Bioinformatics, 33(15), 2017, 2345–2353
doi: 10.1093/bioinformatics/btx171
Advance Access Publication Date: 27 March 2017
Original Paper

Systems biology

Sequential computation of elementary modes and minimal cut sets in genome-scale metabolic networks using alternate integer linear programming

Hyun-Seob Song[1,*], Noam Goldberg[2], Ashutosh Mahajan[3] and Doraiswami Ramkrishna[4]

[1]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352, USA, [2]Department of Management, Bar-Ilan University, Ramat Gan 52900, Israel, [3]Industrial Engineering and Operations Research, IIT Bombay, Powai, Mumbai 400076, India and [4]School of Chemical Engineering, Purdue University, West Lafayette, IN 47907, USA

*To whom correspondence should be addressed.
Associate Editor: Cenk Sahinalp

Received on July 5, 2016; revised on March 8, 2017; editorial decision on March 21, 2017; accepted on March 23, 2017

Abstract

Motivation: Elementary (flux) modes (EMs) have served as a valuable tool for investigating structural and functional properties of metabolic networks. Identification of the full set of EMs in genome-scale networks remains challenging due to combinatorial explosion of EMs in complex networks. It is often, however, that only a small subset of relevant EMs needs to be known, for which optimization-based sequential computation is a useful alternative. Most of the currently available methods along this line are based on the iterative use of mixed integer linear programming (MILP), the effectiveness of which significantly deteriorates as the number of iterations builds up. To alleviate the computational burden associated with the MILP implementation, we here present a novel optimization algorithm termed alternate integer linear programming (AILP).

Results: Our algorithm was designed to iteratively solve a pair of integer programming (IP) and linear programming (LP) to compute EMs in a sequential manner. In each step, the IP identifies a minimal subset of reactions, the deletion of which disables all previously identified EMs. Thus, a subsequent LP solution subject to this reaction deletion constraint becomes a distinct EM. In cases where no feasible LP solution is available, IP-derived reaction deletion sets represent minimal cut sets (MCSs). Despite the additional computation of MCSs, AILP achieved significant time reduction in computing EMs by orders of magnitude. The proposed AILP algorithm not only offers a computational advantage in the EM analysis of genome-scale networks, but also improves the understanding of the linkage between EMs and MCSs.

Availability and Implementation: The software is implemented in Matlab, and is provided as supplementary information.

Contact: hyunseob.song@pnnl.gov

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

1 Introduction

Elementary (flux) modes (EMs) is a useful metabolic pathway concept that has been widely used in a broad range of applications, including (i) the analysis of network properties such as robustness (Behre *et al.*, 2008; Wilhelm *et al.*, 2004) and fragility (Klamt, 2006; Klamt and Gilles, 2004), (ii) the estimation of metabolic states such as flux distribution (Kurata *et al.*, 2007; Wiback *et al.*, 2004) and enzyme expression pattern (Stelling *et al.*, 2002) and (iii) guiding metabolic engineering of microbes (Trinh *et al.*, 2006, 2008). EM was also used as an essential element of dynamic metabolic modeling frameworks (Ramkrishna and Song, 2012; Song, et al., 2013), such as macroscopic bioreaction models (Provost and Bastin, 2004; Provost *et al.*, 2006), hybrid cybernetic models (HCMs) (Franz *et al.*, 2011; Kim *et al.*, 2008; Song *et al.*, 2009), and lumped HCMs (L-HCMs) (Song and Ramkrishna, 2010, 2011).

To date, EM analysis has been applied mostly to moderate-size networks with a focus on central carbon metabolism. A standalone pathway analysis tool useful for this purpose is *Metatool* (von Kamp and Schuster, 2006), further development of which is being made as part of a more general network analysis package, *CellNetAnalyzer* (Klamt and von Kamp, 2011). *Efmtool* (Terzer and Stelling, 2008, 2010; Terzer *et al.*, 2009) is a more efficient tool that leads to computation of up to millions of pathways. *Efmtool* has also been used in combination with other complementary algorithms for improved efficiency. For example, *regEfmtool* (Jungreuthmayer *et al.*, 2013), that adapted *Efmtool* to integrate with Boolean transcriptional regulatory networks, showed effectiveness in computing feasible EMs by early exclusion of infeasible modes. In a similar context, *tEMA* enabled effective computation of thermodynamically feasible EMs by incorporating metabolomic data into *Efmtool* (Gerstl *et al.*, 2015). Hunt *et al.* (2014) developed the network splitting algorithm, the integration of which with *Efmtool* led to complete identification of the full set of EMs from a genome-scale network. Despite these continued encouraging advances, application of these nullspace-based algorithms (Klamt *et al.*, 2007; Urbanczik and Wagner, 2005a,b; Wagner, 2004) to genome-scale networks remains challenging in general, due to a computational burden arising from the combinatorial explosion of EMs (Klamt and Stelling, 2002).

In a broad range of studies on metabolic network analysis, researchers seek a small subset of pathways that is specifically relevant for a given problem. As an extreme, flux balance analysis (FBA) (Orth *et al.*, 2010) estimates flux distribution using a single optimal pathway identified from linear programming (LP). Flux variability analysis (FVA) (Mahadevan and Schilling, 2003) accounts for other equivalent optimal pathways as well (albeit not all). It is also common that metabolic networks are analyzed in a low dimensional space composed of a few selected reactions, such as phenotypic space (Gayen and Venkatesh, 2006), yield space (Song and Ramkrishna, 2009), conversion cone (Urbanczik and Wagner, 2005a,b), and projected cone (Marashi *et al.*, 2012). Decomposition of an experimentally measured or computationally estimated flux (or yield) vector into a subset of EMs is another example (Badsha *et al.*, 2014; Chan and Ji, 2011; Jungers *et al.*, 2011; Song and Ramkrishna, 2009; Soons *et al.*, 2010). There has also been an increasing interest in random sampling of EMs to use as a reprehensive subset of the whole (Bohl *et al.*, 2010; Kaleta *et al.*, 2009; Machado *et al.*, 2012).

In the above applications where only a small subset of EMs needs to be known, optimization-based algorithms that enable sequential computation of EMs can serve as an alternative to nullspace-based simultaneous enumeration. Optimization problems for this purpose have been formulated commonly as a mixed integer linear programming (MILP), while other formulations were also considered (Bohl *et al.*, 2010; Kaleta *et al.*, 2009; Pey *et al.*, 2015; Quek and Nielsen, 2014). Example applications of MILP-based algorithms include identification of the first K-shortest EMs (de Figueiredo *et al.*, 2009), filtered computation of EMs that satisfy a prescribed criterion (e.g. EMs producing a specific metabolic product with a minimum yield) (Pey and Planes, 2014) and decomposition of a given flux vector into a subset of EMs (Badsha *et al.*, 2014; Chan and Ji, 2011; Chan *et al.*, 2014).

Flexible formulation of MILP that enables selective identification of EMs comes at a computational price. In comparison to LP (considered in FBA), the size of MILP is double due to the introduction of integer variables as indicators for zero and non-zero fluxes. Further, obtaining accurate solutions can be difficult due to the difference in the scales of flux and integer variables. Thus, the effectiveness of MILP significantly decreases as the number of computed EMs increases.

