ABSTRACT

Title of Document:	INVESTIGATING THE METABOLIC LANDSCAPE ALTERATIONS IN POPLAR CELLS INDUCED BY CARBON AND NITROGEN DEFICIENCY VIA IMPROVED ^{13}C METABOLIC FLUX ANALYSIS METHODOLOGY
	Xiaofeng Zhang, Doctor of Philosophy, 2015
Directed By:	Ganesh Sriram, Associate Professor Department of Chemical and Biomolecular Engineering

Plants are considered biological factories with their ability of converting solar energy into chemical energy in the form of various commercially valuable products, such as food, biofuel and pharmaceuticals. The yields of these products are directly influenced by the level of nitrogen nutrient supply. However, both biological and industrial nitrogen fixation are energetically expensive and thus managing the nitrogen cycle has been identified as one of the 14 grand challenges by the National Academy of Engineering (NAE). Therefore it is desirable to investigate how plants themselves adapt to nitrogen deficient environment and improve their nitrogen use efficiency (NUE). A powerful tool to study metabolism is isotope-assisted metabolic flux analysis (isotopic MFA), which quantifies intracellular chemical reaction rates (fluxes) via isotopic labeling experiments (ILEs) and subsequent mathematical modeling. In ILEs the labeling patterns of the metabolites can be measured at either

isotopic steady state or isotopic instationary state. Between these two methods, collecting data during isotopic instationary state saves experimental time, but is computationally more challenging due to that instationary MFA involves solving ordinary differential equations (ODEs). In this study, we firstly developed an approach that combined the concept of "originomer" with an analytical based solution method to improve computational efficiency of instationary MFA. Simulation results showed that this approach reduced computational time by 23-fold for certain realistic metabolic network. Secondly, we managed to solve an intrinsic problem that affect steady state MFA in fed-batch cell culture environment – the influence of unlabeled biomass that are present before applying isotopic tracers in an ILE. We proposed a full "reflux" metabolic network model that significantly improved the accuracy of evaluated fluxes when compared to the models without "reflux". Finally, we investigated the ability of adapting nutrient deficiencies and the NUE-improving mechanisms in suspension cells of poplar, a woody perennial tree capable of efficiently managing its nitrogen reserves. Through (i) steady-state ^{13}C MFA and (ii) transcripomic profiling via microarray on poplar cells growing under different carbon (C) and nitrogen (N) supply levels, we found a plastidic localization of oxidative pentose phosphate pathway (oxPPP), as well as a lower oxPPP flux under low nitrogen supply. Gene expression data also points to possible NUE improving mechanisms employed by poplar cells. We hope this study will shed light on potential metabolic engineering directions to improve NUE in plants.

INVESTIGATING THE METABOLIC LANDSCAPE ALTERATIONS IN POPLAR CELLS INDUCED BY CARBON AND NITROGEN DEFICIENCY VIA IMPROVED ^{13}C METABOLIC FLUX ANALYSIS METHODOLOGY

By

Xiaofeng Zhang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctorate of philosophy
2015

Advisory Committee:
Associate Professor Ganesh Sriram, Dept. of Chemical and Biomolecular Engineering, Committee Chair
Associate Professor Gary Coleman, Dept. of Plant Science and Landscape Architecture
Associate Professor Panagiotis Dimitrakopoulos, Dept. of Chemical and Biomolecular Engineering
Assistant Professor Amy Karlsson, Dept. of Chemical and Biomolecular Engineering
Professor Zhongchi Liu, Dept. of Cell Biology and Molecular Genetics

Dedication

I dedicate this work to my parents for their unconditional, endless love and support.

Acknowledgements

Firstly and foremost, I would like to thank my advisor, **Dr. Ganesh Sriram**, for his excellent guidance throughout my Ph.D. study, his encouragement when I faced difficulties in my research work and his help in my career development.

I thank **Dr. Gary Coleman** for his guidance and suggestions in my poplar projects.
I thank **Dr. Rongshuang Lin** for her help and tutorial on RNA extraction technique and microarray data processing.

I acknowledge my funding source for this work, **the A. James Clark Distinguished Fellowship, the Maryland Industrial Partnerships, the Minta Martin Foundation** and **the National Science Foundation (IOS – 0922650 and CBET - 1134115)**.

I thank my co-workers in the lab, **Shilpa Nargund, Ashish Misra, Yuting Zheng, Andrew Quinn, Navadeep Boruah** and all the undergraduate researchers for their help, suggestions and the creation of a friendly, productive atmosphere in the lab.

Finally, I wish to thank my Ph.D. proposal and defense committee members, **Dr. William E. Bentley, Dr. Gary Coleman, Dr. Panagiotis Dimitrakopoulos, Dr. Amy Karlsson, Dr. Zhongchi Liu, Dr. Srinivasa Raghavan** for their valuable comments on my thesis work.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1. Background and Motivation

Plants are biological factories with their ability of converting solar energy into chemical energy in the form of various commercially valuable products. The yields of these products are directly influenced by the availability of carbon and nitrogen nutrients. The atmosphere primarily consists of nitrogen gas (N_2, ~78%), oxygen gas (O_2, ~21%), carbon dioxide (CO_2, ~0.039%) and small amount of other gases. Although plants are capable of fixing atmospheric CO_2 through photosynthesis, most plants are not able to directly utilize N_2 for biomass production. The reduction of N_2 to ammonia (NH_4^+), the nitrogen nutrient form that can be assimilated by plants, is accomplished either industrially or biologically. However, since the bond energy of N_2 is as high as 942 KJ/mol, nitrogen fixation is energetically expensive. For example, the industrial N_2 fixation to produce fertilizers, also known as the "Haber process", requires reactors operating at a temperature of 400-500 °C and a pressure of 150-250 atm, with a relatively low yield of ~15% in each reactor (Marschner, 2012). Biologically, fixing one mole of N_2 demands the energy equivalent to the hydrolysis of 16 moles of ATP (Lehninger, 2005). These energetic drawbacks, combined with the detrimental influence on the environment of using chemically synthesized fertilizers, makes it crucial to understand how plants themselves adapt to nitrogen deficient environment and how to improve their NUE. It has been estimated that an increase of 1% of the NUE by crops leads to annual savings of ~234 million dollars in fertilizer production (Raun and Johnson, 1999).

Wooden perennial trees are generally capable of managing their nitrogen cycle efficiently. Poplar, in particular, is able to recycle around 80% of its nitrogen reserves from senescing leaves during fall and store them in the form of bark storage proteins (BSPs) (Cantón et al., 2005; Pregitzer et al., 1990; Zhu and Coleman, 2001). There stored nitrogen can be remobilized in spring for *de novo* biomass synthesis. This efficient nitrogen recycling mechanism reduces the consumption of ATPs and reductants needed to assimilate nitrogen nutrient from the environment (Nunes-Nesi et al., 2010) and thus promote growth of the plant. In fact, poplar is one of the most fast growing perennial trees in the world (Stanturf et al., 2001), which makes it an ideal crop for various applications including biofuel and paper production (Sannigrahi et al., 2010). Together with the fact that the full genome of poplar has been sequenced (Tuskan et al., 2006), studying metabolism in poplar cells is of particular interest.

Despite the importance of nitrogen as a nutrient, it is currently extremely difficult to predict the relationship between nitrogen deficiency and the concomitant poor performance of plants (Amtmann and Armengaud, 2009). This is due to the complexity of the metabolic networks underlying nitrogen metabolism and the inextricable links between carbon and nitrogen metabolism. Both carbon and nitrogen are important constituting elements of enzymes and cofactors needed for various cellular activities such as photosynthesis, respiration and biomass synthesis (Gao et al., 2008; Kruger and von Schaewen, 2003). The availability of carbon nutrient directly affect the production rate of ATP and NADPH required for nitrogen

assimilation (Allen and Young, 2013). Conversely, the nitrogen availability has been shown to affect the abundances of enzyme that catalyze the reactions in respiratory bypass pathways (Escobar et al., 2006). Besides, multiple studies have shown that the carbon and nitrogen balance regulate gene expressions. For example, through transcriptomic profiling Gutiérrez *et al.* (2007) found that carbon and nitrogen interactions affect expression levels of various genes encoding central carbon metabolism and protein catabolic pathways in *Arabidopsis*. Several other studies concluded that the carbon and nitrogen balance affect the expression levels of a putative nitrogen transporter (Little et al., 2005), a putative glutamate receptor (Kang and Turano, 2003) and a putative methyltransferase (Gao et al., 2008, p. 1) in *Arabidopsis.* Therefore comprehensive studies integrating carbon and nitrogen metabolism are desirable.

Previous work on understanding plant metabolism under different nitrogen sources or provisions are mainly focused on metabolomics, proteomics or gene expression data (Allen et al., 2011; Hockin et al., 2011; Noorhana, 2011; Truong et al., 2013; Zhila et al., 2005). Recently, researchers have started applying ^{13}C isotope-assisted MFA tool to study the influence of nitrogen source on central carbon metabolism in plants at a molecular level (Allen and Young, 2013; Masakapalli et al., 2013). For example, Masakapalli *et al*. (2013) found that by replacing ammonium with nitrate in the culturing media, the oxidative pentose phosphate pathway flux increased by 50% in *Arabidopsis* cells. Allen and Young's study (2013) on soybean embryos showed that reducing the ratio of carbon and nitrogen from 37:1 to 13:1 led

to a 90% increase in downstream TCA cycle flux, which promoted the contribution of malic enzyme to pyruvate production.

1.2. Application and recent development of metabolic flux analysis

MFA is the major computational and analytical method applied in this work. It is a vital tool in metabolic engineering due to its ability to quantify the relative rates of chemical reactions in an intracellular metabolic network (Ahn and Antoniewicz, 2011; Kruger and Ratcliffe, 2009; Leighty and Antoniewicz, 2012; Schwender, 2008). MFA is also capable of elucidating previously unknown pathways (e.g. Schwender et al., 2004; Sriram et al., 2007a). Flux values obtained from MFA, if combined with metabolite concentration data, constitute a minimal set of information needed to fully characterize metabolism (Stephanopoulos, 1999).

Besides, flux data can reflect the effect of genetic or environmental perturbations and thus identify metabolic engineering strategies (Sauer, 2006; Stephanopoulos, 1999; Stephanopoulos and Vallino, 1991). Many recent studies apply metabolic flux analysis for strain development. For examples, via ^{13}C MFA several articles reported the redox (NADPH/NADH) imbalance was the major bottleneck for the biosynthesis of long chain fatty acids (He et al., 2014; Ranganathan et al., 2012; Wasylenko et al., 2015). They proposed different strategies to resolve the deficiency of the reductants, including the diversion of fluxes from glycolysis to serine synthesis pathway (Ranganathan et al., 2012) and Entner–Doudoroff pathway

(He et al., 2014; Ranganathan et al., 2012), or a more active pentose phosphate pathway (He et al., 2014; Wasylenko et al., 2015).

The direct role of MFA is to quantify fluxes through certain intracellular and extracellular measurements. Two major approaches in MFA are flux balance analysis (FBA) and isotope-assisted MFA. Typical FBA only requires extracellular measurements such as uptake rates of nutrient sources and secretion rates of biomass (Orth et al., 2010; Ramakrishna et al., 2000; Varma and Palsson, 1994). It involves balancing the abundances of all the metabolites in the metabolic network (details are discussed in **Sec. 1.2.3**), which leads to a system of linear algebraic equations. Objection functions such as maximization of the growth rate of the organism, or maximization of the secretion rate of certain cellular products, are required to solve for fluxes. However, the assumed objection functions are sometimes inaccurate, which makes the flux estimation results biased (Chen et al., 2011; Wu et al., 2015). **Fig. 1.1** shows a phenotype space representation of the FBA method. The surfaces of the polyhedron represent thermodynamic or artificial constraints. Any point in the polyhedron region represents a theoretically feasible flux solution, with the apexes defining the optimal solution based on an objection function.

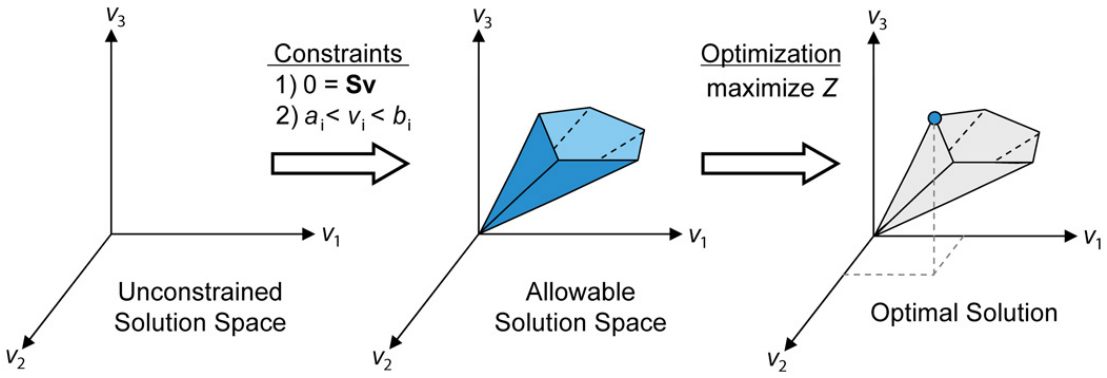


Figure 1.1. A typical phenotype space of a metabolic network.
The three dimensional axis represent three independent fluxes. The flux phenotype space can be represented by a polyhedron. Each surface represents a biological constraint (e.g. the thermodynamic irreversibility or feasible range of a flux). Any flux distribution that falls within the phenotype space is theoretically feasible. The apexes represent the metabolic state where certain flux is maximized.
Figure retrieved from (Orth et al., 2010)

Isotope-assisted MFA, specifically ^{13}C MFA, is another approach that can provide more intracellular details than FBA. In general, ^{13}C MFA detects the labeling response of intermediates of a metabolic network to certain types of labeled substrate (Wiechert, 2001) in an ILE. It solves for fluxes through a mathematical model that correlates this labeling response to fluxes. This approach is becoming very powerful with the widespread application of gas chromatography-mass spectrometry (GC-MS) (Ahn and Antoniewicz, 2011; Leighty and Antoniewicz, 2012; e.g. Wittmann and Heinzle, 1999), the development of tandem MS (Choi and Antoniewicz, 2011) and nuclear magnetic resonance (NMR) techniques (Goudar et al., 2010; Sriram et al., 2007a; Szyperski, 1995).

The labeling patterns of intracellular metabolites can be measured either at isotopically steady-date or isotopically instationary state. At isotopically instationary

state, the labeling patterns of metabolites evolve with time, but fluxes and metabolite concentrations (pool sizes) are constant. The transient labeling states of metabolites are then used to evaluate fluxes via instationary MFA (iMFA) (Nöh et al., 2006; Wiechert and Noh, 2005; Young et al., 2008). iMFA is different from steady-state MFA on two major counts. (i) Experimentally, pool size data and multiple sampling through the isotopic instationary state are required (Nöh et al., 2006; Wiechert and Noh, 2005). Although some pool sizes can be regarded as variables and estimated during flux evaluation, the results may be imprecise due to large confidence intervals (Young et al., 2008). (ii) Computationally, the extension from MFA to iMFA requires solving systems of ordinary differential equations (ODEs) during each step of a global optimization algorithm, which is a significantly expensive task (Wiechert and Noh, 2005; Young et al., 2008). Many efficient computational tools have been developed to improve the efficiency of simulation in iMFA (Nöh et al., 2006; Nöh and Wiechert, 2004). However, all these methods are based on numerical solvers for ODEs and thus the computational time and precision depends on the integration step size. Thus developing efficient computational tools for iMFA is of particular interest.

1.3. Methods: Mathematical modeling of ^{13}C MFA

1.3.1. Establishing metabolic networks from metabolic pathway databases

Several free online databases provide comprehensive information on metabolic pathways and genomic information for various organisms, such as Kyoto Encyclopedia of Genes and Genomes (KEGG) (Masoudi-Nejad et al., 2008) and

MetaCyc Encyclopedia of metabolic pathways (Zhang et al., 2005). National Center for Biotechnology Information (NCBI) Genome site and Uniprot (Apweiler et al., 2004) contain detailed information on protein sequences and gene annotations. We established real metabolic reaction networks based on the information from these databases.

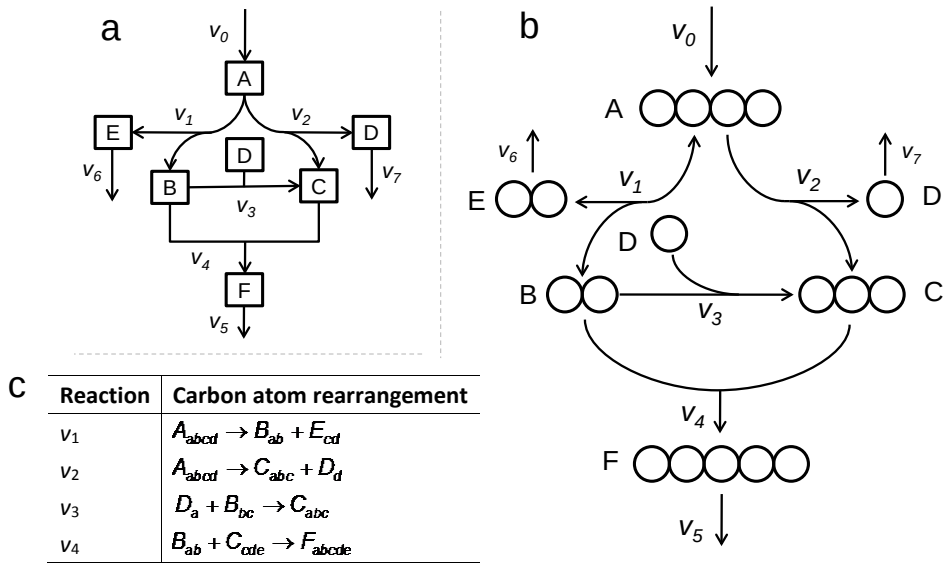
Moreover, since plant cells are eukaryotes, their chemical reactions are compartmentalized in different organelles such as cytosol, mitochondrion and chloroplast. Some information of the compartmentation is available from above online databases, or published genome scale models (e.g. Poolman et al., 2009). If these information are not available for certain organisms, we can iteratively revise the established metabolic network models during MFA until the best flux estimates are achieved.

1.3.2. Balancing intracellular metabolites

Here we use a simple metabolic network **(Fig. 1.2a)** to illustrate the protocol of MFA. In this network, metabolite A is the carbon source. Two cleavage reactions v_1 and v_2 convert A to B and E, A to C and D, respectively. Two condensation reaction convert B and D to C, B and C to D, respectively. Reaction v_0 is the influx of A and reaction v_5, v_6, v_7 are the effluxes of F, E and D. Here we use v_0 to v_7 to denote both reaction names and flux values. This network consists of 8 chemical or transport reactions. However, not all of the fluxes are independent variables. For example, the measurements of v_0 and v_1 automatically gives v_2 as $v_2 = v_0 – v_1$. In other words, this

network has a degree of freedom smaller than 8. The purpose of balancing all the metabolites is to determine the degree of freedom of the flux space and express all the "dependent" fluxes with respect to the independent fluxes.

The balancing of metabolites is based on the conservation of mass. One of the important assumption in isotope-assisted MFA is the achievement of metabolic steady state, during which the metabolites can neither be accumulated nor be depleted. Based on this assumption, we list the material balancing equations for all the metabolites in **Table 1.1.**



Reaction	Carbon atom rearrangement
v_1	$A_{abcd} \rightarrow B_{ab} + E_{cd}$
v_2	$A_{abcd} \rightarrow C_{abc} + D_{d}$
v_3	$D_{a} + B_{bc} \rightarrow C_{abc}$
v_4	$B_{ab} + C_{cde} \rightarrow F_{abcde}$

Figure 1.2. Simple metabolic network for illustration of MFA.
(a) Squares represent metabolites and arrows represent chemical reactions. Fluxes are denoted by v_0 to v_7. **(b)** Circles represent carbon atoms. Other notations are same as **(a)**. **(c)** Carbon atom arrangements of the chemical reaction.

Table 1.1. Materials balance equations for the metabolites in Fig 1.2.

