Metabolic flux elucidation for large-scale models using ^{13}C labeled isotopes

Patrick F. Suthers[a], Anthony P. Burgard[b], Madhukar S. Dasika[a], Farnaz Nowroozi[c], Stephen Van Dien[b], Jay D. Keasling[c], Costas D. Maranas[a],*

[a]*Department of Chemical Engineering, The Pennsylvania State University, University Park, PA 16802, USA*
[b]*Genomatica, Inc, 5405 Morehouse Drive, Suite 210, San Diego, CA 92121, USA*
[c]*Department of Chemical Engineering, University of California - Berkeley, Gilman Hall, Berkeley, CA 94720-1462, USA*

Abstract

A key consideration in metabolic engineering is the determination of fluxes of the metabolites within the cell. This determination provides an unambiguous description of metabolism before and/or after engineering interventions. Here, we present a computational framework that combines a constraint-based modeling framework with isotopic label tracing on a large scale. When cells are fed a growth substrate with certain carbon positions labeled with ^{13}C, the distribution of this label in the intracellular metabolites can be calculated based on the known biochemistry of the participating pathways. Most labeling studies focus on skeletal representations of central metabolism and ignore many flux routes that could contribute to the observed isotopic labeling patterns. In contrast, our approach investigates the importance of carrying out isotopic labeling studies using a more comprehensive reaction network consisting of 350 fluxes and 184 metabolites in *Escherichia coli* including global metabolite balances on cofactors such as ATP, NADH, and NADPH. The proposed procedure is demonstrated on an *E. coli* strain engineered to produce amorphadiene, a precursor to the antimalarial drug artemisinin. The cells were grown in continuous culture on glucose containing 20% [U-^{13}C]glucose; the measurements are made using GC–MS performed on 13 amino acids extracted from the cells. We identify flux distributions for which the calculated labeling patterns agree well with the measurements alluding to the accuracy of the network reconstruction. Furthermore, we explore the robustness of the flux calculations to variability in the experimental MS measurements, as well as highlight the key experimental measurements necessary for flux determination. Finally, we discuss the effect of reducing the model, as well as shed light onto the customization of the developed computational framework to other systems.
© 2007 Elsevier Inc. All rights reserved.

Keywords: Metabolic flux analysis; Isotope labeling; Constraint-based modeling; Nonlinear optimization; Statistical analysis

1. Introduction and objectives

In recent years, high-throughput methods have enabled rapid advances in the rate of data generation in genomics, transcriptomics, and proteomics. Despite a number of advances in the experimental measurement and analysis techniques metabolic flux elucidation (i.e., metabolic flux analysis (MFA), Vallino and Stephanopoulos, 1993) have not yet become commonplace as the above-mentioned omics analyses. This is because flux elucidation is context-specific requiring sensitive NMR and/or gas chromatography–mass spectrometry (GC–MS) measurements, accurate metabolic network reconstructions and powerful computational techniques to match experimental observables with underlying models.

The elucidation of metabolic fluxes is important for a number of reasons. First, the set of fluxes through a cell's metabolic pathways is a key descriptor of its physiology (Nielsen, 2003) and for evaluating mechanisms for metabolic engineering (Bailey, 1991; Stephanopoulos and Vallino, 1991). While gene transcripts, protein levels and metabolite concentrations tend to vary in a seemingly unpredictable fashion, metabolic fluxes are the only

---

*Corresponding author.
  E-mail address: costas@psu.edu (C.D. Maranas).

invariant that seems to most effectively capture a cell's metabolic state. They provide the only unambiguous means for pinpointing the effect of engineering interventions (e.g., knock-outs/ins, up/down regulations) on cellular metabolism alluding to their effectiveness and suggesting additional engineering strategies. While optimization-based hypotheses such as biomass yield maximization (Varma and Palsson, 1994), minimization of metabolic adjustments (MOMA) (Segre et al., 2002) and others have in certain cases led to accurate flux predictions, they cannot be relied upon to always reliably deduce the correct flux distributions. Flux measurements (especially internal ones) are needed to provide a reference state (i.e., for MOMA), test the effectiveness of different maximization hypotheses and provide clues as to how to construct new ones.

The potential of using ^{13}C-labeled isotopes to elucidate metabolic pathways has long been recognized in the metabolic engineering community. Early efforts relied on NMR spectra of metabolites which were related to the underlying pathways used to create them (Jeffrey et al., 1991). NMR spectra could also be used to elucidate the flux through metabolic pathways (Bacher et al., 1998; Kelleher, 2001). NMR-based recapitulations of central metabolism fluxes in *Escherichia coli* were accomplished using uniformly labeled glucose (Szyperski, 1995). This work was later expanded to resolve central metabolism fluxes in *E. coli* under a variety of conditions (Sauer et al., 1999). The central metabolic fluxes and assumptions about reversibility in *Bacillus subtilis* were also explored using NMR data that were analyzed within an isotope isomer-(isotopomer)-balancing framework (Dauner et al., 2001). These experimental protocols were subsequently employed to examine the changes in flux distributions in *E. coli* when pyruvate kinase activity was removed (Emmerling et al., 2002).

The use of NMR is hampered by the fact that it is time consuming and requires high concentrations of the metabolites. Alternatively, fluxes can be deduced by generating GC–MS spectra of primarily amino acids (and other metabolites) to de-convolute the metabolic flux distributions. Based on the observed fragment weight patterns the distribution of isotopes can then be deduced. One of the earliest such contributions involved using ^{13}C-labeling MFA GC–MS to determine the central metabolism fluxes for *Corynebacterium glutamicum* (Marx et al., 1996) using enrichment data for 11 amino acids. Similar approaches were used for profiling strains from successive generations of *C. glutamicum* (Wittmann and Heinzle, 2002). Notably ^{13}C-labeling based MFA has also been informative when characterizing regulatory mutations (Van Dien et al., 2003).

Both GC–MS and/or NMR spectra information must first be mapped onto metabolic fluxes before any elucidation procedure can be deployed. One of the first such modeling contributions is the introduction of atom mapping matrices (AMMs) (Zupke and Stephanopoulos, 1994) that tracks the transfer of carbon atoms from reactants to products. This concept was subsequently generalized in the form of isotopomer mapping matrices (IMMs) by Schmidt et al. (1997). The use of IMMs enables the formulation of all isotopomer mass balances of a metabolite pool as closed-form nonlinear algebraic equations. The variables in these representations include the metabolic fluxes and the isotopomer distribution vectors (IDVs) that quantify the fraction of each metabolite being present in a particular isotope form. These modeling developments were used to elucidate central metabolic maps of *E. coli* (Schmidt et al., 1999).

A potential problem with the use of IMMs is that even for a given flux distribution the identification of the underlying IDVs yields a set of equations which remain nonlinear. The elegant cumomer concept (Wiechert et al., 1999) was later introduced to first prove that there exists a unique IDV assignment that satisfies any given feasible flux distribution and subsequently devise an IDV identification procedure by solving a cascade of linear equations. However, the nonlinear coupling between metabolic fluxes and IDVs remains. An alternative method for reducing the dimensional space of the isotopomer problem is the concept of the theoretical bondomer (van Winden et al., 2002). A limitation of this method of analysis is that it requires the use of a single, uniformly labeled substrate. Also, Forbes et al. (2001) introduced the isotopomer path tracing concept which identifies all isotopomer paths that produce an observable isotopomer. Most recently, the elemental metabolic unit (EMU) framework has been developed that dramatically reduces the number of variables necessary to calculate the mass isotopomer distribution of a measured metabolite (Antoniewicz et al., 2007a).

The above-described modeling developments link isotopomer fractions (codified though a variety of different variable sets) and metabolic flux information into a set of algebraic equations that allowed for the straightforward identification of IDVs given a feasible flux distribution. The next step is to solve for the fluxes that best explain the observable data. Most approaches rely on gradient-based minimization searches that minimize the sum of the squares of the differences between measurements and observations. These include the Levenberg–Marquardt algorithm (Zhao and Shimizu, 2003), the generalized reduced gradient method (Klapa et al., 2003) and trust region methods (Yang et al., 2004). Efforts to decrease the computation time led to the development of analytical derivation techniques for the Jacobian matrix (Wittmann and Heinzle, 2002). Alternatively, evolutionary algorithms such as simulated annealing employed by Schmidt et al. (1999) attempt to avoid being trapped in local minima. A potential shortcoming of these approaches is that by relying on essentially local optimization procedures they may get trapped in sub-optimal solutions. To remedy this global optimization approaches relying on branch and bound (e.g., BARON, Tawarmalani and Sahinidis, 2004, a deterministic global optimization package) coupled with

convex relaxation of the problem have recently been used to quantify the fluxes (Ghosh et al., 2005, 2006; Riascos et al., 2005).

While the mathematical rigor gained by pinpointing the globally optimal solution is noteworthy, the practical benefit of this is diluted by the fact that experimental error in the measurements may render the true flux distribution a sub-optimal solution. This motivates the importance of assessing and quantifying the impact of measurement uncertainty on all obtained solutions. Earlier on, linearized statistics were used (Dauner et al., 2001; Emmerling et al., 2002; Wiechert and de Graaf, 1997). In addition, Monte-Carlo stochastic simulation has been used to examine how the elucidated fluxes change when uncertainty in the form of normally distributed noise is added to the data (Forbes et al., 2001; Wittmann and Heinzle, 2002; Zhao and Shimizu, 2003). More recently new techniques were introduced that allow for determination of the upper and lower bounds that do not use local estimates of the standard deviations and can therefore be non-uniform (Antoniewicz et al., 2006). Instead, the flux whose sensitivity is under investigation is increased step-by-step until a threshold χ^2 value is reached.

A common feature key to all the studies so far is that they focus on relatively small subsets of cellular metabolism such as the tricarboxylic acid (TCA) cycle, the pentose-phosphate pathway (PPP), or core metabolism using a coarse description of metabolism with 15–75 reactions at most. This lumping of reaction steps not only obscures detail but also could lead to the erroneous conclusion that the available data are sufficient to elucidate a unique flux distribution. Notably, when lumped metabolic models are projected onto large-scale models significant ambiguity in flux allocation is revealed. Inherently, any bias in the generation of the lumped metabolic model is thus propagated when the flux distribution is elucidated. In addition, considering individual parts of metabolism in isolation of the rest may lead to the elucidation of flux distributions that are not physiologically relevant. A consequence of this is the inability to handle cofactor balancing which in many cases is the bottleneck in the desired product generation. These limitations motivated in this paper the construction of a detailed isotopomer mapping network reaction consisting of 238 reactions (350 fluxes) and 184 metabolites in E. coli which fully accounts for cofactor balancing and biomass drain requirements.

When faced with a much larger isotope mapping model, the computational challenge of resolving fluxes becomes much more pronounced, notably in the determination of the global optimum. To this end, we have devised an iterative optimization procedure for identifying a number of flux distributions that are (local) minima to the minimum square discrepancy problem. After enumerating a multitude of solutions we can then use statistical analysis and biological insight to home in to the physiologically relevant one(s). Finally, by feeding all identified solutions

to an F-test based significance procedure, confidence levels can be assigned to different flux distributions that are supported by the data.