In this article, we proposed a *de novo* optimization concept to alleviate these problems. Instead of simultaneously solving flux and indicator variables together, we split MILP into IP and LP to alternately solve them as separate optimization problems. In each step of iteration, the IP identifies a minimal set of reactions, whose deletion disables all of the previously found EMs, thus guaranteeing that the subsequent LP solution to be a distinct EM. In comparison to the MILP formulation, tandem implementation of two half-sized optimization problems (i.e. IP and LP) is advantageous in reducing computational burden. As another key advantage, independent implementation of IP and LP removes the issue caused by different scales of the flux and integer variables, leading to more accurate solutions in comparison to MILP. This fundamentally new formulation we developed here was termed alternate integer linear programming (AILP) in contrast to MILP.

In addition to EMs, the AILP algorithm identifies 'minimal cut sets (MCSs)' during iteration. This desirable by-product is useful in metabolic engineering applications, particularly for identifying knockout targets in designing or optimizing strains (Burgard *et al.*, 2003; Choon et al., 2014). MCSs can be stated as minimal sets of reactions, the deletion of which prevents a metabolic network from achieving a prescribed objective (Klamt and Gilles, 2004); in a more general term, they can be defined as those, whose deletion disables the operation of arbitrary sets of EMs (Klamt, 2006). This implies that a set of IP-identified reactions for deletion represents an MCS, if the subsequent LP solution is infeasible. At every iteration, therefore, the AILP algorithm generates either an EM (if LP solution is feasible) or an MCS (if infeasible). Despite such an additional capability of computing MCSs, AILP achieved substantial time reduction in computing EMs in comparison to MILP. For example, we observed more than an order-of-magnitude time reduction in computing several thousands of EMs. The level of time reduction is shown more significant as the number of computed EMs increases. Beyond the computational efficiency, AILP offers other useful advantages, which we discussed through the case studies of moderate-size and genome-scale metabolic networks.

2 Materials and methods

We briefly discuss general aspects of EMs to provide background knowledge. For a metabolic network composed of n_r (internal and exchange) fluxes and m intracellular metabolites, steady-state flux distributions represent any feasible flux vectors that satisfy mass

balances of intracellular metabolites. In general, an infinite number of solutions arise from such an underdetermined system with $n_r > m$, forming an unbounded polyhedral cone in flux space (thus called *flux cone*). EMs are a special subset (among the feasible flux vectors of the cone) that collects all pathways (or sub-networks) holding non-decomposability property, i.e. pathways composed of the *minimal* number of reactions, the removal of any of which disables their operation in steady state (Klamt and Stelling, 2003). The standard way to calculate EMs is the Double-Description method (Motzkin *et al.*, 1953).

EMs can be computed in a different way based on the relationship with extreme currents (ECs) (Clarke, 1988). As a key aspect, ECs are convex bases in a flux space *expanded* by splitting reversible reactions into irreversible (forward and backward) pairs in the network The set of steady-state flux vectors obtained from the network reconfigured as such forms the flux cone **F** defined below:

$$\mathbf{F} \equiv \{\mathbf{x} \in \mathbf{R}^n : \ \mathbf{Ax} = \mathbf{0}; \quad \mathbf{0} \leq \mathbf{x}\} \tag{1}$$

where $\mathbf{A} \in R^{m \times n}$ is the stoichiometric matrix representing the balances of n ($\geq$ n$_r$) fluxes (denoted by $\mathbf{x}$) around m intracellular metabolites. Geometrically, ECs are edge vectors of the flux cone **F**.

Identification of ECs can also be formulated as a problem of identifying vertices of a polyhedron, rather than edges of a flux cone. A common way for this is to constrain a certain flux (e.g. the rate of carbon uptake or biomass production). Once ECs are identified, it is straightforward to convert them into EMs through the projection back onto the original flux space. In this procedure, therefore, ECs can be considered intermediate solutions before EMs.

In the following, we provide two alternative optimization formulations (MILP and AILP) to identify vertices of the polyhedron. As a pre-requisite for both MILP and AILP algorithms, we split reversible reactions into irreversible pairs, by which all fluxes can be assumed non-negative without loss of generality. We first use an LP to get the first optimal EM that maximizes or minimizes a given objective function; then iteratively implement MILP or AILP to sequentially compute alternative optimal or suboptimal EMs.

2.1 LP for identifying the first EM

Inspired by the work elsewhere (Song *et al.*, 2014), we formulated the following flux-minimization LP problem to obtain an EM:

$$\min \sum_{i=1}^{n} w_i x_i \tag{2}$$

such that

$$\mathbf{Ax} = \mathbf{0}, \quad \mathbf{0} \leq \mathbf{x} \tag{3}$$

$$x_{\text{ref}} = \sigma \tag{4}$$

$$x_j = 0, \quad j \in I_D \tag{5}$$

where w_i is the weight to the ith flux, x_{ref} is a reference flux that is fixed with an arbitrary positive constant σ, and I_D denotes the indices of reactions, the fluxes of which are forced to be zero. The intersection between the flux cone and the hyperplane represented by Equations (3) and (4), respectively, defines a polyhedron, the vertices of which correspond to EMs.

The property of the EM can be determined through the constraints in Equations (4) and (5), as well as the values of weighting factors in Equation (2). For example, in a metabolic network with two substitutable carbon sources S_1 and S_2 (the uptake rates of which are denoted by x_1 and x_2, respectively), an EM that converts S_1 to biomass with the highest yield can be computed by setting $w_1 = W(\gg 1)$, $w_i = 1\,(i = 2, 3, \ldots, n)$, $x_2 = 0$, and $x_B = \sigma$, where W is an arbitrarily large constant and x_B is the rate of biomass production. This setting uses a flux minimization problem to solve a yield maximization problem. Throughout this article with a focus on a yield maximization problem, we chose biomass as a target product of interest; used 0.1 for σ and an appropriately larger value ($\geq$1000) for W.

2.2 MILP scheme

Once the first EM (with the highest yield of biomass) is identified, iterative use of MILP can generate a series of EMs in a descending order of their yield values. With $(K-1)$ EMs known, the Kth EM can be computed by adding the following constraints to the LP formulation in the previous section:

$$0 \leq x_i \leq M y_i, \quad i = 1, 2, \ldots, n \tag{6}$$

$$\sum_{i \in I_{NZ,k}} y_i \leq |I_{NZ,k}| - 1, \quad k = 1, 2, \ldots, K-1 \tag{7}$$

where y_i is a binary integer identifier variable associated with x_i, M is a sufficiently large constant, $I_{NZ,k}$ is the indices of non-zero flux elements of the kth EM, and $|I_{NZ,k}|$ denotes the length of $I_{NZ,k}$ (i.e. the *number* of non-zero fluxes in $\mathbf{x}_k$). Inequality condition (6) enforces $y_i = 1$ when the corresponding x_i is non-zero, while y_i can be free (i.e. either 0 or 1) when x_i is zero. Taken together, conditions (6) and (7) force at least one of the non-zero fluxes (i.e. non-zero basic variables) of the past solutions to become zero (i.e. to become a non-basic variable) in the current Kth EM, thus guaranteeing all subsequently ensuing solutions to be distinct. In the interest of EMs that produce biomass or ATP, the stopping criterion can be the lowest values of their yields. Alternatively, the iteration can stop when the number of computed EMs reaches a certain number.

