ARTICLE

BIOTECHNOLOGY and BIOENGINEERING

An Elementary Metabolite Unit (EMU) Based Method of Isotopically Nonstationary Flux Analysis

Jamey D. Young, Jason L. Walther, Maciek R. Antoniewicz, Hyuntae Yoo, Gregory Stephanopoulos

Department of Chemical Engineering, Massachusetts Institute of Technology, Building 56 Room 469C, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139; telephone: 617-253-4583; fax: 617-253-3122; e-mail: gregstep@mit.edu

Received 5 March 2007; revision received 25 July 2007; accepted 8 August 2007

Published online 4 September 2007 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/bit.21632

ABSTRACT: Nonstationary metabolic flux analysis (NMFA) is at present a very computationally intensive exercise, especially for large reaction networks. We applied elementary metabolite unit (EMU) theory to NMFA, dramatically reducing computational difficulty. We also introduced block decoupling, a new method that systematically and comprehensively divides EMU systems of equations into smaller subproblems to further reduce computational difficulty. These improvements led to a 5000-fold reduction in simulation times, enabling an entirely new and more complicated set of problems to be analyzed with NMFA. We simulated a series of nonstationary and stationary GC/MS measurements for a large *E. coli* network that was then used to estimate parameters and their associated confidence intervals. We found that fluxes could be successfully estimated using only nonstationary labeling data and external flux measurements. Addition of near-stationary and stationary time points increased the precision of most parameters. Contrary to prior reports, the precision of nonstationary estimates proved to be comparable to the precision of estimates based solely on stationary data. Finally, we applied EMU-based NMFA to experimental nonstationary measurements taken from brown adipocytes and successfully estimated fluxes and some metabolite concentrations. By using NFMA instead of traditional MFA, the experiment required only 6 h instead of 50 (the time necessary for most metabolite labeling to reach 99% of isotopic steady state).

Biotechnol. Bioeng. 2008;99: 686–699.

KEYWORDS: metabolic engineering; metabolic flux analysis; isotopically nonstationary; elementary metabolite units; Escherichia coli; brown adipocytes

Introduction

The goal of metabolic flux analysis (MFA) is the quantitative determination of intracellular fluxes. Results are typically obtained by (1) introducing a labeled substrate into a cell culture at metabolic steady state, (2) allowing the system to reach an isotopic steady state, (3) measuring relative labeling in metabolic intermediates and byproducts, and (4) computationally processing these measurements to estimate fluxes (Wiechert, 2001). An understanding of these fluxes can facilitate identification of possible genetic targets for optimization of a particular phenotype (Kiefer et al., 2004). Similarly, flux data can shed light on the metabolic impact of environmental conditions or previous genetic modifications (Klapa et al., 2003; Sauer et al., 1997; Wittmann and Heinzle, 2002).

Nonstationary metabolic flux analysis (NMFA) is similar to MFA with the provision that metabolite labeling is sampled and measured during the transient period before the system reaches an isotopic steady state. NMFA offers significant advantages as compared to MFA:

- NMFA experiments are much less costly (in terms of both time and money) since one does not need to wait for isotopic steady state to be established (Nöh et al., 2007).
- NMFA is particularly suited for systems that cannot be held at a metabolic steady state indefinitely (e.g., primary cells or animal studies) because experimental durations are greatly reduced.
- In some cases, NMFA identifies metabolic fluxes with greater precision because some isotopically transient measurements have greater sensitivities to fluxes (Nöh et al., 2006).

This article contains Supplementary Material available via the Internet at http://www.interscience. wiley.com/jpages/0006-3592/suppmat.

Jamey D. Young and Jason L. Walther contributed equally to this work.

Hyuntae Yoo's present address is Institute for Systems Biology, 1441 North 34th Street, Seattle, WA 98103.

Correspondence to: G. Stephanopoulos

Contract grant sponsor: National Institutes of Health

Contract grant number: DK-75850; DK-072856

- In some cases, NMFA measurement data can be used to estimate metabolite concentrations in addition to fluxes.
- ^{13}C NMFA can successfully estimate fluxes in systems that rely solely on single-carbon substrates (e.g., photoautotrophs and methylotrophs) whereas at isotopic steady state, metabolites are uniformly labeled and no new information is generated by a ^{13}C tracer (Shastri and Morgan, 2007).

MFA and NMFA are concerned with solving an "inverse problem" in which fluxes (and in the case of NMFA, concentrations) are estimated from metabolite labeling distributions by means of an iterative least-squares fitting procedure. At each iteration, a "forward problem" must be solved in which metabolite labeling distributions are simulated for a given metabolic network and a given set of parameter estimates. The mismatch between the simulated and experimental measurements is assessed and the parameter estimates are updated to achieve an improving fit.

In the context of MFA, the forward problem can be represented by systems of linear algebraic equations. NMFA, on the other hand, requires the solution of systems of ordinary differential equations, a significantly more difficult task. This additional complexity means that the algorithms for NMFA must be carefully designed so that the computational expense for large metabolic networks does not become prohibitive. Currently, state-of-the-art algorithms (using cumomer fractions as state variables) require more than an hour to simulate isotopic labeling of a realistic network model (Nöh et al., 2006; Wiechert et al., 1999).

In this work, we propose a new approach based upon the Elementary Metabolite Unit (EMU) framework (Antoniewicz et al., 2007a) that efficiently and robustly handles the inverse problem of NMFA by solving the forward problem thousands of times faster than currently available methods. Because of these improvements in the NMFA model, we are able to show for the first time that fluxes and concentrations can be estimated from nonstationary data for realistically sized metabolic networks in short amounts of time.

Methods

EMU Network Decomposition

The nonstationary treatment presented here is built using the mass isotopomer distributions (MIDs) of EMUs as state variables (Antoniewicz et al., 2007a). An EMU is defined as a distinct subset of a metabolite's atoms. EMUs can exist in a variety of mass states depending on their isotopic compositions. An EMU in its lowest mass state is referred to as M+0, while an EMU that contains one additional atomic mass unit (e.g., due to the presence of a ^{13}C atom in place of a ^{12}C atom) is referred to as M+1, with higher mass states described accordingly. An MID is a vector that contains the fractional abundance of each mass state of an EMU.

The goal of an NMFA simulation is the calculation of metabolite labeling patterns that are measurable by mass spectroscopy; that is, the MIDs of a certain subset of EMUs in the system. While the total number of all possible EMUs in a network is equal to the number of isotopomers or cumomers, in most cases only a small fraction of EMUs is required to simulate measurable MIDs.

EMUs of metabolites in a common reaction network can be assembled into an analogous EMU network composed of EMU reactions where the MIDs of upstream EMUs affect the MIDs of downstream EMUs. Oftentimes, EMU networks can be decoupled into separate and smaller subnetworks. Decoupling of EMU reactions based on (1) EMU size and (2) network connectivity has previously been discussed (Antoniewicz et al., 2007a). (EMU size is defined as the number of atoms comprising a particular EMU.) Because MIDs of EMUs depend only upon MIDs of equally sized or smaller EMUs, the EMU network can be partitioned into size-based networks, each containing equally sized EMUs and depending on inputs only from smaller sized EMUs. If smaller, completely independent EMU subnetworks can be identified within these size-based networks, even further decoupling can occur. Computational costs can therefore be decreased in two ways: first, the total size of the system can be reduced, and second, the system can be divided into smaller subsystems that cumulatively can be solved more quickly.

Block Decoupling

We propose a systematic and comprehensive method of EMU reaction network decoupling in which metabolite units are grouped into blocks. A block is defined as a set of EMUs whose MIDs are mutually dependent within the context of the EMU reaction network. Thus, by definition all EMUs within a particular block (1) are of the same size, (2) mutually approach an isotopic steady state, and (3) must be solved for simultaneously and not sequentially. Blocks can be arranged such that each is a self-contained subproblem depending only upon the outputs of previously solved blocks. This lets us work with smaller and more tractable matrices, greatly increasing computational efficiency.

To arrange EMUs into blocks, we first regard the EMU reaction network as a directed graph in which nodes represent EMUs and edges represent EMU reactions. An $N \times N$ adjacency matrix is then constructed for the directed graph, where N is the total number of EMUs. In short, a nonzero entry $a(i, j)$ of the adjacency matrix indicates the dependence of the ith EMU's MID on the jth EMU's MID. We then perform a Dulmage–Mendelsohn decomposition on the adjacency matrix, returning an upper block triangular matrix from which the diagonal blocks are extracted (Dulmage and Mendelsohn, 1958; Pothen and Fan, 1990).

Simple Network

A simple metabolic network appears in Figure 1a as an example. Figure 1b delineates the atom transitions for the network. Hypothetical metabolite C is assumed to be measurable by GC/MS. After EMU decomposition, the nonstationary system can be described in terms of 16 EMUs. This represents a 44% reduction in state variables from the 29 cumomer fractions required to simulate the system with the cumomer method. After decoupling based on EMU size and connectivity, these 16 state variables can be separated into four smaller subproblems (see Fig. 2a).

