

RESEARCH ARTICLE

Efficient Modeling of MS/MS Data for Metabolic Flux Analysis

Naama Tepper, Tomer Shlomi*

Department of Computer Science, Technion, Israel Institute of Technology, Haifa, Israel

* tomersh@cs.technion.ac.il

Abstract

Metabolic flux analysis (MFA) is a widely used method for quantifying intracellular metabolic fluxes. It works by feeding cells with isotopic labeled nutrients, measuring metabolite isotopic labeling, and computationally interpreting the measured labeling data to estimate flux. Tandem mass-spectrometry (MS/MS) has been shown to be useful for MFA, providing positional isotopic labeling data. Specifically, MS/MS enables the measurement of a metabolite tandem mass-isotopomer distribution, representing the abundance in which certain parent and product fragments of a metabolite have different number of labeled atoms. However, a major limitation in using MFA with MS/MS data is the lack of a computationally efficient method for simulating such isotopic labeling data. Here, we describe the tandem approach for efficiently computing metabolite tandem mass-isotopomer distributions in a metabolic network, given an estimation of metabolic fluxes. This approach can be used by MFA to find optimal metabolic fluxes, whose induced metabolite labeling patterns match tandem mass-isotopomer distributions measured by MS/MS. The tandem approach is applied to simulate MS/MS data in a small-scale metabolic network model of mammalian methionine metabolism and in a large-scale metabolic network model of *E. coli*. It is shown to significantly improve the running time by between two to three orders of magnitude compared to the state-of-the-art, cumomers approach. We expect the tandem approach to promote broader usage of MS/MS technology in metabolic flux analysis. Implementation is freely available at www.cs.technion.ac.il/~tomersh/methods.html



OPEN ACCESS

Citation: Tepper N, Shlomi T (2015) Efficient Modeling of MS/MS Data for Metabolic Flux Analysis. PLoS ONE 10(7): e0130213. doi:10.1371/journal.pone.0130213

Editor: Petras Dzeja, Mayo Clinic, UNITED STATES

Received: February 27, 2015

Accepted: May 16, 2015

Published: July 31, 2015

Copyright: © 2015 Tepper, Shlomi. This is an open access article distributed under the terms of the [Creative Commons Attribution License](http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Metabolic flux analysis (MFA) is a method for quantifying in vivo metabolic fluxes that is commonly used to address problems in biotechnology and medicine [1–6]. It involves feeding cells with isotopic labeled nutrients (e.g. ¹³C labelled substrates), measuring metabolite isotopic labeling, and applying computational methods to estimate fluxes [1, 7–9].

MFA is based on the key observation that metabolite isotopic labeling patterns are uniquely determined by the distribution of metabolic flux in the network [10]. It is typically implemented as a non-convex optimization problem, searching for the most likely distribution of fluxes that would give rise to metabolite isotopic labeling that optimally matches experimental

Table 1. The distribution of isotopomers of metabolite A (shown in Fig 2) within tandemers of A, defined with respect to $A_{(2,3)}^{(2,3,4)}$.

Isotopomers	Tandemers
0000	$[M + 0] > [m + 0]$
0001	$[M + 1] > [m + 0]$
0010	$[M + 1] > [m + 1]$
0011	$[M + 2] > [m + 1]$
0100	$[M + 1] > [m + 1]$
0101	$[M + 2] > [m + 1]$
0110	$[M + 2] > [m + 2]$
0111	$[M + 3] > [m + 2]$
1000	$[M + 0] > [m + 0]$
1001	$[M + 1] > [m + 0]$
1010	$[M + 1] > [m + 1]$
1011	$[M + 2] > [m + 1]$
1100	$[M + 1] > [m + 1]$
1101	$[M + 2] > [m + 1]$
1110	$[M + 2] > [m + 2]$
1111	$[M + 3] > [m + 2]$

Isotopomers are represented by sequences of zeroes and ones, denoting non-labeled and labeled atoms, respectively.

doi:10.1371/journal.pone.0130213.t001

measurements. The running time of MFA methods becomes a major bottleneck when analyzing large-scale metabolic networks, consisting of hundreds of reactions, when repeatedly applied to compute accurate flux confidence intervals [11, 12], and in experimental design of isotopic labeling experiments [13]. The major factor that affects the performance of MFA implementations is the time required to simulate metabolite isotopic labeling for a candidate flux distribution.

A distinct labeling pattern of a certain metabolite is called an *isotopomer*, while the distribution of abundances of all isotopomers is referred to as, *isotopomer distribution*. A metabolite with n carbons has 2^n distinct isotopomers; for example, as shown in Table 1, a metabolite having 4 carbons has 16 possible isotopomers. Previous studies have suggested the cumomers [14] and fluxomers [15] approaches for efficiently simulating the isotopomer distributions of all metabolites in a metabolic network given a flux vector. However, as the number of distinct isotopomers of a metabolite is exponentially dependent on the number of carbons that it has, these methods require a huge number of variables and may become computationally intractable.

Measuring the complete isotopomer distribution of metabolites is technically infeasible. Instead, mass-spectrometry is typically used to measure the relative abundance of a given metabolite having different number of labeled atoms (i.e. zero labeled atoms, one, two, etc). A set of isotopomers of a certain metabolite having the same mass is referred to as *mass-isotopomers*, while the relative abundance of mass-isotopomers denoted *mass-isotopomer distribution*. Notably, a mass-isotopomer distribution provides limited information on positional isotopic labeling, as isotopomers with the same number of labeled atoms have the same mass regardless of their position. Mass-isotopomer distributions can be calculated given the complete isotopomer distributions (by summing the abundances of isotopomers having the same number of labeled atoms). Alternatively, they can be directly and efficiently computed via the EMU approach [1].

Information on the positional labeling of metabolites can be obtained by tandem mass-spectrometry (i.e. MS/MS) and was previously shown to significantly improve quantification of metabolic fluxes via MFA [12, 16–18]. It works by isolating a single parent ion from the full spectrum and measuring its mass, followed by a collision that yields product ions whose mass is also measured. It can hence be employed to derive the mass-isotopomer distribution of a metabolite of interest and that of a collisional fragment. Most importantly, MS/MS can further measure the abundance of specific transitions from certain parent to product mass-isotopomers, referred to as *tandem mass-isotopomers* (also denoted here as *tandemers*, for short). We denote the number of labeled atoms of a parent ion by $M+0$ (having no labeled atoms), $M+1$ (having one labeled atom), etc., and the number of labeled atoms of a product ion by $m+0$, $m+1$, etc. We denote a transition from parent mass-isotopomer $M+i$ to product mass-isotopomer $m+j$, by $[M+i] > [m+j]$. This provides additional information on positional labeling beyond that available via the mass-isotopomer distribution of the parent and product fragments separately (describing the abundance of a metabolite having various combinations of specific mass-isotopomers for the parent and product molecules). The relative abundance of all tandemers is referred to as a *tandem mass-isotopomer distribution* (or *tandemer distribution*, for short).

