

Systems biology

CycleFreeFlux: efficient removal of thermodynamically infeasible loops from flux distributions

Abdelmoneim Amer Desouki^{1,4}, Florian Jarre², Gabriel Gelius-Dietrich¹ and Martin J. Lercher^{1,3}

¹Institute for Computer Science, ²Institute for Mathematics, ³Cluster of Excellence on Plant Sciences, Heinrich Heine University, 40225 Düsseldorf, Germany, and ⁴Faculty of Computers and Informatics, Suez Canal University, Ismailia, Egypt

Associate Editor: Jonathan Wren

Received on April 4, 2014; revised on January 27, 2015; accepted on February 10, 2015

Abstract

Motivation: Constraint-based metabolic modeling methods such as Flux Balance Analysis (FBA) are routinely used to predict metabolic phenotypes, e.g. growth rates, ATP yield or the fitness of gene knockouts. One frequent difficulty of constraint-based solutions is the inclusion of thermodynamically infeasible loops (or internal cycles), which add nonbiological fluxes to the predictions.

Results: We propose a simple postprocessing of constraint-based solutions, which removes internal cycles from any given flux distribution $v^{(0)}$ without disturbing other fluxes not involved in the loops. This new algorithm, termed CycleFreeFlux, works by minimizing the sum of absolute fluxes $\|v\|_1$ while (i) conserving the exchange fluxes and (ii) using the fluxes of the original solution to bound the new flux distribution. This strategy reduces internal fluxes until at least one reaction of every possible internal cycle is inactive, a necessary and sufficient condition for the thermodynamic feasibility of a flux distribution. If alternative representations of the input flux distribution in terms of elementary flux modes exist that differ in their inclusion of internal cycles, then CycleFreeFlux is biased towards solutions that maintain the direction given by $v^{(0)}$ and towards solutions with lower total flux $\|v\|_1$. Our method requires only one additional linear optimization, making it computationally very efficient compared to alternative strategies.

Availability and implementation: We provide freely available R implementations for the enumeration of thermodynamically infeasible cycles as well as for cycle-free FBA solutions, flux variability calculations and random sampling of solution spaces.

Contact: lercher@cs.uni-duesseldorf.de

1 Introduction

Constraint-based methods for the functional analysis of metabolic networks (CBM or COBRA methods) rely on stoichiometric equations and do not require kinetic information for individual reactions. They are thus applicable to genome-scale metabolic systems (Lewis *et al.*, 2012). Flux-balance analysis (FBA), which involves the maximization of a fitness-related objective function in steady-state conditions (Orth *et al.*, 2010), has been successfully applied to predict

phenotypes such as growth rates, ATP yield and the effects of gene knockouts. Recent publications have applied FBA to predict drug targets (Colijn *et al.*, 2009; Plata *et al.*, 2010), while others have used FBA to study the evolution of metabolic systems (Barve and Wagner, 2013; Pál *et al.*, 2005, 2006).

FBA predicts the maximal value of the objective function—often the yield of biomass production (Orth *et al.*, 2010; Schuster *et al.*, 2008)—together with a set of flux values for all reactions included

in the metabolic system (a ‘flux distribution’) (Colijn *et al.*, 2009). One of the difficulties of FBA predictions is that they frequently include thermodynamically infeasible internal cycles, i.e. sets of reactions that together carry a flux that has no influence on the exchange reactions of the model (De Martino *et al.*, 2013; Price *et al.*, 2002; Schellenberger *et al.*, 2011; Schilling *et al.*, 2000; Wright and Wagner, 2008). These are metabolic ‘perpetual motion machines’ and do not occur in biological reality. Such internal cycles thus distort predicted flux distributions and should be removed from the predictions. Thermodynamically infeasible cycles affect not only the predictions of FBA, but also those of many other constraint-based analysis methods. In particular, thermodynamically infeasible cycles distort the output of sampling algorithms that aim to characterize the steady-state solution space, and lead to the assignment of unrealistically high values to the flux ranges that can be carried by individual reactions in flux variability analysis (FVA) (Mahadevan and Schilling, 2003).

At least for FBA, there exists a strategy that ensures the thermodynamic feasibility of the flux distribution: the minimization of the sum of absolute fluxes under the additional constraint of an optimal value for the original objective function (minimization of total flux, MTF) (Holzhütter, 2004; Schuetz *et al.*, 2012). This method aims to approximate a flux distribution that minimizes cellular investment into enzymes, and is hence sometimes called parsimonious FBA (Lewis *et al.*, 2010). As a side effect of its parsimonious strategy, MTF also removes any internal cycles. MTF solutions do not represent the full thermodynamically feasible solution space of the original FBA problem; hence, they cannot be used to characterize this full space, which is the purpose of random sampling algorithms and FVA calculations.

Several approaches for the identification and/or exclusion of thermodynamically infeasible cycles have been proposed (De Martino *et al.*, 2013; Schellenberger *et al.*, 2011; Wright and Wagner, 2008). The most widely used method, IL-COBRA (Schellenberger *et al.*, 2011), is based on the integration of thermodynamic constraints with FBA into a mixed-integer linear problem (MILP). However, this approach is computationally expensive, resulting in runtimes that severely limit its applicability to large-scale studies.

