Counting and correcting thermodynamically infeasible flux cycles in genome-scale metabolic networks

(Article begins on next page)

08 December 2024

Article

Counting and Correcting Thermodynamically Infeasible Flux Cycles in Genome-Scale Metabolic Networks

Daniele De Martino [1,2]**, Fabrizio Capuani** [1]**, Matteo Mori** [1]**, Andrea De Martino** [1,2,3,†,*] **and Enzo Marinari** [1,2,†,*]

[1] Dipartimento di Fisica, Sapienza Università di Roma, p.le A. Moro 2, Roma 00185, Italy

[2] Center for Life Nano Science@Sapienza, Istituto Italiano di Tecnologia, v. Regina Elena 291, Roma 00151, Italy

[3] CNR-IPCF, UOS di Roma, Dipartimento di Fisica, Sapienza Università di Roma, Roma 00185, Italy

[†] These authors contributed equally to this work.

[*] Authors to whom correspondence should be addressed;
E-Mails: andrea.demartino@roma1.infn.it (A.D.M.); enzo.marinari@uniroma1.it (E.M.);
Tel.: +39-06-4991-4394 (A.D.M); +39-06-4991-4363 (E.M.); Fax: +39-06-4957-697 (A.D.M.);
+39-06-9725-7050 (E.M.).

Abstract: Thermodynamics constrains the flow of matter in a reaction network to occur through routes along which the Gibbs energy decreases, implying that viable steady-state flux patterns should be void of closed reaction cycles. Identifying and removing cycles in large reaction networks can unfortunately be a highly challenging task from a computational viewpoint. We propose here a method that accomplishes it by combining a relaxation algorithm and a Monte Carlo procedure to detect loops, with *ad hoc* rules (discussed in detail) to eliminate them. As test cases, we tackle (a) the problem of identifying infeasible cycles in the *E. coli* metabolic network and (b) the problem of correcting thermodynamic infeasibilities in the Flux-Balance-Analysis solutions for 15 human cell-type-specific metabolic networks. Results for (a) are compared with previous analyses of the same issue, while results for (b) are weighed against alternative methods to retrieve thermodynamically viable flux patterns based on minimizing specific global quantities. Our method, on the one hand, outperforms previous techniques and, on the other, corrects loopy solutions to Flux Balance Analysis. As a byproduct, it also turns out to be able to reveal possible inconsistencies in model reconstructions.

---

1. Introduction

Starting from the discovery by Lavoisier concerning the relation between respiration and combustion, thermodynamics stands as a key physical framework for understanding metabolism and physiology, from single cell to whole organisms. When applied to a given metabolic reaction network, at the simplest level, thermodynamics requires that, in non-equilibrium steady states, fluxes of matter proceed downhill in the underlying Gibbs (free) energy landscape. Violations of this rule (which corresponds to nothing but the second law of thermodynamics) are signaled by the existence of unphysical cycles in flux configurations [1]. In the current era of metabolic genome-scale reconstructed networks, the implementation of such a constraint in computational models of a cell's metabolism has far-reaching implications [2], ranging from the physical feasibility of flux configurations [3] to the estimation of metabolite levels [4], the assignment of directionality for reactions and pathways [5] and the characterization of the overall chemical energy balance [6]. Accounting for thermodynamics in genome-scale models, however, poses considerable practical problems both for algorithms and for CPU costs.

The reference modeling scheme that we shall consider here is given by the so-called constraint-based models [7], widely employed in the literature to describe the operation of a biochemical reaction network at steady states with time-independent metabolite levels. While building a detailed model of metabolism presupposes knowledge of the kinetic parameters and reaction mechanisms [8], and should possibly take into account stochasticity [9] and spatial diffusion [10,11], constraint-based models focus on well-mixed non-equilibrium steady states (NESSs) for the reaction fluxes, to recover what fundamental information comes from the underlying stoichiometry alone. For a given stoichiometric matrix $\mathbf{S} = \{S_{mr}\}$ that accounts for the stoichiometric coefficient of metabolite m in reaction r (with the usual sign convention to distinguish products from substrates), a flux vector $\mathbf{v} = \{v_r\}$ represents a non-equilibrium steady state if it enforces the balance of metabolite levels $\mathbf{c} = \{c_m\}$, *i.e.*, if:

$$\dot{\mathbf{c}} \equiv \mathbf{S}\mathbf{v} = \mathbf{0} \tag{1}$$

In usual applications, physiological aspects constrain fluxes to vary with certain ranges, so that bounds of the type $v_r \in [v_r^{\min}, v_r^{\max}]$ are normally prescribed for every reaction, r. Such bounds may reflect, for instance, the fact that certain processes are known to be physiologically irreversible (e.g., $v_r \geq 0$) or are required to occur at precise rates (as can be the case for maintenance reactions). From a geometric point of view, under Equation (1) and the bounds on fluxes, the space of possible NESSs is represented by a convex polytope. If all flux configurations inside this volume could be considered as physically realizable solutions, one might assess the "typical" productive capabilities of the network by sampling them using a controlled algorithm [12]. Unluckily, this route often turns out to be computationally too expensive for large enough systems. Alternatively, one may search for the state(s) that maximize the value of certain biologically motivated objective functions, which can usually be cast in the form of a linear combination

of fluxes that represents the selective production of a given set of metabolites. The flux configurations that maximize such a linear functional can be retrieved with the methods of linear programming [13], the textbook case being growth yield maximization for bacterial cells in culture. Such a framework, known as Flux Balance Analysis (FBA) [14], has been shown to be predictive in many instances, even under genetic and/or environmental perturbations [15] (possibly with small modifications).

Solutions of Equation (1) are in general not guaranteed to be thermodynamically viable. Frameworks, like FBA, can however be modified to include thermodynamic constraints directly in order to generate thermodynamically viable flux configurations, for instance, by resorting to empirical data to estimate the chemical potentials of metabolites [16] and infer reaction reversibility more precisely [17,18]. As a matter of fact, a large part of thermodynamic inconsistencies appear to be due to fallacious direction assignments. Models of this type, however, require prior biochemical information that is often scarce or unavailable [19]. To overcome these difficulties, new methods were devised that detect infeasible loops leveraging only on the constraint based model, *i.e.*, on the structure on the metabolic network alone [20–22]. Although these methods release the need for experimental knowledge, the direct detection of infeasible loops is a computationally demanding task that limits their applicability. It is therefore important to devise algorithms that are able to identify and remove thermodynamic inconsistencies from solutions of Equation (1) or, more generally, from generic flux patterns.

Checking the thermodynamic feasibility of a flux pattern can be made straightforward. Denoting by $\mathbf{v}'$ a flux vector, from which we exclude uptakes and every reaction that cannot be associated directly with a thermodynamic constraint (like biomass production, "effective" reactions with non-integer stoichiometry or null fluxes), let us define the matrix $\mathbf{\Omega} = \{\Omega_{mr}\}$ with elements $\Omega_{mr} = -\text{sign}(v'_r)S_{mr}$. Note that the minus sign in the definition of $\mathbf{\Omega}$ is needed to connect the directions of the reactions to the corresponding Gibbs energy differences. The thermodynamic feasibility of $\mathbf{v}'$ is easily seen to be guaranteed (see the Supporting Text for a toy example) if a non-zero vector $\boldsymbol{\mu} = \{\mu_m\}$ (of chemical potentials) exists, such that [23]:

$$\boldsymbol{\mu}\mathbf{\Omega} > \mathbf{0} \tag{2}$$

(Note that, for physiologically realism, one would like the individual μ_m's to lie in specific ranges. As we shall not be concerned here with reconstructing the cellular Gibbs energy landscape [24], this aspect will be neglected in what follows.) Solving Equation (2) can be done very efficiently, e.g., *via* relaxation algorithms [13]. By Gordan's theorem of the alternatives (see e.g., [25]), if Equation (2) has no solution, then necessarily its dual system:

$$\mathbf{\Omega}\mathbf{k} = \mathbf{0} \tag{3}$$

with $\mathbf{k} = \{k_r\}$ possesses at least one non-zero solution with $k_r \geq 0$ for each r. It is easy to understand that such vectors $\mathbf{k}$ represent closed cycles of reactions that could in principle be able to perform work without using free energy, contradicting the laws of thermodynamics (note, however, that the topology of such cycles may turn out to be remarkably complex; see e.g., [24]). The problem posed by thermodynamics can then be seen as that of identifying and removing such loops.

