Fast thermodynamically constrained flux variability analysis

Arne C. Müller1,2,3,* and Alexander Bockmayr1,4

1Department of Mathematics and Computer Science, Freie Universität Berlin, 14195 Berlin, 2International Max Planck Research School for Computational Biology and Scientific Computing (IMPRS-CBSC), Max Planck Institute for Molecular Genetics, 14195 Berlin, 3Berlin Mathematical School (BMS), 10623 Berlin and 4DFG-Research Center Matheon, 10623 Berlin, Germany

ABSTRACT

Motivation: Flux variability analysis (FVA) is an important tool to further analyse the results obtained by flux balance analysis (FBA) on genome-scale metabolic networks. For many constraint-based models, FVA identifies unboundedness of the optimal flux space. This reveals that optimal flux solutions with net flux through internal biochemical loops are feasible, which violates the second law of thermodynamics. Such unbounded fluxes may be eliminated by extending FVA with thermodynamic constraints.

Results: We present a new algorithm for efficient flux variability (and flux balance) analysis with thermodynamic constraints, suitable for analysing genome-scale metabolic networks. We first show that FBA with thermodynamic constraints is NP-hard. Then we derive a theoretical tractability result, which can be applied to metabolic networks in practice. We use this result to develop a new constraint programming algorithm Fast-tFVA for fast FVA with thermodynamic constraints (tFVA). Computational comparisons with previous methods demonstrate the efficiency of the new method. For tFVA, a speed-up of factor 30–300 is achieved. In an analysis of genome-scale metabolic networks in the BioModels database, we found that in 485 of 716 networks, additional irreversible or fixed reactions could be detected.

Availability and implementation: Fast-tFVA is written in C++ and published under GPL. It uses the open source software SCIP and libSBML. There also exists a Matlab interface for easy integration into Matlab. Fast-tFVA is available from page.mi.fu-berlin.de/arnem/fast-tfva.html.

Contact: arne.mueller@fu-berlin.de; Alexander.Bockmayr@fu-berlin.de
Supplementary information: Supplementary data are available at Bioinformatics online.

Received on June 22, 2012; revised on January 13, 2013; accepted on February 3, 2013

1 INTRODUCTION

Flux Balance Analysis (FBA) is a widely used method to analyse the capabilities of a metabolic network (Durot et al., 2009; Mahadevan and Schilling, 2003; Orth et al., 2010; Price et al., 2004; Schuster et al., 2007; Terzer et al., 2009; Teusink et al., 2009; Varma and Palsson, 1994). The strength of FBA is that it allows predicting growth rates of cells accurately (Edwards et al., 2001). FBA is based on the steady-state assumption, i.e. every metabolite that is produced, must also be consumed at the same rate. Flux through the network is enabled by exchange reactions, such as uptake of nutrients and production of biomass. Usually, next to bounds on the fluxes, stoichiometric constraints are the only constraints given, and optimal production of biomass can be computed by solving a linear program (LP).

Frequently, there is not only one optimal flux distribution that achieves optimal biomass production. FBA, however, computes only one such solution. Elementary flux modes (Schuster and Hilgetag, 1994; Schuster et al., 2000) or extreme pathways (Schilling et al., 2000) are commonly used tools to analyse the whole flux space. But, in practice, the number of elementary flux modes grows exponentially with the number of reactions in the network. Hence, enumeration of elementary flux modes becomes infeasible for genome-scale networks. In addition, we are usually not interested in the raw data of all elementary flux modes (resp. extreme pathways), but only in specific properties (Driouch et al., 2012; Haus et al., 2008; Orman et al., 2012; Schwartz and Kanehisa, 2006). Therefore, sampling methods (Schellenberger and Palsson, 2009), elementary flux patterns (Kaleta et al., 2009) and flux variability analysis (FVA) (Burgard et al., 2001; Mahadevan and Schilling, 2003) have been developed.