Metabolites	Balancing equations
A	$v_0 - v_1 - v_2 = 0$
B	$v_1 - v_3 - v_4 = 0$
C	$v_2 + v_3 - v_4 = 0$
D	$v_2 - v_3 - v_7 = 0$
E	$v_1 - v_6 = 0$
F	$v_4 - v_5 = 0$

These linear equations can be expressed in a matrix form as,

$$\begin{bmatrix} 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \end{bmatrix} \bullet \begin{bmatrix} v_0 \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad [1.1]$$

in which the 6 by 8 matrix is the "stoichiometric matrix", represented by $\boldsymbol{S}$, and the 8 by 1 vector is the "flux vector", represented by $\boldsymbol{v}$. Therefore we have

$$\boldsymbol{S} \bullet \boldsymbol{v} = \boldsymbol{0} \qquad [1.2]$$

In the stoichiometric matrix, the entries in each row correspond to one metabolite, and entries in each column correspond to one flux variable. In this linear system, the number of independent flux variables, or free fluxes, is calculated by the formula,

$$Number\ of\ \ free\ \ fluxes = Number\ of\ \ fluxes - Rank\ of\ \ \boldsymbol{S} \qquad [1.3]$$

The selection of independent flux set is not unique. A classic method to determine a valid set of free fluxes and subsequently solve this linear system is to use

Gaussian Elimination, which converts ***S*** to a upper triangular matrix through linear transformations. Here Eq. [1.2] can be transformed into,

$$\begin{bmatrix} 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 3/2 & 0 & -1 & -1/2 \\ 0 & 0 & 0 & 0 & 0 & -1 & 2/3 & 1/3 \end{bmatrix} \bullet \begin{bmatrix} v_0 \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad [1.4]$$

Therefore, if the effluxes of D (v_7) and E (v_6) are measured, all the other fluxes can be calculate as,

$$\begin{aligned} v_0 &= \frac{4}{3}v_6 + \frac{2}{3}v_7 \\ v_1 &= v_6 \\ v_2 &= \frac{1}{3}v_6 + \frac{2}{3}v_7 \\ v_3 &= \frac{1}{3}v_6 - \frac{1}{3}v_7 \\ v_4 &= \frac{2}{3}v_6 + \frac{1}{3}v_7 \\ v_5 &= \frac{2}{3}v_6 + \frac{1}{3}v_7 \end{aligned} \quad [1.5]$$

In this example, the “dependent” fluxes are solved manually through Gaussian Elimination. For more complex and realistic metabolic networks, this process requires the application of computational tools such as MATLAB.

1.3.3. Balancing isotopomers of intracellular metabolites

In the metabolic network **Fig. 1.2b.**, assume that compounds A to F consist of certain numbers of carbon atoms as denoted in the figure. Carbon atom rearrangements in the chemical reactions are shown in **Fig. 1.2c**. The substrate is an isotopically labeled tracer of compound A (i.e. certain carbon atoms in A are ^{13}C isotopes while others are ^{12}C isotopes). As soon as the chemical reactions proceed, the carbon atoms of all the intermediate metabolites B to F can only have two isotopic states: labeled (^{13}C) or unlabeled (^{12}C). This yields a total of 2^n labeling patters for a metabolite that consists of n carbon atoms and these labeling patters are referred to as isotope isomers, or isotopomers. At isotopic steady state, the formation rate of an isotopomer is identical to its depletion rate. **Table 1.2** lists the balance equations of the total of eight isotopomers of metabolic C as an example.

Table 1.2. Materials balance equations of the isotopomers of metabolite C. "1" represents labeled carbon atoms and "0" represents unlabeled carbon atoms. Thus each combination of "0" and "1" denotes an isotopomer. Uppercase notations denote the isotopomer names and lowercase notations denote the fractions of the corresponding isotopomers (isotopomer abundances).

Isopomers of C	Balancing equations
C_{000}	$v_4 c_{000} = v_2(a_{0000} + a_{0001}) + v_3 b_{00} d_0$
C_{001}	$v_4 c_{001} = v_2(a_{0010} + a_{0011}) + v_3 b_{01} d_0$
C_{010}	$v_4 c_{010} = v_2(a_{0100} + a_{0101}) + v_3 b_{10} d_0$
C_{011}	$v_4 c_{011} = v_2(a_{0110} + a_{0111}) + v_3 b_{11} d_0$
C_{100}	$v_4 c_{100} = v_2(a_{1000} + a_{1001}) + v_3 b_{00} d_1$
C_{101}	$v_4 c_{101} = v_2(a_{1010} + a_{1011}) + v_3 b_{01} d_1$
C_{110}	$v_4 c_{110} = v_2(a_{1100} + a_{1101}) + v_3 b_{10} d_1$
C_{111}	$v_4 c_{111} = v_2(a_{1110} + a_{1111}) + v_3 b_{11} d_1$

1.3.4. Applying an optimization procedure to calculate fluxes

The isotopomer balance equations correlate isotopomer abundance, a measurable parameter, to intracellular fluxes that cannot be measured directly. Metabolic flux analysis relies on an optimization procedure to search for the best flux estimates **(Fig 1.3)**. Specifically, initial values are assigned to a set of valid free fluxes and all the dependent fluxes are calculated as discussed in **Sec. 1.3.2.** Isotopomer abundances are then solved iteratively from their balance equations (e.g. **Table 1.2**) with different initial guesses of the free fluxes. The best flux estimates are determined when the difference between the calculated isotopomers and the measured isotopomers are minimized.

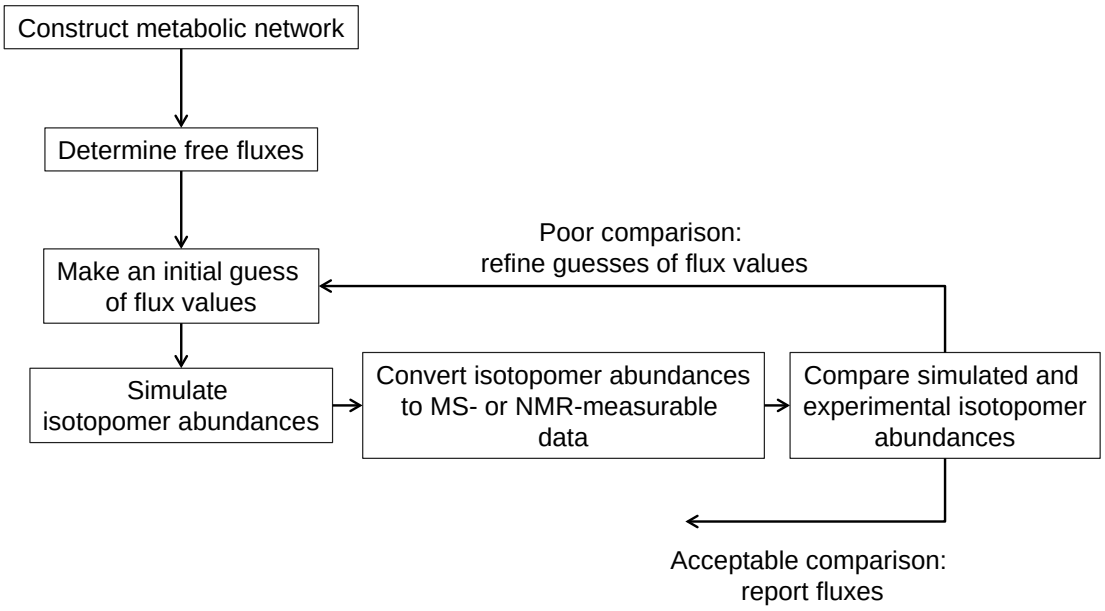


Figure 1.3. General flow chart of MFA protocol

That solving for fluxes in MFA is an optimization procedure is mainly due to two reasons. Firstly, in most cases not every single isotopomer abundance is

experimentally measurable, which makes it impossible to solve for fluxes directly. GC-MS and NMR are currently the most widely used technique to measure isotopic data, but neither of the methods is capable of measuring individual isotopomers directly. For example, GC-MS (the technique applied in this study) data illustrates the mass isotopomer distributions (MIDs) of a metabolite. It can determine how many carbon atoms in a metabolite are labeled, whereas it cannot determine which carbon atom is labeled. Therefore, the data it measures is usually a combination of several isotopomer abundances. **Table 1.3** displays the linear conversion from isotopomer abundances to MIDs for a metabolite containing three carbon atoms. C (M + *k*) (k=0, 1, 2, 3) represents the mass isotopomer in which *k* carbon atoms are labeled. Among them only c (M+0) and c (M+3) are identical to an individual isotopomer abundance. c (M+1) and c (M+2), however, are both equal to the sum of three individual isotopomer abundances.

Table 1.3. The conversion between isotopomer abundances and MIDs for metabolite C.
Uppercase notations denote the variables' names and lowercase notations denote the abundances of the corresponding variables.

MIDs of C	Balancing equations
C (M+0)	$c(M+0) = c_{000}$
C (M+1)	$c(M+1) = c_{001} + c_{010} + c_{100}$
C (M+2)	$c(M+2) = c_{011} + c_{101} + c_{110}$
C (M+3)	$c(M+3) = c_{111}$

Additionally, current GC-MS technique is able to detect a vast number of metabolites, including but not limited to intermediate central carbon metabolites, amino acids **(Sec. 3.3 and Sec. 4.3)** and fatty acids (Wu et al., 2005). The large MID

dataset usually makes flux variables overdetermined by isotopomer balance equations. The attempt of utilizing all available measurement data in order to reduce flux estimation error also leads to the optimization procedure.

Several softwares have been developed to perform MFA, such as 13CFlux2 (Weitzel et al., 2013, p. 2), OpenFlux (Quek et al., 2009), Metran (not available publically) and NMR2Flux (not available publically). In this study we used NMR2Flux, a home grown program developed by Sriram *et al.* (Sriram and Shanks, 2004). Briefly, this program accepts metabolic network stoichiometries and carbon atom rearrangements supplied by the user, employs cumulative isotopomer (cumomer, discussed in **Sec. 1.3.5**) balancing to simulate isotopomer abundances corresponding to any feasible set of fluxes in the network. Given a set of measured isotopomer abundances, the program uses a global optimization routine based on simulated annealing to iteratively evaluate flux values that best account for the measured data. It can also perform statistical analysis on estimated flux values via Monte-Carlo algorithm.

1.3.5. Linearization of isotopomer balance equations.

One major step in MFA is to solve the isotopomer balance equations. However, solving for isotopomer abundances directly is computationally difficult because: (1) some of the balance equations are non-linear, which makes most analytical solution methods (e.g. LU decomposition) not applicable; (2) all the balance equations need to be solved simultaneously. Linearization of the isotopomer

balance equations can be accomplished by introducing a new type of labeling variable, such as the “cumulative isotopomer” concept (Wiechert et al., 1999), the “elementary metabolite units” concept (M. R. Antoniewicz et al., 2007) the “bond isomer” concept (Sriram et al., 2004; van Winden et al., 2002) and the “fluxomer” concept (Srour et al., 2011).

We applied the “cumulative isotopomer” (cumomer) concept to linearize the isotopomer balance equations in this study. Unlike isotopomers that distinguish carbon atoms between “labeled” and “unlabeled”, cumomers distinguish them between “labeled” and “labeled or unlabeled”. After performing a simple linear transformation converting isotopomer abundances to cumomer abundances, this new method decomposes the non-linear isotopomer balance equations into a series of linear systems that can be solved in a cascaded manner. The total number of balance equations remain the same, whereas solving several linear systems requires significantly less computational power compared to solving a single, large non-linear system (Wiechert et al., 1999).

1.4. Outline of work performed in this study

MFA is a powerful technique to study metabolism. However, the requirement of intensive computational power by MFA, especially instationary MFA, makes it difficult to be applied to complex but realistic metabolic networks. In the first chapter of this work, we introduced two novel approaches to improve the computational efficiency of MFA. Firstly, we presented the concept of “originomer”, which uses the origins of atoms, instead of the isotopic nature of atoms, as labeling variables. A

major advantage of this novel concept is that it eliminates the influence of natural abundances of atoms on the labeling and that it can be applied to most ILE scenarios. Secondly, we applied an analytical algorithm that solves the systems of ODEs resulting from instationary “cumulative originomer” balances by using eigensystems of the coefficient matrices. Comparisons between ILE simulations employing the traditional cumomer concept and the novel originomer concept, as well as comparisons between traditional numerical solutions and the new analytical solution method for a realistic metabolic network show that both improvements significantly reduced the computational time.

In the second chapter, we aimed at illustrating and solving an intrinsic problem associated with MFA on cells or tissues in “fed-batch” culturing environment: the influence of initial present biomass on the accuracy of flux estimations. Through three parallel ^{13}C ILEs with different types of labeled glucose as the major carbon source, we demonstrated that the mixing behavior of this initial present biomass with newly synthesized biomass involves “dilution” and “reflux”, which can significantly affect isotopomer abundances. Then we proposed a full “reflux” metabolic network model that accounts for measured isotopomer data much more satisfactory than three other previously reported strategies. Estimated flux results from our methodology using the reflux model were similar to those from another methodology involving minimal reflux, which further validated the application of the “reflux” concept.

Finally, we reported a study that investigated the metabolic adaptation to carbon and nitrogen deficiencies by poplar cells and the mechanism to improve their NUE under low nitrogen supply. To accomplish this, we concurrently performed (i) steady-state ^{13}C metabolic flux analysis using multiple isotope labels and (ii) transcriptomic profiling using cDNA microarrays. ^{13}C flux analysis revealed the absolute flux through the oxidative pentose phosphate pathway (oxPPP) to be substantially (~3-fold) lower under the low-nitrogen conditions and the relative flux partitioning between the tricarboxylic acid cycle and anaplerotic pathways to vary considerably, Together, the flux and gene expression data suggested a plastidic localization of the oxPPP as well as transcriptional regulation of some major flux alterations including that in the oxPPP. The gene expression data also pointed to NUE-improving mechanisms such as redirection of additional carbon to aromatic metabolic pathways and recruitment of photosynthetic and light-harvesting proteins as nitrogen sinks in nitrogen-abundant conditions.

Chapter 2: Originomer: A Novel Concept Enabling Efficient Analytical Solutions For Instationary Isotope-Assisted Metabolic Flux Analysis

Xiaofeng Zhang[1,2], Ian C. MacIntire[1] and Ganesh Sriram[1,2]

GS conceived this work; GS, XZ and ICM designed it. XZ performed most of the computations; ICM performed initial versions of them. XZ wrote the manuscript and GS critically edited it.

[1]Department of Chemical and Biomolecular Engineering, University of Maryland

[2]Principal investigator and corresponding author

Abstract

Isotope-assisted metabolic flux analysis (MFA), with its ability to quantify carbon traffic through intracellular metabolic networks, is a powerful tool in metabolic engineering. However, implementing MFA in in organisms with complex biochemistries is time-consuming because of large computational costs. Moreover, instationary MFA (iMFA) exacerbates this situation because it necessitates the solution of ordinary differential equations instead of linear equations. To improve the computational efficiency of MFA and iMFA, we present two independent, novel approaches. The first approach is a concept called “originomer”, which uses the origins of atoms, instead of the isotopic nature of atoms, as labeling variables. Just as isotopomers are extended to cumomers or elementary metabolite units, it is possible to extend originomers to cumulative originomers and elementary originomer units (EOUs). Our second approach is an eigensystem method which facilitates analytical solutions of cumulative originomer or EOU balances in iMFA. Using this approach enabled us to represent labeling variables as sums of exponential terms. Simulation results for several example networks confirmed the improved efficiency resulting from the application of both approaches. While the analytical solution method does not save time for all metabolic networks, it can save time for many realistic iMFA scenarios without loss of accuracy. For example, a comparison between a cumulative originomer-based analytical solution method and the conventionally used cumomer-based numerical method for the tricarboxylic acid cycle showed that the approaches introduced in this work lead to a 23-fold reduction in computational time. We anticipate that in the future, these methods will facilitate the extension of iMFA to

more complex metabolic networks and multi-element labeling experiments (e.g. ^{13}C, ^{15}N).

2.1. Introduction

Metabolic flux analysis (MFA) is a vital tool in metabolic engineering due to its ability to quantify the relative rates of chemical reactions in an intracellular metabolic network (Ahn and Antoniewicz, 2011; Kruger and Ratcliffe, 2009; Leighty and Antoniewicz, 2012; Schwender, 2008). MFA is also capable of elucidating previously unknown pathways (Schwender et al., 2004; Sriram et al., 2007a). Besides, flux data can reflect the effect of genetic or environmental perturbations and thus identify metabolic engineering strategies (Sauer, 2006; Stephanopoulos, 1999; Stephanopoulos and Vallino, 1991). Flux values obtained from MFA, if combined with metabolite concentration data, constitute a minimal set of information needed to fully characterize metabolism (Stephanopoulos, 1999).

Two major approaches in MFA are flux balance analysis (FBA) and isotope-assisted flux analysis. FBA is relatively straightforward to implement because it requires minimal experimental measurements (Orth et al., 2010; Ramakrishna et al., 2000; Varma and Palsson, 1994). Isotope-assisted MFA, specifically ^{13}C MFA, is another approach that can provide more intracellular detail than FBA. In general, 13C MFA detects the labeling response of intermediates of a metabolic network to certain types of labeled substrate (Wiechert, 2001). This approach is becoming very powerful with the widespread application of gas chromatography-mass spectrometry (GC-MS) (Ahn and Antoniewicz, 2011; Leighty and Antoniewicz, 2012; Wittmann and Heinzle,

1999), the development of tandem MS (Choi and Antoniewicz, 2011) and nuclear magnetic resonance (NMR) techniques to measure isotopic labeling patterns (Goudar et al., 2010; Szyperski, 1995).

Interpreting GC-MS or NMR measurements to obtain fluxes requires the introduction of state variables connecting the flux values to the measurements. The most straightforward idea is to balance all atoms of the metabolites in the network (Zupke and Stephanopoulos, 1994). An extension of this idea that makes use of a larger experimental dataset is to balance all isotopomers of the metabolites (Schmidt et al., 1997). However, the nonlinear nature of the resultant algebraic systems requires iterative solutions and therefore makes these methods cumbersome. To surmount this problem, researchers have introduced concepts such as cumulative isotopomers (cumomers) (Wiechert et al., 1999), elementary metabolite units (EMUs) (M. R. Antoniewicz et al., 2007) and bondomers (Sriram and Shanks, 2004; van Winden et al., 2002). Wiechert et al. (1999) pioneered the concept of cumomer. A cumomer is a linear combination of certain isotopomers. This transformed variable enables the decoupling of the bilinear isotopomer balance equations to a cascade of linear equations that can be solved sequentially. However, this improvement does not reduce the number of state variables because the number of cumomers of a metabolite is identical to the number of its isotopomers. (A metabolite with n carbon atoms has 2^n isotopomers and 2^n cumomers.) Recently, a new concept called fluxomer decouples bilinear terms by using a single variable that consolidates both fluxes and isotopomers (Srour et al., 2011), resulting in balances that can be solved sequentially.

To reduce the number of labeling variables in certain types of isotope labeling experiments (ILEs), van Winden et al. (2002) introduced and Sriram and Shanks (2004) advanced another concept called bondomer, which keeps track of the integrities of bonds in the substrate molecule instead of the labeling states of single atoms. This concept factors out the natural abundance of unusual isotopes (e.g. ^{13}C and ^{15}N) and can thus lead to a ~three-fold reduction in the number of labeling variables in realistic metabolic networks (Sriram and Shanks, 2004). However, a shortcoming of the bondomer concept is that it can only be applied to ILEs in which only a single uniformly labeled (carbon) source and its naturally abundant counterpart are supplied. Subsequently, Antoniewicz et al. (2007) introduced the EMU concept. By grouping isotopomers according to the way they participate in metabolism, this concept eliminates the labeling information of non-measurable molecules that do not affect the labeling states of measurable metabolites. Thus, this approach drastically reduces the number of labeling variables. Although the EMU concept can be applied to any labeling experiment, a limitation of it is that unlike bondomers, it does not factor out natural abundance. Due to the complementary shortcomings of both these methods, it is highly desirable to extend the bondomer concept to other types of labeling experiments and to ultimately combine this "extended bondomer" concept with EMUs. Toward this goal, we propose to introduce the concept of "originomers".

Another trend in ^{13}C MFA besides the development of efficient network decomposition techniques is instationary MFA (iMFA). As discussed in **Sec. 1.2.2**,

iMFA saves experimental time, but is computationally more challenging due to the requirement of solving ODEs. In this work, we introduce two novel approaches to improve the computational efficiency of MFA. Firstly, we present the concept of originomers, which uses the origins of atoms, instead of the isotopic nature of atoms, as labeling variables. A major advantage of this novel concept is that it eliminates the influence of natural abundances of atoms on the labeling, just as the bondomer concept does. However, the originomer concept is applicable to all labeling experiment scenarios and reduces the number of labeling variables or the number of significant labeling variables in many commonly used labeling experiments. Secondly, we introduce an analytical algorithm that solves the systems of ODEs resulting from instationary cumulative originomer balances by using eigensystems of the coefficient matrices. Comparisons between ILE simulations employing the traditional cumomer concept and the novel originomer concept, as well as comparisons between traditional numerical solutions and the new analytical solution method for an example network comprising the tricarboxylic acid (TCA) cycle and glyoxylate shunt show that both improvements significantly reduce the computational time.

2.2. Materials and Methods

2.2.1. Originomers track the origins of atoms in metabolites

The originomer concept enables the extension of a bondomer-like idea to several commonly employed labeling experiments. In contrast to isotopomers, which are isomers of a metabolite that differ in the isotopic nature of their individual atoms, originomers are isomers of a metabolite that differ in the origins of their individual

atoms (**Fig. 2.1A**). To analyze an ILE by using the originomer concept, one atom of a substrate is designated "special". Atoms of intracellular metabolites are then classified according to whether they originated from this special atom or did not originate from it. These two "origin states" apply to every atom in the metabolic network. Thus, a metabolite with *n* atoms of a certain element can have 2^n isomers that differ in the origin states of their atoms. We define these new types of isomers as "originomers". For example, in an ILE in which 1-^{13}C glucose is the sole carbon source, it is convenient to designate the labeled (C-1) atom of glucose as "special". All the carbon atoms of an intracellular metabolite, e.g. triose phosphate (T3P) in glycolysis, can then be in one of two states: (i) they originated from C-1 of glucose and are thus special (S), or they did not originate from C-1 of glucose and are thus non-special (N). In this manner, T3P can have eight originomers, just as it can have eight isotopomers (**Fig. 2.1B, Fig. 2.1C**). To distinguish originomer notation from standard isotopomer notation $T3P_{101}$ (1: labeled; 0: unlabeled) (Wiechert et al., 1999), we denote originomers of T3P as $T3P_{SNS}$ (S: special; N: non-special).