Here, we propose a new algorithm, termed CycleFreeFlux, which removes all thermodynamically infeasible cycles from any given flux distribution with a single linear optimization step; this makes it orders of magnitude faster than previous approaches when applied to sampling or flux variability analyses of genome-scale metabolic networks.

2 Results

2.1 Characterization of thermodynamically feasible flux distributions

To motivate the CycleFreeFlux algorithm, we first formally characterize thermodynamically feasible flux distributions. Consider the following standard FBA problem (bold face indicates vectors):

$$\begin{aligned} & \max \mathbf{c}^T \mathbf{v} \\ & \text{subject to: } \mathbf{S} \mathbf{v} = 0 \\ & \mathbf{l} \leq \mathbf{v} \leq \mathbf{u} \end{aligned} \quad (1)$$

Here, \mathbf{S} is a stoichiometric matrix, where each row corresponds to one metabolite, and each column contains the stoichiometric coefficients of one reaction; \mathbf{v} is a vector of fluxes through these reactions; $\mathbf{S} \mathbf{v} = 0$ enforces a steady state in which all internal metabolite

concentrations are constant; \mathbf{l} and \mathbf{u} are vectors of lower and upper bounds for \mathbf{v} , respectively, where \leq is applied element-wise; and $\mathbf{c}^T \mathbf{v}$ with a constant weight vector \mathbf{c} is a linear combination of fluxes that is maximized (the ‘gain’ of \mathbf{v}). In many applications, this objective function contains just one hypothetical flux that corresponds to the rate of biomass production, $\mathbf{c}^T \mathbf{v} = v_{bio}$ (Schuster *et al.*, 2008).

In the following, we will call a flux distribution ‘feasible’ if it is nonzero and adheres to all constraints in the FBA problem (1) (i.e. if it lies in the space of feasible solutions of (1)). $\text{support}(\mathbf{v}')$ denotes the set of indices j with $v_j' \neq 0$. A flux distribution \mathbf{v}' is thermodynamically feasible if, and only if, there exists an assignment of free energies \mathbf{G} to the metabolites such that the free energy change caused by each active reaction is strictly negative, i.e. $v_j' \Delta G_j < 0$ for all $j \in \text{support}(\mathbf{v}')$, with $\Delta G_j = \mathbf{s}_j^T \mathbf{G}$, where \mathbf{s}_j denotes the j -th column of \mathbf{S} (Beard *et al.*, 2002; Schellenberger *et al.*, 2011).

We will explicitly consider the free energies of external metabolites in order to deal correctly with thermodynamically infeasible flux distributions that involve transport reactions (reactions that shuttle metabolites between internal compartments and the external compartment). For each external metabolite, we add one reversible ‘exchange’ reaction between the external compartment and an additional ‘NULL’ compartment that is not explicitly modeled. Many genome-scale metabolic reconstructions already define exchange reactions in that way. For any given free energy of a metabolite present in the external compartment, we can set a hypothetical free energy outside the external compartment that drives the exchange reaction in the desired direction. Thus, exchange reactions as defined here are always thermodynamically feasible. When considering thermodynamical feasibility, we only need to examine internal reactions (including transport reactions), and it is hence convenient to partition the stoichiometric matrix into an internal and an exchange part, $\mathbf{S} = [\mathbf{S}_{int}, \mathbf{S}_{ex}]$.

We call a nonzero flux distribution $\Delta \mathbf{v}$ ‘internal’ if, and only if, it fulfills the conditions $\mathbf{S} \Delta \mathbf{v} = 0$, $\min \{l_i, 0\} \leq v_i \leq \max \{0, u_i\}$ for all fluxes v_i , and if all its exchange fluxes are zero. Thus, internal flux distributions do not change any internal or external metabolite concentrations; they can be thought of as combinations of internal cycles that collectively neither consume nor produce anything. If the constraints of (1) do not enforce any nonzero fluxes, then internal flux distributions are also feasible.

As free energy is a state variable of metabolites, the free energy changes of the reactions active in $\Delta \mathbf{v}$ must add to zero in steady state: $\Delta \mathbf{v} \Delta \mathbf{G} = 0$ (Beard *et al.*, 2002; Schellenberger *et al.*, 2011). Thus, $\Delta v_k \Delta G_k \geq 0$ for at least one reaction $k \in \text{support}(\Delta \mathbf{v})$. There is hence no thermodynamic driving force for at least one reaction in $\text{support}(\Delta \mathbf{v})$; consequently, all internal flux distributions are thermodynamically infeasible.

The following theorem characterizes thermodynamically feasible flux distributions:

Theorem: A flux distribution $\mathbf{v} \neq 0$ that is feasible for (1) is thermodynamically feasible if, and only if, there does not exist any internal flux distribution $\Delta \mathbf{v} \neq 0$ with $\text{support}(\Delta \mathbf{v}) \subset \text{support}(\mathbf{v})$ and $\Delta v_k v_k \geq 0$ for all k .