FVA determines the maximum and minimum values of all the fluxes that will satisfy the constraints and allow for the same optimal objective value. For example, it is known that FBA is unreliable in predicting the flux value of by-products (Khannapho et al., 2008). In such cases, FVA can be applied to predict the range of possible by-product production rates under maximal biomass production, which can be linked to gene expression data (Bilu et al., 2006). Variations of FVA can also be used to determine blocked or unessential reactions (Burgard et al., 2004).

In FVA, however, the following problem arises: It can happen that the network contains internal cycles, i.e. there exist non-zero steady-state fluxes involving only internal reactions. In most metabolic models, only bounds on the flux value of exchange reactions are given (in addition to the zero bounds for the irreversible reactions). This leads to unbounded fluxes through reactions contained in such internal cycles, which is of course not realistic. To remove this issue, a general form of thermodynamic constraints can be added to the model.

*To whom correspondence should be addressed.
(Beard et al., 2002; Qian and Beard, 2005), extending the standard non-negativity constraints for thermodynamic irreversibility of individual reactions. It has been shown that the flux values computed with thermodynamic constraints are more consistent, for example, with respect to gene expression data (Feist et al., 2007; Fleming et al., 2009; Jol et al., 2012; Price et al., 2006; Schellenberger et al., 2011a). In recent years, there has been increasing interest in thermodynamic constraints, as they can also be used to link metabolite concentrations to flux modes (Henry et al., 2006; Hoppe et al., 2007; Kümmel et al., 2006a, b; Singh et al., 2011).

In this article, we present a new method for efficient thermodynamically constrained FBA and FVA. We will work with the relaxed form of thermodynamic constraints, as it was introduced by Beard et al. (2004). There, thermodynamic constraints are formulated as follows:

\[
SJ = 0 \quad \text{steady-state assumption (1)}
\]
\[
\ell \leq J \leq u \quad \text{flux bounds, e.g. irrev. (2)}
\]
\[
\Delta \mu, J_i < 0 \lor J_i = 0 \quad \forall i \in \mathcal{I} \quad \text{thermodynamic constraint (3)}
\]
\[
\Delta \mu^T = \mu^T S_{\mathcal{I}} \quad \text{potential differences (4)}
\]

Here \(S\) denotes the stoichiometric matrix, \(J\) the flux vector, \(\mathcal{M}\) the set of metabolites, \(\mathcal{R}\) the set of all reactions, \(\mathcal{E}\) the set of exchange reactions, \(\mathcal{I} = \mathcal{R} \setminus \mathcal{E}\) the set of internal reactions and \(\mu\) the chemical potential of each metabolite (Gibbs free energy of formation). The operator \(\cdot^T\) denotes transposition. If the concentrations and equilibrium constants are known (Jankowski et al., 2008; Mavrovouniotis, 1990; Noor et al., 2012a), it is possible to further constrain \(\mu\). However, because this information is often not available, we will work here without additional constraints on \(\mu\).

In a thermodynamically constrained flux balance problem, we maximize a linear objective function \(c^T J\), where \(c \in \mathbb{R}^\mathcal{R}\), with respect to the constraints (1)-(4). In thermodynamically constrained flux variability analysis (tFVA), we analyse the flux variability within the feasible (respectively optimal) solutions of a thermodynamically constrained flux balance problem, i.e. for all \(i \in \mathcal{R}\), we solve the optimization problems \(\max\{\pm J_i|((1)-(4))\}\) (respectively \(\max\{\pm J_i|((1)-(4))\}\)).

To simplify notation, we do not only use reactions as indices (e.g. \(J_i\) to denote flux through reaction \(i\)) but also sets of reactions to index sub-vectors. For example, \(J_{\mathcal{E}}\) denotes the flux vector on the internal reactions only and \(S_{\mathcal{I}}\) contains only the columns corresponding to internal reactions.

**Definition 1** (Thermodynamically Feasible Flux). A flux vector \(J\) is called thermodynamically feasible if there exists a vector \(\mu\) such that (1), (3) and (4) are satisfied. If additionally (2) holds, \(J\) will be called a thermodynamically feasible flux that satisfies bounds \(\ell\) and \(u\).