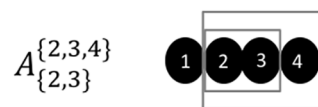
Currently, there is no method for efficiently simulating tandemer distributions. Previous applications of MFA given MS/MS data have inefficiently computed the complete isotopomer distributions for all metabolites in the network (for example, via cumomers [18]) in order to simulate tandemer distributions. Here, we present the tandemer method for efficiently simulating MS/MS measurements (i.e. tandemer distributions) of metabolites in a metabolic network. It builds upon and extends ideas set forward by the EMU method which efficiently simulates mass-isotopomer measurements [1].

Theory

A formal definition of tandemers

We denote a *metabolite fragment pair (MFP)* of a metabolite A with parent fragment N and product fragment K by $A_{K,N}^N$, with $N \subseteq \{1..n\}$ (where n is the total number of atoms), and $K \subseteq N$. For example, Fig 1a shows metabolite A having 4 carbons and an associated MFP $A_{\{2,3\}}^{\{2,3,4\}}$, where $N = \{2,3,4\}$ and $K = \{2,3\}$. A tandemer of metabolite A , $[M+i] > [m+j]$ with respect to a

a



b

$$[A]_{\{2,3\}}^{\{2,3,4\}} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \end{matrix} & \begin{bmatrix} [A]_{\{2,3\}}^{\{2,3,4\}}(0,0) & 0 & 0 \\ [A]_{\{2,3\}}^{\{2,3,4\}}(1,0) & [A]_{\{2,3\}}^{\{2,3,4\}}(1,1) & 0 \\ 0 & [A]_{\{2,3\}}^{\{2,3,4\}}(2,1) & [A]_{\{2,3\}}^{\{2,3,4\}}(2,2) \\ 0 & 0 & [A]_{\{2,3\}}^{\{2,3,4\}}(3,2) \end{bmatrix} \end{matrix}$$

Fig 1. (a) Metabolite A and its MFP $A_{\{2,3\}}^{\{2,3,4\}}$ with parent fragment $N = \{2,3,4\}$ and product fragment $K = \{2,3\}$. (b) The tandemer distribution matrix $[A]_{\{2,3\}}^{\{2,3,4\}}$. The abundance of infeasible tandemers in $[A]_{\{2,3\}}^{\{2,3,4\}}$ is, by definition, zero.

doi:10.1371/journal.pone.0130213.g001

MFP A_K^N is defined as a set of isotopomers of A having $0 \leq i \leq |N|$ labeled atoms within the parent fragment N, and $0 \leq j \leq |K|$ labeled atoms within the product fragment K. A tandem is considered feasible if it does not represent an empty set of isotopomers, i.e. when j is no larger than i (as the product fragment K is enclosed within the parent fragment N), and no smaller than $i - (|N| - |K|)$ (when all atoms that are in the parent but not in the product fragment are labeled). The number of feasible tandemers for A_K^N is hence $(|N| - |K| + 1)(|K| + 1)$ [18].

The entire tandem distribution of A, with respect to the MFP A_K^N , can be represented by a *tandemer distribution matrix* $[A]_{\{2,3\}}^{\{2,3,4\}}$ with $|N| + 1$ rows (representing the number of labeled atoms in the parent fragment; from zero to $|N|$), and $|K| + 1$ columns (representing the number of labeled atoms in the product fragment), such that $[A]_{\{2,3\}}^{\{2,3,4\}}(i,j)$ is the relative abundance of tandem $[M + i] > [m + j]$. As entries in $[A]_{\{2,3\}}^{\{2,3,4\}}$ represent distinct events whose sum is 1, the matrix represents a probability distribution. Notably, the abundance of infeasible tandemers is by definition zero. Fig 1b shows the tandem distribution matrix $[A]_{\{2,3\}}^{\{2,3,4\}}$, while Table 1 shows the corresponding feasible tandemers. For example, given a metabolite A having 4 carbons, where the abundance of the isotopomer 1001 is 0.6, that of 0101 is 0.3 and 1011 is 0.1 (and the abundance of all other isotopomers of A is zero), the the tandem distribution matrix of MFP $A_{\{2,3\}}^{\{2,3,4\}}$ is:

$$[A]_{\{2,3\}}^{\{2,3,4\}} = \begin{matrix} & & 0 & 1 & 2 \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \end{matrix} & \begin{bmatrix} 0 & 0 & 0 \\ 0.6 & 0 & 0 \\ 0 & 0.4 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{matrix}$$

Calculating metabolite tandem distributions

Under isotopic steady-state, for metabolite B that is produced solely through one biochemical reaction with a single substrate, A, the tandem distribution matrix $[B]_{N'}^M$ is equal to $[A]_{N'}^{M'}$, where atoms in M' and N' are mapped to atoms in M and N, respectively (Fig 2a). We refer to $A_{N'}^{M'}$ as the *substrate MFP* of $B_{N'}^M$ via reaction i, and denote it by $S(B_{N'}^M, i)$.

