Fast thermodynamically constrained flux variability analysis

Arne C. Müller[1,2,3,*] and Alexander Bockmayr[1,4]

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

Associate Editor: Mario Albrecht

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:

$$SJ = 0 \qquad \text{steady-state assumption} \qquad (1)$$

$$\ell \leq J \leq u \qquad \text{flux bounds, e.g. irrev.} \qquad (2)$$

$$\Delta\mu_i J_i < 0 \vee J_i = 0 \ \forall i \in \mathcal{I} \quad \text{thermodynamic constraint} \qquad (3)$$

$$\Delta\mu^T = \mu^T S_{\mathcal{I}} \qquad \text{potential differences} \qquad (4)$$

$$J \in \mathbb{R}^{\mathcal{R}}, \ \mu \in \mathbb{R}^{\mathcal{M}}, \ \Delta\mu \in \mathbb{R}^{\mathcal{I}}$$

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 μ 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 μ. However, because this information is often not available, we will work here without additional constraints on μ.

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 $\max\{\pm J_i | (1)-(4), c^T J = \text{opt}\}$).

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, $J_{\mathcal{I}}$ 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 μ such that (1), (3) and (4) are satisfied. If additionally (2) holds, J will be called a* thermodynamically feasible flux that satisfies bounds ℓ and u.

By multiplying μ^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 $a \in \{0, 1\}^{\mathcal{I}}$, 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):

$$SJ = 0$$

$$\ell_i \leq J_i \leq u_i \ \forall i \in \mathcal{R}$$

$$\ell_i(1 - a_i) \leq J_i \leq u_i a_i \ \forall i \in \mathcal{I} \qquad (5)$$

$$-Ma_i + \varepsilon \leq \Delta\mu_i \leq M(1 - a_i) - \varepsilon \ \forall i \in \mathcal{I} \qquad (6)$$

$$\Delta\mu^T = \mu^T S_{\mathcal{I}}$$

$$J \in \mathbb{R}^{\mathcal{R}}, \ \mu \in \mathbb{R}^{\mathcal{M}}, \ \Delta\mu \in \mathbb{R}^{\mathcal{I}}, \ a \in \{0, 1\}^{\mathcal{I}}$$

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* $\mathcal{N} = \left(\mathcal{M}, \ \mathcal{R} = \mathcal{I} \dot{\cup} \mathcal{E}, \ S \in \mathbb{Q}^{M \times R}\right)$
  - *Objective reaction* $r \in \mathcal{R}$

- *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* $\mathcal{N} = \left(\mathcal{M}, \ \mathcal{R} = \mathcal{I} \dot{\cup} \mathcal{E}, \ S \in \mathbb{R}^{M \times R}\right)$ *be a metabolic network. A reaction* $r \in \mathcal{R}$ *is* contained in an internal cycle *if there exists a* $J \in \mathbb{R}^{\mathcal{I}}$ *with* $S_{\mathcal{I}} J = 0$ *and* $J_r > 0$.

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

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

THEOREM 2. *Let* $\mathcal{N} = \left(\mathcal{M}, \ \mathcal{R} = \mathcal{I} \dot{\cup} \mathcal{E}, \ S \in \mathbb{R}^{M \times R}\right)$ *be a metabolic network with lower and upper flux bounds* $\ell, u \in \mathbb{R}^{\mathcal{R}}$ *and a linear objective function* $c \in \mathbb{R}^{\mathcal{R}}$. *Let* $C \subseteq \mathcal{I}$ *be the set of reactions contained in internal cycles, let* $F \subseteq \mathcal{R}$ *be the set of reactions that are flux-forcing and let* $O \subseteq \mathcal{R}$ *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. size of the network* $\mathcal{N}$ *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. ($\mathbb{1}$ denotes a vector where all entries are 1.)

Input: A steady-state flux J
repeat
  $I^+ := \{i \in \mathcal{I} : J_i \geq 0\}$
  $I^- := \{i \in \mathcal{I} : J_i \leq 0\}$
  $L := \arg\max \left\{ \begin{matrix} \mathbb{1}L_{I^+} - \mathbb{1}L_{I^-} : & S_{\mathcal{I}}L = 0, \\ & J_{I^-} \leq L_{I^-} \leq 0, \\ J_{\mathcal{I}} := J_{\mathcal{I}} - L & J_{I^+} \geq L_{I^+} \geq 0 \end{matrix} \right\}$
until $\mathbb{1}L_{I^+} - \mathbb{1}L_{I^-} = 0$
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 $\mathcal{C}$ 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 $C \in \mathcal{C}$, 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* $B \subset \mathcal{R} \times \{-, +\}$ (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.
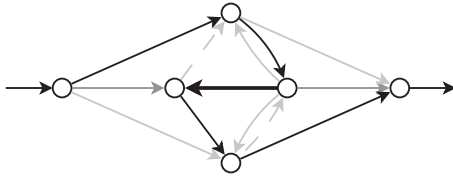
Fig. 1. Blocking set (dashed arrows) for a given thermodynamically feasible flux (marked in black). Reversible reactions are drawn using two arrows. The bold arrow indicates the objective reaction
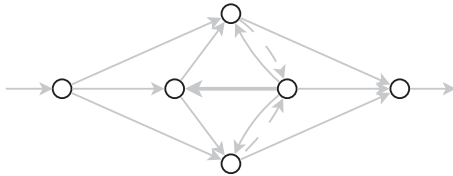