By multiplying \(\mu^T\) from the left side with \(S_{\mathcal{I}}\), the potential differences for internal reactions are obtained. This is equivalent to the often found formulation \(K \Delta \mu = 0\), where \(K\) is the null-space matrix of \(S_{\mathcal{I}}\). The motivation behind Equation (3) is that normally a chemical reaction carries flux if and only if it reduces Gibbs free energy (Alberty, 2003; Beard et al., 2002; Qian and Beard, 2005). Since many reactions are catalysed by enzymes, however, it can happen that an enzyme is not present, for example because of regulatory control and hence, the corresponding reaction is not possible and does not carry flux, even if there is a negative potential difference. Thus, zero flux in Equation (3) is always allowed as well. Note that other formulations of thermodynamic constraints, for example the one used by Fleming et al. (2012), do not have this property. Hence, the different results have to be applied with care.

Using Boolean variables \(a \in [0,1]^T\), a sufficiently large constant \(M > 0\) and a small constant \(\varepsilon > 0\), this formulation can be translated into a mixed integer linear program (MILP), where \(a_i = 0\) indicates that a forward flux \(J_i > 0\) on reaction \(i\) is not allowed, and \(a_i = 1\) forbids a backward flux \(J_i < 0\). This is done by replacing (3) with the inequalities (5) and (6):

\[
SJ = 0
\]
\[
\ell_i \leq J_i \leq u_i \quad \forall i \in \mathcal{R}
\]
\[
\ell_i(1-a_i) \leq J_i \leq u_i a_i \quad \forall i \in \mathcal{R}
\]
\[
-Ma_i + \varepsilon \leq \Delta \mu_i \leq M(1-a_i) - \varepsilon \quad \forall i \in \mathcal{R}
\]
\[
\Delta \mu^T = \mu^T S_{\mathcal{I}}
\]

This formulation is used in a similar form by Schellenberger et al. (2011a) for the COBRA toolbox and many others (Beard et al., 2004; Cogne et al., 2011; Henry et al., 2007; Hoppe et al., 2007). It can be shown that if \(S\) does not contain any zero-columns, this MILP formulation is equivalent to the original one (Müller, 2012). Although solving MILPs is NP-hard in general, practical tests have shown that current MILP solvers are able to optimize flux with respect to the above formulation also on genome-scale networks like the iAF1260 reconstruction of *Escherichia coli* (Schellenberger et al., 2011a).

Beard et al. (2004) observed that a steady-state flux vector \(J\) [i.e. which satisfies Equation (1)] is thermodynamically feasible if and only if there is no internal cycle contained by \(J\). A sketch of the proof using oriented matroids can be found in (Beard et al., 2004). A detailed version using LP-duality is available in Müller (2012) and was also recently published by Noor et al. (2012b). This result is used by Yang et al. (2005) to detect reactions that are irreversible due to thermodynamic constraints. In particular, one of their methods simply runs tFVA.

This article is organized as follows: In Section 2.1, we show that the thermodynamically constrained flux balance problem is NP-complete, and thus, MILP is an appropriate tool. In Section 2.2, we derive a tractability result and analyse how much this applies to current genome-scale models. These theoretical results
are the backbone of a new algorithm, which is described in Section 3. In Section 4, we apply our new method to study the importance of thermodynamic constraints in the analysis of genome-scale metabolic networks.

2 METHODS

2.1 NP-completeness

In FVA, a series of optimization problems are solved. The flux through each reaction in the network is maximized and minimized. In the following, we will see that already determining whether there exists a positive flux through a given reaction is NP-complete. Thus, the optimization problem is NP-hard, and it is unlikely that an efficient algorithm exists for the general case.

Problem 1 (Thermoflux).

- Instance:
  - Metabolic network \( N = (M, R = I \cup E, S \in \mathbb{Q}^{M \times R}) \)
  - Objective reaction \( r \in R \)

- Question: Does a thermodynamically feasible flux \( J \geq 0 \) with \( J_r > 0 \) exist?