By applying Dulmage–Mendelsohn decomposition and block decoupling to the simple network, we can achieve even further system reduction. Figure 3 shows this decomposition and the resulting blocks in matrix form. Block decoupling improves upon previous methods, enabling the 16 essential EMUs to be divided among eight subproblems instead of four (see Fig. 2b). Table I provides a detailed comparison of the model reductions achieved by cumomer and EMU decompositions both with and without block decoupling.

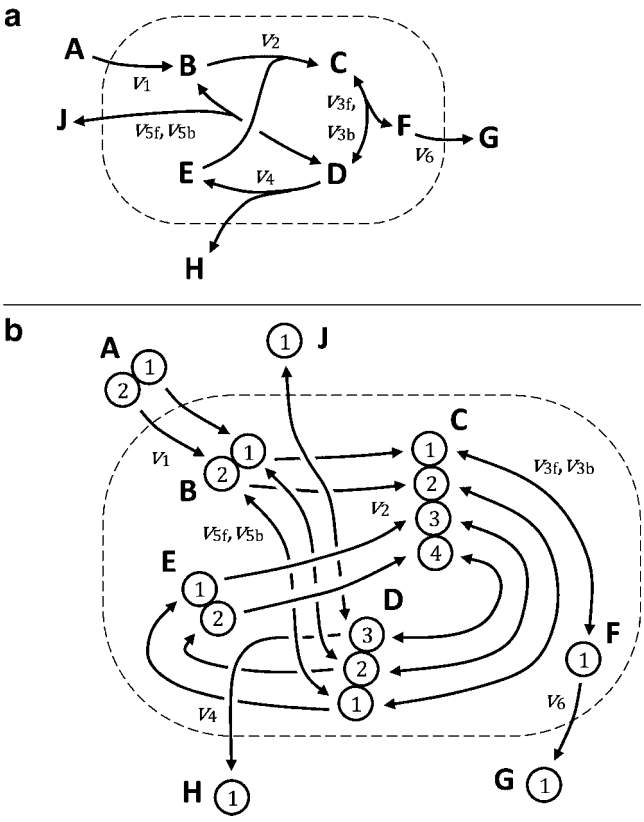

Figure 1. **a** A simple example network used to illustrate EMU network decomposition. The network fluxes are assumed to be constant since the system is at a metabolic steady state. Extracellular metabolites A and J are assumed to be at a fixed state of isotopic labeling to which intracellular metabolites B, C, D, E, and F adapt over time. **b** Atom transitions for the simple example network.

Simulation of Metabolite Labeling

Decomposition of a network into blocks of EMUs generates a cascaded system of ordinary differential equations, where level n of the cascade represents the network of EMUs within the nth block. Each system has the following form:

$$\mathbf{C}_n \cdot \frac{\mathrm{d}\mathbf{X}_n}{\mathrm{d}t} = \mathbf{A}_n \cdot \mathbf{X}_n + \mathbf{B}_n \cdot \mathbf{Y}_n \tag{1}$$

The rows of the state matrix $\mathbf{X}_n$ correspond to MIDs of EMUs within the nth block. The input matrix $\mathbf{Y}_n$ is analogous but with rows that are MIDs of EMUs that are previously calculated inputs to the nth block. The concentration matrix $\mathbf{C}_n$ is a diagonal matrix whose elements are concentrations corresponding to EMUs in $\mathbf{X}_n$. Finally, the system matrices $\mathbf{A}_n$ and $\mathbf{B}_n$ describe the network as follows:

$$\mathbf{A}_n(i,j) = \begin{cases} -\text{sum of fluxes consuming } i\text{th EMU in } \mathbf{X}_n & i = j \\ \text{flux to } i\text{th EMU in } \mathbf{X}_n \text{ from } j\text{th EMU in } \mathbf{X}_n & i \neq j \end{cases} \tag{2}$$

$$\mathbf{B}_n(i,j) = \text{flux to } i\text{th EMU in } \mathbf{X}_n \text{ from } j\text{th EMU in } \mathbf{Y}_n \tag{3}$$

The least-squares fitting algorithm employed to solve the inverse problem requires repeated calculation of first-order derivatives, that is, sensitivities of simulated measurements with respect to fluxes and concentrations. To this end, implicit differentiation of Equation (1) yields

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathbf{X}_n}{\partial \mathbf{p}} = {} & \mathbf{C}_n^{-1} \cdot \mathbf{A}_n \cdot \frac{\partial \mathbf{X}_n}{\partial \mathbf{p}} + \frac{\partial(\mathbf{C}_n^{-1} \cdot \mathbf{A}_n)}{\partial \mathbf{p}} \cdot \mathbf{X}_n \\ & + \mathbf{C}_n^{-1} \cdot \mathbf{B}_n \cdot \frac{\partial \mathbf{Y}_n}{\partial \mathbf{p}} + \frac{\partial(\mathbf{C}_n^{-1} \cdot \mathbf{B}_n)}{\partial \mathbf{p}} \cdot \mathbf{Y}_n \end{aligned} \tag{4}$$

where $\mathbf{p}$ is a vector of metabolic fluxes and concentrations.

We integrate the system with a customized ordinary differential equation solver that discretizes Equations (1) and (4) by applying a first-order hold equivalent with adaptive step size control (Powell et al., 1990). This method is A-stable, simple to code, and enables large time steps by making use of partial analytical solutions to the system equations. More details are available in the appendix.

Flux and Concentration Estimation

Fluxes and concentrations are estimated by minimizing the difference between measured and simulated data according to the following equation (Antoniewicz et al., 2006; Nöh et al., 2006):

$$\begin{aligned} &\min_{\mathbf{u},\mathbf{c}} \Phi = [\mathbf{m}(\mathbf{u},\mathbf{c},t) - \hat{\mathbf{m}}(t)]^T \cdot \Sigma_{\mathbf{m}}^{-1} \cdot [\mathbf{m}(\mathbf{u},\mathbf{c},t) - \hat{\mathbf{m}}(t)] \\ &\text{s.t. } \mathbf{N} \cdot \mathbf{u} \geq 0, \ \mathbf{c} \geq 0 \end{aligned} \tag{5}$$

