux cone of metabolic networks.** *Discrete Appl Math* 2009, **157**:2257–2266.
21. Burgard AP, Nikolaev EV, Schilling CH, Maranas CD: **Flux coupling analysis of genome-scale metabolic network reconstructions.** *Genome Res* 2004, **14**(2):301–312.
22. Notebaart RA, Kensche PR, Huynen MA, Dutilh BE: **Asymmetric relationships between proteins shape genome evolution.** *Genome Biol* 2009, **10**:R19.
23. Pál C, Papp B, Lercher MJ: **Adaptive evolution of bacterial metabolic networks by horizontal gene transfer.** *Nat Genet* 2005, **37**:1372–1375.
24. Yizhak K, Tuller T, Papp B, Ruppin E: **Metabolic modeling of endosymbiont genome reduction on a temporal scale.** *Mol Syst Biol* 2011, **7**:479.
25. Notebaart RA, Teusink B, Siezen RJ, Papp B: **Co-regulation of metabolic genes is better explained by flux coupling than by network distance.** *PLoS Comput Biol* 2008, **4**:e26.
26. Montagud A, Zelezniak A, Navarro E, de Córdoba, P F, Urchueguía JF, Patil KR: **Flux coupling and transcriptional regulation within the metabolic network of the photosynthetic bacterium Synechocystis sp. PCC6803.** *Biotechnol J* 2011, **6**:330–342.
27. Szappanos B, Kovács K, Szamecz B, Honti F, Costanzo M, Baryshnikov A, Gelius-Dietrich G, Lercher M, Jelasky M, Myers C, Andrews B, Boone C, Oliver S, Pál C, Papp B: **An integrated approach to characterize genetic interaction networks in yeast metabolism.** *Nat Genet* 2011, **43**(7):656–662.
28. Suthers PF, Chang YJ, Maranas CD: **Improved computational performance of MFA using elementary metabolite units and flux coupling.** *Metab Eng* 2010, **12**:123–128.
29. Bundy JG, Papp B, Harmston R, Browne RA, Clayson EM, Burton N, Reece RJ, Oliver SG, Brindle KM: **Evaluation of predicted network modules in yeast metabolism using NMR-based metabolite profiling.** *Genome Res* 2007, **17**:510–519.
30. Lee DS, Park J, Kay KA, Christakis NA, Oltvai ZN, Barabasi AL: **The implications of human metabolic network topology for disease comorbidity.** *Proc Natl Acad Sci USA* 2008, **105**(29):9880–9885.
31. David L, Marashi SA, Larhlimi A, Mieth B, Bockmayr A: **FFCA: a feasibility-based method for flux coupling analysis of metabolic networks.** *BMC Bioinf* 2011, **12**:236.
32. Xi Y, Chen YPP, Qian C, Wang F: **Comparative study of computational methods to detect the correlated reaction sets in biochemical networks.** *Brief Bioinf* 2011, **12**(2):132–150.
33. Larhlimi A, Bockmayr A: **A new approach to flux coupling analysis of metabolic networks.** In *Computational Life Sciences II, Second International Symposium (CompLife 2006)*. Cambridge, UK, Volume 4216 of Lecture Notes in Computer Science. 2006:205–215.
34. Larhlimi A: **New concepts and tools in constraint-based analysis of metabolic networks.** *PhD thesis*, Freie Universität Berlin 2008.
35. Schrijver A: *Theory of Linear and Integer Programming*. NY, USA: John Wiley & Sons Inc.; 1986.
36. Pfeiffer T, Sánchez-Valdenebro I, Nuño JC, Montero F, Schuster S: **METATOOL: for studying metabolic networks.** *Bioinformatics* 1999, **15**:251–257.
37. Klamt S, Stelling J, Ginkel M, ED G: **FluxAnalyzer: exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps.** *Bioinformatics* 2003, **19**(2):261–269.
38. Gagneur J, Klamt S: **Computation of elementary modes: a unifying framework and the new binary approach.** *BMC Bioinf* 2004, **5**:175.
39. Urbanczik R, Wagner C: **Functional stoichiometric analysis of metabolic networks.** *Bioinformatics* 2005, **21**(22):4176–4180.
40. Löfberg J: **Mexclp** 2006, [http://control.ee.ethz.ch/johanni/clp.php].

doi:10.1186/10.1186/1471-2105-13-57

Cite this article as: Larhlimi et al.: F2C2: a fast tool for the computation of flux coupling in genome-scale metabolic networks. *BMC Bioinformatics* 2012 **13**:57.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