Theorem 1. Thermoflux is NP-complete.

The proof of this result can be found in the Supplementary Material.

2.2 Tractability

It was observed early on by Price et al. (2002) that a thermodynamically infeasible flux can be turned feasible by removing internal cycles. For the following tractability result, it is important to understand when these internal cycles can be removed without changing the flux through the objective reactions. Reactions that are contained in internal cycles and reactions that are flux-forcing are problematic.

Definition 2. Let \( N = (M, R = I \cup E, S \in \mathbb{R}^{M \times R}) \) be a metabolic network. A reaction \( r \in R \) is contained in an internal cycle if there exists a \( J \in \mathbb{R}^I \) with \( S_I J = 0 \) and \( J_r > 0 \).

Given lower and upper flux bounds \( \ell, u \in \mathbb{R}^R \), a reaction \( r \) is called flux-forcing if \( \ell_r > 0 \) or \( u_r < 0 \).

For a linear objective function \( c \in \mathbb{R}^R \), a reaction \( r \) is called objective if \( c_r \neq 0 \).

Theorem 2. Let \( N = (M, R = I \cup E, S \in \mathbb{R}^{M \times R}) \) be a metabolic network with lower and upper bounds \( \ell, u \in \mathbb{R}^R \) and a linear objective function \( c \in \mathbb{R}^R \). Let \( C \subseteq I \) be the set of reactions contained in internal cycles, let \( F \subseteq R \) be the set of reactions that are flux-forcing and let \( O \subseteq R \) be the set of objective reactions.

Given a steady-state flux \( J \) with \( \ell \leq J \leq u \), a thermodynamically feasible flux \( J^* \) with \( c^T J = c^T J^* \) and \( \ell \leq J^* \leq u \) can be computed in polynomial time w.r.t. the size of the network \( N \) if \( C \cap (F \cup O) = \emptyset \).

The key to the proof of Theorem 2 is Algorithm 1, which gives the wanted polynomial-time algorithm. The complete proof can be found in the Supplementary Material.

Algorithm 1: This algorithm computes a thermodynamically feasible flux out of a possibly thermodynamically infeasible flux, if the conditions of Theorem 2 are satisfied. It runs in polynomial time. (\( \mathbb{R} \) denotes a vector where all entries are 1.)

Input: A steady-state flux \( J \)

Repeat

\[
\begin{align*}
I^+: & = \{ i \in I : J_i \geq 0 \} \\
I^-: & = \{ i \in I : J_i \leq 0 \} \\
L: & = \arg \max \{ 1 L_I - 1 L_F : S_I L = 0, \ J_F \leq L_F \leq 0, \ J_F \geq L_F \geq 0 \} \\
\text{until} \ 1 L_I - 1 L_F = 0
\end{align*}
\]

Return \( J \)

It should be noticed that this theorem is similar to the result by Fleming et al. (2012). Although these authors use a different definition of thermodynamic feasibility, they also propose a method that keeps the flux through exchange reactions invariant.

Since we can compute optimal steady-state fluxes in polynomial time by solving an LP, it follows that we can compute an optimal thermodynamically feasible flux in polynomial time, if all of the reactions contained in internal cycles are neither flux-forcing nor contained in the objective function. In practice, however, there will be reactions contained in internal cycles, and we have also to consider this case.

For the following result, we need to quantify how often a reaction is contained in internal cycles. For this, it comes in very handy that internal cycles are simply steady-state flux vectors that do not use exchange reactions. Hence, we can describe the space of all internal cycles by elementary flux modes (Schuster and Hilgetag, 1994; Schuster et al., 2000). The elementary flux modes that do not contain exchange reactions will be called internal circuits or simply circuits. They may also be interpreted in terms of oriented matroid theory, which was introduced into metabolic network analysis by Oliveira et al. (2001) and used by Beard et al. (2004) to describe the internal circuits. The number of internal circuits containing the given reaction will be the quantification measure.