DOI 10.1002/bit

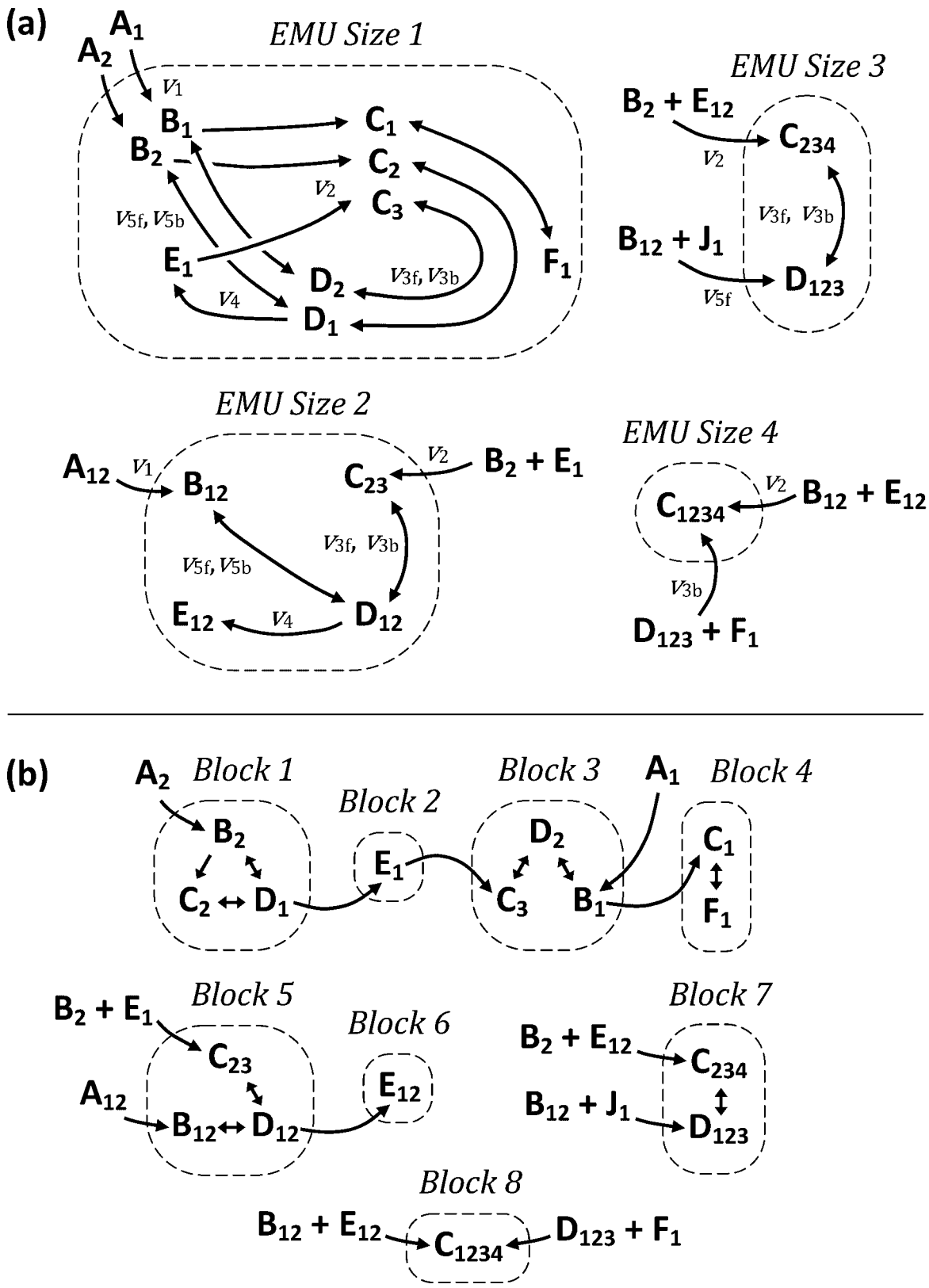

Figure 2. **a** EMU network decomposition for the simple example network (Fig. 1) generated to simulate the labeling of metabolite C. The EMU reaction network was decoupled based on EMU size and network connectivity. **b** EMU network decomposition for the same network using block decoupling. Subscripts refer to the atoms of a compound that are contained within the EMU.

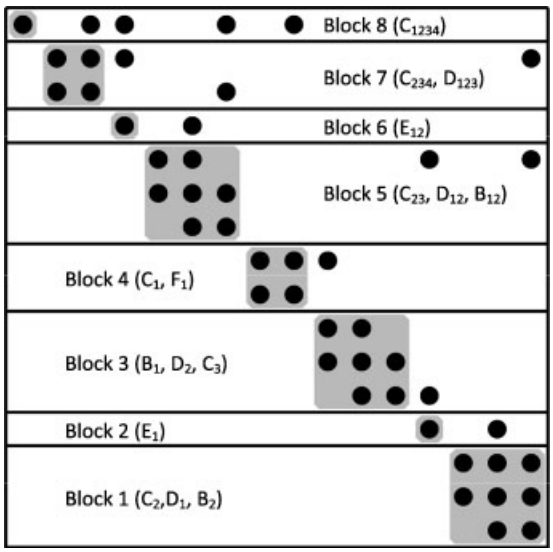

Figure 3. Dulmage–Mendelsohn decomposition of an adjacency matrix representing the EMU reaction network for the simple example network described in Figure 1. A nonzero entry (denoted by a black circle) at the *i*th row and *j*th column of the matrix represents the dependence of the *i*th EMU's MID on the *j*th EMU's MID. The upper block triangular matrix resulting from the Dulmage–Mendelsohn decomposition can be separated into blocks as indicated by bold lines. Blocks can be solved in a sequential order, beginning at the lower right-hand corner and working upwards.

where Φ is the objective function to be minimized, $\mathbf{u}$ is a vector of free fluxes, $\mathbf{c}$ is a vector of metabolite concentrations, t is time, $\mathbf{m}(\mathbf{u}, \mathbf{c}, t)$ is a vector of simulated measurements, $\hat{\mathbf{m}}(t)$ is a vector of observed measurements, $\boldsymbol{\Sigma}_{\mathbf{m}}$ is the measurement covariance matrix, and $\mathbf{N}$ is the nullspace of the stoichiometric matrix. We have implemented a reduced gradient method to handle the linear constraints of this problem within a Levenberg–Marquardt nonlinear least-squares solver (Gill et al., 1981; Madsen et al., 2004).

Calculation of parameter standard errors requires the inverse Hessian of Φ, which becomes ill-conditioned when some parameters are poorly identifiable. Because the Hessian is obtained by numerically integrating the measurement sensitivities in Equation (4), it is contaminated by numerical errors in the nonstationary case. Upon matrix inversion, even small errors can greatly distort standard error estimates, rendering them nearly meaningless. As such, we compute nonlinear flux confidence intervals (using parameter continuation around the optimal solutions) instead of relying upon local standard errors (Antoniewicz et al., 2006). These confidence intervals, though more computationally expensive to obtain than local standard errors, yield a significantly more reliable and realistic description of the true parameter uncertainty.

Table I. A comparison of modeling approaches to simulate the dynamic labeling of a simple example network.

Model	Cumomer	EMU	EMU
Decoupling method	Size	Size/connectivity	Blocks
(Size) # of vars	(1) 12	(1) 9	(1) 3, 3, 2, 1
	(2) 11	(2) 4	(2) 3, 1
	(3) 5	(3) 2	(3) 2
	(4) 1	(4) 1	(4) 1
Total variables	29	16	16

EMU network decomposition followed by block decoupling minimizes the number of state variables both in the overall system and within any one subproblem. The subproblems are listed by EMU size (or in the case of cumomers, by weight). The EMU sizes (or cumomer weights) are indicated within parentheses and the number of variables within each subproblem follow. For instance, the entry "(2) 3, 1" indicates that there are two subproblems involving EMUs of size 2. One subproblem contains three variables and the other only one.

Results

Small *E. coli* Network

To compare the performance of our approach to prior methods, we reconstructed the simplified *E. coli* model described by Nöh consisting of 28 free fluxes and 16 metabolite pools (Nöh et al., 2006). Application of the EMU-based algorithm to this system leads to a 5000-fold reduction in the computational time required for parameter and sensitivity simulation of the forward problem (from 83 min on an AMD Opteron 2000+ down to 1 s on a 2.0 GHz T2500 dual core processor). Whereas computational time for parameter estimation via cumomers was conjectured to be 24–48 h, we estimated fluxes and concentrations in less than 1 min, beginning from a randomized set of initial parameters.

Large *E. coli* Network

Because of the increased efficiency of EMU-based NMFA, we chose to apply our method to a larger and more realistic *E. coli* network. Specifically, we modeled the central metabolism of a strain capable of producing high levels of 1,3-propanediol (PDO) using a network that includes 35 free fluxes and 46 metabolite pools (Antoniewicz et al., 2007b). (A complete list of reactions and atom transitions is available in Table S-I of Supplementary Materials.) The size of this problem can be reduced by over 90% via EMU decomposition (relative to isotopomers/cumomers) and can be further parsed into 47 subproblems with block decoupling (compared to only 14 with decoupling by size and network connectivity). Block decoupling led to a 27% decrease in computational time relative to decoupling based only upon size and connectivity. Table II provides a detailed comparison of the model reductions achieved by cumomer and EMU decompositions both with and without block decoupling.

To investigate the relationship between sampling times and parameter identifiability, we generated a series of simulated data sets. We drew flux values from a previously

DOI 10.1002/bit

Table II. A comparison of modeling approaches to simulate the dynamic labeling of 33 GC/MS fragments in the large *E. coli* metabolic network.

Model	Cumomer	EMU	EMU
Decoupling Method	Size	Size/connectivity	Blocks
(Size) # of Vars	(1) 54	(1) 146	(1) 117, 24, 1 × 5
	(2) 241	(2) 90	(2) 34, 22, 9, 5 × 2, 4, 2, 1 × 9
	(3) 527	(3) 47	(3) 26, 5 × 2, 4, 1 × 7
	(4) 771	(4) 12, 8, 1 × 2	(4) 8, 6, 5, 1 × 3
	(5) 876	(5) 5, 1 × 4	(5) 5, 1 × 4
	(6) 832	(6) 2	(6) 2
	(7) 655	(7) none	(7) none
	(8) 404	(8) 1	(8) 1
	(9) 183	(9) 1	(9) 1
	(10) 57		
	(11) 11		
	(12) 1		
Total variables	4,612	318	318
Simulation time	Not available	22 seconds	16 seconds

Subproblems are specified as in Table I. Multiple occurrences of a particular size of subproblem are indicated with the multiplication symbol; that is, "5 × 2" indicates that two subproblems of size 5 exist within the system. EMU network decomposition dramatically reduces the number of state variables within the overall system. Block decoupling further simplifies the system by minimizing the number of state variables within any one subproblem, reducing computational time by an additional 27%.

published stationary MFA experiment involving the aforementioned PDO-producing strain (Antoniewicz et al., 2007b) and metabolite concentration values from various literature sources on both *E. coli* and *S. cerevisae* (Buchholz et al., 2001; Gombert et al., 2001). Three different sets of measurements were simulated:

- *Short nonstationary experiment*: Two sets of replicate measurements were taken every second for 15 s following the introduction of tracer to the culture (for a total of 30 sets of measurements). All measured metabolite fragments remained isotopically transient during this regime.