Finding all cycles in a directed (bipartite) network is, at the heart, an integer programming problem in the NP-hard (Non-deterministic Polynomial time class) [26], which suggests that using deterministic algorithms to find loops in large enough networks may be unwise. However, for networks in which reliable prior thermodynamic information is available, the complexity of loop counting can be

significantly reduced, and indeed, in some cases, the problem has already been tackled (although, in our view, not fully solved) in genome-scale networks with some degree of success [24,27]. By contrast, in large networks lacking detailed thermodynamic information, like the human metabolic networks, the implementation of thermodynamic constraints requires the development of algorithms that are able to handle more difficult instances of the loop counting problem. Luckily, in many hard computational problems where the use of exact algorithms is prevented by CPU costs, stochastic methods have proven to be effective. A biologically relevant case is represented by the problem of sampling solutions of Equation (1), for which Monte Carlo [28,29] and message-passing techniques [30] are being employed instead of deterministic methods (as the latter presuppose the enumeration of a possibly exponential number of vertices of the polytope). It is simple to guess that a similar strategy might be employed (as we shall see, with some care) for the analysis of the solutions of Equation (3), *i.e.*, to identify reaction cycles.

The strategy we present here combines a relaxation algorithm and a Monte Carlo method to allow for the thorough analysis of thermodynamic infeasibilities on genome-scale metabolic networks of unprecedented size. More precisely, loops will be found by applying Monte Carlo to Equation (3) with a reduced search space obtained by analyzing how relaxation behaves when applied to Equation (2). Once a loop is found, it can be removed in several ways, provided they do not violate any of the constraints other than thermodynamic (e.g., mass balance). We shall discuss and compare different approaches: more precisely, a "local" rule that exploits, in essence, the fact that fluxes in cycles are defined up to a constant and a "global" rule, based on the minimization of an overall function of the fluxes. The method will be used to analyze different types of networks of a large size. Specifically, we shall first identify all loops in the metabolic network of *E. coli* [31], then focus on amending the FBA solutions of 15 different human metabolic network models derived from the genome-scale Reactome Recon-2 [32], all bearing a specified objective function. Such solutions turn out to be rich with infeasible cycles, which we are able to find and correct.

The structure and rationale of the method we propose are discussed in detail in Section 2, together with a brief summary of the network reconstructions we shall employ. Section 3 exposes our results, while our conclusions are reported in Section 4.

2. Materials and Methods

2.1. Materials: Metabolic Network Reconstructions

The human Reactome Recon-2 [32] has been reconstructed by a community that merged and integrated existing global human metabolic networks and transcriptional information on specific human cell types. Authors verified the quality of Recon-2 by determining how many tasks the network was able to perform. A task can be as simple as the transformation of a metabolite by a single enzyme or by a complex pathway—like fermentation or oxidative phosphorylation—or as complex as the production of the building blocks, energy, cofactors, *etc.* required for cell duplication, *i.e.*, biomass. For *E. Coli* and, in general, for unicellular organisms, biomass yield is a valuable objective function for the FBA framework [33], since its maximization essentially equals growth maximization at fixed nutrient intake.

Although it is unlikely that, in normal circumstances, cells in a multicellular organism maximize the biomass yield, we stick to it as the FBA objective function, as, for our purposes, the objective function can be seen merely as a tool to obtain motivated flux patterns for thermodynamic analysis.

In addition to the global reconstruction, [32] provides a collection of 65 drafts of cell-specific networks, which are derived from Recon-2 by means of an automatic procedure that utilizes proteomic data [33,34]. We focused on 15 networks with the ability to produce biomass, which we list in the first column of Table 1 together with the number of reactions (N) and metabolites (M) included in each case. We have used, in particular, the network reconstructions in SBML (System biology Markup Language) format from [32] and resorted to the COBRA (COnstraints Based Reconstruction Analysis) Toolbox [33] for the FBA analysis and to produce the stoichiometric matrix and the list of metabolites and reactions to be used in our analysis.

We have also analyzed the reconstructed metabolic network of the bacterium *E. coli* derived in [31], consisting of 2,382 reactions (including 305 uptakes) among 1,668 metabolites. In this model, 548 reactions are putatively reversible.

In each case, the key information we employed is encoded in the stoichiometric matrix $\mathbf{S}$.

2.2. Methods

2.2.1. Algorithm for Thermodynamic Analysis: Structure

Before describing the algorithm in detail, we briefly recall the idea behind the procedure. We do not directly assess whether the flux configuration is loop free, but we try to compute the chemical potentials that satisfy Equation (2), which is a computationally easier problem to solve. If such a a solution exists, we are guaranteed that the flux configuration does not contain infeasible loops. It can be demonstrated [24] that the relaxation method described below always converges polynomially to a solution; a lack of convergence signals the presence of infeasible loops. Even when the relaxation method does not converge, it still provides us with a list of reactions that are likely to contain infeasible loops. To this limited subset of reactions, we can directly apply algorithms to find loop-free solutions.

The overall structure of the algorithm is reported in Figure 1.

In a few words, and referring to points A, B, C.1, C.2 and D shown explicitly in the flow chart:

(A) Input: the input information includes a stoichiometric matrix, $\mathbf{S}$, a flux vector, $\mathbf{v}$ (e.g., a solution of FBA) and a prior vector, $\boldsymbol{\mu}$, of chemical potentials. Initialize an integer variable, t, at $t = 0$ (relaxation steps) and an empty list.

(B) Compute the matrix, $\boldsymbol{\Omega}$, and evaluate the thermodynamic constraints (2), *i.e.*, compute $\boldsymbol{\mu}\boldsymbol{\Omega}$. If they are satisfied, *i.e.*, if $\boldsymbol{\mu}\boldsymbol{\Omega} > \mathbf{0}$, go to (D); otherwise, register the least unsatisfied constraint (l.u.c.), *i.e.*, the value of the index, r, for which the corresponding components of the vector, $\boldsymbol{\mu}\boldsymbol{\Omega}$, is smallest (more negative). Insert it into the list, and increase the t variable by 1; if $t \leq T$, with T, a pre-defined large parameter, go to (C.1); otherwise, go to (C.2).

(C.1) Update the vector, $\boldsymbol{\mu}$, by performing a single step of the relaxation algorithm described in Section 2.2.2. ; update the list by inserting the new l.u.c. and go back to (B).

(C.2) Perform a Monte Carlo computation, as described in Section 2.2.3. , in order to find a solution of system Equation (3), namely $\Omega k = 0$, including only the reactions appearing in the list. Once a solution is found, correct the associated cycle as described in Section 2.2.4. and 2.2.5. ; re-initialize t, empty the list and go back to (B).

(D) Output: a thermodynamically feasible flux vector.