We observed that the number of internal circuits is small for many genome-scale networks. See the Supplementary Materials for more details.

Theorem 3. If the number of internal circuits containing flux-forcing or objective reactions is bounded by a constant, the thermodynamically constrained flux balance (and flux variability) problem can be solved in polynomial time w.r.t. the size of the network.

Proof. Let \( F \) resp. \( O \) denote again the set of flux-forcing resp. objective reactions. Let \( C \) be the set of internal circuits that contain a reaction of \( F \cup O \).

We know that the optimal thermodynamically feasible solution must not contain any internal circuits. Hence, in every internal circuit \( C \in C \), we can find a reaction that does not carry any flux in the direction of the circuit. Any such set of reaction/circuit direction pairs will be called a blocking set \( B \subseteq R \times \{-, +\} \) (cf. Figs 1 and 2). We can enforce such a blocking set by adding sign constraints to the reactions in \( B \), i.e. if \( (r, +) \in B \), we add \( J_r \leq 0 \) to the problem and if \( (r, -) \in B \), we add \( J_r \geq 0 \).

It follows that after we enforced a blocking set, there exist no internal circuits that contain reactions of \( F \cup O \) anymore. Hence, by Theorem 2 we can use linear programming to solve the thermodynamically constrained flux balance problem.
By brute force, we only need to enumerate all blocking sets to find one giving us the optimal solution (see Algorithm 2). For each \( C \in \mathcal{C} \), we have at most \(|C| - 1\) reactions that can be blocked (it does not make sense to block flux-forcing or objective reactions). This way, we only need to enumerate at most \( \prod_{C \in \mathcal{C}} (|C| - 1) \) blocking sets. Since the size of each circuit is bounded by the number of reactions and the number of circuits is bounded by assumption, we obtain a polynomial running time algorithm.

**Algorithm 2: General solving procedure**

- **Input:** objective function \( c \), lower bounds \( lb \), upper bounds \( ub \)
- **maxFlux := \(-\infty\)**
- **for all blocking set \( B \) do**
  - \( \ell := lb, u := ub \)
  - **for each \((\ell, +)\) in B do** \( w := 0 \)
  - **for each \((\ell, -)\) in B do** \( \ell_i := 0 \)
  - \( J = \) solve ordinary FBA with bounds \( \ell, u \)
  - \( \text{maxFlux} = \max(\text{maxFlux}, c^T J) \)
- **end for**
- **return maxFlux**

In case of the *E.coli iAF1260* model, the bound of Theorem 3 tells us that we only need to analyse at most 1680 different blocking sets (for optimization on one reaction without internal flux-forcing reactions).

### 3 IMPLEMENTATION

To implement Algorithm 2 and use the result of Theorem 3, we still have to find a way for enumerating all blocking sets. To do this, we used the *constraint integer programming* (CP) framework Scip (Achterberg, 2009). Scip can not only solve *mixed integer programs* (Wolsey, 1998) but also offers an interface for adding more general and complex constraints. Hence, we implemented a constraint handler and primal heuristic for Scip.

Our algorithm (see Algorithm 3) works directly on the flux variables and does not use artificial Boolean variables. It basically implements a branch-and-bound strategy. As we have seen in Theorem 3, we need not block all internal circuits, but only those that contain objective or flux-forcing reactions.

- With highest priority, the constraint handler branches on circuits that contain objective reactions. If this is achieved, every flux through objective reactions must be sourced by exchange reactions. This usually bounds the flux to realistic values, and we may be able to reject the current blocking set, because a better solution has already been found.
- If no circuits containing objective reactions are left, we branch on circuits containing flux-forcing reactions.
- It never happens that we need to branch on circuits that contain neither objective reactions nor flux-forcing reactions. This is because of the heuristic we also implemented.