2.3 AILP scheme

In contrast with MILP that determines flux and identifier variables together in a single optimization problem, the AILP formulates an LP and an IP separately for tandem implementation. The IP identifies a minimal set of reactions, the deletion of which deactivates all of the previously found EMs, so that the subsequent LP solution is guaranteed to be distinct from the previously identified EMs. The form of LP in the AILP algorithm is the same as the one formulated in Equations (2) to (5) except that I_D in Equation (5) is updated in every step based on the output from IP.

The IP is formulated with no direct reference to a metabolic network, which leads the subsequent LP solution to become infeasible in some cases. As addressed earlier, feasible LP solutions are EMs; if no feasible LP solutions are found, LP-derived constraints become MCSs. In the case that $(K-1)$ EMs and $(L-1)$ MCSs are found, we solve the following IP towards the Kth EM or the Lth MCS depending the type of the subsequent LP solution, i.e.

$$\min \sum_i d_i (\equiv J_{\text{IP}}) \tag{8}$$

such that

$$\sum_{i \in I_{NZ,k}} d_i \geq 1, \quad k = 1, 2, \ldots, K-1 \tag{9}$$

$$\sum_{i \in I_{C,l}} d_i \leq |I_{C,l}| - 1, \quad l = 1, 2, \ldots, L-1 \tag{10}$$

where d_i is the binary integer variable denoting whether the ith flux is to be deleted (i.e. 1) or not (i.e. 0) in the subsequent LP problem, and $I_{NZ,k}$ and $I_{C,l}$ respectively denote the indices of non-zero elements of the kth EM and the indices of removed reactions that led to the lth MCS. Constraint (9) implies that at least one of the non-zero elements in each of the previous EMs is forced to be *deactivated*. In constraint (10), one of the reactions contained in each of the previously identified MCSs is forced to be *activated*. This constraint is to avoid unfruitful attempts of deleting supersets of MCSs (i.e. CSs), which are not our interest. While mentioned as IP for simplicity, Equations (8) to (10) represents a binary IP (BIP) problem because integer variables d_i are allowed to take only 0 or 1.

The problem formulated above is translated into a classical combinatorial question known as 'set-covering' or 'hitting-set problem,' which is a basic model widely used in mathematical and engineering applications (Ceria *et al.*, 1997) as well as in calculating MCSs. In the situation where all EMs are known, Hädicke and Klamt (2011) developed the adaptive Berge algorithm to calculate hitting sets that remove target EMs without destroying a given number of desirable EMs, which were termed constrained MCSs (cMCSs). Enumeration of cMCSs was later formulated as a BIP problem (Jungreuthmayer and Zanghellini, 2012; Jungreuthmayer *et al.*, 2013). Calculating hitting sets in our problem is relatively simpler as there is no need to impose those constraints. Thus, our algorithm can be viewed as a special case of the BIP for computing cMCS.

As illustrated in Figure 1, the AILP primarily uses LP to identify EMs in combination with IP. The role of this auxiliary algorithm (i.e. IP) is to provide constraints to make the next EM to be distinct during iteration. Combining LP with a secondary algorithm for the sequential computation of EMs is a basic structure commonly used in other related approaches. The work by Kaleta *et al.* (2009) may be considered one of the first attempts to use LP to calculate EMs from genome-scale metabolic networks. They additionally used a genetic algorithm to explore the solution space in search of EMs of interest. The TreeEFM algorithm by Pey *et al.* (2015) or the algorithm by Quek and Nielsen (2014) improved the computational efficiency by combining LP with a tree search procedure. In contrast, our method used IP, which desirably enables the systematic identification of MCSs, in addition to EMs. In this regard, the AILP holds a connection to the approach by Haus *et al.* (2008). They also incrementally computed MCSs and EMs, but using a completely different algorithm by Fredman and Khachiyan (1996). Based on the dual relationship between MCSs and EMs, they formulated a problem of finding minimal true assignments (for MCSs) and (the complement of) maximal false assignments (for EMs) in Boolean representation of reactions. The key distinction of our approach from Haus *et al.* lies in the use of an optimization algorithm, by which the selective identification of EMs and MCSs can be directed into certain sets of interest through the flexible design of objective functions and constraints. Setting of the objective function in (8), for example, enables finding blocked reactions with a minimal size. Therefore, when blocked reactions turn out to be MCSs, they will be MCSs with the smallest size, which is a critically useful aspect for metabolic engineering application. Similarly, AILP can selectively identify MCSs with a certain range of size.

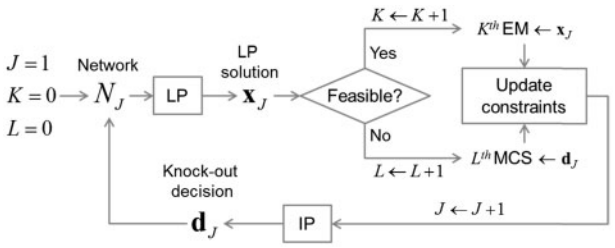

Fig. 1. Iterative procedures of computing EMs and MCSs in AILP. Indices J, K, and L denote the numbers of total iteration, calculated EMs and calculated MCSs, respectively. Symbols $\mathbf{d}_J$, N_J, and $\mathbf{x}_J$ represent the IP-derived reaction deletion set, the resulting subnetwork, and the subsequent LP solution at the Jth iteration

2.4 Metabolic network models used for testing algorithms

We used small- and large-scale metabolic network models to comparatively evaluate AILP and MILP algorithms. Small-size networks include the central carbon metabolism of *Escherichia coli* growing glucose (Carlson and Srienc, 2004) and genetically engineered yeast strain consuming glucose and xylose (Song *et al.*, 2009). *E. coli* network represents both aerobic and anaerobic growth on glucose. The original *E. coli* network contained 36 metabolites and 45 reactions, to which we added the oxygen exchange reaction (R98: OXY_ext = OXY) for the sake of convenience. The network of recombinant yeast describes the anaerobic growth on glucose or xylose. We used the recombinant yeast network without modification, which was composed of 30 metabolites and 37 reactions. After splitting reversible reactions into irreversible pairs, the numbers of reactions of the small-size *E. coli* and yeast networks were expanded to 63 and 51, respectively.

As large-size network examples, we chose genome-scale networks of *Saccharomyces cerevisiae*, iND750 (Duarte *et al.*, 2004) and *E. coli* iAF1260 (Feist *et al.*, 2007), which were composed of 1061 metabolites and 1266 reactions and 1668 metabolites and 2382 reactions, respectively. We obtained the Systems Biology Markup Language (SBML) files of iND750 and iAF1260 from BiGG (Schellenberger *et al.*, 2010). After splitting of reversible reactions, the sizes of the iND750 and iAF1260 were expanded to contain 1702 and 2956 reactions, respectively.

We ran MILP and AILP algorithms using the CPLEX (ILOG, Mountain View, CA) solvers. The entire program was written in MATLAB (Mathworks, Inc., Natick, MA). We conducted all computations on a workstation with Intel Xeon CPU 2.4 GHz processor and 128 GB RAM. We provided metabolic network models (small-size networks as *Metatool* files; genome-scale networks as SBML files) (Supplementary Dataset S1) and the MATLAB scripts used for the reproduction of the results (Supplementary Dataset S2).