For metabolite C that is produced solely through one reaction with two substrates, A and B, $[C]_N^M$ can be calculated based on two tandem distribution matrices: $[A]_{N_1}^{M_1}$ (where atoms in M_1 and N_1 are mapped to atoms in M and N, respectively) and $[B]_{N_2}^{M_2}$ (where atoms of M_2 and N_2 are similarly mapped to M and N; Fig 2b). $A_{N_1}^{M_1}$ and $B_{N_2}^{M_2}$ are further referred to as the substrate MFPs of C_N^M via reaction i, and denote by $S_1(C_N^M, i)$ and $S_2(C_N^M, i)$ respectively. Specifically, $[C]_N^M(i, j)$ can be calculated as following:

$$C(i, j) = \sum_{\substack{0 \leq i_1 < m_1 \\ 0 \leq i_2 < m_2 \\ i_1 + i_2 = i}} \sum_{\substack{0 \leq j_1 < n_1 \\ 0 \leq j_2 < n_2 \\ j_1 + j_2 = j}} A(i_1, j_1) \cdot B(i_2, j_2) \tag{1}$$

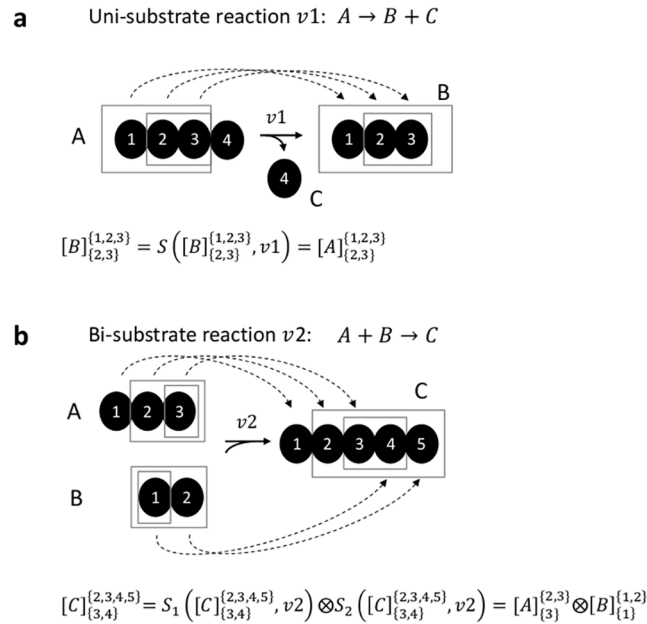


Fig 2. Calculating the tandemer distribution matrix of a product MFP based on tandemer distribution matrices of substrate MFPs in uni-substrate (a) and bi-substrate (b) reactions.

doi:10.1371/journal.pone.0130213.g002

We refer to $[C]_N^M$ as being equal to the Cauchy product between matrices $A^{m_1 \times n_1}$ and $B^{m_2 \times n_2}$, denoted $A \otimes B$ (extending the definition of Cauchy product between vectors to matrices).

For example, let us consider the bi-substrate reaction shown in Fig 2b, in which A (having 3 carbons) is condensed with B (having 2 carbons) to make C, with carbons from A mapped to the first 3 carbons in C and atoms from B mapped to the last 2 carbons in C. In this case, the tandemer distribution matrix for $C_{\{3,4\}}^{\{2,3,4,5\}}$, can be calculated based on the Cauchy product of the tandemer distributions of the substrate MFP's $A_{\{3\}}^{\{2,3\}}$ and $B_{\{1\}}^{\{1,2\}}$.

Under isotopic steady-state, a tandemer distribution matrix for the MFP A_K^N is determined based on the corresponding tandemer distributions of all of its substrate MFPs according to the following balance equation:

$$[A]_K^N \sum_{i \in \left\{ \begin{array}{l} \text{reactions} \\ \text{producing A} \end{array} \right\}} v_i = \sum_{i \in \left\{ \begin{array}{l} \text{Uni-substrate} \\ \text{reactions} \\ \text{producing A} \end{array} \right\}} [S(A_K^N, i)] \cdot v_i + \sum_{i \in \left\{ \begin{array}{l} \text{Bi-substrate} \\ \text{reactions} \\ \text{producing A} \end{array} \right\}} [S_1(A_K^N, i)] \otimes [S_2(A_K^N, i)] \cdot v_i \quad (2)$$

where v_i is the flux through reaction i .

An algorithm for simulating tandemer distributions

Our goal is to efficiently simulate the tandemer distribution for a pre-defined set of metabolites (for which corresponding experimental data might be available). We assume that a metabolic network model with reaction atom-mappings (describing the mapping of substrate to product metabolite atoms in each reaction) and candidate fluxes are given. To address this problem, we

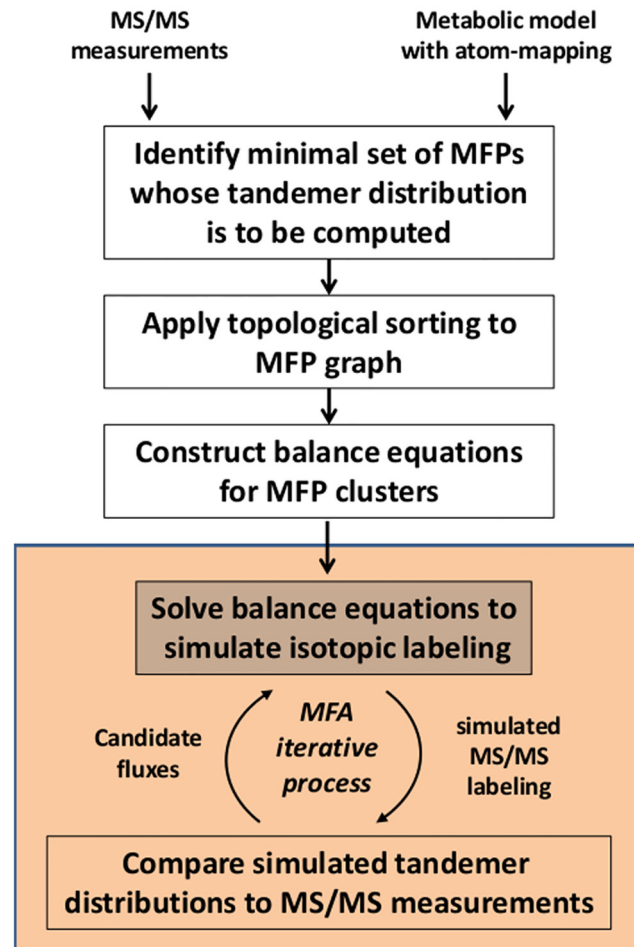


Fig 3. An outline of the tandemers approach. First, given MS/MS measurements and a metabolic model, a minimal set of MFPs is identified constructing an MFP graph. Second, MFPs are clustered and sorted and third, isotopic balance equations are formulated for each MFP cluster. Given a candidate flux vector, tandemer distributions are calculated by solving the set of isotopic balance equations.

doi:10.1371/journal.pone.0130213.g003

present the tandemers approach, whose outline is shown in Fig 3. A detailed explanation of the various steps of the algorithms is provided below, while a Matlab implementation is available at: www.cs.technion.ac.il/~tomersh/methods.html.