Proof: We may assume, without loss of generality, that $v_k \neq 0$ for all k . (If not delete the k -th component of \mathbf{v} and the k -th column of \mathbf{S} .) We may further assume that $v_k > 0$ for all k . (If not replace the k -th component ‘ v_k ’ by ‘ $-v_k$ ’ and the k -th column ‘ \mathbf{s}_k ’ of \mathbf{S} by ‘ $-\mathbf{s}_k$ ’.) Then, by definition, \mathbf{v} is thermodynamically feasible, if, and only if, there exists a vector of free energies \mathbf{G} such that $\mathbf{S}_{int}^T \mathbf{G} < 0$ (i.e. each component of the vector $\mathbf{S}_{int}^T \mathbf{G}$ must be negative).

By Gordan's theorem (see, e.g. Theorem 2.2.1 in (Borwein and Lewis, 2006)), this is the case, if, and only if, there does not exist a vector $\mathbf{x} \geq 0$ with $\mathbf{x} \neq 0$ and $S_{int}\mathbf{x} = 0$. Identifying \mathbf{x} with $\Delta\mathbf{v}$ the claim follows.

We have thus shown that a feasible flux distribution \mathbf{v} is thermodynamically feasible if and only if it cannot be 'reduced' by subtracting an internal flux distribution $\Delta\mathbf{v}$.

2.2 Removing internal cycles from a given flux distribution

The theorem provides the motivation for the CycleFreeFlux algorithm. We aim to reduce a given flux distribution $\mathbf{v}^{(0)}$ to its thermodynamically feasible part. To achieve this goal, CycleFreeFlux minimizes the sum of absolute fluxes while (i) all exchange fluxes are kept constant, and (ii) no internal flux is allowed to change direction or increase in size. If the input flux distribution is the output of a previous optimization, CycleFreeFlux additionally constrains the value of the objective function $\mathbf{c}^T\mathbf{v}$ to its optimal value. We assume that there are no lower bounds $l_i > 0$ and no upper bounds $u_i < 0$, i.e. all fluxes are allowed to be nonactive; otherwise, these constraints may enforce thermodynamically infeasible fluxes, and then $\mathbf{v}^{(0)}$ cannot be reduced to a feasible solution that is also thermodynamically feasible.

We thus solve the following linear optimization problem:

$$\begin{aligned} \min \quad & \sum_i |v_i| \\ \text{subject to:} \quad & S\mathbf{v} = 0 \\ & 0 \leq v_i \leq v_i^{(0)} \text{ for } i \text{ with } v_i^{(0)} \geq 0 \\ & v_i^{(0)} \leq v_i \leq 0 \text{ for } i \text{ with } v_i^{(0)} < 0 \\ & v_j = v_j^{(0)} \text{ for all exchange fluxes } v_j \\ & \mathbf{c}^T\mathbf{v} = \mathbf{c}^T\mathbf{v}^{(0)} \end{aligned} \quad (2)$$

As fluxes are not allowed to change directions, we can transform (2) trivially to a linear problem by replacing the sum over the absolute values by two sums over positive and negative fluxes, respectively: $\sum_i |v_i| = \sum_{i \text{ with } v_i^{(0)} > 0} v_i - \sum_{i \text{ with } v_i^{(0)} < 0} v_i$.

The flux distribution \mathbf{v} resulting from this optimization is 'structurally consistent' with $\mathbf{v}^{(0)}$. Here, we define a flux distribution \mathbf{v} as structurally consistent with an input flux distribution $\mathbf{v}^{(0)}$ if \mathbf{v} neither increases (in absolute value) nor inverts any fluxes compared to $\mathbf{v}^{(0)}$. CycleFreeFlux outputs a minimal (in terms of the sum of absolute fluxes $\|\mathbf{v}\|_1$) flux distribution that is structurally consistent with the input flux distribution $\mathbf{v}^{(0)}$.

Corollary: A flux distribution $\mathbf{v} \neq 0$ that is feasible for (1) (with $l \leq 0 \leq u$) is thermodynamically feasible if, and only if, the output of the CycleFreeFlux algorithm (2) with input $\mathbf{v}^{(0)} = \mathbf{v}$ is \mathbf{v} itself.

Proof: To simplify the presentation assume again that $\mathbf{v}^{(0)} \geq 0$. If CycleFreeFlux returns a flux $\mathbf{v} \neq \mathbf{v}^{(0)}$, then, by construction, $\Delta\mathbf{v} := \mathbf{v}^{(0)} - \mathbf{v} \geq 0$ and $\text{support}(\Delta\mathbf{v}) \subset \text{support}(\mathbf{v}^{(0)})$. Thus, by the theorem, $\mathbf{v}^{(0)}$ is not thermodynamically feasible. Conversely, if $\mathbf{v}^{(0)}$ is not thermodynamically feasible, then, by the theorem, there exists an internal flux distribution $\Delta\mathbf{v} \neq 0$ with $\text{support}(\Delta\mathbf{v}) \subset \text{support}(\mathbf{v}^{(0)})$. Then, $\mathbf{v}^{(0)} - \epsilon\Delta\mathbf{v}$ is feasible for (2) for small $\epsilon > 0$, reducing the objective value in (2), so that CycleFreeFlux will not return $\mathbf{v} = \mathbf{v}^{(0)}$ as output.