Figure 1. Flowchart of the algorithm for counting and removing cycles employed in this study. See text for details.
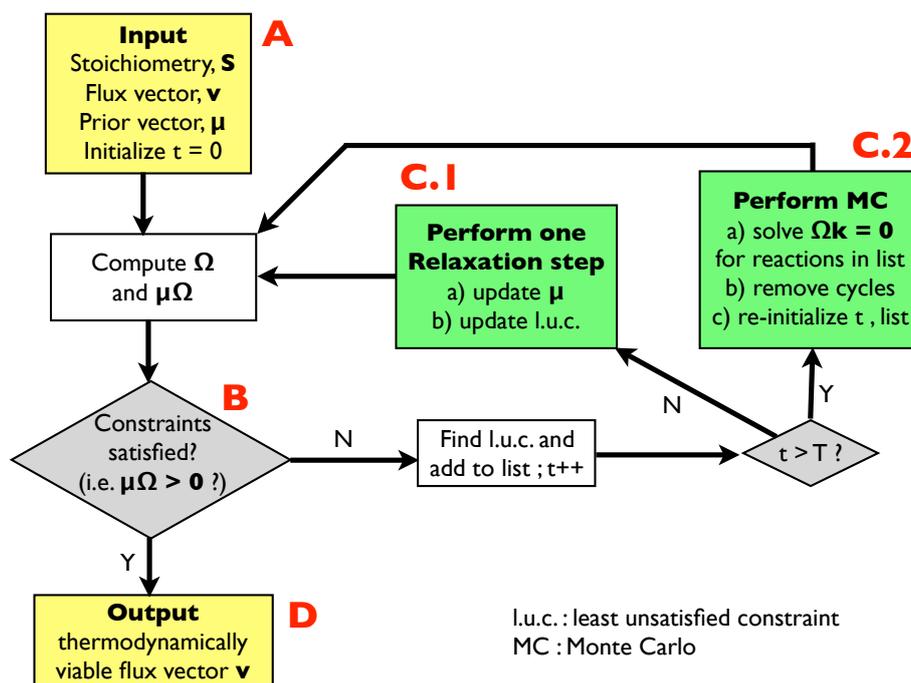
In the following sections, we shall describe the sub-procedures (relaxation method, Monte Carlo and cycle removal) of the algorithm in detail. A C++ code, which performs each of the above steps, is provided as Supporting Material. It is worth pointing out that the present study is not concerned with the calculation of realistic chemical potentials. Rather, we simply require their existence in order for the flux configuration to be feasible. In this case, the prior vector of chemical potentials can be arbitrary, e.g., constant or composed by independent and identically distributed random variables. If complemented with specific experimentally determined or computationally estimated priors for the chemical potentials, however, the relaxation method included in the above algorithm generates, as a by-product, a free energy vector compatible with the final flux configuration and can, thus, be employed to refine experimental data on the free energy of the formation of metabolites and/or on their levels [24]. Better priors ultimately allow one to obtain more precise estimates for the real chemical potentials, but are essentially irrelevant for the convergence of the relaxation method. In what follows, we shall neglect this aspect, which is discussed in depth in [24], and focus exclusively on the retrieval of cycles.

2.2.2. Checking Thermodynamic Viability by Relaxation

This routine, corresponding to point C.1 of the flow chart, allows one to retrieve a solution of Equation (2) starting from a vector of chemical potentials that is not a solution thereof. For simplicity, we construct an initial vector made of uniformly distributed random numbers. At any step, t, of the procedure, given a chemical potential vector $\boldsymbol{\mu}(t)$, the relaxation algorithm corrects the l.u.c. of Equation (2) through the dynamics defined by:

$$r_t = \arg\ \min_r \sum_m \Omega_{mr}\mu_m(t) \tag{4}$$

$$\mu_m(t+1) = \mu_m(t) + \alpha\Omega_{mr_t} \quad \forall m \tag{5}$$

where r_t is the index of the l.u.c. added to the list at the t-th step and $\alpha > 0$ is a constant. As explained in [24], the above step simply shifts chemical potentials in a direction that will improve the l.u.c. The parameter α can be chosen in different ways, from a suitably small constant (as in the so-called MinOver (minimum overlap) scheme [35]), to a quantity proportional to the amount by which the constraint is violated (as in the Motzkin scheme [13]). The Minover scheme returns a solution that is typically closer to the prior, but here, we utilized the faster Motzkin scheme. The above algorithm is known to converge to a solution of Equation (2), upon iteration, in polynomial time, if and only if a solution exists. If convergence fails, instead, by the Gordan theorem, the reaction pattern contains infeasible cycles. Convergence of relaxation therefore guarantees feasibility of a flux vector. In the presence of loops, an iteration of the above dynamics typically cycles among inconsistent constraints. Therefore, by keeping track of the l.u.c. over the iterations, *i.e.*, by recording the series $\{r_t\}$ for $t \geq 0$ (corresponding to the content of the list described in the flow chart), one can build a list of reactions that are candidates for being responsible for the failed convergence. If relaxation does not converge to a solution in a reasonable time (denoted as T above), we look for infeasible cycles, *i.e.*, for solutions of Equation (3), within such a restricted list. To this aim, we employ a Monte Carlo method.

2.2.3. Identifying Loops by Monte Carlo

As said above, cycles generically correspond to solutions of Equation (3) with $\mathbf{k} \geq \mathbf{0}$. As the stoichiometric coefficients are typically integers, one can focus on searching solutions with k_r non-negative integers for each r. To this aim, the following method (borrowed from the standard statistical physics toolbox) can be employed. Starting from Equation (3), note that the function:

$$E(\mathbf{k}) = \sum_m \left(\sum_r \Omega_{mr}k_r\right)^2 \tag{6}$$

vanishes when $\mathbf{k}$ defines a flux cycle. Because $E \geq 0$, then infeasible loops correspond to the minima of E, and loop finding amounts to locating the minima of E in the search space $k_r \in \{0, 1, 2, \ldots\}$ for each r. Monte Carlo methods are ideally suited to tackle this type of problem [36]. In brief, such methods (the most famous of which is possibly the Metropolis scheme) generically generate vectors $\mathbf{k}$ distributed according to:

$$P(\mathbf{k}) \propto e^{-\beta E(\mathbf{k})} \tag{7}$$

where $\beta > 0$ is an externally fixed parameter. When $\beta \to \infty$, the above measure concentrates around the minima of E. One possibility to make sure that large enough values of β are reached is to initialize the Monte Carlo simulation at some small value of β and, then, increase β in a controlled way, initializing each time from the configuration retrieved at the previous value of β ("simulated annealing"). This is precisely the approach we have employed here: in order to identify the infeasible loops, we have performed iterated Metropolis-based annealings to minimize the fictitious "energy" (6). (The increase in performance warranted by the annealing procedure compared to the simple Metropolis scheme at a fixed temperature is discussed in the Supporting Text.)

It is worth noting that, based on the above discussion of the relaxation method, the number of reactions to be included in the above procedure equals the number of distinct reactions appearing in the list, which is usually much smaller than N. To give an idea, in the study of *E. coli*, whose results are reported below, our lists ended up containing at most 50 reactions, to be compared with the over 2,000 that form the genome-scale reconstruction. Hence, the computational costs of the Monte Carlo step of our algorithms are overall modest.

2.2.4. Correcting the Flux Configuration: Local Strategy

Once a flux cycle has been identified, there are multiple ways to remove it and re-organize the flux pattern, while still preserving all constraints and, eventually, the values of objective functions.

To clarify the situation, consider the following simple example with four reactions, pictured in Figure 2.

Figure 2. Example of a toy reaction network. The black dots are two metabolites, to each of which corresponds a mass balance constraint. Each line is labeled with the name of the flux carried by a reaction, and the arrow indicates the conventional forward direction of the fluxes. Evidently, a thermodynamically infeasible cycle is present if v_1 and v_2 have the same sign.
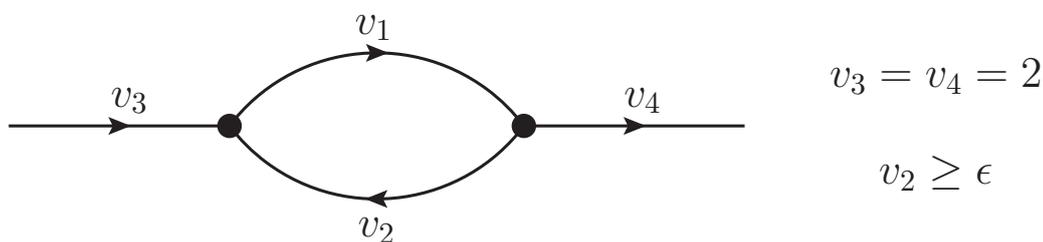
v_1 and v_2 are "internal" fluxes, while v_3 and v_4 are an intake and an outtake flux, respectively. The stoichiometric matrix $\mathbf{S}$ and the flux vector $\mathbf{v}$ reads:

$$\mathbf{S} = \begin{pmatrix} -1 & 1 & 1 & 0 \\ 1 & -1 & 0 & -1 \end{pmatrix}, \qquad \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix} \tag{8}$$

With this formalism, the mass balance constraints (1) are given by the homogeneous equation $\mathbf{S}\mathbf{v} = \mathbf{0}$. We also fix the value of the uptakes to, for example, $v_3 = v_4 = 2$, and add a lower bound on the first flux, $v_2 \geq \epsilon$ (with ϵ as a constant).

