

Fast Thermodynamically Constrained Flux Variability Analysis

Arne C. Müller^{1,2,4,*}, Alexander Bockmayr^{1,3}

Abstract

Motivation: Flux variability analysis (FVA) is an important tool to further analyze the results obtained by flux balance analysis (FBA) on genome-scale metabolic networks. Standard FVA may predict unbounded fluxes through some reactions in the network even if the nutrient uptake rate is bounded. These fluxes violate the second law of thermodynamics. They may be eliminated by extending flux variability analysis with thermodynamic constraints.

Results: We present a new algorithm for efficient flux variability (and flux balance) analysis with thermodynamic constraints, suitable for analyzing genome-scale metabolic networks. We first show that flux balance analysis 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 flux variability analysis 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 out 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

1 Introduction

Flux Balance Analysis (FBA) is a widely used method to analyze the capabilities of a metabolic network (Varma and Palsson, 1994; Mahadevan and Schilling, 2003; Price *et al.*, 2004; Terzer *et al.*, 2009; Durot *et al.*, 2009; Schuster *et al.*, 2007; Teusink *et al.*, 2009; Orth *et al.*, 2010). The strength of FBA is that it allows predicting growth rates of cells very accurately (Edwards *et al.*, 2001). Flux Balance Analysis 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

*to whom correspondence should be addressed

¹ Department of Mathematics and Computer Science, Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany

² International Max Planck Research School for Computational Biology and Scientific Computing (IMPRS-CBSC), Max Planck Institute for Molecular Genetics, Ihnestr 63-73, D-14195 Berlin, Germany

³ DFG-Research Center Matheon, Berlin, Germany

⁴ Berlin Mathematical School (BMS), Berlin, Germany

and Hilgetag, 1994; Schuster *et al.*, 2000) or extreme pathways (Schilling *et al.*, 2000) are commonly used tools to analyze 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 (Schwartz and Kanehisa, 2006; Haus *et al.*, 2008; Driouch *et al.*, 2012; Orman *et al.*, 2012). Therefore, sampling methods (Schellenberger and Palsson, 2009), elementary flux patterns (Kaleta *et al.*, 2009) and flux variability analysis (Mahadevan and Schilling, 2003; Burgard *et al.*, 2001) have been developed.

Flux Variability Analysis (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 very 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, thermodynamic constraints can be added to the model (Beard *et al.*, 2002; Qian and Beard, 2005). It has been shown that the flux values computed with thermodynamic constraints are more consistent, for example, with respect to gene expression data (Schellenberger *et al.*, 2011a; Feist *et al.*, 2007; Fleming *et al.*, 2009; Price *et al.*, 2006; Jol *et al.*, 2012). In recent years there has been increasing interest in thermodynamic constraints, since they can also be used to link metabolite concentrations to flux modes (Kümmel *et al.*, 2006a,b; Hoppe *et al.*, 2007; Henry *et al.*, 2006; Singh *et al.*, 2011).

In this paper, we present a new method for efficient thermodynamically constrained flux balance and flux variability analysis. 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:

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

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

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

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

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

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 (Mavrouniotis, 1990; Jankowski *et al.*, 2008; Noor *et al.*, 2012a), it is possible to further constrain μ . However, since 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 $\max\{\pm J_i \mid (1) - (4)\}$ (respectively $\max\{\pm J_i \mid (1) - (4), c^T J = \text{opt}\}$).

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* . \square

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 Eq. 3 is that normally a chemical reaction carries flux if and only if it reduces Gibbs free energy (Beard *et al.*, 2002; Alberty, 2003; Qian and Beard, 2005). Since many reactions are catalyzed 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 Eq. 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):

$$\begin{aligned} SJ &= 0 \\ \ell_i &\leq J_i \leq u_i \quad \forall i \in \mathcal{R} \\ \ell_i(1 - a_i) &\leq J_i \leq u_i a_i \quad \forall i \in \mathcal{I} \\ -Ma_i + \varepsilon &\leq \Delta\mu_i \leq M(1 - a_i) - \varepsilon \quad \forall i \in \mathcal{I} \\ \Delta\mu^T &= \mu^T S_{\mathcal{I}} \\ J \in \mathbb{R}^{\mathcal{R}}, \mu &\in \mathbb{R}^{\mathcal{M}}, \Delta\mu \in \mathbb{R}^{\mathcal{I}}, a \in \{0, 1\}^{\mathcal{I}}. \end{aligned} \tag{5}$$

$$\tag{6}$$

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; Hoppe *et al.*, 2007; Henry *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 *E. coli* (Schellenberger *et al.*, 2011a).

Beard *et al.* (2004) observed that a *steady-state flux vector* J (i.e., which satisfies Eq. 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 paper is organized as follows: In Sect. 2.1, we show that the thermodynamically constrained flux balance problem is NP-complete and thus, MILPs are an appropriate tool. In Sect. 2.2, we derive a tractability result and analyze how much this applies to current genome-scale models. These theoretical results are the backbone of a new algorithm, which is described

in Sect. 3. In Sect. 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 is 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 $\mathcal{N} = (\mathcal{M}, \mathcal{R} = \mathcal{I} \dot{\cup} \mathcal{E}, S \in \mathbb{Q}^{M \times R})$
 - 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 $\mathcal{N} = (\mathcal{M}, \mathcal{R} = \mathcal{I} \dot{\cup} \mathcal{E}, S \in \mathbb{R}^{M \times R})$ 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 $\mathcal{N} = (\mathcal{M}, \mathcal{R} = \mathcal{I} \dot{\cup} \mathcal{E}, S \in \mathbb{R}^{M \times R})$ 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. the size of the network \mathcal{N} if $C \cap (F \cup O) = \emptyset$. □

The key to the proof of Thm. 2 is Alg. 1 which gives the wanted polynomial-time algorithm. The complete proof can be found in the supplementary material.