Figure 1 illustrates our strategy on a simple toy model with five reactions, v_1, \dots, v_5 . All steady-state flux distributions in this model can be expressed as non-negative linear combinations of the three

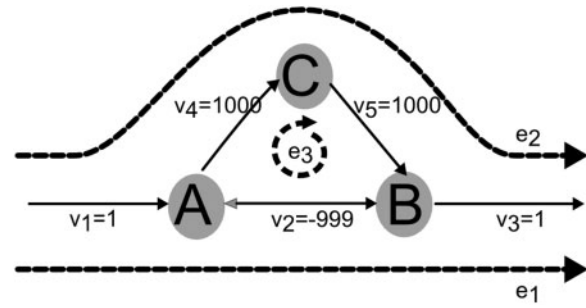


Fig. 1. A toy model illustrating our algorithms. Three metabolites (A, B, C) can be interconverted by three internal reactions (v_2, v_4, v_5). A can be taken up from the environment through the exchange reaction v_1 , while B can be secreted through v_3 . All fluxes except v_2 are unidirectional, allowed to proceed only from left to right (which is also the nominal flux direction of v_2). The steady-state flux distributions of this model can be expressed as positive linear combinations of the elementary flux modes $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ (dashed lines). The cycle \mathbf{e}_3 involves only internal fluxes (i.e. it is a type III elementary flux mode), and is thus thermodynamically infeasible

elementary flux modes (Schuster and Hilgetag, 1994) $\mathbf{e}_1, \mathbf{e}_2$ and \mathbf{e}_3 . \mathbf{e}_1 and \mathbf{e}_2 involve the exchange fluxes v_1 and v_3 and are thus type I elementary flux modes (analogous to type I extreme pathways (Price et al., 2002)). In contrast, the loop \mathbf{e}_3 involves only the internal fluxes v_2, v_4 and v_5 and is hence type III (i.e. thermodynamically infeasible).

The fluxes shown in the figure are our input flux distribution, $\mathbf{v}^{(0)}$. To apply the CycleFreeFlux algorithm, we constrain the internal reactions v_4 and v_5 to non-negative values ($0 \leq v_4 \leq 1000$ and $0 \leq v_5 \leq 1000$), while the third internal reaction, v_2 , is constrained to nonpositive values ($-999 \leq v_2 \leq 0$). At the same time, we fix the exchange reactions at their input values, $v_1 = v_3 = 1$. Minimizing the total flux under these conditions reduces the flux through the internal loop down to the point where v_2 becomes zero, and hence the algorithm puts out $\mathbf{v} = \mathbf{e}_2$ as the thermodynamically feasible part of the input flux distribution $\mathbf{v}^{(0)} = \mathbf{e}_2 + 999\mathbf{e}_3$.

Note that the decomposition of the input flux distribution $\mathbf{v}^{(0)}$ in terms of the elementary modes is not unique: we can also write $\mathbf{v}^{(0)} = \mathbf{e}_1 + 1000\mathbf{e}_3$. Thus, $\mathbf{v} = \mathbf{e}_1$ would be an alternative thermodynamically feasible solution consistent with the input flux distribution $\mathbf{v}^{(0)}$. This alternative inverts the direction of one internal flux, v_2 , and is thus not structurally consistent with $\mathbf{v}^{(0)}$.

The preference for structurally consistent solutions causes a corresponding bias if there exist alternative decompositions of $\mathbf{v}^{(0)}$ into elementary flux modes that are structurally inconsistent with $\mathbf{v}^{(0)}$; in addition, CycleFreeFlux is biased towards solutions with lower total flux $\|\mathbf{v}\|_1$. That there always exists a structurally consistent solution is caused by the fact that internal flux distributions are cycles, and hence it is always possible to route the thermodynamically feasible flux through the cycle section that runs in the same direction. The condition of structural consistency with an input flux distribution is what differentiates cycleFreeFlux from MTF.

We can write the thermodynamically feasible output flux distribution $\mathbf{v} = \sum_i \alpha_i \mathbf{e}_i^{(I)}$ in terms of type I elementary flux modes, and the internal flux distribution $\Delta\mathbf{v} = \sum_i \beta_i \mathbf{e}_i^{(III)}$ in terms of type III elementary flux modes (Price et al., 2002). This allows us to write

$$\begin{aligned} \mathbf{v}^{(0)} &= \sum_i \alpha_i \mathbf{e}_i^{(I)} + \sum_j \beta_j \mathbf{e}_j^{(III)} \\ \text{with } \mathbf{v} &= \sum_i \alpha_i \mathbf{e}_i^{(I)} \end{aligned} \quad (3)$$

i.e. there exists a decomposition of the input flux distribution into

type I and type III elementary flux modes such that the output flux distribution consists exactly of the type I contributions.

Note that if the constraints in (1) entail lower flux bounds $l_i > 0$ and/or upper bounds $u_i < 0$, then these enforce nonzero fluxes through the corresponding reactions. Such bounds are rarely used in FBA-type analyses. However, if such bounds are used and are chosen inappropriately, they may result in nonzero activities of thermodynamically infeasible cycles. These will be removed by our algorithm, resulting in an output flux distribution that will not be feasible for the original problem. That such an inconsistency may be present can be detected by comparing the output of the CycleFreeFlux algorithm to the constraints of the original problem to check its feasibility. However, if one is certain that the constraints of (1) do not enforce internal cycles, then corresponding modifications of the constraints in (2) to handle this case are straightforward.