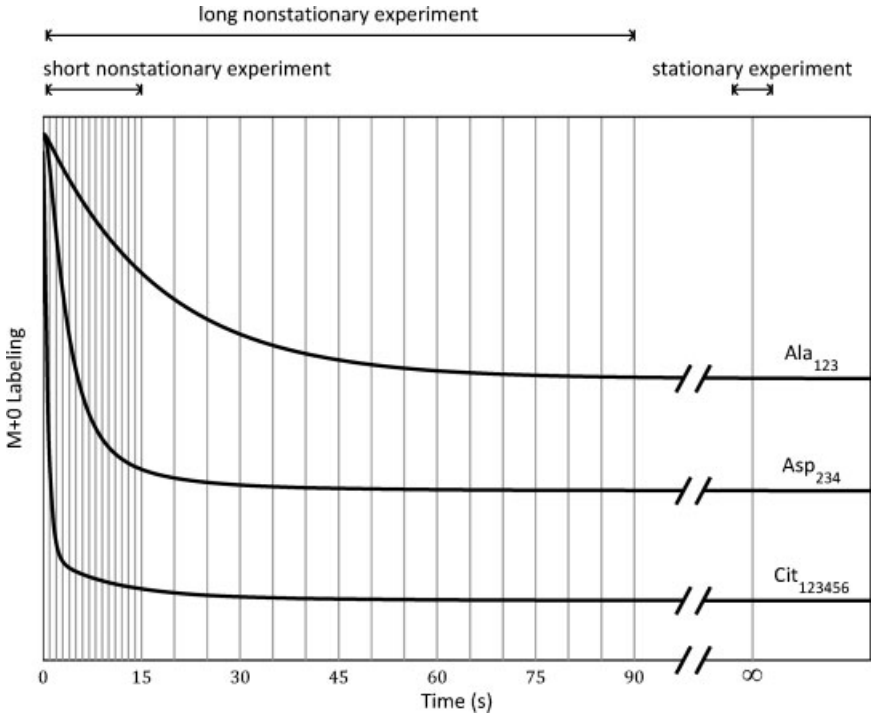

Figure 4. Measurement time points for the three simulated experiments involving the large *E. coli* model. Time points are indicated by gray vertical lines. M+0 labeling profiles for three representative measured metabolite fragments are overlaid to convey the time scale of the isotopic transience in the system.

- *Long nonstationary experiment*: One set of measurements was taken every second for 15 s following the introduction of tracer. For the next 75 s, measurements were then taken every 5 s, giving a total of 30 sets of measurements. By the end of this period, all measured metabolite fragments were within 99% of isotopic steady state.
- *Stationary experiment*: Measurements were conducted at a time sufficiently large such that all metabolite labeling was assumed constant. Thirty replicate sets of measurements were made such that the total number of measurements in all experiments was equal.

The timelines of these three experiments and their corresponding labeling measurements are illustrated in Figure 4. Table III lists the available measurements, which include 33 mass spectroscopy fragments and seven external fluxes. We assume that no metabolite concentrations are measurable. Standard errors of 5% for external fluxes and 0.3–1 mol% for GC/MS MIDs were assumed and introduced randomly and normally.

Fluxes and concentrations were estimated for each experiment. One forward nonstationary simulation of metabolite labeling required 16 s of computational time. Parameter estimation, beginning with a randomly distributed guess of concentrations and fluxes, ran in under 15 min. Nonlinear confidence intervals were also calculated for each estimated parameter. The results have been grouped as net fluxes, exchange fluxes, and metabolite concentrations and displayed in Figs. 5–7, respectively. (Results are also listed in Tables S-III, S-IV, and S-V of Supplementary Materials.) The wide majority of parameter values were recovered within their respective 95% confidence intervals; those values that were not recovered still fell reasonably near their estimated intervals. Exchange fluxes were scaled according to

$$v_{\mathrm{xch}}^{[0,1]} = \frac{v_{\mathrm{xch}}}{v_{\mathrm{xch}} + v_{\mathrm{ref}}} \tag{6}$$

where v_{xch} is the unscaled exchange flux, v_{ref} is the reference flux (in this case, glucose uptake), and $v_{\mathrm{xch}}^{[0,1]}$ is the scaled exchange flux (Wiechert and de Graaf, 1997).

Net fluxes were estimated most accurately and precisely. Most estimated net flux values fell within 10% of the actual values and possessed confidence intervals ranging between ±5 and ±20%. While the stationary measurements generated estimates significantly closer to actual values, confidence intervals across all three experiments were comparable in width.

Exchange flux estimation was considerably more difficult. Out of the 24 total exchange fluxes, 10 were unidentifiable (or nearly unidentifiable). Even when confidence intervals were obtained, they tended to be extremely broad. For the majority of these intervals, only an upper or a lower bound could be found. Overall, none of the three experiments could claim significantly more precise confidence intervals. On a parameter-by-parameter basis, however, precision varied greatly between the different experiments.

Both lower and upper bounds were successfully found for 14 metabolite concentrations. Upper bounds were identified for the remaining concentrations. Confidence intervals ranged between ±5 and ±25% except for pyruvate (±75%). The long nonstationary experiment consistently produced narrower confidence intervals than the short experiment. (Obviously, no metabolite concentrations could be obtained in the stationary experiment.)

Table III. External fluxes and free amino acid fragment MIDs measured in the simulated study of the large *E. coli* network.

Measurable external fluxes

Metabolite	Flux
Gluc	$Gluc_{ext} \rightarrow G6P$
Cit	$Cit_{ext} \rightarrow Cit$
CO_2	$CO_2 \rightarrow CO_{2,ext}$
Glyc	$Glyc_{ext} \rightarrow Glyc$
Ac	$Ac \rightarrow Ac_{ext}$
PDO	$PDO \rightarrow PDO_{ext}$
Biomass	Various metabolites → Biomass

Measurable GC/MS fragments

Metabolite	Mass	Carbons	Formula
AKG	346	12345	$C_{14}H_{28}O_5NSi_2$
Ala	232	23	$C_{10}H_{26}ONSi_2$
Ala	260	123	$C_{11}H_{26}O_2NSi_2$
Asp	302	12	$C_{14}H_{32}O_2NSi_2$
Asp	376	12	$C_{16}H_{38}O_3NSi_3$
Asp	390	234	$C_{17}H_{40}O_3NSi_3$
Asp	418	1234	$C_{18}H_{40}O_4NSi_3$
Cit	459	123456	$C_{20}H_{39}O_6Si_3$
Glu	330	2345	$C_{16}H_{36}O_2NSi_2$
Glu	432	12345	$C_{19}H_{42}O_4NSi_3$
Gly	218	2	$C_9H_{24}ONSi_2$
Gly	246	12	$C_{10}H_{24}O_2NSi_2$
Ile	200	23456	$C_{11}H_{26}NSi$
Ile	274	23456	$C_{13}H_{32}ONSi_2$
Leu	274	23456	$C_{13}H_{32}ONSi_2$
Mal	419	1234	$C_{18}H_{39}O_5Si_3$
Met	218	2345	$C_{10}H_{24}NSiS$
Met	292	2345	$C_{12}H_{30}ONSi_2S$
Met	320	12345	$C_{13}H_{30}O_2NSi_2S$
Phe	234	23456789	$C_{14}H_{24}NSi$
Phe	302	12	$C_{14}H_{32}O_2NSi_2$
Phe	308	23456789	$C_{16}H_{30}ONSi_2$
Phe	336	123456789	$C_{17}H_{30}O_2NSi_2$
Pyr	174	123	$C_6H_{12}O_3NSi$
Ser	288	23	$C_{14}H_{34}ONSi_2$
Ser	302	12	$C_{14}H_{32}O_2NSi_2$
Ser	362	23	$C_{16}H_{40}O_2NSi_3$
Suc	289	1234	$C_{12}H_{25}O_4Si_2$
Thr	376	234	$C_{17}H_{42}O_2NSi_3$
Thr	404	1234	$C_{18}H_{42}O_3NSi_3$
Tyr	302	12	$C_{14}H_{32}O_2NSi_2$
Val	260	2345	$C_{12}H_{30}ONSi_2$
Val	288	12345	$C_{13}H_{30}O_2NSi_2$