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

Algorithm 1 This algorithm computes a thermodynamically feasible flux out of a possibly thermodynamically infeasible flux, if the conditions of Thm. 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

$$I^+ := \{i \in \mathcal{I} : J_i \geq 0\}$$

$$I^- := \{i \in \mathcal{I} : J_i \leq 0\}$$

$$L := \arg \max \left\{ \begin{array}{l} \mathbb{1}_{I^+} - \mathbb{1}_{I^-} : S_{\mathcal{I}}L = 0, \\ J_{I^-} \leq L_{I^-} \leq 0, \\ J_{I^+} \geq L_{I^+} \geq 0 \end{array} \right\}$$

$$J_{\mathcal{I}} := J_{\mathcal{I}} - L$$

until $\mathbb{1}_{I^+} - \mathbb{1}_{I^-} = 0$

return J

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.* \square

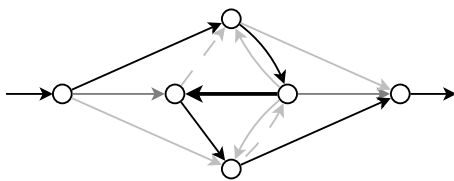


Figure 1: Blocking set (dashed arrows) for a given thermodynamically feasible flux (marked in black). Reversible reactions are drawn using two arrows. The bold arrow indicates the objective reaction.

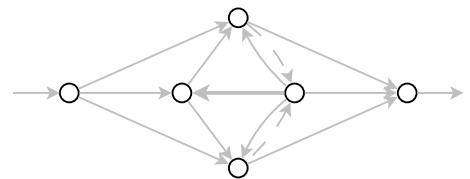


Figure 2: If a “wrong” blocking set is chosen, only suboptimal flux (here: no flux) may be possible.

