OXFORD

Systems biology

Accelerating flux balance calculations in genome-scale metabolic models by localizing the application of loopless constraints

Siu H. J. Chan, Lin Wang, Satyakam Dash and Costas D. Maranas*

Department of Chemical Engineering, The Pennsylvania State University, University Park, PA 16802, USA

*To whom correspondence should be addressed.
Associate Editor: Jonathan Wren

Abstract

Background: Genome-scale metabolic network models and constraint-based modeling techniques have become important tools for analyzing cellular metabolism. Thermodynamically infeasible cycles (TICs) causing unbounded metabolic flux ranges are often encountered. TICs satisfy the mass balance and directionality constraints but violate the second law of thermodynamics. Current practices involve implementing additional constraints to ensure not only optimal but also loopless flux distributions. However, the mixed integer linear programming problems required to solve become computationally intractable for genome-scale metabolic models.

Results: We aimed to identify the fewest needed constraints sufficient for optimality under the loopless requirement. We found that loopless constraints are required only for the reactions that share elementary flux modes representing TICs with reactions that are part of the objective function. We put forth the concept of localized loopless constraints (LLCs) to enforce this minimal required set of loopless constraints. By combining with a novel procedure for minimal null-space calculation, the computational time for loopless flux variability analysis (ll-FVA) is reduced by a factor of 10–150 compared to the original loopless constraints and by 4–20 times compared to the current fastest method Fast-SNP with the percent improvement increasing with model size. Importantly, LLCs offer a scalable strategy for loopless flux calculations for multi-compartment/multi-organism models of large sizes, for example, shortening the CPU time for ll-FVA from 35 h to less than 2 h for a model with more than 10^4 reactions.

Availability and implementation: Matlab functions are available in the Supplementary Material or at https://github.com/maranasgroup/lll-FVA

Contact: costas@psu.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

A genome-scale model (GSM) provides an inventory of reactions for a given organism that allows for the analysis of cellular metabolism and the design of gene modulation strategies for bioproduction. Despite extensive manual curations, thermodynamically infeasible cycles (TIC) often exist in GSMs because of overly permissive reaction inclusion or directionalities that can affect flux range calculations using flux balance analysis (FBA, Orth *et al.*, 2010) and/or flux

variability analysis (FVA, Gudmundsson and Thiele, 2010; Mahadevan and Schilling, 2003). A TIC is an internal cycle in the metabolic network satisfying mass balances and directionality constraints without involving any exchange reactions, i.e. no system input/output is required. For example, a TIC is formed by the following three reactions:

i. H_2O + Glutamate + NAD^+ ↔ H^+ + NH_4^+ + NADH + 2-Oxoglutarate
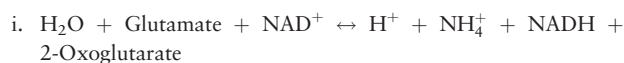
ii.   Alanine + 2-Oxoglutarate ↔ Pyruvate + Glutamate
iii. H_2O + Alanine + NAD^+ ↔ H^+ + Pyruvate + NH_4^+ + NADH

The cycle can carry an arbitrarily large flux when performing FBA despite the fact that any turn around the cycle does not produce or consume any metabolites. Thus, it violates the second law of thermodynamics as the net change of Gibbs free energy is zero. FVA of the TIC participating reactions always predicts unbounded flux ranges. The unbounded fluxes may affect regulatory constraints by making the regulated fluxes essentially unresponsive to any imposed regulation (Dash *et al.*, 2014). In addition, strain design algorithms such as OptForce (Chowdhury *et al.*, 2014; Ranganathan *et al.*, 2010) which rely on precise flux range calculations can also be adversely affected by the presence of TICs.

While TICs can be eliminated by simply *a priori* restricting the directionality of some reactions participating in TICs, this may also rule out biologically realistic phenotypes. In the aforementioned example, thermodynamics dictates that all three reactions are reversible under standard cellular concentrations. Several methods have been proposed to identify and eliminate TICs without overrestricting directionalities. The min-sum flux procedure first identifies the maximum biomass yield and then imposes this as a requirement while minimizing the sum of fluxes in the network (Holzhütter, 2004). A variation of this principle is adopted in parsimonious FBA (Lewis *et al.*, 2010). The min-sum flux criterion, however, can become too restrictive ruling out other possibly physiologically meaningful flux distributions and does not necessarily eliminate all TICs. Thermodynamic metabolic flux analysis is a method which takes metabolite concentrations into account to determine directionality, but relies on *a prior* knowledge of the standard Gibbs free energy and physiologically relevant ranges for metabolite concentrations (Henry *et al.*, 2007). Post-processing TIC removal has also been proposed in the recent CycleFreeFlux framework (Desouki *et al.*, 2015). CycleFreeFlux detects and removes TICs in a given flux distribution by solving an additional linear programming (LP) problem and uses an iterative algorithm to calculate FVA ranges for reactions participating in TICs. However, the optimality of the flux distribution with respect to the objective function after TIC removal cannot be guaranteed. Alternatively, loopless FVA (ll-FVA) directly computes the flux range for a reaction in the absence of any TICs (Maranas and Zomorrodi, 2016; Schellenberger *et al.*, 2011):

$$\min/\max v_j$$

$$\text{s.t.} \sum_{j\in J} S_{ij}v_j = 0, \qquad \forall i \in \mathbf{I} \qquad (1)$$

$$LB_j \le v_j \le UB_j, \qquad \forall j \in \mathbf{J} \qquad (2)$$

$$-M(1-y_j) \le v_j \le My_j, \qquad \forall j \in \mathbf{J}^{\text{int}} \qquad (3)$$

$$-My_j + \varepsilon(1-y_j) \le G_j \le -\varepsilon y_j + M(1-y_j), \quad \forall j \in \mathbf{J}^{\text{int}} \qquad (4)$$

$$\sum_{j\in J^{\text{int}}} N_{jr}^{\text{int}} G_j = 0, \qquad \forall r \in \{1,\dots,R\} \qquad (5)$$

$$y_j \in \{0,1\}, G_j \in \mathbb{R}, \qquad \forall j \in \mathbf{J}^{\text{int}}$$

$$v_j \in \mathbb{R}, \qquad \forall j \in \mathbf{J}$$

where **I** is the set of metabolites, **J** is the set of reactions, $\mathbf{J}^{\text{int}}$ is the set of internal reactions (non-exchange reactions), $\mathbf{S} = [S_{ij}]_{i\in I, j\in J}$ is the stoichiometric matrix (S-matrix) of the network, v_j is the flux of