The originomer concept reduces the number of labeling variables in an ILE because although a given metabolite can have equal numbers of originomers and isotopomers, not all originomers of all metabolites may be feasible in a real metabolic network. This is because the originomer concept factors out the natural abundances of unusual isotopes and thus, "non-special" is absolute. Therefore, in an ILE in which 1-^{13}C glucose is the sole carbon source, exactly one originomer of glucose (Glc_{SNNNNN}) is supplied. Conversely, "unlabeled" is not absolute because of the natural abundance

of isotopes such as ^{13}C. Thus, feeding 1-^{13}C glucose actually means feeding 63 isotopomers of glucose (all except Glc_{000000}), many of which may be present in measurable quantities.

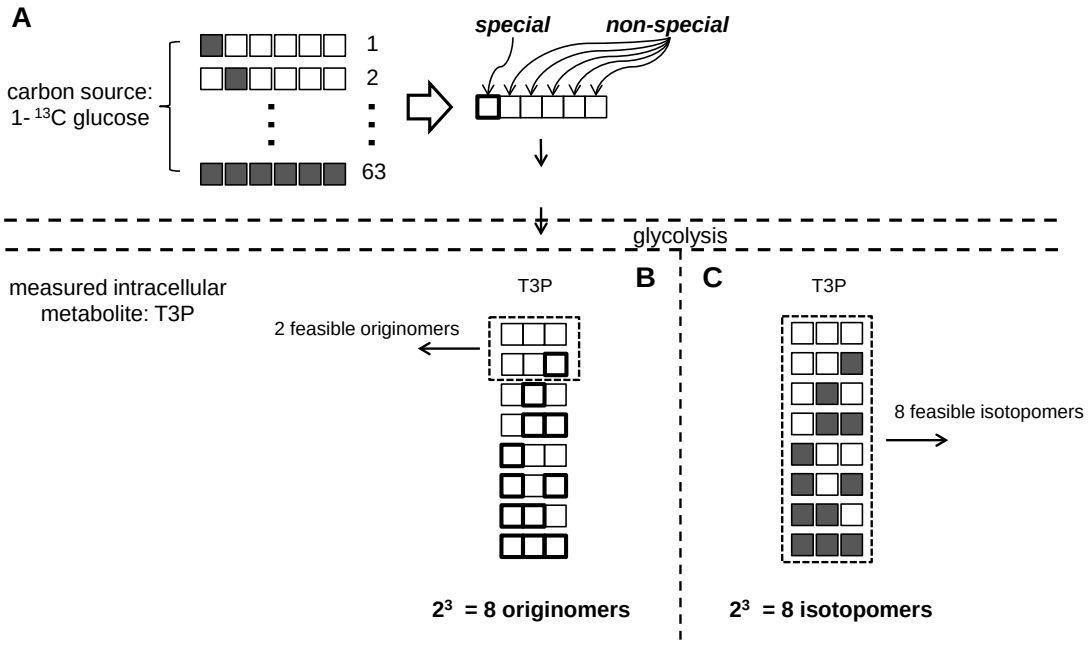


Figure 2.1. The originomer concept.
Originomers are isomers of a metabolite that differ in the origins of their individual atoms. **(A)** To analyze an ILE by using the originomer concept, one or more atoms of (a) substrate(s) (C-1 of glucose in this example) is (are) designated "special". Atoms of intracellular metabolites (e.g. T3P in glycolysis) are then classified according to whether they originated from this special atom or did not originate from this atom (non-special). These two "origin states" apply to every atom in the metabolic network. **(B)** Thus, a metabolite with *n* atoms of an element can thus have 2^n isomers that differ in the origin states of their atoms. We define these isomers as "originomers". Squares with bold borders denote special atoms; squares with light borders denote non-special atoms. **(C)** Originomers contrast with isotopomers, which are isomers of a metabolite that differ in the isotopic nature of their individual atoms. Filled squares denote ^{13}C atoms; empty squares denote ^{12}C atoms. Abbreviation: T3P, triose phosphate.

Furthermore, analogous to the extension from isotopomer to cumomer (Wiechert et al., 1999), we can extend the originomer concept to cumulative originomer. Using the symbol "?" to represent an atom whose origin state (S or N) is

undetermined or irrelevant, the notation $T3P_{?S?}$ represents a cumulative originomer of T3P in which the second carbon is special, whereas the first and third atoms may be special or non-special. Also analogous to cumomers, we define the order of cumulative originomers as the number of atoms that are known to be special. To further reduce computational time, we could extend the originomer concept to elementary originomer units (EOUs) similar to the extension of isotopomer to EMU (M. R. Antoniewicz et al., 2007). However, this extension is not discussed in this work.

2.2.2. Originomer enumeration

To harness the potential of the originomer concept in reducing labeling variables, it is necessary to first enumerate all the originomers in a metabolic network and then write balance equations by disregarding the infeasible originomers. This enumeration-elimination procedure is similar to that used in bondomer balancing (Sriram and Shanks, 2004). We demonstrate originomer enumeration through an illustrative metabolic network (**Fig. 2.2**). Here, 1-^{13}C-labeled M is the carbon source; A, B, C are intermediates; D, E, F are products of the network. Only the reaction v_1 (A → B) is assumed reversible, whereas all other reactions are assumed irreversible. **Table 2.1** lists the carbon atom rearrangements of this network. Toward enumerating originomers, we designate carbon atom C-1 of the substrate M (the only atom that is labeled) as the special atom. Then, on working through the network by starting from the carbon source originomer M_{SNNN} and using the carbon atom rearrangements, we realize that only 13 originomers (A_{SNNN}, A_{NNNN}, B_{SN}, B_{NN}, C_{SNN}, C_{NNN}, D_S, D_N, E_{NN},

F_{SNSNN}, F_{SNNNN}, F_{NNSNN}, F_{NNNNN}) are feasible in this network, out of the entire set of 66 originomers of all the metabolites (A, B, C, D, E, F). For a complex and more realistic network, such an enumeration step may need to be performed *in silico* instead of manually.

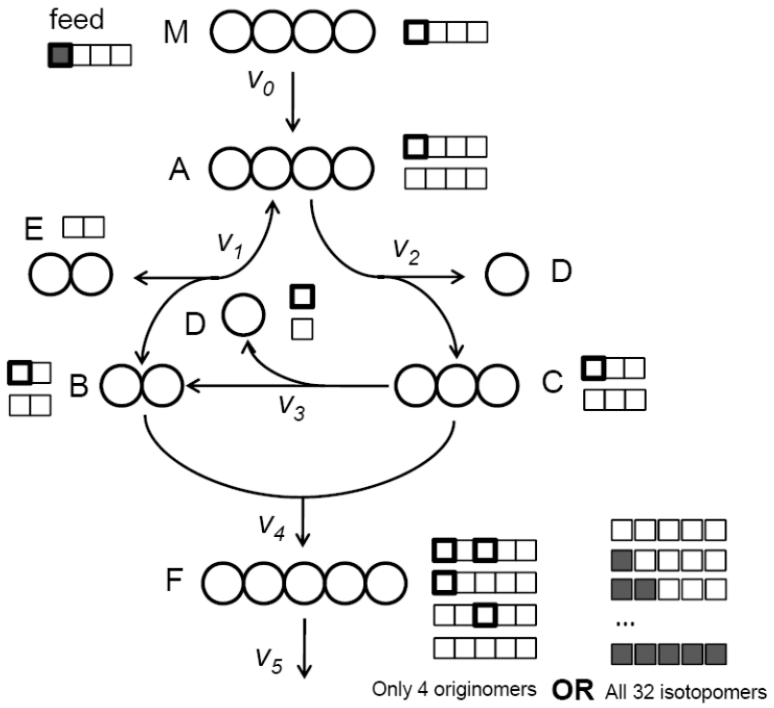


Figure 2.2. Metabolic network for illustration of originomers and isotopomers. This simple, hypothetical network contains seven intracellular metabolites (A to F), a single carbon source (M) and a single metabolite (F) whose labeling state is measured. Circles or squares represent carbon atoms and arrows represent reactions, whose fluxes are v_0 to v_5. Originomers and isotopomers are depicted beside the corresponding metabolites. Squares with bold borders represent "special" atoms (see text for definition), whereas filled squares represent labeled (^{13}C) atoms. Due to the natural abundance of ^{13}C in the "unlabeled" carbon atoms of the feed, all the possible (32) isotopomers of F are feasible in this network. However, only four originomers of F are feasible, and contain the same labeling information as the isotopomers.

Table 2.1. Stoichiometry and carbon atom rearrangements of the example network

Reaction	Carbon atom rearrangement
v_0	$M_{abcd} \rightarrow A_{abcd}$
v_1	$A_{abcd} \rightarrow B_{ab} + E_{cd}$
v_2	$A_{abcd} \rightarrow C_{abc} + D_d$
v_3	$C_{abc} \rightarrow D_a + B_{bc}$
v_4	$B_{ab} + C_{cde} \rightarrow F_{abcde}$

After determining a set of free fluxes and using a stoichiometric matrix to express all other fluxes in terms of the free fluxes, we can now write originomer balances for the network and decompose these balances into a series of cascaded cumulative originomer networks by following procedures analogous to cumomer decomposition (Wiechert et al., 1999). These cumulative originomer networks have a cascaded structure; therefore they should be solved from order 1 upwards to the highest order in the network. For example, the 1st-order (feasible) cumulative originomers $A_{S???}$, $B_{S?}$, $C_{S??}$, $F_{S????}$ and $F_{??S??}$ (abundances indicated below by lowercase letters) can be balanced as:

$$P_a \frac{da_{S???}}{dt} = -\left(v_1^+ + v_2\right) a_{S???} + v_1^- b_{S?} + v_0 m_{S???} \quad [2.1]$$

$$P_b \frac{db_{S?}}{dt} = -\left(v_1^- + v_4\right) b_{S?} + v_1^+ a_{S???} \quad [2.2]$$

$$P_c \frac{dc_{S??}}{dt} = -\left(v_3 + v_4\right) c_{S??} + v_2 a_{S???} \quad [2.3]$$

$$P_f \frac{df_{S???}}{dt} = -v_5 f_{S????} + v_4 b_{S?} \quad [2.4]$$

$$P_f \frac{df_{??S??}}{dt} = -v_5 f_{??S??} + v_4 c_{S??} \quad [2.5]$$

These are identical to the balances on the corresponding 1st-order cumomers (A_{1XXX}, B_{1X}, C_{1XX}, D_1, F_{1XXXX}, F_{XX1XX}), with "1" in the cumomer balances replaced by

"S" in the cumulative originomer balances. As with cumomer balances, the 1st-order cumulative originomer balances are "self-contained" and linear in the unknown variables, so that they can be solved by matrix inversion. Although the 2nd-order cumulative originomer balances contain bilinear terms, these terms consist only of 1st-order cumulative originomer abundances:

$$P_f \frac{df_{S?S??}}{dt} = -v_5 f_{S?S??} + v_4 b_{S?} c_{S??} \qquad [2.6]$$

Therefore, once the 1st-order cumulative originomer balances have been solved, the set of 2nd-order cumulative originomer balances becomes linear in the unknown variables and can be solved analytically in a straightforward manner.

A compact way to represent these balance equations is the following matrix representation:

$$\mathbf{P}\frac{d\,{}^{r}\mathbf{x}}{dt} = {}^{r}\mathbf{A}\,{}^{r}\mathbf{x} - {}^{r}\mathbf{b} \qquad [2.7]$$

where $\mathbf{P}$ is a diagonal matrix containing pool sizes of the metabolites (irrelevant in steady-state MFA but indispensable in iMFA), ${}^{r}\mathbf{A}$ is a coefficient matrix for the current order, entries of which are metabolic fluxes, vector ${}^{r}\mathbf{b}$ collects all the terms obtained from entry fluxes or bilinear terms involving lower-order cumulative originomers.

After using Eq. [2.7] to solve for the originomer abundances, we can convert them to isotopomer abundances measurable by mass spectrometry or nuclear magnetic resonance. For this, we multiply the originomer abundances by a mapping matrix whose elements, composed of natural abundances of carbon atoms, indicates

the labeling probabilities of the carbon atoms in each originomer. Therefore, each row of this matrix represents the contribution of different originomer abundances to a given isotopomer abundance.

In the above example, we designated the ^{13}C-labeled atom in the substrate as the special atom. However, the designation of special atoms is flexible to some extent. Importantly, multiple atoms from the same substrate can be simultaneously designated special. For ILEs involving a single carbon source it is usually convenient to designate the labeled atoms of this carbon source as special (**Fig. 2.3, case I**). However, for ILEs involving multiple carbon sources, it may be more advantageous to designate all the atoms of the smaller, unlabeled substrate molecule as special (**Fig. 2.3, cases II and III**). Furthermore, in combined carbon-nitrogen MFA in which a ^{13}C,^{15}N-labeled amino acid is fed as a carbon and a nitrogen source, it is beneficial to regard the nitrogen atom as special (**Fig. 2.3, case IV**). However, it is meaningless to: (i) simultaneously designate a labeled and an unlabeled atom from the same substrate together as special; and (ii) simultaneously designate atoms from multiple substrates as special. These designations will result in valid originomer networks; however, the originomers may not meaningfully correlate with isotopomers. Ultimately, the goal of designating special atoms is to reduce the number of feasible labeling variables.

2.2.3. Solution of instationary originomer balances

Our second approach is to improve the computational efficiency of iMFA by determining analytical solutions for the instationary cumulative originomer balances in Eq. [2.7], given that these equations are 1st-order, linear, inhomogeneous ODEs with constant coefficients. The general idea of this method is to diagonalize the coefficient

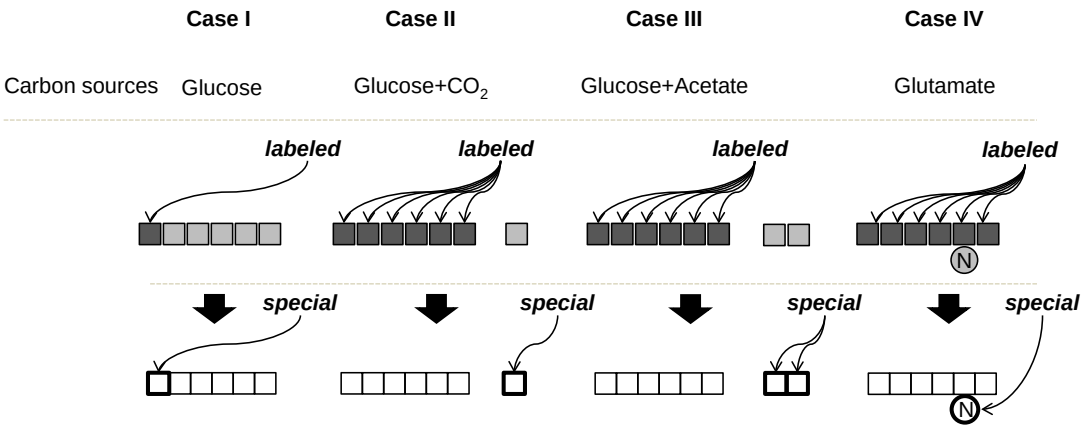


Figure 2.3. Several ways of designating special atoms.
Case I. In a labeling experiment in which 1-^{13}C glucose is the sole carbon source, it is convenient to designate the C-1 of glucose as special. Cases II and III. Labeling experiments to investigate co-metabolism of two carbon sources may use U-^{13}C versions of one of them and naturally abundance versions of others, e.g. U-^{13}C glucose + naturally abundant CO_2 or U-^{13}C glucose + naturally abundant acetate. In such cases, it is convenient to designate all the carbon atoms of the substrate with fewer carbon atoms as special. Case IV. Experiments to investigate carbon-nitrogen interactions may use labeled carbon and nitrogen source, e.g. U-^{13}C glucose and $^{15}NO_3^-$. Even here, the substrate with fewer labeled atoms ($^{15}NO_3^-$) can be designated special.

matrix **A** by using eigenvalues and eigenvectors (or eigensystems). Then, we can decouple the ODE system to a set of ODEs that can be solved separately and analytically.

The eigenvalues λ and eigenvectors $\mathbf{x}$ of a matrix $\mathbf{A}$ are defined so that:

$$\mathbf{Ax} = \lambda\mathbf{x} \qquad [2.8]$$

After left multiplying the inverse of the pool size matrix $\mathbf{P^{-1}}$ to both sides of Eq. [2.7], we obtain (the order number *r* is deleted in Eq. [2.9] for simplicity):

$$\frac{d\mathbf{x}}{dt} = \mathbf{A_1 x} + \mathbf{b_1} \qquad [2.9]$$

Using the eigenvalues and eigenvectors of $\mathbf{A_1}$ enables a transformation of variables:

$$\frac{d\mathbf{y}}{dt} = \mathbf{Dy} + \boldsymbol{\beta} \qquad [2.10]$$

where $\mathbf{x} = \mathbf{Qy}$, in which $\mathbf{Q}$ is the eigenmatrix of $\mathbf{A_1}$ and $\mathbf{D}$ is a diagonal matrix whose entries are eigenvalues of $\mathbf{A_1}$. By expanding Eq. [2.10], we decouple the ODE system to a list of ODEs:

$$\begin{aligned} \frac{dy_1}{dt} &= \lambda_1 y_1 + \beta_1 \\ \frac{dy_2}{dt} &= \lambda_2 y_2 + \beta_2 \\ &\cdots \\ \frac{dy_n}{dt} &= \lambda_n y_n + \beta_n \end{aligned} \qquad [2.11]$$

The solutions of these equations are:

$$y_j = e^{\lambda_j t}\left(c_j + \int e^{-\lambda_j t}\beta_j dt\right) \qquad [2.12]$$

After substituting the values of transformed variables y_j back to $\mathbf{x} = \mathbf{Qy}$, we obtain the cumulative originomer abundances $\mathbf{x}$.

The constants c_j are determined from expressions of the transformed variables y_j (Eq. [2.12]). ODEs with respect to y are decoupled and thus, the integration

constants can be determined directly from Eq. [2.12] given the initial conditions of y_j: $y_j\big|_{t=0} = 0$. This condition arises from the fact that before the instationary experiment, no special atom should have been present in the network. For example, for 1st-order cumulative originomer abundances, the constants c_j are:

$$c_{j,\,order=1} = \frac{\beta_{j,\,order=1}}{\lambda_{j,\,order=1}} \qquad [2.13]$$

Similarly, for higher order cumulative originomer abundances, constants c_j are functions of eigenvectors of the coefficient matrix $\mathbf{A_1}$ of the lower order and eigenvalues of matrix $\mathbf{A_1}$ of the current order.

$$c_{j,\,order=2} = \frac{f\left(\beta_{j,\,order=1}\right)}{\lambda_{j,\,order=2}} \qquad [2.14]$$

This analytical solution method for instationary balances is equally applicable to cumomers, EMUs, bondomers and originomers. However, due to natural abundance, using it with cumomers and possibly EMUs may result in a very large number of terms in Eq. [2.12], making the method impractical. Below, we show that the reduction in labeling variables afforded by the originomer concept renders the analytical solution method practical for various realistic iMFA scenarios.

In this work, we used MATLAB 7.0 and a computer with an i5 760 CPU to solve algebraic and ODE systems of originomer or cumomer balances. To measure computation times for various solution methodologies, we solved the relevant equations or executed the relevant process 1000 times (so as to obtain robust estimates of computation time) and employed the tic/toc command in MATLAB

2.3. Results

We applied the originomer concept and the analytical solution method to three typical metabolic networks encompassing central carbon and primary nitrogen metabolism. Network 1 comprises the TCA cycle and glyoxylate shunt with 1-^{13}C acetate as substrate (**Fig. 2.4**), network 2 features primary nitrogen metabolism with $^{15}NH_4^+$ and naturally abundant glutamine as substrates (**Fig. 2.5**); and network 3 exemplifies mixotrophic algal metabolism with U-^{13}C glucose and naturally abundant CO_2 as substrates, featuring glycolysis, pentose phosphate pathway (PPP) and photosynthetic carbon assimilation (**Fig. 2.6**). For all these networks, we first demonstrated the computational efficiency of the originomer concept by comparing the total number of variables that needed to be balanced using the cumomer and originomer concepts. Subsequently, we demonstrate our analytical solution method for network 1 and compare computation times for solving balance equations using the standard numerical method and our analytical method.

2.3.1. The originomer concept reduces the number of labeling variables

Network 1: TCA cycle and glyoxylate shunt.