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. Fig. 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 Thm. 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 Alg. 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  $\ell, u$ 
    maxFlux =  $\max(\maxFlux, c^T J)$ 
end for
return maxFlux

```

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

3 Implementation

To implement Alg. 2 and use the result of Thm. 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 Alg. 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 Thm. 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

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. $S_{\mathcal{I}}L = 0, 1 \geq \text{sign}(J_i)L_i \geq 0 \quad \forall i \in \mathcal{I}$.

if no non-zero internal cycle L is found **then** run heuristic (Alg. 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

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 Alg. 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 analyzed further. If the network contains circuits with objective reactions, we usually have a very large flux through an internal cycle and thus, a very bad dual bound. Although Alg. 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 Alg. 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 Alg. 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, 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 Thm. 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

Fig. 3 and Tab. 1. By Thm. 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 most likely due to numerical issues of the big-M formulation, since the results of our code 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 to two different kinds of MILP formulations found in the literature; we did not compare it to nonlinear formulations, since most nonlinear 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 Fig. 7 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 to the one implemented in the COBRA-Toolbox (Schellenberger *et al.*, 2011a). The CP algorithm is 30-300 times faster than the COBRA-Implementation (see Tab. 1). The only trouble maker is the *H. sapiens* reconstruction that did not finish within 2 hours 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 analyzed 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 17 Sep 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 very bad, since 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 orders better than the bounds computed by ordinary FVA. 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 Sect. 2, our algorithm works best if the number of reactions in internal cycles is small. In addition, as it can be seen in Fig. 3, the genome-scale networks in the BioModels database may be divided into two categories: networks with less than 600 reactions in internal cycles and networks with significantly more reactions. In what follows, we analyzed those networks with less than 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

Table 1: Running Times of FVA on different metabolic networks. 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.

Model	method	internal circuits (in internal circuits)	reactions (in internal circuits)	whole flux space	only optimal fluxes w.r.t. biomass production
<i>E. coli</i> iJR904	CP	19	1075 (40)	11s	11s
	COBRA			963s	Error
<i>E. coli</i> iAF1260	CP	38	2382 (68)	71s	51s
	COBRA			11162s	15049s
<i>E. coli</i> iJO1366	CP	46	2583 (76)	92s	79s
	COBRA			25869s	Error
<i>H. pylori</i> iT341	CP	8	554 (22)	4s	4s
	COBRA			122s	176s
<i>H. sapiens</i> Recon. 1	CP	?	3742 (958)	> 2h	> 2h
	COBRA			Error	Error
<i>M. barkeri</i> iAF692	CP	31	690 (30)	7s	7s
	COBRA			204s	332s
<i>M. tuberculosis</i> iNJ661	CP	66	1025 (53)	22s	19s
	COBRA			844s	1316s
<i>S. aureus</i> iSB619	CP	3	743 (7)	3s	2s
	COBRA			179s	198s
<i>S. cerevisiae</i> iND750	CP	43	1266 (64)	44s	34s
	COBRA			1662s	Error

reaction (e.g. biomass production), we also analyzed the optimal flux space with FVA and tFVA. We ran the computations of FVA and tFVA for at most 30 minutes.

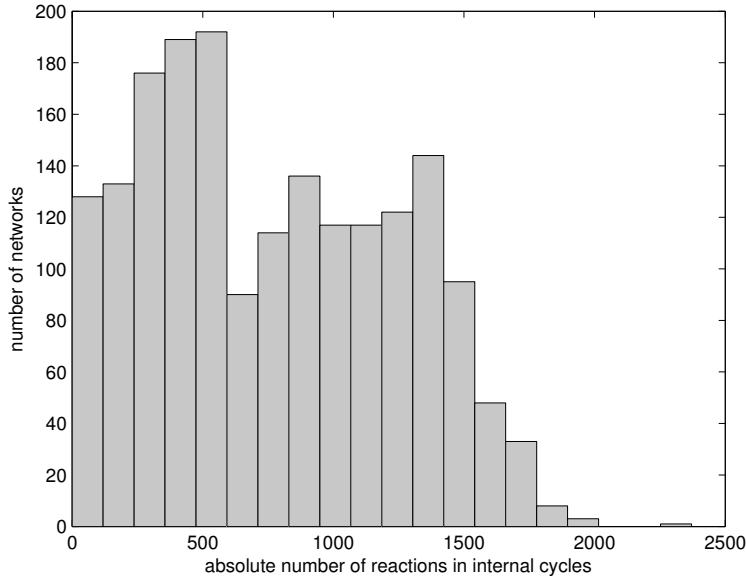


Figure 3: The genome-scale networks available in the `BioModels` database seem to separate into two classes: Networks with few (less than 600) reactions in internal cycles and networks with many more reactions in internal cycles.

We obtained that 829 (45% of all genome-scale networks in the `BioModels` database) had less than or equal to 600 reactions in internal cycles. Of these we were able to analyze 716 networks in the given time limit. We obtained the following results for the whole flux space (resp. the optimal flux space):

- in 386 (resp. 387) networks, flux through at least 1 reaction was additionally fixed due to thermodynamic constraints.
- In 481 (resp. 485) networks, at least 1 additional reaction was detected to be irreversible due to thermodynamic constraints.