DOI 10.1002/bit

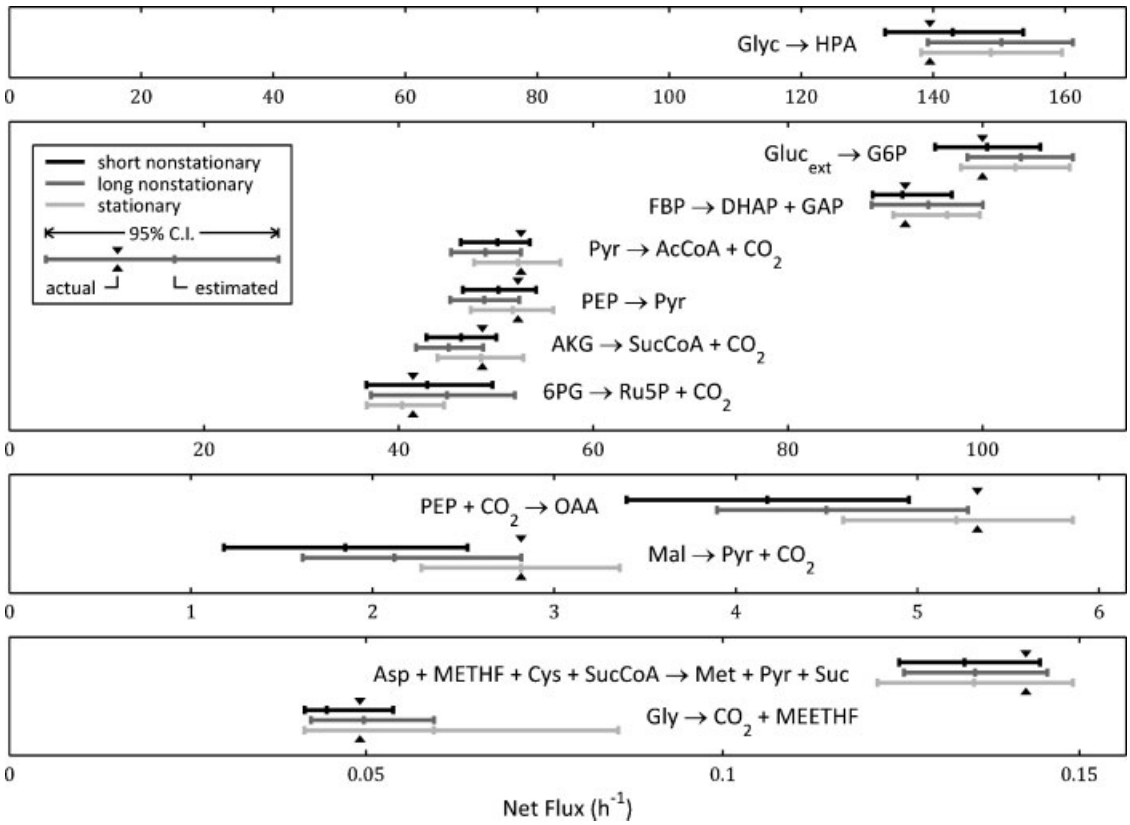

Figure 5. A comparison of estimated independent net fluxes in the large *E. coli* network using short nonstationary, long nonstationary, and stationary measurements. Actual flux values are indicated by black arrowheads. Accurate 95% confidence intervals are indicated by horizontal error bars while hash marks within these bars indicate estimated flux values. Values are in units of h^{-1} and are scaled such that the $Gluc_{ext} \rightarrow G6P$ flux is 100 h^{-1}.

Brown Adipocytes

EMU-based NMFA was lastly applied to actual experimental data in order to estimate fluxes. Brown adipocytes were cultured as described by Yoo et al. (2007). The adipocyte metabolic network is shown in Figure 8 (Antoniewicz et al., 2007c). (A detailed list of atom transitions for each reaction can be found in Table S-II of Supplementary Materials.) After cells were mature and at a metabolic steady state, the growth medium was replaced with assay medium containing [U-^{13}C]glucose and unlabeled glutamine. At 2, 4, and 6 h, cells were quenched and intracellular metabolites were extracted, derivatized, and analyzed with GC/MS. The metabolite fragments that were subsequently measured are listed in Table IV. Standard errors of at most 1.5 mol% were assumed for all GC/MS measurements.

Over the three time points we collected 249 mass isotopomer abundances while the model attempted to fit 12 free fluxes and 14 metabolite concentrations. Hence, the system possessed $249-12-14=223$ redundant measurements and the expected lower and upper bounds of the 95% confidence region were 183 and 266, assuming that the minimized sum of squared residuals in Equation (5) follows a χ^2 distribution. A nonstationary flux estimation was conducted to fit the model to these measurements. The resulting minimized sum of squared residuals was 251, indicating that the fit was statistically acceptable. The estimated labeling profiles are shown together with the measurements in Figure 9. Net and exchange flux estimates are presented in Figure 10. These parameters' estimated numerical values along with their accurate 95% confidence intervals are listed in Table V. Exchange fluxes are calculated according to Equation (6), where v_{ref} is the glucose uptake flux.

Discussion

Our experiments involving the large *E. coli* model demonstrated that metabolic parameters can be successfully estimated using only external flux measurements and nonstationary labeling measurements. We were able to estimate all net fluxes, most exchange fluxes, and some metabolite concentrations with no need for concentration measurements or stationary flux measurements. This same theme was seen in the adipocyte experiment; with only

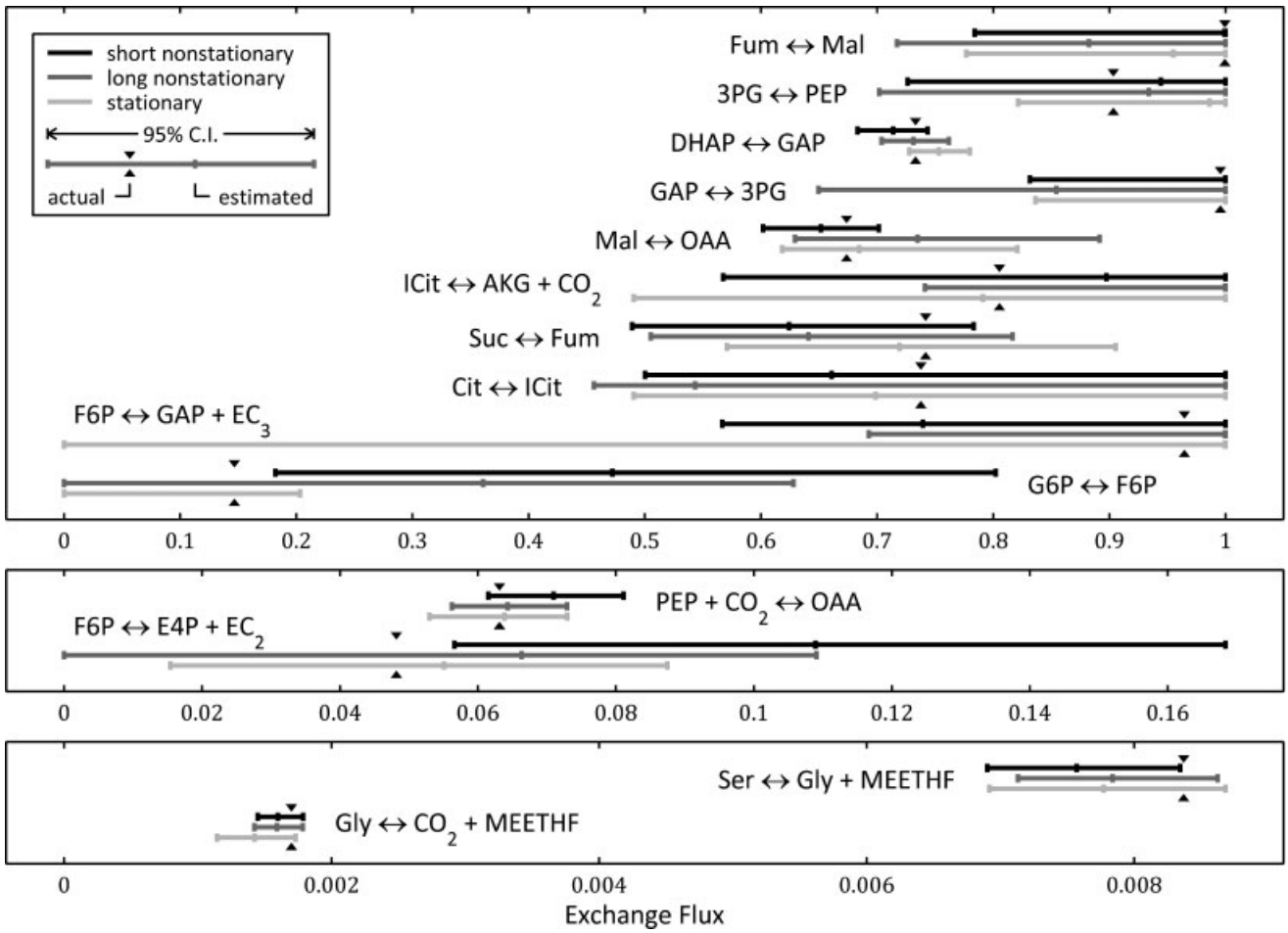

Figure 6. A comparison of estimated exchange fluxes in the large *E. coli* network using short nonstationary, long nonstationary, and stationary measurements. Actual flux values are indicated by black arrowheads. Accurate 95% confidence intervals are indicated by horizontal error bars while hash marks within these bars indicate estimated flux values. Values are dimensionless and are scaled according to Eq. (6). Unidentifiable and nearly unidentifiable fluxes were omitted.

labeling data taken during the nonstationary regime, all net fluxes, some exchange fluxes, and three metabolite concentrations were estimated. In this experiment, parameters were less identifiable because only three measurement time points were available and no external fluxes were measured.