Futile cycles involve the consumption and/or production of ‘currency metabolites’ such as ATP. Futile cycles are often thermodynamically feasible. They are not internal flux distributions, and do not require special treatment in our formalism.

2.3 Speed

Because of their different strategies, the output of CycleFreeFlux and Il-FBA cannot be compared directly: Il-FBA (Schellenberger et al., 2011) directly solves an FBA problem within the subspace of thermodynamically feasible flux distributions; in contrast, CycleFreeFlux takes any given steady-state flux distribution (which may or may not be the result of an FBA calculation) and removes its thermodynamically infeasible contributions.

As the CycleFreeFlux algorithm is a postprocessing step consisting of a single linear optimization, it increases the computation time approximately 2-fold compared to standard FBA, and is thus very similar to an MTF strategy in terms of run times. Table 1 compares the run time of CycleFreeFBA (standard FBA followed by the CycleFreeFlux algorithm) to that of the Il-FBA algorithm proposed by Schellenberger et al. (2011). For better comparability, both algorithms were implemented in R and are run on the same metabolic networks, using the same numerical solver. We used the default environments and biomass reactions supplied by the BIGG database (Schellenberger et al., 2010) for the *Escherichia coli* models and by the original publication for the yeast model (Mo et al., 2009).

For the genome-scale networks of *E.coli* (Feist et al., 2007) and *Saccharomyces cerevisiae* (Mo et al., 2009), cycleFreeFlux is 400–2500 times faster than the alternative Il-FBA algorithm, which solves a mixed-integer linear problem instead of a standard linear programming problem.

2.4 Cycle-free sampling

The CycleFreeFlux algorithm can be applied to any given steady-state flux distribution. Thus, it can not only remove internal cycles from FBA solutions, but it can also be directly applied to random samples of the solution space (Bordel et al., 2010). To remove thermodynamically infeasible cycles from sampled flux distributions, we reduce each sample to its contributions from type I elementary flux modes as explained above.

Flux samples taken from the two solution spaces including and excluding internal cycles are compared in Figure 2 for the genome-scale metabolic model of *E.coli* (Feist et al., 2007). The sampled fluxes involved in thermodynamically infeasible internal cycles (Fig. 2a) are reduced (Fig. 2b) to the ‘normal’ range observed for reactions not involved in active loops (Fig. 2c; these nonloop fluxes

Table 1. Runtime comparison between Il-FBA (Schellenberger et al., 2011) and CycleFreeFBA

Model	Reactions no	Solver	Il-FBA ^a	CycleFreeFlux ^a
Ec_core	95	GLPK	0.08	0.03
Ec_core	95	CPLEX	0.42	0.04
iMM904	1577	GLPK	225	0.53
iMM904	1577	CPLEX	172	0.16
iAF1260	2382	GLPK	1099	0.79

^aRun times in seconds on a standard laptop with core i7 processor and 8 GB RAM.

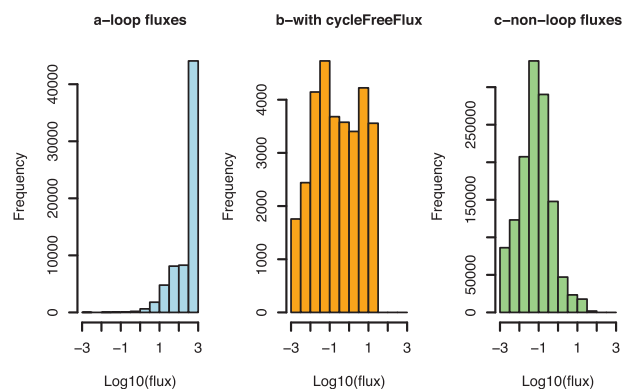


Fig. 2. Cycle-free random sampling of the solution space. (a) Histogram of the sampled flux values for fluxes involved in internal cycles. (b) The same fluxes after reduction to the cycle-free solution subspace with the CycleFreeFlux algorithm. (c) Flux values of reactions not involved in active internal cycles. Random samples were taken from the solution space of the *E.coli* genome-scale metabolic network iAF1260 in a glucose-limited aerobic medium. Flux values affected by involvement in active internal cycles were identified as those that changed through the application of CycleFreeFlux; these are summarized in (a), while all the remaining (unchanged) fluxes are summarized in (c).

are identified as those that remain unchanged when applying the CycleFreeFlux algorithm).

The sampling of the thermodynamically feasible solution space aims at covering all steady-state solutions with the same finite probability. This aim is not fully achieved through the CycleFreeFlux algorithm, as can be seen from the example in Figure 1. Ideally, the two type I elementary flux modes, e_1 and e_2 , should be sampled in equal proportions. However, any initial sample that contains a non-zero flux around e_3 will be reduced to a flux distribution that only contains e_2 , biasing the samples towards this flux mode. Any point in the solution space that is involved in an internal cycle and simultaneously has a nonunique representation in terms of elementary flux modes may cause a violation of uniform sampling. We note, however, that uniform sampling is required only when volume-related features are investigated. In many applications, the bias introduced by cycleFreeFlux may be preferable to the bias caused by the inclusion of thermodynamically infeasible flux distributions.

