BIOINFORMATICS ORIGINAL PAPER

Vol. 29 no. 23 2013, pages 3045–3052
doi:10.1093/bioinformatics/btt516

Systems biology

Advance Access publication October 10, 2013

An integrated computational approach for metabolic flux analysis coupled with inference of tandem-MS collisional fragments

Naama Tepper[1] and Tomer Shlomi[1,2,*]

[1]Department of Computer Science, Technion–Israel Institute of Technology, Haifa 32000, Israel and
[2]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA

Associate Editor: John Hancock

ABSTRACT

Motivation: Metabolic flux analysis (MFA) is a commonly used approach for quantifying metabolic fluxes based on tracking isotope labeling of metabolite within cells. Tandem mass-spectrometry (MS/MS) has been recently shown to be especially useful for MFA by providing rich information on metabolite positional labeling, measuring isotopic labeling patterns of collisional fragments. However, a major limitation in this approach is the requirement that the positional origin of atoms in a collisional fragment would be known a priori, which in many cases is difficult to determine.

Results: Here we show that MS/MS data could also be used to improve flux inference even when the positional origin of fragments is unknown. We develop a novel method, metabolic flux analysis/unknown fragments, that extends on 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. MFA/UF is shown to markedly improve flux prediction accuracy in a simulation model of gluconeogenesis and using experimental MS/MS data in *Bacillus subtilis*.

Availability and Implementation: Freely available at www.cs.technion.ac.il/~tomersh/methods.html

Contact: tomersh@cs.technion.ac.il

Received on June 27, 2013; revised on August 19, 2013; accepted on August 28, 2013

1 INTRODUCTION

Metabolic flux analysis (MFA) enables the determination of *in vivo* metabolic fluxes and is commonly used to address problems in biotechnology and medicine (Antoniewicz *et al.*, 2007; Boghigian *et al.*, 2010; Jin *et al.*, 2004; Sauer, 2006; Sillers *et al.*, 2009; Wiechert, 2002). It is based on feeding cells with isotope-labeled nutrients, measuring isotopic labeling patterns of intracellular metabolites and applying computational approaches that analyze metabolite labeling data to estimate fluxes (Antoniewicz *et al.*, 2007; Wiechert *et al.*, 1999). Biotechnological applications of MFA are especially diverse and include metabolic engineering, pathway discovery, filling gaps in annotated genomes and mapping of environmental microbial metabolism (Creek *et al.*, 2012; Kelleher, 2001; Tang *et al.*, 2012; You *et al.*, 2012). The measurement of isotopic labeling patterns uses either NMR (Burgess *et al.*, 2003; Jones *et al.*, 1997) or mass-spectrometry coupled with liquid (LC-MS) or gas (GC-MS) chromatography (Burgess *et al.*, 2003; Kleijn *et al.*, 2007). NMR provides detailed information on positional labeling of metabolites. However, NMR analysis requires a large amount of sample, long analysis times and expensive equipment (Jones *et al.*, 1997). On the other hand, mass spectrometry is more readily accessible, although it typically provides limited information on positional labeling, as isotopomers (i.e. distinct labeling patterns) with the same number of labeled atoms have the same mass regardless of their position. A set of isotopomers with the same mass is referred to as *mass-isotopomers*, whereas the relative abundance of a metabolite's mass-isotopomers is denoted as *mass-isotopomer distribution*.

Tandem mass-spectrometry (MS/MS; e.g. GC-MS/MS and LC-MS/MS) provides information on metabolite positional labeling that can be used to improve flux inference (Choi and Antoniewicz, 2011; Jeffrey *et al.*, 2002; Ruhl *et al.*, 2012). 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. We denote the mass-isotopomers of a parent ion by M0 (having no labeled atoms), M1 (having one labeled atom) and so forth, and the mass-isotopomers of a product ion by m0, m1 and so forth. Various MS/MS ion scanning modes enable to measure the mass-isotopomer distribution of a parent molecule as well as that of a fragmented product. Ruhl *et al.* have recently used LC-MS/MS to measure the mass-isotopomer distribution of various metabolite fragments in *Bacillus subtilis*, which were shown to improve flux inference (Ruhl *et al.*, 2012). MS/MS can further measure the abundance of specific collisional transitions from certain parent to product mass-isotopomers, referred to as *tandem mass-isotopomers* [e.g. a transition from parent mass-isotopomer M1 to product mass-isotopomer m1, denoted (M1 > m1)]. The relative abundance of all tandem mass-isotopomers is referred to as *tandem mass-isotopomer* distribution. The importance of tandem mass-isotopomer distribution data for flux inference was also recently demonstrated (Choi and Antoniewicz, 2011).

A current limitation of using MS/MS data for MFA is that the positional origin of atoms in a collisional fragment must be known a priori, which in some cases may be difficult to infer. One method for inferring fragment positions involves injecting the mass-spec with multiple isotopically labeled standards (Choi *et al.*, 2012). However, this approach is limited by the lack of availability of such standards for most metabolites, as well as the high price of existing isotopic labeled standards. Another approach is to computationally predict fragment positions using rule-based and combinatorial fragmentation methods

*To whom correspondence should be addressed.

(Heinonen *et al*., 2008). Ruhl *et al*. have recently applied such methods to examine the position of 172 MS/MS collisional fragments of 20 metabolites in *B.subtilis*. However, considering that these methods predicted multiple possible positions for most fragments and that further experimental validation of predicted positions is of need, the unique position of just ~10% of the fragments was determined and used for flux inference.

Here, we show that MS/MS data can be highly valuable for MFA, even when fragment positions are unknown. Specifically, we suggest an extension to the standard MFA approach, metabolic flux analysis/unknown fragments (MFA/UFs), which uses steady-state MS/MS data to jointly infer the most likely fluxes in a metabolic network and the positional origin of fragment atoms. The method is first applied to predict fluxes in a simulation model of gluconeogenesis given metabolite tandem mass-isotopomer distribution data. Then, it is applied to predict fluxes in *B.subtilis* given measured mass-isotopomer distributions for both intact metabolites and their collisional fragments. In both cases, an improved flux prediction performance of MFA/UF in comparison with the standard MFA approach is demonstrated.