3 Results

We illustrated the concept of the AILP algorithm using a toy network (Fig. 2). All fluxes were treated as non-negative by splitting the reversible reaction (i.e. R_4) into forward and backward pair (denoted by $R_{4,f}$ and $R_{4,b}$) (Fig. 2a). We then formulated a biomass-yield maximizing LP problem for the reconfigured network (Fig. 2b). The LP is solved under reaction deletion constraint as

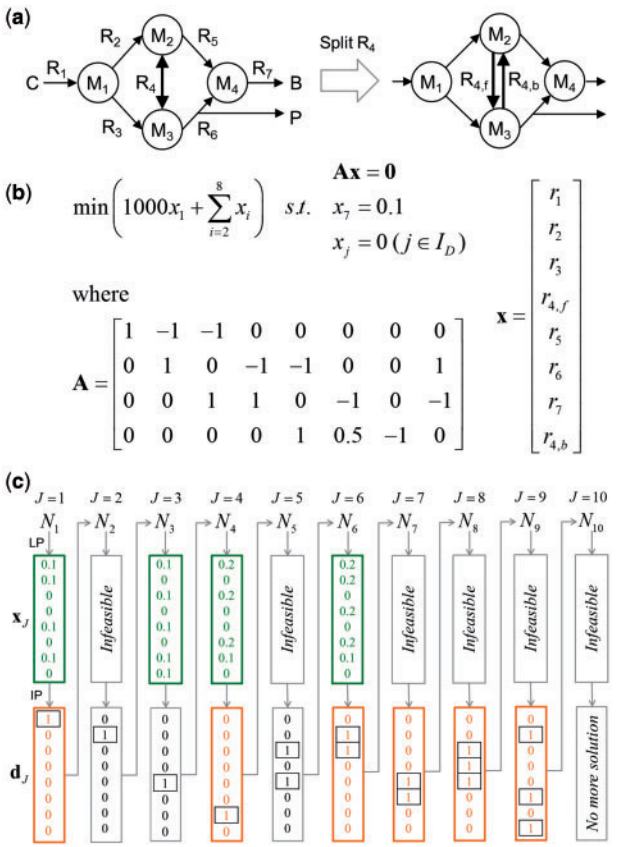

Fig. 2. Tutorial example problem considered for illustrating the concept of the proposed AILP algorithm: (**a**) the original and augmented network, (**b**) the formulation of LP for computing EMs, and (**c**) step-by-step procedures of computing EMs and MCSs. N_J, $\mathbf{x}_J$, and $\mathbf{d}_J$, respectively denote the metabolic network and the corresponding LP solution at the Jth iteration, and the IP-determined integer vector whose non-zero elements represent the list of reactions to be deleted in the next iteration

denoted by I_D, which is updated at every iteration by solving the associated IP problem formulated in Equations (8) to (10). The iterative implementation of the AILP was described as a process of generating sets of LP ($\mathbf{x}_J$) and IP solutions ($\mathbf{d}_J$) in series (Fig. 2c).

We denoted the intact network by N_1 and subnetworks by N_2 to N_{10}. Subnetworks were created at every step by deactivating a minimal number of reactions (i.e. boxed elements in $\mathbf{d}_J$'s) using the IP. In this example, the complete identification of EMs and MCSs required 10 iterations in total (note that the last iteration was necessary to confirm no further LP nor IP solution to be available). Non-zero elements in $\mathbf{d}_1$, $\mathbf{d}_4$, and $\mathbf{d}_6-\mathbf{d}_9$ (in orange boxes) that led to infeasible LP solutions are MCSs represent MCSs, the deletion of which prevents the network from producing biomass. Among four feasible LP solutions (in green boxes) (i.e. EMs), $\mathbf{x}_1$ and $\mathbf{x}_3$ represent two alternative optimal pathways of the highest biomass yield; $\mathbf{x}_4$ and $\mathbf{x}_6$ the next suboptimal. Identification of EMs in a decreasing order is only a special case we encountered in this specific example because the order is in general irregular in AILP. The final forms of EMs and MCSs can be obtained simply by the projection of $\mathbf{x}_J(J=1,3,4,6)$ and $\mathbf{d}_J(J=1,4,6-9)$ back onto the original flux space.

3.1 Basic properties of AILP solutions

We applied both AILP and MILP-based algorithms to central carbon metabolic networks of *E. coli* and recombinant yeast described in the previous section. While simultaneous computation of EMs (using *Metatool* or *Efmtool*) would be the most efficient for such small networks, the goal here is to discuss basic properties of AILP in comparison to the typical MILP implementation.

First, we checked how correctly AILP and MILP algorithms are able to generate the intended solutions. This examination using small-size networks is important for assuring the reliability of the algorithms when applied to genome-scale networks. For this purpose, we considered the following three cases: (i) yeast growth on glucose and (ii) yeast growth on xylose, (iii) *E. coli* growth on glucose; examined how the parameter sets in AILP and MILP algorithms affect the correctness of solutions. While not explicitly shown in Equations (6) to (10), the common parameter in both algorithms is the threshold value of ε required for determining fluxes to be zero or non-zero. This parameter is to account for possible errors in numerical precision, e.g. caused by truncation of stoichiometric coefficients and flux variables. Another parameter, that appears only in the MILP but not in AILP, is M contained in (6). The value of M should be large enough so as not to cut off any part of the solution space, but setting it above a certain value may make the problem *ill-conditioned*. Similar issues associated with the choice of M have also been discussed elsewhere (David and Bockmayr, 2014). Appropriate values for ε and M would vary depending on a problem setting (e.g. a chosen value of σ and lower and upper bounds of fluxes). With a focus on the computation of biomass-producing EMs, we evaluated the solution accuracy of MILP and AILP. Both algorithms accurately computed EMs for the two yeast networks with parameter settings considered in Table S1. For the *E. coli* network (which is relatively larger in size than yeast networks), however, MILP failed to completely identify EMs using the same range of parameters considered in the previous case, while AILP worked correctly. The ε value for MILP that we found worked the best was 10^{-3}, which was larger than the value for AILP by more than two-orders. The impact of M (another MILP parameter) was not pronounced in the range of 10^3 and 10^5 in these examples.

Next, we compared the difference between MILP and AILP in the order of EM identification using the case of the small *E. coli* network as an example. As expected, MILP generated EMs in a monotonically descending order of biomass yield (Fig. 3a top). In contrast, AILP successfully identified the complete set, but in an irregular order (middle). This should not be considered a disadvantage because a simple post-sorting results in ordered EMs. It is important to note that AILP is also able to selectively collect EMs that satisfy a certain criterion, for example, normalized biomass yield ($Y_{B,rel}$) ≥ 0.5 (bottom). For simplicity, we denoted the results of AILP subject to this constraint by a subscript. That is, AILP_0 and $\text{AILP}_{0.5}$ in Figure 3 denote the results from AILP when $Y_{B,rel}>0$ and $Y_{B,rel}\geq 0.5$ were imposed as constraint, respectively. MCSs identified under the constraint of $Y_{B,rel}\geq 0.5$ denote the sets of reactions, the deletion of which disables all EMs whose biomass yield is above or equal to 0.5. This also implies that the maximum obtainable biomass yield from MCS-deleted subnetworks (or mutant networks) is less than 0.5.