The simulated *E. coli* experiments do indicate that there are some limitations when relying only upon nonstationary data. Confidence intervals produced from the short nonstationary measurements were the least likely to recapture original parameter values. Across the three experiments, only five of the original parameter values fell significantly outside of their respective 95% confidence intervals and all five of these intervals were generated from the short nonstationary experiment. Intervals from the long nonstationary and stationary experiments, however, sucessfully recovered the actual values.

Previous research has indicated that more flux information may be available when drawing measurements from the isotopically nonstationary regime of a labeling experiment, leading to more precise confidence intervals (Nöh et al., 2006). Our analysis of the large *E. coli* network demonstrated that, in general, the nonlinear confidence intervals of almost all parameters estimated in both the short and long nonstationary experiments were comparable to those of the stationary experiment. This indicates that the gains due to additional sensitivity in the nonstationary measurements are most likely offset by the increased number of total parameters that need to be estimated by NMFA. There are three likely reasons for the discrepancy between our results and prior findings:

- Most importantly, the total number of measurements were not standardized across all conditions in Nöh's study. Confidence intervals for the stationary case relied upon only one set of measurements whereas the nonstationary cases with multiple time points comprised multiple sets. To ensure a fair comparison, we used

DOI 10.1002/bit

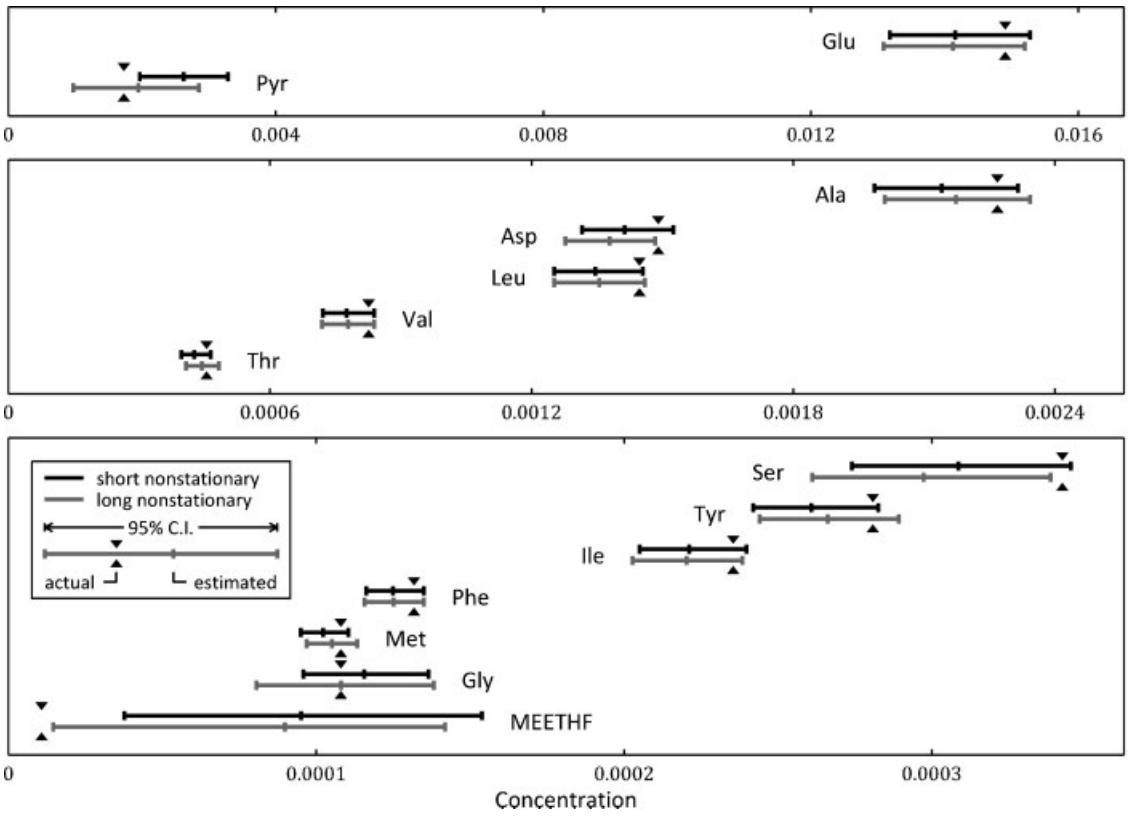

Figure 7. A comparison of estimated metabolite concentrations in the large *E. coli* network using short nonstationary and long nonstationary measurements. Actual concentration values are indicated by black arrowheads. Accurate 95% confidence intervals are indicated by horizontal error bars while hash marks within these bars indicate estimated concentration values. Values are dimensionless and are scaled such that the $Gluc_{ext} \rightarrow$ G6P flux is 100 h^{-1}. Concentrations with lower bounds of zero were omitted.

replicate measurements so that the total quantity of measurements was equal in every simulated experiment.

- Our rigorous and accurate determination of nonlinear confidence intervals revealed that the linearized standard errors employed in previous studies are unreliable and often much smaller than the true parameter uncertainty.
- Our nonstationary estimations did not include concentration measurements while Nöh's estimations did. These additional measurements also play a role in the smaller standard errors predicted by Nöh.

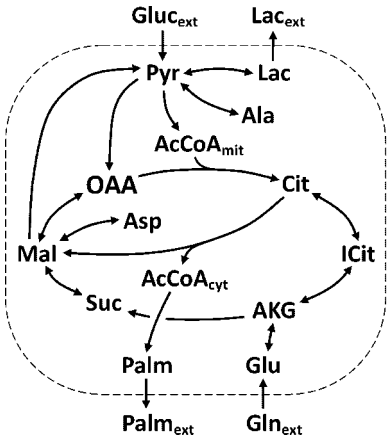

Figure 8. A simplified model of brown adipocyte metabolism. (Atom transitions can be found in Table S-II of Supplementary Materials.).

Table IV. Metabolite fragment MIDs measured in the brown adipocyte study.

Measurable GC/MS fragments			
Metabolite	Mass	Carbons	Formula
Ala	232	23	$C_{10}H_{26}ONSi_2$
Ala	260	123	$C_{11}H_{26}O_2NSi_2$
Asp	316	234	$C_{15}H_{34}O_2NSi_2$
Asp	418	1234	$C_{18}H_{40}O_4NSi_3$
Cit	459	123456	$C_{20}H_{39}O_6Si_3$
Glu	330	2345	$C_{16}H_{36}O_2NSi_3$
Glu	432	12345	$C_{19}H_{42}O_4NSi_3$
Lac	233	23	$C_{10}H_{25}O_2Si_2$
Lac	261	123	$C_{11}H_{25}O_3Si_2$
Mal	419	1234	$C_{18}H_{39}O_5Si_3$
Palm	270	1–16	$C_{17}H_{34}O_2$
Pyr	174	123	$C_6H_{12}O_3NSi$

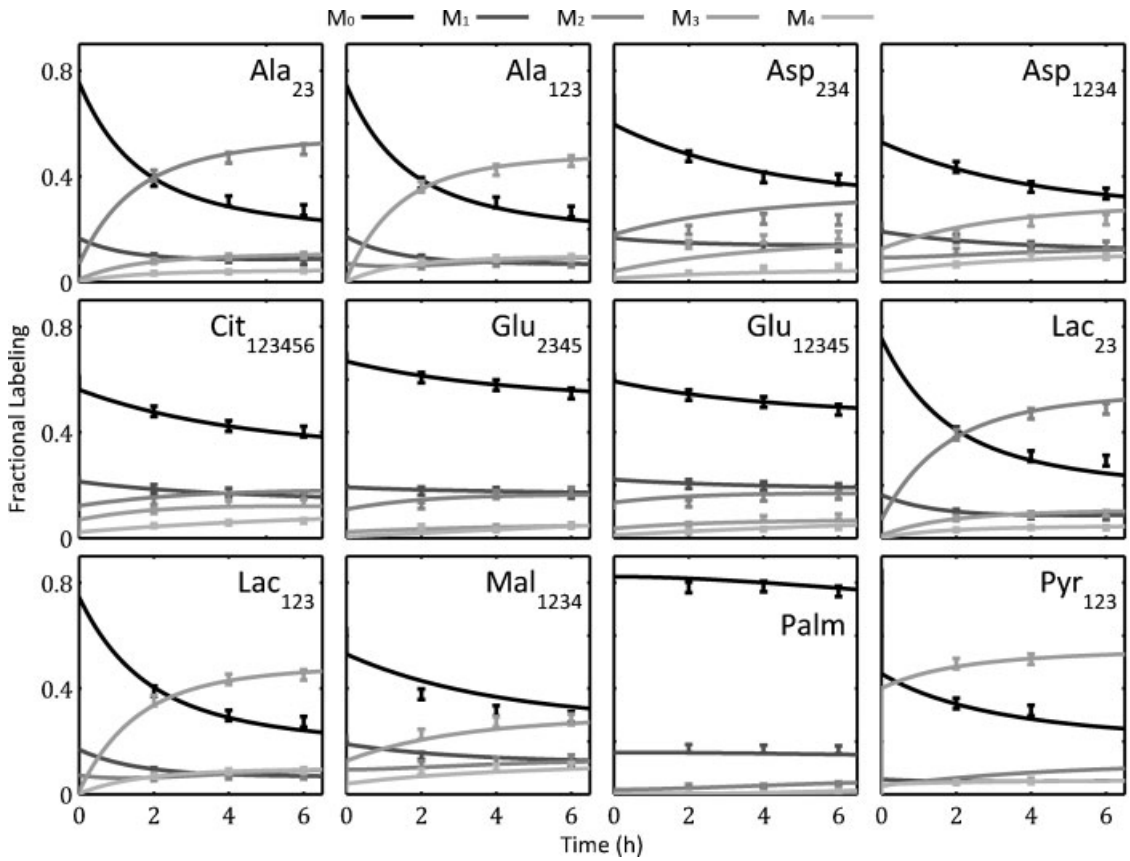

Figure 9. Fitted MIDs versus time for the measured metabolite fragments overlaid with actual measurements from the nonstationary brown adipocyte experiment. Error bars indicate 95% confidence intervals for the measurements.