The specific experimental system examined is part of a larger effort to develop strains of E. coli for the production of terpenoid compounds. Terpenoids are a class of isoprenoids often isolated from plants, and are currently used for a variety of applications including anticancer and antimicrobial drugs. For example, artemisinin is a powerful natural antimalarial drug first isolated from wormwood (Dhingra et al., 2000), and Taxol is a drug effective in cancer treatment extracted from the Pacific yew (Cragg, 1998). Because compounds such as these are normally produced in extremely small quantities, purification from biological material is difficult and resource consuming. Furthermore, chemical synthesis of terpenoids is expensive and inefficient (Avery et al., 1992). Production of these compounds in a microbial host would eliminate many of these problems, and a strain of E. coli producing amorphadiene, the precursor to artemisinin, as a model terpenoid compound has recently been developed (Martin et al., 2003). Flux analysis was performed in order to better understand how the metabolic physiology of the organism changes with the introduction of heterologous pathways, and to evaluate how productivity could be improved.

In the following sections we first describe the isotopomer model construction procedure, followed by the optimization procedure and statistical analysis techniques. Comprehensive results are next presented for the amorphodiene producing strain followed by a discussion and summary.

2. Materials and methods

2.1. Experimental system

The labeling experiments were performed in the chemostat with amorphadiene-producing strains of E. coli. The amorphadiene-producing strain (Martin et al., 2003) harbored three plasmids: pMevT, expressing an exogenous pathway for the synthesis of mevalonate from acetyl-CoA; pMBIS, expressing an exogenous pathway for the conversion of mevalonate to the isoprenoid precursor IPP as well as the native E. coli isoprenoid genes idi and ispA; and pADS, expressing the amorphadiene synthase gene for the conversion of farnesyl diphosphate to amorphadiene.

Chemostat cultivations were carried in mineral salts media (Neidhardt et al., 1974) supplemented with thiamine, iron and micronutrients, using a 1-L benchtop fermentor (Sartorius, Göttingen, Germany) at a working volume of approximately 600 mL. The carbon source used was glucose, supplied at 20% uniformly ^{13}C-labeled (U-^{13}C-glucose), with an overall concentration of 10.0 g/L. pH was controlled at 7.0 by the addition of NaOH, the temperature was set at 37 °C, and the vessel was aerated with sterile air. The agitation rate was set at 500 rpm and the dissolved oxygen level was at >50% of saturation at all times. The vessel was inoculated with 30 mL of culture

already grown for several generations at the appropriate label concentration. Once stable optical density was reached, steady state was insured by waiting an additional three vessel volume changes before sampling was begun. Flowrate was maintained to give a dilution rate of $0.19\,h^{-1}$, and the average OD_{600} of the culture during the sampling period was 3.9. After taking each 50 mL sample, the vessel was immediately filled to the original volume using fresh media, and the system was allowed to return to steady state (three volume changes) before the next sample.

Samples were harvested, total protein was extracted and hydrolyzed and the resulting amino acids were derivatized with N-(tert-butyldimethylsilyl)-N-methyl-trifluroacetamide (TBDMS-FA) (Siddiquee et al., 2004). Derivatized amino acids were analyzed using a Hewlett-Packard HP 5971A gas chromatograph-quadrupole mass selective detector (electron impact), equipped with a DB-5 column (Agilent Technologies). Each derivatized sample was injected four times, thus providing at least eight total data points for the calculation of means and variances. Raw mass isotopomer data were corrected for naturally occurring ^{13}C in the derivatization reagents and non-carbon isotopes in the entire fragment using an infinite dimensional matrix calculus method (Wahl et al., 2004). In addition, the aqueous phase of culture supernatants was assayed for residual glucose (2300 STAT Plus glucose analyzer, YSI Life Sciences, Yellow Springs, Ohio) and acetate (kit from R-Biopharm AG, Darmstadt, Germany), while the concentration of amorphadiene was determined in the organic phase. Amorphadiene was measured by first diluting 10 μl of the dodecane phase in 990 μl ethyl acetate, and then quantifying using GC–MS by comparison to an amorphadiene standard run on GC–MS.

2.2. Construction of large-scale isotopomer mapping model

The isotopomer model was constructed using as a starting point the large-scale stoichiometric model of *E. coli* metabolism, *i*JR904 (Reed et al., 2003), which has been successfully applied to predict the phenotypes of various *E. coli* strains, both wild-type and mutant, under certain conditions (Fong and Palsson, 2004; Ibarra et al., 2002). The *i*JR904 model contains 931 intracellular reactions although many of these reactions do not contribute to the labeling patterns of the examined amino acids. Accordingly, in this study, we developed a smaller, though still biologically comprehensive, metabolic model for the flux elucidation. The model reduction procedure first involved removing from *i*JR904 all blocked reactions defined as reactions that cannot carry flux during aerobic growth on glucose due to stoichiometric limitations (Burgard et al., 2004). Examples of blocked reactions include transporter reactions for components absent from the media such as xylose, glycerol, or fructose, and reactions involved in their utilization. This enabled the removal of nearly one third of the reactions from *i*JR904. In addition, reactions in the cell envelope, membrane lipid,

nucleotide, and cofactor biosynthetic pathways were not explicitly included in the model enabling the removal of approximately another one third of the reactions from *i*JR904. However, because these biosynthetic pathways can affect amino acid labeling patterns by draining precursor metabolites away from central metabolism, the biomass equation, which was based largely on the one described in Edwards and Palsson (2000), was modified to account for these additional drains on central metabolism. Lastly, we assumed that the catabolic reactions in the nucleotide salvage pathways do not contribute significantly to the labeling patterns of the amino acids.

The model constructed for this study contained 238 reactions (including biomass) and 184 metabolites, as well as 32 exchange fluxes. This total includes a set of reactions absent from *i*JR904 that enable amorphadiene production via the non-native mevalonate pathway (Martin et al., 2003). Reversible reactions were broken into their forward and backward components to enable the investigation of how reaction reversibility affects the labeling patterns of the amino acids. The network contained 80 such reversible reactions, bringing the total number of flux describing variables to 350. The model included all reactions of Embden–Meyerhoff–Parnas (EMP) and Entner–Doudoroff (ED) glycolysis, the tricarboxylic acid (TCA) cycle, and the PPP. In addition, all of the anaplerotic reactions and amino acid biosynthesis and degradation pathways were included. Finally, the model enforced the explicit balancing of all metabolic cofactors (e.g., ATP, NADH, NADPH) by including reactions for energy generation via substrate-level and oxidative phosphorylation as well as transhydrogenase activity.

The AMMs, which describe the transfer of carbon atoms from the reactants to products (Zupke and Stephanopou-los, 1994), were calculated using Pipeline Pilot^{TM} (SciTegic Inc., San Diego, CA), a high-throughput data analysis and mining system for chemoinformatic applications, using a structural matching algorithm. AMMs are matrices where the columns represent carbon atoms in the reactant and the rows represent carbon atoms in the product. The numbering scheme is consistent with the order of atoms represented in the molecule (.mol) file. There is one AMM per reactant–product pair. The input to the program was a list of reactions with associated reactants and products and their KEGG ID numbers (http://www.genome.jp/kegg/ligand.html). The program then extracted the appropriate molecule files (.mol format) and calculated the predicted AMM as well as a score indicating the quality of the match. The above procedure was applied for the reactions in the *E. coli* model described above, and the results checked manually based on known biochemistry. Approximately 80% of the auto-generated AMMs were found to be correct while the remaining were subsequently corrected manually. These corrected AMMs are summarized in compact atom transformation form, using letters to represent the carbon atoms, in the Supplementary material.
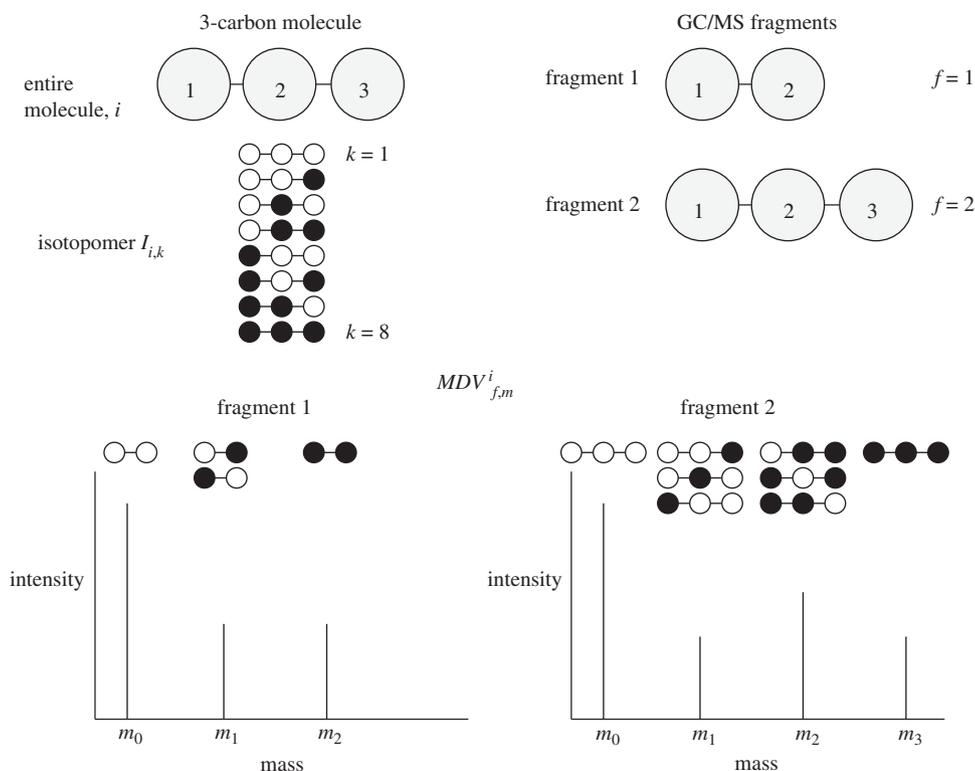
Fig. 1. Pictoral representation of the variable I_{ik} and the parameter MDV for a sample molecule. For a molecule, i, that contains three carbons there are 2^3, i.e., 8, different labeling patterns that make up the k members of the isotope fraction I_{ik}. The relative fraction of each is contained in the I_{ik}. During GC–MS, the molecule is derivitized and can be fragmented into different-sized species, f, that are then analyzed. The mass distribution vector, MDV, contains the information about the relative fractions of each fragment that contains m labeled carbons. Because only the total mass of each instance of a fragment is determined, the mass distribution vector is made by collecting various isotopomers by mass, as shown for two different fragments.

IMMs were then calculated directly from the AMMs using the algorithm introduced by Van Dien et al. (2003). IMMs indicate the possible product isotopomers that can be created from each reactant isotopomer (Schmidt et al., 1997). IMMs were used in the isotopomer balance equations to determine the IDVs for each metabolite. The numbering scheme shown in Fig. 1 is used for the isotopomers.

The final step of the initial analysis was to calculate the mass distribution vector (MDV) for all observed products from the mass spectrometry data (Wittmann and Heinzle, 1999). For each amino acid that was detected, elements of the IDV with the same number of labeled carbons were summed to yield an element of the MDV, as shown in Fig. 1. For amino acid fragments (for example, a common fragment is missing the carboxy-terminal carbon), the procedure was modified to include only the relevant carbon atoms. The process was facilitated by a Matlab program that makes use of matrix algebra (Van Dien et al., 2003; Wittmann and Heinzle, 1999).