In the network shown in **Fig. 2.4**, 1-^{13}C acetate is the carbon source (this enters the network as 1-^{13}C acetyl-CoA) and oxaloacetate is the product. We designated the C-1 atom of acetate (acetyl-CoA) as special. **Table 2.2** shows carbon atom rearrangements for the reactions in this network. Because succinate is a symmetric molecule, its carbon atoms C-1 and C-4 are chemically identical, as are C-2 and C-3. Therefore, we consolidated the two cumulative originomers $SUC_{S???}$ and

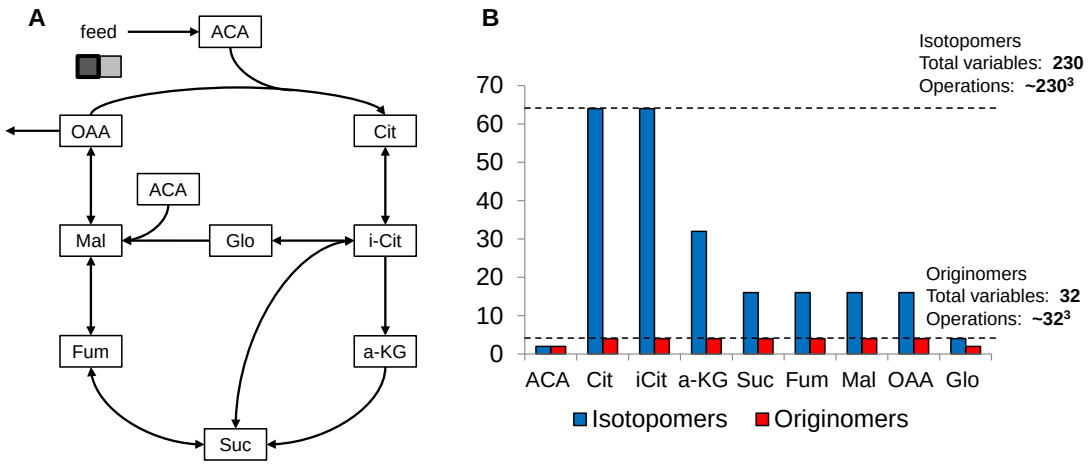


Figure 2.4. The originomer concept factors out natural abundances to reduce the total number of variables.
(A) Metabolic network comprising TCA cycle and glyoxylate shunt, with 9 metabolites and 8 fluxes. 1-^{13}C acetyl-CoA is the substrate and oxaloacetate is the product. **(B)** Designation of C-1 of acetyl-CoA as special results in much fewer feasible originomers than feasible isotopomers. The total number of originomers (32) is 7-fold lower than the total number of isotopomers (230). This translates to a significantly greater reduction in the number of operations as the number of operations in matrix inversion scales with the cube of the number of rows or columns. The lower dashed line indicates the highest number of originomers of any metabolite (8 for Cit, iCit, α-KG, Suc, Fum, Mal and OAA); the higher dashed line indicates the highest number of isotopomers for any metabolite (64 for Cit and iCit). Abbreviations: ACA, acetyl-CoA; Cit, citrate; Fum, fumarate; Glo, glyoxylate; iCit, isocitrate; Mal, malate; OAA, oxaloacetate; Suc, succinate; α-KG, α-ketoglutarate.

$SUC_{???S}$ into a single entity, representing it as $SUC_{S???}$. Originomer enumeration for this network indicates that only 32 originomers and 12 cumulative originomers are feasible (listed in **Table 2.3**, summarized in **Table 2.4**), with eight 1st-order cumulative originomers and four 2nd-order cumulative originomers (**Table 2.3**), out of the complete set of 230 possible cumulative originomers of all metabolites. Due to natural abundance of ^{13}C, this network has 230 non-zero isotopomers and 230 non-zero cumomers. Thus, employing the originomer concept can save substantial computation time for this network. This especially happens because matrix inversion

is the most computationally expensive process in the solution of compartmented systems such as cumomer (and analogously, cumulative originomer) balances. Since the inversion of an $n \times n$ matrix requires ~$O(n^3)$ multiplications (Press et al., 2007), reducing the number of labeling variables, as in this example, can substantially reduce computation time.

Table 2.2. **Carbon atom rearrangements for TCA cycle with glyoxylate shunt.**

Reaction	Carbon atom rearrangement
v_1	$ACA_{ab} + OAA_{cdef} \rightarrow AKG_{fedba} + CO_{2c}$
v_2	$AKG_{abcde} \rightarrow \frac{1}{2} SUC_{bcde} + \frac{1}{2} SUC_{edcb} + CO_{2a}$
v_3	$\frac{1}{2} SUC_{abcd} + \frac{1}{2} SUC_{dcba} \rightarrow MAL_{abcd}$
v_4	$MAL_{abcd} \rightarrow OAA_{abcd}$
v_5	$ACA_{ab} + OAA_{cdef} \rightarrow \frac{1}{2} SUC_{abdc} + \frac{1}{2} SUC_{cdba} + GLO_{ef}$
v_6	$ACA_{ab} + GLO_{cd} \rightarrow MAL_{dcba}$

Table 2.3. Cumulative originomer balances for TCA cycle with glyoxylate shunt.

Order	Cumulative originomer	Balance equation
1	$AKG_{S???}$	$P_{akg} \frac{dAKG_{S???}}{dt} = v_1 OAA_{S???} - v_2 AKG_{S???}$
	$AKG_{???S}$	$P_{akg} \frac{dAKG_{???S}}{dt} = v_1 ACA_{S?} - v_2 AKG_{???S}$
	$SUC_{S???}$	$P_{suc} \frac{dSUC_{S???}}{dt} = v_2 AKG_{???S} + v_3^- \left(MAL_{S???} + MAL_{???S}\right) + v_5^+ \left(ACA_{S?} + OAA_{S???}\right) - \left(v_3^+ + v_5^-\right) SUC_{S???}$
	$MAL_{S???}$	$P_{mal} \frac{dMAL_{S???}}{dt} = \frac{1}{2} v_3^+ SUC_{S???} + v_4^- OAA_{S???} + v_6 GLO_{?S} - \left(v_3^- + v_4^+\right) MAL_{S???}$
	$MAL_{???S}$	$P_{mal} \frac{dMAL_{???S}}{dt} = \frac{1}{2} v_3^+ SUC_{S???} + v_4^- OAA_{???S} + v_6 ACA_{S?} - \left(v_3^- + v_4^+\right) MAL_{???S}$

	$OAA_{S???}$	$P_{oaa}\frac{dOAA_{S???}}{dt} = v_4^+ MAL_{S???} + v_5^- SUC_{S???} - \left(v_4^- + v_5^+ + v_1\right) OAA_{S???}$
	$OAA_{???S}$	$P_{oaa}\frac{dOAA_{???S}}{dt} = v_4^+ MAL_{???S} + v_5^- GLO_{?S} - \left(v_4^- + v_5^+ + v_1\right) OAA_{???S}$
	$GLO_{?S}$	$P_{gly}\frac{dGLY_{?S}}{dt} = v_5^+ OAA_{???S} - \left(v_5^- + v_6\right) GLO_{?S}$
2	$AKG_{S???S}$	$P_{akg}\frac{dAKG_{S??S}}{dt} = v_1 ACA_{S?} OAA_{S???} - v_2 AKG_{S??S}$
	$SUC_{S??S}$	$P_{suc}\frac{dSUC_{S??S}}{dt} = v_3^- MAL_{S??S} + v_5^+ ACA_{S?} OAA_{S???} - \left(v_3^+ + v_5^-\right) SUC_{S??S}$
	$MAL_{S??S}$	$P_{mal}\frac{dMAL_{S??S}}{dt} = v_3^+ SUC_{S??S} + v_4^- OAA_{S??S} + v_6 ACA_{S?} GLO_{?S} - \left(v_3^- + v_4^+\right) MAL_{S???}$
	$OAA_{S??S}$	$P_{oaa}\frac{dOAA_{S??S}}{dt} = v_4^+ MAL_{S??S} + v_5^- GLO_{?S} SUC_{S???} - \left(v_4^- + v_5^+ + v_1\right) OAA_{S??S}$

Table 2.4. Number of labeling variables that need to be balanced if the concepts of isotopomer, cumomer, EMU and originomer, cumulative originomer and EOU are invoked.

When using the originomer concept, we considered only metabolic feasible originomers, cumulative originomers and EOUs. We performed a comparison between EMU and EOU only for network 1, where we assumed that the mass isotopomer distributions of OAA and αKG are measureable.

Network → **Number of variables ↓**	**Network 1: TCA cycle (Fig. 4)**	**Network 2: Primary nitrogen metabolism (Fig. 5)**	**Network 3: Glycolysis, PPP, photosynthesis (Fig. 6)**
Isotopomers	230	369	314
Originomers	32	24	71
Cumomers	230	369	314

Cumulative originomers	12	15	71
EMUs	124	–	–
EOUs	56	–	–
Comment			Originomer reduces higher order variables

Network 2: Primary nitrogen metabolism. The originomer concept is particularly powerful in analyzing ^{15}N labeling experiments in which the natural abundances of both ^{13}C and ^{15}N have to be considered. Nitrogen usually enters metabolism in the form of one atom (NO_3^-, NH_4^+ or rarely, glutamate) or two atoms (glutamine or rarely, urea). Since the incoming nitrogen atoms are not chemically bonded to each other in any of these substrates, it is impossible to obtain detailed flux information from steady-state MFA. Because the originomer concept factors out the influence of natural abundances of atoms, we can neglect the naturally abundant carbon atoms and design a decomposition procedure only for nitrogen atoms. This becomes a crucial advantage in instationary ^{15}N labeling experiments, where using the isotopomer or cumomer concepts may lead to a very large number of labeling variables.

To illustrate this, we analyzed a simple, but relevant metabolic network of nitrogen metabolism in plant cells (**Fig. 2.5**). This model features two compartments: cytosol and plastid, and accounts for pools of five amino acids, γ-aminobutyric acid (GABA) and ammonium. The substrates of this network are $^{15}NH_4^+$ and naturally

abundant glutamine. The steady-state labeling patterns in this network enable identification of the relative contributions of glutamine and ammonium fluxes to nitrogen-containing metabolites, whereas the instationary labeling patterns enable evaluation of nitrogen fluxes. Originomer enumeration with designation of the two nitrogen atoms in glutamine as special led to only 15 feasible originomers, which is a 96% reduction over the 369 isotopomers or cumomers feasible for this network (**Table 2.4**).

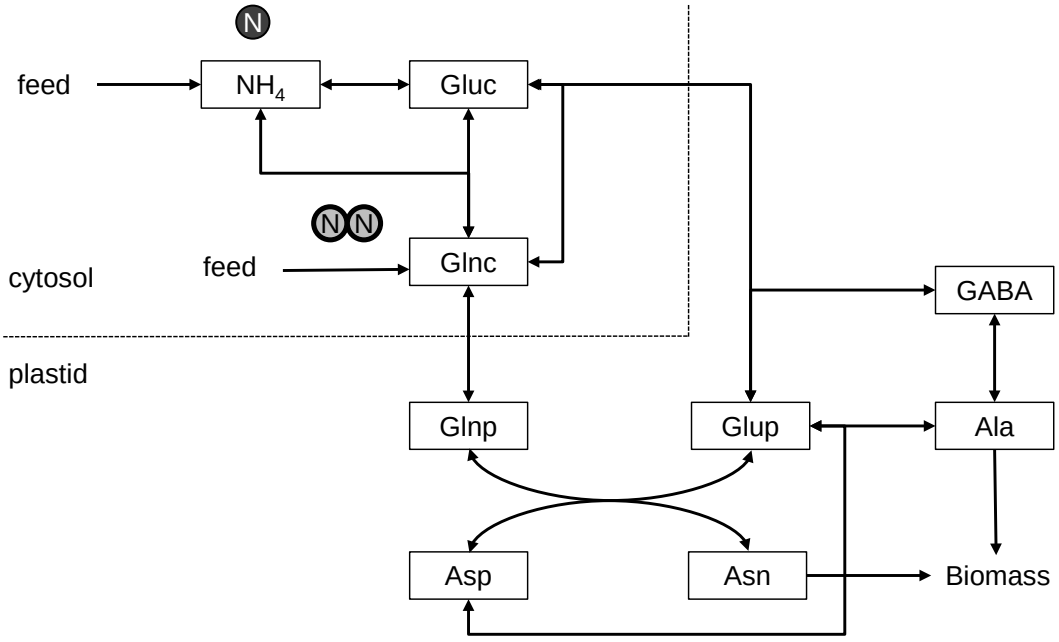


Figure 2.5. Plant nitrogen metabolism (simplified model) with two compartments: the cytosol and the plastid.
$^{15}NH_4^+$ and naturally abundant glutamine are the substrates. The atoms of the substrate glutamine are designated special. Abbreviations: Ala, alanine; Asn, asparagine; Asp, aspartate; GABA, γ-aminobutyric acid; Glnc, cytosolic glutamine; Glnp, plastidic glutamine; Gluc, cytosolic glutamate; Glup, plastidic glutamate.

Network 3. Mixotrophic algal metabolism with glycolysis, PPP and photosynthesis. Our third illustrative network features glycolysis, PPP and photosynthetic carbon assimilation in a typical model of algal mixotrophic

metabolism, and is supplied with U-^{13}C glucose and atmospheric CO_2 (**Fig. 2.6**). This network is particularly complex due to significant carbon atom rearrangement and reversible reactions. Due to this complexity, we find that on designating CO_2 carbon as the special atom, the number of cumulative originomers with non-zero abundance is identical to the number of cumomers. Despite this, employing cumulative originomers is still advantageous because the aforementioned choice of special atom results in a cumulative originomer distribution almost completely concentrated in orders 1 and 2. Thus hardly any information will lost if cumulative originomer balances beyond order 2 are not solved. Conversely, using cumomer representation results in measurable abundances for cumomers of all orders, making it necessary to balance all cumomers.

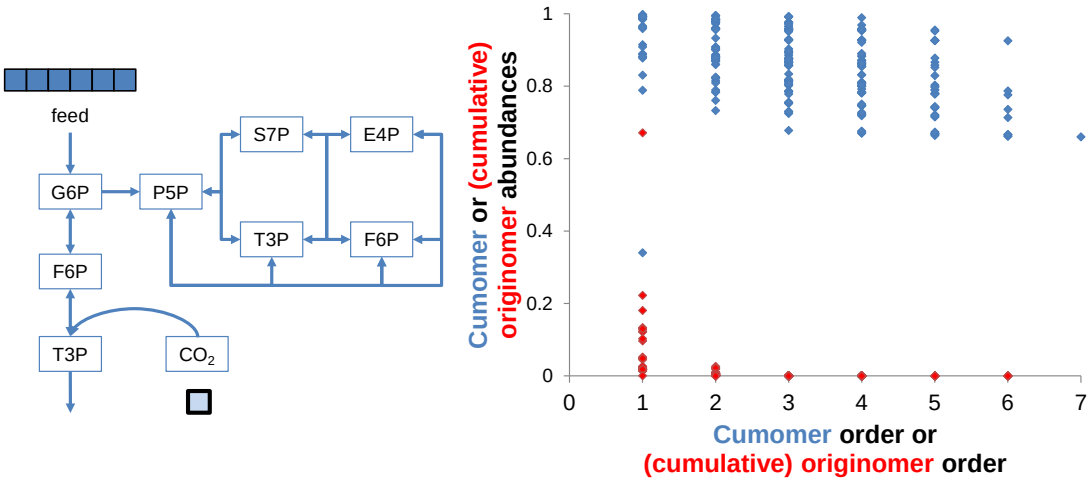


Figure 2.6. The originomer concept can reduce higher-order variables.
(A) Metabolic network comprising glycolysis and PPP in a photosynthetic organism, with 7 metabolites and 8 fluxes. U-^{13}C glucose and naturally abundant CO_2 are the substrates and T3P is the product. (B) Comparison of distributions of labeling variables among different orders by using isotopomer and originomer concepts. The total number of labeling variables is identical for isotopomers and originomers. However, using originomer representation (with CO_2 carbon as the special atom)

generates higher-order cumulative originomers at such low abundances that can be neglected.

An explanation for this observation is that a mixotrophic ILE feeding **(a)** U-^{13}C glucose and $^{12}CO_2$ (not naturally abundant) is a complement or "mirror image" of an ILE feeding **(b)** U-^{12}C glucose (not naturally abundant) and $^{13}CO_2$. Thus, analyzing ILE **(a)**, which features hardly any cumomers above order 2, is equivalent to analyzing ILE **(b)**. However, completely ^{12}C-labeled substrates are quite expensive, often more so than ^{13}C-labeled ones. For example, U-^{12}C glucose is significantly more expensive than U-^{13}C glucose (data from catalog of Cambridge Isotope Laboratories, Andover, MA, retrieved in April 2013). Furthermore, substrates in which some atoms are ^{13}C and others are purely ^{12}C are not available to the extent of our knowledge. Thus, substrates containing ^{12}C atoms unadulterated by natural abundance are not a practical option in designing ILEs; therefore, researchers have to resort to cheap, naturally abundant substrates. With the introduction of natural ^{13}C abundance, ILEs **(a)** and **(b)** are no longer mirror images of each other nor can be analyzed equivalently using the cumomer or EMU concepts. The originomer concept is able to establish an equivalence between the two ILEs, so that cumulative originomer solutions can be stopped at order 2 with negligible loss in accuracy. As discussed below, the exclusion of higher-order labeling variables is especially beneficial in iMFA that employs analytically solutions.

2.3.2. The originomer concept facilitates time-saving analytical solutions for iMFA

Using the TCA cycle-glyoxylate shunt network as an example, we compared relative errors and computation times of the (cumulative) originomer and (cumulative) isotopomer approaches, as well as between analytical and numerical solutions (**Fig. 2.7**). The relative errors of numerical solutions are represented by the difference between calculation results by numerical solutions and by analytical solutions. As expected, the relative error of the numerical solution increases significantly as the tolerance increases (**Fig. 2.7** bars/left axis). However, the number of integrations steps is not sensitive to tolerance, decreasing from 113 to 105 as the tolerance increases by three orders of magnitude (**Fig. 2.7** squares/right axis). A comparison of computational times (**Fig. 2.7** triangles/right axis) reveals that the analytical solution method is ~20-fold more efficient than the numerical method at 0.001 tolerance for this network (2.34 s vs. ~48.0 s for a system of 11 originomer balance equations). Moreover, the results also show that using the analytical solution with originomer decomposition leads to 2.56-fold faster than with cumomer decomposition (2.34 s vs. 6.00 s).

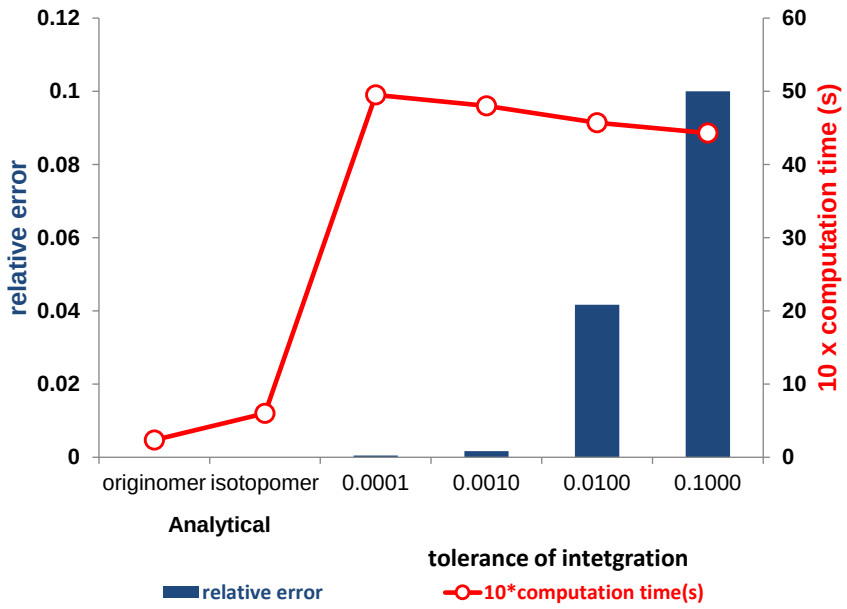


Figure 2.7. Originomer-based solutions are faster than isotopomer-based ones; analytical solutions are faster than numerical ones.
The left axis and histograms depict the relative errors of the analytical method employing (cumulative) originomers or isotopomers when compared with the analytical method. The right axis shows the number of integration steps (squares) and 10 × the computation times (triangles) for the various methods and their combinations. We estimated computation time as explained in Methods.

2.4. Discussion

In this work, we introduced two new concepts: a new labeling variable called originomer and a new analytical solution method for iMFA. The originomer concept factors out natural abundance from ILE calculations, and thus: (i) reduces the number of labeling variables and (ii) facilitates a more intuitive, metabolism-centric interpretation of isotope labeling data. Thereby, it enables faster computation *without loss of accuracy* as well as more intuitive experimental design. The analytical solution method for iMFA can, for many commonly performed ILEs, substantially cut down computation time.

2.4.1. Originomers reduce computation time.

There are several advantages of employing the originomer concept and its derived concepts cumulative originomers and EOUs over isotopomers and its derived concepts cumomers and EMUs. Foremost, in an isotopomer or cumomer model, a complete set of isotopomers or cumomers of every metabolite in the network must be included because of the natural abundance of ^{13}C, ^{15}N or other unusual isotopes. (The exceptions to this are rare cases wherein the abundance of higher-order cumomers is so small that they need not be simulated.) By contrast, natural labeling abundance does not complicate originomer balancing since it is the origins of the atoms in a metabolite that are considered rather than the labeling states. Therefore, several originomers are infeasible in many typical metabolic networks, as we demonstrated for network 1 (**Fig. 2.4**) and network 2 (**Fig. 2.5**). Secondly, even if most or all originomers in a network are feasible, choosing an appropriate special atom can lead to originomer abundances that are concentrated in the first few orders, so that the balance equations need not be solved for higher orders. We exemplified this in network 3, which is a typical scenario in mixotrophic metabolism.