2 METHODS

2.1 The mathematical formulation of MFA/UF

MFA aims to find a steady-state flux distribution (v), satisfying stoichiometric mass-balance constraint (and potentially other constraints on nutrient uptake and by-product secretion rates), such that metabolite labeling patterns, which are uniquely determined by the flux distribution (Mollney *et al*., 1999; Wiechert *et al*., 1997; Wiechert *et al*., 1999), would optimally match experimental measurements. Specifically, given experimental mass-isotopomer distribution MID_i^{exp} for each metabolite i in some metabolite set E, MFA is formulated as an optimization function that aims to maximize the log likelihood of these measured mass-isotopomers. Assuming a Gaussian error model for measured mass-isotopomer data, maximum log likelihood is obtained by minimizing the variance-weighted sum of squared residuals between measured and predicted ($MID_i^{pred}(v)$) mass-isotopomer distributions:

$$\min_v \sum_{i \epsilon E} \left[\left(MID_i^{exp} - MID_i^{pred}(v) \right)^T \cdot VAR_i^{-1} \cdot \left(MID_i^{exp} - MID_i^{pred}(v) \right) \right]$$

s.t.

$$S \cdot v = 0$$

where VAR_i represents the measurement variance matrix (having measurement variances on the diagonal), and S represents a stoichiometric matrix. In all applications of MFA/UF performed here, we assumed a variance of 0.2 mol% in measured isotope distributions, following Choi and Antoniewicz (2011). The mass-isotopomer distribution for metabolite i ($MID_i^{pred}(v)$) can be efficiently computed for a given flux vector v (while evaluating the objective function of the above optimization problem) via either elementary metabolite unit (Antoniewicz *et al*., 2007) or based on the complete isotopomer distribution of metabolite i (denoted ID_i) derived by either cummomer (Wiechert *et al*., 1999) or fluxomer (Srour *et al*., 2011) approach. Given the isotopomer distribution of metabolite i, its mass-isotopomer distribution can be calculated as following:

$$MID_i^{pred}(v) = C_i ID_i(v) \quad (1)$$

where C_i is a matrix in which $C_{i,j,k} = 1$, if the k'th isotopomer of metabolite i has j labeled atoms.

Applying MFA given a mass-isotopomer distribution for a fragment F of metabolite i (denoted $MID_{i,F}^{exp}$) would require computing the corresponding mass-isotopomer distribution for that fragment based on the flux vector v ($MID_{i,F}^{pred}(v)$), which could also be easily calculated from the predicted isotopomer distribution:

$$MID_{i,F}^{pred}(v) = C_i^F ID_i(v) \quad (2)$$

where C_i^F is a matrix in which $C_{i,j,k}^F = 1$, if fragment F of the k'th isotopomer of metabolite i has j labeled atoms. Similarly, given a tandem mass-isotopomer distribution for fragment F of metabolite i (representing the abundance of detected transitions from a specific parent to product mass-isotopomers in metabolite i; denoted $TMID_{i,F}^{exp}$), it can be calculated from the isotopomer distribution as following:

$$TMID_{i,F}^{pred}(v) = C_i^F ID_i(v) \quad (3)$$

where in this case, C_i^F is a matrix in which $C_{i,j,k}^F = 1$ if the k'th isotopomer of metabolite i is consistent with the j'th tandem isotopomer, in respect to fragment F of metabolite I, i.e. if (i) the total number of labeled atoms in the isotopomer is the same as in the parent ion of the tandem isotopomer, (ii) the number of labeled atoms in that isotopomer, within the fragment F, is the same as in the product ion of that tandem isotopomer. An example for how C_i^F is defined is shown in Equation (5) later in the text (see also Choi and Antoniewicz, 2011). Notably, the predicted tandem-mass isotopomer distribution for metabolite i in respect to fragment F can be adjusted for natural isotope abundances as shown in Choi and Antoniewicz (2011).

Given a mass-isotopomer distribution or tandem mass-isotopomer distribution for some unknown fragment of metabolite i [denoted $(T)MID_i^{exp}$ for simplicity], assuming that each possible fragment F (within a predefined fragments list) can be produced by the MS/MS collision with probability λ_i^F , the corresponding prediction [denoted $(T)MID_i^{pred}$] can be calculated as following:

$$(T)MID_i^{pred}(v,\lambda) = \sum_{F \in \left\{ \begin{array}{c} fragments\ of \\ metabolite\ i \end{array} \right\}} \lambda_i^F C_i^F ID_i(v) \quad (4)$$

here, we suggest a variant of MFA, MFA/UF, that simultaneously searches for a flux vector v and fragment probabilities λ (also referred to as *fragment weight variables*) that maximizes the log likelihood of the measured mass-isotopomer or tandem mass-isotopomer distributions:

$$\min_{v,\lambda} \sum_{i \epsilon E} \left[\left((T)MID_i^{exp} - (T)MID_i^{pred}(v,\lambda_i) \right)^T \cdot VAR_i^{-1} \cdot \left((T)MID_i^{exp} - (T)MID_i^{pred}(v,\lambda_i) \right) \right]$$

s.t.

$$S \cdot v = 0,$$

$$\sum_{F \in \left\{ \begin{array}{c} fragments\ of \\ metabolite\ i \end{array} \right\}} \lambda_i^F = 1, \text{ for each metabolite } i \epsilon E$$

To solve the above optimization problem, we used Matlab's implementation of the sequential quadratic optimization (SQP) algorithm. To overcome local minima, the SQP optimization was run multiple times, starting from random starting points.

To compute flux confidence intervals for a given reaction by MFA/UF (or by our implementation of MFA), we used the likelihood ratio test to compare the maximum log-likelihood estimation, computed by the above SQP optimization, to that obtained when constraining the flux to higher or lower values. Specifically, we iteratively run the SQP optimization to compute the maximum log-likelihood estimation while constraining the flux to increasing (and then decreasing) values (with a step size equal to 5% of the flux predicted in the initial maximum log-likelihood estimation). The confidence interval bounds were determined based on the 95%

quantile of χ^2-distribution with one degree of freedom (Antoniewicz *et al.*, 2006). The results of this iterative process (i.e. the probability of a reaction having a range of possible flux rates) were also used to calculate the standard deviation of flux predictions. A similar method was used to compute confidence intervals for fragment weights (i.e. λ variables). Although the running time of each SQP optimization was on the order of a few minutes, computing accurate confidence intervals with the described approach is computationally intensive and was run on a computer cluster with ~100 CPU cores (taking a total of ~30 h for the analysis presented later in the text).