Il-COBRA (Schellenberger et al., 2011) also suggests to remove internal cycles from sampled flux distributions in a postprocessing step, minimizing the distance of the flux vectors between the sampled flux vector and the cycle-free subspace. As can be seen from the example in Figure 1, this strategy introduces a bias that is very similar to the one caused by the CycleFreeFlux algorithm. Thus, currently no published method exists that guarantees uniform

sampling of the thermodynamically feasible solution space; as CycleFreeFlux increases the time for sampling only about 2-fold compared to methods that include internal cycles, we consider CycleFreeFlux the method of choice.

2.5 Cycle-free flux variability analysis

Flux variability analysis (FVA) calculates upper and lower bounds for steady-state fluxes through each reaction at the optimal value of the objective function (Mahadevan and Schilling, 2003). Thus, FVA results can be used to characterize the space of alternative optimal solutions to an FBA problem. Thermodynamically infeasible internal cycles are unbounded except by *a priori* constraints, leading to artificial bounds for all reactions involved in such cycles (Mahadevan and Schilling, 2003).

To exclude internal cycles from FVA, we propose CycleFreeFVA, an extension of the CycleFreeFlux algorithm. Standard FVA performs two linear optimizations for each reaction R_i , one maximization and one minimization (Mahadevan and Schilling, 2003). We extend this approach by iteratively performing maximizations and removing cycles. To obtain the flux variability for reaction R_i excluding thermodynamically infeasible internal cycles (i.e. based only on type I elementary flux modes), we apply the following algorithm. Any constraints added in one iteration are maintained through further iterations, and the flux distribution $\mathbf{v}^{(0)}$ is updated accordingly in each iteration:

```
Repeat: {
    find a steady state flux distribution  $\mathbf{v}^{(0)}$  with maximal flux
     $v_i^{(0)}$  through  $R_i$ ;
    apply the cycleFreeFlux algorithm to  $\mathbf{v}^{(0)}$ , resulting in a
    new flux distribution  $\mathbf{v}^{(1)}$ ;
    if ( $v_i^{(0)} = v_i^{(1)}$ ) {
        exit iteration (as  $v_i^{(0)}$  is not affected by internal
        cycles);
    } else {
        apply the CycleFreeFlux algorithm to  $\mathbf{v}^{(0)}$  while
        constraining  $v_i = v_i^{(0)}$ , resulting in a new flux dis-
        tribution  $\mathbf{v}^{(2)}$  that contains one internal cycle
        involving  $R_i$ ;
        break the internal cycle: constrain the fluxes  $v_j$  that
        are zero in  $\mathbf{v}^{(1)}$  but not in  $\mathbf{v}^{(2)}$  to  $v_j = 0$ ;
    }
}
```

The upper bound for v_i excluding internal cycles is now in $v_i^{(0)}$. In sum, after maximizing the flux through reaction R_i , the algorithm checks whether the reaction is involved in internal cycles by applying the CycleFreeFlux algorithm to the resulting flux distribution $\mathbf{v}^{(0)}$; if the flux v_i through R_i is reduced through the removal of internal cycles, then the original maximum $v_i^{(0)}$ was indeed influenced by such cycles. We identify one such cycle involving v_i by fixing v_i at its original (cycling) maximum $v_i^{(0)}$ and then removing all other cycles; afterwards, we break this cycle by reducing its flux as little as possible, but as much as necessary. These steps are repeated until all loops involving v_i have been removed.

The calculation of the lower bound for v_i excluding internal cycles is performed in the same way, replacing maximizations with minimizations.

To see how the algorithm works, it is again instructive to consider the toy model in Figure 1. Let us assume that we want to calculate flux variability within the steady-state solution space to an FBA problem with v_3 as the objective function and $v_1 \leq 1$. $v_1 = 1$ and $v_3 = 1$ are then fixed by the boundary conditions, and we need to

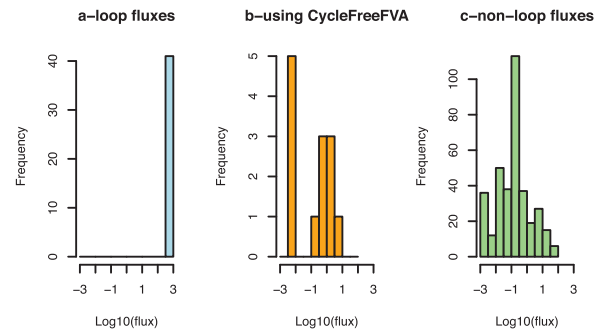


Fig. 3 Cycle-free flux variability analysis. (a) Histogram of the maximal flux values that result from involvement in internal cycles. (b) Maximal flux values of the same reactions after excluding internal cycles with the CycleFreeFVA algorithm. (c) Maximal flux values of reactions not affected by internal cycles. Analyses were performed for the *E.coli* genome-scale metabolic network iAF1260 in glucose-limited aerobic medium. Reactions affected by involvement in internal cycles were identified as those for which the maximum flux differed between standard FVA and CycleFreeFVA; these are summarized in (a) and (b), while the remaining (unaffected) reactions are summarized in (c).