reaction j, LB_j and UB_j are the lower and upper bounds respectively, $\mathbf{N}^{\text{int}} = [N_{jr}^{\text{int}}]_{j\in J, r\in\{1,\dots R\}}$ is a null-space matrix of rank R of the S-matrix for internal reactions $\mathbf{S}^{\text{int}} = [S_{ij}]_{i\in I, j\in J^{\text{int}}}$, G_j is a continuous variable associated with internal reaction j analogous to the change in Gibbs free energy of the reaction, M is a large positive constant, ε is a small positive constant, y_j is a binary variable. Figure 1A shows a toy network and the associated S-matrix. There are three TICs: R1 + R2, R5 + R6 and R5 + R7, which are identified in the null-space matrix (Fig. 1B). The null-space matrix also contains an infeasible loop R3 + R4 (2nd column) because R3 and R4 form a cycle if they are reversible reactions. ll-FVA formulates a mixed integer linear programming (MILP) problem with the number of free binary variables equal to the number of non-zero rows in the null-space matrix (7 in the toy network, see Fig. 1B). It becomes more time consuming for larger models, e.g. 10 h for the *E.coli* iJO1366 model (Orth *et al.*, 2011) as reported in Saa and Nielsen (2016), which presented a novel approach to largely reduce the time for ll-FVA by invoking a Fast-Sparse Null-space Pursuit (Fast-SNP) algorithm. Fast-SNP selects a minimal sparse null-space basis $\mathbf{N}^{\text{int}}$ that spans a subspace containing all possible internal loops and excludes infeasible loops when the reaction directionality is specified (e.g. R3 + R4 in the toy network; see Fig. 1B–C). In this way, variables G_j and y_j become uncoupled from the rest of the formulation and can be pre-calculated for all reactions j that do not participate in any TIC, for example, G_{R3} and G_{R4}, as the coefficients for rows R3, R4 are zero in the null-space matrix. This reduces the number of binary variables from seven to five. In general, the number of binary variables being active in ll-FVA (the key determinant of the MILP complexity) is reduced to the number of reactions participating in any TICs using Fast-SNP.

In addition to null-space reduction, it is intuitively true that constraints for loopless solutions are redundant while performing FVA analysis for reactions not present in TICs (e.g. R3 in Fig. 1A). Moreover, when finding the flux range for a reaction in a TIC (e.g. R7 of TIC R5 + R7 in Fig. 1A), the constraints for restricting other independent TICs (e.g. R1+R2 in Fig. 1A) can also be removed to simplify the model. As a result, a complete set of loopless constraints defined in ll-FVA is not necessary for analyzing every reaction. Herein, we generalize these observations by introducing the concept of localized loopless constraints (LLCs) along with a novel algorithm to compute a minimal null-space basis. LLCs are constraints that are only invoked for reactions present in the objective function. The MILP problem has a reduced number of binary variables with LLCs imposed during FVA compared to ll-FVA and finds an optimal flux distribution that has no TIC involving the reactions present in the objective function, thus attaining the same flux range as the original ll-FVA problem. By solving an additional LP problem to remove other independent TICs, an optimal loopless flux distribution can be computed. We prove that the minimum number of the required binary variables to enforce a loopless requirement when maximizing or minimizing reaction j is equal to the number of reactions that share an elementary flux mode of a TIC with reaction j. We prove that as long as the reactions in the objective function and ATP maintenance (ATPM, usually the only reaction with an active lower bound) are not in any TICs, the optimum solution value is unaffected without the constraints for loopless solutions. By using LLCs and a novel null-space algorithm, we are able to further accelerate loopless flux calculations significantly. The null-space calculation time is reduced by 4–1000 times compared to the current available fastest procedure Fast-SNP (Saa and Nielsen, 2016). For the models previously tested using Fast-SNP, LLCs exhibits an improvement of 4 to 10 times in the overall computational time compared to Fast-SNP and 10–150 times compared to the use of the
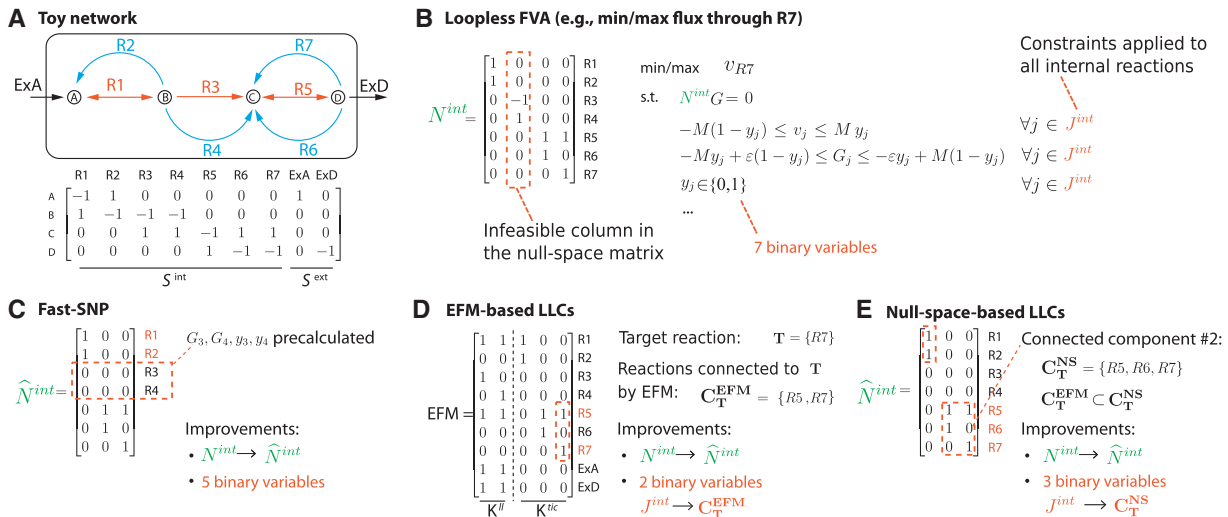
Fig. 1. Toy network for illustrating the idea of localized loopless constraints. (**A**) A toy network with TICs and the associated stoichiometric matrix. (**B**) The original constraints for loopless flux calculations imposed on all internal reactions. (**C**) The previously proposed Fast-SNP (Saa and Nielsen, 2016) to find a minimal null-space. G_{R3}, G_{R4}, y_{R3}, y_{R4} become uncoupled from the rest of the formulation and can be pre-calculated. (**D**) The proposed LLCs using elementary flux modes that represent TICs. Constraints are imposed only on reactions that are connected to the target reaction R7 by any EFM. The number of binary variables is reduced to two. (**E**) The proposed LLCs based on the connected components of the null-space. The number of binary variables is three, equal to the size of the connected component containing the target reaction R7. EFM calculations are not required

original constraints (3)–(4) for loopless solutions. We tested community models consisting of multiple *E.coli* and observed 8 to 20-fold improvement by using LLCs compared to Fast-SNP. This implies that LLCs is a tractable and scalable strategy for loopless flux calculations in multi-compartment/multi-organism models. The Matlab functions for the COBRA Toolbox (Heirendt *et al.*, 2017) are available in https://github.com/maranasgroup/lll-FVA.

2 Materials and methods

In this section, the concept and applications of LLCs for finding loopless flux distributions given a minimal null-space matrix characterizing all TICs are developed in Sections 2.1–2. 4. A novel algorithm for computing a minimal null-space follows in Section 2.5. The overall procedure is summarized in Section 2.6. Throughout this work, a flux distribution $\mathbf{v} = [v_j]_{j \in \mathbf{J}}$ is defined as a vector satisfying the steady-state condition Eq. (1) and the bound constraint Eq. (2). $\mathbf{v}^{int} = [v_j]_{j \in \mathbf{J}^{int}}$ denotes the vector containing the fluxes of all internal reactions in $\mathbf{J}^{int}$. The proofs for the propositions presented are given in Supplementary Methods. The network in Figure 1 is used as a running example to explain the concepts and definitions.

2.1 Thermodynamically infeasible cycles

Thermodynamically infeasible cycles (TICs) are defined as follows:

Definition 1
A TIC is a nonzero flux distribution $\mathbf{v}$ such that $\mathbf{S}^{int}\mathbf{v}^{int} = 0$.