Fig. 2. If a 'wrong' blocking set is chosen, only suboptimal flux (here: no flux) may be possible

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 $(i, +) \in B$ **do** $u_i := 0$
  for each $(i, -) \in B$ **do** $\ell_i := 0$
  $J = $ solve ordinary FBA with bounds ℓ, u
  maxFlux $= \max($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.

Algorithm 3: Sketch of the constraint handler that enforces thermodynamic feasibility

Input: objective function c, flux bounds ℓ, u
Run ordinary FBA to compute a steady-state flux J.
Find internal cycle L containing an objective reaction / a flux-forcing reaction by maximizing flux through the respective reactions w.r.t. $S_{\mathcal{I}} L = 0, 1 \geq \text{sign}(J_i) L_i \geq 0 \;\; \forall i \in \mathcal{I}$.
if no non-zero internal cycle L is found **then** run heuristic (Algorithm 1)
else
  for $i : L_i > 0$ **do**
    if $\ell_i \leq 0$ **then** create child node with $u_i = 0$.
  end for
  for $i : L_i < 0$ **do**
    if $u_i \geq 0$ **then** create child node with $\ell_i = 0$.
  end for
  Add created child nodes to branch and bound tree and continue solving nodes.
end if

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
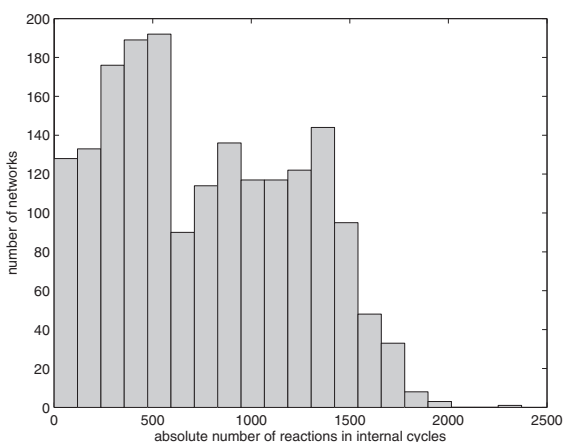
Fig. 3. The genome-scale networks available in the BioModels database seem to separate into two classes: Networks with few (<600) reactions in internal cycles and networks with many more reactions in internal cycles

reactions, sometimes on circuits containing flux-forcing reactions and never on circuits containing none of the two.

Next to the actual solving routine, the result of Theorem 2 can also be used to speed-up tFVA significantly. Gudmundsson and Thiele (2010) observed that for ordinary FVA, a speed-up of factor 100 could be achieved by simply warm starting the LP computation necessary for FBA. Previous implementations always created new instances of the LP problem that the LP-solver had to solve from scratch. In warm starting, the LP-solution of the previous iteration is used as the starting point of the simplex algorithm used by the LP-solver. In the case of tFVA, however, we do not solve LPs anymore, hence this result is not directly applicable. On the other hand, many genome-scale metabolic models only contain few reactions involved in internal cycles, see Figure 3 and Table 1. By Theorem 2, it follows that for most reactions, the LP-solution can easily be transformed into a thermodynamically feasible one without having to start any MILP- or CP-solver. Thus, the warm-starting effect can also be used for tFVA.

4 DISCUSSION

4.1 Run time

We tested the correctness and run time of our implementation on various networks of the BiGG database (Schellenberger *et al.*, 2011b). For tFVA, we compared our results with those generated by the COBRA toolbox and obtained a difference of the order 10^{-3}. This difference is likely due to numerical issues of the big-M formulation (Bradley *et al.*, 1977), as the results of our code for different orderings of the reactions and metabolites only vary in the order of 10^{-8}. The *E.coli iAF1260* model was used for a more detailed analysis. We compared the CP approach with two different kinds of MILP formulations found in the literature; we did not compare it with non-linear formulations, as most non-linear solvers do not return globally optimal solutions. All computations were run on an Intel Core i5-2400S (2.5 GHz, 4 cores, 6 MiB L3 cache) with 4GB RAM. Both MILP formulations were solved using Gurobi (www.gurobi.com); the CP formulation was solved using Scip together with the public-domain

solver Soplex (Achterberg, 2009). As can be seen in Figure 3 of the Supplementary Material, the CP approach outperforms the MILP formulations in nearly all cases.