find the variability of the remaining three reactions. In standard FVA, we would conclude that v_4 can vary within the full range allowed by the *a priori* constraints, say, $0 \leq v_4 \leq 1000$. Now let us apply the above algorithm for v_4 . The first flux distribution with maximal flux through v_4 will be the one indicated in Figure 1, $\mathbf{v}^{(0)} = \mathbf{e}_2 + 999\mathbf{e}_3$. Application of the CycleFreeFlux algorithm reduces this to $\mathbf{v}^{(1)} = \mathbf{e}_2$. v_4 changed its value between $\mathbf{v}^{(0)}$ and $\mathbf{v}^{(1)}$, indicating that we're not done yet. As the model contains only one internal cycle, we have $\mathbf{v}^{(2)} = \mathbf{v}^{(0)}$. The only reaction that is zero in $\mathbf{v}^{(1)}$ but not in $\mathbf{v}^{(2)} = \mathbf{v}^{(0)}$ is v_2 ; this reaction we now constrain to zero. The maximal value for v_4 is now 1. Exactly the same happens when we consider v_5 .

What about v_2 ? Its maximum, 1, is not affected by the internal cycle \mathbf{e}_3 . Its minimum is again realized in the flux distribution $\mathbf{v}^{(0)} = \mathbf{e}_2 + 999\mathbf{e}_3$ indicated in Figure 1. Again, we have $\mathbf{v}^{(1)} = \mathbf{e}_2$ and $\mathbf{v}^{(2)} = \mathbf{v}^{(0)}$. Reducing the flux through the cycle as little as possible, but as much as necessary, we identify the lower limit $0 \leq v_2$.

We applied the CycleFreeFVA algorithm to the genome-scale model of *E.coli* (Feist et al., 2007). Figure 3 shows the distribution of the maximal fluxes for reactions that can be affected by internal cycles with standard FVA (Fig. 3a) and with CycleFreeFVA (Fig. 3b), and compares those to the distribution for fluxes not affected by internal cycles (Fig. 3c).

2.6 Enumeration of internal cycles

To enumerate thermodynamically infeasible cycles, we propose a method based on an extension of the CycleFreeFVA approach. We perform CycleFreeFVA for each internal flux v_i , and we simply store the cycles removed in each iteration (i.e. $\Delta \mathbf{v} := \mathbf{v}^{(2)} - \mathbf{v}^{(1)}$). Because the same cycle may be identified for each reaction involved in the cycle, we need to restrict the total set of identified internal cycles to unique reaction sets.

Each set of reactions identified through this algorithm represents a unique internal flux distribution that cannot be reduced while maintaining the flux through one of its reactions (the one used for its identification); thus, it is an 'elementary' internal cycle (type III elementary flux mode).

Can we be sure to identify all internal cycles in this way? For each reaction involved in internal cycles, CycleFreeFVA identifies at least one such cycle. Thus, each internal cycle that contains at least

one reaction not shared with another cycle must be identified as soon as this unique reaction is considered. However, it is possible to construct examples where all reactions in one loop are also involved in other loops; in this case, the corresponding loop might be missed by our algorithm. The enumeration algorithm is thus only approximate: we may fail to identify some internal cycles if these share all their reactions with other internal cycles. On the other hand, we can be sure that all reactions involved in internal cycles are identified, and the algorithm is exact in this sense.

We used our algorithm to identify thermodynamically infeasible cycles in the genome-scale metabolic model for *E.coli* (Feist et al., 2007). The identification of internal cycles in this network took less than five minutes on a standard laptop (core i7 processor and 8GB RAM). Excluding trivial cycles involving only two reactions (i.e. forward and backward reactions of the same stoichiometry), we identified 26 nontrivial thermodynamically infeasible cycles. This is very close to the number reported by (Wright and Wagner, 2008), 27; as the detailed results of (Wright and Wagner, 2008) are not available, we could not examine if the one additional cycle consists entirely of reactions involved in the other cycles, which would explain why it is not contained in our results.

Recently, (De Martino et al., 2013) reported a lower bound of 189 nontrivial thermodynamically infeasible loops (≥ 2 reactions each) in the same metabolic network model. 11 of these internal cycles were also identified by our approach; the remaining 178 cycles violate at least one *a priori* thermodynamic constraint, i.e. at least one reaction in the cycle carries flux in a direction not allowed by the published model (Feist et al., 2007).