3 RESULTS

3.1 An example of using MS/MS data for inferring flux rates and position of collisional fragments

Figure 1 shows an example metabolic network, demonstrating how MS/MS data can be used to jointly infer metabolic fluxes and position of collisional fragments. In the example, a 3-carbon metabolite (*A*), labeled in carbons 1 and 2, is being taken up from the medium in a known rate (*v1*). Reaction *v4* transforms metabolite *A* to *D*, making an isotopomer of *D* that is labeled in the first two out of three carbons (denoted *D110*). A pathway through *v2* makes *D* that is labeled in the first carbon (*D100*), and a pathway through v3 makes *D* that is labeled in the third carbon (*D001*). Hence, based on isotopomer balance considerations, all fluxes can be determined given the *D*'s isotopomer distribution (denoted ID_D; representing the relative abundance of each isotopomer of *D*): $v2=v1*ID_{D100}$, $v3=v1*ID_{D001}$, and $v4=v1*ID_{D110}$. Given the mass-isotopomer distribution of *D* (denoted MID_D), standard MFA cannot determine the specific fluxes through *v2* and *v3*, as both *D100* (made through v2) and *D001* (made through v3) belong to the M1 mass-isotopomer of *D*. Next we examine whether the abundances of all three isotopomers of *D* (and hence all fluxes in this network) can be inferred given: (i) a tandem mass-isotopomer distribution of *D* with either known or unknown fragment positions and (ii) a mass-isotopomer distribution for both intact *D* and its fragment, whose position is either known or not.

First, we assume that a tandem mass-isotopomer distribution for metabolite *D* with a 2-carbon fragment *F* is given (denoted $TMID_{D,F}$). Assuming, that the position of fragment *F* is known and that *F* consists of the first two carbons of *D* (denoted $D_{1,2}$), the abundance of all three isotopomers can be inferred [based on Equation (3)]:

$$\begin{bmatrix} TMID_{D,F,M0>m0} \\ TMID_{D,F,M1>m0} \\ TMID_{D,F,M1>m1} \\ TMID_{D,F,M2>m1} \\ TMID_{D,F,M2>m2} \\ TMID_{D,F,M3>m2} \end{bmatrix} = \begin{bmatrix} 0 \\ ID_{D001} \\ ID_{D100} \\ 0 \\ ID_{D110} \\ 0 \end{bmatrix} \quad (5)$$

Alternatively, we assume that the MS/MS collision makes either one of two fragments: $D_{1,2}$ and $D_{2,3}$, or a combination of both with probabilities $\lambda_{1,2}$ and $\lambda_{2,3}$, respectively (where $\lambda_{1,2}+\lambda_{2,3}=1$). Under this assumption, both the isotopomer distribution and fragment probabilities can be calculated as follows [based on Equation (4)]:

$$\begin{bmatrix} TMID_{D,F,M0>m0} \\ TMID_{D,F,M1>m0} \\ TMID_{D,F,M1>m1} \\ TMID_{D,F,M2>m1} \\ TMID_{D,F,M2>m2} \\ TMID_{D,F,M3>m2} \end{bmatrix} = \lambda_{1,2} \begin{bmatrix} 0 \\ ID_{D001} \\ ID_{D100} \\ 0 \\ ID_{D110} \\ 0 \end{bmatrix} + \lambda_{2,3} \begin{bmatrix} 0 \\ ID_{D100} \\ ID_{D001} \\ ID_{D110} \\ 0 \\ 0 \end{bmatrix} \quad (6)$$

As shown, tandem mass-isotopomer [*M2>m2*] is produced solely from the $D_{1,2}$ fragmentation of *D110*, and [*M2>m1*] produced from the $D_{2,3}$ fragmentation of *D110*; hence, the abundance of *D110* is equal to the sum of abundances of [*M2>m2*] and [*M2>m1*] (as $\lambda_{1,2}+\lambda_{2,3}=1$). The fragment probabilities can hence be calculated based on the relative abundance of the two tandem mass-isotopomers: $\lambda_{1,2}=[M2>m2]/([M2>m2]+[M2>m1])$. Given the

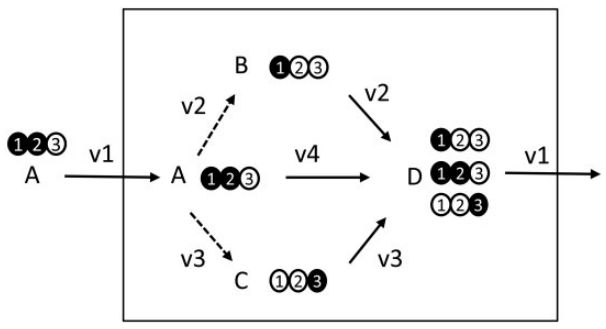

Fig. 1. An example metabolic network model and the labeling pattern of its metabolites. As shown, metabolite *D* is labeled in its first two carbons (for *D* made via reaction *v4*), only in its first carbon (for *D* made via *v2*) or only in its last carbon (for *D* made via *v3*)

Table 1. Four examples of tandem mass-isotopomer distributions for metabolite D in the example network (Fig. 1) and the corresponding predictions of collisional fragment probabilities and fluxes [as inferred from Equation (4)]

	Parameter	Case I	Case II	Case III	Case IV
Tandem mass-isotopomer distribution of metabolite D	[M0>m0]	0	0	0	0
	[M1>m0]	0.1	0.1	0.2	0
	[M1>m1]	0.3	0.3	0.2	0.4
	[M2>m1]	0	0.15	0.3	0.15
	[M2>m2]	0.6	0.45	0.3	0.45
	[M3>m3]	0	0	0	0
Fragment probabilities	$\lambda_{1,2}$	1	0.75	0.5	
	$\lambda_{2,3}$	0	0.25	0.5	
Fluxes	*v2*	0.3	0.4	[0–0.4]	
	v3	0.1	0	[0–0.4]	
	v4	0.6	0.6	0.6	