In the toy network, for example, $v_{R1} = v_{R2} = 1$ is a TIC and $v_{ExA} = v_{R1} = v_{R3} = v_{R5} = v_{ExD} = 1$ is not. We now define the relation of a flux distribution containing a sub flux distribution.

Definition 2
A flux distribution $\mathbf{v}$ is said to *contain* another flux distribution $\mathbf{v}'$, denoted by $\mathbf{v}' \preccurlyeq \mathbf{v}$, if each flux v'_j in $\mathbf{v}'$ is either zero or has the same sign as v_j with a smaller or equal magnitude, i.e.

$$0 \leq \text{sgn}(v_j)v'_j \leq \text{sgn}(v_j)v_j, \ \forall j \in \mathbf{J}$$

where $\text{sgn}(v_j) = 1$ if $v_j \geq 0$ and $= -1$ if $v_j < 0$. Furthermore, $\mathbf{v}$ is said to *properly contain* $\mathbf{v}'$, denoted by $\mathbf{v}' \prec \mathbf{v}$, if in addition to $\mathbf{v}' \preccurlyeq \mathbf{v}$, there exists at least one reaction having nonzero flux in $\mathbf{v}$ with zero flux in $\mathbf{v}'$, i.e.

$$\exists j \in \mathbf{J} \ \ \text{s.t.} \ \ v'_j = 0 \text{ and } v_j \neq 0$$

The relation $\mathbf{v}' \preccurlyeq \mathbf{v}$ used in this article has a meaning different from the common usage of $\mathbf{v}'$ being component-wise less than $\mathbf{v}$. Instead, in the context of this work, it implies that $\mathbf{v}'$ and $\mathbf{v}$ have the same dimensions and the magnitude of all fluxes in $\mathbf{v}'$ are between zero and the corresponding value in $\mathbf{v}$. The relation $\mathbf{v}' \prec \mathbf{v}$ implies that in addition to $\mathbf{v}' \preccurlyeq \mathbf{v}$ as defined, $\mathbf{v}'$ and $\mathbf{v}$ also satisfy that the set of reactions with zero fluxes for $\mathbf{v}$ is a proper subset of the set of reactions with zero fluxes for $\mathbf{v}'$. We also invoke the concept of elementary flux modes (EFMs), which are flux distributions not properly containing any nonzero flux distribution (Schuster and Hilgetag, 1994).

Definition 3
A flux distribution $\mathbf{e}$ is an EFM if $\mathbf{v} \prec \mathbf{e} \ \Rightarrow \ \mathbf{v} = 0$.

In other words, if a flux distribution $\mathbf{e}$ is an EFM then the only flux distribution it properly contains is $\mathbf{v} = 0$.

2.2 Localized loopless constraints

In this section, we present the localized loopless constraints (LLCs) for inactivating TICs that involve a specific subset of reactions. Let $\mathbf{J}^{tic} = \{j | \exists \text{ TIC } \mathbf{v} \ \text{ such that } \ v_j \neq 0\}$ be the set of reactions participating in any TICs (e.g. $\mathbf{J}^{tic} = \{R1, R2, R5, R6, R7\}$ in the toy network of Fig. 1A). Let $\mathbf{K} = \{1, \dots, K\}$ be the index set for all EFMs $\mathbf{e}_1, \dots, \mathbf{e}_K$, $\mathbf{K}^{ll} = \{k \in \mathbf{K} | \exists j \in \mathbf{J} \backslash \mathbf{J}^{int} \ \text{ such that } \ e_{jk} \neq 0\}$ be the set of loopless EFMs. $\mathbf{K}^{tic} = \mathbf{K} \backslash \mathbf{K}^{ll}$ is then the set of all TIC EFMs (see Fig. 1D). Assume that only specific TICs involving a set of target reactions in $\mathbf{T} \subset \mathbf{J}^{tic}$ are required to be inactivated in the flux distribution, i.e. TICs not containing any reactions in $\mathbf{T}$ can still be part

of the flux distribution. Denote the set of TIC EFMs involving any reactions in $\mathbf{T}$ by $\mathbf{K}_{\mathbf{T}}^{\text{tic}} = \{k \in \mathbf{K}^{\text{tic}} | \exists \, j \in \mathbf{T} \text{ such that } e_{jk} \neq 0\}$. By the 'no-cancellation' rule (Schuster *et al.*, 2002) used extensively before (Carlson, 2009; Chan and Ji, 2011; Chan *et al.*, 2014; Ip *et al.*, 2011; Schwartz and Kanehisa, 2005, 2006; Zhao and Kurata, 2009), any flux distribution $\mathbf{v}$ can be decomposed into EFMs as follows:

$$
\begin{aligned}
\mathbf{v} &= \mathbf{v}^{\text{ll}} + \mathbf{v}^{\text{tic}} \\
&= \mathbf{v}^{\text{ll}} + \mathbf{v}^{\text{tic, nontarget}} + \mathbf{v}^{\text{tic, target}} \\
&= \sum_{k \in \mathbf{K}^{\text{ll}}} \alpha_k \mathbf{e}_k + \sum_{k \in \mathbf{K}^{\text{tic}} \setminus \mathbf{K}_{\mathbf{T}}^{\text{tic}}} \alpha_k \mathbf{e}_k + \sum_{k \in \mathbf{K}_{\mathbf{T}}^{\text{tic}}} \alpha_k \mathbf{e}_k, \quad \alpha_k \geq 0, \ \forall k \in \mathbf{K}
\end{aligned}
\tag{6}
$$

where $\mathbf{v}^{\text{ll}}$ is the loopless part of the flux distribution $\mathbf{v}$ and $\mathbf{v}^{\text{tic}}$ consists of TICs only. $\mathbf{v}^{\text{tic}}$ can be decomposed into TICs not involving any target reactions in $\mathbf{T}$ ($\mathbf{v}^{\text{tic, nontarget}}$) and TICs through reactions in $\mathbf{T}$ ($\mathbf{v}^{\text{tic, target}}$). From Eq. (6), blocking the target set of EFMs $\mathbf{K}_{\mathbf{T}}^{\text{tic}}$ is sufficient to ensure $\mathbf{v}^{\text{tic, target}} = 0$ and thus eliminate TICs through reactions in $\mathbf{T}$ from $\mathbf{v}$. Let $\mathbf{C}_{\mathbf{T}}^{\text{EFM}} = \{j \in \mathbf{J}^{\text{tic}} | \exists \, k \in \mathbf{K}_{\mathbf{T}}^{\text{tic}}$ such that $e_{jk} \neq 0\}$ be the set of reactions connected to $\mathbf{T}$ by any target TIC EFMs. It contains all reactions that can have nonzero fluxes in $\mathbf{v}^{\text{tic, target}}$. In the toy network, R7 is used as an example target reaction, i.e. $\mathbf{T} = \{\text{R7}\}$. From the EFM matrix, only the fifth EFM is a TIC EFM that involves R7, therefore, $\mathbf{K}_{\mathbf{T}}^{\text{tic}} = \{5\}$ and $\mathbf{C}_{\mathbf{T}}^{\text{EFM}} \{\text{R5, R7}\}$ (see Fig. 1E). Proposition 1 states that the LLCs imposed on $\mathbf{C}_{\mathbf{T}}^{\text{EFM}}$, which are Eqs. (7)–(8), eliminate TICs involving $\mathbf{T}$.