It is highly desirable to reduce computation time in MFA and iMFA because of the large number of operations inherent in the evaluation of metabolic fluxes from isotope labeling data. Even at isotopic steady state, this process is a challenging parameter estimation problem in an extensive parameter space (Antoniewicz et al., 2006), wherein the fluxes are parameters that have to be iteratively estimated from

isotope label data and other measurements. For example, accurate flux evaluation for a medium-sized plant metabolic network (with ~1000 isotopomers, ~100 fluxes and ~30 flux parameters) requires hundreds or thousands of iterations of this parameter estimation process. Each iteration simulates isotopomer or (cumomer or EMU) abundances from guessed flux values using matrix inversions that require $O(N^3)$ operations where N is approximately the number of labeling variables balanced. Furthermore, this cycle of iterations has to be repeated several times to calculate flux confidence intervals or to check for multiple flux solutions. Thus, reducing the number of labeling variables has enormous impact on the computation time.

2.4.2. Originomers permit a more intuitive, metabolism-centric interpretation of isotope labeling data.

The originomer and cumulative originomer concepts track the metabolic origins of atoms in a network, which depend only on metabolism (i.e. the structure of the metabolic network and the values of fluxes in it). On the other hand, the isotopomer and cumomer concepts track isotopes through a network, which depend not only on metabolism but also on the natural abundance of isotopes such as ^{13}C and ^{15}N. Thus, the originomer concept decouples (or "cleans") the effects of natural abundance from those of metabolism. Although the EMU concept is more metabolism-centric than the cumomer concept, it still does not decouple natural abundance from metabolism. Additionally, because originomers and their abundances depend only on metabolism, they permit a more intuitive interpretation of labeling data. For example, if a researcher conducting a mixotrophic experiment with U-^{13}C

glucose and naturally abundant CO_2 wishes to evaluate how much carbon the two substrates contribute to various metabolites, the answer can be *directly* obtained from originomer abundances. Furthermore, because many originomers will be infeasible in real metabolic networks, measuring a significant abundance for these originomers will unequivocally indicate that the metabolic network is incorrect and needs to be refined. Moreover, we anticipate that new EMU-based methods for the design of isotope labeling experiments (Crown et al., 2012; Crown and Antoniewicz, 2012a) may benefit from the fact that originomers and EOUs are metabolism-centric. Finally, in the analysis of network 3 (**Fig. 2.6**; see **Results**), we observed that two ILEs that are mirror images of each other, e.g. {U-^{13}C glucose and $^{12}CO_2$} vs. {U-^{12}C glucose and $^{13}CO_2$}, are exactly equivalent. Thus, the ILE whose analysis is mathematically more convenient may be analyzed instead of the other. However, it is practical to use naturally abundant substrates instead of pure ^{12}C substrates and the equivalence between the ILEs disappears when ^{12}C is adulterated by natural abundance. Nonetheless, the originomer concept still establishes an equivalence between the ILEs, as we demonstrated with network 3.

2.4.3. Originomers permit analytical solutions to iMFA

We established an analytical solution strategy for iMFA that enables obtaining explicit analytical solutions for labeling variables. These expressions only need to be evaluated at select time points at which metabolites were sampled. This is a tremendous advantage over the currently used numerical method, in which the solution of ODEs involving thousands of variables within each iteration of an

optimization loop becomes a major bottleneck. However, a potential disadvantage of our analytical solution strategy is that the analytical expressions for labeling variables could involve a very large number of exponential terms. Since many originomers in a metabolic network are clearly infeasible, they enable a realistic implementation of this analytical solution strategy. Furthermore, as we have shown, the analytical solutions are exact whereas the numerical ones are slightly inexact.

2.4.4. Cases when the originomer method is not applicable

As demonstrated in our examples, the originomer method is best applied to ILEs in which a single label of a single carbon source is fed. Thus, an *apparent* shortcoming of this concept is that it is not applicable to ILEs where multiple labels of a carbon source (e.g. 1-^{13}C glucose and U-^{13}C glucose) are fed as a mixture. However, recent studies on ILE design have shown that feeding multiple labels in parallel experiments provides more information than feeding a mixture of the labels in a single experiment (Crown and Antoniewicz, 2012b; Nargund and Sriram, 2013). This is very likely because when the labels are fed together, isotopomers arising from one label mask those arising from the other (Nargund and Sriram, 2013). Thus, performing parallel ILEs, each with a single label, is actually desirable. The originomer concept becomes very appropriate here, because each of the parallel ILEs can now be analyzed by designating a unique special atom (or set of special atoms) appropriate to that ILE. The only exception would be an ILE in which a U-^{13}C substrate diluted with its naturally abundant version is fed; the bondomer concept

(Sriram and Shanks, 2004; van Winden et al., 2002) provides a metabolism-centric approach for this type of ILE.

2.4.5. Analytical solution method for iMFA

We showed that the analytical solution method improves computational efficiency by 23-fold over the numerical method. We substantiated this by estimating the number of operations required for the analytical and numerical solution methods for network 1, and found that the analytical method requires 18-fold lesser operations than the numerical method.

A major barrier in that analytical solution method is the huge number of exponential terms involved in Eq. [2.12]. In the homogeneous part of the ODE, or the first term of Eq. [2.12], the number the exponential terms is equal to the number of variables at that cumulative originomer order. However, in the inhomogeneous part, or second term, that number is determined by the number of exponential terms present in β_k, which arises from the bilinear term $\mathbf{b}$ in the balance Eq. [2.7]. Entries of the vector $\mathbf{b}$ are multiplications of two lower orders cumulative originomer abundances, and this is the major cause for the numerous exponential terms. For example, in the condensation reaction $A + B \rightarrow C$, the cumulative originomer abundances of C will depend on those of A and B. If the analytical expressions for certain same-order cumulative originomers of A and B have 100 exponential terms each, the inhomogeneous part of the corresponding cumulative originomer of C will consist of at least $C_{101}^2 = 5050$ exponential terms, and more if different-order

cumomers of A and B are being considered. This problem is the intrinsic bottleneck of the analytical solution method. Even if this does not contribute much to the computational cost, significant memory must be allocated to store these exponential terms, especially for higher-order cumulative originomers. Fortunately, it is in such cases that originomer decomposition shows obvious superiority over isotopomer decomposition. Once an appropriate special atom is chosen, many cumulative originomers are either infeasible or most higher-order cumulative originomer abundances have negligible abundances, making the analytical solution still feasible as we demonstrated for network 1.

2.5. Conclusions

In this work we presented two novel approaches to facilitate faster computation and more intuitive interpretation of ILEs. The originomer and its extension, the cumulative originomer, factor out natural abundances of unusual isotopes. At best, the cumulative originomer concept significantly reduces either the total number of labeling variables in an ILE or the number of higher-order labeling variables. At worst, this concept yields the same number of equations as the popular cumomer concept. Furthermore, it is possible to extend this concept to an EMU-like concept called EOU, which has further potential to enhance computational efficiency. The originomer and its derivative concepts can find widespread application in isotopic steady-state and instationary ILEs, from single substrate to parallel, multi-substrate scenarios. We also developed an analytical solution method for ODEs arising in iMFA. This method is feasible and advantageous for many commonly employed instationary ILEs. The superiority of analytical solutions over numerical

solutions results from the fact that the analytical expressions need to be evaluated only for the time points at which metabolites were sampled instead of at every integration step. The analytical solution method can result in large numbers of exponential terms; however, this problem is alleviated by combining it with the originomer concept. Our simulation results show that for typical networks, this combination reduces computation time by over 90%.

Chapter 3: Flux and reflux: Metabolite reflux in plant suspension cells and its implications on isotope-assisted metabolic flux analysis Title of Chapter 3

This chapter is part of an article that has been published in journal *Molecular Biosystems*. DOI: 10.1039/C3MB70348G.

Shilpa Nargund[1,2,&], Ashish Misra[1,3,&], Xiaofeng Zhang[1,&], Gary D. Coleman[4] and Ganesh Sriram[1,5]

SN, AM, XZ, GDC and GS conceived this study; SN, AM, XZ and GS designed it. GDC provided poplar cell suspension cultures; SN, AM and XZ jointly performed experiments and analyzed data. XZ wrote substantial portions of the manuscript. XZ and GS revised the manuscript after peer review. SN, AM, XZ contributed equally;

[1]Department of Chemical and Biomolecular Engineering, University of Maryland Current affiliations: [2]Syngene International Limited, Bengaluru, India; [3]DBT-ICT Centre for Energy Biosciences, Institute of Chemical Technology, Mumbai, India [4]Department of Plant Science & Landscape Architecture, University of Maryland [5]Principal investigator and corresponding author

Abstract

Isotope-assisted metabolic flux analysis (MFA) is a powerful methodology to quantify intracellular fluxes through isotope labeling experiments (ILEs). In batch cultures, which are often convenient, inexpensive or inevitable especially for eukaryotic systems, MFA is complicated by the presence of the initially present biomass. This unlabeled biomass may either mix with the newly synthesized labeled biomass or reflux into the metabolic network, thus masking the true labeling patterns in the newly synthesized biomass. Here, we report a detailed investigation of such metabolite reflux in cell suspensions of the tree poplar. In ILEs supplying 28% or 98% U-^{13}C glucose as the sole organic carbon source, biomass components exhibited lower ^{13}C enrichments than the supplied glucose as well as anomalous isotopomers not explainable by simple mixing of the initial and newly synthesized biomass. These anomalous labeling patterns were most prominent in a 98% U-^{13}C glucose ILE. By comparing the performance of light- and dark- grown cells as well as by analyzing the isotope labeling patterns in aspartic and glutamic acids, we eliminated photosynthetic or anaplerotic fixation of extracellular $^{12}CO_2$ as explanations for the anomalous labeling patterns. We further investigated four different metabolic models for interpreting the labeling patterns and evaluating fluxes: (i) a carbon source (glucose) dilution model, (ii) an isotopomer correction model with uniform dilution for all amino acids, (iii) an isotopomer correction model with variable dilution for different amino acids, and (iv) a comprehensive metabolite reflux model. Of these, the metabolite reflux model provided a substantially better fit for the observed labeling patterns (sum of squared residues: 538) than the other three models whose sum of

squared residues were (i) 4,626, (ii) 4,983, and (iii) 1,748, respectively. We compared fluxes determined by the metabolite reflux model to those determined by an independent methodology involving an excessively long ILE to wash out initial biomass and a minimal reflux model. This comparison showed identical or similar distributions for a majority of fluxes, thus validating our comprehensive reflux model. In summary, we have demonstrated the need for quantifying interactions between initially present biomass and newly synthesized biomass in batch ILEs, especially through the use of ≈100% U-^{13}C carbon sources. Our ILEs reveal a high amount of metabolite reflux in poplar cell suspensions, which is well explained by a comprehensive metabolite reflux model.

3.1. Introduction

Isotope-assisted metabolic flux analysis (isotope MFA) is a powerful tool for investigating metabolic networks because it can estimate intracellular chemical reaction rates (fluxes) (Ahn and Antoniewicz, 2011; Kruger and Ratcliffe, 2009; Leighty and Antoniewicz, 2012; Schwender, 2008) and elucidate new metabolic pathways (Schwender et al., 2004; Sriram et al., 2007a). The basis of this methodology is the isotope labeling experiment (ILE), which involves supplying one or more isotopically labeled substrates to a cell or tissue culture, followed by measurement of isotope labeling patterns (isotopomers) in synthesized products such as biomass components or intracellular metabolites (Wiechert, 2001). Isotope MFA involves interpretation of these labeling patterns through a mathematical model of metabolism to ultimately result in metabolic flux maps for the cell or tissue of interest.

This modeling component serves as a bridge that connects the measurable quantities in an ILE, including isotopic labeling patterns and extracellular or biomass fluxes, to the underlying intracellular fluxes.

Two major approaches in isotope MFA are (i) steady-state MFA (e.g. Wiechert, 2001), in which the labeling states of cellular products are measured after isotopic steady state is reached, and (ii) instationary MFA (e.g. Wiechert and Noh, 2005; Wiechert and Nöh, 2013), in which the labeling states are intentionally measured during the approach to isotopic steady state. Of these, steady-state MFA is currently more widespread due to two reasons: first, its experimental implementation is easier and second, it is computationally less challenging as it does not involve solutions of large systems of ordinary differential equations within a global optimization algorithm (Nöh and Wiechert, 2006; Wiechert and Noh, 2005). For both types of MFA, it is ideal to measure isotope labeling patterns in intracellular metabolites. However, in steady-state isotope MFA, biomass components such as proteinogenic amino acids and sugars are used as surrogates for intracellular metabolites (Dauner et al., 2001; Sriram et al., 2008, 2004; Szyperski, 1995). This is because these compounds are synthesized from intracellular metabolite precursors distributed throughout primary metabolism. Furthermore, cellular protein and sugars are much more abundant than intracellular metabolites and can be hydrolyzed to amino acids (Klapa et al., 2003) or sugar hydrolysates (Sriram et al., 2007b) directly without employing any extraction steps. In this case, information on fluxes is contained in the isotopically labeled biomass synthesized *de novo* during the ILE

from the supplied carbon sources. Biomass present before the commencement of the ILE (hereafter, "initial biomass") is labeled at the natural ^{13}C abundance of 1.1%, and is thus effectively unlabeled. Often, this initial biomass can constitute up to ~10% of the biomass harvested at the end of the ILE and can interfere with the interpretation of labeling patterns. Continuous culture, such as in a chemostat, affords the washing out of the initial biomass and has been employed for many microbial systems. However, continuous culture is infeasible, unsuitable or expensive for many plant, mammalian and even microbial cell types, leaving batch culture as the only available option. In fact, steady-state isotope MFA in batch culture has been reported in studies investigating the metabolic landscapes of the bacterium *E. coli* (Chen et al., 2011; Shen et al., 2013), cell cultures of the plant *Arabidopsis thaliana* (Masakapalli et al., 2013; Nargund and Sriram, 2013) and various mammalian cells (Metallo et al., 2009; Omasa et al., 2009; Quek et al., 2010).

Unless batch culture is performed for long periods of time extending into multiple subculture cycles, the naturally abundant initial biomass can never be washed out. Thus, the biomass harvested at the conclusion of the ILE will consist of a mixture of initial (naturally abundant) and newly synthesized (^{13}C-labeled) portions (**Fig. 3.1**). This raises an important question: does the initial biomass interact with the newly synthesized biomass? If there is no interaction, but only simple mixing between the initial and newly synthesized biomass, the ratio of the initial biomass to the newly synthesized biomass can be used to factor out the isotopomer abundances of the initial biomass from those measured in the harvested biomass. However,

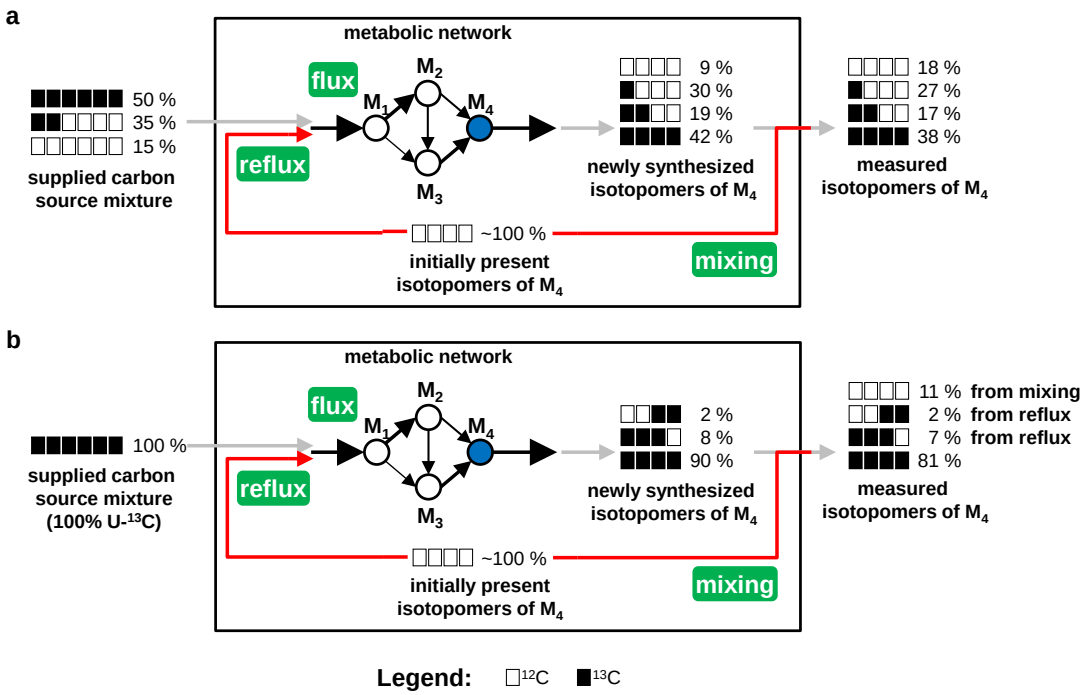


Figure 3.1. Metabolic flux and reflux in isotope labeling experiments (ILEs). (a) In ILEs, a combination of isotopically labeled (^{13}C or ^{12}C) carbon sources is fed to a metabolic network. The metabolite molecules synthesized by the network from the ^{13}C or ^{12}C atoms of the carbon sources will consist of isotopically distinct isomers (isotopomers). Information on intracellular fluxes that is concealed in the isotopomer patterns and abundances can be unraveled by mathematical modeling. In batch ILEs, the biomass present at the beginning of the ILE may interfere with the ILE. This initial biomass, labeled at the natural abundance of ^{13}C (~1.1%), may either mix with the newly synthesized isotopomers ("mixing" arrow) or reflux into the metabolic network ("reflux" arrow). The isotopomers of a hypothetical four-carbon metabolite M_4 illustrate these interactions. (b) An ILE supplying a single, ~100% U-^{13}C carbon source will produce isotopomers that clearly distinguish mixing from reflux, as well as quantify the extent of these interactions. In such an ILE, biomass that is newly synthesized from the supplied carbon source will be represented by U-^{13}C isotopomers. Mixing of the initial biomass with the newly synthesized biomass will result in isotopomers that are almost entirely ^{12}C, whereas reflux of the initial biomass will result in isotopomers that are predominantly ^{13}C but contain a few ^{12}C atoms. Small rectangles represent carbon atoms (black: ^{13}C, white: ^{12}C). Small circles represent entire metabolite molecules. The newly synthesized and initially present intracellular isotopomers correspond the metabolite M_4, shown shaded in gray.

interaction between the initial biomass and the newly synthesized biomass (or "reflux" of the initial biomass into the metabolic network) can result in additional labeling patterns that can only be factored out by a metabolic model that mimics this interaction. Previous isotope MFA studies have used data correction strategies toward factoring out the effect of the initial biomass. A straightforward method is to assume that the initial biomass does not reflux into the metabolic network and that the measured isotopomers are a weighted average of the known, naturally abundant isotopomers of the initial biomass and the unknown, isotopomers of the newly synthesized biomass. In this case, a simple algebraic operation decouples the isotopomer abundances of the initial biomass from those of the newly synthesized biomass. This "dilution rate" method has been reported in multiple studies (Dauner et al., 2001; Sriram et al., 2007a, 2004). Another approach is to assume a small influx of the naturally abundant version of the supplied carbon source (e.g. glucose), the value of this influx being a parameter estimated by MFA (M. Antoniewicz et al., 2007). Recently, researchers have begun hypothesizing interactions between the initial biomass and the newly synthesized biomass, and modeling these interactions as influxes of unlabeled amino acids. Such models either assume identical influxes for each amino acid on the basis of the culture growth rate (Lonien and Schwender, 2009), or different influxes for each amino acid, which are estimated as part of the flux evaluation procedure (Kruger et al., 2011; Masakapalli et al., 2013). Although these solutions have partially solved the intrinsic shortcoming of steady state MFA in batch-like cell cultures resulted from initial biomass, their validity, performance and

ability to obtain accurate flux estimates have not yet been comprehensively investigated.

We report an experimental and computational investigation of the interactions between initial and newly synthesized biomass, or metabolite reflux, in batch-grown cell suspensions of the forest tree poplar, a potential biofuel feedstock(Sannigrahi et al., 2010). We performed ILEs by supplying three labeled varieties of ^{13}C glucose. Of these, a 98% U-^{13}C glucose ILE revealed the presence of numerous, anomalous mass isotopomers that appeared to have been formed due to metabolite reflux. By comparing labeling patterns in light- and dark- grown cells as well as by interpreting the labeling patterns of tricarboxylic acid (TCA) cycle metabolites, we determined that these mass isotopomers were neither contributed by photosynthesis nor by anaplerotic CO_2 assimilation. On elimination of these processes, metabolite reflux remained the sole cause for the anomalous mass isotopomers. We then compared the ability of three previously reported flux-isotopomer models to account for the observed mass isotopomer data. These models accounted for the anomalies by (i) using a uniform or non-uniform (for different amino acids) dilution rate to correct the isotopomer abundances, (ii) assuming a small influx of naturally abundant glucose, or (iii) accounting for the reflux by full-fledged reflux model in which reflux is mimicked by influxes of several amino acids and metabolites. Our results show significant metabolite reflux in poplar suspension cells, with the 98% U-^{13}C glucose ILE being most sensitive to the effects of reflux. Among the four models we examined, the comprehensive reflux model accounts best for the observed anomalies in the labeling patterns. While flux results from this model compare qualitatively with

those obtained from the labeling patterns of intracellular metabolites, there are some numerical differences between fluxes. This points to the need for future extensions of the metabolite reflux model.