Note: In 'Case I', the non-zero abundance of [M2>m2] and zero abundance of [M2>m1] suggests the existence of a single fragment of *D*, $D_{1,2}$, and hence the flux through reactions *v2* and *v3* is proportional to the abundance of [M1>m1] and [M1>m0], respectively. In 'Case II', the ratio between [M2>m2] and [M2>m1] is 3:1, suggesting that fragment $D_{1,2}$ is three times more frequent than $D_{2,3}$. In 'Case III', the probability of the two fragments is equal ($\lambda_{1,2}=0.5$ and $\lambda_{2,3}=0.5$), and it is hence impossible to differentiate between flux going through reactions *v2* and *v3* [which is reflected by the multiple solutions to Equation (7)]. In 'Case IV', there is no solution to Equation (7), suggesting that one of the undelaying assumptions regarding the network structure, reactions atom mapping or possible fragment positions is incorrect.

inferred fragment probabilities, ID_{D100} ID_{D100} and ID_{D001} linearly depend on the abundances of [*M1>m0*] and [*M1>m1*] and are straightforwardly calculated. Table 1 shows several examples of possible tandem mass-isotopomers of *D*, and the corresponding fragment probabilities and fluxes derived based on Equation (6). Two notable examples are 'Case III', where the probability of the two fragments is equal to 0.5, and hence Equation (6) is underdetermined and cannot be used to uniquely infer ID_{D100} and ID_{D001}. In 'Case IV', there is no solution to Equation (6), suggesting that at least one the undelaying assumptions regarding the network structure, reactions atom-mapping or considered fragment positions of *D* is incorrect.

Next, we consider the case where a mass-isotopomer distribution for both the intact *D* (denoted MID_D) and a 2-carbon fragment *F* of *D* are given (denoted $MID_{D,F}$). If the fragment position is known to be $D_{1,2}$, for example, then all three isotopomers of *D* can be easily inferred (not shown). Alternatively, assuming the fragment position is unknown, and may be either $D_{1,2}$ or $D_{2,3}$ (with probabilities $\lambda_{1,2}$ and $\lambda_{2,3}$, respectively), as in the above example, $MID_{D,F}$ $MID_{D,F}$ can be expressed as following [based on Equation (4)]:

$$\begin{bmatrix} MID_{D,F,M0} \\ MID_{D,F,M1} \\ MID_{D,F,M2} \end{bmatrix} = \lambda_{1,2} \begin{bmatrix} ID_{D001} \\ ID_{D100} \\ ID_{D110} \end{bmatrix} + \lambda_{2,3} \begin{bmatrix} ID_{D100} \\ ID_{D110} + ID_{D001} \\ 0 \end{bmatrix} \quad (7)$$

Considering that the mass-isotopomer distribution of the intact *D* can be expressed as:

$$\begin{bmatrix} MID_{D,M0} \\ MID_{D,M1} \\ MID_{D,M2} \\ MID_{D,M3} \end{bmatrix} = \begin{bmatrix} 0 \\ ID_{D100} + ID_{D001} \\ ID_{D110} \\ 0 \end{bmatrix} \quad (8)$$

both the fragment probabilities and abundances of all three isotopomers of *D* can be inferred from equations in Equations (7) and (8).

3.2 Applying MFA/UF in a simulation model of gluconeogenesis

To evaluate the performance of MFA/UF, we applied it as well as on standard MFA on a simplified model of gluconeogenesis shown in Figure 2 (adapted from (Choi and Antoniewicz, 2011). The stoichiometry and atom transitions for the network reactions are given in Choi and Antoniewicz (2011). Aspartate and acetyl-CoA are provided as substrates, whereas glucose and carbon dioxide are products. We assumed a mixture of aspartate isotopomers in the media as in Choi and Antoniewicz (2011), with the following composition: 30% [1, 2-^{13}C] aspartate, 20% [4-^{13}C] aspartate, 25% [2, 3, 4-^{13}C] aspartate and 25% unlabeled aspartate, while provided acetyl-coA is unlabeled. The carbon dioxide pool is assumed to remain non-labeled, owing to high rate of mixing with the non-labeled, atmospheric CO_2. In all applications of MFA and MFA/UF, citrate synthase flux

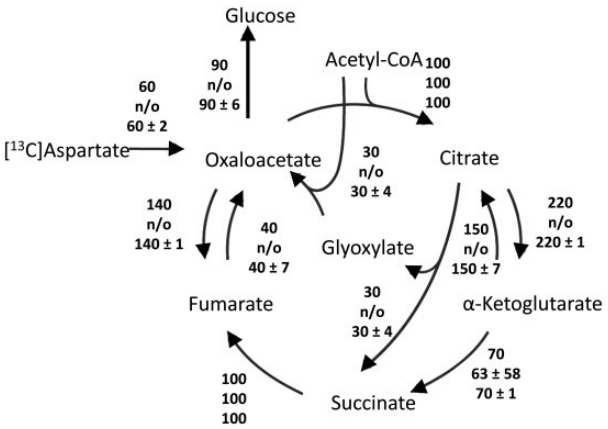

Fig. 2. A metabolic network of gluconeogenesis, with TCA cycle and glyoxylate shunt [adapted from (Choi and Antoniewicz, 2011)]. For each reaction, the assumed flux, predicted flux via MFA given a mass-isotopomer distribution for citrate and predicted flux via MFA/UF given a tandem mass-isotopomer distribution for citrate with the fragment $citrate_{1,2,3}$ are shown (from top to bottom). Reactions for which the predicted flux standard deviation is as high as the assumed flux were considered non-observable (and marked 'n/o')

(making citrate from oxaloacetate and acetyl-CoA) was assumed to be equal to 100 ± 0.1.