Proposition 1

A flux distribution $\mathbf{v}$ does not contain any TICs involving reactions in $\mathbf{T}$ if Eq. (3) and Eqs. (4)–(5) restricted to $j \in \mathbf{C}_{\mathbf{T}}^{\text{EFM}}$ are satisfied:

$$
-M(1 - y_j) \leq v_j \leq M y_j, \qquad \forall j \in \mathbf{C}_{\mathbf{T}}^{\text{EFM}} \tag{7}
$$

$$
-M y_j + \varepsilon(1 - y_j) \leq G_j \leq -\varepsilon y_j + M(1 - y_j), \quad \forall j \in \mathbf{C}_{\mathbf{T}}^{\text{EFM}} \tag{8}
$$

$$
\sum_{j \in \mathbf{J}^{\text{int}}} N_{jr}^{\text{int}} G_j = 0, \qquad \forall r \in \{1, \ldots, R\} \tag{9}
$$

$$
G_j \in \mathbb{R}, \qquad \forall j \in \mathbf{J}^{\text{int}}
$$

$$
y_j \in \{0, 1\}, \qquad \forall j \in \mathbf{C}_{\mathbf{T}}^{\text{EFM}}
$$

See Supplementary Methods for the proof. The idea is to derive the equations by constraining $\mathbf{v}^{\text{tic, target}}$ in Eq. (6) to zero. The LLCs and the proof in Supplementary Methods refine the original constraints and the previously presented proof Noor *et al.* (2012). In the toy network, by Proposition 1, Eqs. (7)–(8) constraining only {R5, R7} are sufficient to prevent any TIC involving R7. However, if $\mathbf{T} = \{\text{R5}\}$, since there is one TIC EFM connecting R5 and R6 and one connecting R5 and R7, $\mathbf{C}_{\mathbf{T}}^{\text{EFM}} = \{\text{R5, R6, R7}\}$ is required to be constrained in Eqs. (7)–(8) to prevent TICs involving R5.

2.3 Determination of the target reaction set for applying LLCs

Proposition 1 implies that the number of binary variables can be reduced if only the specific TICs involving the target reactions in $\mathbf{T}$ (e.g. R5 and R7 in the toy network) are required to be inactivated. Indeed in many applications, the optimal objective function value of a MILP with the original constraints for loopless solutions is equal

to that with suitably selected LLCs. The optimal loopless solution can also be derived from the corresponding partially loopless solution. Consider the following general LP problem for finding a flux distribution $\mathbf{v}$:

$$
\begin{aligned}
\min &\sum_{j \in \mathbf{J}} c_j v_j \\
\text{s.t.} &\sum_{j \in \mathbf{J}} S_{ij} v_j = 0, \qquad \forall i \in \mathbf{I} \\
&\sum_{j \in \mathbf{J}} a_{pj} v_j \leq b_p, \quad \forall p \in \{1, \ldots, P\} \\
&v_j \in \mathbb{R}, \qquad \forall j \in \mathbf{J}
\end{aligned}
\tag{10}
$$

where P is the number of additional constraints. From Eq. (6), by applying LLCs, we already have $\mathbf{v}^{\text{tic, target}} = \mathbf{0}$. Obviously, both $\mathbf{v}^{\text{ll}}$ and $\mathbf{v}^{\text{tic, nontarget}}$ satisfy the mass balance equation. Therefore, as long as $\mathbf{v}^{\text{tic, nontarget}}$ does not contribute to the optimal objective function value (i.e. $\sum_{j \in \mathbf{J}^{\text{tic}}} c_j v_j^{\text{tic, nontarget}} = 0$) and $\mathbf{v}^{\text{ll}}$ alone satisfies the additional P constraints (i.e. $\sum_{j \in \mathbf{J}} a_{pj} v_j \leq b_p$ for all p), then $\mathbf{v}^{\text{ll}}$ is an optimal feasible solution to problem (10). Proposition 2 establishes a useful sufficient condition for using LLCs based on this idea.

Proposition 2

Denote the sets of reactions whose forward and reverse directions participate in TICs respectively by $\mathbf{J}_{\text{fwd}}^{\text{tic}} = \{j \in \mathbf{J}^{\text{tic}} | \exists \, k \in \mathbf{K}^{\text{tic}}$ s.t $\cdot \ e_{jk} > 0\}$ and $\mathbf{J}_{\text{rev}}^{\text{tic}} = \{j \in \mathbf{J}^{\text{tic}} | \exists \, k \in \mathbf{K}^{\text{tic}}$ s.t. $e_{jk} < 0\}$. For the LP problem (10), assume that the target reaction set $\mathbf{T} \subset \mathbf{J}^{\text{tic}}$ contains all reactions in TICs satisfying one of the following three conditions:

i. $j \in \mathbf{J}_{\text{fwd}}^{\text{tic}}$ and $c_j < 0$
ii. $j \in \mathbf{J}_{\text{rev}}^{\text{tic}}$ and $c_j > 0$
iii. Conditions (A) and (B) are satisfied:

   (A) $(j \in \mathbf{J}_{\text{fwd}}^{\text{tic}}$ and $\exists \, p$ s.t. $a_{pj} < 0)$ or $(j \in \mathbf{J}_{\text{rev}}^{\text{tic}}$ and $\exists \, p$ s.t. $a_{pj} > 0)$
   (B) $\exists \, p$ s.t. $a_{pj} \neq 0$ and $(a_{pj'} \neq 0$ for some $j' \neq j$ or $b_p < 0)$

Then the optimal objective function value of the LP problem (10) constrained with the localized loopless constraints Eqs. (7)–(9) is equal to the optimal objective function value of the LP problem (10) constrained with the original loopless constraints Eqs. (3)–(5).

Conditions (I) and (II) state that a reaction in $\mathbf{J}_{\text{fwd}}^{\text{tic}}$ or in $\mathbf{J}_{\text{fwd}}^{\text{rev}}$ needs to be put in the target set $\mathbf{T}$ only when the reaction flux is being maximized or minimized, respectively. In the toy network, $\mathbf{J}_{\text{fwd}}^{\text{tic}} = \{\text{R1, R2, R5, R6, R7}\}$ and $\mathbf{J}_{\text{rev}}^{\text{tic}}$ is empty. Though R1 and R5 are reversible, the reverse direction does not participate in any TIC. Condition (III.A) excludes a reaction that is in TICs from $\mathbf{T}$ even when it is constrained in the problem as long as it participates in TICs only in the forward (or reverse) direction and meanwhile the corresponding constraint coefficients for that reaction are all non-negative (or non-positive). Condition (III.B) excludes most of the bound constraints on fluxes in Eq. (2) with $LB_j \leq 0$ and $UB_j \geq 0$. For example, $v_{\text{R2}} \leq 1$ and $v_{\text{R2}} \geq 0$ do not satisfy conditions (III.A) and (III.B) respectively. These constraints do not necessitate putting R2 in $\mathbf{T}$. For $v_{\text{R2}} \geq 1$ and $-v_{\text{R2}} + v_{\text{R4}} \leq 0$, both of them satisfy conditions (III.A) and (III.B). If any one of the two constraints is in the problem, R2 must be included in $\mathbf{T}$. The proof is provided in Supplementary Methods. An optimal loopless flux distribution $\mathbf{v}^{\text{ll}}$ can be obtained by removing all TICs not involving any target reactions $\mathbf{v}^{\text{tic, nontarget}}$ contained in the localized loopless flux distribution