We also compared the computational efficiency between MILP and AILP in calculating biomass-producing EMs from the *E. coli* network. As shown in Figure 3b shows the elapsed time of MILP (t_{MILP}) exponentially increased with the number of calculated EMs. The computational time of AILP to obtain biomass-producing EMs (t_{AILP_0}) was 1.5 times shorter than that of MILP, while the time to selectively compute EMs with $Y_{B,rel}\geq 0.5$ ($t_{AILP_{0.5}}$) was the almost the same between AILP and MILP. The initial inefficiency of AILP as shown by the abrupt increase of t_{AILP_0} and $t_{AILP_{0.5}}$ at discrete

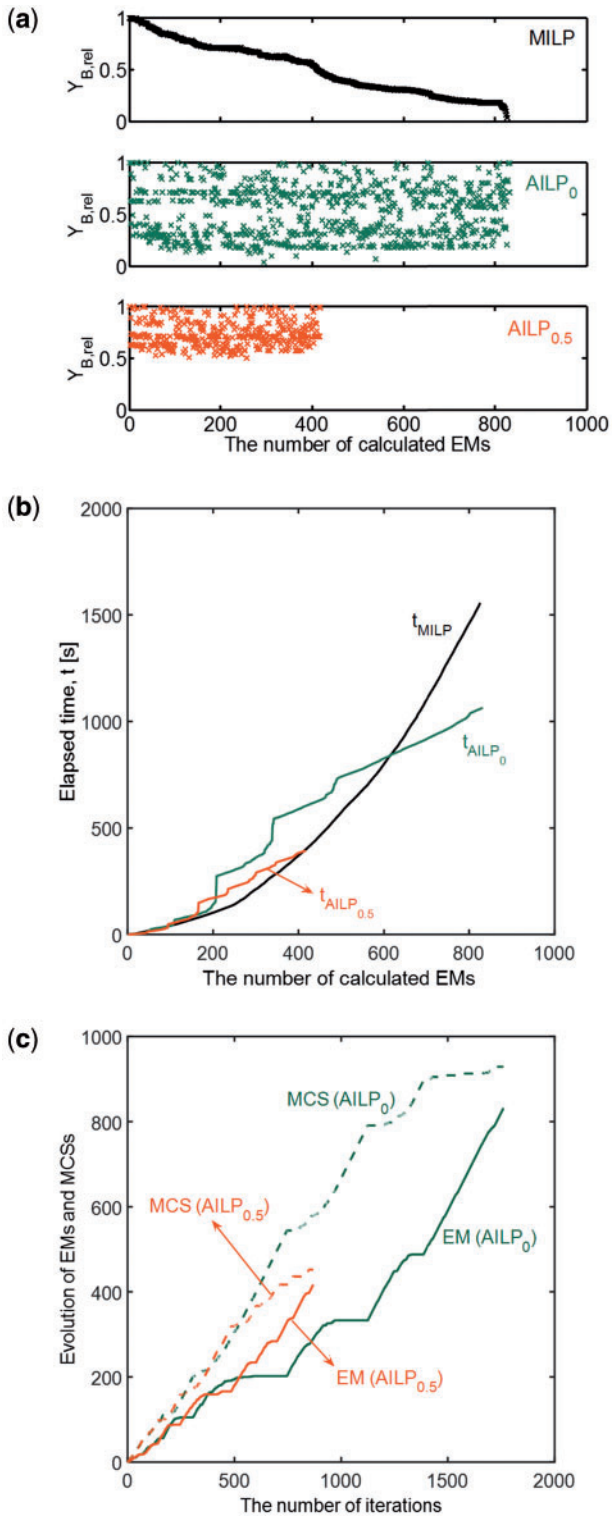

Fig. 3. Comparison of MILP and AILP in computing EMs from the small-size *E. coli* network: (**a**) pattern of EM identification: MILP (top) and AILP (middle and bottom), (**b**) profiles of elapsed time for computing EMs from MILP and AILP, and (**c**) evolution of MCSs and EMs with iterations. The subscripts in $AILP_0$ and $AILP_{0.5}$ denote the constraints imposed on AILP such as $Y_{B,rel} \geq 0$ and $Y_{B,rel} \geq 0.5$

points is due to additional computation of MCSs, i.e. when IP-derived reaction deletion constraints do not lead to feasible LP solutions. The linkage between MCS occurrence and EM computation was provided in Supplementary Text.

In Figure 3c, we showed how AILP alternately generated MCSs and EMs with iterations. The evolutions of MCSs and EMs followed non-linear patterns composed of repeated sets of linearly-evolving and stagnant regimes. The occurrence of these two regimes was exactly opposite between EMs and MCSs, i.e. EMs linearly evolved whenever MCSs stopped advancing and *vice versa*. The evolution patterns from $AILP_0$ and $AILP_{0.5}$ were initially similar, while they eventually bifurcated after about 500 iterations. Interestingly, in both cases, we observed the computation of MCSs was terminated earlier than EMs, which made EM profiles linearly uprising toward the end of iterations. This might be case-dependent as we observed the opposite (i.e. earlier termination of EMs than MCSs) in the analysis of the tutorial network (Fig. 2c).

In Supplementary Table S2, we considered various other cases with different biomass yield constraints to compare the computational efficiency between AILP with MILP. For fair comparison, we split the total time of AILP (t_{AILP}) into the time consumed for computing EMs ($t_{AILP,EM}$) and the time consumed for computing MCSs ($t_{AILP,MCS}$) to evaluate the relative performance of MILP and AILP in terms of $t_{MILP}/t_{AILP,EM}$ and t_{MILP}/t_{AILP}. As expected, the range of $t_{MILP}/t_{AILP,EM}$ was higher (i.e. from 1.3 to 2.7) than that of t_{MILP}/t_{AILP} (i.e. from 0.4 to 1.5). Although the overall computational efficiency of AILP was shown greater than that of MILP despite additional computation of MCSs in these cases, we expect that this difference may not be significant for small-size networks in general. It is also obvious that in comparison to nullspace-based algorithms, the use of optimization algorithms is not computationally advantageous for the analysis of small scale networks. For example, *Efmtool* required less than 2 s to compute the full set of EMs (that includes non-biomass-producing modes as well) from the same network considered here, while AILP took way longer times even for computing only subsets (Supplementary Table S2). In the following section, we comparatively test the effectiveness of AILP and MILP for complex and large-size metabolic networks.

3.2 Evaluation of computational efficiency of AILP and MILP using genome-scale networks

We applied AILP to genome-scale networks of *S. cerevisiae* iND750 (Duarte *et al.*, 2004) and *E. coli* iAF1260 (Feist *et al.*, 2007). In computing EMs, we considered glucose as a sole carbon source and took the maximization of biomass yield as the objective function, in both cases.