The evaluation of MFA/UF and MFA involved the following procedure: (i) We assumed that metabolic fluxes are known for all reactions in the network, considering the fluxes from Choi and Antoniewicz (2011) (shown in Fig. 2). (ii) Given these fluxes, we computed the mass-isotopomer distribution for five metabolites: citrate, α-ketoglutarate, succinate, fumarate and oxaloacetate. We further computed tandem-mass isotopomer distributions for citrate considering 4 different fragments of size 4, for α-ketoglutarate considering 3 fragments of size 3, and for succinate, fumarate and oxaloacetate considering 3 fragments of size 2, giving a total of 16 tandem mass-isotopomer distributions. (iii) We applied MFA given the mass-isotopomer distribution of citrate, α-ketoglutarate, succinate, fumarate or oxaloacetate (a total of 5 runs), and applied MFA/UF given each of the 16 tandem mass-isotopomer distributions described earlier in the text. In the application of MFA/UF, we assumed that the fragment position is unknown, providing the algorithm with a list of possible fragments in each run (assuming the same four possible fragments for citrate and three fragments for α-ketoglutarate, succinate, fumarate and oxaloacetate used earlier in the text). (iv) Reassuringly, in all applications of MFA and MFA/UF, the predicted 95% flux confidence intervals encompassed the assumed fluxes (used in the generation of labeling data), and hence the predictive performance of the two approaches was evaluated based on the size of the confidence intervals (expressed in percentage of the assumed flux).

We find that the prediction performance of MFA/UF, given tandem mass-isotopomer distributions (assuming fragment positions are unknown), is markedly higher than that of standard MFA given the mass-isotopomer distribution data (Fig. 3). Specifically, confidence interval sizes derived by MFA/UF are ~2-fold smaller on average than those derived by MFA (across all reactions and applications of MFA/UF and MFA). Given 11

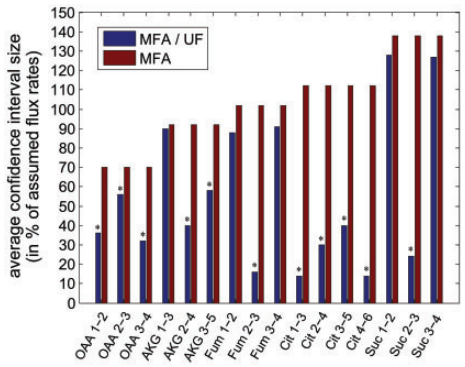

Fig. 3. Average size of flux confidence intervals in the gluconeogenesis model predicted by MFA based on mass-isotopomer distributions of intact metabolites and by MFA/UF given tandem mass-isotopomer distributions of various collisional fragments of oxaloacetate, α-ketoglutarate, fumarate, citrate and succinate (fragment carbons in metabolite M denoted by Mx-y). Applications of MFA/UF that result in significantly smaller flux confidence intervals compared with those obtained by MFA (based on Wilcoxon ranked sum test) are marked with an asterisk. Metabolite abbreviations: OAA, oxaloacetate; AKG, α-ketoglutarate; Fum, fumarate; Cit, citrate; Suc, succinate

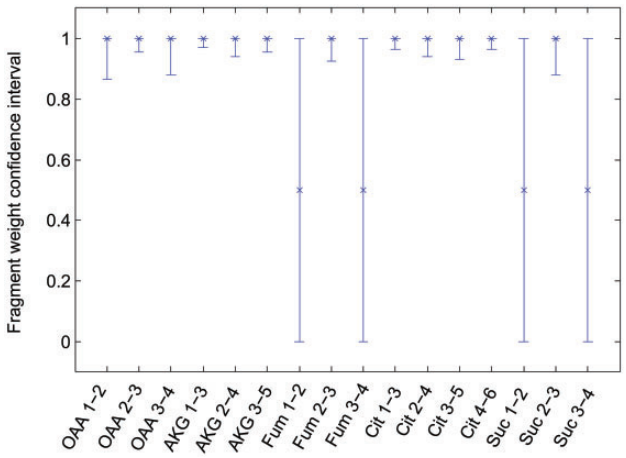

Fig. 4. Predicted fragment weights by MFA/UF given tandem mass-isotopomer distributions of various collisional fragments of oxaloacetate, α-ketoglutarate, fumarate, citrate and succinate. For each application of MFA/UF given a tandem mass-isotopomer distribution for a metabolite with a certain fragment, the confidence interval of the weight variable corresponding to that specific fragment is shown. A value of 1 represents a prediction of the correct fragment being the only product of the collision, whereas lower values represent the possible existence of other fragments. Metabolite abbreviations are as in Figure 3

of the 16 tandem-mass isotopomer distributions, the confidence interval sizes predicted by MFA/UF are significantly smaller than those predicted by MFA-based labeling data of the corresponding intact metabolites (Wilcoxon ranked sum $P<0.05$; FDR corrected for multiple testing; marked with asterisk in Fig. 3). For example, the confidence intervals predicted by MFA/UF given the tandem mass-isotopomer distribution of citrate with the fragment $citrate_{1,2,3}$ are of an average size of 14%, whereas those predicted by MFA are of an average size of 112% (show in Fig. 2).

Next, we explored whether MFA/UF correctly inferred the true fragment positions when applied to predict fluxes with each of the 16 given tandem mass-isotopomer distributions. We found that for citrate, α-ketoglutarate and oxaloacetate, the most likely solution predicted by MFA/UF correctly identified the position of the single fragment used to simulate the tandem mass-isotopomer distribution (with the corresponding fragment weight variable λ being equal to 1; Fig. 4). In simulations with tandem mass-isotopomer data for fumarate and succinate, the correct position of some fragments could not be inferred owing to structural symmetry of these metabolites. For example, for fumarate, MFA/UF cannot differentiate between a 2-carbon fragment consisting of carbons 1 and 2 ($Fum_{1,2}$) or carbons 3 or 4 ($Fum_{3,4}$).