2.3. Mathematical analysis of flux elucidation

The problem of calculating the fluxes and labeling patterns for a given set of GC–MS data is formulated as a nonlinear optimization problem (FluxCalc) that is given in Appendix A. This problem minimizes the sum of the variance-weighted differences between the experimental data and the calculated values for the MDVs of the measured amino acid fragments by solving isotope balances (see Fig. 2). (FluxCalc) is solved multiple times using CONOPT version 3 accessed within the GAMS modeling environment. The uniformly labeled glucose was modeled to be 99% isotopically pure, with the ^{12}C impurities being equally distributed among each of the carbons. Only isotope forms having five labeled carbons were considered for the impurities; each of the six possibilities was equally likely. The initial flux distributions are provided by solving (FluxInit) that generates a set of random feasible flux distributions; it is also briefly described in Appendix A, using CPLEX version 10, also accessed within GAMS. Given to the non-convex nature of (FluxCalc), the size of the resulting formulation and the experimental variability in the experimental MDV, we pursue the identification of as many local optima solutions as possible.

2.4. Probability distribution of least squares objective function

The iterative procedure described in the previous section identifies many different local optima with potentially highly varying flux distributions. Due to the variability and measurement imprecision of the mass spectrometry

$$\boxed{\sum_{j|S_{ij}>0}\left(S_{ij}\cdot v_j \cdot \prod_{i'|S_{i'j}<0}\left(\sum_{k'}IMM^j_{i'\to i,k'\to k}\cdot I_{i',k'}\right)\right)} \qquad \boxed{\sum_{j|S_{ij}<0}\left(S_{ij}\cdot v_j \cdot I_{ik}\right)}$$

A

D

$$S_{E,1}\cdot v_1 \cdot \left(\sum_{k'}IMM^1_{A\to E,k'\to k}\cdot I_{Ak'}\right)$$

v_1

v_3

$(S_{E,3}\cdot v_3)\cdot I_{Ek}$

E

I_{Ek}

v_2

v_4

B + C

F + G

$$S_{E,2}\cdot v_2 \cdot \left(\sum_{k'}IMM^2_{B\to E,k'\to k}\cdot I_{Bk'}\right)\cdot\left(\sum_{k'}IMM^2_{C\to E,k'\to k}\cdot I_{Ck'}\right)$$

$(S_{E,4}\cdot v_4)\cdot I_{Ek}$
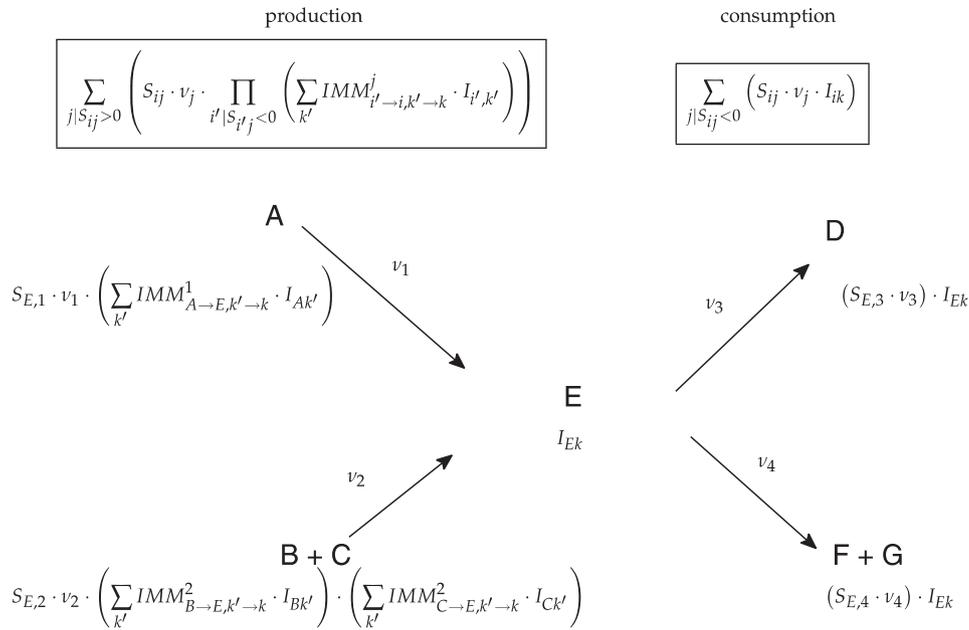
Fig. 2. The isotope labeling of species is governed by an isotopomer distribution vector balance and the isotope mapping matrix. Two reactions, j_1 and j_2, form the species E; the first is unimolecular and the second is bi-molecular. E is also consumed by two reactions j_3 and j_4; again, the first is unimolecular and the second is bi-molecular. The isotope labeling of E is given by I_{Ek}. Each carbon-containing compound has an associated isotope fraction for each labeling pattern, I_{ik}. The isotope labeling balance of the system is at the top, with the first term of the right-hand side of Eq. (3) on the left and the second term on the right. The product symbol $\prod$ is used for term by term multiplication. The isotope labeling, I_{ik}, and the isotope mapping matrix, $IMM^j_{i'\to i,k'\to k}$, that describes the contributions from the k' isotopes forms of the reactant i' on the k isotopes forms of product E are shown for the two production reactions. Note that the unimolecular reaction j_1 creates bilinear terms, and the bi-molecular reaction j_2 creates trilinear terms, as indicated. For the consumption terms, on the right, only the overall labeling patterns of E are considered.

measurements that yield $MDV^{i,\exp}_{f,m,r}$, the key question is how many (if any) of these flux distributions are statistically indistinguishable representations of the metabolic state of the system. To address this question, validation of the fit was performed using a χ^2-based goodness-of-fit test.

2.5. Identification of lower/upper bounds on elucidated fluxes given an objective value cutoff

Upon identifying the top scoring solution, a confidence level (i.e., 95%) may be imposed using the method from Antoniewicz et al. (2006) to determine an objective function cutoff. The flux ranges can be calculated by maximizing/minimizing each *net flux* and exchange flux separately subject to the constraints of formulation (FluxCalc) along with the following limit on the objective function: $z^{(l)}\leqslant z^{\text{cutoff}}$ to obtain the upper and lower bounds ($v^{\text{U,isotopomer}}_j$ and $v^{\text{L,isotopomer}}_j$, respectively). We refer to this formulation as (FluxRange). As in the case of (FluxCalc) formulation, (FluxRange) is non-convex. Therefore, it is solved iteratively from multiple initial guesses and only the best solutions are recorded. Fluxes whose maximization and minimization under (FluxRange) yield exactly the same value are referred to as *fully resolved*. Fluxes that yield solutions of the (FluxRange) formulation that are identical to the upper and lower bounds found by the relaxed problem for which only overall mass balance is

considered ($v^{\text{U,stoichiometry}}_j$ and $v^{\text{L,stoichiometry}}_j$, respectively) are referred to as *unresolved*, whereas fluxes that yield narrower ranges are referred to as *partially resolved*. These flux ranges were quantified by defining the variable d, the degree of resolution:

$$d_j = 1 - \frac{(v^{\text{U}}_j - v^{\text{L}}_j)^{\text{isotopomer}}}{(v^{\text{U}}_j - v^{\text{L}}_j)^{\text{stoichiometry}}}, \quad \forall j|(v^{\text{U}}_j - v^{\text{L}}_j)^{\text{stoichiometry}}\neq 0.$$

(1)

The higher the value of d is for a flux, the greater its resolution is. Thus, the degree of resolution is equal to zero for those fluxes that are unresolved, between zero and one for those that are partially resolved, and equal to one for those that are fully resolved.

3. Results and discussion

Three macroscopic measurements from the chemostat experiment were used to constrain the system. These measurements are comprised the initial glucose concentration minus the residual glucose concentration (6.1 ± 0.2 g/L), the cell density (1.6 ± 0.2 g/L), and the final amorphadiene concentration (0.077 ± 0.005 g/L) all of which were taken in replicate. The dilution rate was held constant at $0.19\,\text{h}^{-1}\pm0.00\,\text{h}^{-1}$. The glucose entering the system was assumed to be converted to biomass, amorphadiene, CO_2, or acetate. Although the fluxes towards biomass and

amorphadiene were fixed based on the experimental measurements, the employed metabolic model was free to appropriately partition the fluxes towards CO_2 and acetate while attempting to match the observed labeling patterns of the amino acids. The measured acetate flux (10.2–12.8 mM/h) was not used as an explicit constraint on the system, but rather was left as a final consistency check. All computational results are reported using a basis glucose uptake rate of 10 mM/h.

Using GC–MS, we observed the mass distributions (MDV) for fragments of 13 amino acids collected from two samples taken at different time points. Each derivatized sample was injected four times and the standard deviations for each sample and an overall standard deviation were calculated. In general, the overall standard deviations of the mass distributions among replicate measurements were less than the instrument error of 0.4%, which is consistent with previous estimates of instrument error (Wittmann et al., 2002). Next, the measurement error was estimated as the quadrature addition of the instrument error and the standard deviation (Taylor, 1990). After correcting for naturally occurring ^{13}C, we validated the internal consistency of the MDV distributions of the fragments for each amino acid.

The 32 fragments used for MFA are listed in Table 1. These fragments contain 162 mass entries giving $162 - 32 = 130$ independent mass measurements for use in the MFA. The mean values of the experimental mass measurements and corresponding estimates of the measurement error are given in Table 2. The best computationally derived MDV values (weighted residual sum of squares = 56.4) are also given for each fragment. Although the unweighted differences are quite small, we performed a goodness-of-fit analysis. The isotope model contains 130 free fluxes. Using direct experimental flux measurements, three internal fluxes (glucose uptake, amorphadiene production, and biomass production) and 22 exchange fluxes were set, giving a total of 105 independent flux variables for (FluxCalc) to determine. A χ^2 distribution with $130 - 105 = 25$ degrees of freedom yields a maximum allowable objective value of 37.7, at the 95% confidence level. Thus the model fit is statistically not acceptable $(P(\chi^2[25] < 56.4) = 0.9997)$ implying significant errors in the measurements. As seen in the table, most of the error lies in the serine and lysine fragments. A normal probability plot of the weighted residuals was linear.