Both the *E. coli* and adipocyte experiments confirm that NMFA can yield important information regarding metabolite concentrations even when those concentrations are not directly measured. Admittedly, most metabolite concentrations in both the large *E. coli* and brown adipocyte estimations are unidentifiable. The reason is twofold. First, the time intervals between measurements are too large for these pools to influence observable changes in labeling; that is, these pools appear to be at a pseudo-steady state relative to the sampling time scale. Second, because labeling is measured in only a subset of metabolites, there is considerable ambiguity with regard to the concentrations of unmeasured metabolites. At the very least, upper bounds can be determined for all concentrations. We were able to estimate both upper and lower bounds for metabolites that (1) were at high enough concentrations to influence the labeling dynamics on a time scale similar to the sampling time scale and (2) were the targets of labeling measurements or were "sandwiched" between other metabolites where labeling measurements were available. In the *E. coli* experiment, the concentrations of almost all metabolites whose labeling was measured could be estimated with a 95% confidence interval on the order of $\pm 10\%$ of the estimated value. The adipocyte experiment generated three concentration estimates with nonzero lower bounds. However, these confidence intervals were much less informative (between ± 25 and $\pm 55\%$ of the estimated value), most likely because labeling was measured at only three widely spaced time points as opposed to 15 closely spaced time points in the short nonstationary *E. coli* case.

Several factors frame the brown adipocyte system as an ideal subject for nonstationary analysis. Studying the estimated parameters in the system, we find that the measured metabolites do not reach a stationary labeling state (99% of the steady-state value) until between 35 and 55 h, except for palmitate, which requires 250 h due to its slow turnover in these cells. In such a case, the utility of nonstationary flux analysis becomes quite apparent. Flux estimates can be obtained in only a fraction of the experimental time, leading to large cost and time savings. Moreover, even if money and time were not limiting factors, it is highly unlikely that the adipocytes could be kept metabolically or phenotypically stable for such lengthy durations, making NMFA not just a convenient tool in this situation, but an essential one.

Finally, we have shown that application of the EMU framework to NMFA results in dramatic improvements in network decomposition and parameter estimation. These

DOI 10.1002/bit

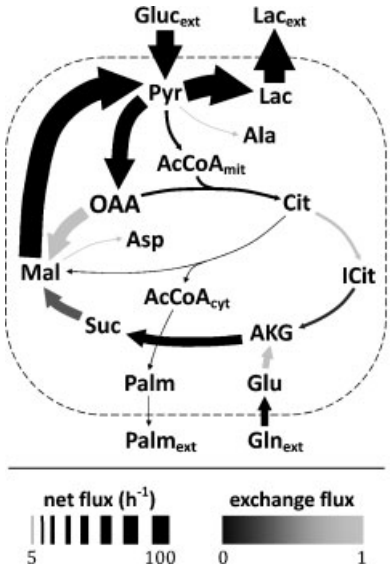

Figure 10. Visualization of flux estimates for nonstationary flux analysis of brown adipocytes. Net flux directionality is indicated by arrows and net flux magnitude is proportional to line thickness while exchange flux magnitude correlates to arrow color. Net flux values are in units of h^{-1} and have been scaled such that the ½ $Gluc_{ext} \rightarrow$ Pyr flux is 100 h^{-1}. Exchange flux values are dimensionless and have been scaled according to Eq. (6).

Table V. Estimated net fluxes (→), exchange fluxes (↔), and concentrations (pool) and their respective 95% confidence intervals for brown adipocyte metabolism.

Parameter	Value	Interval
Pyr → $AcCoA_{mit}$	17.0	[14.3, 22.1]
Suc → Mal	54.0	[45.6, 60.4]
Mal → OAA	−73.1	[−92.4, −55.8]
$AcCoA_{cyt}$ → ⅛ Palm	8.6e-3	[0, 11.9]
Pyr ↔ Lac	0.25	[0, 0.49]
Pyr ↔ Ala	0.00	[0, 0.46]
Cit ↔ ICit	1.00	[0.18, 1]
Glu ↔ AKG	1.00	[0.77, 1]
ICit ↔ AKG	0.23	[0.18, 1]
Suc ↔ Mal	0.53	[0.45, 0.60]
Mal ↔ OAA	1.00	[0.80, 1]
Mal ↔ Asp	0.07	[0, 1]
$AcCoA_{cyt}$ pool	1.0e-5	[0, 6.1]
$AcCoA_{mit}$ pool	126.4	[82.1, 189.9]
AKG pool	3.6e-4	[0, 23.7]
Ala pool	2.2e-3	[0, 75.0]
Asp pool	4.0e-5	[0, 40.0]
Cit pool	2.2e-4	[0, 13.0]
Glu pool	3.6e-4	[0, 23.7]
ICit pool	2.1e-4	[0, 13.0]
Lac pool	178.4	[112.9, 225.6]
Mal pool	8.1e-4	[0, 40.0]
OAA pool	6.7e-4	[0, 40.0]
Palm pool	6.3e-2	[0, 90.3]
Pyr pool	2.5e-3	[0, 64.3]
Suc pool	208.1	[111.6, 323.8]

Concentrations are dimensionless and net fluxes have units of h^{-1}; both have been scaled such that the ½ $Gluc_{ext} \rightarrow$ Pyr flux is 100 h^{-1}. Exchange fluxes are dimensionless and scaled according to Equation (6).

advances make feasible an entirely new realm of problems in nonstationary flux analysis. Previously, for instance, the analysis of systems with complicated reaction networks, multiple isotopic tracers, or large molecules were impractical targets for NMFA. By shifting to an EMU framework, these kinds of problems are now solvable. The EMU framework also makes possible the calculation of accurate confidence intervals for parameters estimated by NMFA, a computationally intensive exercise that otherwise would be infeasible.

Nomenclature

3PG	3-phosphoglycerate
6PG	6-phosphogluconate
Ac	acetate
AcCoA	acetyl coenzyme A
AKG	α-ketoglutarate
Ala	alanine
Arg	arginine
Asn	asparagine
Asp	aspartate
Cit	citrate
CO_2	carbon dioxide
Cys	cystine
Cyt	cytosolic
DHAP	dihydroxyacetone phosphate
E4P	erythrose 4-phosphate
EC_2	transketolase 2-carbon intermediate
EC_3	transaldolase 3-carbon intermediate
EMU	elementary metabolite unit
Ext	external
F6P	fructose 6-phosphate
FBP	fructose 1,6-biphosphate
FTHF	formyltetrahydrofolate
Fum	fumarate
G6P	glucose 6-phosphate
GAP	glyceraldehyde 3-phosphate
GC/MS	gas chromatography/mass spectroscopy
Gln	glutamine
Glu	glutamate
Gluc	glucose
Gly	glycine
Glyc	glycerol
Glyc3P	glycerol 3-phosphate
HPA	3-hydroxypropionaldehyde
ICit	isocitrate
Ile	isoleucine
KDPG	2-keto-3-deoxy-6-phosphogluconate
Lac	lactate
Leu	leucine
LL-DAP	LL-diaminopimelate
Mal	malate
MEETHF	methylenetetrahydrofolate
Met	methionine
METHF	methyltetrahydrofolate

MFA	metabolic flux analysis
Mit	mitochondrial
NMFA	nonstationary metabolic flux analysis
OAA	oxaloacetate
Palm	palmitate
PDO	1,3-propanediol
PEP	phosphoenolpyruvate
Phe	phenylalanine
Pro	proline
Pyr	pyruvate
R5P	ribose 5-phosphate
Ru5P	ribulose 5-phosphate
S7P	seduheptulose 7-phosphate
Ser	serine
Suc	succinate
SucCoA	succinyl coenzyme A
Thr	threonine
Tyr	tyrosine
Val	valine
X5P	xylulose 5-phosphate

This study was supported by a Graduate Research Fellowship from the National Science Foundation (JLW). The authors would like to thank Prof. Joanne K. Kelleher for her valuable advice and insights regarding brown adipocyte metabolism.