3 Conclusions

The CycleFreeFlux algorithm described above, MTF (Holzhütter, 2004), and ll-FBA (Schellenberger et al., 2011) each put out one thermodynamically feasible flux distribution. However, these algorithms are not directly comparable, as they differ in both their goals and their inputs. MTF delivers very special flux distributions, which minimize the total sum of absolute fluxes $\|v\|_1$, constrained only by the optimal value of the objective function. In contrast, cycleFreeFlux takes any given steady-state flux distribution (which may or may not be the result of an FBA calculation) and reduces it to its loopless contribution. Finally, ll-FBA solves an FBA problem such that the resulting flux distribution already lies in the thermodynamically feasible subspace. Given that the MTF flux distribution will often be biologically more realistic than many alternative solutions to an FBA problem (Schuetz et al., 2012), MTF (or parsimonious FBA) should be the method of choice when just one thermodynamically feasible solution to an FBA problem is required. In contrast, CycleFreeFlux and ll-COBRA provide means to characterize the full space of thermodynamically feasible (optimal) steady-state solutions. CycleFreeFlux solutions are biased towards solutions that run in the same direction as internal cycles with which they overlap and towards solutions with lower total flux $\|v\|_1$ whenever alternative decompositions into elementary modes exist; in many applications, these biases will not be important.

The CycleFreeFlux algorithm and its extensions allow the reliable identification and exclusion of thermodynamically infeasible internal cycles. We implemented corresponding functions as an extension package for *sybil* (Gelius-Dietrich et al., 2013), an open-source library for constraint-based analyses in the framework of the R-project for statistical computing (R Development Core Team 2013). The implemented CycleFreeFlux functions are orders of

magnitude faster than previously available algorithms for the generation of thermodynamically feasible random flux samples and flux variability analyses and are freely available from CRAN (<http://cran.r-project.org/web/packages/sybilcycleFreeFlux/>).

Acknowledgements

A.A.D. and M.J.L. conceived of the project and wrote the manuscript. A.A.D. carried out the work. G.G.D. participated in the discussions and helped with the implementation of the *sybil* extension. F.J. developed the formal proof to the Theorem.

Funding

We thank Stefan Schuster for helpful discussions. A.A.D. was supported by a DAAD scholarship. M.J.L. was supported by the German Research Foundation (DFG grants EXC 1028 and CRC 680).

Conflict of Interest: none declared.

References

- Barve,A. and Wagner,A. (2013) A latent capacity for evolutionary innovation through exaptation in metabolic systems. *Nature*, **500**, 203–206.
- Beard,D.A. et al. (2002) Energy balance for analysis of complex metabolic networks. *Biophys. J.*, **83**, 79–86.
- Bordel,S. et al. (2010) Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes. *PLoS Comput. Biol.*, **6**, e1000859.
- Borwein,J. and Lewis,A. (2006) *Convex Analysis and Nonlinear Optimization*. 2nd edn. CMS Books in Mathematics. Springer, New York, pp. 23–27.
- Colijn,C. et al. (2009) Interpreting expression data with metabolic flux models: predicting *Mycobacterium tuberculosis* mycolic acid production. *PLoS Comput. Biol.*, **5**, e1000489.
- De Martino,D. et al. (2013) Counting and correcting thermodynamically infeasible flux cycles in genome-scale metabolic networks. *Metabolites*, **3**, 946–966.
- 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.
- Gelius-Dietrich,G. et al. (2013) *sybil* - Efficient constraint-based modelling in R. *BMC Syst. Biol.*, **7**, 125.
- Holzhütter,H.-G. (2004) The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *Federation Eur. Biochem. Soc. J.*, **271**, 2905–2922.
- Lewis,N. E. et al. (2010) Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol. Syst. Biol.*, **6**, 390.
- Lewis,N.E. et al. (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat. Rev. Microbiol.*, **10**, 291–305.
- Mahadevan,R. and Schilling,C. (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.*, **5**, 264–276.
- Mo,M.L. et al. (2009) Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst. Biol.*, **3**, 37.
- Orth,J.D. et al. (2010) What is flux balance analysis? *Nat. Biotechnol.*, **28**, 245–248.
- Pál,C. et al. (2005) Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat. Genet.*, **37**, 1372–1375.
- Pál,C. et al. (2006) An integrated view of protein evolution. *Nat. Rev. Genet.*, **7**, 337–348.
- Plata,G. et al. (2010) Reconstruction and flux-balance analysis of the *Plasmodium falciparum* metabolic network. *Mol. Syst. Biol.*, **6**, 408.
- Price,N.D. et al. (2002) Extreme pathways and Kirchhoff's second law. *Biophys. J.*, **83**, 2879–2882.
- R Development Core Team. (2013) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

- Schellenberger, J. *et al.* (2010) BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions Database. *BMC Bioinformatics*, **11**, 213.
- Schellenberger, J. *et al.* (2011) Elimination of thermodynamically infeasible loops in steady-state metabolic models. *Biophys. J.*, **100**, 544–553.
- Schilling, C.H. *et al.* (2000) Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J. Theor. Biol.*, **203**, 229–248.
- Schuetz, R. *et al.* (2012) Multidimensional Optimality of Microbial Metabolism. *Science*, **336**, 601–604.
- Schuster, S. and Hilgetag, C. (1994) On elementary flux modes in biochemical reaction systems at steady state. *J. Biol. Syst.*, **2**, 165–182.
- Schuster, S. *et al.* (2008) Is maximization of molar yield in metabolic networks favoured by evolution? *J. Theor. Biol.*, **252**, 497–504.
- Wright, J. and Wagner, A. (2008) Exhaustive identification of steady state cycles in large stoichiometric networks. *BMC Systems Biol.*, **2**, 61.