The heuristic basically runs Algorithm 1. This means, if the network contains no circuits with objective or flux-forcing reactions, the heuristic will output an optimal flux, and the current branch of the search tree does not need to be analysed further. If the network contains circuits with objective reactions, we usually have a large flux through an internal cycle and thus, a bad dual bound. Although Algorithm 1 may also find an optimal solution in this case, the solver will not know it (the dual bound will be larger than the value of the solution found). To prove optimality, the solver will have to do the branching nonetheless. This is why we do not run the heuristic in this case. The heuristic may also fail if the network does contain circuits with flux-forcing reactions and no circuits with objective reactions. However, we may be lucky, and Algorithm 1 produces a solution that still satisfies the flux bounds. This solution is thus proven to be optimal. Therefore, we implemented a slight modification of Algorithm 3 that additionally runs the heuristic if there exist circuits with flux-forcing reactions, but no circuits with objective reactions.

Thus, the heuristic is run at those nodes of the search-tree that do not contain any circuits with objective reactions. This way, we usually only need to branch on circuits containing objective reactions.
The warm-starting effect can also be used for tFVA. It can easily be transformed into a thermodynamically feasible one. By Theorem 2, it follows that for most reactions, the LP-solution involved in internal cycles, see Figure 3 and Table 1. By this result is not directly applicable. On the other hand, many genome-scale metabolic models only contain few reactions in internal cycles. For the whole (resp. the optimal) flux space, we obtained 716 networks in the given time limit. For the whole (resp. the optimal) flux space, we obtained 829 (45% of all genome-scale networks in the BioModels database) had reactions where FVA cannot predict bounds different from those given in the original FBA problem. These bounds tend to be uninformative, as usually the only small bounds are given on exchange reactions. tFVA, however, will compute bounds for these internal reactions that depend on the bounds on the exchange reactions, in particular the bounds on nutrient uptake. Therefore, the bounds obtained by tFVA for reactions contained in internal cycles will be better than the bounds computed by ordinary FVA by orders of magnitude. Irreversible reactions and reactions with fixed fluxes are a special kind of reactions with improved bounds, which we now investigate in more detail.

As explained in Section 2, our algorithm works best if the number of reactions in internal cycles is small. In addition, as it can be seen in Figure 3, the genome-scale networks in the BioModels database may be divided into two categories: networks with <600 reactions in internal cycles and networks with many more reactions in internal cycles. These bounds tend to be uninformative, as usually the only small bounds are given on exchange reactions. tFVA, however, will compute bounds for these internal reactions that depend on the bounds on the exchange reactions, in particular the bounds on nutrient uptake. Therefore, the bounds obtained by tFVA for reactions contained in internal cycles will be better than the bounds computed by ordinary FVA by orders of magnitude. Irreversible reactions and reactions with fixed fluxes are a special kind of reactions with improved bounds, which we now investigate in more detail.

As explained in Section 2, our algorithm works best if the number of reactions in internal cycles is small. In addition, as it can be seen in Figure 3, the genome-scale networks in the BioModels database may be divided into two categories: networks with <600 reactions in internal cycles and networks with many more reactions in internal cycles. For each network, we ran FVA and tFVA on the whole flux space. If the model also contained an objective function on some reaction bounds computed by ordinary FVA by orders of magnitude. Irreversible reactions and reactions with fixed fluxes are a special kind of reactions with improved bounds, which we now investigate in more detail.

As explained in Section 2, our algorithm works best if the number of reactions in internal cycles is small. In addition, as it can be seen in Figure 3, the genome-scale networks in the BioModels database may be divided into two categories: networks with <600 reactions in internal cycles and networks with many more reactions in internal cycles. For each network, we ran FVA and tFVA on the whole flux space. If the model also contained an objective function on some reaction bounds computed by ordinary FVA by orders of magnitude. Irreversible reactions and reactions with fixed fluxes are a special kind of reactions with improved bounds, which we now investigate in more detail.

As explained in Section 2, our algorithm works best if the number of reactions in internal cycles is small. In addition, as it can be seen in Figure 3, the genome-scale networks in the BioModels database may be divided into two categories: networks with <600 reactions in internal cycles and networks with many more reactions in internal cycles. For each network, we ran FVA and tFVA on the whole flux space. If the model also contained an objective function on some reaction bounds computed by ordinary FVA by orders of magnitude. Irreversible reactions and reactions with fixed fluxes are a special kind of reactions with improved bounds, which we now investigate in more detail.