2.3.1. Identifying a minimal set of MFPs whose tandemer distributions are needed to simulate MS/MS measurements. The identification of a minimal set of MFP's in the metabolic network whose tandemer distributions would enable simulating the tandemer distribution of a given set of metabolites is done using a recursive procedure, in a similar manner to that presented in the EMU approach [1]. Specifically, starting with a list of MFPs for which tandemer distributions are available, we traverse the metabolic network to iteratively add substrate MFPs (via the definition of substrate MFPs provided above). This procedure ends when no more new MFPs can be added to the list.

The total number of MFPs depends on many factors, including the structure of the metabolic network, number of atoms per metabolite, reaction atom mappings, and number of metabolites for which experimental MS/MS data is available as input. In theory, for each

metabolite with n carbons, the worst case number of possible MFPs could go up to:

$$\sum_{i=1}^n \sum_{j=1}^i \binom{n}{i} \binom{i}{j} < 3^n \tag{3}$$

Hence, theoretically, the number of MFPs for a certain metabolite found by the recursive procedure described above may exceed the number of possible isotopomers for that metabolite (which is 2^n). In practice, as we show below, applying this method on various metabolic networks, the number of MFPs found per metabolite is markedly lower than the number of isotopomers, resulting in improved running time compared to existing methods.

2.3.2. MFP’s clustering and ordering. In a uni-substrate reaction producing MFP B_K^N , the size of the parent fragment in B_K^N equals that of its substrate $A_{K'}^N = S(B_K^N, i)$ (i.e. $|N'| = |N|$), while in a bi-substrate reaction, the size of the parent fragment is larger than that of both its substrate MFPs (noting that a bi-substrate reaction in which one of the substrate MFP has a parent fragment of size zero can be regarded as a uni-substrate reaction). Hence, a tandem distribution with parent fragment of a certain size is linearly dependent upon tandem distributions with parent fragment of the same size, and non-linearly dependent (via Cauchy product) only upon tandem distributions with smaller parent fragments (see Eq 2). Therefore, all tandem distributions can be computed by sequentially solving linear balance equations for sets of tandem distributions with increasing sizes (as Cauchy product of tandem distribution matrices of smaller parent fragments can be computed before reaching tandem distributions of larger parent fragments; similarly to the EMU [1] and cumomer [19] approaches).

Following a method proposed by [20], we divided the linear balance equations into smaller sets of equations that can be sequentially solved. Specifically, we construct a directed graph whose nodes represent the identified MFPs and edges connect an MFP A_K^N with its substrate $S(A_K^N, i)$ for uni-substrate reactions, and with both its substrates $S_1(A_K^N, i)$ and $S_2(A_K^N, i)$ for bi-substrate reactions. Notably, for a given MFP in this graph, its parent fragment size is not smaller than those of all MFPs having a directed edge towards it (see Fig 2). Hence, decomposing the graph into strongly connected components [21] and applying topological sorting on the identified clusters [22] leads to a cascade of MFP clusters, each consisting of MFPs having the same parent fragment size, with clusters ordered according to a non-decreasing fragment parent size (see Example below in Section 2.4). Iterating through the list of clusters, all tandem distributions in a given cluster can be calculated via a set of linear equations based on tandem distributions inferred in previous clusters (via Eq (2); where non-linear terms associated with bi-substrate reactions are calculated based on tandem distributions inferred in previous clusters)[22].

2.3.3. Formulating and solving a series of linear balance equations for MFPs in each cluster. Tandem distributions for MFPs in a cluster are linearly dependent on each other, given the corresponding tandem distributions for MFPs in previous clusters. Notably, the set of balance equations for tandem distribution matrices for MFPs in the i ’th cluster can be formulated as following:

$$A_i X_i = B_i Y_i \tag{4}$$

where X_i is a matrix whose rows represent the tandem distributions for MFPs in the i ’th cluster (i.e. each tandem distribution represented as a row vector; removing infeasible tandemers), Y_i is a matrix whose rows represent tandem distributions and Cauchy product of tandem distributions computed for previous clusters, and A_i and B_i consist of corresponding fluxes.

Solving a set of balance equations for MFPs in the i 'th cluster requires calculating the inverse of A_i , whose number of rows (and columns) is equal to the number of MFPs in the cluster. As the cumomers approach also involves grouping cumomers in clusters and inverting flux matrices (similar to A_i , here) whose size depends on the number of cumomers in each cluster, the running time of cumomers and tandemers approaches can be compared in terms of cumomers and MFPs cluster sizes, and the time needed to invert the corresponding flux matrices. Theoretically, an n fold reduction in the number of MFPs versus cumomers in a cluster should result in n^3 improvement in running time. Notably, the number of non-feasible tandemers in each tandem distribution matrix (represented by the number of column of X_i has a negligible effect on the time require to solve Eq (4), which is dominated by the n^3 time required to calculate the inverse of A_i).

An example of the tandemers approach on a toy metabolic network

In this section we describe the application of the tandemers approach on a toy metabolic network shown in Fig 4a, where atom mappings are given in Fig 4b. We assume that metabolic fluxes as well as the labeling pattern of A are known (considering that isotopically labeled A is given the growth media) and aim to compute the tandem distribution of a $E_{\{2,3\}}^{\{1,2,3,4\}}$ (assumed to be measured via MS/MS). Applying the recursive procedure described in the previous section, starting from $E_{\{2,3\}}^{\{1,2,3,4\}}$, we identify a total of 10 MFPs (for metabolites other than A) whose tandem distribution is needed to compute that of $E_{\{2,3\}}^{\{1,2,3,4\}}$. The MFP graph and its three connected components are shown in Fig 4c. Notably, cluster (III) depends on clusters (I) and (II), while the latter clusters are mutually independent.