In Fig. 4 we provide further information on the number of additionally fixed or irreversible reactions when analyzing 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 *S. 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 reaction 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.

4.3 Conclusion

We presented a new algorithm for solving thermodynamically constrained flux variability analysis. The run time experiments showed an enormous speed-up compared to previous imple-

Model	fixed	opt. fixed	irrev.	opt. irrev.
<i>E. coli</i> iJR904	11	30	6	5
<i>E. coli</i> iAF1260	9	41	8	6
<i>E. coli</i> iJO1366	13	44	8	7
<i>H. pylori</i> iT341	6	16	3	4
<i>M. barkeri</i> iAF692	1	7	8	9
<i>M. tuberculosis</i> iNJ661	6	13	8	13
<i>S. aureus</i> iSB619	0	5	0	1
<i>S. cerevisiae</i> iND750	5	15	13	19

Table 2: Number of additionally fixed and irreversible reactions due to thermodynamic constraints. *Opt. fixed* and *opt. irrev.* denote the number of additionally fixed resp. irreversible reactions in the space of fluxes with optimal biomass production.

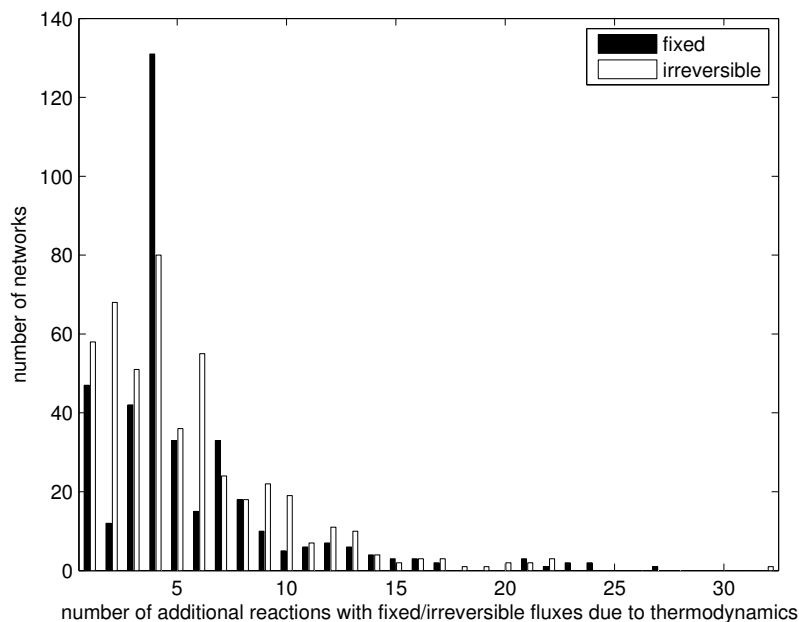


Figure 4: Distribution of networks with additional fixed resp. irreversible reactions

mentations. This way, thermodynamically constrained flux variability analysis can now be run on large sets of networks and also 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 does not allow additional information on metabolite concentrations to be included into the model. We plan to work on the incorporation of metabolite concentrations as a next step of our work. However, this extension will not be trivial. The main workhorse of the given method was the heuristic. When integrating metabolite concentrations, the main properties that made the heuristic work will not be applicable anymore.

Acknowledgement

This work was funded by the Berlin Mathematical School in terms of a PhD stipend.

Appendix: Supplementary Material

This section contains the proofs of our mathematical results and additional information on the computational experiments. We always assume that our model is finite, i.e., there is only a finite number of reactions and a finite number of metabolites. For reasons of notation, we define

$$\begin{aligned} \text{sign}(v) &:= (\{i : v_i > 0\}, \{i : v_i < 0\}) \\ (A^+, A^-) \subseteq (B^+, B^-) &:\Leftrightarrow A^+ \subseteq B^+ \wedge A^- \subseteq B^- \end{aligned}$$

For the following proofs, Thm. 0 is an essential ingredient. For proofs of Thm. 0 we refer the reader to Beard *et al.* (2004); Müller (2012) and the recent publication by Noor *et al.* (2012b).

Theorem 0 *Given a metabolic network $\mathcal{N} = (\mathcal{M}, \mathcal{R} = \mathcal{I} \dot{\cup} \mathcal{E}, S)$, a flux vector $J \in \mathbb{R}^{\mathcal{R}}$ is thermodynamically feasible if and only if it holds for all $v \in \mathbb{R}^{\mathcal{I}}$ with $\text{sign}(v) \subseteq \text{sign}(J_{\mathcal{I}})$ and $S_{\mathcal{I}}v = 0$ that $v = 0$. \square*

NP-completeness

We show that determining whether there exists a positive flux through a given reaction is NP-complete. To do so, we first establish that the problem THERMOFLUX is NP-hard (Proposition 1). Then, we prove that the problem also lies in NP, completing the NP-completeness proof (Theorem 1).

Problem 1 (Thermoflux)

- *Instance:*
 - Metabolic network $\mathcal{N} = (\mathcal{M}, \mathcal{R} = \mathcal{I} \dot{\cup} \mathcal{E}, S \in \mathbb{Q}^{M \times R})$
 - Objective reaction $r \in R$
- *Question: Does a thermodynamically feasible flux $J \geq 0$ with $J_r > 0$ exist?*

Proposition 1 THERMOFLUX is NP-hard. \square

PROOF We will reduce DISJOINT-PATH to THERMOFLUX. Fortune *et al.* (1980) showed that DISJOINT-PATH is NP-hard; hence, the reduction of DISJOINT-PATH to THERMOFLUX will prove NP-hardness of THERMOFLUX.

Problem 2 (Disjoint-Path)

- *Instance:*
 - Digraph $G = (V, A)$
 - $s_1, t_1, s_2, t_2 \in V$
- *Question: Do vertex disjoint paths from s_1 to t_1 and s_2 to t_2 exist?*

Let $G = (V, A)$ be the given digraph and let $s_1, t_1, s_2, t_2 \in V$ be the vertices we want to find disjoint paths for. Let I be the incidence matrix of G .

We now construct a metabolic network \mathcal{N} by adding an input reaction a , an output reaction o , and a target reaction r . The reaction a will deliver flow to s_1 . Reaction r will be the reaction



Figure 5: Corresponding Instances of THERMOFLUX and DISJOINT-PATH

we are interested in in THERMOFLUX and which will transport flow from t_1 to s_2 . Reaction o will serve as the output flow reaction from t_2 . Formally, we have

$$\mathcal{N} = \{V, A \cup \{a, o, r\}, S\}, \text{ with}$$

$$S_A = I, S_a = e_{s_1}, S_o = -e_{t_2}, S_r = e_{s_2} - e_{t_1},$$

where e_i denotes the i -th unit-vector. The internal reactions of \mathcal{N} are $\mathcal{I} = A \cup \{r\}$ and the external reactions are $\mathcal{E} = \{a, o\}$. See Fig. 5 for an illustration of the construction.