As explained in Section 2, our algorithm works best if the number of reactions in internal cycles is small. In addition, as it can be seen in Figure 3, the genome-scale networks in the BioModels database may be divided into two categories: networks with <600 reactions in internal cycles and networks with many more reactions in internal cycles. For each network, we ran FVA and tFVA on the whole flux space. If the model also contained an objective function on some reaction bounds computed by ordinary FVA by orders of magnitude. Irreversible reactions and reactions with fixed fluxes are a special kind of reactions with improved bounds, which we now investigate in more detail.

As explained in Section 2, our algorithm works best if the number of reactions in internal cycles is small. In addition, as it can be seen in Figure 3, the genome-scale networks in the BioModels database may be divided into two categories: networks with <600 reactions in internal cycles and networks with many more reactions in internal cycles. For each network, we ran FVA and tFVA on the whole flux space. If the model also contained an objective function on some reaction bounds computed by ordinary FVA by orders of magnitude. Irreversible reactions and reactions with fixed fluxes are a special kind of reactions with improved bounds, which we now investigate in more detail.

As explained in Section 2, our algorithm works best if the number of reactions in internal cycles is small. In addition, as it can be seen in Figure 3, the genome-scale networks in the BioModels database may be divided into two categories: networks with <600 reactions in internal cycles and networks with many more reactions in internal cycles. For each network, we ran FVA and tFVA on the whole flux space. If the model also contained an objective function on some reaction bounds computed by ordinary FVA by orders of magnitude. Irreversible reactions and reactions with fixed fluxes are a special kind of reactions with improved bounds, which we now investigate in more detail.

As explained in Section 2, our algorithm works best if the number of reactions in internal cycles is small. In addition, as it can be seen in Figure 3, the genome-scale networks in the BioModels database may be divided into two categories: networks with <600 reactions in internal cycles and networks with many more reactions in internal cycles. For each network, we ran FVA and tFVA on the whole flux space. If the model also contained an objective function on some reaction bounds computed by ordinary FVA by orders of magnitude. Irreversible reactions and reactions with fixed fluxes are a special kind of reactions with improved bounds, which we now investigate in more detail.
In 481 (resp. 485) networks, at least one additional reaction was detected to be irreversible due to thermodynamic constraints.

In Figure 4, we provide further information on the number of additionally fixed or irreversible reactions when analysing the whole flux space. The plot looks similar for the optimal flux space. Additional details can be found in the Supplementary Material.

Finally, we ran the same analysis for the networks of the BiGG database and obtained that in all of them, except Staphylococcus aureus iSB619, additional fixed and irreversible reactions were detected, see Table 2. In the optimal flux space, also in S. aureus iSB619, additional fixed and irreversible reactions were found. These results show that thermodynamic constraints provide useful information even for well-curated models like those in the BiGG database. Using our algorithm, tFVA can now be performed in a routine manner.