After the elimination of the uptakes, the internal stoichiometric matrix and flux vector (which, for clarity, in this section, we denote as $\mathbf{S}^{int}$ and $\mathbf{v}^{int}$) are given by:

$$\mathbf{S}^{int} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}, \qquad \mathbf{v}^{int} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \tag{9}$$

Since we fixed the values of v_3 and v_4, we can move them to the right hand side of the mass balance constraints, obtaining:

$$\mathbf{S}^{int}\mathbf{v}^{int} = \mathbf{u}, \qquad \mathbf{u} = \begin{pmatrix} -v_3 \\ v_4 \end{pmatrix} \tag{10}$$

Equation (10) must be solved together with the constraint $v_2 \geq \epsilon$, which we will treat separately.

Given a solution $\mathbf{v}$ of Equation (10), any linear combination of the form:

$$\mathbf{v}' \equiv \mathbf{v}'(\mathbf{L}) = \mathbf{v} + \sum_a L^a \mathbf{n}^a \tag{11}$$

still satisfies the mass balance Equation (10), provided $\mathbf{n}^a$ is in the right null space of the internal stoichiometric matrix, *i.e.*, it is a solution of $\mathbf{S}^{int}\mathbf{n}^a = 0$, and $\mathbf{L} = \{L^a\}$ is a family of real numbers. If we now recall that loops $\mathbf{k}$ are non negative solutions of Equation (3), *i.e.*, of $\mathbf{\Omega}\mathbf{k} = \mathbf{0}$, where the matrix $\mathbf{\Omega}$ is built from $\mathbf{S}^{int}$ as $\Omega_{mr} = -\text{sign}(v_r)S_{mr}^{int}$, we see that, by construction, the relation $n_r^a = \text{sign}(v_r)k_r^a$ connects infeasible loops $\mathbf{k}^a$ ($a = 1, 2, \ldots$) and the solutions of $\mathbf{S}^{int}\mathbf{n}^a = 0$. In other terms, the presence of cycles causes a degeneracy in flux patterns, as there are many ways to assign fluxes to reactions in a cycle.

Equation (11) can be used to correct the infeasible loops, while still satisfying all mass balance constraints. The simplest possible correction scheme is based on the idea that by properly fixing the value of the coefficients, L^a, one can lift the degeneracy and rid the flux configuration of loops. There is, however, a major caveat. To make it explicit, we note that (a) the signs of the fluxes v_r' depend on the choice of the coefficients $\mathbf{L}$, so that the matrix $\mathbf{\Omega}$ will also depend on it; and (b) it is not guaranteed that the new fluxes $\mathbf{v}'$ will vary within the same bounds as $\mathbf{v}$. This means that equation (3) must be solved together with all other constraints which are not related to the stoichiometry, such as sign constraints. We will write these constraints in a general fashion as $\mathbf{A}\mathbf{v}' \geq \mathbf{b}$. In our example, this matrix inequality reduces to $v_2 \geq \epsilon$.

With this notation, the space of vectors $\mathbf{L}$ yielding thermodynamically feasible solutions is given by $\mathcal{C}_1 \cap \mathcal{C}_2$, where:

$$\mathcal{C}_1 = \{\mathbf{L} : \mathbf{A}\mathbf{v}' \geq \mathbf{b}\} \tag{12}$$

$$\mathcal{C}_2 = \{\mathbf{L} : \nexists\, \mathbf{k} \geq 0 \ \text{such that} \ \mathbf{\Omega}(\mathbf{L})\mathbf{k} = 0\} \tag{13}$$

The first set contains all constraints that are not related to stoichiometry, while the second one contains the thermodynamic ones. Now, unluckily, these two sets may or may not have points in common,

depending on the properties of the network and on the additional constraints (in other words, it may not be possible to choose **L** properly).

Suppose now that, in our example, we are given the flux vector $\mathbf{v}^\star = (3,1)$ as the solution to Equation (10). We see that the vector $\mathbf{n} = (1,1)$ is in the null space of $\mathbf{S}^{int}$ and that the new vector $\mathbf{v}' = \mathbf{v}^\star + L\mathbf{n}$ still satisfies (10). Furthermore, since both $v_1^\star$ and $v_2^\star$ are positive, $\mathbf{n}$ itself identifies a loop, *i.e.*, it is a solution of (3) (for, in this case, $\mathbf{\Omega} = -\mathbf{S}^{int}$). It can be easily checked that, in this example, as long as the constraint $v_2 \geq \epsilon$ is not taken into account, we can pick L in the interval $[-3, -1]$ to get rid of the cycle. The choice is arbitrary and produces a fully directional flux pattern, such that reactions v_1 and v_2, if both active, operate in the same direction. The constraint $v_2 \geq \epsilon$, however, implies $L \geq \epsilon - 1$. The sets, $\mathcal{C}_1$ and $\mathcal{C}_2$, are then given by:

$$\mathcal{C}_1 = \{L : L \geq \epsilon - 1\} \tag{14}$$

$$\mathcal{C}_2 = \{L : -3 \leq L \leq -1\} \tag{15}$$

Therefore, we see that if $\epsilon < 0$, there are infinitely many values of L that remove the cycle, whereas the cycle cannot be removed if $\epsilon > 0$. In other words, each time a flux is constrained to keep a pre-defined sign, there is no guarantee that the loops involving this reaction can be corrected by simply lifting the flux degeneracy associated with them. The value $\epsilon = 0$ plays here a particular role, since $v_2 \geq 0$ is an irreversibility constraint. In this case, the intersection of the two sets above imposes $L = -1$. This is indeed the most common kind of constraint on the fluxes and allows for the simplest unambiguous loop removal strategy: setting the smallest flux to zero without changing signs to any of the other fluxes involved (as occurs in the present case upon choosing $L = -1$).

From this last observation, we can deduce the following general loop removal strategy (which we refer to as the "local" correction strategy): for a given loop $\mathbf{k}^a$, choose the value of L^a that sets to zero the flux of the reaction whose absolute value is the smallest, *i.e.*,:

$$L^a = -\min_{r\,:\,k_r^a > 0} \frac{|v_r|}{k_r^a} \tag{16}$$

With this choice, every constraint of the form $v_r \geq 0$ will still be satisfied, and at least one loop will be removed. This strategy tends, in a sense, to minimize the distance between the original flux pattern and the corrected one. (We will quantify this aspect more precisely below.) On the other hand, if bounds like $v_r \geq \epsilon$ with $\epsilon > 0$ are present, a more careful analysis is required. We finish by noting that the most important instance of a constraint of the latter type in models of metabolism is represented by the ATP maintenance flux.