For example, the isotopic balance equation for tandem distribution matrices in cluster (I) is formulated as following, according to Eq (4):

$$\begin{aligned}
 & \begin{bmatrix} -(v_1 + v_5^b) & 0 & 0 & v_5^b \\ v_3 & -(v_2 + v_3 + v_4^b) & v_4^b & 0 \\ 0 & v_4^f & -(v_2 + v_4^f) & 0 \\ v_5^f & v_6 & 0 & -(v_5^f + v_6) \end{bmatrix} \\
 & \cdot \begin{bmatrix} [B]_{\{4\}}^{\{3,4\}}(0,0) & [B]_{\{4\}}^{\{3,4\}}(1,0) & [B]_{\{4\}}^{\{3,4\}}(1,1) & [B]_{\{4\}}^{\{3,4\}}(2,1) \\ [C]_{\{2\}}^{\{1,2\}}(0,0) & [C]_{\{2\}}^{\{1,2\}}(1,0) & [C]_{\{2\}}^{\{1,2\}}(1,1) & [C]_{\{2\}}^{\{1,2\}}(2,1) \\ [D]_{\{2\}}^{\{1,2\}}(0,0) & [D]_{\{2\}}^{\{1,2\}}(1,0) & [D]_{\{2\}}^{\{1,2\}}(1,1) & [D]_{\{2\}}^{\{1,2\}}(2,1) \\ [E]_{\{4\}}^{\{3,4\}}(0,0) & [E]_{\{4\}}^{\{3,4\}}(1,0) & [E]_{\{4\}}^{\{3,4\}}(1,1) & [E]_{\{4\}}^{\{3,4\}}(2,1) \end{bmatrix} \\
 & = \begin{bmatrix} -v_1 & 0 & 0 \\ 0 & -v_2 & 0 \\ 0 & 0 & -v_2 \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} [A]_{\{4\}}^{\{3,4\}}(0,0) & [A]_{\{4\}}^{\{3,4\}}(1,0) & [A]_{\{4\}}^{\{3,4\}}(1,1) & [A]_{\{4\}}^{\{3,4\}}(2,1) \\ [A]_{\{2\}}^{\{1,2\}}(0,0) & [A]_{\{2\}}^{\{1,2\}}(1,0) & [A]_{\{2\}}^{\{1,2\}}(1,1) & [A]_{\{2\}}^{\{1,2\}}(2,1) \\ [A]_{\{3\}}^{\{3,4\}}(0,0) & [A]_{\{3\}}^{\{3,4\}}(1,0) & [A]_{\{3\}}^{\{3,4\}}(1,1) & [A]_{\{3\}}^{\{3,4\}}(2,1) \\ 0 & 0 & 0 & 0 \end{bmatrix}
 \end{aligned}$$

Considering that the tandem distribution $[B]_{\{4\}}^{\{3,4\}}$ is calculated as part of cluster (I) and $[C]_{\{1\}}^{\{1,2\}}$ as part of cluster (II), enables to calculate the Cauchy product between the two matrices prior

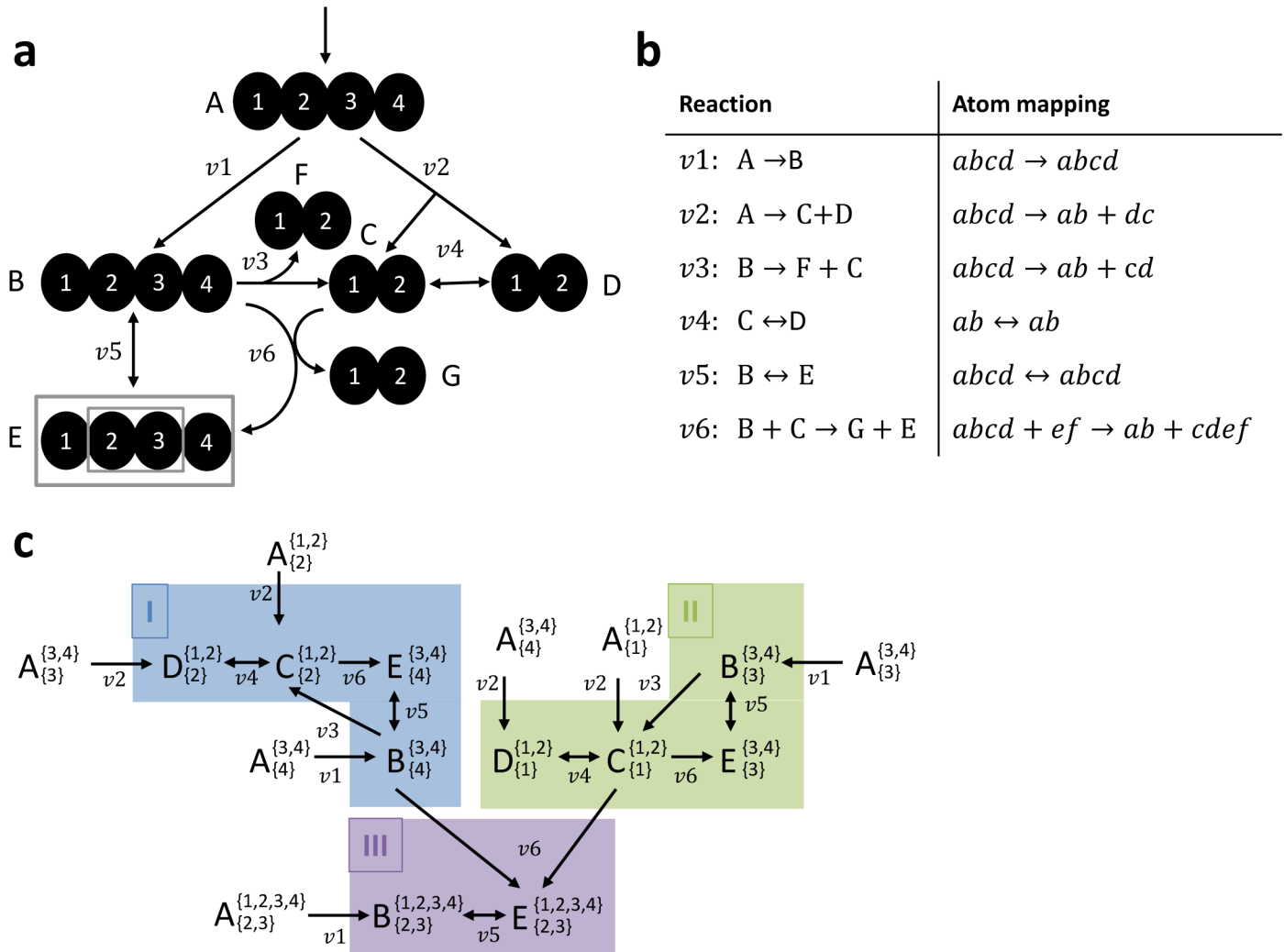


Fig 4. (a) A toy metabolic network, where the labeling pattern of A that is supplied in the media is assumed to be known, and the tandemer distribution of $[E]_{\{2,3\}}^{\{1,2,3,4\}}$ is to be calculated; (b) Atom mapping for network reactions; (c) An MFP graph and three strongly connected components whose numbering is determined via topological sorting.