### Table 1. Running times of FVA on different metabolic networks

<table>
<thead>
<tr>
<th>Model</th>
<th>Method</th>
<th>Internal circuits</th>
<th>Reactions (in internal circuits)</th>
<th>Whole flux space</th>
<th>Only optimal fluxes w.r.t. biomass production</th>
</tr>
</thead>
<tbody>
<tr>
<td>Escherichia coli iJR904</td>
<td>CP COBRA</td>
<td>19</td>
<td>1075 (40)</td>
<td>11 s 963 s</td>
<td>11 s Error</td>
</tr>
<tr>
<td>Escherichia coli iAF1260</td>
<td>CP COBRA</td>
<td>38</td>
<td>2382 (68)</td>
<td>71 s 11162 s</td>
<td>51 s 15 049 s</td>
</tr>
<tr>
<td>Escherichia coli iJO1366</td>
<td>CP COBRA</td>
<td>46</td>
<td>2583 (76)</td>
<td>92 s 25 869 s</td>
<td>79 s Error</td>
</tr>
<tr>
<td>Helicobacter pylori iT341</td>
<td>CP COBRA</td>
<td>8</td>
<td>554 (22)</td>
<td>4 s 122 s</td>
<td>4 s 176 s</td>
</tr>
<tr>
<td>Homo sapiens Recon. 1</td>
<td>CP COBRA</td>
<td>?</td>
<td>3742 (958)</td>
<td>&gt;2 h Error</td>
<td>&gt;2 h Error</td>
</tr>
<tr>
<td>Methanosarcina barkeri iAF692</td>
<td>CP COBRA</td>
<td>31</td>
<td>690 (30)</td>
<td>7 s 204 s</td>
<td>7 s 332 s</td>
</tr>
<tr>
<td>Mycobacterium tuberculosis iNJ661</td>
<td>CP COBRA</td>
<td>66</td>
<td>1025 (53)</td>
<td>22 s 844 s</td>
<td>19 s 1316 s</td>
</tr>
<tr>
<td>Staphylococcus aureus iSB619</td>
<td>CP COBRA</td>
<td>3</td>
<td>743 (7)</td>
<td>3 s 179 s</td>
<td>2 s 198 s</td>
</tr>
<tr>
<td>Saccharomyces cerevisiae iND750</td>
<td>CP COBRA</td>
<td>43</td>
<td>1266 (64)</td>
<td>44 s 162 s</td>
<td>34 s Error</td>
</tr>
</tbody>
</table>

FVA was performed on the whole flux space and on the subspace of the fluxes with optimal biomass production rate. The COBRA implementation used the Gurobi solver. The COBRA method sometimes failed to produce a solution, because Gurobi returned infeasible for problems that are actually feasible.

### Table 2. Number of additionally fixed and irreversible reactions due to thermodynamic constraints

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli iJR904</td>
<td>11</td>
<td>30</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>E. coli iAF1260</td>
<td>9</td>
<td>41</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>E. coli iJO1366</td>
<td>13</td>
<td>44</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>H. pylori iT341</td>
<td>6</td>
<td>16</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>M. barkeri iAF692</td>
<td>1</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>M. tuberculosis iNJ661</td>
<td>6</td>
<td>13</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td>S. aureus iSB619</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>S. cerevisiae iND750</td>
<td>5</td>
<td>15</td>
<td>13</td>
<td>19</td>
</tr>
</tbody>
</table>

*Opt. fixed and Opt. irrev. denote the number of additionally fixed, resp. irreversible, reactions in the space of fluxes with optimal biomass production.*

In 481 (resp. 485) networks, at least one additional reaction was detected to be irreversible due to thermodynamic constraints.

In Figure 4, we provide further information on the number of additionally fixed or irreversible reactions when analysing the whole flux space. The plot looks similar for the optimal flux space. Additional details can be found in the Supplementary Material.

Finally, we ran the same analysis for the networks of the BiGG database and obtained that in all of them, except Staphylococcus aureus iSB619, additional fixed and irreversible reactions were detected, see Table 2. In the optimal flux space, also in S. aureus iSB619, additional fixed and irreversible reactions were found.

### 5 CONCLUSION

We presented a new algorithm for thermodynamically constrained FVA. Compared with previous implementations, an enormous speed-up was obtained. tFVA can now be performed for a large number of networks and as a subroutine for other methods. As shown in our experiments, many network reconstructions, in particular also well-curated models, may be improved by executing our algorithm.
Our current algorithm works without additional information on metabolic concentrations. However, in more and more cases, such information becomes available. In future work, we plan to extend our method by incorporating metabolic concentrations and other experimental data (Kühl et al., 2006a; Pei et al., 2011; Shinar and Feinberg, 2010).

Funding: This work was supported by Berlin Mathematical School (in terms of a PhD stipend).

Conflicts of Interest: none declared.

REFERENCES