We also performed run time tests for tFVA and compared our algorithm with the one implemented in the COBRA toolbox (Schellenberger *et al.*, 2011a). The CP algorithm is 30–300 times faster than the COBRA implementation (see Table 1). The only trouble maker is the *Homo sapiens* reconstruction that did not finish within 2 h in our implementation. We were also not able to confirm the run time result by Schellenberger *et al.* (2011a) on that network, because COBRA (version 2.0.3) with Gurobi 5.0.0 terminated with an error message, probably due to numerical instabilities of the MILP formulation in the Gurobi solver.

4.2 Irreversible and fixed flux rates due to thermodynamics

In a second study, we analysed how much information on irreversibility and fixed flux rates of reactions can be gained by adding thermodynamic constraints. Fixed reactions are sometimes the focus of network analysis, as in Hädicke *et al.* (2011). We used the genome-scale networks in the BioModels database (http://www.ebi.ac.uk/biomodels-main, available on September 17, 2012). Nearly all of these networks contain reactions in internal cycles (see Fig. 3). Hence, nearly every network of the BioModels database has 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 significantly more reactions. In what follows, we analysed those networks with <600 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 (e.g. biomass production), we also analysed the optimal flux space with FVA and tFVA. We ran the computations of FVA and tFVA for at most 30 min.

We obtained that 829 (45% of all genome-scale networks in the BioModels database) had ≤600 reactions in internal cycles. Of these, we were able to analyse 716 networks in the given time limit. For the whole (resp. the optimal) flux space, we obtained the following results:

- In 386 (resp. 387) networks, flux through at least one reaction was additionally fixed due to thermodynamic constraints.

Table 1. Running times of FVA on different metabolic networks

Model	Method	Internal circuits	Reactions (in internal circuits)	Whole flux space	Only optimal fluxes w.r.t. biomass production
Escherichia coli iJR904	CP COBRA	19	1075 (40)	11 s 963 s	11 s Error
Escherichia coli iAF1260	CP COBRA	38	2382 (68)	71 s 11 162 s	51 s 15 049 s
Escherichia coli iJO1366	CP COBRA	46	2583 (76)	92 s 25 869 s	79 s Error
Helicobacter pylori iIT341	CP COBRA	8	554 (22)	4 s 122 s	4 s 176 s
Homo sapiens Recon. 1	CP COBRA	?	3742 (958)	>2 h Error	>2 h Error
Methanosarcina barkeri iAF692	CP COBRA	31	690 (30)	7 s 204 s	7 s 332 s
Mycobacterium tuberculosis iNJ661	CP COBRA	66	1025 (53)	22 s 844 s	19 s 1316 s
Staphylococcus aureus iSB619	CP COBRA	3	743 (7)	3 s 179 s	2 s 198 s
Saccharomyces cerevisiae iND750	CP COBRA	43	1266 (64)	44 s 1662 s	34 s Error

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.
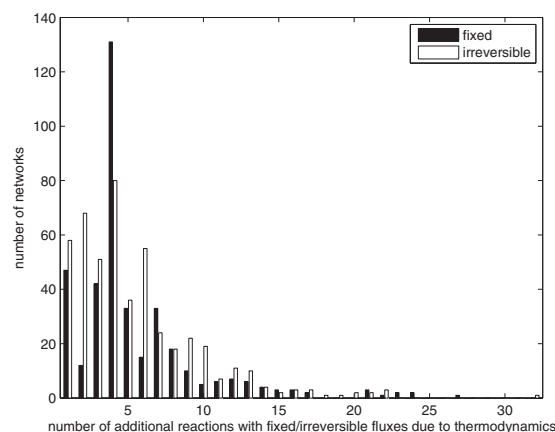
Fig. 4. Distribution of networks with additional fixed (resp. irreversible) reactions

- 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.

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

Model	Fixed	Opt. fixed	Irrev.	Opt. Irrev.
E. coli iJR904	11	30	6	5
E. coli iAF1260	9	41	8	6
E. coli iJO1366	13	44	8	7
H. pylori iIT341	6	16	3	4
M. barkeri iAF692	1	7	8	9
M. tuberculosis iNJ661	6	13	8	13
S. aureus iSB619	0	5	0	1
S. cerevisiae iND750	5	15	13	19

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

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.