3.2. Materials and Methods

3.2.1. Poplar cell suspension culture and growth rate measurement

Poplar (*Populus. tremula* × *Populus. alba* clone 717-1B4) cell suspensions were grown at 20 °C in 125 mL Erlenmeyer flasks shaken at 125 rpm on an orbital shaker. The flasks were placed in continuous light from cool-white fluorescent light (Ecolux® Technology Plant and Aquarium F40T12 bulbs (photosynthetically active radiation of 200-300 μmol m^{-2} s^{-1}, incident radiation on flasks measured to be 28.3 ± 2.6 μmol $m^{-2}s^{-1}$) or continuous dark; the dark condition was realized by wrapping the flasks in aluminum foil. The suspensions were subcultured every 7 d by transferring 600 mg cells into 30 mL Murashige and Skoog medium (Phytotechnology Laboratories, Shawnee Mission, KS) containing 20 g L^{-1} glucose as the sole organic carbon source.

To quantify growth, several flasks of cells were subcultured in parallel. Three biological replicates were harvested at 7 d by vacuum filtration through glass microfiber filter paper (Whatman, Piscataway, NJ). Fresh and dry weights of the harvested cell pellets were obtained by weighing them before and after lyophilization, respectively.

3.2.2. Parallel ILEs

We performed three parallel ILEs by growing cells in continuous light, replacing the usual 20 g L^{-1} naturally abundant glucose in the medium with 20 g L^{-1} of: (i) 98% U-^{13}C glucose, (ii) 100% 1-^{13}C glucose or (iii) a mixture of 28% U-^{13}C and 72% naturally abundant glucose. The 98% U-^{13}C glucose ILE was also performed in continuous dark. Isotopically labeled glucose was obtained from Cambridge Isotope Laboratories (Andover, MA). Each ILE was represented by three biological duplicates, which were harvested on day 7 of culture. Cells were harvested from the ILEs by vacuum filtration through glass microfiber filter paper, and were immediately frozen in liquid nitrogen to arrest metabolism. The frozen cell pellets were then lyophilized and stored at –80°C until further analysis. We further performed three 21-d ILEs by supplying the same three types of ^{13}C labeled glucose. In these ILEs, we transferred cells to fresh media every 7 d, harvested cells at 21 d and analyzed the mass isotopomers of soluble metabolites (free amino acids) from the harvested cells.

3.2.3. Measurement of isotopomer abundances of proteinogenic amino acids and intracellular metabolites

To analyze proteinogenic amino acids, 20 mg of ground, lyophilized cells were vacuum-hydrolyzed with 6 N HCl for 5 h at 160 °C in hydrolysis tubes (Thermo Fisher Scientific, Rockford, IL). The hydrolysate was dried overnight in a RapidVap evaporator (Labconco, Kansas City, MO), reconstituted in 1 mL water, filtered, lyophilized and stored in -80 °C until further analysis. To improve their volatility in

the GC, the amino acids were derivatized with 100 µL of N-(tert-butyldimethylsilyl)-N-methyltrifluoroacetamide (MTBSTFA; Sigma-Aldrich, St. Louis, MO) in 100 µL dimethylformamide (Thermo Fisher Scientific) by heating at 70 °C for 1 h. Intracellular metabolites were extracted and quantified as described by Fiehn and co-workers(Fiehn et al., 2000). Briefly, 20 mg of ground, lyophilized cells were contacted with 1 mL methanol and 50 µL water, heated at 75 °C for 15 min and then centrifuged, after which the supernatant containing the soluble metabolites was collected. These steps were repeated once, following which the supernatant was dried, re-dissolved in 100 µL of a 20 mg mL^{-1} solution of methoxyamine hydrochloride (MP Biomedicals, Solon, OH) in pyridine (EMD Chemicals, Gibbstown, NJ) and heated at 30 °C for 90 min. The resulting soluble metabolites were derivatized by adding 100 µl MTBSTFA and heating at 70 °C for 1 hr. Soluble metabolites were quantified by gas-chromatography as described below, using norleucine (Sigma-Aldrich, St. Louis, MO) as an internal standard.

A Varian 450 gas chromatograph (GC) connected to a Varian 300 mass spectrometer (MS) (Bruker Corporation, Fremont, CA) was used to measure the abundances and mass isotopomer distributions of amino acids. The GC was equipped with a VF-5ms column (30 m × 0.25 mm × 0.25 µm; Bruker). One microliter of each sample was injected in to the GC at a split ratio of 50:1, with helium as the carrier gas at a flow rate at 1.0 mL min^{-1}. To analyze proteinogenic amino acids, the oven temperature was programmed as follows: an initial temperature of 150 °C for 2 min, increased to 230 °C at 3 °C min^{-1}, then to 240 °C at 2 °C min^{-1} and finally to 275 °C

at 10 °C min^{-1} and a constant temperature of 275 °C for 6 min, amounting to a run time of 43.2 min. To analyze intracellular metabolites, the oven temperature was programmed as follows: an initial temperature of 150 °C for 2 min, increased to 250 °C at 3 °C min^{-1}, then finally to 275 °C at 10 °C min^{-1} and a constant temperature of 275 °C for 5 min. The MS was operated in electron ionization mode, with the ion source at 280 °C. Isotopomer abundances were recorded under selected ion monitoring mode.

The Varian MS Workstation software (version 6.9.3) was used together with the NIST mass spectral library (National Institute of Standards and Technology, Gaithersburg, MD) to identify and quantify the peaks in the MS spectra. An in-house MATLAB (The Mathworks, Natick, MA) script(Sriram et al., 2008) was used to correct the mass isotopomer distributions for the natural abundances of isotopes of hydrogen, nitrogen, oxygen, sulfur, silicon and non-metabolic carbon. By acquiring and processing spectra of amino acid standards of defined compositions, we verified that both the MS measurements and the natural abundance correction program were accurate (data not shown). The thus corrected mass isotopomer distributions from the three parallel ILEs are listed in Supplementary **Table S1**. This correction different from the data correction strategy described in Introduction because the data correction strategy aims at elucidating the effects of initial ^{12}C naturally abundant biomass in a ILE, while this MATLAB program corrects for all non-metabolic carbon atoms in a measurable molecule resulted from derivatization.

3.2.4. Flux evaluation from isotopomer data

To estimate intracellular metabolic fluxes, the corrected mass isotopomer distributions were processed by our flux evaluation program NMR2Flux+. Details about NMR2Flux+ and its capabilities are described elsewhere(Nargund and Sriram, 2013; Sriram et al., 2008, 2004). Briefly, this program accepts metabolic network stoichiometries and carbon atom rearrangements supplied by the user, employs cumomer balancing to simulate isotopomer distributions and isotopomer abundances corresponding to any feasible set of fluxes in the network. Given a set of isotopomer abundances, the program uses a global optimization routine based on simulated annealing to iteratively evaluate flux values that best account for the supplied isotopomer abundances. NMR2Flux+ assesses goodness-of-fit using the sum of squared residuals (SSR) metric, calculated as:

$$\mathrm{SSR} = \sum_{j} \frac{\left(I_j - I_j^x\right)^2}{\sigma_j^{\,2}}$$

where the index j cycles through all the (mass) isotopomer abundances, I^x represents measured isotopomer abundances, I represents simulated isotopomer abundances corresponding to the evaluated set of fluxes and σ represents the measured standard deviations of the isotopomer abundances. For all metabolic models, the SSR was evaluated for 405 isotopomer abundances (see Supplementary **Table S1**) measured from three ILEs on 98% 1-^{13}C glucose, 28% U-^{13}C glucose and 98% U-^{13}C glucose. The dataset used for flux evaluation includes all the proteinogenic amino acid fragments detected by GC-MS (a total of 30 fragments across 12 amino acids: Ala, Asp, Gly, Glu, His, Ile, Leu, Phe, Pro, Ser, Thr, Val; Supplementary **Table S1**).

To obtain standard deviations for the evaluated fluxes, we used a bootstrap Monte-Carlo algorithm that operated as follows: (i) the original isotopomer abundance dataset was perturbed 280 times, resulting in 280 synthetic datasets normally distributed around the original dataset; (ii) NMR2Flux+ independently evaluated fluxes from each synthetic dataset to obtain 280 synthetic flux distributions; (iii) these synthetic distributions were used to generate standard deviations for the evaluated fluxes. Fluxes and standard deviations evaluated by NMR2Flux+ for the four metabolic models in this work – uniform isotopomer correction (Iso_corr[U]), variable isotopomer correction model (Iso_corr[V]), glucose dilution (Gluc_dilu) and amino acid reflux (AA_in) (see Results for details) – are listed in Supplementary **Table S3**. The metabolic networks corresponding to these four models are listed in Supplementary **Tables S4, S5, S6** and **S7**, respectively. All NMR2Flux+ computations were performed on three identical workstations equipped with an Intel Xeon E5620 CPU and 32 GB of RAM.

3.3. Results

3.3.1. Proteinogenic amino acids from a 7-d batch ILE exhibit anomalous 13C enrichments and mass isotopomer distributions

The net ^{13}C enrichments (evaluated as explained in **Fig. 3.2**) of proteinogenic amino acid fragments from poplar suspension cultures grown on 28% or 98% U-^{13}C glucose for 7 d in batch cultures were considerably lower (22%-28%) (**Fig. 3.2**) than that of the supplied glucose. For growth on 28% U-^{13}C glucose, the enrichments ranged from 22%-23% for isoleucine, lysine and proline fragments to 27%-28% for glutamate and

alanine fragments. Proportionally, for growth on 98% U-^{13}C glucose, the enrichments ranged from 76% for isoleucine, lysine and proline fragments to 92%-98% for glutamate and alanine fragments. These enrichment data were reproduced in several labeling experiments performed weeks apart from each other. The observed ^{13}C

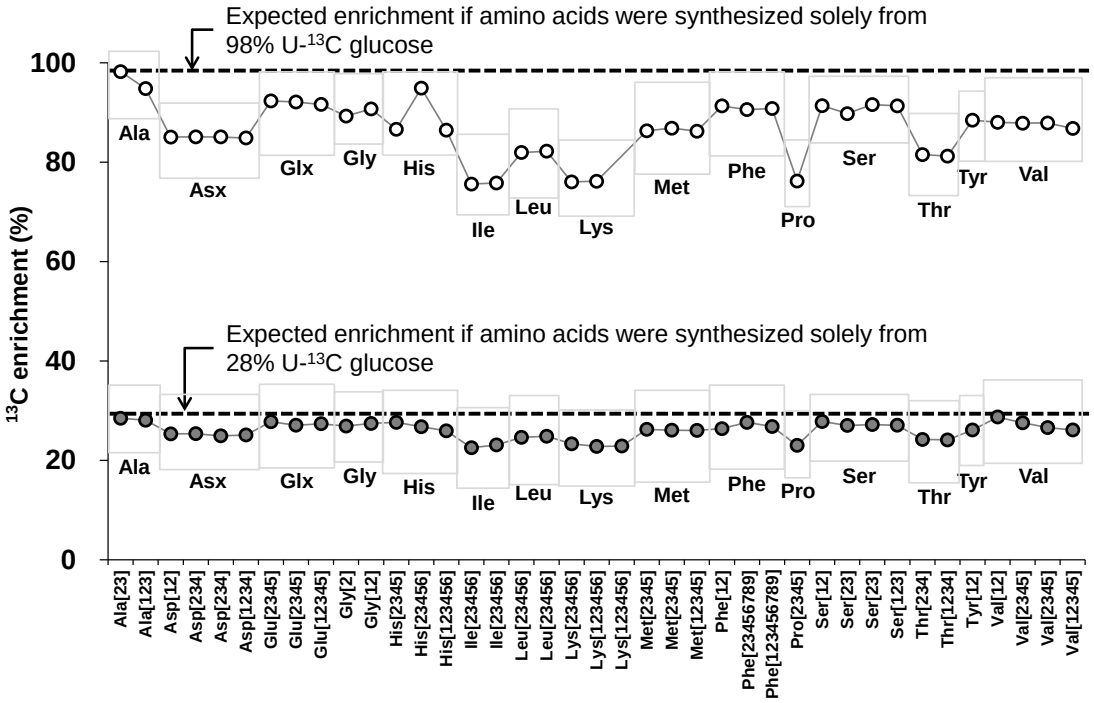


Figure 3.2. Proteinogenic amino acid fragments from poplar cell suspensions grown on U-^{13}C glucose are less^{13}C-enriched than the supplied glucose.
For a fragment containing *n* metabolic carbon atoms, the ^{13}C enrichment is defined as:

$$^{13}C\ enrichment = \frac{0 \times A_{m+0} + 1 \times A_{m+1} + \cdots + n \times A_{m+n}}{n}$$

where A_{m+i} is the (relative) abundance of a mass isotopomer in which *i* of the *n* carbon atoms are ^{13}C. Thus, the ^{13}C enrichment simply represents the fraction of atoms in a fragment that are ^{13}C. The ^{13}C enrichments of various proteinogenic amino acid mass spectral fragments (horizontal axis) were calculated from the mass isotopomer distributions of the fragments. Fragments belonging to the same amino acid are grouped within boxes. Two ILEs were performed, using 98% U-^{13}C glucose (white circles) or 28% U-^{13}C glucose (gray circles) as the sole organic carbon source. The dashed lines denote the ^{13}C enrichments expected for each ILE if the amino acids were entirely synthesized from the supplied glucose. All data points have error bars, although most error bars are too small to be visible.

enrichments are anomalous because they contradict the expectation that the non-photosynthetic poplar cell suspensions should synthesize biomass components entirely from the supplied glucose. A recent study (Wasylenko and Stephanopoulos, 2013) shows that enzymes can discriminate against ^{13}C atoms principally because ^{13}C atoms form stronger bonds than ^{12}C atoms, and that this isotopic effect can contribute up to 0.8% in GC-MS measurements when particular ^{13}C tracers are used. Nevertheless, the observed discrepant ^{13}C enrichments (up to 20%) are too large to be explained by the difference in enzyme affinities for ^{12}C and ^{13}C. Additionally, a previous study (Kruger et al., 2007) demonstrated that feeding ^{13}C-labeled carbon sources to Arabidopsis cell suspensions does not perturb metabolism.

One explanation for the discrepant ^{13}C enrichments is that they were due to mixing of initial biomass with the newly synthesized biomass without reflux of the initial biomass. However, an examination of the mass isotopomer abundances from a 98% U-^{13}C ILE (**Fig. 3.3** and Supplementary **Table S1**) revealed that this explanation is invalid. Mixing between the initial biomass and the newly synthesized biomass at a ratio of 1:9 (as obtained from dry weight measurements before and after the ILE) would result in the least massive (entirely ^{12}C) and most massive (entirely ^{13}C) mass isotopomers for each fragment (left panels of **Fig. 3.3**). Contrastingly, reflux of the initial biomass would result in breakdown of the initial biomass and condensation of the ensuing fragments with ^{13}C labeled fragments from the supplied glucose, thus producing a range of mass isotopomers for each fragment (right panels of **Fig. 3.3**).

The measured mass isotopomer abundances (center panels of **Fig. 3.3**) clearly agree with the latter possibility. This suggests reflux of the initial biomass in addition to simple mixing. It may appear that two competing hypotheses – photosynthetic or anaplerotic fixation of $^{12}CO_2$ – could potentially explain the anomalous isotopomers.

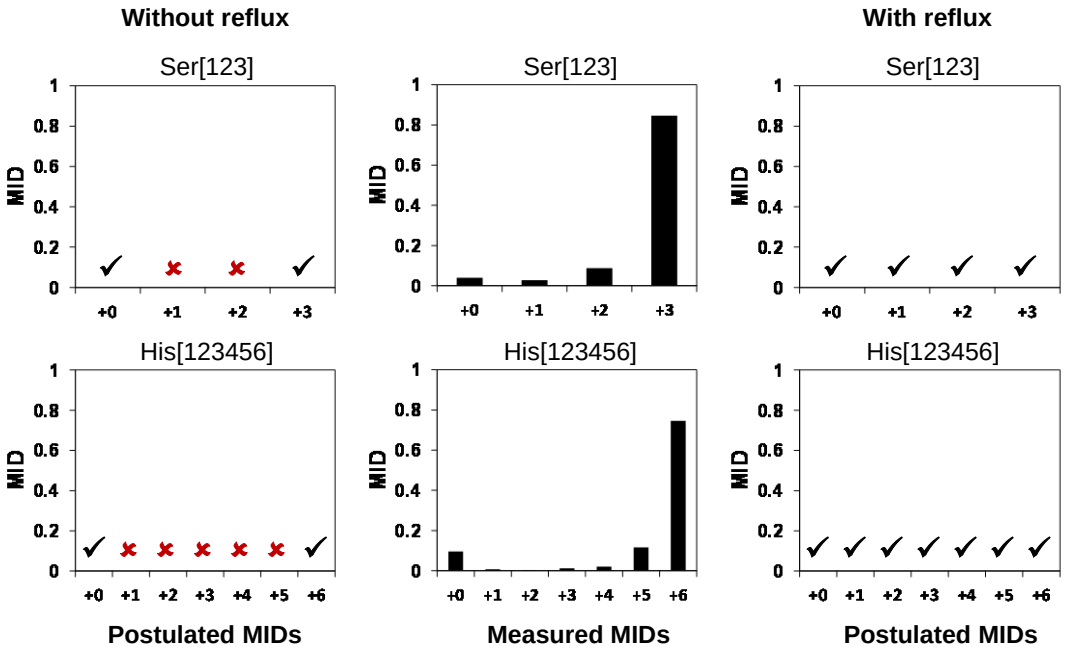


Figure 3.3. Metabolites from a 100% U-^{13}C ILE contain anomalous mass isotopomers not explained by mixing of initial and newly synthesized biomass. This figure depicts mass isotopomer abundances (MIDs) of two illustrative fragments Ser[123] and His[123456]. The MIDs on the left were postulated by assuming mixing between the initial biomass and the newly synthesized biomass (at a ratio of 1:9 obtained from dry weight measurements) but no metabolite reflux. In this case, the initial biomass is naturally abundant (lowest mass isotopomers) and the newly synthesized biomass is fully ^{13}C-labeled (highest mass isotopomers). The MIDs in the center were experimentally measured in a 98% U-^{13}C glucose ILE. The MIDs on the right were postulated assuming metabolite reflux (right). The mass isotopomers between the lowest and highest masses are anomalous and cannot be explained by mixing between the initial and newly synthesized biomass.
The black "✓" marks indicate isotopomers postulated to be feasible, whereas the red "×" marks indicate isotopomers postulated to be infeasible under the relevant scenario. All data points have error bars, although most error bars are too small to be visible.

However, as explained below, further experiments and a detailed examination of isotopomer data did not support these hypotheses.

3.3.2. Photosynthetic $^{12}CO2$ assimilation does not explain anomalous isotopomers

To assess if photosynthetic $^{12}CO_2$ fixation explained the anomalous isotopomers, we performed 98% U-^{13}C glucose ILEs for 7 d in continuous light or continuous dark. Photosynthetic $^{12}CO_2$ fixation would occur in the light but not in the dark. Therefore, if it were responsible for the anomalous ^{13}C enrichments of **Fig. 3.2**, only light-grown cells would exhibit substantially lower enrichments than the supplied glucose. Furthermore, since photosynthesis will result in $^{12}CO_2$ incorporation at C-1 of the photosynthetic product 3-phosphoglycerate, the light-grown cells should exhibit lower ^{13}C enrichment on C-1 of Ser (an amino acid derived from 3-phosphoglycerate) than dark-grown cells. Contrary to this expectation, intracellular metabolites (amino acids) from light- and dark-grown cell suspensions exhibited nearly identical ^{13}C enrichments (**Fig. 3.4**), including Ser. This clearly demonstrates that the discrepant ^{13}C enrichments cannot be explained by photosynthetic fixation of extracellular $^{12}CO_2$. Although the light- and dark-grown cell suspensions exhibit some differences between their glutamate and glutamine enrichments, this difference cannot be explained by photosynthetic CO_2 fixation as these metabolites are not generated in the Calvin cycle. Instead, it might be a result of a possible regulation of glutamine synthetase by light, which has been shown in multiple studies on *Arabidopsis thaliana* (Oliverira *et al*., 2001; Peterman and Goodman, 1991).