Lemma 1 *There exist disjoint paths from s_1 to t_1 and from s_2 to t_2 in G if and only if there exists a thermodynamically feasible flow through r in \mathcal{N} .*

PROOF \Rightarrow : Let $P_1, P_2 \subset A$ be the two disjoint paths. By removing cycles in P_1, P_2 , we can assume w.l.o.g. that P_1, P_2 are simple. Set $J_i = 1$ for all $i \in P_1 \cup P_2 \cup \{a, r, o\}$ and $J_i = 0$ else. Since P_1, P_2 are paths, we have $IJ_A = -e_{s_1} + e_{t_1} - e_{s_2} + e_{t_2}$. It follows that $SJ = 0$, hence J is a steady-state flux.

J is also thermodynamically feasible, since P_1, P_2 are vertex disjoint and simple, hence every vertex on the path $P_1 \cup P_2 \cup \{a, r, o\}$ has degree 2. Thus, no internal cycle is contained in J and J is thermodynamically feasible by Thm. 0.

\Leftarrow : Let J be a thermodynamically feasible flux with $J_r > 0$. Since a, o are the only exchange reactions, both must carry flux as well. Define $Q := \{e \in A : J_e > 0\}$. Assume there exists a path P in (V, Q) from s_1 to t_2 that does not contain r . Let $j_P := \min\{J_e : e \in P\}$ the flux through this path. Since J is thermodynamically feasible, $\tilde{J} := J - j_P (e_{P \cup \{a, o\}})$ is also a thermodynamically feasible flux (by Thm. 0) with $J_r > 0$.

Thus, we can assume w.l.o.g. that every path from s_1 to t_2 contains r . Since $J_r > 0$, there exists such a path. Let us denote this path by P . Since J is thermodynamically feasible, P is simple, i.e., it does not contain any cycles. It follows that $P \setminus \{r\}$ is the union of two vertex disjoint paths, one going from s_1 to t_1 and the other going from s_2 to t_2 . ■

We have now shown the reduction of DISJOINT-PATH to THERMOFLUX. Hence, by the NP-hardness result of Fortune et al. Fortune *et al.* (1980), it follows that THERMOFLUX is NP-hard. ■

For practical applications we can assume that the stoichiometric matrix contains only rational coefficients. In this case we can also show NP-completeness.

Theorem 1 THERMOFLUX is NP-complete. □

PROOF Since we already proved NP-hardness, it suffices to show that THERMOFLUX is in NP. We can easily check (using one matrix multiplication) if a given flux J satisfies the steady-state

assumption. For thermodynamic feasibility, it follows by Thm. 0 that we only need to check if there exists no internal cycle contained in J . This can be done by solving the following LP:

$$\begin{aligned} \max \text{sign}(J_{\mathcal{I}})L : S_{\mathcal{I}}L &= 0, \\ 1 &\geq \text{sign}(J_i)L_i \geq 0 \quad \forall i \in \mathcal{I} \end{aligned}$$

Since each component of L is restricted to be either positive or negative, it follows that the objective value is greater than zero if and only if the flux is thermodynamically infeasible. Since solving LPs is in P, we can check solutions in polynomial time and THERMOFLUX is in NP. ■

Tractability

Theorem 2 *Let $\mathcal{N} = (\mathcal{M}, \mathcal{R} = \mathcal{I} \cup \mathcal{E}, S \in \mathbb{R}^{M \times R})$ 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. the size of the network \mathcal{N} if $C \cap (F \cup O) = \emptyset$.* □

To prove this result, we consider the Alg. 1 from the main article.

Lemma 2 (Correctness) *When Algorithm 1 terminates, it returns a thermodynamically feasible flux J^* with $\text{sign}(J^*) \subseteq \text{sign}(J)$, $|J_r^*| \leq |J_r|$ for all $r \in \mathcal{R}$, and $J_{\mathcal{R} \setminus C} = J_{\mathcal{R} \setminus C}^*$.* □

PROOF Let J^i denote the value of J at the end of the i -th iteration of the algorithm. Let L^i denote the value of L computed in the i -th iteration. Hence, we have by definition of the algorithm

$$\begin{aligned} J^0 &= J \\ J_{\mathcal{I}}^i &= J_{\mathcal{I}}^{i-1} - L^i, \quad J_{\mathcal{E}}^i = J_{\mathcal{E}}^{i-1} \\ J^n &= J^* \quad \text{if } L^n = 0. \end{aligned}$$

L^i always exists, because the LP that computes L is always feasible, since 0 is a feasible solution. Since $L^n = 0$, it follows by Thm. 0 that J^n is thermodynamically feasible. Since L^i is a feasible solution of the LP, it satisfies $J_{I^-}^i \leq L_{I^-}^i \leq 0$ and $J_{I^+}^i \geq L_{I^+}^i \geq 0$ and thus $\text{sign}(J^i) \subseteq \text{sign}(J^{i-1})$, $|J_r^i| \leq |J_r^{i-1}|$ for all $r \in \mathcal{R}$ and all $i = 1, \dots, n$. By definition of C , we have $\text{supp}(L^i) \subseteq C$ for all $i = 1, \dots, n$. This implies $J_{\mathcal{R} \setminus C}^i = J_{\mathcal{R} \setminus C}^{i-1}$. The lemma follows by induction. ■

Lemma 3 (Termination and running time) *Algorithm 1 runs in polynomial time.* □

PROOF In each iteration the flux through one cycle is reduced by its maximal amount. Thus, the flux of one reaction (which wasn't zero before) becomes zero. Cycles are only searched on those reactions that have nonzero flux. Thus, reactions that didn't carry any flux will also not carry any flux after the iteration. Hence, the number of reactions carrying flux strictly decreases with each iteration. It follows that the algorithm terminates after at most $|\mathcal{I}|$ iterations. Since we can solve LPs in polynomial time, the algorithm runs in polynomial time. ■