Given that no single solution emerged as a clear-cut candidate for the correct flux distribution, we set out to determine if our model contained degrees of freedom that had no bearing on the carbon labeling and thus made the goodness-of-fit test too conservative. First, we tested whether the observed SSE are indicative of true elucidation of metabolic fluxes and not an artifact due to the numerical scaling of the problem. To do this we generated random but feasible flux distributions and then calculated the corresponding SSE and contrasted them against the ones obtained for the solutions obtained with (FluxCalc). As

Table 1
Fragments of the amino acids experimentally measured by GC–MS and simulated using the IMM models

Amino acid[a]	Monitored ions	Amino acid carbon atoms[b]	Fragment
Val-159	186–192	1-2-4-5	$M-C_7H_{15}O_2Si$
Val-85	260–266	1-2-4-5	$M-C_5H_9O$
Val-57	288–295	1-2-3-4-5	$M-C_4H_9$
Gly-85	218–221	2	$M-C_5H_9O$
Gly-57	246–249	1-2	$M-C_4H_9$
Ala-85	232–235	1-3	$M-C_5H_9O$
Ala-57	260–264	1-2-3	$M-C_4H_9$
Glu-159 R	330–335	1-2-4-5	$M-C_7H_{15}O_2Si$
Glu-85 R	404–409	1-2-4-5	$M-C_5H_9O$
Glu-57 R	432–438	1-2-3-4-5	$M-C_4H_9$
Asp-159 R	316–320	1-2-4	$M-C_7H_{15}O_2Si$
Asp-57 R	418–423	1-2-3-4	$M-C_4H_9$
Asp-15 R	460–465	1-2-3-4	$M-CH_3$
Met-159	218–223	1-3-4-5	$M-C_7H_{15}O_2Si$
Met-85	292–297	1-3-4-5	$M-C_5H_9O$
Met-57	320–325	1-2-3-4-5	$M-C_4H_9$
Leu-159	200–205	1-2-4-5-6	$M-C_7H_{15}O_2Si$
Leu-85	274–280	1-2-4-5-6	$M-C_5H_9O$
Ile-159	200–205	1-2-4-5-6	$M-C_7H_{15}O_2Si$
Ile-85	274–280	1-2-4-5-6	$M-C_5H_9O$
Pro-159	184–189	1-3-4-5	$M-C_7H_{15}O_2Si$
Ser-159 R	288–291	1-3	$M-C_7H_{15}O_2Si$
Ser-85 R	362–365	1-3	$M-C_5H_9O$
Ser-57 R	390–394	1-2-3	$M-C_4H_9$
Thr-85 R	376–379	1-2-4	$M-C_5H_9O$
Thr-57	290–294	1-2-3-4	$M-C_4H_9$
Thr-57 R	404–408	1-2-3-4	$M-C_4H_9$
Phe-159	234–243	1-2-3-4-5-6-7-9	$M-C_7H_{15}O_2Si$
Phe-85	308–317	1-2-3-4-5-6-7-9	$M-C_5H_9O$
Phe-57	336–345	1-2-3-4-5-6-7-8-9	$M-C_4H_9$
Lys-159 R	329–335	1-2-4-5-6	$M-C_7H_{15}O_2Si$

[a] The number after the amino acid is the size of the fragment removed from the molecule with both N and C groups derivatized. An R following a fragment denotes that the R-group is also derivatized.

[b] The carbon numbering follows that in the .mol file from the KEGG database (Kanehisa et al., 2006). For those fragments that have a removed C, it corresponds to that of the carboxyl.

seen in Fig. 3A, the root mean square error for the inferred fluxes was always much smaller than those for randomly generated feasible fluxes. This lack of overlap suggested that the obtained low SSE values were not caused by the numerical scaling of the problem but instead were driven by correctly apportioning fluxes in accordance with the observed experimental data and suggested that the goodness-of-fit test might have been too conservative.

Thus, the next step was to characterize the tightness of elucidation (i.e., resolvability) of each flux in the network. First, we calculated the variability in all fluxes subject to only network stoichiometry, the three fixed macroscopic measurements (i.e., glucose, biomass, and amorphadiene), and the fixed exchange fluxes. This involved solving the (FluxRange) formulation *without* imposing the isotopomer balance constraints. The flux ranges for central metabolism (see Fig. 4A) are provided in the third and fourth columns of Table 3 while the ranges for the entire network are

Table 2
Experimentally measured and simulated mass distributions (mol %) of amino acid fragments

Fragment		M + 0	M + 1	M + 2	M + 3	M + 4	M + 5	M + 6	M + 7	M + 8	M + 9
Val-159	Meas[a]	61.9 ± 0.6	5.3 ± 0.6	27.6 ± 0.5	2.1 ± 0.4	3.1 ± 0.4					
	Sim[b]	62.2	5.6	27.9	1.2	3.1					
Val-85	Meas	61.7 ± 0.5	6.0 ± 0.6	27.8 ± 0.5	1.4 ± 0.4	3.1 ± 0.4					
	Sim	62.2	5.6	27.9	1.2	3.1					
Val-57	Meas	59.6 ± 0.6	7.5 ± 0.5	16.4 ± 0.7	12.9 ± 0.6	1.1 ± 0.4	2.6 ± 0.4				
	Sim	59.5	7.5	16.1	13.1	1.1	2.7				
Gly-85	Meas	80.9 ± 0.5	19.2 ± 0.5								
	Sim	80.6	19.4								
Gly-57	Meas	77.8 ± 0.5	5.6 ± 0.5	16.7 ± 0.4							
	Sim	77.9	5.6	16.5							
Ala-85	Meas	78.7 ± 0.5	3.6 ± 0.6	17.8 ± 0.5							
	Sim	78.9	3.5	17.6							
Ala-57	Meas	75.0 ± 0.8	6.2 ± 0.8	3.7 ± 0.8	15.1 ± 0.7						
	Sim	75.5	6.1	3.4	15.1						
Glu-159 R	Meas	58.1 ± 0.7	13.5 ± 1.1	23.3 ± 0.7	2.9 ± 0.5	2.2 ± 0.4					
	Sim	58.2	13.4	23.4	2.8	2.2					
Glu-85 R	Meas	58.1 ± 0.8	13.6 ± 1.2	23.4 ± 0.7	2.7 ± 0.5	2.1 ± 0.4					
	Sim	58.2	13.4	23.4	2.8	2.2					
Glu-57 R	Meas	51.9 ± 0.7	15.3 ± 0.8	21.4 ± 0.4	8.1 ± 0.4	2.3 ± 0.4	1.1 ± 0.4				
	Sim	52.1	15.1	21.5	7.9	2.3	1.0				
Asp-159 R	Meas	65.7 ± 0.6	16.1 ± 0.7	11.9 ± 0.4	6.2 ± 0.5						
	Sim	66.1	16.2	11.8	5.9						
Asp-57 R	Meas	60.7 ± 0.5	16.5 ± 0.5	10.6 ± 0.6	9.6 ± 0.5	2.7 ± 0.4					
	Sim	60.9	16.5	10.4	9.7	2.5					
Asp-15 R	Meas	60.8 ± 0.6	16.3 ± 0.5	10.6 ± 0.6	9.6 ± 0.5	2.8 ± 0.5					
	Sim	60.9	16.5	10.4	9.7	2.5					
Met-159	Meas	53.6 ± 0.5	25.8 ± 0.5	12.7 ± 0.5	6.9 ± 0.5	1.1 ± 0.4					
	Sim	53.3	25.9	12.6	7.0	1.1					
Met-85	Meas	53.2 ± 1.0	25.8 ± 0.6	12.8 ± 0.5	7.0 ± 0.6	1.2 ± 0.7					
	Sim	53.3	25.9	12.6	7.0	1.1					
Met-57	Meas	49.2 ± 0.6	24.9 ± 0.6	11.6 ± 0.5	9.8 ± 0.5	4.0 ± 0.5	0.5 ± 0.4				
	Sim	49.1	25.1	11.6	9.8	3.9	0.5				
Leu-159	Meas	50.5 ± 0.5	15.9 ± 0.6	23.9 ± 0.5	6.3 ± 0.4	2.9 ± 0.4	0.6 ± 0.4				
	Sim	50.1	16.6	23.6	6.4	2.7	0.6				
Leu-85	Meas	50.3 ± 0.5	16.4 ± 0.5	23.7 ± 0.5	6.3 ± 0.4	2.8 ± 0.4	0.6 ± 0.4				
	Sim	50.1	16.6	23.6	6.4	2.7	0.6				
Ile-159	Meas	52.0 ± 0.5	15.1 ± 0.7	21.7 ± 0.4	7.9 ± 0.5	2.3 ± 0.4	1.0 ± 0.4				
	Sim	52.1	15.3	21.5	7.9	2.3	1.0				
Ile-85	Meas	51.5 ± 0.5	15.5 ± 0.7	21.7 ± 0.4	8.0 ± 0.4	2.3 ± 0.4	1.0 ± 0.4				
	Sim	52.1	15.3	21.5	7.9	2.3	1.0				
Pro-159	Meas	58.8 ± 0.7	13.3 ± 1.1	23.1 ± 0.7	2.6 ± 0.5	2.1 ± 0.5					
	Sim	58.2	13.4	23.4	2.8	2.2					
Ser-159 R	Meas	77.5 ± 0.7	6.2 ± 0.8	16.3 ± 0.7							
	Sim	78.3	4.7	17.0							
Ser-85 R	Meas	77.4 ± 0.8	6.2 ± 0.9	16.4 ± 0.7							
	Sim	78.3	4.7	17.0							
Ser-57 R	Meas	75.0 ± 0.5	6.1 ± 0.5	3.7 ± 0.5	15.2 ± 0.5						
	Sim	75.7	5.7	3.4	15.2						
Thr-85 R	Meas	66.0 ± 0.5	16.4 ± 0.7	11.7 ± 0.5	5.9 ± 0.5						
	Sim	66.0	16.4	11.8	5.8						
Thr-57	Meas	61.0 ± 0.5	16.5 ± 0.6	10.3 ± 0.5	9.5 ± 0.5	2.7 ± 0.4					
	Sim	60.8	16.6	10.5	9.6	2.5					
Thr-57 R	Meas	60.6 ± 0.5	16.7 ± 0.5	10.5 ± 0.6	9.6 ± 0.5	2.6 ± 0.4					
	Sim	60.8	16.6	10.5	9.6	2.5					
Phe-159	Meas	43.6 ± 0.7	10.7 ± 0.7	20.9 ± 0.7	9.5 ± 0.7	8.8 ± 0.7	3.1 ± 0.7	2.7 ± 0.7	0.4 ± 0.7	0.3 ± 0.7	
	Sim	43.4	11.3	19.9	9.3	9.1	3.4	2.9	0.4	0.3	
Phe-85	Meas	43.4 ± 0.7	10.8 ± 0.7	21.0 ± 0.7	9.4 ± 0.7	8.7 ± 0.7	3.1 ± 0.7	2.6 ± 0.7	0.3 ± 0.7	0.6 ± 0.7	
	Sim	43.4	11.3	19.9	9.3	9.1	3.4	2.9	0.4	0.3	
Phe-57	Meas	42.4 ± 0.5	11.3 ± 0.4	12.3 ± 0.4	16.7 ± 0.4	8.5 ± 0.4	4.1 ± 0.4	2.9 ± 0.4	1.3 ± 0.4	0.3 ± 0.4	0.3 ± 0.4
	Sim	42.0	11.6	12.1	16.1	9.0	4.2	3.0	1.4	0.4	0.3
Lys-159 R	Meas	51.9 ± 0.8	15.5 ± 1.1	21.6 ± 0.7	7.9 ± 0.7	2.2 ± 0.7	1.1 ± 0.7				
	Sim	50.1	17.8	21.8	7.2	2.3	0.7				

[a]Meas are the values measured experimentally, reported as mean ± experimental error. The mass distributions are corrected for naturally occuring isotopes.