doi:10.1371/journal.pone.0130213.g004

to solving the balance equations for cluster (III). Given the Cauchy product between $[B]_{\{4\}}^{\{3,4\}}$ and $[C]_{\{1\}}^{\{1,2\}}$, the isotopic balance equations for cluster (III) are linear:

$$[B]_{\{2,3\}}^{\{1,2,3,4\}} \cdot (v_1 + v_5^b) = [A]_{\{2,3\}}^{\{1,2,3,4\}} \cdot v_1 + [E]_{\{2,3\}}^{\{1,2,3,4\}} \cdot v_5^b$$

$$[E]_{\{2,3\}}^{\{1,2,3,4\}} \cdot (v_5^f + v_6) = [B]_{\{2,3\}}^{\{1,2,3,4\}} \cdot v_5^f + [B]_{\{4\}}^{\{3,4\}} \otimes [C]_{\{1\}}^{\{1,2\}} \cdot v_6$$

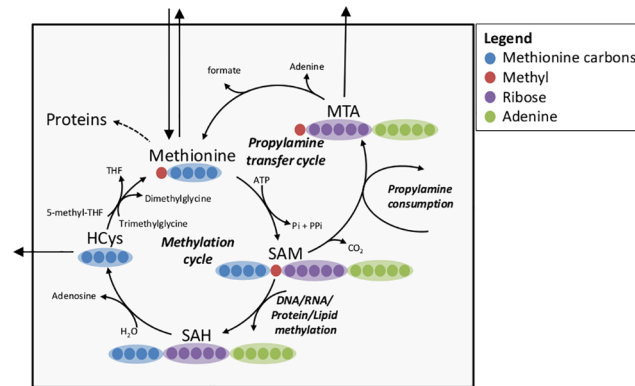


Fig 5. Methionine metabolism, including transmethylation cycle, polyamine biosynthesis and methionine salvage cycle. Metabolites abbreviations: SAM: S-Adenosylmethionine; SAH: S-Adenosylhomocysteine; HCys: L-Homocysteine; MTA: Methylthioadenosine.

doi:10.1371/journal.pone.0130213.g005

Results

Applying the tandemers method on a small-scale model of mammalian methionine metabolism

To demonstrate the applicability of the tandemers method for efficiently computing experimental MS/MS data in ^{13}C labeling experiments, we applied it on a simplified metabolic network model of mammalian cellular metabolism of methionine (Fig 5, S1 File). Methionine metabolism involves two partially overlapping cyclic pathways for transmethylation (of protein and DNA) and propylamine transfer (for polyamine biosynthesis). The metabolic donor of both the methyl and propylamine groups is S-adenosylmethionine (SAM), which has 15 carbons (hence, a high-carbon metabolite). The product metabolites S-adenosylhomocysteine (SAH) and methylthioadenosine (MTA) also have high number of carbons, 14 and 11 carbons respectively (as one additional SAM carbon is oxidized to CO_2 prior to the propylamine transfer). Considering that the number of isotopomers of a metabolite with n carbons is 2^n , explicitly modeling the entire isotopomer distribution of all five metabolites in this network (as done in the cumomers method) would require 52,306 variables.

To apply the tandemers method, we utilized experimentally determined fluxes in this network as input [23]. We assume that the carbon labeling pattern of metabolites that are outside the scope of the model, including that of media methionine, ATP, and 5-methyl-tetrahydrofolate (i.e. the labeling of the methyl group in 5-methyl-THF) are known. First, we applied the tandemers method to compute the tandemer distribution of all 5 metabolites in the network, assuming that parent fragments are the intact metabolites, and that product fragments are the adenine group in SAM, SAH, and MTA, and the four methionine carbons other than the methyl group for methionine and HCys. The resulting number of MFPs whose tandemer distributions are found to be required to compute the tandemer distributions of the given MFPs is 35. These are divided into clusters, such that the largest cluster has only 4 MFPs. In comparison, applying the cumomers method given the same input data would require a total of 52,306 cumomers (the same as the number of isotopomers), divided into clusters whose largest one has 10,197 cumomers.

Next, we ran both the tandemers and cumomers methods multiple times, choosing a different subset of metabolites to calculate their tandemer distribution in each run. The parent

Table 2. Comparison of the performance of the cumomers and tandemers methods in calculating tandem distributions on mammalian methionine and *E. coli* networks.

	Mammalian methionine metabolism model			<i>E. coli</i> model		
	Variable count	Maximal cluster size	Running time	Variable count	Maximal cluster size	Running time
Cumomers	52,306	10,197	0.68	19,404	4,016	3.3
Tandemers	33	4	0.00046	695	32	0.01

doi:10.1371/journal.pone.0130213.t002

fragment was assumed to be the intact metabolite and the product fragment was chosen to be the ribose, adenine, propylamine, or the four methionine carbons other than the methyl group. We find an average number of only 33 MFPs per run of the tandemers method, while the cumomers method requiring of 52,306 cumomer variables regardless of assumed input. The average running time of the tandemers method was found to be 0.00046 seconds, while that of the cumomers method was 0.68 seconds, ~1500-fold higher (Table 2). Notably, the significant reduction in running time will be especially important for designing optimal isotope tracing experiments, requiring numerous (many thousands for large-scale networks) repeated simulations of isotope labeling patterns for possible flux distributions [13].

Applying the tandemers approach in a large-scale metabolic network model of *E. coli*

To further demonstrate the applicability of the tandemers approach, we applied it on a large-scale metabolic network model of *E. coli* [13]. This isotopomer model accounts for glycolysis, TCA cycle, pentose phosphate pathway, oxidative phosphorylation, pyruvate metabolism, anaerobic reactions and other central and biosynthetic pathways, with a total of 206 metabolites and 405 reactions. Notably, metabolites with a high number of carbons were not included in this network reconstruction to facilitate the application of the cumomers approach (by lumping surrounding reactions together) [13]. The total number of isotopomers in this large-scale network is hence surprisingly low, reaching 19,404 isotopomers (much lower than that in the small-scale methionine network applied above). Hence, even though this network model is substantially larger than the methionine network, we do not expect the tandemers approach to demonstrate the same level of improvement compared to the cumomers method.