3.3 Applying MFA/UF to predicting metabolic fluxes in *B.subtilis* based on measured MS/MS data

To further demonstrate the applicability of MFA/UF, we applied it to predict metabolic fluxes in central metabolism of *B.subtilis*, using recent MS/MS mass-isotopomer measurements (Ruhl *et al.*, 2012). Measurements were done with a growth media consisting of a mixture of 20% (w/w) [U-^{13}C] and 80% [1-^{13}C] glucose. The used metabolic model includes 42 reactions in glycolysis, pentose-phosphate pathway and TCA cycle (Ruhl *et al.*, 2012). For seven metabolites in the network (glucose-6-phosphate, pentose-5-phosphate, phosphoenolpyruvate, pyruvate, oxoglutarate, succinate, oxaloacetate), a mass-isotopomer distribution is available for both the intact metabolite and for between 1 and 4 known fragments. In all applications of MFA and MFA/UF to predict fluxes in this model, metabolite uptake and secretion rates as well as growth-associated biomass demand were fixed based on experimental measurements (Ruhl *et al.*, 2012). We assume experimental error of 20% in measurement of metabolite uptake and secretion and biomass demand, and hence allow predicted rates through the corresponding reactions to deviate by up to 20% from the measured fluxes. The abundances of extracellular glucose and CO_2 isotopomers used as input for all MFA and MFA/UF analysis were adjusted to account for natural isotopic abundances and impurity of labeled glucose [as done in Ruhl *et al.* (2012)].

To evaluate the performance of MFA/UF versus that of standard MFA in predicting fluxes in this model, we applied both approaches given labeling data for randomly sampled sets of metabolites. Specifically, we sample sets of metabolites with 2, 3 and 4 metabolites (5 sets of each) and considered also 7 additional sets each having a single metabolite (resulting in a total of 22 metabolite sets). Given each metabolite set, we performed the following analysis: (i) Run MFA given the labeling data of the intact metabolites; (ii) Run MFA/UF given the labeling data for both intact metabolites and their fragments, assuming fragment positions are unknown. In this case, for each fragment with a given mass-isotopomer distribution, MFA/UF considered all possible positions such that fragment carbons are connected in the parent molecule (considering between 1 and 4 possible positions per fragment). (iii) To evaluate the accuracy of predictions made by MFA (in i) and by MFA/UF (in ii), we compared the

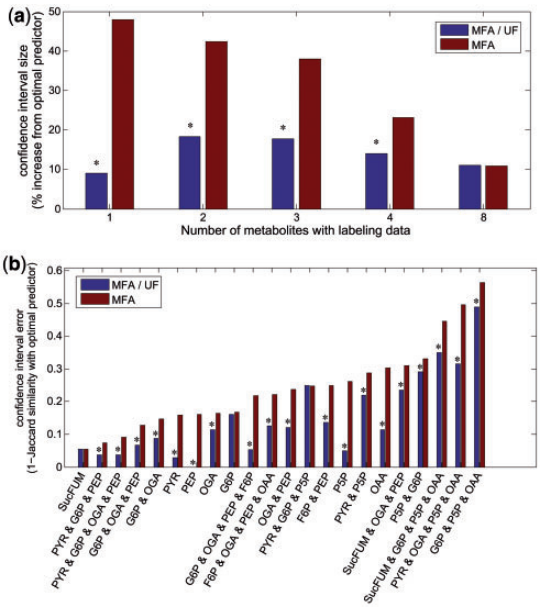

Fig. 5. (**a**) Size of flux confidence intervals in *B.subtilis* predicted by MFA based on MS data and by MFA/UF given MS/MS data for randomly chosen sets of metabolites of different sizes (*x*-axis). The *y*-axis represents the average percentage increase in confidence interval size predicted by MFA and MFA/UF compared with those predicted by the optimal predictor (i.e. MFA using MS/MS data, assuming fragment positions are known). Metabolite set sizes for which the confidence interval sizes predicted by MFA/UF are significantly smaller than those predicted by MFA (based on Wilcoxon ranked sum test) are marked with an asterisk. (**b**) The average error in flux confidence intervals predicted by MFA given MS data and MFA/UF given MS/MS data (*y*-axis) for randomly chosen sets of metabolites (*x*-axis). The error measure is defined as 1 minus the Jaccard similarity coefficient between a confidence interval predicted by MFA or MFA/UF and that derived by the optimal predictor. Applications of MFA/UF that result in significantly smaller flux confidence intervals errors than those predicted by MFA (based on Wilcoxon ranked sum test) are marked with an asterisk. Metabolite abbreviations: G6P, glucose-6-phosphate; P5P, pentose-5-phosphate; PEP, phosphoenolpyruvate; PYR, pyruvate; OGA, oxoglutarate; SucFUM succinate/fumarate; OAA, oxaloacetate

flux predictions made by these methods with those obtained by an *optimal* predictor, which is MFA given labeling data for both the intact metabolites and their fragments, assuming fragment positions are known.

Examining the size of the 95% flux confidence intervals predicted by MFA/UF and MFA (without using fragment labeling data), we find that the confidence intervals derived by MFA/UF are markedly smaller (Fig. 5a). For example, given the labeling pattern of just single metabolites, the confidence intervals derived by MFA/UF are just 9% larger on average than those obtained by the optimal predictor, whereas confidence intervals predicted by MFA without fragment labeling data are 47% larger on average than those of the optimal predictor. Notably though, the average confidence interval size predicted by MFA decrease as labeling data for more metabolites is given as input. Given the labeling data for all metabolites, MFA/UF's and MFA's predicted confidence interval sizes are roughly the same (11% higher than the optimal predictor).

To further evaluate the predicted fluxes by MFA/UF versus those of MFA (with no fragment labeling data), we compared the confidence interval derived by these methods with those obtained by the optimal predictor based on Jaccard coefficient (i.e. size of intersection divided by size of union). The error in predicted flux confidence intervals was defined as one minus the Jaccard coefficient. We find that in 19 of 22 applications of MFA/UF, its prediction error is significantly lower than that of MFA (Wilcoxon ranked sum $P<0.05$; FDR corrected for multiple testing; Fig. 5b). Overall, we find an average reduction of 40% in prediction error in MFA/UF versus standard MFA. A notable example is the flux prediction given the labeling data for phosphoenolpyruvate, where MFA/UF predicts the same flux confidence intervals as those derived by the optimal predictor, whereas MFA predicts markedly different confidence intervals, with an average error of 0.18 (Fig. 5b). As another example, Table 2 shows fluxes and confidence intervals predicted by MFA, MFA/UF and the optimal predictor given labeling data for pyruvate, oxoglutarate, pentose-5-phosphate and oxaloacetate. In this case, we highlight several reactions for which MFA but not MFA/UF predicted confidence interval falsely consists of a zero flux.