[b]Sim are the values simulated by the isotope model, reported for the lowest objective (56.4).
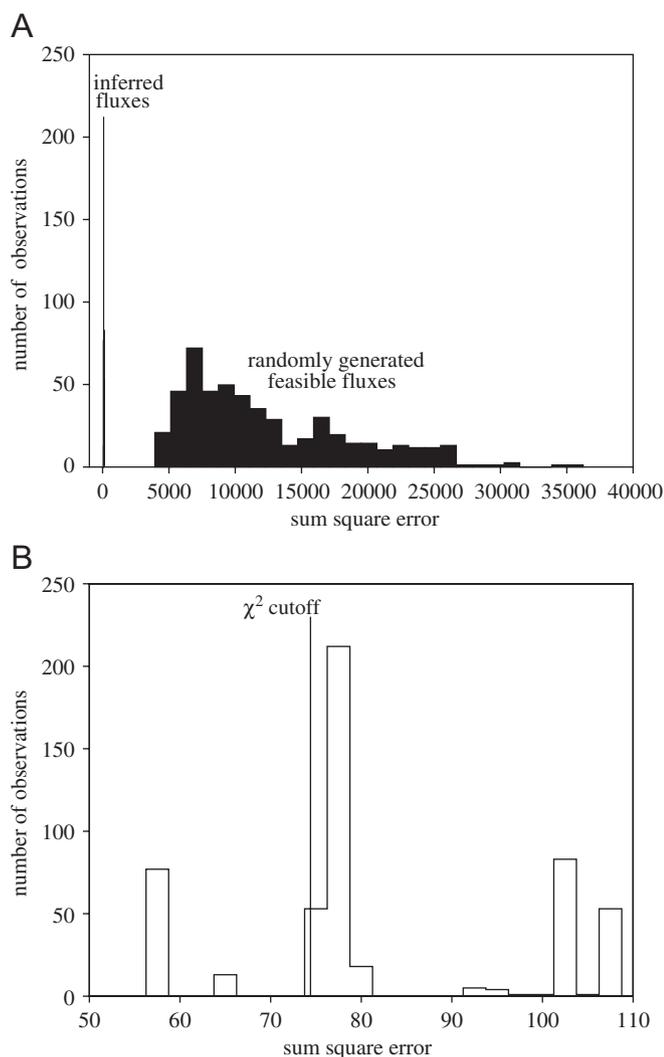
Fig. 3. Inferred fluxes and propagation of error. Panel (A) shows that the objective values determined for inferred fluxes do not overlap those from randomly generated feasible fluxes, suggesting that the fit of the data requires correctly apportioning fluxes in accordance with the observed experimental data. The same total of solutions occur for each histogram. Panel (B) shows that these solutions cluster tightly for the most part and that many of solutions could be statically acceptable once experimental error is propagated onto the objective values. Those to the left of the maximal χ^2 value (solid vertical line) are retained for use as inputs into (FluxRange) whereas those to the right are discarded.

calculated flux ranges for central metabolism under this scenario are provided in the fifth and sixth columns of Table 3, while the remaining flux ranges are included as supplementary material. The identified classifications of the fluxes based on their resolution status are summarized in Table 4. Of the 145 net fluxes that were not fixed due to stoichiometry, 131 were partially resolved with bound ranges spanning magnitudes from 0.2% to 98% of the flux range allowable by the stoichiometry alone. These ranges can be characterized by defining the degree of resolution, as shown in Fig. 5. Examples of partially resolved fluxes with significantly narrower (and more biologically realistic) ranges after the imposition of the isotopomer balancing constraints included the PPP reactions (RPE, RPI, TKT1, TKT2, and TAL), the TCA cycle reactions (ACONT, ICDHy, AKGD, and SUCOAS), and the anaplerotic reactions (ICL, MALS, PPC, and PPCK).

Closer investigation of columns five and six of Table 4 reveals that additional manual intervention is needed when resolving fluxes based on a large model, particularly in cases where there exist pathways with redundant labeling patterns. For example, the labeling pattern of pyruvate is unaffected regardless of whether glucose is metabolized via EMP glycolysis (GAPD, PGK, PGM, ENO, and PYK) or via the methylglyoxal pathway (MGSA, LGTHL, GLYOX, and LDH_D2). In fact, one feasible, though obviously unrealistic, flux distribution involved a high gluconeogenic flux (~20 mM/h) coupled to an even higher flux (~40 mM/h) through the methylglyoxal pathway. In cases such as these, care must be taken to prune the model using biological intuition without compromising its ability to match the experimental labeling observations. Although this involves making assumptions about the current metabolic system, it provides a systematic way of cataloging these assumptions that can be later verified via expression measurements, enzyme assays, or additional labeling studies with different labeled substrates. Candidate reactions for removal must have (i) flux ranges that start at zero and (ii) no effect on the objective value of any (FluxCalc) solutions upon removal. Thus by using information from the large-scale model, reactions can be identified whose removal does not affect the quality of agreement with the data. This removal of 33 fluxes in effectively redundant pathways from the model (see Table 5) selected using biological intuition and *the requirement of not worsening the objective* led to the derivation of a 'reduced' model. We note that with the removal of these reactions that have no bearing on the effective carbon labeling, given the substrate labeling, that the model contains fewer degrees of freedom. There are 99 free fluxes in the reduced model, giving a total of 74 independent flux variables for (FluxCalc) to determine. An updated goodness-of-fit test (having $130 - 74 = 56$ degrees of freedom) is passed, with a maximum allowable objective of 74.4. Thus, we can accept the reduced model as statically valid ($P(\chi^2[56] < 56.4) = 0.5401$).

available as supplementary material. Table 4 reveals that 125 fluxes of the total 270 net fluxes are fixed by stoichiometry alone. However, as expected, nearly all central metabolic fluxes span wide ranges of values motivating the need for ^{13}C-based MFA.

We next reran the (FluxRange) formulation after including the isotopomer balancing constraints to determine (i) which fluxes (if any) were fully resolved; (ii) how many of them were only partially resolved; and (iii) if there were any remaining fluxes in the model that are completely unaffected by the isotopomer data. A cut-off objective value of 66.0 was determined for (FluxRange) as the 95% percentile using an *F*-test statistic (Kleijn et al., 2006). The

As shown in Fig. 3B, most of the 521 inferred optima had objective values in the range 56–107 with most (all but 100) outside the significance level of 74.4. The lowest objective value obtained was 56.4 which occurred six times (for comparison, 56.5 occurred 20 times). A total of 49 different optimum values were observed when rounding to 0.1; 12 of these were below the statistical cutoff. Given the plethora of identified locally minimum solutions with very similar objective function values acceptable by χ^2 criteria *but with quite different flux distributions*, the next step was to deduce flux ranges in equivalent solutions given the experimental error in the measurements. This result further validated the decision to identify as many statistically equivalently solutions as possible rather than homing in on a single

solution that may miss many of the attributes of the physiologically relevant one.

Reapplication of the (FluxRange) formulation, using an updated cutoff of 60.6 owing to the increased degrees of freedom, lead to far more biologically realistic ranges for the glycolytic fluxes GAPD, PGK, PGM, and ENO as shown in columns seven and eight of Table 3. Note that this reduced model also depends upon the specific experimental labeling conditions; that is, a flux that is redundant for one experimental condition may not be generically so. For instance the ED pathway was set to zero because the current experiment was performed with a mixture of uniformly and un-labeled glucose meaning that it cannot distinguish between ED and EMP glycolysis. This assumption led to

Fig. 4. Central metabolism flux maps. The flux map of central metabolism for the full model (panel A) shows the metabolite names (sans-serif typeface) and reaction names (smaller, serif typeface). Irreversible reactions have single, filled, arrows. Reversible reactions have two arrows: the forward direction is filled and the backward direction is unfilled. In comparison, most of the reactions are irreversible for the smaller basic model (panel B).

Fig. 4. (*Continued*)

a tighter flux range for the oxidative branch of the PPP (G6PDHy, PGL, and PGDH). Improvement in the overall degree of resolution was also observed for this reduced model as seen in Fig. 5. We again observed that some fluxes that had consistent values for the optimum solutions (e.g., ICL had a value of zero) did in fact have a range of values they could take when accounting for measurement noise, thus emphasizing the importance of range calculations. Of especial note is the acetate transport flux. Although the allowable stoichiometric range remained the same as for the full model (from 0 to 12.3 mM/h), the allowable range for the isotopomer model shrank to a range from 8.4 to

11.5 mM/h. This narrow range was in good agreement with the experimentally observed values (10.2–12.8 mM/h), and thus served as a validation of the model.

Finally, we tested the value added by using a relatively large-scale model of metabolism in contrast to existing metabolic representations used in similar efforts (Christensen and Nielsen, 1999; Fischer et al., 2004; Ghosh et al., 2005). Instead of choosing one such representation, we decided to simply prune additional reactions from our model to make it resemble previously employed metabolic representations. In contrast with the reduced model described above, the construction of the 'basic' model

Table 3
Flux ranges

Abbreviation[a]	Subsystem	Relaxed model[b,c]		Full model		Reduced model		Basic model	
		Min	Max	Min	Max	Min	Max	Min	Max
ICL	Anaplerotic reactions	0.0	12.3	0.0	1.0	0.0	0.4	0.0	0.0
MALS	Anaplerotic reactions	0.0	12.3	0.0	1.0	0.0	0.4	0.0	0.0
ME1x	Anaplerotic reactions	0.0	75.9	0.0	4.0	0.0	3.7	0.0	0.0
PPA	Anaplerotic reactions	0.8	129.4	0.8	107.2	0.8	80.1	0.8	83.1
PPC	Anaplerotic reactions	0.0	130.1	1.1	6.8	1.1	5.5	1.4	5.5
PPCK	Anaplerotic reactions	0.0	128.7	0.0	5.4	0.0	4.0	0.0	4.0
ACONT	Citrate cycle (TCA)	0.5	12.8	0.8	4.3	0.9	3.2	0.9	3.1
AKGD	Citrate cycle (TCA)	0.0	12.3	0.0	2.7	0.0	2.7	0.1	2.6
CITL	Citrate cycle (TCA)	0.0	128.7	0.0	128.6	0.0	128.7	0.0	0.0
CS	Citrate cycle (TCA)	0.5	141.0	0.8	109.4	1.4	82.4	0.9	3.1
FRD2	Citrate cycle (TCA)	0.0	68.7	0.0	58.6	0.0	0.0	0.0	0.0
FRD3	Citrate cycle (TCA)	0.0	68.7	0.0	58.6	4.0	41.5	0.0	0.0
FUM	Citrate cycle (TCA)	0.4	281.5	1.0	230.0	0.7	14.3	0.8	5.5
ICDHy	Citrate cycle (TCA)	0.5	12.8	1.1	3.2	0.5	3.2	0.9	3.1
MDH	Citrate cycle (TCA)	−59.1	281.5	−1.4	230.0	−1.1	14.3	0.8	5.5
SUCD1i	Citrate cycle (TCA)	0.0	77.3	2.6	60.5	5.1	43.8	0.4	2.6
SUCOAS	Citrate cycle (TCA)	−12.3	18.4	−4.0	0.1	−2.6	11.6	−2.5	0.0
ENO	Glycolysis/gluconeogenesis	−44.3	18.8	−40.3	18.4	16.5	18.4	16.6	18.8
FBA	Glycolysis/gluconeogenesis	−0.3	9.6	0.2	9.6	7.7	9.6	7.9	9.6
FBP	Glycolysis/gluconeogenesis	0.0	128.7	0.0	107.3	0.0	79.8	0.0	0.0
G1PP	Glycolysis/gluconeogenesis	0.0	128.7	0.0	107.3	0.0	0.0	0.0	0.0
GAPD	Glycolysis/gluconeogenesis	−32.3	19.0	−20.1	19.0	17.2	19.0	17.3	19.0
GLCP	Glycolysis/gluconeogenesis	0.0	128.7	0.0	107.3	0.0	0.0	0.0	0.0
GLCS1	Glycolysis/gluconeogenesis	0.1	128.7	0.1	107.4	0.1	0.1	0.1	0.1
GLGC	Glycolysis/gluconeogenesis	0.1	128.7	0.1	107.4	0.1	0.1	0.1	0.1
HEX1	Glycolysis/gluconeogenesis	0.0	136.2	0.0	115.9	0.0	10.0	0.0	10.0
PDH	Glycolysis/gluconeogenesis	0.0	27.1	0.0	16.4	0.0	16.4	0.0	15.3
PFK	Glycolysis/gluconeogenesis	0.0	138.3	0.0	116.4	7.8	88.3	7.9	9.6
PGI	Glycolysis/gluconeogenesis	0.0	9.9	0.0	9.9	4.4	9.9	4.7	9.9
PGK	Glycolysis/gluconeogenesis	−32.3	19.0	−20.1	19.0	17.2	19.0	17.3	19.0
PGM	Glycolysis/gluconeogenesis	−44.3	18.8	−42.3	18.4	16.5	18.4	16.6	18.8