We applied the tandemers method 1000 times to compute the tandem distribution of randomly chosen sets of metabolites (having between 1 to 20 metabolites). The average number of resulting MFPs was 695, with a maximal MFP cluster size of 32. In comparison, applying the cumomers approach resulted in 19,404 cumomers, and a maximal cluster size of 4,016. The average running time of the tandemers and cumomers methods are 0.01 and 3.3 seconds, respectively, representing a ~300-fold improvement by the tandemers method (Table 2).

For the tandemers methods, the number of variables represents the number of MFP whose tandem distribution is calculated; for the cumomers approach, it represents the number of cumomers. The number of these variables corresponds to the size of the flux matrices whose inverse is calculated by each method, and is hence proportional to overall running time (see Section 2.3.3). We report the average variable count, clustersize, and running time for the cumomers and tandemers methods in multiple simulations given different sets of metabolites and collisional fragments, as described above. Notably, considering that MFA applications and especially experimental design of isotope tracing experiments require thousands of repeated simulations of metabolite isotopic labeling, the ~1500-fold and 300-fold improvement in

running time observed in the mammalian methionine network and on the *E. coli* network, respectively, is of a major practical importance.

Discussion

Tandem MS holds great promise for metabolic flux analysis as it provides information on metabolite positional labeling [12, 16–18]. However, a major limitation in using MFA with MS/MS data is the lack of a computationally efficient method for simulating isotopic labeling data measurable via MS/MS. State-of-the-art methods such as cumomers that enable to simulate MS/MS data requires simulating the abundance of all distinct isotopomers, whose number is exponentially dependent on the number of atoms in each metabolite. Here, we described the tandemers approach that is specifically designed for efficiently computing tandem mass-isotopomer distributions measurable via MS/MS, demonstrating a roughly two to three orders of magnitude improvement in running time compared to the cumomers approach. The tandemers approach is especially useful when analyzing metabolic networks with metabolites having a high number of carbons, where modeling the entire isotopomer distribution may become computationally intractable.

In our application of the tandemers method on a metabolic network of mammalian methionine metabolism and for *E. coli*, we computed tandemer distributions for MFPs in which the parent fragment was the entire metabolite. This represents the case where no in-source fragmentation occurs during MS ionization, which is typically the situation with LC-MS. However, in-source fragmentation can occur (mostly with GC-MS), leading to measurement of tandemer distributions with parent fragment that is not the entire metabolite. Such in-source fragmentation can provide further useful information on positional labeling, for example, as was recently used to infer all distinct isotopomers of aspartate [24]. Obviously, the tandemers approach may also be used to compute tandemer distributions for MFPs with parent fragments that are not the entire metabolite.

In a recent study, we described a method, Metabolic Flux Analysis/Unknown Fragments (MFA/UF), capable of using MS/MS data to improve flux inference even when the positional origin of fragments is unknown [12]. MFA/UF extends upon standard MFA and jointly searches for the most likely metabolic fluxes together with the most plausible position of collisional fragments that would optimally match measured MS/MS data. To simulate MS/MS data given candidate fluxes and candidate collisional fragments, MFA/UF utilized the cumomers approach to simulate MS/MS labeling data. Considering that the tandemers approach was shown here to outperform the cumomers method, integrating it within MFA/UF is expected to lead to a significant improvement in running time.

Considering that a major current complication in utilizing MS/MS data in metabolic flux analysis involves the lack of computationally efficient methods for simulating such experimental measurements, we expect the tandemers approach to promote broader usage of this technology.

Supporting Information

S1 File. Methionine metabolism model.
(XLSX)

Author Contributions

Conceived and designed the experiments: NT TS. Performed the experiments: NT TS. Analyzed the data: NT TS. Wrote the paper: NT TS.