2.2.5. Correcting the Flux Configuration: Global Strategy

Other possible loop-removing procedures are based on the minimization of some norm of the fluxes, as suggested in [37]. Let us elaborate this idea further and consider the function $Q_p(\mathbf{v}) = \sum_r |v_r|^p$ with $p \geq 1$, representing, for different p's, different norms of the flux vector $\mathbf{v}$ (Q_1 is the so-called "Taxicab" norm, Q_2 is the square of the Euclidean norm, *etc.*). Suppose we have an FBA solution $\mathbf{v}^\star$ that minimizes Q_p. If $\{\mathbf{n}^a\}$ is the set of all null space vectors of the stoichiometric matrix, we can

construct a new solution $\mathbf{v}' = \mathbf{v}^\star + \sum_a L^a \mathbf{n}^a$ and compute the partial derivative of Q_p with respect to a coefficient, L^a. The derivatives vanish when evaluated at $\mathbf{L} = \mathbf{0}$:

$$\left.\frac{\partial}{\partial L^a} Q_p\left(\mathbf{v}^\star + \sum_b L^b \mathbf{n}^b\right)\right|_{\mathbf{L}=\mathbf{0}} = \sum_r p|v_r^\star|^{p-1}\mathrm{sign}(v_r^\star)n_r^a = 0 \tag{17}$$

We see immediately that the quantities $k_r^a = \mathrm{sign}(v_r^\star)n_r^a$ cannot have a definite sign. Restricting the sum to all terms with $n_r^a \neq 0$, we have two cases:

- If at least one of the fluxes $v_r^\star$ is zero, this reaction cannot be involved in any cycle. In particular, $\mathbf{n}^a$ is not associated with a loop.

- If all fluxes are non-zero, the vector $\mathbf{k}^a$ cannot have a definite sign (positive or negative), since the sum of its entries, namely Equation (17), weighted with some positive coefficients, is zero.

Therefore, the vector $\mathbf{v}^\star$ that minimizes Q_p does not contain cycles.

The argument can be easily extended to include irreversibility constraints. Let $\mathbf{v}^\star$ denote the flux configuration that minimizes $Q_p(\mathbf{v})$ with irreversibility constraints $v_r \geq 0$ for the reactions, r, belonging to the set $\mathcal{I} = \{r_1,\ r_2,\ \ldots\}$. We shall instead denote by $\mathcal{I}_0 = \{r'_1,\ r'_2,\ \ldots\} \subseteq \mathcal{I}$ the set of irreversible reactions for which $v_r = 0$ in $\mathbf{v}^\star$. Clearly, $\mathbf{v}^\star$ also minimizes Q_p subject to the stronger constraints $v_r = 0$ for $r \in \mathcal{I}_0$ and $v_r > 0$ for $r \in \mathcal{I} \setminus \mathcal{I}_0$. Given this, one can now proceed along the same lines as before, because, for any vector $\mathbf{n}^a$ in the null space of $\mathbf{S}^{int}$:

- If some reaction, r, for which $n_r^a \neq 0$ is forced to have zero flux, since $r \in \mathcal{I}_0$, then $\mathbf{n}^a$ is not associated with a cycle;

- Otherwise, we can demonstrate that $\mathbf{n}^a$ does not correspond to a cycle by taking the partial derivative of $Q_p(\mathbf{v}^* + L^a \mathbf{n}^a)$ as done above.

Problems may arise, as before, when boundary conditions like $v_r \geq \epsilon$ with $\epsilon > 0$ have to be considered. In particular, if the flux of a variable thus bounded is fixed to take the value ϵ, there is the possibility that the cycle cannot be removed. In the example discussed above (see Figure 2):

- If $\epsilon < -1$, then $v_2 > \epsilon$ and the Q_p minimization yields $\mathbf{v}^\star = (1, -1)$;

- If $-1 \leq \epsilon \leq 0$, then $v_2 = \epsilon$, but the flux configuration $\mathbf{v}^\star = (1 - \epsilon, \epsilon)$ is still feasible (in particular, the configuration is feasible for $\epsilon = 0$);

- If $\epsilon > 0$, then $v_2 = \epsilon$, and the optimal flux configuration is not feasible.

In summary, the global minimization of the norm Q_p produces thermodynamically feasible flux patterns, provided they are allowed by the constraints. If not, the minimization can get rid of all loops not involving reactions constrained to keep the same sign. We shall term the cycle-removal strategy based on minimizing a norm as the "global" strategy.

3. Results

3.1. A Test: Identifying Infeasible Loops in the E. Coli Network iAF1260

As a proof of principle, we have applied our method to search and enumerate all independent infeasible loops of a large metabolic network reconstruction for the bacterium *E. Coli*, the iAF1260 [31]. Other authors have attempted to solve the same enumeration problem before (see, e.g., [27]). We note, however, that our method is radically different in that we make use of the theorem of alternatives and do not directly search for loops on the graph, which is the more standard route [38], or rely on subsequent optimizations that reduce the search space [27,39]. In the present case, we characterize cycles in an ensemble of net-flux patterns generated randomly by assigning a specific operating direction to each reaction according to its reversibility. More precisely, random flux patterns are generated by simply assigning an operation direction for each reaction as follows: if the reaction is irreversible, we pick the allowed direction; if the reaction is reversible, we select the forward or the reverse direction randomly with a probability of 1/2 (note that direction assignments suffice to pose the problem of thermodynamic feasibility. In this way, *all* reactions are active, a worst-case scenario with respect to a growth-yield optimizing state that normally only requires the operation of around 30% of the reactions, implying (in our case) a much larger number of loops and, in principle, higher computational costs for loop counting. For each configuration, we look for and eliminate cycles until the material flow is thermodynamically consistent, recording the cycles that we have detected. Finally, we keep only independent loops by applying Gaussian elimination (*i.e.*, we exclude from our list loops that can be decomposed as the sum of, say, two simpler loops).

In Figure 3, we display the number of independent loops that we identify as a function of the number of random configurations tested.

Figure 3. The number of independent loops identified in the metabolic network of *E. Coli* iAF1260 as a function of the number of random configurations tested. Results are obtained by jackknifing over 10 configurations. The black line represents the average number of loops, while the blue lines represent the extremes of the error bars at each point.
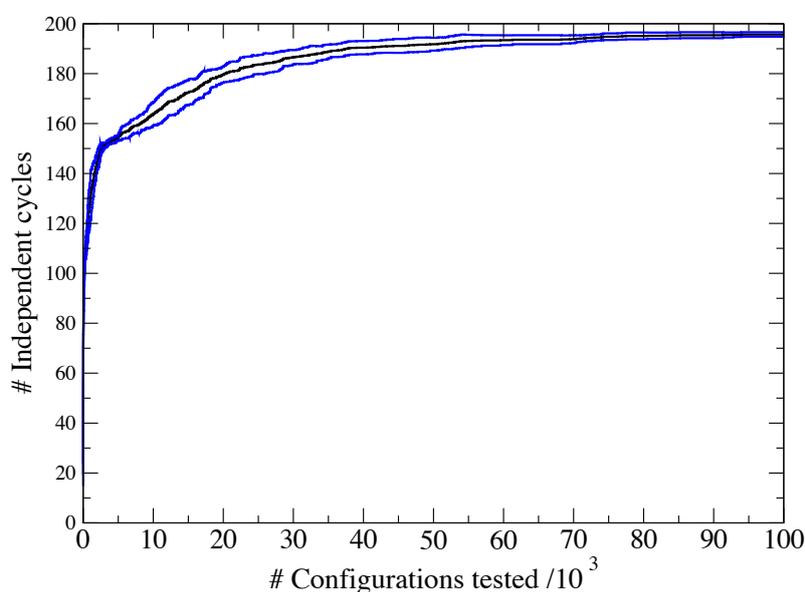
We identify 196 loops (189 of which turn out to be of a size of three or more) after having generated about 80,000 random configurations, and no new loops appear upon enlarging the test ensemble. The loops thus found are listed in Supporting File 1, and a histogram of the cycle lengths (in terms of the number of reactions involved) is displayed in Figure 4.