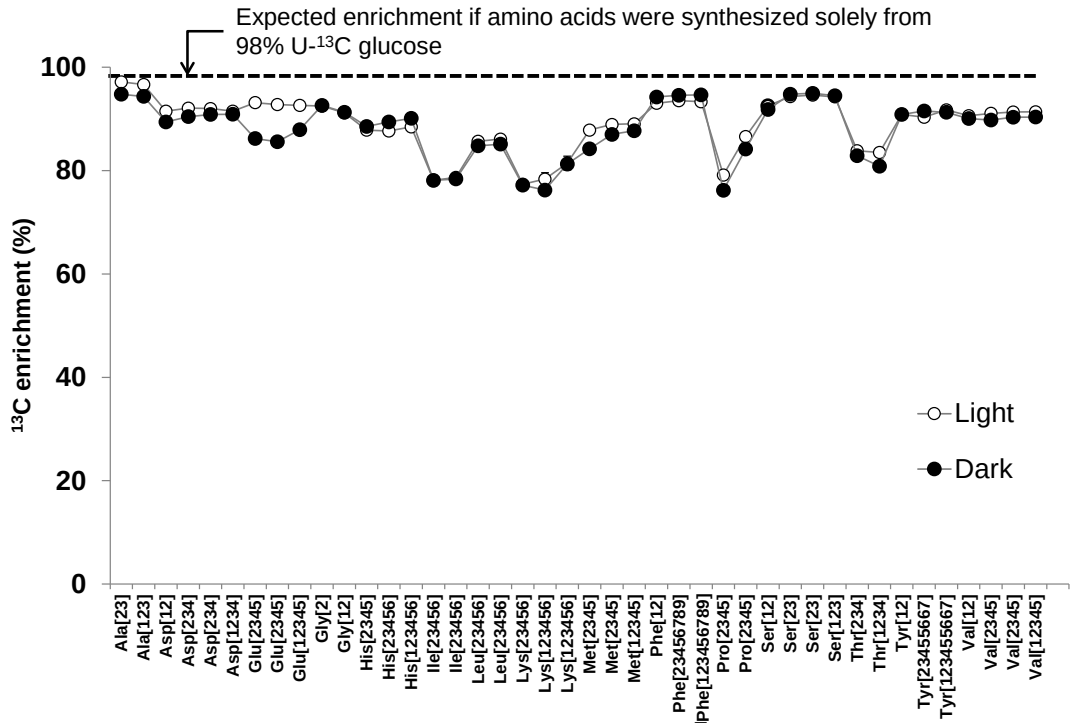


Figure 3.4. Photosynthetic $^{12}CO_2$ assimilation does not explain anomalous isotopomers.
The ^{13}C enrichments (vertical axis) of various proteinogenic amino acid mass spectral fragments (horizontal axis) were calculated from the mass isotopomer distributions of the fragments. The dashed lines denote the ^{13}C enrichments expected for each ILE if the amino acids were entirely synthesized from the supplied glucose.
Photosynthetic assimilation of atmospheric $^{12}CO_2$ may be one possible explanation for the anomalous isotopomers observed in **Fig. 3.3**. To test this, we measured ^{13}C enrichments of proteinogenic amino acid fragments between poplar cell suspensions grown on 98% U-^{13}C glucose in continuous light (white circles) and continuous dark (black circles). The similarity of ^{13}C enrichments between light and dark conditions indicates that photosynthetic carbon assimilation did not occur; therefore photosynthesis cannot explain the anomalous isotopomers. All data points have error bars, although most error bars are too small to be visible.

3.3.3. Anaplerotic 12CO2 assimilation does not fully explain anomalous isotopomers

Anaplerotic reactions interconvert three-carbon glycolytic metabolites such as phosphoenolpyruvate or pyruvate and four-carbon tricarboxylic acid (TCA) cycle metabolites such as oxaloacetate or malate, by fixing or releasing CO_2. Significant

flux through these reactions can be expected during growth as they are necessary for biomass synthesis and for shuttling reductants between organelles(Sriram et al., 2007a). To determine whether net anaplerotic incorporation of extracellular $^{12}CO_2$ was solely responsible for the anomalous isotopomers, we scrutinized the mass isotopomers of aspartate and glutamate (**Fig. 3.5**). The significant anomalous isotopomers of these amino acids include the *M*+1 isotopomers of aspartate[12] and aspartate[234], the *M*+2 and *M*+3 isotopomers of glutamate[2345] as well as the *M*+3 and *M*+4 isotopomers of glutamate[12345] (**Fig. 3.5c**).

In a 98% U-^{13}C glucose ILE, anaplerotic fixation of $^{12}CO_2$ would cause ^{13}C depletion of oxaloacetate (OAA; the keto acid precursor of aspartate) at C-4 and at C-1 due to exchange between OAA and the symmetrical molecules fumarate and succinate. It would also cause ^{13}C depletion of α-ketoglutarate (AKG; the keto acid precursor of glutamate) at C-1 due to the forward TCA cycle (**Fig. 3.5a**). Anaplerotic $^{12}CO_2$ fixation, even if followed by infinite turns of the TCA cycle, cannot result in ^{13}C depletion at C-2 or C-3 of OAA or aspartate nor at C-2, C-3, C-4 or C-5 of AKG or glutamate. Hence, anaplerotic $^{12}CO_2$ fixation cannot produce the *M*+1 isotopomer of aspartate[12] or aspartate[234], nor the *M*+1, *M*+2 or *M*+3 isotopomers of glutamate[2345] or glutamate[12345] (**Fig. 3.5b**). However, the significant observed abundances of these isotopomers imply that even if significant anaplerotic $^{12}CO_2$ occurred during the ILEs, it would not explain all anomalous isotopomers of aspartate and glutamate.

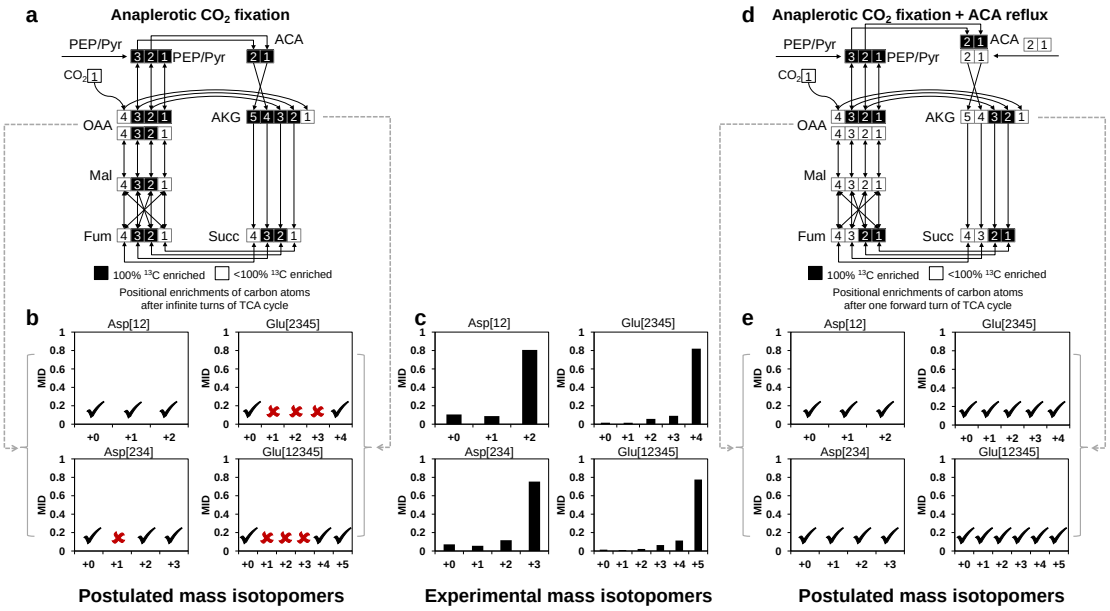


Figure 3.5. Anomalous isotopomers of glutamate (Glu) and aspartate (Asp) are not fully explained by anaplerotic $^{12}CO_2$ assimilation, but are explained by acetyl-CoA (ACA) reflux.

One possible explanation for the anomalous isotopomers in **Fig. 3.3** is that $^{12}CO_2$ was assimilated by anaplerotic reactions that interconvert three-carbon glycolytic metabolites (e.g. phosphoenolpyruvate or pyruvate) and four-carbon TCA cycle metabolites (e.g. oxaloacetate or malate). However, this possibility cannot explain all the anomalous isotopomers of Glu (derived from the metabolic precursor α-ketoglutarate [AKG]) and Asp (derived from the precursor oxaloacetate [OAA]). For example, in the 100% U-^{13}C ILE, anaplerotic $^{12}CO_2$ fixation **(a)** will result in AKG with ^{12}C on C-1 and OAA with ^{12}C on C-1 and/or C-4, even after infinite turns of the TCA cycle. This will result in mass isotopomer distributions of Glu and Asp **(b)** lacking containing only the isotopomers indicated with a black "✓" and lacking the mass isotopomers indicated with a red "✗". The isotopomer measurements from the 100% U-^{13}C ILE **(c)** do not agree with this prediction. An alternative model **(d)** featuring anaplerotic $^{12}CO_2$ fixation and reflux of acetyl-CoA (ACA, obtainable from the degradation of several amino acids) predicts isotopomer abundances **(e)** consistent with measurements. Small rectangles represent carbon atoms (black: ^{13}C, white: ^{12}C). On the horizontal axis, the numbers "+0", "+1", etc. represent *M*+0, *M*+1, etc. mass isotopomers. All data points have error bars, although most error bars are too small to be visible.

On exploring various metabolic models that could explain the anomalous isotopomers of aspartate and glutamate, we arrived at a model that included reflux of acetyl-CoA (ACA) in addition to anaplerotic $^{12}CO_2$ fixation (**Fig. 3.5d**). This is

consistent with amino acid reflux, as the degradation of many amino acids including glutamate, serine, alanine, threonine, valine and tyrosine produces pyruvate, ACA or fatty acids, and the degradation of fatty acids produces ACA (degradation pathways obtained from the MetaCyc metabolic pathway database (Caspi et al., 2011)). By tracing carbon atom rearrangements, it is clear that this model results in ^{13}C depletions at appropriate carbon atoms in the TCA cycle metabolites to satisfactorily explain all the anomalous isotopomers of aspartate and glutamate (**Fig. 3.5e**). Simulations using a metabolic model (development explained below) also showed that the anomalous isotopomers were best explained by the influx of naturally abundant ACA, but neither 1-^{13}C nor 2-^{13}C ACA. In summary, these results eliminated photosynthesis and anaplerotic $^{12}CO_2$ fixation as explanations for the anomalous ^{13}C enrichments and isotopomers, leaving only metabolite reflux as an explanation. In particular, the reflux of ACA that can be derived from degradation of amino acids and lipids, was crucial to explain several anomalous isotopomers.

3.3.4. Metabolic model incorporating reflux quantitatively explains isotopomer data better than previously reported models

We developed a comprehensive flux-isotopomer model to simulate isotopomer abundances and infer a carbon traffic in the batch-grown poplar cell suspensions (**Fig. 3.6**). Our model consists of plant primary metabolic pathways including glycolysis, the pentose phosphate pathway (PPP), the TCA cycle, multiple anaplerotic reactions as well as reactions leading from intracellular metabolites to biomass components such as proteinogenic amino acids and fatty acids. Furthermore

our model is compartmentalized, with the compartmental organization of pathways based on previous plant metabolic network models published by one of the us (GS) (Iyer et al., 2008; Nargund and Sriram, 2013; Sriram et al., 2007a, 2004) and one of the models evaluated by Ratcliffe and co-workers (Masakapalli et al., 2010). Accordingly, our model features duplicated (cytosolic and plastidic) glycolysis and pentose phosphate pathways connected by multiple intercompartmental transport fluxes as well as a mitochondrial TCA cycle. U-^{13}C glucose and $^{12}CO_2$ (though anaplerotic fixation) are the principal carbon sources in the model. The model incorporates metabolite reflux in the form of influxes of naturally abundant amino acids and metabolites resulting from the degradation of biomass components. As suggested by the anomalous isotopomers (**Fig. 3.3**), the amino acids entering the network are histidine, serine, glycine, leucine, alanine, aspartate/asparagine, threonine, methionine, isoleucine, leucine and glutamate/glutamine. The biomass component degradation products entering the model are fructose-6-phosphate (degradation product of sugars), cytosolic and plastidic pyruvate (degradation of serine, alanine, leucine and glutamate via citramalate pathway (Wu et al., 2009)), 3-phosphoglycerate (degradation of serine and glycine), succinate (degradation of valine) and ACA (degradation of fatty acids and acetate/propanoate resulting from the degradation of glutamate, threonine, valine and tyrosine). These influxes are marked in red in **Fig. 3.6**. The complete metabolic model incorporating metabolite reflux, hereafter designated “AA_in”, is listed in Supplementary **Table S7**.

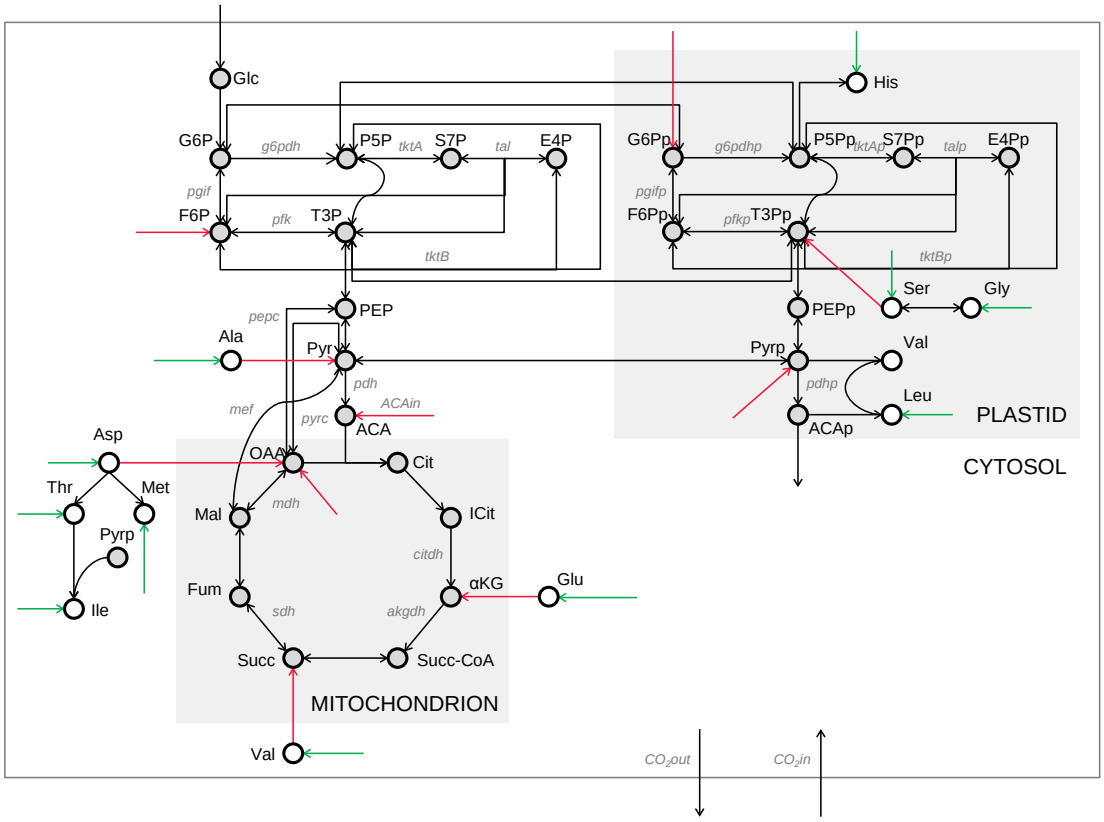


Figure 3.6. Metabolic network model incorporating amino acid reflux.
This three-compartment metabolic network is a model of primary metabolism in poplar cell suspension. It includes glycolysis and PPP in the cytosol and the plastid, TCA cycle in the mitochondrion as well as amino acids synthesis and reflux reactions. Gray circles represent intracellular metabolites and white circles represent amino acids derived from them. Arrows indicate biochemical reactions, with green arrows representing amino acid "dilution" or "mixing" and red arrows representing amino acid "reflux". Names of significant or interesting reactions (as designated by us) are shown adjacent to their arrows. The stoichiometries, carbon atom rearrangements and fluxes evaluated for all reactions are listed in Supplementary **Table S6** and **S7**. Standard three-letter abbreviations are used for amino acids. Abbreviations for other metabolite names: ACA, acetyl-CoA; AKG, α-ketoglutarate; Cit, citrate; E4P, erythrose-4-phosphate; F6P, fructose-6-phosphate; Fum, fumarate; G6P, glucose-6-phosphate; Glc, glucose; ICit, isocitrate; Mal, malate; OAA, oxaloacetate; P5P, pentose-5-phosphate (representing ribose-5-phosphate, ribulose-5-phosphate and xylulose-5-phosphate); PEP, phosphoenolpyruvate; Pyr, pyruvate; S7P, sedoheptulose-7-phosphate; Succ, succinate; Succ-CoA, succinyl-CoA; T3P, triose-3-phosphate (representing glyceraldehyde-3-phosphate, dihydroxyacetone phosphate and 3-phosphoglycerate). For metabolites duplicated in more than one compartment, compartmentalization is indicated by subscripts: c, cytosol; p, plastid.

We constructed this metabolic network model iteratively. Beginning with a simpler network, we attempted various refinements to the network and accepted those refinements that better accounted for the isotopomer abundance measurements (mathematically, those refinements that lowered the SSR of the model). In particular, a single-compartment model did not yield a satisfactory SSR, and introducing cytosol-versus-plastid compartmentalization lowered the SSR. Similarly, we introduced the amino acid influxes iteratively into successive versions of the model. We allowed the flux evaluation program NMR2Flux+ to find, through a simulated annealing global optimization algorithm, values of the metabolic pathway fluxes as well as the influxes, so as to result in the lowest possible SSR. We constrained the amino acid influxes to a lower bound of zero and an upper bound of the order of 12% of their rate of incorporation into biomass. We calculated the upper bound from measurements of (i) the growth rate (data not shown), (ii) the ratio of the initial biomass to newly synthesized biomass (1:9) as evaluated from growth rate measurements and (iii) the biomass composition (data not shown). However, the upper bounds for some influxes had to be increased, especially in case of the ACA influx, to achieve a low SSR. The evaluated fluxes and influxes, as well as their standard deviations (estimated as described in Materials and Methods) are listed in Supplementary Tables **S3 and S7**.

We compared the ability of the AA_in model to account for the measured isotopomer abundances with that of previously reported models (Lonien and Schwender, 2009) that employed other mechanisms to model the effect of the initial

biomass. One previously employed approach is adjustment of the purity (^{13}C enrichment) of the labeled carbon source. Our model "Gluc_dilu" (listed in Supplementary **Table S5**) uses this approach to model the effect of the initial biomass instead of the metabolite influxes described above. Another previous approach is the correction of the measured MIDs by a factor based on the ratio of the initial to final biomass. Our uniform isotopomer correction model "Iso_corr(U)" (listed in Supplementary **Table S4**) uses this approach. Instead of introducing additional metabolite influxes, it assumes that all proteinogenic amino acid isotopomer measurements are uniformly diluted by natural abundance in the ratio 1:9, which is the measured ratio of initial biomass to newly synthesized biomass. An improved version of this model, the variable isotopomer correction model "Iso_corr(V)", allows different dilution ratios or mixing coefficients for different amino acids. These dilution ratios are parameters in the flux model and are estimated by our flux evaluation program NMR2Flux+ (the evaluated ratios are in Supplementary **Table S6**). This new Iso_corr(V) model is clearly an improvement over the original isotopomer correction model, as it reflects the expectation that different amino acids may be diluted by naturally abundant intracellular pools to different extents upon introduction of the labeled carbon source.

To evaluate the performance of the models, we calculated an SSR based on a set of 405 isotopomer abundances of all the 30 amino acid fragments detected on GC-MS measured from three ILEs using 98% 1-^{13}C glucose, 28% U-^{13}C glucose and 98% U-^{13}C glucose. As evident from the parity plot **Fig. 3.7**, the AA_in model fits the

isotopomer measurements significantly better than the other three models. Quantitatively, the AA_in model yielded an SSR of 538, whereas the Gluc_dilu model gave an SSR of 4,626, the Iso_corr(U) model gave an SSR of 4,983, and the Iso_corr(V) gave an SSR of 1,748 (**Fig. 3.7**, inset). Remarkably, SSR corresponding to the 98% U-^{13}C glucose ILE was the lowest in case of the AA_in model, whereas the two alternative models provided substantially poorer fits and thereby larger SSRs. The statistically acceptable SSR value for 405 redundant measurements is 462. The AA_in model comes reasonably close to this, given that it modeled a batch culture at a pseudo-steady state.