PPS	Glycolysis/gluconeogenesis	0.0	128.7	0.0	108.4	0.0	79.8	0.0	0.0
PYK	Glycolysis/gluconeogenesis	0.0	144.2	0.0	112.3	3.5	92.7	4.6	16.6
TPI	Glycolysis/gluconeogenesis	−41.6	9.6	−29.4	9.6	7.8	9.6	7.9	9.6
ATPS4r	Oxidative phosphorylation	−24.2	108.4	−24.2	106.6	−24.2	60.3	−24.2	63.5
CYTBO3	Oxidative phosphorylation	27.9	77.2	31.1	77.2	31.1	43.7	31.1	44.2
FDH2	Oxidative phosphorylation	0.0	27.1	0.0	27.1	0.0	15.3	0.0	15.3
LDH_D2	Oxidative phosphorylation	0.0	68.7	0.0	54.5	0.0	0.0	0.0	0.0
NADH10	Oxidative phosphorylation	0.0	68.7	0.0	58.8	0.0	0.0	0.0	0.0
NADH12	Oxidative phosphorylation	0.0	68.7	0.0	54.8	0.0	35.0	0.0	42.1
NADH6	Oxidative phosphorylation	0.0	68.7	0.0	53.7	0.0	35.0	0.0	42.1
NADH7	Oxidative phosphorylation	0.0	68.7	0.0	58.8	0.0	0.0	0.0	0.0
NADH8	Oxidative phosphorylation	0.0	68.7	0.0	58.8	3.0	41.5	0.0	0.0
NADH9	Oxidative phosphorylation	0.0	68.7	0.0	58.3	0.0	0.0	0.0	0.0
POX	Oxidative phosphorylation	0.0	27.1	0.0	16.1	0.0	15.3	0.0	0.0
SUCD4	Oxidative phosphorylation	−0.1	77.2	2.3	60.4	5.0	43.7	0.3	44.2
THD2	Oxidative phosphorylation	0.0	257.3	0.0	214.8	0.0	159.7	0.0	164.7
THD5	Oxidative phosphorylation	0.0	272.6	0.0	229.6	0.0	163.9	0.0	169.9
TRDR	Oxidative phosphorylation	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
EDA	Pentose-phosphate cycle	0.0	9.9	0.0	9.9	0.0	0.0	0.0	0.0
G6PDHy	Pentose-phosphate cycle	0.0	9.9	0.0	9.9	0.0	5.5	0.0	5.2
PGDH	Pentose-phosphate cycle	0.0	9.9	0.0	5.5	0.0	5.5	0.0	5.2
PGDHy	Pentose-phosphate cycle	0.0	9.9	0.0	9.9	0.0	0.0	0.0	0.0
PGL	Pentose-phosphate cycle	0.0	9.9	0.0	9.9	0.0	5.5	0.0	5.2
RPE	Pentose-phosphate cycle	−0.2	6.4	−0.2	4.5	−0.2	3.4	−0.2	3.2
RPI	Pentose-phosphate cycle	−3.5	−0.2	−2.6	−0.2	−2.1	−0.2	−1.9	−0.2
TAL	Pentose-phosphate cycle	−0.1	3.2	−0.1	2.3	−0.1	1.8	−0.1	1.7
TKT1	Pentose-phosphate cycle	−0.1	3.2	−0.1	2.3	−0.1	1.8	−0.1	1.7
TKT2	Pentose-phosphate cycle	−0.2	3.1	−0.2	1.7	−0.2	1.6	−0.2	1.6
ACt6	Transport, extracellular	−12.3	0.0	−11.6	0.0	−11.5	−8.4	−11.5	0.0
GLCpts	Transport, extracellular	0.0	10.0	0.0	10.0	0.0	10.0	0.0	10.0
GLCt2	Transport, extracellular	0.0	10.0	0.0	10.0	0.0	10.0	0.0	10.0
AMDNt	Amorphadiene	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
BIOMASS	Biomass	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5

[a]All flux values have units of mM/h and are relative to a glucose uptake rate of 10 mM/h.

[b]The relaxed model only uses stoichiometric data and not the ^{13}C labeling data and serves as the allowable bounds for the fluxes. Shown here is the relaxed model of the full model.

[c]Flux abbreviations correspond to the names in Reed et al. (2003) and are shown in Fig. 4.

Table 4
Flux resolution classification

	Number of fluxes		
	Full	Reduced	Basic
Set by experimental conditions	3	3	3
Pruned by experimental conditions/reductions	22	51	61
Stoichiometrically fixed	100	94	96
Fully resolved	0	3	0
Partially resolved	131	110	88
Unresolved	14	9	22

Fig. 5. Flux degree of resolution. The proportion of fluxes with improved (higher) degrees of resolution considerably increases for the reduced model (open rectangles), as compared with the full model (closed rectangles), once ambiguities in the flux elucidations are identified and the network is selectively pruned with biological insight.

was not based on using information from the elucidation in the large model to decide which reactions can be removed. A pictorial representation of the basic model is shown in Fig. 4B. Specifically, we further pruned 26 transformations (see Table 4) consisting primarily of reactions typically assumed inactive for aerobic growth on glucose such as fumarate reductase (FRD) and phosphoenolpyruvate synthase (PPS) and also anapleurotic reactions such as malic enzyme (ME1x) and the glyoxylate shunt (ICL). Furthermore, many reactions were treated as irreversible as was assumed in previous studies. In doing so, the residual sum of squares for the measured labeling patterns of the amino acids worsened from 56.4 to 559 in the best case, indicating clearly that significant information is lost when neglecting potentially active pathways present in the large-scale model. Notably, all solutions to the (FluxCalc) problem were well outside of the goodness-of-fit cut-off objective value used in the previous cenarios, and the model was also statistically unacceptable ($P(\chi^2[80] < 558) = 1.0$). A new F-test cut-off was calculated and implemented for (FluxRange) as before. Using the

Table 5
Fluxes manually set or pruned

Abbreviation[a]	Value	Model
EX_glc	−10	All
BIOMASS_Ec	0.46	All
EX_amdn	0.1114	All
EX_ala-L, EX_arg-L, EX_asn-L, EX_asp-L, EX_cys-L, EX_etoh, EX_for, EX_gln-L, EX_glu-L, EX_gly, EX_ile-L, EX_lac-L, EX_leu-L, EX_lys-L, EX_met-L, EX_phe-L, EX_pro-L, EX_ser-L, EX_succ, EX_thr-L, EX_tyr-L, EX_val-L	0	All
MGSA, LGTHL, GLYOX, LDH_D2	0	Reduced, basic
PGDHY, EDA	0	Reduced, basic
FRD2	0	Reduced, basic
ADHEr_b, ADHEr_f	0	Reduced, basic
LDH_D_b, LDH_D_f	0	Reduced, basic
G1PP	0	Reduced, basic
METAT, ADMDC, SPMS, MTAN, MTRK, MTRI, MDRPD, DKMPPD, DKMPD2, P5CR, UNK3	0	Reduced, basic
GLCP	0	Reduced, basic
ASPT	0	Reduced, basic
SERD_L	0	Reduced, basic
TRPAS1	0	Reduced, basic
FTHFD	0	Reduced, basic
NADH7, NADH9, NADH10	0	Reduced, basic
PGMT_f	0	Reduced, basic
VALTA_f	0	Reduced, basic
ME1x	0	Basic
ICL, MALS	0	Basic
POX	0	Basic
FBP	0	Basic
PPS	0	Basic
CITL	0	Basic
NADH8, FRD3	0	Basic
THRA_f, THRA_b	0	Basic
FBA_b, GAPD_b, PGK_b, PGM_b, ENO_b, SUCOAS_f, ICDHy_b, ACONT_b, G6PDHy_b, ASPTA1_f, PGI_b	0	Basic

[a]Flux abbreviations correspond to the names in Reed et al. (2003). EX_amdn is the amorphadiene exchange flux. Forward and backwards directions of reversible fluxes are denoted by appending_f or _b, respectively.

basic model and new objective value cut-off, many partially resolved fluxes had tighter ranges than for the complete model as shown in columns nine and ten of Table 3. Although this may seem to be a desirable attribute, in reality it means that by excluding some reactions the

smaller model cannot discern certain flux distributions that lead to wider ranges in the elucidated fluxes and thus loss of potentially useful information (Table 4). The impact of removing each reaction individually on the objective was also examined. Removal of FRD had the largest impact, resulting in a 24% increase in objective value, but removal of ME1x and the assumption of irreversibility for PGM and ENO also caused 9% increases, by not allowing changes in the labeling of PEP by the TCA cycle to affect the labeling of the 3PG amino acids. Other reactions had smaller effects individually.

4. Summary

In this paper we introduced a large-scale isotopomer mapping model accounting for 238 reactions, 350 fluxes and 184 metabolites which is at least fourfold larger than existing mapping models. This model is available as supplementary material. The novel aspect of this model is the simultaneous presence of explicit cofactor balancing and biomass drain used in conjunction with isotope data. While others have previously introduced models that included cofactor balances or isotope data (van Winden et al., 2003) their simultaneous use was never exploited for flux elucidation. The construction procedure of the isotopomer mapping model by reducing a genome-scale one enabled the seamless integration of MFA results with FBA analyses. The model was adapted for the amorphodiene producing strain (Martin et al., 2003) where fragments from 13 amino acids were analyzed using GC–MS. We found that many (i.e., tens) of different local minima solutions fit the experimental data equally well. This motivated the use of a goodness-of-fit test to determine significance. We found that many of the local minima were statistically indistinguishable from one another implying that given the available data significant ambiguity remained in flux elucidation. This ambiguity was subsequently quantified by identifying lower and upper bounds on all the fluxes given a required level of fit with the experimental data. Using these results as a starting point we subsequently removed unneeded, unobservable, or effectively redundant pathways (from the perspective given the experimental data) from the model by making sure that their removal does not worsen the fit. Subsequently, with fewer equivalent routes, the model was found to adequately describe the data. The advantage of starting with a more complete model and subsequently pruning inactive pathways was quantitatively demonstrated by simply removing pathways typically absent in existing isotopomer mapping models. This led to a statistically significant worse fit and allowed us to pinpoint the reactions whose removal was not warranted by the available data.