Figure 4. Histogram of the length (number of reactions involved) of the 196 independent cycles detected in *E. Coli* iAF1260.
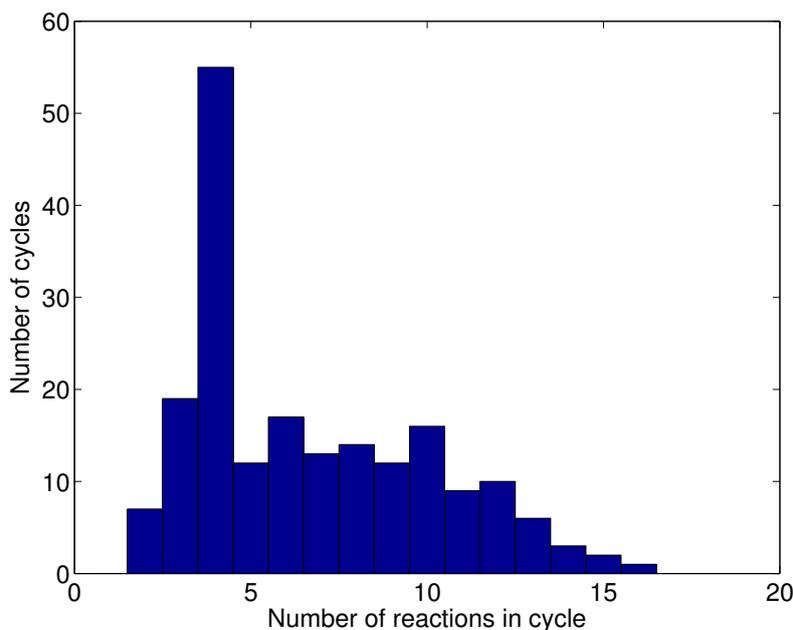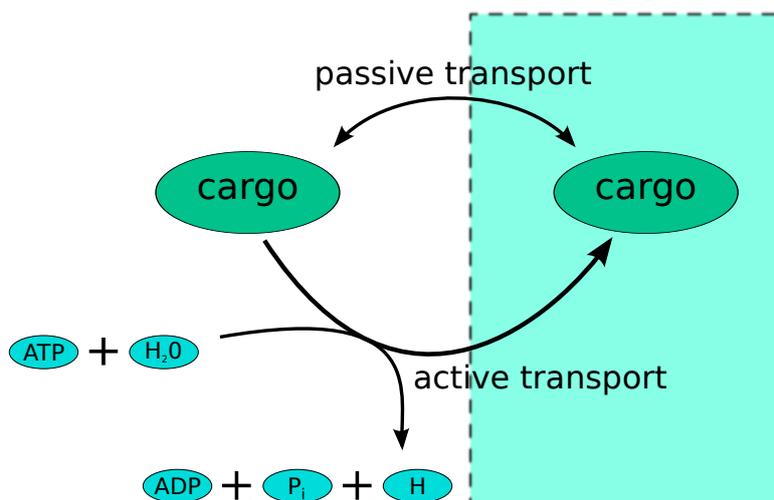

We note that, in [27], 591 cycles were identified, 564 of which are, however, formed by two reactions, mostly originating from the fact that reversible reactions were, in that study, split in two separate processes (forward and reverse). Therefore, only 27 of those cycles were formed by three reactions or more. Because we do not split reversible reactions, we find only seven cycles of a length of two and 189 loops of a length at least equal to three. We note that these 189 cycles span 396 reactions altogether. This suggests that, for the technique employed in [27], some loops were undetectable once the exhaustive search had been restricted to 50 reactions. We stress, however, that the procedure discussed in [27] is, in principle, exact, and once the restriction is removed, it might be able to identify more cycles involving at least three reactions.

3.2. Inconsistencies in the FBA Solution for the Overall Human Reactome Recon-2

We now move on to the identification of thermodynamic infeasibilities in the human Reactome Recon-2 [32]. In specific, we have analyzed the feasibility of flux patterns defined by solving FBA on the entire Reactome, using the "biomass" reaction that comes with the reconstruction as the objective function. This section provides a concrete example of an inconsistency that is unrecoverable without correcting basic structural information concerning the network. It should be kept in mind, however, that the physiologically relevant metabolic networks that can be obtained from Recon-2 are the cell-type specific ones, which will be discussed in the following section.

As almost all metabolic objective functions, the biomass reaction of Recon-2 contains ATP hydrolysis, representing the energetic requirements associated with cell duplication, which are not explicitly accounted for by the flux organization. As such, requirements are typically large: the stoichiometry of ATP in the biomass reaction is often two orders of magnitude larger than that of the other chemical species. Hence, ATP tends to be the limiting factor for biomass production, and FBA solutions will often organize metabolic fluxes, so as to produce as much ATP as possible. This however turns out to lead, in Recon-2, to a violation of thermodynamics. In particular, in the FBA solution for Recon-2, we detect a huge number of cycles involving the active and passive transport of a metabolite through a membrane, as, e.g., for the transport of stearoyl-CoA (stcoa) from cytosol (c) to peroxisomes (x), namely (see Figure 5).

Figure 5. Typical structure of an infeasible loop created by two reversible transports across a membrane-enclosed compartment: a passive one (by diffusion) and an active one (requiring the expenditure of energy). If the active process is allowed to reverse and the cargo re-enters the cell or compartment *via* diffusion, an infeasible loop is generated that builds ATP from ADP without energetic costs.

$$stcoa[c] \quad \rightleftarrows \quad stcoa[x] \tag{18}$$

$$adp[c] + h[c] + pi[c] + stcoa[x] \quad \rightleftarrows \quad h2o[c] + atp[c] + stcoa[c] \tag{19}$$

Note that both reactions are listed as reversible. When the chemical potential difference drives stcoa from peroxisome to cytoplasm, the cell can actively transport stearoyl-CoA to peroxisomes by consuming ATP. By reverting both reactions, however, the cell could produce ATP at no expense. This is precisely the type of solution that we obtain when we maximize the biomass yield in Recon-2.

ATP-coupled reactions are a common, though not the only, source of thermodynamic inconsistencies that can be spotted in Recon-2 (see Supporting File 2 for the complete list of cycles we identified in the Recon-2-derived cell-type specific networks. It is however important to stress that they are spurious and may be identified easily by complementing Recon-2 with a maintenance reaction that mimics the energy

expenditure associated with basal processes (similar to those that are present in bacterial metabolic networks) and even cured automatically (or with an automated procedure) by fixing the directionality of active transports directly in the reconstruction (when possible).

3.3. Correcting Infeasible Loops in FBA Solutions for Cell-Type Specific Human Metabolic Networks

In this section, we focus on finding and correcting infeasible loops in FBA solutions of the cell-type specific human metabolic networks obtained by Recon-2. We have restricted our attention to 15 networks carrying an objective function, representing, respectively, cerebral cortex neuronal cell, liver bile duct cell, cervix uterine squamous epithelial cell, kidney tubule cell, gall bladder cell, lung macrophage, small intestine glandular cell, rectum glandular cell, smooth muscle cell, urinary bladder urothelial cell, pre- and post-menopause uterus glandular cell, pancreatic exocrine glandular cell, tonsil germinal cell and squamous epithelial cell. We first computed the FBA solutions for each of the networks *via* the COBRA Toolbox [33]. We have computed optimal solutions with respect to the biomass objective function. The choice is mainly motivated by the fact that maximizing the biomass yield represents a network-wide goal with respect to the more specific tasks described by other objective functions included in the reconstructions. We stress however that for our present purposes, the objective function merely provides a means of obtaining flux patterns; hence, the particular choice we made is immaterial for the problem we consider. Subsequently, we identified infeasible loops using the method described in Section 2.2.3. and, finally, corrected thermodynamic inconsistencies using both the local and global strategies described in Sections 2.2.4. and 2.2.5. (in the latter case, minimizing the Q_1 norm, while fixing sinks, uptakes and objective function to the values of the FBA solution). In particular, with the local strategy, we eliminate one infeasible loop at a time, making sure that no constraint is violated by the corrected solutions, including the value of the objective function. We note, however, that the local strategy does not return a unique thermodynamically consistent network, since the final flux pattern may depend on the order with which loops are removed. We shall see that, quite generically, this strategy produces flux patterns that are more similar to the original (infeasible) solutions than those generated by the global correction strategy.