PROOF (PROOF OF THEOREM 2) From Lemma 2 and Lemma 3, we conclude that Alg. 1 returns a thermodynamically feasible flux J^* after a polynomial running time with $J_{\mathcal{R} \setminus C} = J_{\mathcal{R} \setminus C}^*$. Since $C \cap (F \cup O) = \emptyset$, it follows that $c^T J = c^T J^*$ and $\ell_r \leq J_r^* = J_r \leq u_r$ for all $r \in F$. For $r \notin F$ and $J_r \geq 0$, we have $\ell_r \leq 0 \leq J_r^* \leq J_r \leq u_r$. Analogously, for $r \notin F$ and $J_r \leq 0$, we have $\ell_r \leq J_r \leq J_r^* \leq 0 \leq u_r$. Thus $\ell_r \leq J_r^* \leq u_r$ is satisfied for all $r \in \mathcal{R}$. ■

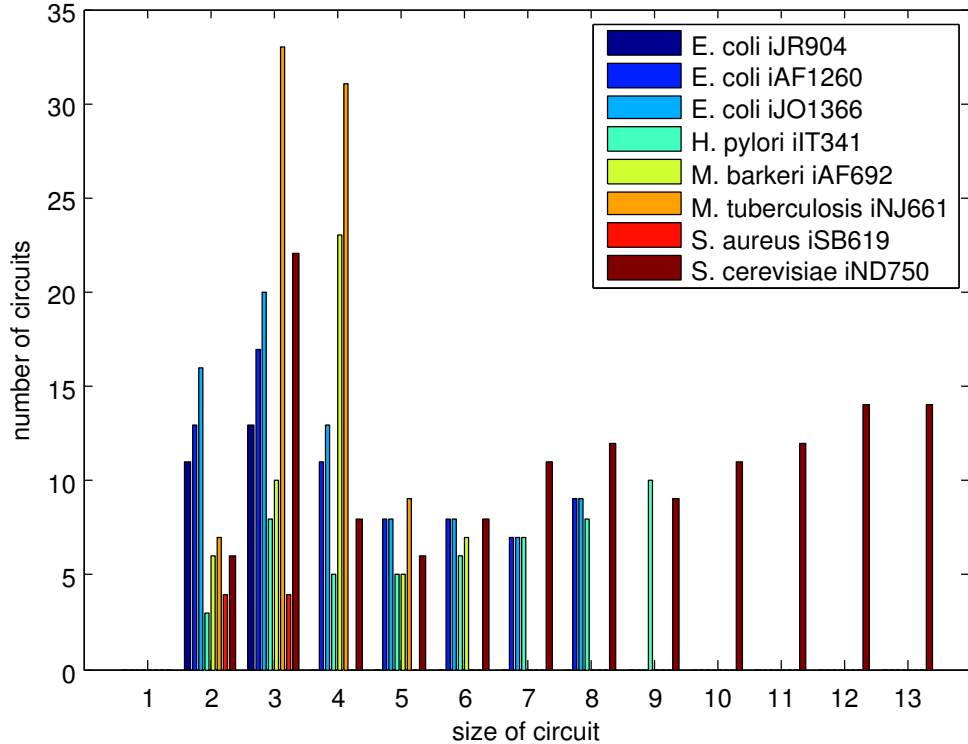


Figure 6: Sizes of the internal circuits in metabolic network models from the BiGG database

Tractability for the Networks from the BiGG database

Using the WW-Algorithm by Wright and Wagner (2008), we can compute the internal circuits for many genome-scale networks. It turns out that all models of the BiGG database (Schellenberger *et al.*, 2010) (except Human Recon. 1, where the WW-Algorithm fails) contain only relatively few internal circuits, see Fig. 6.

In Theorem 3 we derived a runtime bound of

$$\prod_{C \in \mathcal{C}} (|C| - 1), \tag{7}$$

where \mathcal{C} is the set of internal circuits that contain a flux forcing or objective reaction.

We observe that the run time increases exponentially with the number of internal circuits that contain objective or flux forcing reactions. We only were able to obtain a polynomial running time by assuming that the number of these internal circuits is bounded by a constant. This is of course not true in general. However, Fig. 6 shows that the number is small for many networks. Thus, the assumption is not violated too much, which explains why the algorithm still works very well for many networks, such as for those from the BiGG database. In particular, we observed that for *E. coli* iAF1260 the largest number of circuits that contained a common reaction was 6.

In Fig. 7 we can also see how the theoretical runtime bound behaves compared to the actual running time. Assuming that the actual running time is approximately proportional to the theoretical bound, we drew linear correlation lines. We see, that as expected the runtime of the CP method is correlated to the theoretical bound, but also the runtime of the MILP methods from the literature are correlated to the theoretical bound.

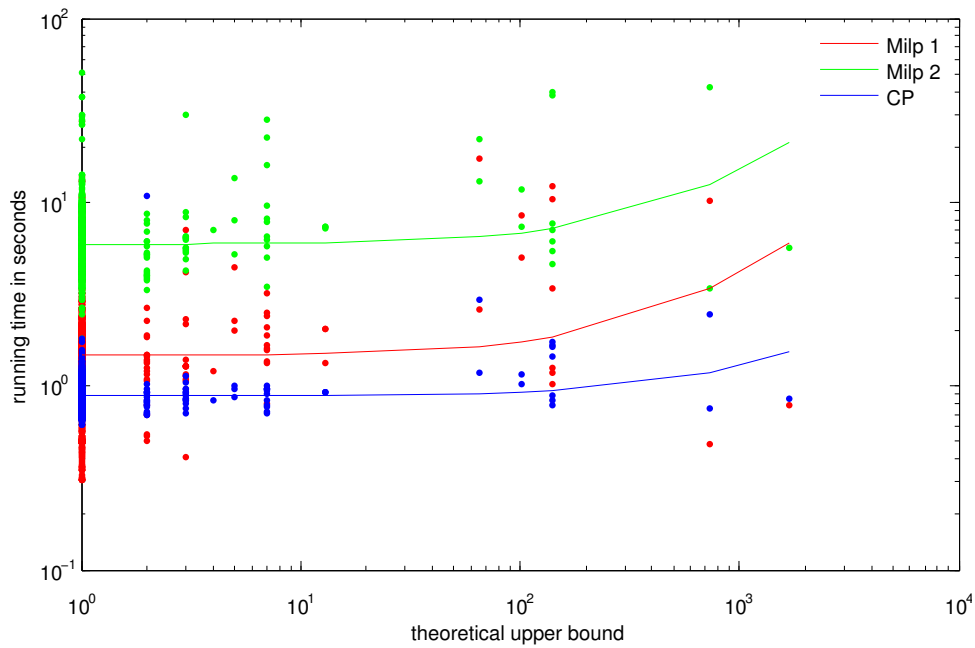


Figure 7: Comparison of different formulations of thermodynamic constraints. Milp1 is the formulation used by Schellenberger *et al.* (2011a), Milp2 is the formulation used by Hoppe *et al.* (2007); Cogne *et al.* (2011). CP is the new method presented in this paper. For each reaction of *E. coli iAF1260*, corresponding to its run time in seconds and theoretical bound, a dot is drawn (color according to the method). The lines are linear regression lines through these points. Because this plot also shows reactions that need no branching, the theoretical bound (Eq. 7) was increased by one for each reaction.

MILP Formulations used in the Runtime-Comparison

Milp 1 This is the formulation used by Schellenberger *et al.* (2011a). K is the precomputed nullspace matrix of $S_{\mathcal{I}}$, $M = 101$ and $\varepsilon = 1$.

$$\begin{aligned} \max \quad & c^T J \\ \text{s.t.} \quad & SJ = 0, \quad \ell^J \leq J \leq u^J \\ & K\Delta\mu = 0 \\ & \ell_i^J(1 - a_i) \leq J_i \leq u_i^J a_i \quad \forall i \in \mathcal{I} \\ & -Ma_i + \varepsilon \leq \Delta\mu_i \leq M(1 - a_i) - \varepsilon \quad \forall i \in \mathcal{I} \\ & a \in \{0, 1\}^{\mathcal{I}} \end{aligned}$$

Milp 2 Instead of using the null space matrix, we can also directly formulate the space of potential differences using the potentials μ . A formulation of this kind is used by Hoppe *et al.* (2007) and Cogne *et al.* (2011).

$$\begin{aligned} \max \quad & c^T J \\ \text{s.t.} \quad & SJ = 0, \quad \ell^J \leq J \leq u^J \\ & \Delta\mu^T = \mu^T S_{\mathcal{I}} \\ & \ell_i^J(1 - a_i) \leq J_i \leq u_i^J a_i \quad \forall i \in \mathcal{I} \\ & -Ma_i + \varepsilon \leq \Delta\mu_i \leq M(1 - a_i) - \varepsilon \quad \forall i \in \mathcal{I} \\ & \ell^\mu \leq \mu \leq u^\mu \\ & a \in \{0, 1\}^{\mathcal{I}} \end{aligned}$$

The Effect of Warm-Starting

The effect of the warm-starting can be observed by comparing Fig. 7 and Tab. 1 in the main article. According to Fig. 7 the speed-up should amount to at most factor 10. However, for FVA we obtain a speedup of factor 30 – 300. Hence, the remaining speed-up (of a factor of at least 15) for FVA can only be due to warm-starting.

References