Appendix

Customized Differential Equation Solver

EMU labeling states and sensitivities are described by Equations (1) and (4) and can be simplified as shown below:

$$\frac{d\mathbf{X}_n}{dt} = \mathbf{F}_n \cdot \mathbf{X}_n + \mathbf{G}_n \tag{A-1}$$

$$\frac{d}{dt}\frac{\partial \mathbf{X}_n}{\partial \mathbf{p}} = \mathbf{F}_n \cdot \frac{\partial \mathbf{X}_n}{\partial \mathbf{p}} + \mathbf{H}_n \tag{A-2}$$

by making use of the following substitutions:

$$\mathbf{F}_n = \mathbf{C}_n^{-1} \cdot \mathbf{A}_n \tag{A-3}$$

$$\mathbf{G}_n = \mathbf{C}_n^{-1} \cdot \mathbf{B}_n \cdot \mathbf{Y}_n \tag{A-4}$$

$$\mathbf{H}_n = \frac{\partial \mathbf{F}_n}{\partial \mathbf{p}} \cdot \mathbf{X}_n + \frac{\partial \mathbf{G}_n}{\partial \mathbf{p}} \tag{A-5}$$

The functions $\mathbf{G}_n$ and $\mathbf{H}_n$ potentially comprise convolutions of MIDs belonging to EMUs of previously solved blocks (a result of EMU condensation reactions) and as such Equations (A-1) and (A-2) lack analytical solutions. Partial analytical solutions, however, can be written as

$$\mathbf{X}_n(t_1) = \mathbf{e}^{\mathbf{F}_n \cdot \Delta t} \cdot \mathbf{X}_n(t_0) + \int_0^{\Delta t} \mathbf{e}^{\mathbf{F}_n \cdot (\Delta t - \tau)} \cdot \mathbf{G}_n(\tau + t_0) \cdot d\tau \tag{A-6}$$

$$\left.\frac{\partial \mathbf{X}_n}{\partial \mathbf{p}}\right|_{t_1} = \mathbf{e}^{\mathbf{F}_n \cdot \Delta t} \cdot \left.\frac{\partial \mathbf{X}_n}{\partial \mathbf{p}}\right|_{t_0} + \int_0^{\Delta t} \mathbf{e}^{\mathbf{F}_n \cdot (\Delta t - \tau)} \cdot \mathbf{H}_n(\tau + t_0) \cdot d\tau \tag{A-7}$$

where the initial state of the system at time t_0 is assumed to be known and Δt is defined as $t_1 - t_0$. Again, the integrals in Equations (A-6) and (A-7) lack analytical solutions. Instead, we evaluate at discrete points by applying a first-order-hold equivalent with adaptive step size control to numerically integrate and solve the problem (Powell et al., 1990). This discretized approximation can be expressed as follows:

$$\begin{aligned}(\mathbf{X}_n)_{k+1} = {} & \Phi_n \cdot (\mathbf{X}_n)_k + \Gamma_n \cdot (\mathbf{G}_n)_k \\ & + \Omega_n \cdot [(\mathbf{G}_n)_{k+1} - (\mathbf{G}_n)_k]\end{aligned} \tag{A-8}$$

$$\begin{aligned}\left(\frac{\partial \mathbf{X}_n}{\partial \mathbf{p}}\right)_{k+1} = {} & \Phi_n \cdot \left(\frac{\partial \mathbf{X}_n}{\partial \mathbf{p}}\right)_k + \Gamma_n \cdot (\mathbf{H}_n)_k \\ & + \Omega_n \cdot [(\mathbf{H}_n)_{k+1} - (\mathbf{H}_n)_k]\end{aligned} \tag{A-9}$$

where the transition matrices $\mathbf{\Phi}_n$, $\mathbf{\Gamma}_n$, and $\mathbf{\Omega}_n$ are functions of fluxes, concentrations, and the time step magnitude according to the following relationship:

$$\begin{bmatrix} \mathbf{\Phi}_n & \mathbf{\Gamma}_n & \mathbf{\Omega}_n \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \exp\left(\begin{bmatrix} \mathbf{F}_n \cdot \Delta t & \mathbf{I} \cdot \Delta t & 0 \\ 0 & 0 & \mathbf{I} \\ 0 & 0 & 0 \end{bmatrix}\right) \tag{A-10}$$

where the exponential function refers to the matrix exponential (Golub and Van Loan, 1996). At each time point, $\mathbf{Y}_n$s and $\mathbf{X}_n$s are calculated in ascending order until the EMUs representing all desired measurements are obtained.

References

Antoniewicz MR, Kelleher JK, Stephanopoulos G. 2006. Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements. Metab Eng 8(4):324–337.

Antoniewicz MR, Kelleher JK, Stephanopoulos G. 2007a. Elementary metabolite units (EMU): A novel framework for modeling isotopic distributions. Metab Eng 9(1):68–86.

Antoniewicz MR, Kraynie DF, Laffend LA, González-Lergier J, Kelleher JK, Stephanopoulos G. 2007b. Metabolic flux analysis in a

DOI 10.1002/bit

nonstationary system: Fed-batch fermentation of a high yielding strain of *E. coli* producing 1,3-propanediol. Metab Eng 9(3):277–292.

Antoniewicz MR, Yoo H, Kelleher JK, Stephanopoulos G. 2007c. Global flux quantification and statistical analysis in stable-isotope studies quantifies effects of IRS1 knockout in brown adipose cells. J Biol Chem In Submission.

Buchholz A, Takors R, Wandrey C. 2001. Quantification of intracellular metabolites in *Escherichia coli* K12 using liquid chromatographic-electrospray ionization tandem mass spectrometric techniques. Anal Biochem 295(2):129–137.

Dulmage AL, Mendelsohn NS. 1958. Coverings of bipartite graphs. Canad J Math 10:517–534.

Gill PE, Murray W, Wright MH. 1981. Practical optimization. London: Academic Press.

Golub GH, Van Loan CF. 1996. Matrix computations. Baltimore: Johns Hopkins University Press.

Gombert AK, Moreira dos Santos M, Christensen B, Nielsen J. 2001. Network identification and flux quantification in the central metabolism of *Saccharomyces cerevisiae* under different conditions of glucose repression. J Bacteriol 183(4):1441–1451.

Kiefer P, Heinzle E, Zelder O, Wittmann C. 2004. Comparative metabolic flux analysis of lysine-producing *Corynebacterium glutamicum* cultured on glucose or fructose. Appl Environ Microb 70(1):229–239.

Klapa MI, Aon JC, Stephanopoulos G. 2003. Systematic quantification of complex metabolic flux networks using stable isotopes and mass spectrometry. Eur J Biochem 270(17):3525–3542.

Madsen K, Nielsen HB, Tingleff O. 2004. Methods for non-linear least squares problems. 2nd edition. Lecture notes, Technical University of Denmark. http://www2.imm.dtu.dk/pubdb/p.php?3215.

Nöh K, Wahl A, Wiechert W. 2006. Computational tools for isotopically instationary ^{13}C labeling experiments under metabolic steady state conditions. Metab Eng 8(6):554–577.

Nöh K, Grönke K, Luo B, Takors R, Oldiges M, Wiechert W. 2007. Metabolic flux analysis at ultra short time scale: Isotopically non-stationary ^{13}C labeling experiments. J Biotechnol 129(2):249–267.

Pothen A, Fan CJ. 1990. Computing the block triangular form of a sparse matrix. T Math Soft 16(4):303–324.

Powell JD, Franklin GF, Workman ML. 1990. Digital control of dynamic systems. Reading, MA: Addison-Wesley.

Sauer U, Hatzimanikatis V, Bailey JE, Hochuli M, Szyperski T, Wüthrich K. 1997. Metabolic fluxes in riboflavin-producing *Bacillus subtilis*. Nat Biotechnol 15(5):448–452.

Shastri AA, Morgan JA. 2007. A transient isotopic labeling methodology for ^{13}C metabolic flux analysis of photoautotrophic microorganisms. Phytochemistry 68:2302–2312.

Wiechert W. 2001. ^{13}C Metabolic flux analysis. Metab Eng 3(3):195–206.

Wiechert W, de Graaf AA. 1997. Bidirectional reaction steps in metabolic networks: I. Modeling and simulation of carbon isotope labeling experiments. Biotechnol Bioeng 55(1):101–117.

Wiechert W, Möllney M, Isermann N, Wurzel M, de Graaf AA. 1999. Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems. Biotechnol Bioeng 66(2):69–85.

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 Microb 68(12):5843–5859.

Yoo H, Antoniewicz MR, Stephanopoulos G, Kelleher JK. 2007. Quantifying reductive carboxylation flux of glutamine to lipid in a brown adipocyte cell line. J Biol Chem In Submission.