$\mathbf{v}^*$ which is the solution of problem (10) with LLCs, i.e. Eqs. (7)–(9) imposed. $\mathbf{v}^{\text{tic, nontarget}}$ is computed by maximizing the internal fluxes contained in $\mathbf{v}^*$ (Desouki *et al.*, 2015):

$$\mathbf{v}^{\text{tic, nontarget}} = \underset{v_j, j \in \mathbf{J}}{\text{argmax}} \sum_{j \in \mathbf{J}^{\text{int}}} \text{sgn}(v_j^*) v_j$$

$$\text{s.t.} \quad \sum_{j \in \mathbf{J}} S_{ij} v_j = 0, \qquad\qquad \forall i \in \mathbf{I}$$

$$0 \le \text{sgn}(v_j^*) v_j \le \text{sgn}(v_j^*) v_j^*, \quad \forall j \in \mathbf{J}^{\text{int}}$$

$$v_j = 0 \qquad\qquad \forall j \in \mathbf{J} \backslash \mathbf{J}^{\text{int}}$$

$$\tag{11}$$

The loopless flux distribution is then calculated by $\mathbf{v}^{\text{ll}} = \mathbf{v}^* - \mathbf{v}^{\text{tic, nontarget}}$. In the toy network, Proposition 2 states that applying LLCs with $\mathbf{T} = \{R7\}$ is sufficient to find the maximum of v_{R7}. Solving with LLCs may result at the flux distribution:

R1	R2	R3	R4	R5	R6	R7	ExA	ExD
[1000	1000	0	0	1000	1000	0	0	0]^T

Despite the fact that it contains TICs, LLCs on R5, R7 prevent any TICs through R7. The completely loopless flux distribution $\mathbf{v} = \mathbf{0}$ can be obtained by solving problem (11). In this way, the number of binary variables in solving MILP is reduced from 7 using the original null-space (see Fig. 1B) to 5 using the Fast-SNP null-space (see Fig. 1C), and further down to 2 using LLCs (see Fig. 1D). We define the active fraction f as the fraction of active binary variables relative to the number of reactions in TICs in an MILP problem for loopless flux calculation, i.e.

$$f = |\mathbf{C}_{\mathbf{T}}^{\text{EFM}}| / |\mathbf{J}^{\text{tic}}|$$

For finding the range for R7, i.e. $\mathbf{T} = \{R7\}$, we have $f = 0.4$. Taking $\mathbf{T} = \{R5\}$ as another example, although R5 is in TICs and is reversible, only the forward direction of R5 is in TICs. When minimizing v_{R5} ($c_{R5} = 1$ and R5 $\notin \mathbf{J}_{\text{rev}}^{\text{tic}}$ do not satisfy any of the conditions in Proposition 2), solving the LP problem gives the same minimum for v_{R5} as solving with the original constraints for loopless solutions. In this case, $f = 0$. A common constraint in metabolic models that satisfies condition (III) of Proposition 3 is the positive lower bound for the ATP maintenance (ATPM) reaction. Nonetheless, ATPM appearing in a TIC implies that it is coupled to an energy generating cycle, which should be resolved by manual curation (Fritzemeier *et al.*, 2017). Therefore, Proposition 2 guarantees that in a well-curated model free of ATP-generating cycles, the standard FBA and FVA can be performed without any constraints for loopless solution when the objective function does not concern any reactions in TICs (i.e. $\mathbf{T}$ is empty). In other applications where dense objective functions are used, all loopless constraints may be required depending on the sign of the objective coefficients for reactions in TICs. In this case, the improvement brought by LLCs may not be significant. Fortunately, the min-sum-flux objective function, which is one of the most frequently used dense objective functions, does not require any LLCs because it is equivalent to splitting each reversible reaction into two irreversible reactions for the forward and reverse directions and minimizing the sum of fluxes, i.e. $\mathbf{J}_{\text{rev}}^{\text{tic}}$ is empty and $c_j = 1$ for all j. Therefore, none of the conditions in Proposition 2 is satisfied. Another common situation requiring LLCs is the presence of positive lower or negative upper bounds on internal fluxes derived from experimental data (e.g. V_{max}) for reactions participating in TICs, which satisfies condition (III) in Proposition 2.

2.4 Finding a superset for reactions connected to the target set

One challenge in implementing LLCs is to determine $\mathbf{C}_{\mathbf{T}}^{\text{EFM}}$, the set of reactions connected to $\mathbf{T}$ by any TIC EFMs. This entails the use of the entire set of EFMs which in some cases could be computationally intractable due to combinatorial explosion (Klamt and Stelling, 2002). However, the minimal null-space matrix offers a computationally efficient way to find a superset of $\mathbf{C}_{\mathbf{T}}^{\text{EFM}}$ that is smaller than $\mathbf{J}^{\text{tic}}$ when computing EFMs is not preferred. Using the minimal null-space matrix, we can define that two reactions are connected if (i) they both have nonzero values in a column in the null-space matrix or (ii) if there is a reaction connected to both of them (i.e. the connectivity is transitive). Figure 1E visualizes the relation in the toy network. Pairs R5 + R6 and R5 + R7 satisfy condition (i). Therefore, R5 is connected to both R6 and R7. By condition (ii), R6 and R7 are also connected via R5. Under this relation of connection, the reactions in TICs $\mathbf{J}^{\text{tic}}$ can be partitioned into L connected components $\mathbf{J}_l^{\text{tic}} \subset \mathbf{J}^{\text{tic}}$ for $l = 1, \ldots, L$ (e.g. $\mathbf{J}_1^{\text{tic}} = \{R1, R2\}$ and $\mathbf{J}_2^{\text{tic}} = \{R5, R6, R7\}$ in Fig. 1E). For any union of connected components $\mathbf{C}_{\mathbf{T}}^{\text{NS}} = \cup_{q=1}^{Q} \mathbf{J}_{l_q}^{\text{tic}}$ where $1 \le l_1, \ldots, l_Q \le L$, we prove that if $\mathbf{C}_{\mathbf{T}}^{\text{NS}}$ contains $\mathbf{T}$, then it also contains $\mathbf{C}_{\mathbf{T}}^{\text{EFM}}$ (see Supplementary Methods for proof), i.e.

$$\mathbf{T} \subset \mathbf{C}_{\mathbf{T}}^{\text{NS}} \quad \Rightarrow \quad \mathbf{C}_{\mathbf{T}}^{\text{EFM}} \subset \mathbf{C}_{\mathbf{T}}^{\text{NS}}$$

Imposing LLCs on $\mathbf{C}_{\mathbf{T}}^{\text{NS}}$ is thus sufficient to ensure the absence of TICs through reactions in $\mathbf{T}$. In the toy network, we can apply LLCs on the connected component $\mathbf{C}_{\mathbf{T}}^{\text{NS}} = \mathbf{J}_2^{\text{tic}} = \{R5, R6, R7\}$ to ensure no TICs through R7. In this case, we still have an active fraction $f = 0.6$.