- Achterberg, T. (2009). SCIP: Solving constraint integer programs. *Mathematical Programming Computation*, **1**(1), 1–41. <http://mpc.zib.de/index.php/MPC/article/view/4>.
- Alberty, R. A. (2003). *Thermodynamics of Biochemical Reactions*. Massachusetts Institute of Technology, Cambridge, MA.
- Beard, D. A., dan Liang, S., and Qian, H. (2002). Energy balance for analysis of complex metabolic networks. *Biophysical Journal*, **83**, 79–86.
- Beard, D. A., Babson, E., Curtis, E., and Qian, H. (2004). Thermodynamic constraints for biochemical networks. *Journal of Theoretical Biology*, **228**, 327–333.
- Bilu, Y., Shlomi, T., Barkai, N., and Ruppin, E. (2006). Conservation of expression and sequence of metabolic genes is reflected by activity across metabolic states. *PLoS Computational Biology*, **2**(8), e106.
- Burgard, A. P., Vaidyaraman, S., and Maranas, C. D. (2001). Minimal reaction sets for escherichia coli metabolism under different growth requirements and uptake environments. *Biotechnology Progress*, **17**, 791–797.
- Burgard, A. P., V., N. E., Schilling, C. H., and Maranas, C. D. (2004). Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Research*, **14**(2), 301312.

- Cogne, G., Rügen, M., Bockmayr, A., Tital, M., Dussap, C.-G., Cornet, J.-F., and Legrand, J. (2011). A model-based method for investigating bioenergetic processes in autotrophically growing eukaryotic microalgae: Application to the green algae *Chlamydomonas reinhardtii*. *Biotechnol Progress*, **27**(3), 631–640.
- Driouch, H., Melzer, G., and Wittmann, C. (2012). Integration of in vivo and in silico metabolic fluxes for improvement of recombinant protein production. *Metabolic Engineering*, **14**(1), 4758.
- Durot, M., Bourguignon, P.-Y., and Schachter, V. (2009). Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS Microbiology Reviews*, **33**, 164–90.
- Edwards, J. S., Ibarra, R. U., and Palsson, B. O. (2001). In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature Biotechnology*, **19**, 125–130.
- Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., Broadbelt, L. J., Hatzimanikatis, V., and Palsson, B. O. (2007). A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*, **3**(121).
- Fleming, R., Thiele, I., and Nasheuer, H. (2009). Quantitative assignment of reaction directionality in constraint-based models of metabolism: Application to *Escherichia coli*. *Biophysical Chemistry*, **145**, 47–56.
- Fleming, R. M., Maes, C. M., Saunders, M. A., Ye, Y., and Palsson, B. O. (2012). A variational principle for computing nonequilibrium fluxes and potentials in genome-scale biochemical networks. *Journal of Theoretical Biology*, **292**, 71–77.
- Fortune, S., Hopcroft, J., and Wyllie, J. (1980). The directed subgraph homeomorphism problem. *Theoretical Computer Science*, **10**(111-121).
- Gudmundsson, S. and Thiele, I. (2010). Computationally efficient flux variability analysis. *BMC Bioinformatics*, **11**(489).
- Hädicke, O., Grammel, H., and Klamt, S. (2011). Metabolic network modeling of redox balancing and biohydrogen production in purple nonsulfur bacteria. *BMC Systems Biology*, **5**(150).
- Haus, U.-U., Klamt, S., and Stephen, T. (2008). Computing knock-out strategies in metabolic networks. *Journal of Computational Biology*, **15**(3), 259–268.
- Henry, C., Jankowski, M., Broadbelt, L., and Hatzimanikatis, V. (2006). Genome-scale thermodynamic analysis of *Escherichia coli* metabolism. *Biophysical Journal*, **90**(4), 1453–1461.
- Henry, C. S., Broadbelt, L. J., and Hatzimanikatis, V. (2007). Thermodynamics-based metabolic flux analysis. *Biophysical Journal*, **92**, 1792–1805.
- Hoppe, A., Hoffmann, S., and Holzhütter, H.-G. (2007). Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks. *BMC Systems Biology*, **1**(23).
- Jankowski, M. D., Henry, C. S., Broadbelt, L. J., and Hatzimanikatis, V. (2008). Group contribution method for thermodynamic analysis of complex metabolic networks. *Biophysical Journal*, **95**, 1487–1499.
- Jol, S. J., Kümmel, A., Terzer, M., Stelling, J., and Heinemann, M. (2012). System-level insights into yeast metabolism by thermodynamic analysis of elementary flux modes. *PLoS Computational Biology*, **8**(3).
- Kaleta, C., de Figueiredo, L. F., and Schuster, S. (2009). Can the whole be less than the sum of its parts? Pathway analysis in genome-scale metabolic networks using elementary flux patterns. *Genome Research*, **19**, 1872–1883.
- Khannapho, C., Zhao, H., Bonde, B. L., Kierzek, A. M., Avignone-Rossa, C. A., and Bushell, M. E. (2008). Selection of objective function in genome scale flux balance analysis for process feed development in antibiotic production. *Metabolic Engineering*, **10**(5), 227–233.
- Kümmel, A., Panke, S., and Heinemann, M. (2006a). Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Molecular Systems Biology*, **2**(2006.0034).
- Kümmel, A., Panke, S., and Heinemann, M. (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. *Metabolic Engineering*, **5**, 264–276.