4 DISCUSSION

Tandem-MS data holds great promise in improving the accuracy of metabolic flux inference, by providing higher measurement precision over standard MS (Kiefer *et al.*, 2007) and providing information on metabolite positional labeling (Choi and Antoniewicz, 2011; Jeffrey *et al.*, 2002). To make usage of MS/MS data for flux inference, previous computational approaches require data on the positional origin of atoms in collisional fragments, which is difficult to obtain in many cases. Here, we show that tandem MS data can be highly useful even when collisional fragment positions are unknown, suggesting a computational framework, MFA/UF, that simultaneously determines both flux rates and likely fragment positions. MFA/UF can be applied on both measurements of tandem mass-isotopomer distributions (as demonstrated with the gluconeogenesis simulation model) and mass-isotopomer distribution for collisional fragments (as shown with experimental data in *B.subtilis*). For both types of MS/MS data, MFA/UF is shown to markedly improve flux prediction accuracy compared with standard MFA, relying solely on MS data.

The applicability of MFA/UF was demonstrated for cases in which MS/MS analysis does not induce in-source fragmentation and hence the mass of the parent, non-fragmented metabolite is measured. However, it can be easily adapted to handle the following cases: (i) Standard MS analysis (either GC-MS or LC-MS) in which the mass-isotopomer distribution of an in-source fragment (with an unknown position) is given. In this case, MFA/UF would consider alternative in-source rather than post-source fragments, without requiring any change to the MFA/UF formulation itself. (ii) MS/MS analysis in which the position of an in-source fragment is known though that of the post-source collisional fragment is unknown. This case would require defining C_i^F, in Equation (3), considering the known

Table 2. Fluxes and confidence intervals predicted by MFA, MFA/UF and the optimal predictor (i.e. MFA given labeling data for both the intact metabolites and their fragments, assuming fragment positions are known) in *B.subtilis* [see complete model in (Ruhl *et al.*, 2012)], given labeling data for pyruvate, oxoglutarate, pentose-5-phosphate and oxaloacetate

Reaction	MFA		MFA/UF		Optimal predictor	
	Flux	Confidence interval	Flux	Confidence interval	Flux	Confidence interval
G6P –> F6P	5.7	3.7, 8	5.6	4.4, 6.6	5.3	3.9, 6.2
F6P –> FBP	6.7	6, 7.5	6.6	6.1, 7.1	6.6	6.1, 7
FBP –> GAP + DHAP	6.7	5.9, 7.5	6.6	6.1, 7	6.6	6.1, 7
DHAP –> GAP	6.7	5.9, 7.5	6.7	6.1, 7.1	6.6	6.1, 7
GAP –> BPG	13.6	12.5, 14.6	13.6	12.8, 14.2	13.6	12.8, 14.1
BPG –> PGA	13.6	12.5, 14.6	13.6	12.8, 14.2	13.5	12.8, 14.1
PGA –> PEP	12.0	10.7, 13.2	11.9	10.7, 12.8	12.0	11.1, 12.5
PEP –> PYR	11.4	9.8, 12.5	11.2	9.8, 12.1	11.3	10.2, 12
G6P –> P6G	2.3	0, 4.1	2.4	1.4, 3.6	2.7	1.8, 3.8
P6G –> P5*P* + CO2	2.3	0, 4.3	2.4	0.9, 3.8	2.7	1.8, 3.8
P5*P* + E4P –> GAP + F6P	0.3	−0.5, 1	0.3	−0.1, 0.7	0.5	0.2, 0.9
(2) P5P –> S7*P* + GAP	0.7	0, 1.4	0.7	0.4, 1.1	0.8	0.5, 1.3
GAP + S7P –> E4*P* + F6P	0.7	0, 1.4	0.7	0.3, 1.3	0.9	0.5, 1.3
PYR –> AcCoA + CO2	6.8	5.2, 7.9	6.4	4.8, 7.6	6.7	5.6, 7.3
OAA + AcCoA –> CitICit	3.2	1.7, 4.5	2.8	1.2, 4	2.8	1.4, 3.8
CitICit –> OGA + CO2	3.3	1.7, 4.3	2.9	1.2, 4	2.8	1.4, 3.8
OGA –> SucFUM + CO2	2.4	0.9, 3.5	2.1	0.5, 3	1.9	0.5, 3
SucFUM –> MAL	1.2	0.5, 1.8	1.1	0.3, 1.5	1.0	0.2, 1.5
MAL –> OAA	1.3	−0.3, 2.4	1.2	−0.4, 2.5	1.2	−0.4, 2.5

Note: Reactions for which standard MFA but not MFA/UF predicted confidence interval falsely crosses 0 are colored.

position of the in-source fragment. (iii) MS/MS analysis in which the position of both the in-source and post-source fragments are unknown. Adapting MFA/UF for this case would require changing Equation (4) to account for all combinations of possible in-source and post-source fragment positions. The extent to which MFA/UF would improve flux estimates in the above scenarios is an open question.

Notably, fragment positions inferred by MFA/UF may theoretically be further used by standard MFA for flux analysis via MS/MS data. In practice, this may be complicated by noise in experimental data preventing MFA/UF from uniquely determining fragment positions. For example, in all simulations described earlier in the text in gluconeogenesis, a single collisional fragment was assumed when generating tandem mass-isotopomer distributions, though MFA/UF predicted the potential existence of other low probability fragments (in cases not involving symmetric metabolite structures where specific fragment positions could not be inferred; Fig. 4). The probability of such additional possible fragments is taken into account by MFA/UF when computing the most likely fluxes and confidence intervals.