2.5 Algorithm for minimal null-space

In addition to introducing LLCs, we propose here a novel algorithm for computing a minimal null-space. The current best method Fast-SNP constructs a minimal null-space by iteratively solving LP problems to find a new feasible basis vector not lying in the null-space under construction until no new basis vector is found. Instead, we propose to solve a single MILP problem that find a maximal TIC such that each reaction in $\mathbf{J}^{\text{tic}}$ carries nonzero flux. A minimal null-space basis is then calculated from the submatrix $\mathbf{S}^{\text{tic}} = [S_{ij}]_{i \in \mathbf{I}, j \in \mathbf{J}^{\text{tic}}}$ using a sparse LU factorization algorithm termed LUSOL (Gill *et al.*, 1987) implemented in the COBRA toolbox. The procedure recovers the null-space in a significantly shorter time for large models compared to Fast-SNP. By introducing constraints similar to Eqs. (3)–(4) to model the flux direction, the MILP problem is formulated as follows:

$$\min \sum_{j \in \mathbf{J}^{\text{int}}} (z_j^+ + z_j^-)$$

$$\text{s.t.} \sum_{j \in \mathbf{J}^{\text{int}}} S_{ij} v_j = 0, \qquad\qquad \forall i \in \mathbf{I}$$

$$-M \delta_j^L \le v_j \le M \delta_j^U, \qquad\qquad \forall j \in \mathbf{J}^{\text{int}}$$

$$\varepsilon - M(z_j^+ + 1 - \delta_j^U) \le v_j \le -\varepsilon + M(z_j^- + 1 - \delta_j^L), \ \forall j \in \mathbf{J}^{\text{int}}$$

$$z_j^+, \ z_j^- \ge 0 \qquad\qquad \forall j \in \mathbf{J}^{\text{int}}$$

$$v_j \in \mathbb{R} \qquad\qquad \forall j \in \mathbf{J}^{\text{int}}$$

$$z_j^+, \ z_j^- \in \{0, 1\} \qquad\qquad \forall j \in \mathbf{J}^{\text{int}} \ if \ \delta_j^L = \delta_j^U = 1$$

$$\tag{12}$$

where $\delta_j^L = 1$ if $LB_j < 0$ and zero otherwise whereas $\delta_j^U = 1$ if $UB_j > 0$ and zero otherwise. Thus, for a reversible reaction j, $\delta_j^L = \delta_j^U = 1$. The proof that all reactions in TICs have nonzero fluxes in the

solution of problem (12) is provided in Supplementary Methods. Problem (12) forces each reaction in TICs to have a nonzero flux by minimizing $z_j^+ + z_j^-$. $z_j^+ = 0$ implies that reaction j can have positive flux in TICs and $z_j^- = 0$ implies that reaction j can have negative flux. For each irreversible reaction j, one of the z_j^+, z_j^- can be predetermined ($z_j^+ = 0$ if $\delta_j^U = 0$ and $z_j^- = 0$ if $\delta_j^L = 0$) and only the other needs to be determined. Modeling z_j^+, z_j^- as continuous variables for an irreversible reaction j suffices to force $\nu_j \neq 0$ if it is feasible. For a reversible reaction j, z_j^+, z_j^- are required to be binary. A special property of problem (12) is that during branch and bound for solving MILP, branching down (take z_j^+ or $z_j^- = 0$) whenever possible until integer feasibility can always lead to an optimal solution (see Supplementary Methods for more detailed analysis) implying that at most $2n_{rev}$ relaxed LPs need to be solved for n_{rev} reversible reactions. Solving problem (12) is similar to an iterative LP procedure but it takes advantage of the built-in structure of any modern MILP solver. In practice, problem (12) is pre-solved as LPs to determine $j \in J_{\text{fwd}}^{\text{tic}}$ and $j \in J_{\text{fwd}}^{\text{rev}}$ simultaneously (see Supplementary Methods).

2.6 Overall procedure

Combining all the methods presented, we propose the following procedure for performing loopless flux balance calculations:

1. Solve problem (12) to identify $\mathbf{J}^{\text{tic}}$ and compute a minimal null-space matrix $\mathbf{N}^{\text{int}}$ using the submatrix $\mathbf{S}^{\text{tic}} = [S_{ij}]_{i \in I, j \in J^{\text{tic}}}$.
2. Identify the set of reactions whose forward directions are in TICs $j \in J_{\text{fwd}}^{\text{tic}}$ and whose reverse directions are in $j \in J_{\text{fwd}}^{\text{rev}}$ by FVA on reactions in $\mathbf{J}^{\text{tic}}$ with all exchange reactions shut down.
3. Find all connected components $\mathbf{J}_l^{\text{tic}} \subset \mathbf{J}^{\text{tic}}$ from the null-space.
4. For each loopless flux calculation, determine the target set $\mathbf{T}$ for applying LLCs using the conditions in Proposition 2.
5. Find the minimum $\mathbf{C}_{\mathbf{T}}^{\text{NS}} = \cup_{q=1}^{Q} \mathbf{J}_{l_q}^{\text{tic}}$ such that $\mathbf{T} \subset \mathbf{C}_{\mathbf{T}}^{\text{NS}}$. For each $q = 1, \ldots, Q$, calculate the complete set of EFMs for connected component $\mathbf{J}_{l_q}^{\text{tic}}$ using the corresponding columns from the S-matrix $\mathbf{S}$. Determine $\mathbf{C}_{\mathbf{T}}^{\text{EFM}}$ from the EFMs.
6. Solve problem (10) with constraints (7)–(9) on $\mathbf{C}_{\mathbf{T}}^{\text{EFM}}$ to obtain the localized loopless flux distribution $\mathbf{v}^*$.
7. (optional) Solve problem (11) to obtain $\mathbf{v}^{\text{tic, nontarget}}$. The completely loopless optimal flux distribution is given by $\mathbf{v}^{\text{ll}} = \mathbf{v}^* - \mathbf{v}^{\text{tic, nontarget}}$.

See Supplementary Methods for more details on Step 1–3. If the calculation of EFMs in Step 5 is not preferable, one can skip the EFM calculation and solve the MILP problem with LLCs imposed on $\mathbf{C}_{\mathbf{T}}^{\text{NS}}$ instead of $\mathbf{C}_{\mathbf{T}}^{\text{EFM}}$. We have implemented the procedure in MATLAB using the COBRA Toolbox (Heirendt *et al.*, 2017) and the optimization solver Gurobi (http://www.gurobi.com). *EFMtool* was used for EFM computations (Terzer and Stelling, 2008).

3 Results

3.1 Single-organism models

Using the seven models tested in Saa and Nielsen (2016) excluding the toy model therein (see Table 1), we compared the performance between (i) ll-FVA using the original loopless constraint (referred to as ll-FVA), (ii) ll-FVA with Fast-SNP preprocessing (referred to as Fast-SNP), (iii) FVA with null-space-based LLCs, i.e. imposing LLCs on $\mathbf{C}_{\mathbf{T}}^{\text{NS}}$ in Step 6 in Section 2.6 without EFM calculations (referred to as NS-LLC) and (iv) FVA with EFM-based LLCs (referred to as EFM-LLC). In the models tested, NS-LLC and EFM-LLC show 10–150× reduction in computational needs (in CPU time) compared to ll-FVA and 4–10× reduction compared to Fast-SNP (Fig. 2A; see

Supplementary Table S1 for the numerical values). The time for NS-LLC preprocessing is in general small and accounts for only <0.5% of the total CPU time (Fig. 2B). In contrast, the time for EFM-LLC preprocessing which includes EFM calculation is more model-dependent and accounts for 12–64% of the total run time respectively, for the *E.coli* core model and yeast6 models. In return, EFM-LLC required an MILP solution time 2.5 times faster than that of NS-LLC for the yeast6 model due to the overall smaller active fraction f (fraction of active 0-1 variables relative to the number of reactions in TICs after LLC preprocessing) using the information from EFMs (last two columns in Fig. 2B). For all models except yeast6, a significant amount of the solution time was spent on solving problems without any integer variables (active fraction $f = 0$), i.e. LP problems only. We also tested the performance of the proposed approach to obtain completely loopless flux distributions (Step 7 in Section 2.6) when maximizing the biomass production against ll-FBA with/without Fast-SNP preprocessing. Except for the *E.coli* core model for which all methods performed similarly, 3–500-fold improvements were observed for other models (Supplementary Table S2). Since all biomass reactions are not part of any TICs, finding a loopless flux distribution required solving only two LP problems [the standard FBA without any constraints restricting TICs and problem (11) for removing independent TICs], in contrast to solving an MILP problem in other methods.