The number of EMs can be scaled to millions in genome-scale networks, but we confined the analysis to the cases of computing the first several thousand EMs. We compared computational efficiency between MILP and AILP when 6000 EMs were enumerated from iND750 (Fig. 4a) and 3000 EMs from iAF1260 (Fig. 4b). In both cases, MILP showed sharp exponential growth in computational time with the increasing number of computed EMs; in contrast, the increase of computational time in AILP was moderate (top panels). We also displayed the ratio of computational times between MILP (t_{MILP}) and and AILP (t_{AILP}) (bottom panels). In the case of iND750 (bottom of the Fig. 4a), the time reduction by AILP showed three distinct phases: (i) an initial abrupt rise and drop during the computation of 1000 EMs, (ii) almost constant until 3000 EMs were calculated and (iii) gradual increase afterwards. Time reduction by AILP was about 10 to 11 times around the peak and about four times when 6000 EMs was calculated. The difference between $AILP_0$ and $AILP_{0.5}$ was insignificant. We observed a similar pattern for

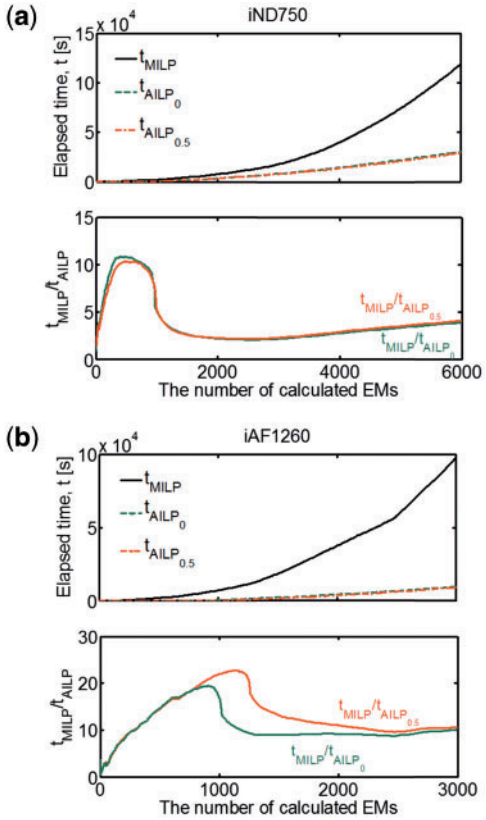

Fig. 4. Comparison of MILP and AILP in computing EMs from genome-scale networks: (**a**) *S. cerevisiae* network iND750 and (**b**) *E. coli* network iAF1260. Top panels compare elapsed times of MILP and AILP; bottom panels show time reduction by AILP. The subscripts in $AILP_0$ and $AILP_{0.5}$ denote the constraints of $Y_{B,rel} \geq 0$ and $Y_{B,rel} \geq 0.5$, respectively, imposed during AILP implementation

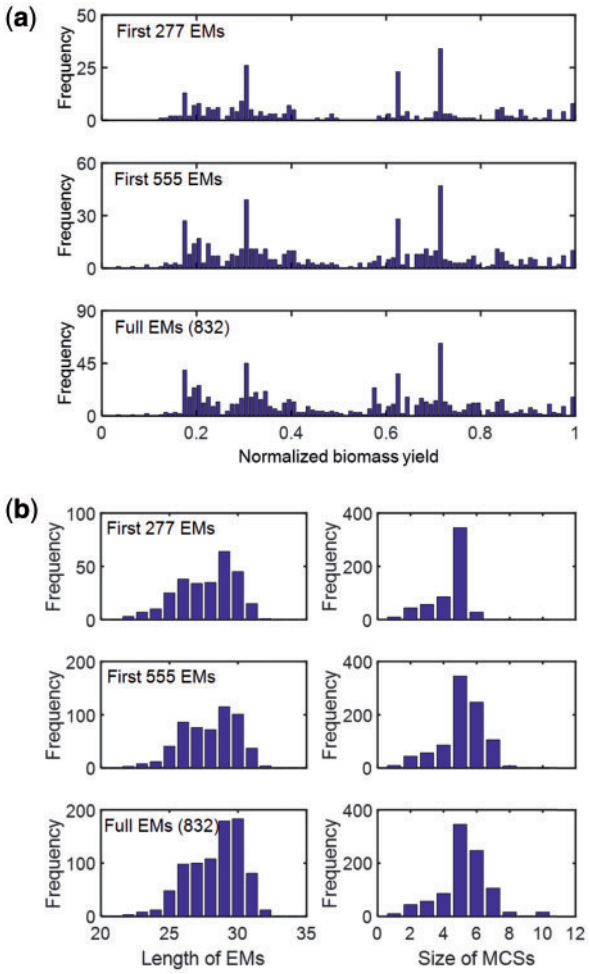

Fig. 5. Analysis of EMs and MCSs computed from the small-size *E. coli* network: distribution of (**a**) normalized biomass yields and (**b**) EM length and MCS size. The distribution was compared among the three different sizes of EM subsets (as denoted by the number inside each panel) and among three MCS subsets computed along with the corresponding EM subsets

iAF1260, while t_{MILP}/t_{AILP} showed appreciable difference between $AILP_0$ and $AILP_{0.5}$ in this case (bottom of Fig. 4b): AILP reduced the computational time by a factor of 19 ($AILP_0$) and 23 ($AILP_{0.5}$) around the peaks and by a factor of about 10 ($AILP_0$) and 11 ($AILP_{0.5}$) at the computation of 3000 EMs.

The effectiveness of AILP in computing a given number of EMs was shown greater when network size was larger. Time reduction by AILP in computing 3000 EMs for iAF1260 was more than twice of that for iND750. The relative performance of AILP was also shown greater as the iteration goes on. In the case of iND750, $t_{MILP}/t_{AILP_{0.5}}$ increased (almost in a linear fashion) from 2.3 to 4.1 when 3000 and 6000 EMs were calculated, respectively. Similarly, $t_{MILP}/t_{AILP_{0.5}}$ for iAF1260 was 9.7 with 2500 EMs, but became 10.6 with 3000 EMs. When extrapolated to many more EMs, the effectiveness of AILP in time reduction in comparison to MILP is estimated to be orders of magnitude. These results highlight the circumstances where the utility of AILP can be fully beneficial.

3.3 Properties of initial subsets of EMs as samples

As discussed earlier, AILP generates EMs irregularly (while not randomly). This feature brings a possibility of using an initial set of EMs as samples of the entire set without having to wait for the completion of long-lasting computation. We examined the extent to which initially computed subsets of EMs may represent to the entire set with respect to distributions of relative fluxes of consumption and production of extracellular metabolites including biomass.

Figure 5a shows the distribution of normalized biomass yield of EMs obtained from the small-scale *E. coli* network. The distributions were similar between the full 832 biomass-producing EMs (bottom), and EMs sampled by taking initial two-thirds (middle) and one-third (top) of the whole. We made the same observation for all other exchange fluxes in the small-size *E. coli* network (Supplementary Fig. S1). In Figure 5(b), we also provided the distributions of EM length (left column) and MCS size (right column). The distribution of EM length was also similar among three cases. Notably, we observed no appreciable bias into smaller-size EMs even in the first subset (i.e. the 277 EMs identified during the initial iterations). In contrast, the distribution of MCS size developed longer tails to the right tail as the size of EMs increased, because we formulated IP such that MCSs are identified starting from the smallest size progressively toward longer size.

Similar trends were observed for the genome-scale yeast network iND750 (Supplementary Fig. S2). As the whole set of EMs is not available in this case, we compared EMs when the different number of EMs were taken, i.e. first 2000 (top), 4000 (middle) and 6000 (bottom). As in the case of small *E. coli* network, the distribution of normalized biomass yield (Supplementary Fig. S2a) and EM length (Supplementary Fig. S2b) were consistent across three EM subsets.

Besides biomass yield, we also checked that distributions of all other exchange fluxes did not change much (results not shown).

These results imply that initial subsets of EMs may be used to analyze metabolic networks even without identifying the entire set of EMs. This is important in the analysis of large-scale metabolic networks, for which the enumeration of EMs of interest may take a significantly long time.