A major open question in MFA, which has been partially addressed by recent studies, is how to design labeling experiments that maximize the accuracy of flux inference or revealing new pathways in novel microbes? (Crown and Antoniewicz, 2012; Crown *et al.*, 2012; Mollney *et al.*, 1999; Schellenberger *et al.*, 2012; Tang *et al.*, 2012). The design of a labeling experiment entails finding an optimal set of labeled nutrients, which produce a pattern of intracellular metabolite labeling, which can be decoded to infer relevant fluxes. To improve the utility MS/MS experiments for flux inference, further challenges remain. First, identify specific metabolite fragmentation patterns whose positional labeling can improve flux inference. Second, design specific labeling experiments to identify the position (or validate predicted positions, e.g. by MFA/UF) of collisional fragments that are likely to improve flux inference. For example, to find the position of a collisional fragment that appears in an MS/MS, possible experiments could involve feeding a specific labeled nutrient to a microbe, potentially in combination with one or more gene knockouts, and applying MS/MS analysis.

As the difficulty of identifying the positional origin of collisional fragments is a major bottleneck preventing broader use of MS/MS data for metabolic flux inference, the computational approach presented here could promote broader usage of this technology.

ACKNOWLEDGEMENT

The authors would like to thank Sean Hackett for commenting on the article.

Funding: Israel Ministry of Science (to T.S.).

Conflict of Interest: none declared.

REFERENCES

Antoniewicz,M.R. *et al.* (2006) Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements. *Metab. Eng.*, **8**, 324–337.

Antoniewicz,M.R. *et al.* (2007) Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. *Metab. Eng.*, **9**, 68–86.

Boghigian,B.A. *et al.* (2010) Metabolic flux analysis and pharmaceutical production. *Metab. Eng.*, **12**, 81–95.
Burgess,S.C. *et al.* (2003) Noninvasive evaluation of liver metabolism by 2H and 13C NMR isotopomer analysis of human urine. *Anal. Biochem.*, **312**, 228–234.
Choi,J. and Antoniewicz,M.R. (2011) Tandem mass spectrometry: a novel approach for metabolic flux analysis. *Metab. Eng.*, **13**, 225–233.
Choi,J. *et al.* (2012) Measuring complete isotopomer distribution of aspartate using gas chromatography/tandem mass spectrometry. *Anal. Chem.*, **84**, 4628–4632.
Creek,D.J. *et al.* (2012) Stable isotope-assisted metabolomics for network-wide metabolic pathway elucidation. *Anal. Chem.*, **84**, 8442–8447.
Crown,S.B. and Antoniewicz,M.R. (2012) Selection of tracers for 13C-metabolic flux analysis using elementary metabolite units (EMU) basis vector methodology. *Metab. Eng.*, **14**, 150–161.
Crown,S.B. *et al.* (2012) Rational design of (1)(3)C-labeling experiments for metabolic flux analysis in mammalian cells. *BMC Syst. Biol.*, **6**, 43.
Heinonen,M. *et al.* (2008) FiD: a software for ab initio structural identification of product ions from tandem mass spectrometric data. *Rapid Commun. Mass Spectrom.*, **22**, 3043–3052.
Jeffrey,F.M. *et al.* (2002) 13C isotopomer analysis of glutamate by tandem mass spectrometry. *Anal. Biochem.*, **300**, 192–205.
Jin,E.S. *et al.* (2004) Glucose production, gluconeogenesis, and hepatic tricarboxylic acid cycle fluxes measured by nuclear magnetic resonance analysis of a single glucose derivative. *Anal. Biochem.*, **327**, 149–155.
Jones,J.G. *et al.* (1997) Measurement of gluconeogenesis and pyruvate recycling in the rat liver: a simple analysis of glucose and glutamate isotopomers during metabolism of [1,2,3-(13)C3]propionate. *FEBS Lett.*, **412**, 131–137.
Kelleher,J.K. (2001) Flux estimation using isotopic tracers: common ground for metabolic physiology and metabolic engineering. *Metab. Eng.*, **3**, 100–110.
Kiefer,P. *et al.* (2007) Determination of carbon labeling distribution of intracellular metabolites from single fragment ions by ion chromatography tandem mass spectrometry. *Anal. Biochem.*, **360**, 182–188.
Kleijn,R.J. *et al.* (2007) Metabolic flux analysis of a glycerol-overproducing Saccharomyces cerevisiae strain based on GC-MS, LC-MS and NMR-derived C-labelling data. *FEMS Yeast Res.*, **7**, 216–231.
Mollney,M. *et al.* (1999) Bidirectional reaction steps in metabolic networks: IV. Optimal design of isotopomer labeling experiments. *Biotechnol. Bioeng.*, **66**, 86–103.
Ruhl,M. *et al.* (2012) Collisional fragmentation of central carbon metabolites in LC-MS/MS increases precision of (1)(3)C metabolic flux analysis. *Biotechnol. Bioeng.*, **109**, 763–771.
Sauer,U. (2006) Metabolic networks in motion: 13C-based flux analysis. *Mol. Syst. Biol.*, **2**, 62.
Schellenberger,J. *et al.* (2012) Predicting outcomes of steady-state (1)(3)C isotope tracing experiments using Monte Carlo sampling. *BMC Syst. Biol.*, **6**, 9.
Sillers,R. *et al.* (2009) 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.*, **102**, 38–49.
Srour,O. *et al.* (2011) Fluxomers: a new approach for 13C metabolic flux analysis. *BMC Syst. Biol.*, **5**, 129.
Tang,J.K. *et al.* (2012) Recent advances in mapping environmental microbial metabolisms through 13C isotopic fingerprints. *J. Roy. Soc. Interface*, **9**, 2767–2780.
Wiechert,W. (2002) An introduction to 13C metabolic flux analysis. *Genet. Eng.*, **24**, 215–238.
Wiechert,W. *et al.* (1997) Bidirectional reaction steps in metabolic networks: II. Flux estimation and statistical analysis. *Biotechnol. Bioeng.*, **55**, 118–135.
Wiechert,W. *et al.* (1999) Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems. *Biotechnol. Bioeng.*, **66**, 69–85.
You,L. *et al.* (2012) Metabolic pathway confirmation and discovery through (13)C-labeling of proteinogenic amino acids. *J. Vis. Exp.*, e3583.