3.2 Multi-organism models

The computational performance of Fast-SNP, NS-LLC and EFM-LLC was further compared using models for microbial communities. The nine-species model previously used for modeling the gut microbiota (Chan *et al.*, 2017b) and community models consisting of multiple copies of the *E.coli* iJO1366 (Orth *et al.*, 2011) were tested (Table 2). In the community models, inter-organism TICs are eliminated without the need of over-restricting the directionality of intracellular reactions by imposing suitable penalties on transport reactions, e.g. consuming instead of gaining proton gradient by exporting certain metabolites (Chan *et al.*, 2017a). Otherwise, when the number of organisms increases, extremely large inter-organism TICs (involving hundreds to thousands of reactions) can appear and cause the MILP problems to become intractable. Restricting TICs to appear only within individual organisms and applying LLCs can become a scalable procedure for loopless flux calculations for multi-organism models because the number of binary variables in each MILP problem solved is independent of the number of organisms. Figure 3A shows the total CPU time of loopless FVA for all reactions in TICs. They are the reactions for which non-trivial MILP problems must be solved. Overall, NS-LLC and EFM-LLC show improvements in computation speed of 8–16 times and 8–19 times, respectively, compared to Fast-SNP. In particular for the eight-copy *E.coli* community model Ec(8), Fast-SNP took 35 h of CPU time while EFM-LLC took only 1.8 h. The reduction is caused by the larger extent of binary variable reduction (smaller active fraction f) when solving MILP problems as the model size grows (Fig. 3B), confirming the particular advantage of LLCs for multi-organism and multi-compartment models. Supplementary Figure S3 compares the solution time taken by NS-LLC and EFM-LLC compared to Fast-SNP for the same MILP problems solved, respectively (see SI). They clearly show that as the model size increases, the MILP complexity of Fast-SNP grows more rapidly than NS-LLC and EFM-LLC. The large standard deviations signify the fact that the actual reduction in computational time is very problem-specific.
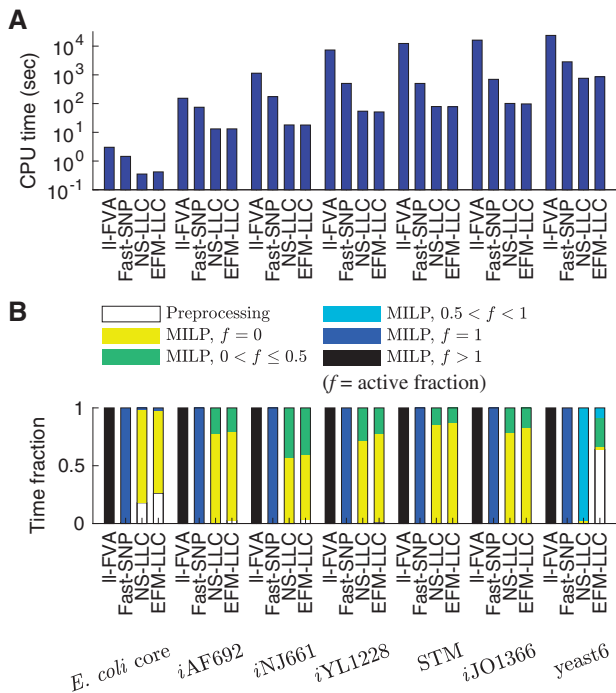
Fig. 2. Performance of the four methods for loopless FVA under comparison on single-organism models. (**A**) Total CPU time. (**B**) Fractional breakdown of the total CPU time into time for preprocessing and solving MILP problems with various degree of active binary variables in terms of the active fraction f

Fig. 3. Performance of Fast-SNP, NS-LLC and EFM-LLC for loopless FVA on community models. (**A**) Total CPU time. (**B**) Fractional breakdown of the total CPU time into time for preprocessing and solving MILP problems with various degree of active binary variables in terms of the active fraction f. Ec(n) is the community model of n copies of the *E.coli* iJO1366 model

Table 1. Single-organism models tested

Model	#metabolites[a]	#reactions[a]	$|\mathbf{J}^{tic}|$
E.coli core	68	87	2
*i*AF692	417	484	30
*i*NJ661	579	740	53
*i*YL1228	830	1223	59
STM	1086	1597	52
*i*JO1366	1136	1679	76
yeast6	756	1018	293

$|\mathbf{J}^{tic}|$ is the number of reactions in TICs.

[a]Numbers of metabolites and reactions after removal of blocked reactions.

3.3 Performance of the proposed null-space algorithm

Figure 4 shows the performance of the proposed algorithm for computing minimal null-space compared to Fast-SNP for all tested single-organism and microbial community models (see Supplementary Figs S1 and S2 for the actual time spent for null-space calculations and other steps). Except for the smallest *E.coli* core model, the proposed algorithm outperformed Fast-SNP and the advantage became more significant as the model size increases, ranging from 4- to >1000-fold reduction in computational time. This suggests that the proposed null-space algorithm scales better than Fast-SNP.

4 Conclusions

In this paper, we propose the use of LLCs to further reduce the computational cost for loopless flux calculations. Notably we proved that for many models simply computing the LP problem without loopless constraints is sufficient to guarantee optimality. An optimal and loopless flux distribution can be derived from the LP solution
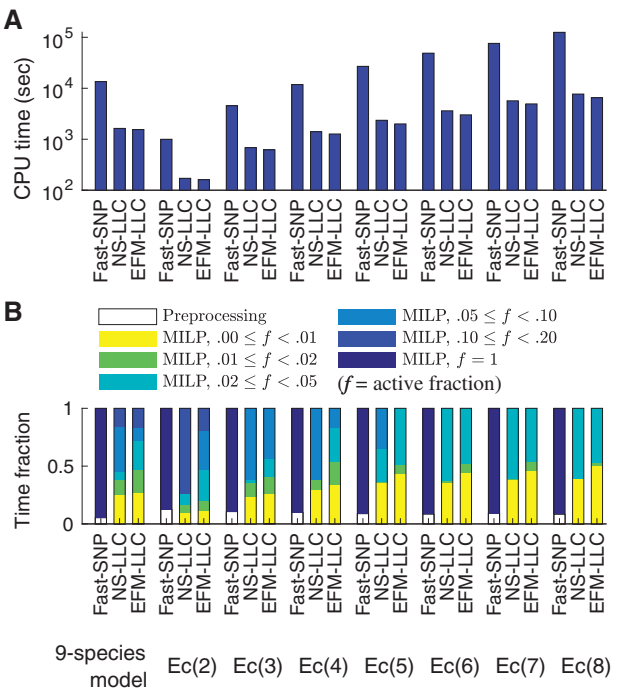
Table 2. Microbial community models tested

Model	#metabolites[a]	#reactions[a]	$|\mathbf{J}^{tic}|$
9-species model	5758	7621	253
Ec(2)	2460	3615	130
Ec(3)	3625	5351	195
Ec(4)	4790	7087	260
Ec(5)	5955	8823	325
Ec(6)	7120	10559	390
Ec(7)	8285	12295	455
Ec(8)	9450	14031	520

Ec(n) is the community model of n copies of the *E.coli* iJO1366 model.