Results are shown in Table 1, where we list different topological quantities (specifically, the overall number of reactions and metabolites and the number of reactions carrying a non-zero flux and that of metabolites that are produced and consumed by at least one reaction) for the original (infeasible) FBA solutions and for the corrected flux patterns, both for the local and global strategies, as well as the number of loops in the FBA solutions that need to be corrected by the local strategy. One sees that the local strategy typically needs to resolve several hundreds of inconsistencies in order to obtain viable solutions and that correction strategies enforce a reduction in the number of active processes, which in certain cases, can be rather dramatic. Supporting File 2 lists the cycles we identified and corrected in each of the 15 metabolic networks we have analyzed. To quantify more precisely the similarity between the solutions thus obtained, we have measured the 'overlap' parameter defined as follows: given two flux configurations, $\mathbf{v}^a = \{v_r^a\}$ and $\mathbf{v}^b = \{v_r^b\}$, we let:

$$q_{ab} = \frac{2}{N} \sum_{r=1}^{N} \frac{v_r^a v_r^b}{(v_r^a)^2 + (v_r^b)^2} \qquad (20)$$

Clearly, $q_{ab} = 1$, if $\mathbf{v}^a = \mathbf{v}^b$, while the more different fluxes are in the two solutions the smaller q_{ab} gets, until $q_{ab} = -1$, if $\mathbf{v}^a = -\mathbf{v}^b$. Larger values of q_{ab}, therefore, generically point to the fact that the two flux vectors are more similar also in terms of their directions. (Note that in computing (20), one should account for the fact that a flux that is null in both solutions contributes one to the above sum. In this study, for numerical reasons, a flux, v_r, is considered to be null whenever $v_r < v_0$, where v_0 is a (small) threshold. Results have been obtained with $v_0 = 10^{-6}$, but they are robust to changes in this value. Values of the overlaps between the three solutions we consider (original FBA, FBA corrected by the local strategy, FBA corrected by the global strategy) are also displayed in Table 1, clarifying that the local strategy applied to our sample always generates flux patterns that are closer to the original (infeasible) solution than those obtained by the global strategy. Nevertheless, the overlap between the locally- and globally-corrected solutions can also be rather large in some cases, suggesting that a common physical, possibly variational, requirement may underlie, to some extent, the two criteria.

The final column of Table 1 shows the sign of the Gibbs energy change of ATP hydrolysis that is obtained in the solution corrected by the global strategy. This provides an interesting check of physiologic consistency, as solutions should be compatible with a spontaneous ATP hydrolysis *in vivo* (*i.e.*, with a negative Gibbs energy difference). We find that only for five models does Q_1 minimization provide (thermodynamically feasible) flux configurations carrying a negative Gibbs energy difference for ATP hydrolysis. A possible, simple to obtain improvement of the method we present indeed includes taking into account physiological aspects when correcting a flux configuration. We stress once more, however, that these types of infeasibilities are due to inconsistent constraints or wrong reversibility assignments that prevent the existence of feasible, energetically realistic flux patterns and can be eliminated already at the stage of network reconstruction. Our main goal here was to show that our method is capable of identifying and correcting loops. By this type of example, we prove that it can furthermore point to possible limitations of the current models.

4. Discussions

Accounting for thermodynamic constraints in stoichiometry-based flux models, though potentially highly rewarding (in terms of the possibility to predict metabolite levels, chemical potentials, reaction free energies and reversibility), is a generically hard task. Methods that integrate directly with the constraints defining the space of viable fluxes are often computationally intensive and either presuppose prior biochemical knowledge or lead to a considerable increase in the number of parameters (or both). The technique presented here makes use of stoichiometry alone (hence, it is essentially a topological method) and allows us to accomplish two goals: on the one hand, counting and listing the infeasible reaction cycles that spur flux configurations derived from thermodynamics-free models; on the other, correcting such infeasibilities in a physically motivated manner. Indeed, we have first analyzed the genome scale metabolic network reconstruction iAf1260 of the bacterium *E. coli*. By simply recording the cycles found in randomly generated flux patterns we are able to uncover a much larger set of (much more complex) loops than previously obtained, also involving a much larger overall number of processes, comparing, in particular, with [27] (in this sense, outperforming previously employed methods). In passing, we note that our method comes with a certificate of completeness for the set of

cycles, which was previously unavailable. Secondly, after showing that cycles plague FBA solutions for the metabolic networks of several different types of human cells (all retrieved from the human Recon-2 Reactome), we have applied our loop-removal strategies in order to obtain thermodynamically viable flux patterns that both preserve the basic constraints of FBA, as well as the value of the objective function. In doing so, some inconsistencies in the reconstructions have been identified, which can easily be eliminated at the level of network building. Quite importantly in our view, we have also discussed the possibility to employ global variational criteria to generate thermodynamically feasible flux configurations. In particular, generalizing a previous observation, we have proven that flux patterns that minimize the p-norms of the fluxes are thermodynamically viable, provided they are allowed by the constraints. Otherwise, this idea can be used (with some care) to remove cycles that do not involve reactions that cannot be inverted or silenced.

The work presented here extends and improves over previous studies and takes several steps to suggest controlled and motivated methods to deal with thermodynamic inconsistencies in large networks of biochemical reactions. Further improvements along the lines discussed above (requiring, e.g., more precise physiological constraints) are clearly possible. Most promisingly, however, we believe that work directed at enhancing the integration of thermodynamic constraints into flux analysis would be extremely important in light of the current efforts aimed at increasing the scope, reach and predictive power of computational models of cellular metabolism. In absence of sufficiently detailed biochemical information about metabolite levels *in vivo* or chemical potentials, general stoichiometry-based techniques must be expected to play a key role in this endeavor.

Table 1. Overview of the results obtained for the human tissue-specific metabolic networks (with the biomass objective function). Columns are as follows. Cell type: abbreviations for the tissue-specific metabolic networks examined, for the full names please refer to the text (head of Section 3.2). N and M: overall number of reactions and metabolites appearing in the network. N_{FBA} and M_{FBA}: number of active reactions and produced/consumed metabolites in the Flux Balance Analysis (FBA) solution. # cycles: number of cycles that the local strategy needs to correct. N_{local} and M_{local}: number of active reactions and produced/consumed metabolites in the FBA solution corrected by the local strategy. N_{global} and M_{global}: the number of active reactions and produced/consumed metabolites in the FBA solution corrected by the global strategy. $q_{FBA,local}$: overlap between the FBA solution and the solution corrected by the local strategy. $q_{FBA,global}$: overlap between the FBA solution and the solution corrected by the global strategy. $q_{local,global}$: overlap between the FBA solution corrected by the local and global strategies. ΔG sign: sign of the free energy difference obtained for the ATP hydrolysis in the solution obtained *via* the global correction strategy.

Cell type	N	M	N_{FBA}	M_{FBA}	# cycles	N_{local}	M_{local}	N_{global}	M_{global}	$q_{FBA,local}$	$q_{FBA,global}$	$q_{local,global}$	ΔG sign
Bile duct	2,076	1,445	1,009	743	215	516	554	367	476	0.706	0.559	0.781	+
Cer. cortex	2,169	1,494	1,231	898	358	818	767	257	320	0.750	0.448	0.629	+
Cerv. ut.	1,774	1,171	1,046	780	194	562	620	339	380	0.666	0.480	0.735	-
Gall blad.	3,073	2,159	1,666	1,284	385	1,514	1,227	254	356	0.751	0.471	0.521	+
Kidney	3,176	2,212	1,695	1,285	414	1,423	1,196	142	449	0.759	0.469	0.551	+
Lung macroph. .	2,810	1,991	1,313	960	223	817	779	606	587	0.765	0.681	0.849	-
Pancreas	2,821	1,951	1,319	948	409	814	797	225	534	0.756	0.534	0.701	+
Rectum	2,976	2,041	1,328	1,135	406	989	1017	259	399	0.765	0.560	0.670	-
Small int.	3,179	2,213	1,385	1,192	405	836	1023	185	206	0.776	0.578	0.745	+
Smooth muscle	1,806	1,222	1,042	796	184	579	607	314	320	0.677	0.501	0.747	+
Tonsil ger.	2,126	1,421	1,178	884	405	881	764	357	412	0.667	0.503	0.644	-
Tonsil sqam.	2,573	1,718	1,719	1,250	423	1,455	1,188	301	403	0.718	0.334	0.430	+
Urot. blad.	2,874	1,965	1,597	1,308	219	1,111	1,158	148	686	0.760	0.450	0.613	+
Uterus post-m.	2,773	1,973	1,266	1,095	305	736	927	303	389	0.763	0.578	0.757	+
Uterus pre-m.	2,793	1,982	1,376	1,157	208	924	1022	259	582	0.785	0.507	0.658	+