- Mavrouniotis, M. L. (1990). Group contributions for estimating standard Gibbs energies of formation of biochemical compounds in aqueous solution. *Biotechnology and Bioengineering*, **36**, 1070–1082.
- Müller, A. (2012). *Thermodynamic Constraints in Metabolic Networks*. Master’s thesis, Freie Universität Berlin, Fachbereich Mathematik und Informatik. <http://page.mi.fu-berlin.de/arnem/theses/master.pdf>.
- Noor, E., Bar-Even, A., Flamholz, A., Lubling, Y., Davidi, D., and Milo, R. (2012a). An integrated open framework for thermodynamics of reactions that combines accuracy and coverage. *Bioinformatics*, **28**(15), 2037–2044.

- Noor, E., Lewis, N. E., and Milo, R. (2012b). A proof for loop-law constraints in stoichiometric metabolic networks. *BMC Systems Biology*, **6**, 140.
- Oliveira, J. S., Bailey, C. G., Jones-Oliveira, J. B., and Dixon, D. A. (2001). An algebraic-combinatorial model for the identification and mapping of biochemical pathways. *Bulletin of Mathematical Biology*, **63**, 1163–1196.
- Orman, M. A., Androulakis, I. P., Berthiaume, F., and Ierapetritou, M. G. (2012). Metabolic network analysis of perfused livers under fed and fasted states: Incorporating thermodynamic and futile-cycle-associated regulatory constraints. *Journal of Theoretical Biology*, **293**, 101–110.
- Orth, J. D., Thiele, I., and Palsson, B. O. (2010). What is flux balance analysis. *Nature Biotechnology*, **28**, 245248.
- Price, N. D., Famili, I., Beard, D. A., and Palsson, B. O. (2002). Extreme pathways and Kirchhoff’s second law. *Biophysical Journal*, **83**, 2879–2882.
- Price, N. D., Reed, J. L., and Palsson, B. O. (2004). Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Reviews Microbiology*, **2**, 886–897.
- Price, N. D., Thiele, I., and Palsson, B. O. (2006). Candidate states of *Helicobacter pylori*’s genome-scale metabolic network upon application of “loop law” thermodynamic constraints. *Biophysical Journal*, **90**, 3919–3928.
- Qian, H. and Beard, D. A. (2005). Thermodynamics of stoichiometric biochemical networks in living systems far from equilibrium. *Biophysical Chemistry*, **114**, 213–220.
- Schellenberger, J. and Palsson, B. O. (2009). Use of randomized sampling for analysis of metabolic networks. *The Journal of Biological Chemistry*, **284**(9), 5457–5461.
- Schellenberger, J., Park, J. O., Conrad, T. M., and Palsson, B. O. (2010). BiGG: a biochemical genetic and genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics*, **11**(213).
- Schellenberger, J., Lewis, N. E., and Palsson, B. O. (2011a). Elimination of thermodynamically infeasible loops in steady-state metabolic models. *Biophysical Journal*, **100**, 544–553.
- Schellenberger, J., Que, R., Fleming, R. M., Thiele, I., Orth, J. D., Feist, A. M., Zielinski, D. C., Bordbar, A., Lewis, N. E., Rahmanian, S., Kang, J., Hyduke, D. R., and Palsson, B. O. (2011b). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA toolbox v2.0. *Nature Protocols*, **6**(9), 1290–1307.
- Schilling, C. H., Letscher, D., and Palsson, B. O. (2000). Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *Journal of Theoretical Biology*, **203**, 229–248.
- Schuster, S. and Hilgetag, C. (1994). On elementary flux modes in biochemical systems at steady state. *J. Biol. Systems*, **2**, 165–182.
- Schuster, S., Fell, D. A., and Dandekar, T. (2000). A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnology*, **18**, 326–332.
- Schuster, S., Pfeiffer, T., and Fell, D. A. (2007). Is maximization of molar yield in metabolic networks favoured by evolution? *Journal of Theoretical Biology*, **252**(3), 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).
- Singh, A., Soh, K. C., Hatzimanikatis, V., and Gill, R. T. (2011). Manipulating redox and atp balancing for improved production of succinate in *E. coli*. *Metabolic Engineering*, **13**, 76–81.
- Terzer, M., Maynard, N. D., Covert, M. W., and Stelling, J. (2009). Genome-scale metabolic networks. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, **1**(3), 285–297.
- Teusink, B., Wiersma, A., Jacobs, L., Notebaart, R., and Smid, E. (2009). Understanding the adaptive growth strategy of *Lactobacillus plantarum* by in silico optimisation. *PLoS Computational Biology*, **5**(6), e1000410.
- Varma, A. and Palsson, B. O. (1994). Metabolic flux balancing: Basic concepts, scientific and practical use. *Nature Biotechnology*, **12**, 994–998.
- Wolsey, L. A. (1998). *Integer Programming*. Wiley-interscience series in discrete mathematics. John Wiley & Sons.
- Wright, J. and Wagner, A. (2008). Exhaustive identification of steady state cycles in large stoichiometric networks. *BMC Systems Biology*, **2**(61).
- Yang, F., Qian, H., and Beard, D. A. (2005). Ab initio prediction of thermodynamically feasible reaction directions from biochemical network stoichiometry. *Metabolic Engineering*, **7**, 251–259.