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 metabolite concentrations. However, in more and more cases, such information becomes available. In future work, we plan to extend our method by incorporating metabolite concentrations and other experimental data (Kümmel *et al.*, 2006a; Pey *et al.*, 2011; Shinar and Feinberg, 2010).

Conflicts of Interest: none declared.

REFERENCES

Achterberg,T. (2009) SCIP: Solving constraint integer programs. *Math. Program. Comput.*, **1**, 1–41.

Alberty,R.A. (2003) *Thermodynamics of Biochemcal Reactions*. Massachusetts Institute of Technology, Cambridge, MA.

Beard,D.A. *et al.* (2002) Energy balance for analysis of complex metabolic networks. *Biophys. J.*, **83**, 79–86.

Beard,D.A. *et al.* (2004) Thermodynamic constraints for biochemical networks. *J. Theor. Biol.*, **228**, 327–333.

Bilu,Y. *et al.* (2006) Conservation of expression and sequence of metabolic genes is reflected by activity across metabolic states. *PLoS Comput. Biol.*, **2**, e106.

Bradley,S.P. *et al.* (1977) *Applied Mathematical Programming*. Addison-Wesley Publishing Company, Reading, MA, pp. 46, 186, chap 2.2, 5.3.

Burgard,A.P. *et al.* (2001) Minimal reaction sets for *Escherichia coli* metabolism under different growth requirements and uptake environments. *Biotechnol. Prog.*, **17**, 791–797.

Burgard,A.P. *et al.* (2004) Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res.*, **14**, 301–312.

Cogne,G. *et al.* (2011) A model-based method for investigating bioenergetic processes in autotrophically growing eukaryotic microalgae: application to the green algae *Chlamydomonas reinhardtii. Biotechnol. Prog.*, **27**, 631–640.

Driouch,H. *et al.* (2012) Integration of in vivo and in silico metabolic fluxes for improvement of recombinant protein production. *Metab. Eng.*, **14**, 4758.

Durot,M. *et al.* (2009) Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS Microbiol. Rev.*, **33**, 164–190.

Edwards,J.S. *et al.* (2001) In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat. Biotechnol.*, **19**, 125–130.

Feist,A.M. *et al.* (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.*, **3**, 121.

Fleming,R. *et al.* (2009) Quantitative assignment of reaction directionality in constraint-based models of metabolism: application to *Escherichia coli. Biophys. Chem.*, **145**, 47–56.

Fleming,R.M. *et al.* (2012) A variational principle for computing nonequilibrium fluxes and potentials in genome-scale biochemical networks. *J. Theor. Biol.*, **292**, 71–77.

Gudmundsson,S. and Thiele,I. (2010) Computationally efficient flux variability analysis. *BMC Bioinformatics*, **11**, 489.

Hädicke,O. *et al.* (2011) Metabolic network modeling of redox balancing and biohydrogen production in purple nonsulfur bacteria. *BMC Syst. Biol.*, **5**, 150.

Haus,U.-U. *et al.* (2008) Computing knock-out strategies in metabolic networks. *J Comput. Biol.*, **15**, 259–268.

Henry,C. *et al.* (2006) Genome-scale thermodynamic analysis of *Escherichia coli* metabolism. *Biophys. J.*, **90**, 1453–1461.

Henry,C.S. *et al.* (2007) Thermodynamics-based metabolic flux analysis. *Biophys. J.*, **92**, 1792–1805.

Hoppe,A. *et al.* (2007) Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks. *BMC Syst. Biol.*, **1**, 23.

Jankowski,M.D. *et al.* (2008) Group contribution method for thermodynamic analysis of complex metabolic networks. *Biophys. J.*, **95**, 1487–1499.

Jol,S.J. *et al.* (2012) System-level insights into yeast metabolism by thermodynamic analysis of elementary flux modes. *PLoS Comput. Biol.*, **8**, 3.

Kaleta,C. *et al.* (2009) Can the whole be less than the sum of its parts? Pathway analysis in genome-scale metabolic networks using elementary flux patterns. *Genome Res.*, **19**, 1872–1883.

Khannapho,C. *et al.* (2008) Selection of objective function in genome scale flux balance analysis for process feed development in antibiotic production. *Metab. Eng.*, **10**, 227–233.

Kümmel,A. *et al.* (2006a) Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolite data. *Mol. Syst. Biol.*, **2**, 2006.0034.

Kümmel,A. *et al.* (2006b) Systematic assignment of thermodynamic constraints in metabolic network models. *BMC Bioinformatics*, **7**, 512.