References

1. Antoniewicz MR, Kelleher JK, Stephanopoulos G. Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. *Metab Eng*. 2007; 9(1):68–86. Epub 2006/11/08. doi: [10.1016/j.ymben.2006.09.001](https://doi.org/10.1016/j.ymben.2006.09.001) PMID: [17088092](https://pubmed.ncbi.nlm.nih.gov/17088092/); PubMed Central PMCID: PMC1994654.
2. Boghigian BA, Seth G, Kiss R, Pfeifer BA. Metabolic flux analysis and pharmaceutical production. *Metab Eng*. 2010; 12(2):81–95. Epub 2009/10/29. doi: [10.1016/j.ymben.2009.10.004](https://doi.org/10.1016/j.ymben.2009.10.004) PMID: [19861167](https://pubmed.ncbi.nlm.nih.gov/19861167/).
3. Jin ES, Jones JG, Merritt M, Burgess SC, Malloy CR, Sherry AD. Glucose production, gluconeogenesis, and hepatic tricarboxylic acid cycle fluxes measured by nuclear magnetic resonance analysis of a single glucose derivative. *Analytical biochemistry*. 2004; 327(2):149–55. Epub 2004/03/31. doi: [10.1016/j.ab.2003.12.036](https://doi.org/10.1016/j.ab.2003.12.036) PMID: [15051530](https://pubmed.ncbi.nlm.nih.gov/15051530/).
4. Sauer U. Metabolic networks in motion: 13C-based flux analysis. *Mol Syst Biol*. 2006; 2:62. Epub 2006/11/15. doi: [10.1038/msb4100109](https://doi.org/10.1038/msb4100109) PMID: [17102807](https://pubmed.ncbi.nlm.nih.gov/17102807/); PubMed Central PMCID: PMC1682028.
5. Sillers R, Al-Hinai MA, Papoutsakis ET. Aldehyde-alcohol dehydrogenase and/or thiolase overexpression coupled with CoA transferase downregulation lead to higher alcohol titers and selectivity in *Clostridium acetobutylicum* fermentations. *Biotechnol Bioeng*. 2009; 102(1):38–49. Epub 2008/08/30. doi: [10.1002/bit.22058](https://doi.org/10.1002/bit.22058) PMID: [18726959](https://pubmed.ncbi.nlm.nih.gov/18726959/).
6. Wiechert W. An introduction to 13C metabolic flux analysis. *Genetic engineering*. 2002; 24:215–38. Epub 2002/11/06. PMID: [12416307](https://pubmed.ncbi.nlm.nih.gov/12416307/).
7. Wiechert W, Mollney M, Isermann N, Wurzel M, de Graaf AA. Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems. *Biotechnol Bioeng*. 1999; 66(2):69–85. Epub 1999/11/24. PMID: [10567066](https://pubmed.ncbi.nlm.nih.gov/10567066/).
8. Möllney M, Wiechert W, Kownatzki D, de Graaf AA. Bidirectional reaction steps in metabolic networks: IV. Optimal design of isotopomer labeling experiments. *Biotechnology and Bioengineering*. 1999; 66(2):86–103. PMID: [10567067](https://pubmed.ncbi.nlm.nih.gov/10567067/).
9. Wiechert W, Siefke C, de Graaf AA, Marx A. Bidirectional reaction steps in metabolic networks: II. Flux estimation and statistical analysis. *Biotechnol Bioeng*. 1997; 55(1):118–35. Epub 1997/07/05. doi: [10.1002/\(SICI\)1097-0290\(19970705\)55:1<118::AID-BIT13>3.0.CO;2-I](https://doi.org/10.1002/(SICI)1097-0290(19970705)55:1<118::AID-BIT13>3.0.CO;2-I) PMID: [18636450](https://pubmed.ncbi.nlm.nih.gov/18636450/).
10. Rantanen A, Rousu J, Jouhten P, Zamboni N, Maaheimo H, Ukkonen E. An analytic and systematic framework for estimating metabolic flux ratios from 13C tracer experiments. *BMC Bioinformatics*. 2008; 9(1):266. doi: [10.1186/1471-2105-9-266](https://doi.org/10.1186/1471-2105-9-266)
11. Antoniewicz MR, Kelleher JK, Stephanopoulos G. Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements. *Metab Eng*. 2006; 8(4):324–37. Epub 2006/04/25. doi: [10.1016/j.ymben.2006.01.004](https://doi.org/10.1016/j.ymben.2006.01.004) PMID: [16631402](https://pubmed.ncbi.nlm.nih.gov/16631402/).
12. Tepper N, Shlomi T. An integrated computational approach for metabolic flux analysis coupled with inference of tandem-MS collisional fragments. *Bioinformatics*. 2013; 29(23):3045–52. Epub 2013/10/15. doi: [10.1093/bioinformatics/btt516](https://doi.org/10.1093/bioinformatics/btt516) PMID: [24123514](https://pubmed.ncbi.nlm.nih.gov/24123514/).
13. Schellenberger J, Zielinski DC, Choi W, Madireddi S, Portnoy V, Scott DA, et al. Predicting outcomes of steady-state 13C isotope tracing experiments using Monte Carlo sampling. *BMC Systems Biology*. 2012; 6(9).
14. Wiechert W, Mollney M, Petersen S, de Graaf AA. A universal framework for 13C metabolic flux analysis. *Metab Eng*. 2001; 3(3):265–83. Epub 2001/07/20. doi: [10.1006/mben.2001.0188](https://doi.org/10.1006/mben.2001.0188) PMID: [11461148](https://pubmed.ncbi.nlm.nih.gov/11461148/).
15. Srour O, Young JD, Eldar YC. Fluxomers: a new approach for 13C metabolic flux analysis. *BMC Syst Biol*. 2011; 5:129. Epub 2011/08/19. doi: [10.1186/1752-0509-5-129](https://doi.org/10.1186/1752-0509-5-129) PMID: [21846358](https://pubmed.ncbi.nlm.nih.gov/21846358/).
16. Jeffrey FM, Roach JS, Storey CJ, Sherry AD, Malloy CR. 13C isotopomer analysis of glutamate by tandem mass spectrometry. *Analytical biochemistry*. 2002; 300(2):192–205. Epub 2002/01/10. doi: [10.1006/abio.2001.5457](https://doi.org/10.1006/abio.2001.5457) PMID: [11779111](https://pubmed.ncbi.nlm.nih.gov/11779111/).
17. Rühl M, Rupp B, Noh K, Wiechert W, Sauer U, Zamboni N. Collisional fragmentation of central carbon metabolites in LC-MS/MS increases precision of 13C metabolic flux analysis. *Biotechnol Bioeng*. 2012; 109(3):763–71. Epub 2011/10/21. doi: [10.1002/bit.24344](https://doi.org/10.1002/bit.24344) PMID: [22012626](https://pubmed.ncbi.nlm.nih.gov/22012626/).
18. Choi J, Antoniewicz MR. Tandem mass spectrometry: a novel approach for metabolic flux analysis. *Metab Eng*. 2011; 13(2):225–33. Epub 2010/12/08. doi: [10.1016/j.ymben.2010.11.006](https://doi.org/10.1016/j.ymben.2010.11.006) PMID: [21134484](https://pubmed.ncbi.nlm.nih.gov/21134484/).
19. Wiechert W. 13C metabolic flux analysis. *Metab Eng*. 2001; 3(3):195–206. Epub 2001/07/20. doi: [10.1006/mben.2001.0187](https://doi.org/10.1006/mben.2001.0187) PMID: [11461141](https://pubmed.ncbi.nlm.nih.gov/11461141/).
20. Young JD, Walther JL, Antoniewicz MR, Yoo H, Stephanopoulos G. An elementary metabolite unit (EMU) based method of isotopically nonstationary flux analysis. *Biotechnology and bioengineering*. 2008; 99(3):686–99. Epub 2007/09/06. doi: [10.1002/bit.21632](https://doi.org/10.1002/bit.21632) PMID: [17787013](https://pubmed.ncbi.nlm.nih.gov/17787013/).

21. Weitzel M, Wiechert W, Nöh K. The topology of metabolic isotope labeling networks. *BMC Bioinformatics*. 2007; 8(315).
22. Kahn AB. Topological sorting of large networks. *Communications of the ACM*. 1962; 5(11):558–62.
23. Shlomi T, Fan J, Tang B, Kruger WD, Rabinowitz JD. An analytical approach for quantifying methionine metabolism in cancer cells. *Anal Chem*. 2014.
24. Choi J, Grossbach MT, Antoniewicz MR. Measuring complete isotopomer distribution of aspartate using gas chromatography/tandem mass spectrometry. *Anal Chem*. 2012; 84(10):4628–32. Epub 2012/04/19. doi: [10.1021/ac300611n](https://doi.org/10.1021/ac300611n) PMID: [22510303](https://pubmed.ncbi.nlm.nih.gov/22510303/).