4 Discussion

There are several unique features of the AILP in comparison to the MILP as summarized as follows. First, while the MILP computes EMs in a descending (or ascending) order of a chosen property (e.g. biomass yield), the pattern of identifying EMs by the AILP is somewhat irregular, although not purely random. Thus, if the iteration is terminated early before meeting the stopping criterion, the AILP provides a subset of EMs, whose property of interest (e.g. biomass yield) is distributed over a given range. These EMs may be used as a sample subset for analyzing metabolic networks. Second, the AILP computes MCSs as a bonus by-product. Unlike the way to generate EMs, the AILP identifies MCSs in a *regular* pattern with respect their sizes. That is, the minimization problem formulated in the IP initially generates MCSs of the smallest size and progressively increases their size. This feature is useful for metabolic engineering because MCS of the smallest size will help to minimize the genetic intervention in designing new strains. Third, the AILP offers significant advantages in the speed of computing EMs over the MILP. While the computational burden will gradually increase both in MILP and AILP due to the increasing size of the constraints as the iteration builds up, the resulting inefficiency is far less for AILP that determines integer and flux variables separately, than MILP that determines them simultaneously. The increasing size of the constraints in the AILP affects only the efficiency of the IP, not of the LP. Also note that even though both MILP and IP are NP-hard problems in theory (Garey and Johnson, 1979), the latter takes much less time to solve in practice. LP, on the other hand, can be solved in time polynomial in size of the input. Finally, the AILP formulates a well-posed optimization problem, while the MILP formulation can be ill-posed, particularly as the iteration goes on, due to two types of mixed variables with different scales. Therefore, scaling is an important issue to resolve in the MILP, but not in the AILP.

In comparison to individual enumeration of EMs and MCSs, the use of the AILP algorithm is also beneficial because it generates both of them, implications of which can be intuitively translated. From a sequential series of EMs and MCSs, one can instantaneously identify (i) through what metabolic pathways the network can achieve a prescribed goal (e.g. biomass yield higher than a given threshold) and (ii) what reaction knockout scenarios will stop the network from achieving it. This capability can facilitate metabolic engineering efforts towards the identification of target pathways and fluxes for amplification or deactivation in developing industrially useful strains. The AILP algorithm is also useful for drug design research as it can identify the smallest subsets of genes and reactions fatal to pathogen growth. Altogether, this information is key input for robustness analysis of metabolic networks, which can be conveniently provided from a single algorithm such as the AILP method we proposed here.

Our work was primarily motivated by the need to identify a chosen subset of EMs to be used as input to dynamic metabolic network modeling such as L-HCM. There is an advantage to using a method that progressively computes EMs with prospects of terminating the process when the ones collected are sufficient for the purpose instead of computing the entire set of EMs. The AILP algorithm can selectively identify those EMs (needed for dynamic metabolic modeling) from a constrained space. This capability can also be useful for other metabolic pathway analysis tools including control-effective fluxes (CEFs) (Stelling *et al.*, 2002) and Computational Approach for Strain Optimization aiming at high Productivity (CASOP) (Hadicke and Klamt, 2010). These frameworks in common estimate a flux distribution in a network by taking a weighted average of EMs. In this case, only a subset of individual EMs with higher weights (such as those that can produce biomass, ATP or any metabolites of interest at higher rates or yields) needs to be identified. The utility of the AILP algorithm also extends to the analysis of metabolic networks composed of multiple interacting organisms (Henry *et al.*, 2016; Song *et al.*, 2014). Various desirable properties of the AILP algorithm discussed in this work will enable facilitating the EM analysis of microbial community networks [e.g. the work by Taffs *et al.* (2009)] to be applicable to more complex systems.

Funding

This research was supported by the US Department of Energy (DOE) Office of Biological and Environmental Research (BER), as part of Foundational Scientific Focus Area (SFA) and Subsurface Biogeochemistry Research Program's SFA at the Pacific Northwest National Laboratory (PNNL) (H.S.S.). The PNNL is operated for DOE by Battelle Memorial Institute under contract DE-AC05-76RL01830. Authors H.S.S. and D.R. also acknowledge support from the Center for Science of Information (MCSoI), a National Science Foundation (NSF) Science and Technology Center, under grant agreement CCF-0939370.

Conflict of Interest: none declared.

References

Badsha,M.B. *et al.* (2014) Complementary elementary modes for fast and efficient analysis of metabolic networks. *Biochem. Eng. J.*, **90**, 121–130.

Behre,J. *et al.* (2008) Structural robustness of metabolic networks with respect to multiple knockouts. *J. Theor. Biol.*, **252**, 433–441.

Bohl,K. *et al.* (2010) CASOP GS: computing intervention strategies targeted at production improvement in genome-scale metabolic networks. *Proceedings of the 25th German Conference on Bioinformatics. Bonn, Gesellschaft f. Informatik*, pp. 71–80.

Burgard,A.P. *et al.* (2003) OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.*, **84**, 647–657.

Carlson,R. and Srienc,F. (2004) Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: identification of reactions. *Biotechnol. Bioeng.*, **85**, 1–19.

Ceria,S. *et al.* (1997) Set covering problem. In: Dell'Amico, M., Maffioli, F. and Martello, S. (eds), *Annotated Bibliographies in Combinatorial Optimization*. John Wiley and Sons, New York, pp. 415–428.

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

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.

Choon,Y.W. *et al.* (2014) A hybrid of bees algorithm and flux balance analysis with OptKnock as a platform for in silico optimization of microbial strains. *Bioproc. Biosyst. Eng.*, **37**, 521–532.

Clarke,B.L. (1988) Stoichiometric network analysis. *Cell Biophys.*, **12**, 237–253.

David,L. and Bockmayr,A. (2014) Computing elementary flux modes involving a set of target reactions. *IEEE ACM Trans. Comput. Biol. Bioinform.*, **11**, 1099–1107.

de Figueiredo,L.F. *et al.* (2009) Computing the shortest elementary flux modes in genome-scale metabolic networks. *Bioinformatics*, **25**, 3158–3165.

Duarte,N.C. *et al*. (2004) Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res*., **14**, 1298–1309.

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

Franz,A. *et al*. (2011) Experimental and theoretical analysis of poly(beta-hydroxybutyrate) formation and consumption in *Ralstonia eutropha*. *Biochem. Eng. J*., **55**, 49–58.

Fredman,M.L. and Khachiyan,L. (1996) On the complexity of dualization of monotone disjunctive normal forms. *J. Algorithm*., **21**, 618–628.

Garey,M.R. and Johnson,D.S. (1979) *Computers and Intractability: A Guide to the Theory of NP-Completeness* W. H. Freemand and Company, New York.

Gayen,K. and Venkatesh,K.V. (2006) Analysis of optimal phenotypic space using elementary modes as applied to *Corynebacterium glutamicum*. *BMC Bioinformatics*, **7**, 445.

Gerstl,M.P. *et al*. (2015) tEFMA: computing thermodynamically feasible elementary flux modes in metabolic networks. *Bioinformatics*, **31**, 2232–2234.

Hadicke,O. and Klamt,S. (2010) CASOP: a computational approach for strain optimization aiming at high productivity. *J. Biotechnol*., **147**, 88–101.