It is important to note that the methods described here could be extended to take advantage of the EMU framework (Antoniewicz et al., 2007a) for the rapid evaluation of the labeling patterns using multiple isotopic tracers by coupling it to the optimization procedure described herein, similar to the very recent work on a 1,3-propanediol system (Antoniewicz et al., 2007b). Also, the developed isotopomer model is not restricted to isotope data generated from GC–MS. For example, NMR data could also be incorporated by mapping the I_{ik} predictions onto the NMR data in an analogous way to that used to link the I_{ik} with the MDV values obtained from the GC–MS data. The NMR data may include the fractional abundance of ^{13}C-atoms at each position or the relative amount of ^{13}C–^{13}C and ^{13}C–^{12}C carbons through COSY NMR. Furthermore, the I_{ik} predictions could also be linked to positional isotopomer data generated by methods such as Fourier transform-ion cyclotron resonance mass spectrometry (FT-ICR MS) (Sleno et al., 2005) which would enable even more accurate metabolic flux elucidations by providing more specific flux labeling information.

In the present study, the inability to uniquely resolve fluxes was a consequence of the limited set of derivatized metabolites (i.e., only 13 metabolites), the limited set of measured fragments, the choice of labeled substrate, and the multiple routes with identical labeling patterns afforded by the large-scale mapping model. While not explicitly covered in this paper the proposed framework can be used to systematically identify how many and which additional metabolites to measure and what substrate carbon labeling patterns to use to uniquely recapitulate all fluxes in the network. For example, the inability to tightly resolve the branch point between the PPP and glycolysis is in this experiment was due to the use of uniformly labeled glucose. Instead, using glucose labeled at position 1, would have enabled the identification of the PPP/glycolysis branch ratio. Note that the isotopomer mapping model can be used to evaluate the changes in fluxes throughout the network upon addition of exogenous genes or deletions made during the engineering of pathways by simply appending or removing the corresponding functionalities from the model. Finally our study demonstrated that even if a smaller mapping model is ultimately used, it is important to start with a more complete reconstruction to base the decisions about which reactions can be eliminated without compromising the quality of agreement with the data.

Acknowledgments

Appendix A

In the context of metabolic flux analysis (MFA) the description of a metabolic network requires the definition of the following sets, variables, and parameters.

Sets:

$I = \{i\}$ set of metabolites
$J = \{j\}$ set of reactions

$J^R = \{j\}$　　　　　set of reversible reactions
$I^F = \{i\}$　　　　　Metabolites present in growth medium
$I^E = \{i\}$　　　　　Metabolites that can cross cell boundaries

Parameters:

S_{ij}　　　　　　　　　stoichiometric matrix

Variables:

v_j　　　　　　　　　Fluxes
b_i　　　　　　　　　exchange fluxes

The set I contains all of the metabolites present while set J enumerates all reactions composing the metabolic network. The set I^F is all of the metabolites that are contained in the growth medium, and set I^E contains all metabolites that can cross cellular boundaries. S_{ij} is the stoichiometric coefficient of metabolite i in reaction j. v_j quantifies the rate of reaction j, and b_i is the rate of transport (active or passive) of metabolite i across cellular boundaries. Reversible reactions are replaced by the difference of the corresponding pair of exchange reactions thus maintaining positivity of all reaction steps present in the model:

$$v_j = v_j^{\text{forward}} - v_j^{\text{backward}}, \quad \forall j \in J^R. \tag{A.1}$$

Using the principle of stoichiometric analysis along with the application of a pseudo-steady-state hypothesis to the intracellular metabolites (Vallino and Stephanopoulos, 1993), an overall flux balance can be written as follows:

$$\sum_j S_{ij} \cdot v_j = b_i, \quad \forall i \in I. \tag{A.2}$$

Metabolites that do not exchange across the cell's boundaries have no exchange flux. That is, $b_i = 0, \forall i \notin I^E$. Positive or negative values for b are allowed when a metabolite can enter or leave the cellular boundaries, respectively.

When ^{13}C substrate labeling is introduced, an additional layer of detail is needed to fully characterize the network. This information includes the substrate(s) labeling patterns and descriptions of the fate of the carbon atoms in each reaction. We express the labeling patterns using the concepts of isotopomer distribution vectors (IDVs) (Schmidt et al., 1997; Wittmann and Heinzle, 2002) and isotopomer mapping matrices (IMMs) (see Section 2.2). The following additional sets, parameters and variables are required for the mathematical quantification.

Sets:

$K = \{k\}$　　　　　set of isotopomers
$N = \{n\}$　　　　　the number of carbon atoms

Parameters:

$IMM^j_{i' \to i, k' \to k}$　　　　　Isotopomer mapping matrix

Variables:

I_{ik}　　　　　　　　isotopomer distribution vector

The set K enumerates all possible labeling patterns for a given metabolite. Variable I_{ik}, referred to as the IDV, is defined as the fraction of metabolite i that exists in the isotopomer form k. There are at most 2^n isotopomers for each metabolite containing n carbons, and the sum of all I_{ik} over k for a given metabolite i is equal to one. Parameter IMM links the specific isotopomer form k' of reactant i' that contribute to the formation of product i in isotopomer form k through reaction j. The corresponding entry for IMM for such an indices combination is equal to one unless it refers to a symmetric molecule i' (e.g., succinate). For symmetric molecules IMMs have a fractional entry to account for the fact that an isotopomer k' may map to more than one isotopomers k for the same product molecule i. Given the above definitions the mass balance for every metabolite i in isotopomer form k can be written as follows:

$$I_{ik} \cdot b_i = \sum_{j \mid S_{ij} > 0} \left(S_{ij} \cdot v_j \cdot \prod_{i' \mid S_{i'j} < 0} \sum_{k'} IMM^j_{i' \to i, k' \to k} \cdot I_{i'k'} \right) + \sum_{j \mid S_{ij} < 0} S_{ij} \cdot v_j \cdot I_{ik}, \quad \forall i \in I, k \in K \tag{A.3}$$

in which the product symbol is used for term by term multiplication. Fig. 1 pictorially illustrates the origin of each term present in the isotopomer balance equation. The balance equation acts as a mixing/splitting node where all generation terms through different reactions j of metabolite i in isotope form k are aggregated and then channeled through to all consuming reactions at a fixed isotopomer fraction. Note that the generation terms gives rise to all nonlinearities between IMMs. Unimolecular reactions yield products of metabolic fluxes times isotopomer fractions (bilinear terms) while bimolecular reactions contribute trilinear terms due to the presence of two separate isotopomer fractions in the product. Thus, the isotopomer balances abstracted through Eq. (A.3) relate the I_{ik} of each intracellular metabolite in terms of fluxes v_j, IMMs, and the isotopomer distribution in the feedstock (Schmidt et al., 1997).

Experimental techniques such as mass spectrometry are used to obtain isotope labeling data. However, instead of providing the isotopomer distribution I_{ik} directly, mass spectrometry provides raw data on the mass distribution vectors (MDV) for each measured metabolite. Because essentially mass spectroscopy acts as a molecular level scale that deduces the weight distribution of each measured metabolite, MDV contain information only about the total number of labeled carbons in a fragment generated by the mass spectrophotometer but not their specific location. These weight differences arise from the fact that different isotope forms may contain different numbers of labeled carbons. Each entry of the MDV contains a group of isotopomers that all have the same mass (Wittmann and Heinzle, 1999). Linking the IDVs and the MDV requires the following additional definitions.

Sets:

$F = \{f\}$	set of fragments
$M = \{m\}$	set of mass fractions
$R = \{r\}$	set of replicants
$I^{\mathrm{M}} = \{i\}$	set of measured metabolites

Parameters:

$IDVMDV^i_{f,k \to m}$	isotopomer grouping matrix
$\sigma^i_{f,m}$	standard deviation

Variables:

$MDV^i_{f,m,r}$	mass distribution vector
$MDV^i_{f,m}$	mean mass distribution vector

Here set F contains all possible fragments generated upon ionization for a measured metabolite i. Set I^{M} is the set of measured metabolites which in the current study are 14 amino acids. Set M represents the mass fractions observed for any given fragment f, as shown in the illustrative example in Fig. 2. The set R represents all measurement replicates, as described above in the Experimental section. Parameter $IDVMDV^i_{f,k \to m}$ links the specific isotopomer form k of metabolite i with the mass distribution m associated with fragment f of metabolite i. For a given fragment, all of the isotopomers that have the same number of labeled carbon atoms in them would be grouped together into the same $MDV^i_{f,m,r}$, as shown for the example in Fig. 2. The variable $MDV^i_{f,m}$ is simply the algebraic mean of the mass distribution m of the replicates for a given fragment f of metabolite i, and $\sigma^i_{f,m}$ is the corresponding experimental error.

Using the notation listed above, the problem of calculating the fluxes and labeling patterns for a given set of GC–MS data is formulated as a nonlinear optimization problem (FluxCalc):

measured amino acid fragments. Eq. (A.1) enforces mass balance on metabolites and Eq. (A.3) does the same for individual isotopomers. The mapping of MDV onto I_{ik} is performed by Eq. (A.5). Eq. (A.6) constrains all the fluxes to be positive, and further restricts them between lower and upper bounds; except for fluxes whose values are fixed, these bounds were 0 and 100 times the observed glucose rate. Eq. (A.7) enforce that I_{ik} remains between zero and one and that their sum is equal to one for each metabolite i.

Given the non-convex nature of (FluxCalc), the size of the resulting formulation and the experimental variability in the experimental MDV, we pursue the identification of as many local optima solutions as possible. To facilitate the enumeration of local optima, the additional formulation (FluxInit) was constructed to generate a set of random feasible flux distributions that are used as initial conditions for the solution of the (FluxCalc) problem. (FluxInit) essentially minimally perturbs an original randomly chosen flux distribution so as to satisfy all metabolite balances. The resulting linear optimization problem is as follows:

$$\text{minimize} \quad z = \sum_j e_j, \quad \forall j \in J \quad \text{(FluxInit)} \tag{A.8}$$

subject to

$$\sum_j S_{ij} \cdot v_j = b_i, \quad \forall i \in I, \tag{A.2}$$

$$v_j^{\mathrm{rand}} - v_j \geqslant e_j \geqslant v_j - v_j^{\mathrm{rand}}, \quad \forall j \in J, \tag{A.9}$$

$$v_j \geqslant 0, \quad \mathrm{LB}_j \leqslant v_j \leqslant \mathrm{UB}_j, \quad \forall j \in J. \tag{A.6}$$

Here, Eq. (A.8) minimizes the sum of the distance from feasibility for each flux v_j encoded within variable e_j. Eq. (A.9) defines variable e_j to be equal to the absolute value of the difference between v_j and v_j^{rand} which assumes values drawn from a random uniform distribution within prespecified upper and lower bounds UB, LB. The fluxes

$$\text{minimize} \quad z = \sum_{i,f} \sum_m \left(\frac{MDV^{i,\exp}_{f,m} - MDV^{i,\sim}_{f,m}}{\sigma^i_{f,m}} \right)^2 \quad \text{(FluxCalc)} \tag{A.4}$$

subject to

$$\sum_j S_{ij} \cdot v_j = b_i, \quad \forall i \in I, \tag{A.1}$$

$$I_{ik} \cdot b_i = \sum_{j|S_{ij}>0} \left(S_{ij} \cdot v_j \cdot \prod_{i'|S_{i'j}<0} \sum_{k'} IMM^j_{i' \to i, k' \to k} \cdot I_{i'k'} \right) + \sum_{j|S_{ij}<0} S_{ij} \cdot v_j \cdot I_{ik}, \quad \forall i \in I, \ k \in K, \tag{A.3}$$

$$MDV^{i,\sim}_{f,m} = \sum_k IDVMDV^i_{f,k \to m} I_{ik},$$

$$\forall m \in M, \ f \in F, \ i \in I^{\mathrm{M}},$$

$$v_j \geqslant 0, \ \mathrm{LB}_j \leqslant v_j \leqslant \mathrm{UB}_j, \ \forall j \in J, \tag{A.6}$$

$$0 \leqslant I_{ik} \leqslant 1, \quad \forall i \in I, \ k \in K$$

$$\sum_k I_{ik} = 1, \quad \forall i \in I. \tag{A.7}$$

(A.5)

In (FluxCalc), the objective function z in Eq. (A.4) is the sum of the discrepancies between the experimental data (exp) and the calculated values (sim) for the MDVs of the

are constrained as in (FluxCalc) by Eq. (A.6). The overall mass balance is applied as before by Eq. (A.6). The overall solution procedure is listed in a step-wise manner below.