Acknowledgments

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Price, N.; Famili, I.; Beard, D.; Palsson, B. Extreme pathways and kirchhoff's second law. *Biophys. J.* **2002**, *83*, 2879–2882.
2. Soh, K.; Hatzimanikatis, V. Network thermodynamics in the post-genomic era. *Curr. Opin. Microbiol.* **2010**, *13*, 350–357.
3. Beard, D.; Babson, E.; Curtis, E.; Qian, H. Thermodynamic constraints for biochemical networks. *J. Theor. Biol.* **2004**, *228*, 327–333.
4. Hoppe, A.; Hoffmann, S.; Holzhutter, H. Including metabolite concentrations into flux balance analysis: Thermodynamic realizability as a constraint on flux distributions in metabolic networks. *BMC Syst. Biol.* **2007**, 1, 23:1–23:12.
5. Qian, H.; Beard, D. Thermodynamics of stoichiometric biochemical networks in living systems far from equilibrium. *Biophys. Chem.* **2005**, *114*, 213–220.
6. Beard, D.; Liang, S.; Qian, H. Energy balance for analysis of complex metabolic networks. *Biophys. J.* **2002**, *83*, 79–86.
7. Palsson, B.O. *Systems Biology: Properties of Reconstructed Networks*; Cambridge University Press: Cambridge, NY, USA, 2006.
8. Bowden, A.C. *Fundamentals of Enzyme Kinetics*; Wiley-Blackwell: Weinheim, Germany 2013.
9. Ge, H.; Qian, M.; Qian, H. Stochastic theory of nonequilibrium steady states. Part II: Applications in chemical biophysics. *Phys. Rep.* **2012**, *510*, 87–118.
10. Frey, E.; Kroy, K. Brownian motion: A paradigm of soft matter and biological physics. *Ann. Phys.* **2005**, *14*, 20–50.
11. Beg, Q.; Vazquez, A.; Ernst, J.; de Menezes, M.; Bar-Joseph, Z.; Barabási, A.L.; Oltvai, Z.N. Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 12663–12668.
12. De Martino, A.; Marinari, E. The solution space of metabolic networks: Producibility, robustness and fluctuations. *J. Phys. Conf. Ser.* **2010**, *233*, 012019.
13. Schrijver, A. *Theory of Linear and Integer Programming*; Wiley: New York, NY, USA, 1986.
14. Orth, J.; Thiele, I.; Palsson, B.O. What is flux balance analysis? *Nat. Biotechnol.* **2010**, *28*, 245–248.
15. Segrè, D.; Vitkup, D.; Church, G. Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 15112–15117.

16. Jankowski, M.; Henry, C.; Broadbelt, L.; Hatzimanikatis, V. Group contribution method for thermodynamic analysis of complex metabolic networks. *Biophys. J.* **2008**, *95*, 1487–1499.

17. Fleming, R.; Thiele, I.; Nasheuer, H. Quantitative assignment of reaction directionality in constraint-based models of metabolism: Application to *Escherichia coli*. *Biophys. Chem.* **2009**, *145*, 47–56.

18. Kummel, A.; Panke, S.; Heinemann, M. Systematic assignment of thermodynamic constraints in metabolic network models. *BMC Bioinforma.* **2006**, 7, 512:1–512:12.

19. Alberty, R.A. *Thermodynamics of Biochemical Reactions*; Wiley: Hoboken, NJ, USA, 2003.

20. Schellenberger, J.; Lewis, N.; Palsson, B.O. Elimination of thermodynamically infeasible loops in steady-state metabolic models. *Biophys. J.* **2011**, *100*, 544–553.

21. Henry, C.; Broadbelt, L.; Hatzimanikatis, V. Thermodynamics-based metabolic flux analysis. *Biophys. J.* **2007**, *92*, 1792–1805.

22. Müller, A.; Brockmayr, A. Fast thermodynamically constrained flux variability analysis. *Bioinformatics* **2013**, *29*, 903–909.

23. Beard, D.A.; Qian, H. *Chemical Biophysics*; Cambridge University Press: Cambridge, NY, USA, 2008.

24. De Martino, D.; Figliuzzi, M.; de Martino, A.; Marinari, E. A scalable algorithm to explore the gibbs energy landscape of genome-scale metabolic networks. *PLoS Comp. Biol.* **2012**, *8*, e1002562.

25. De Martino, D. Thermodynamics of biochemical networks and duality theorems. *Phys. Rev. E* **2013**, *87*, 053108.

26. Johnson, D.B. Finding all the elemtary circuits of a directed graph. *SIAM J. Comput.* **1975**, *4*, 77–84.

27. Wright, J.; Wagner, A. Exhaustive identification of steady state cycles in large stoichiometric networks. *BMC Syst. Biol.* **2008**, 2, 61:1–61:9.

28. Wiback, S.; Famili, I.; Greenberg, H.; Palsson, B.O. Monte Carlo sampling can be used to determine the size and shape of the steady-state flux space. *J. Theor. Biol.* **2004**, *228*, 437–447.

29. Price, N.; Schellenberger, J.; Palsson, B.O. Uniform sampling of steady-state flux spaces: Means to design experiments and to interpret enzymopathies. *Biophys. J.* **2004**, *87*, 2172–2186.

30. Mezard, M.; Montanari, A. *Information, Physics, and Computation*; Oxford University Press: Oxford, NY, USA, 2009.

31. Feist, A.; Henry, C.; Reed, J.; Krummenacker, M.; Joyce, A.; Karp, P.; Broadbelt, L.; Hatzimanikatis, V.; Palsson, B.O. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.* **2007**, 3, 121:1–121:18.

32. Thiele, I.; Swainston, N.; Fleming, R.M.T.; Hoppe, A.; Sahoo, S.; Aurich, M.K.; Haraldsdottir, H.; Mo, M.L.; Rolfsson, O.; Stobbe, M.D.; *et al.* A community-driven global reconstruction of human metabolism. *Nat. Biotechnol.* **2013**, *31*, 419–425.

33. Schellenberger, J.; Que, R.; Fleming, R.M.T.; Thiele, I.; Orth, J.D.; Feist, A.M.; Zielinski, D.C.; Bordbar, A.; Lewis, N.E.; Rahmanian, S.; *et al.* Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox v2.0. *Nat. Protoc.* **2011**, *6*, 1290–1307.

34. Shlomi, T.; Cabili, M.N.; Herrgård, M.J.; Palsson, B.O.; Ruppin, E. Network-based prediction of human tissue-specific metabolism. *Nat. Biotechnol.* **2008**, *26*, 1003–1010.

35. Krauth, W.; Mezard, M. Learning algorithms with optimal stability in neural networks. *J. Phys. A* **1987**, *20*, L745–L752.

36. Binder, K.; Heermann, D.W. *Monte Carlo Simulation in Statistical Physics*; Springer: Heidelberg, Germany, 2002.

37. De Martino, A.; de Martino, D.; Mulet, R.; Uguzzoni, G. Reaction networks as systems for resource allocation: A variational principle for their non-equilibrium steady states. *PLoS One* **2012**, *7*, e39849.

38. Schilling, C.H.; Letscher, D.; Palsson, B.O. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J. Theor. Biol.* **2000**, *203*, 229–248.

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