Hadicke,O. and Klamt,S. (2011) Computing complex metabolic intervention strategies using constrained minimal cut sets. *Metab. Eng*., **13**, 204–213.

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

Henry,C.S. *et al*. (2016) Microbial community metabolic modeling: a community data-driven network reconstruction. *J. Cell. Physiol*., **231**, 2339–2345.

Hunt,K.A. *et al*. (2014) Complete enumeration of elementary flux modes through scalable demand-based subnetwork definition. *Bioinformatics*, **30**, 1569–1578.

Jungers,R.M. *et al*. (2011) Fast computation of minimal elementary decompositions of metabolic flux vectors. *Automatica*, **47**, 1255–1259.

Jungreuthmayer,C. *et al*. (2013) Comparison and improvement of algorithms for computing minimal cut sets. *BMC Bioinformatics*, **14**,

Jungreuthmayer,C. *et al*. (2013) regEfmtool: speeding up elementary flux mode calculation using transcriptional regulatory rules in the form of three-state logic. *Biosystems*, **113**, 37–39.

Jungreuthmayer,C. and Zanghellini,J. (2012) Designing optimal cell factories: integer programming couples elementary mode analysis with regulation. *BMC Syst. Biol*., **6**,

Kaleta,C. *et al*. (2009) EFMEvolver: computing elementary flux modes in genome-scale metabolic networks. *Lecture Notes in Informatics-Proceedings*. pp. 179–189.

Kim,J.I. *et al*. (2008) A hybrid model of anaerobic *E. coli* GJT001: combination of elementary flux modes and cybernetic variables. *Biotechnol. Prog*., **24**, 993–1006.

Klamt,S. (2006) Generalized concept of minimal cut sets in biochemical networks. *Biosystems*, **83**, 233–247.

Klamt,S. and Gilles,E.D. (2004) Minimal cut sets in biochemical reaction networks. *Bioinformatics*, **20**, 226–234.

Klamt,S. *et al*. (2007) Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Syst. Biol*., **1**, 2.

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

Klamt,S. and Stelling,J. (2003) Two approaches for metabolic pathway analysis? *Trends Biotechnol*., **21**, 64–69.

Klamt,S. and von Kamp,A. (2011) An application programming interface for CellNetAnalyzer. *Biosystems*, **105**, 162–168.

Kurata,H. *et al*. (2007) Integration of enzyme activities into metabolic flux distributions by elementary mode analysis. *BMC Syst. Biol*., **1**, 31.

Machado,D. *et al*. (2012) Random sampling of elementary flux modes in large-scale metabolic networks. *Bioinformatics*, **28**, I515–I521.

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.

Marashi,S.A. *et al*. (2012) Analysis of metabolic subnetworks by flux cone projection. *Algorithms Mol. Biol*., **7**, 17.

Motzkin,T.S. *et al*. (1953) The double description method. Contributions to the Theory of Games. *Ann. Math. Stud*., **2**, 57–73.

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

Pey,J. and Planes,F.J. (2014) Direct calculation of elementary flux modes satisfying several biological constraints in genome-scale metabolic networks. *Bioinformatics*, **30**, 2197–2203.

Pey,J. *et al*. (2015) TreeEFM: calculating elementary flux modes using linear optimization in a tree-based algorithm. *Bioinformatics*, **31**, 897–904.

Provost,A. and Bastin,G. (2004) Dynamic metabolic modelling under the balanced growth condition. *J. Process Control*, **14**, 717–728.

Provost,A. *et al*. (2006) Metabolic design of macroscopic bioreaction models: application to Chinese hamster ovary cells. *Bioprocess. Biosyst. Eng*., **29**, 349–366.

Quek,L.E. and Nielsen,L.K. (2014) A depth-first search algorithm to compute elementary flux modes by linear programming. *BMC Syst. Biol*., **8**,

Ramkrishna,D. and Song,H.S. (2012) Dynamic models of metabolism: review of the cybernetic approach. *Aiche J*., **58**, 986–997.

Schellenberger,J. *et al*. (2010) BiGG: a biochemical genetic and genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics*, **11**, 213.

Song,H.-S. *et al*. (2014) Mathematical modeling of microbial community dynamics: a methodological review. *Processes*, **2**, 711–752.

Song,H.S. *et al*. (2013) Modeling metabolic systems: the need for dynamics. *Curr. Opin. Chem. Eng*., **2**, 373–382.

Song,H.S. *et al*. (2009) Systematic development of hybrid cybernetic models: application to recombinant yeast co-consuming glucose and xylose. *Biotechnol. Bioeng*., **103**, 984–1002.

Song,H.S. and Ramkrishna,D. (2009) Reduction of a set of elementary modes using yield analysis. *Biotechnol. Bioeng*., **102**, 554–568.

Song,H.S. and Ramkrishna,D. (2010) Prediction of metabolic function from limited data: lumped hybrid cybernetic modeling (L-HCM). *Biotechnol. Bioeng*., **106**, 271–284.

Song,H.S. and Ramkrishna,D. (2011) Cybernetic models based on lumped elementary modes accurately predict strain-specific metabolic function. *Biotechnol. Bioeng*., **108**, 127–140.

Song,H.S. *et al*. (2014) Prediction of metabolic flux distribution from gene expression data based on the flux minimization principle. *PLOS ONE*, **9**, e112524.

Soons,Z. *et al*. (2010) Selection of elementary modes for bioprocess control. *Comput. Appl. Biotechnol*., **11**, 156–161.

Stelling,J. *et al*. (2002) Metabolic network structure determines key aspects of functionality and regulation. *Nature*, **420**, 190–193.

Taffs,R. *et al*. (2009) In silico approaches to study mass and energy flows in microbial consortia: a syntrophic case study. *BMC Syst. Biol*., **3**, 114.

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

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

Terzer,M. and Stelling,J. (2010) Parallel extreme ray and pathway computation. *Parallel Processing and Applied Mathematics, PAM 2009*. Lecture Notes in Computer Science, Vol. **6068**. Springer, Berlin, Heidelberg, pp. 300–309.

Trinh,C.T. *et al*. (2006) Design, construction and performance of the most efficient biomass producing *E. coli* bacterium. *Metab. Eng*., **8**, 628–638.

Trinh,C.T. *et al*. (2008) Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses. *Appl. Environ. Microbiol*., **74**, 3634–3643.

Urbanczik,R. and Wagner,C. (2005a) Functional stoichiometric analysis of metabolic networks. *Bioinformatics*, **21**, 4176–4180.

Urbanczik,R. and Wagner,C. (2005b) An improved algorithm for stoichiometric network analysis: theory and applications. *Bioinformatics*, **21**, 1203–1210.

von Kamp,A. and Schuster,S. (2006) Metatool 5.0: fast and flexible elementary modes analysis. *Bioinformatics*, **22**, 1930–1931.

Wagner,C. (2004) Nullspace approach to determine the elementary modes of chemical reaction systems. *J. Phys. Chem. B*, **108**, 2425–2431.

Wiback,S.J. *et al*. (2004) Using metabolic flux data to further constrain the metabolic solution space and predict internal flux patterns: the *Escherichia coli* spectrum. *Biotechnol. Bioeng*., **86**, 317–331.

Wilhelm,T. *et al*. (2004) Analysis of structural robustness of metabolic networks. *Syst. Biol. (Stevenage)*, **1**, 114–120.