Step 1: Populate model parameters S_{ij}, $IMM^j_{i' \to i, k' \to k}$, $IDVMDV^i_{f, k \to m}$, $MDV^{i,\exp}_{m,f,r}$, $I_{i \in I^F, k}$ UB_j and LB_j based on the experimental system setup and metabolic network abstraction. Set iteration counter Iter to zero.

Step 2: Initialize v^{rand}_j using a random number generator to construct a uniform distribution such that $LB_j < v^{\mathrm{rand}}_j < UB_j$, $\quad \forall j \in J$.

Step 3: Solve (FluxInit) to determine an initial feasible flux distribution $\to v_j$.

Step 4: Initialize I_{ik} by setting $I_{i,k=1} = 1, \forall i \notin I^F$; $I_{i,k>1} = 0, \forall i \notin I^F$.

Step 5: Solve the nonlinear optimization problem (FluxCalc) to generate a (local) solution v_j, I_{ik} and $MDV^{\mathrm{sim}}_{m,f}$.

Step 6: Store solutions for current iteration, update iteration counter Iter $\leftarrow$ Iter $+$ 1, and return to Step 2 if the maximum iterations limit is not exceeded.

Appendix B. Supplementary data

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ymben.2007.05.005.

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., 2007a. Elementary metabolic units (EMU): a novel framework for modeling isotopic distributions. Metab. Eng. 9, 68–86.

Antoniewicz, M.R., et al., 2007b. Metabolic flux analysis in a nonstationary system: fed-batch fermentation of a high yielding strain of *E. coli* producing 1,3-propanediol. Metab. Eng. 9, 277–292.

Avery, M.A., et al., 1992. Stereoselective total synthesis of (+)-artemisinin, the antimalarial constituent of Artemisia annua L. J. Am. Chem. Soc. 114, 974–979.

Bacher, A., et al., 1998. Elucidation of novel biosynthetic pathways and metabolite flux patterns by retrobiosynthetic NMR analysis. FEMS Microbiol. Rev. 22, 567–598.

Bailey, J.E., 1991. Toward a science of metabolic engineering. Science 252, 1668–1675.

Burgard, A.P., et al., 2004. Flux coupling analysis of genome-scale metabolic network reconstructions. Genome Res. 14, 301–312.

Christensen, B., Nielsen, J., 1999. Isotopomer analysis using GC–MS. Metab. Eng. 1, 282–290.

Cragg, G.M., 1998. Paclitaxel (Taxol): a success story with valuable lessons for natural product drug discovery and development. Med. Res. Rev. 18, 315–331.

Dauner, M., et al., 2001. Metabolic flux analysis with a comprehensive isotopomer model in *Bacillus subtilis*. Biotechnol. Bioeng. 76, 144–156.

Dhingra, V., et al., 2000. Current status of artemisinin and its derivatives as antimalarial drugs. Life Sci, 279–300.

Edwards, J.S., Palsson, B.O., 2000. The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. Proc. Natl. Acad. Sci. USA 97, 5528–5533.

Emmerling, M., et al., 2002. Metabolic flux responses to pyruvate kinase knockout in *Escherichia coli*. J. Bacteriol. 184, 152–164.

Fischer, E., et al., 2004. High-throughput metabolic flux analysis based on gas chromatography–mass spectrometry derived 13C constraints. Anal. Biochem. 325, 308–316.

Fong, S.S., Palsson, B.O., 2004. Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. Nat. Genet. 36, 1056–1058.

Forbes, N.S., et al., 2001. Using isotopomer path tracing to quantify metabolic fluxes in pathway models containing reversible reactions. Biotechnol. Bioeng. 74, 196–211.

Ghosh, S., et al., 2005. Closing the loop between feasible flux scenario identification for construct evaluation and resolution of realized fluxes via NMR. Comput. Chem. Eng. 29, 459–466.

Ghosh, S., et al., 2006. A three-level problem-centric strategy for selecting NMR precursor labeling and analytes. Metab. Eng. 8, 491–507.

Ibarra, R.U., et al., 2002. *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature 420, 186–189.

Jeffrey, F.M., et al., 1991. 13C-NMR: a simple yet comprehensive method for analysis of intermediary metabolism. Trends Biochem. Sci. 16, 5–10.

Kanehisa, M., et al., 2006. From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res. 34, D354–D357.

Kelleher, J.K., 2001. Flux estimation using isotopic tracers: common ground for metabolic physiology and metabolic engineering. Metab. Eng. 3, 100–110.

Klapa, M.I., et al., 2003. Systematic quantification of complex metabolic flux networks using stable isotopes and mass spectrometry. Eur. J. Biochem. 270, 3525–3542.

Kleijn, R.J., et al., 2006. 13C-labeled gluconate tracing as a direct and accurate method for determining the pentose phosphate pathway split ratio in *Penicillium chrysogenum*. Appl. Environ. Microbiol. 72, 4743–4754.

Martin, V.J., et al., 2003. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. Nat. Biotechnol. 21, 796–802 Epub 2003 June 1.

Marx, A., et al., 1996. Determination of the fluxes in the central metabolism of *Corynebacterium glutamicum* by nuclear magnetic resonance spectroscopy combined with metabolite balancing. Biotechnol. Bioeng. 49, 111–129.

Neidhardt, F.C., et al., 1974. Culture medium for enterobacteria. J. Bacteriol. 119, 736–747.

Nielsen, J., 2003. It is all about metabolic fluxes. J. Bacteriol. 185, 7031–7035.

Reed, J.L., et al., 2003. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). Genome Biol. 4, R54.

Riascos, C.A.M., et al., 2005. A global optimization approach for metabolic flux analysis based on labeling balances. Comput. Chem. Eng. 29, 447–458.

Sauer, U., et al., 1999. Metabolic flux ratio analysis of genetic and environmental modulations of *Escherichia coli* central carbon metabolism. J. Bacteriol. 181, 6679–6688.

Schmidt, K., et al., 1997. Modeling isotopomer distributions in biochemical networks using isotopomer mapping matrices. Biotechnol. Bioeng. 55, 831–840.

Schmidt, K., et al., 1999. Quantitative analysis of metabolic fluxes in *Escherichia coli*, using two-dimensional NMR spectroscopy and complete isotopomer models. J. Biotechnol. 71, 175–189.

Segre, D., et al., 2002. Analysis of optimality in natural and perturbed metabolic networks. Proc. Natl. Acad. Sci. USA 99.

Siddiquee, K.A., et al., 2004. Metabolic flux analysis of pykF gene knockout *Escherichia coli* based on 13C-labeling experiments together with measurements of enzyme activities and intracellular metabolite concentrations. Appl. Microbiol. Biotechnol. 63, 407–417.

Sleno, L., et al., 2005. Assigning product ions from complex MS/MS spectra: the importance of mass uncertainty and resolving power. J. Amer. Soc. Mass Spectrom. 16, 183–198.

Stephanopoulos, G., Vallino, J.J., 1991. Network rigidity and metabolic engineering in metabolite overproduction. Science 252, 1675–1681.

Szyperski, T., 1995. Biosynthetically directed fractional 13C-labeling of proteinogenic amino acids. An efficient analytical tool to investigate intermediary metabolism. Eur. J. Biochem. 232, 433–448.

Tawarmalani, M., Sahinidis, N.V., 2004. Global optimization of mixed-integer nonlinear programs: a theoretical and computational study. Math. Programming 99, 563–591.

Taylor, J.C., 1990. Statistical Techniques for Data Analysis. Lewis Publishers, Chelsea, MI.

Vallino, J.J., Stephanopoulos, G., 1993. Metabolic flux distributions in *Corynebacterium glutamicum* during growth and lysine overproduction. Biotechnol. Bioeng. 41, 633–646.

Van Dien, S.J., et al., 2003. Quantification of central metabolic fluxes in the facultative methylotroph methylobacterium extorquens AM1 using 13C-label tracing and mass spectrometry. Biotechnol. Bioeng. 84, 45–55.

van Winden, W.A., et al., 2002. Correcting mass isotopomer distributions for naturally occurring isotopes. Biotechnol. Bioeng. 80, 477–479.

van Winden, W.A., et al., 2003. Metabolic flux and metabolic network analysis of *Penicillium chrysogenum* using 2D [13C, 1H] COSY NMR measurements and cumulative bondomer simulation. Biotechnol. Bioeng. 83, 75–92.

Varma, A., Palsson, B.O., 1994. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. Appl. Environ. Microbiol. 60, 3724–3731.

Wahl, S.A., et al., 2004. New tools for mass isotopomer data evaluation in (13)C flux analysis: mass isotope correction, data consistency checking, and precursor relationships. Biotechnol. Bioeng. 85, 259–268.

Wiechert, W., de Graaf, A.A., 1997. Bidirectional reaction steps in metabolic networks: I. Modeling and simulation of carbon isotope labeling experiments. Biotechnol. Bioeng. 55, 101–117.

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.

Wittmann, C., Heinzle, E., 1999. Mass spectrometry for metabolic flux analysis. Biotechnol. Bioeng. 62, 739–750.

Wittmann, C., Heinzle, E., 2002. Genealogy profiling through strain improvement by using metabolic network analysis: metabolic flux genealogy of several generations of lysine-producing corynebacteria. Appl. Environ. Microbiol. 68, 5843–5859.

Wittmann, C., et al., 2002. In vivo analysis of intracellular amino acid labelings by GC/MS. Anal. Biochem. 307, 379–382.

Yang, T.H., et al., 2004. Metabolic network simulation using logical loop algorithm and Jacobian matrix. Metab. Eng. 6, 256–267.

Zhao, J., Shimizu, K., 2003. Metabolic flux analysis of *Escherichia coli* K12 grown on 13C-labeled acetate and glucose using GC–MS and powerful flux calculation method. J. Biotechnol. 101, 101–117.

Zupke, C., Stephanopoulos, G., 1994. Modeling of isotope distributions and intracellular fluxes in metabolic networks using atom mapping matrices. Biotechnol. Prog. 10, 489–498.