Mahadevan,R. and Schilling,C. (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.*, **5**, 264–276.

Mavrovouniotis,M.L. (1990) Group contributions for estimating standard Gibbs energies of formation of biochemical compounds in aqueous solution. *Biotechnol. Bioeng.*, **36**, 1070–1082.

Müller,A. (2012) *Thermodynamic Constraints in Metabolic Networks*. Master's Thesis. Fachbereich Mathematik und Informatik, Freie Universität Berlin, Berlin.

Noor,E. *et al.* (2012a) An integrated open framework for thermodynamics of reactions that combines accuracy and coverage. *Bioinformatics*, **28**, 2037–2044.

Noor,E. *et al.* (2012b) A proof for loop-law constraints in stoichiometric metabolic networks. *BMC Syst. Biol.*, **6**, 140.

Oliveira,J.S. (2001) An algebraic-combinatorial model for the identification and mapping of biochemical pathways. *Bull. Math. Biol.*, **63**, 1163–1196.

Orman,M.A. *et al.* (2012) Metabolic network analysis of perfused livers under fed and fasted states: incorporating thermodynamic and futile-cycle-associated regulatory constraints. *J. Theor. Biol.*, **293**, 101–110.

Orth,J.D. *et al.* (2010) What is flux balance analysis. *Nat. Biotechnol.*, **28**, 245–248.

Pey,J. *et al.* (2011) Do elementary flux modes combine linearly at the "atomic" level? integrating tracer-based metabolomics data and elementary flux modes. *Biosystems*, **105**, 140–146.

Price,N.D. *et al.* (2002) Extreme pathways and Kirchhoff's second law. *Biophys. J.*, **83**, 2879–2882.

Price,N.D. *et al.* (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat. Rev. Microbiol.*, **2**, 886–897.

Price,N.D. *et al.* (2006) Candidate states of *Helicobacter pylori's* genome-scale metabolic network upon application of "loop law" thermodynamic constraints. *Biophys. J.*, **90**, 3919–3928.

Qian,H. and Beard,D.A. (2005) Thermodynamics of stoichiometric biochemical networks in living systems far from equilibrium. *Biophys. Chem.*, **114**, 213–220.

Schellenberger,J. and Palsson,B.O. (2009) Use of randomized sampling for analysis of metabolic networks. *J Biol. Chem.*, **284**, 5457–5461.

Schellenberger,J. *et al.* (2011a) Elimination of thermodynamically infeasible loops in steady-state metabolic models. *Biophys. J.*, **100**, 544–553.

Schellenberger,J. *et al.* (2011b) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA toolbox v2.0. *Nat. Protoc.*, **6**, 1290–1307.

Schilling,C.H. *et al.* (2000) Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function form a pathway-oriented perspective. *J. Theor. Biol.*, **203**, 229–248.

Schuster,S. and Hilgetag,C. (1994) On elementary flux modes in biochemical systems at steady state. *J. Biol. Syst.*, **2**, 165–182.

Schuster,S. *et al.* (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.*, **18**, 326–332.

Schuster,S. *et al.* (2007) Is maximization of molar yield in metabolic networks favoured by evolution? *J. Theor. Biol.*, **252**, 497–504.

Schwartz,J.-M. and Kanehisa,M. (2006) Quantitative elementary mode analysis of metabolic pathways: the example of yeast glycolysis. *BMC Bioinformatics*, **7**, 186.

Shinar,G. and Feinberg,M. (2010) Structural sources of robustness in biochemical reaction networks. *Science*, **327**, 1389–1391.

Singh,A. *et al.* (2011) Manipulating redox and atp balancing for improved production of succinate in *E.coli. Metab. Eng.*, **13**, 76–81.

Terzer,M. *et al.* (2009) Genome-scale metabolic networks. *Wiley Interdiscip. Rev. Syst. Biol. Med.*, **1**, 285–297.

Teusink,B. *et al.* (2009) Understanding the adaptive growth strategy of *Lactobacillus plantarum* by in silico optimisation. *PLoS Comput. Biol.*, **5**, e1000410.

Varma,A. and Palsson,B.O. (1994) Metabolic flux balancing: basic concepts, scientific and practical use. *Nat. Biotechnol.*, **12**, 994–998.

Wolsey,L.A. (1998) *Integer Programming. Wiley-Interscience Series in Discrete Mathematics.* John Wiley & Sons, New York, NY.

Yang,F. *et al.* (2005) Ab initio predicition of thermodynamically feasible reaction directions from biochemical network stroichiometry. *Metab. Eng.*, **7**, 251–259.