$|\mathbf{J}^{tic}|$ is the number of reactions in TICs.

[a]Numbers of metabolites and reactions after removal of blocked reactions.

by a simple post-processing TIC removal. LLCs offer a scalable way for loopless flux calculations for large multi-organism/multi-compartment genome-scale metabolic models. Since in most tested models except the yeast6 model, EFM-LLC slightly outperformed NS-LLC, we recommend using EFM-LLC in the initial attempt. If the EFM calculation is intractable, NS-LLC can be used. While NS-LLC guarantees an efficient preprocessing (finding connected components from the null-space matrix), it does not ensure the largest degree of binary variable reduction as EFM-LLC does. However, the cost of EFM computation in EFM-LLC preprocessing can be unpredictably high due to combinatorial explosion (Klamt and Stelling, 2002). The information required from the computed TIC EFMs is only whether two reactions are linked by any TIC EFMs. Computing the entire set of TIC EFMs is in most cases unnecessary. If an efficient algorithm for determining whether a TIC EFM exists between any two reactions in TICs can be devised, then the
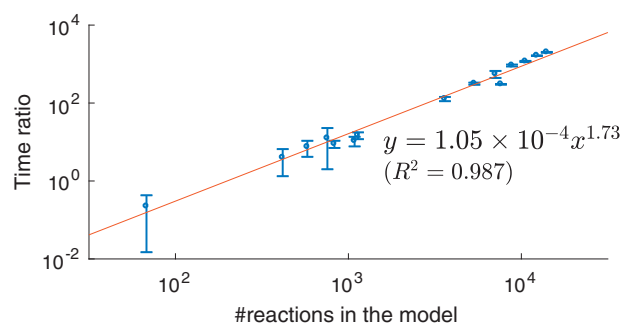
Fig. 4. Time ratios of using Fast-SNP relative to the proposed null-space algorithm for null-space calculation in all models tested

efficiency of loopless flux calculations can be further improved. While we demonstrated primarily the capability of LLCs in accelerating ll-FVA, LLCs are also applicable to other FBA-based calculations. For example, integrating LLCs into the recently proposed semi-thermodynamic FBA (Noor, 2018) can prevent TICs related to the objective function in addition to the prohibition of energy-generating cycles depending on the allowable concentration ranges for a subset of metabolites of interest. LLCs are also useful for regulatory FBA (Covert *et al.*, 2001) by imposing the loopless requirement only for reactions that participate in the regulatory constraints and as well as TICs at the same time. From a theoretical perspective, Fast-SNP introduced the minimal null-space matrix and revealed basic information about TICs in a metabolic network. The decomposition of the null-space matrix into connected components introduced in this study further characterizes the modular structure of TICs and the constituent EFMs of a metabolic network.

Funding

Conflict of Interest: none declared.

References

Carlson,R.P. (2009) Decomposition of complex microbial behaviors into resource-based stress responses. *Bioinformatics*, **25**, 90–97.

Chan,S.H.J. *et al.* (2014) Estimating biological elementary flux modes that decompose a flux distribution by the minimal branching property. *Bioinformatics*, **30**, 3232–3239.

Chan,S.H.J. *et al.* (2017a) Computational modeling of microbial communities. In: Nielsen,J. and Hohmann,S. (eds), *Systems Biology*. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, pp. 163–189.

Chan,S.H.J. *et al.* (2017b) SteadyCom: predicting microbial abundances while ensuring community stability. *PLoS Comput. Biol.*, **13**, e1005539.

Chan,S.H.J. and Ji,P. (2011) Decomposing flux distributions into elementary flux modes in genome-scale metabolic networks. *Bioinformatics*, **27**, 2256–2262.

Chowdhury,A. *et al.* (2014) k-OptForce: integrating kinetics with flux balance analysis for strain design. *PLoS Comput. Biol.*, **10**, e1003487.

Covert,M.W. *et al.* (2001) Regulation of gene expression in flux balance models of metabolism. *J. Theor. Biol.*, **213**, 73–88.

Dash,S. *et al.* (2014) Capturing the response of Clostridium acetobutylicum to chemical stressors using a regulated genome-scale metabolic model. *Biotechnol. Biofuels*, **7**, 144.

Desouki,A.A. *et al.* (2015) CycleFreeFlux: efficient removal of thermodynamically infeasible loops from flux distributions. *Bioinformatics*, **31**, 2159–2165.

Fritzemeier,C.J. *et al.* (2017) Erroneous energy-generating cycles in published genome scale metabolic networks: identification and removal. *PLoS Comput. Biol.*, **13**, e1005494–e1005414.

Gill,P.E. *et al.* (1987) Maintaining LU factors of a general sparse matrix. *Linear Algebra Appl.*, **88-89**, 239–270.

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

Heirendt,L. *et al.* (2017) Creation and analysis of biochemical constraint-based models: the COBRA Toolbox v3.0. ArXiv, 1710.04038.

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

Holzhütter,H.G. (2004) The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *Eur. J. Biochem.*, **271**, 2905–2922.

Ip,K. *et al.* (2011) Analysis of complex metabolic behavior through pathway decomposition. *BMC Syst. Biol.*, **5**, 91.

Klamt,S. and Stelling,J. (2002) Combinatorial complexity of pathway analysis in metabolic networks. *Mol. Biol. Rep.*, **29**, 233–236.

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.

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

Maranas,C.D. and Zomorrodi,A.R. (2016) *Optimization Methods in Metabolic Networks*. John Wiley & Sons, Inc., Hoboken, NJ.

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

Noor,E. (2018) Removing both Internal and Unrealistic Energy-Generating Cycles in Flux Balance Analysis. arXiv Prepr., 1803.04999.

Orth,J.D. *et al.* (2011) A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism–2011. *Mol. Syst. Biol.*, **7**, 535.

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

Ranganathan,S. *et al.* (2010) OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput. Biol.*, **6**, e1000744.

Saa,P.A. and Nielsen,L.K. (2016) Fast-SNP: a fast matrix pre-processing algorithm for efficient loopless flux optimization of metabolic models. *Bioinformatics*, **32**, 3807–3814.

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

Schuster,S. *et al.* (2002) Reaction routes in biochemical reaction systems: algebraic properties, validated calculation procedure and example from nucleotide metabolism. *J. Math. Biol.*, **45**, 153–181.

Schuster,S. and Hilgetag,C. (1994) On elementary flux modes in biochemical reaction systems at steady state. *J. Biol. Syst.*, **02**, 165–182.

Schwartz,J.-M. and Kanehisa,M. (2005) A quadratic programming approach for decomposing steady-state metabolic flux distributions onto elementary modes. *Bioinformatics*, **21**, ii204–ii205.

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

Terzer,M. and Stelling,J. (2008) Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics*, **24**, 2229–2235.

Zhao,Q. and Kurata,H. (2009) Maximum entropy decomposition of flux distribution at steady state to elementary modes. *J. Biosci. Bioeng.*, **107**, 84–89.