Highlights of the flux map evaluated by using the AA_in model (fluxes and statistics shown as blue boxes in **Fig. 3.8**, Supplementary **Table S3**) include substantial net flux through the plastidic PPP (g6pdhp: 0.43 ± 0.23) and relatively lower flux through the cytosolic PPP (0.16 ± 0.07), as well as relatively low anaplerotic cycling (ana_net: 0.04 ± 0.01). To assess the efficacy of the metabolite reflux model, we also performed a second set of three parallel ILEs on 98% 1-^{13}C, 28% U-^{13}C and 98% U-^{13}C glucose on poplar cell suspensions over a 21-d period, by seeding the cells in fresh, appropriately ^{13}C-labeled medium every 7 d. These ILEs, although excessively long, can be expected to be relatively reflux-free because of (i) their length and (ii) the fact that intracellular metabolites will be less affected by reflux and unaffected by mixing (**Fig. 3.1**) as compared to biomass components such as proteinogenic amino acids. We isolated and measuring isotopomer abundances of intracellular metabolites from this experiment, and evaluated fluxes using a metabolic

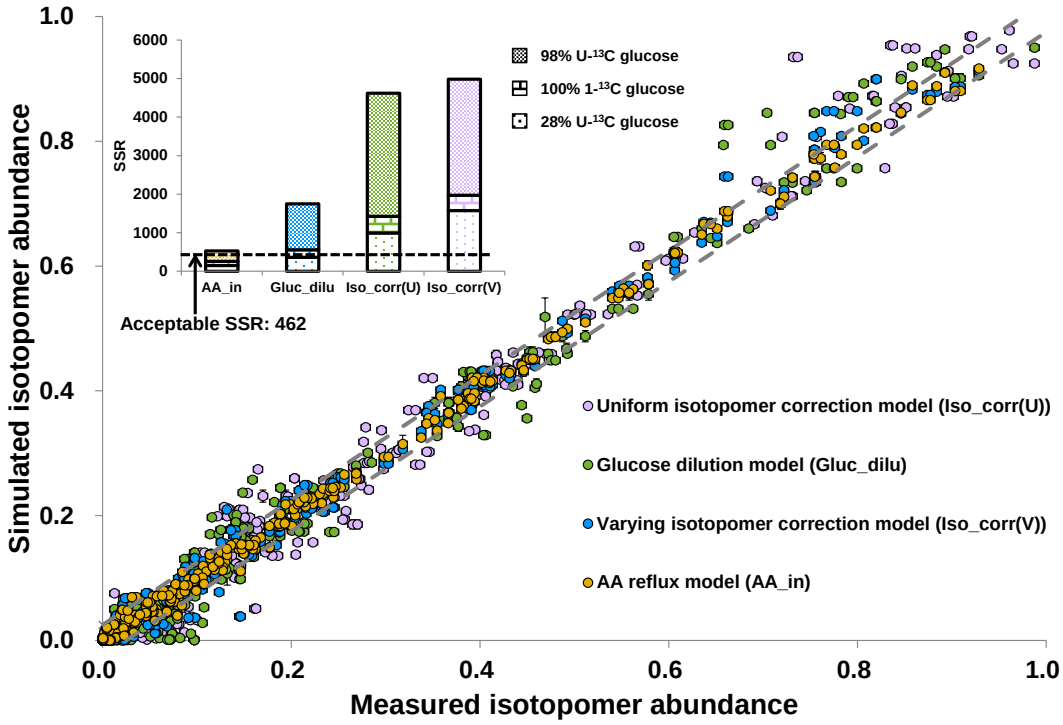


Figure 3.7. Metabolic model incorporating amino acid reflux outperforms models incorporating glucose dilution or isotopomer correction in explaining isotopomer abundances measured from three ILEs.
In the parity (scatter) plot, 405 isotopomer abundances measured from three ILEs (100% 1-^{13}C glucose, 28% U-^{13}C glucose and 98% U-^{13}C glucose) are plotted against isotopomer abundances simulated on the basis of the estimated fluxes. Each ILE was performed in triplicate on batch cultures of poplar cell suspensions. This plot illustrates the ability of four separate metabolic models respectively employing uniform isotopomer correction (Iso_corr[U], lavender points and bars), glucose dilution (Gluc_dilu, green points and bars), variable isotopomer correction (Iso_corr[V], blue points and bars) and amino acid reflux (AA_in, gold points and bars) to model the batch culture. The $y = x$ diagonal line corresponds to a perfect fit. The histogram in the inset depicts the sums of squared residues (SSRs) for the three models. All data points have error bars, although most error bars are too small to be visible.

model that lacked metabolite reflux (except the ACA influx), glucose dilution or isotopomer correction. The resulting flux evaluation yielded a statistically acceptable SSR of 382 for 357 isotopomer abundances, and the corresponding fluxes and their statistics are depicted in **Fig. 3.8** (red boxes). Clearly, the results of both

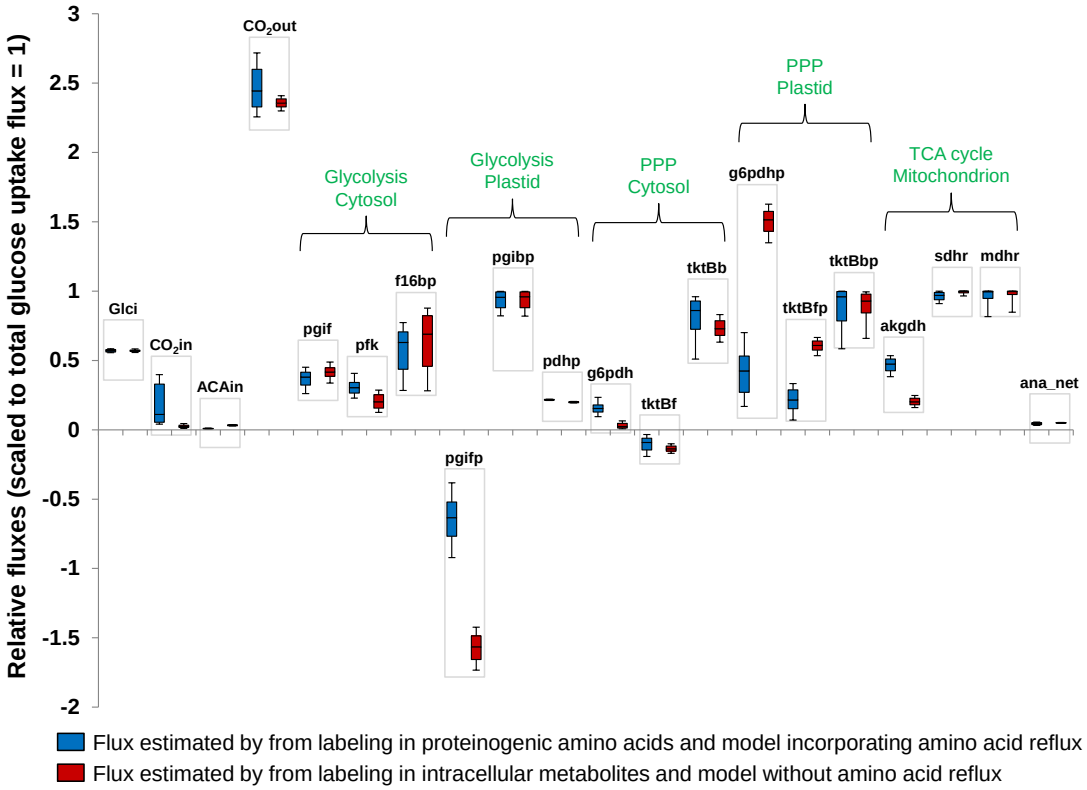


Figure 3.8. Comparison of metabolic fluxes estimated from two independent experiments and models.
Fluxes shown here are identified by name in the metabolic network of **Fig. 3.7** and Supplementary **Table S7**. Fluxes were estimated (i) from a one-week ILE by measuring labeling in proteinogenic amino acids and employing a model incorporating amino acid reflux (blue boxes) as well as (ii) from a three-week ILE by measuring labeling in intracellular metabolites and employing a model without amino acid reflux (red boxes).. For each methodology, flux distributions were obtained by performing 280 bootstrap Monte Carlo simulations that accounted for the larger of the measurement and biological errors in the measured isotopomer abundances. For each flux, the central bar in the box denotes the median of the flux distribution, the lower and upper ends of the box represent the 25th and 75th percentiles of the distribution, and the lower and upper whiskers represent the 10th and 90th percentiles of the distributions. Therefore, shorter boxes and whiskers indicate better flux identifiability. Gray boxes group flux distributions of the same reaction obtained from the two methods. The canonical pathways containing some of the fluxes are indicated above the data points. The flux ana_net is a net anaplerotic CO_2 fixation flux that consolidates the forward and reverse directions of various anaplerotic fluxes in the metabolic model.

methodologies are qualitatively the same, although there are numerical differences between them for some flux values, especially the CO_2 influx (CO_2in, reflux model: 0.18 ± 0.16, model without reflux: 0.03 ± 0.01) and the net flux through the plastidic PPP (g6pdhp, reflux model: 0.43 ± 0.23, model without reflux: 1.50 ± 0.11). Despite this, the estimates of many fluxes (CO_2out, pgif, pgibp, pdhp, tktBb, sdhr, mdhr, ana_net) are similar.

3.5. Discussion

Isotope MFA is a powerful systems biology tool for studying cellular metabolism. It relies on the measurement of labeling patterns of metabolites (isotopomers) generated by the metabolic processing of specific ^{13}C-labeled carbon sources, usually at isotopic steady state. In conjunction with information on carbon atom rearrangements in metabolic reactions, this methodology quantitatively tracks the trajectories of the ^{12}C and ^{13}C atoms through the metabolic network to simulate isotopomer abundances. By fitting these simulated isotopomer abundances to the experimentally measured ones, it evaluates flux values and distributions that best explain the labeling data. Consequently, the flux distributions estimated by isotope MFA are very sensitive to the isotopomer abundances of metabolites. In batch ILEs, an important biological phenomenon that could influence the isotopomer abundance measurements is the presence of the initial biomass and interactions between this and the newly synthesized biomass **(Fig. 3.1)**. Because batch ILEs are convenient or inevitable for many biological systems, it is necessary to quantify the effects of the initial biomass on the isotopomer abundance measurements.

In this work, we show that the use of ≈100% U-^{13}C carbon sources can provide much information on the interactions between the initial and newly synthesized biomass. An ILE using a single U-^{13}C carbon source will typically yield no flux information, as all isotopomers will be ^{13}C-labeled. However, reflux of the naturally abundant initial biomass into metabolism is amplified in such an ILE. The anomalous ^{13}C enrichments **(Fig. 3.2)** and mass isotopomer abundances (**Fig. 3.3**) from our 98% U-^{13}C glucose ILEs suggested that the initial biomass did not remain isolated from the newly synthesized biomass and that it refluxed into the metabolic network. If the initial biomass had remained isolated and had only mixed with the newly synthesized biomass, we would only have observed the least massive (M+0) and most massive (M+n) isotopomers of each fragment. However, isotopomer measurements from the 98% U-^{13}C ILE clearly disagreed with this expectation. We eliminated two competing causes for these isotope labeling trends – photosynthetic and anaplerotic $^{12}CO_2$ incorporation – by performing follow-up ILEs (**Fig. 3.4**) or examining mass isotopomers of aspartate and glutamate (**Fig. 3.5**). Although these results were not able to exclude the "mixing" phenomenon, they did point at reflux of the initial biomass as the major cause of the anomalous isotope labeling patterns.

Previous reports (M. Antoniewicz et al., 2007; Dauner et al., 2001; Sriram et al., 2007a, 2004) of isotope MFA in batch cultures have used two approaches to account for the initial biomass: (i) adjustment of the ^{13}C enrichment of the labeled carbon source, or (ii) correcting the isotopomer abundance measurements assuming

that they are produced by mixing of the initial and newly synthesized biomass. Only one report(Kruger et al., 2011; Masakapalli et al., 2013) used amino acid influxes to model the effect of initial biomass. However, there is no previous report quantifying the extents to which these models account for isotopomer measurements. This is a pressing need, given our 98% U-^{13}C glucose ILEs that are sensitive to metabolite reflux. Our comparison of all these models clearly showed that the model incorporating amino acid and metabolite reflux (AA_in; **Fig. 3.6**) is more accurate than the other three models. It especially outperforms them in explaining the data from the 98% U-^{13}C glucose ILE (**Fig. 3.7**). Two reasons may explain the superior performance of the metabolite reflux model. Foremost, the glucose dilution and isotopomer correction models only use a single dilution factor to account for anomalous isotopomers. In contrast, the reflux model uses several amino acid and metabolite influxes as flux parameters. In other words, we are introducing additional degrees of freedom at reasonable places of our metabolic network. These additional degrees of freedom can explain the anomalous isotopomers observed by us. For example, the reflux model estimated the acetyl-CoA reflux (~0.02 units) to be much higher than histidine reflux (~0.0001 units), indicating that the reflux of acetyl-CoA is much more responsible for the anomalous mass isotopomers than the reflux of His. Nevertheless, increased degrees of freedom are not the sole explanation for the anomalous isotopomer measurements. Specifically, the improved isotopomer correction model Iso_corr(V) does have a large number of degrees of freedom, just as the reflux model. Despite this, the Iso_corr(V) model failed to acceptably fit the labeling data, indicating that the dilution or mixing effect is insufficient to account for

the anomalous isotopomers. The fact that the AA_in model is able to acceptably fit the labeling data, including most of the anomalous isotopomers, emphasizes the evidence for reflux. Moreover, our observation of different ^{13}C enrichments for amino acid fragments sharing a common metabolic precursor (e.g. 27.1%, 27.4%, 27.8% for glutamate[2345] fragments and 22.7%, 23.0% for proline[2345] fragments despite their both being derived from the same metabolic precursor AKG) can only be explained by a comprehensive reflux model.

Being tree cells, poplar cell suspensions are especially appropriate for this study because trees are known to be capable of conserving and reusing nutrients over small and large timescales. Previous studies show that poplar efficiently manages its nitrogen reserves by recycling nitrogen from leaves during senescence in fall, storing this recycled nitrogen in the form of a protein (bark storage protein [BSP]) during winter and remobilizing this stored nitrogen for *de novo* protein synthesis in spring (Cantón et al., 2005; Zhu and Coleman, 2001). Exploring how poplar cells' metabolic landscapes are altered by this nitrogen cycling and storage process is a fascinating research question and metabolic flux analysis holds promise in shedding light into this question. Therefore, our study of biomass reflux model in poplar suspension cells serves as a forerunner of studying the larger problem of nitrogen recycling mechanism in poplar.

The flux distribution obtained by using isotopomer measurements of proteinogenic amino acids from a 7-d ILE and the reflux model agreed qualitatively

with the flux distribution obtained using intracellular metabolites by an independent methodology – 21-d ILEs and a model with minimal reflux (**Fig. 3.8**). Although there were numerical differences between flux values estimated through the two methodologies, many flux values (e.g. ACA reflux, net anaplerotic flux) and all flux trends (e.g. large cyclic PPP in the plastid) were identical for both methodologies. This consistency justifies the application of reflux methodology in analyzing anomalous labeling patterns of proteinogenic amino acids.

3.6. Conclusions

The development and isotope MFA in plants and mammalian cell or tissue cultures, combined with the inevitability of batch culture, necessitates accurate modeling to account for the initial biomass. In this regard, our work addresses two important gaps in knowledge: the use of an ILE supplying a ≈100% U-^{13}C carbon source to accurately quantify the extent of reflux, and the demonstration that a comprehensive reflux model accounts for the isotopomer abundances from such an ILE much more satisfactorily than previously employed models. Furthermore, flux results from our methodology using the reflux model were similar to those from another methodology involving minimal reflux. We anticipate our methodology to be especially useful for isotope MFA of batch cultures of bacterial, plant or mammalian cell or tissue where it is not practical to remove the initial biomass.

Future extensions to our work include targeted ILEs to improve estimates of reflux, especially of the biomass components that reflux to the greatest extent.

Additionally, investigation and explicit incorporation of more nontrivial amino acid degradation pathways similar to the valine ≫ succinate and glutamate ≫ ACA pathways in our model is also necessary.

Chapter 4: Concurrent isotope-assisted metabolic flux analysis and transcriptome profiling reveal responses of poplar cells to altered nitrogen and carbon supply

Xiaofeng Zhang[1,&], Shilpa Nargund[1,2,&], Ashish Misra[1,3,&], Gary D. Coleman[4] and Ganesh Sriram[1,5]

XZ, SN, AM, GDC and GS conceived this study; XZ, SN, AM and GS designed it. GDC provided poplar cell suspension cultures; XZ, SN and AM jointly performed experiments. XZ performed computation and wrote the draft of the manuscript. GS critically edited the manuscript. XZ and GS revised the manuscript after peer review. SN, AM, XZ contributed equally.

[1]Department of Chemical and Biomolecular Engineering, University of Maryland
Current affiliations: [2]Syngene International Limited, Bengaluru, India; [3]DBT-ICT Centre for Energy Biosciences, Institute of Chemical Technology, Mumbai, India
[4]Department of Plant Science & Landscape Architecture, University of Maryland
[5]Principal investigator and corresponding author

Abstract

Reduced nitrogen is indispensable to life. However, its sparing availability in soil, combined with the energetic or environmental drawbacks of chemically synthesized fertilizers, this motivates research into molecular mechanisms to improve plant nitrogen use efficiency (NUE). A systems-level investigation of this problem may benefit from the use of multiple 'omics methodologies. We report such a study on cell suspensions of *Populus trichocarpa*, a woody perennial tree capable of efficiently managing its nitrogen reserves. Acclimation of poplar cell suspensions to four environmental conditions ranging from nitrogen and carbon abundant supply to deficient supply of both elements revealed that the low-nitrogen conditions were associated with substantially higher NUE. To investigate the underlying intracellular mechanisms, we concurrently performed (i) steady-state ^{13}C metabolic flux analysis using multiple isotope labels and (ii) transcriptomic profiling using cDNA microarrays. ^{13}C flux analysis revealed the absolute flux through the oxidative pentose phosphate pathway (oxPPP) to be substantially (~3-fold) lower under the low-nitrogen conditions and the relative flux partitioning between the tricarboxylic acid cycle and anaplerotic pathways to vary considerably, from 84%-16% (abundant carbon and nitrogen) to 55%-45% (deficient carbon and nitrogen). Together, the flux and gene expression data suggested a plastidic localization of the oxPPP as well as transcriptional regulation of some major flux alterations including that in the oxPPP. The gene expression data also pointed to NUE-improving mechanisms such as redirection of additional carbon to aromatic metabolic pathways and extensive downregulation of many genes encoding photosynthetic and light-harvesting proteins

in these non-photosynthetic cells, suggesting the recruitment of these proteins as nitrogen sinks in nitrogen-abundant conditions.

4.1. Introduction

Nitrogen plays an indispensable role in life as it is contained in the building blocks of proteins and nucleic acids, and is a component of many secondary metabolites that defend against stress (Heldt and Piechulla, 2010). Plants facing nitrogen deficiency can present problems ranging from poor yield, poor nutritional quality or taste, difficulty in downstream processing or storage to increased pathogen susceptibility (Amtmann and Armengaud, 2009). Despite the importance of nitrogen as a nutrient, it is currently extremely difficult to predict the relationship between nitrogen deficiency and the concomitant poor performance of plants (Amtmann and Armengaud, 2009) because of the complexity of the metabolic networks underlying nitrogen metabolism and the inextricable links between nitrogen and carbon metabolism. The interaction between nitrogen and carbon metabolism are reflected in various aspects. For example, nitrogen metabolism requires carbon skeletons, energy and reductants, which are generated by carbon metabolism (Gao et al., 2008; Kruger and von Schaewen, 2003). Carbon metabolism also supplies organic acids that facilitate nitrate uptake by acting as counter-ions (Sahrawy et al., 2004). Conversely, nitrogen availability also affects carbon metabolism such as photosynthesis (CO_2 fixation), photorespiration (undesirable O_2 fixation) and respiration which are important cellular functions in plants (Nunes-Nesi et al., 2010). The interactions between C and N themselves respond to a multitude of factors: light and water availability, adenosine triphosphate (ATP), reducing factors (required for assimilation

of inorganic N), and pH balance and consequently they are regulated by various signaling molecules such as nitrate, ammonium, sugars, amino acids (especially glutamine, glutamate and aspartate), and organic acids (Nunes-Nesi et al., 2010). Because nitrogen and carbon metabolism can compete for resources such as ATP (Sahrawy et al., 2004) and reductant (Escobar et al., 2006), their control is likely to be tightly interlinked at the level of gene expression as illustrated by several lines of evidence (Escobar et al., 2006).

Uptake and assimilation of N needs energy in the form of ATP and reducing power (Nunes-Nesi et al., 2010). Studying the existing metabolic machinery that helps recycle N and genetically eliminating possible bottle necks in this process will help trees utilize N more efficiently and make them better biomaterial crops. Poplar, a potential cellulosic biofuel source, is capable of recycling up to 80% of the N in leaves to perennial tissues during leaf senescence through the bark storage proteins (BSPs) (Cantón et al., 2005; Pregitzer et al., 1990). Hence there is tremendous interest in understanding N cycling and C-N interaction in this tree. Probing the intracellular fluxes in poplar cell suspensions (hereafter, “poplar cells” or “cells”) under various N conditions will help understand the N cycling and C-N interactions.

A crucial first step in elucidating nitrogen metabolism is a quantitative understanding of nitrogen fluxes and carbon-nitrogen interactions in plant primary metabolism (Amtmann and Armengaud, 2009), and the relationship between these fluxes and plant growth/productivity. That this problem remains uninvestigated is not

surprising, as the mathematical and experimental tools necessary for this investigation have become available very recently. One example is isotope-assisted or ^{13}C metabolic flux analysis (^{13}C MFA), a vital tool to study metabolism due to its ability to quantify intracellular fluxes by analyzing the labeling patterns of cellular products in isotopic labeling experiments (ILEs). Recently developments in isotope-assisted MFA such as instationary MFA (Nöh and Wiechert, 2006; Wahl et al., 2008) and improved mathematical techniques (M. R. Antoniewicz et al., 2007; Young et al., 2008) have expanded the application of MFA and made it computationally more feasible for analyzing complex metabolic networks.

Previous work on understanding plant metabolism under different nitrogen sources or provisions are mainly focused on metabolomics, proteomics or gene expression data (Allen et al., 2011; Hockin et al., 2011; Noorhana, 2011; Truong et al., 2013; Zhila et al., 2005). Recently, researchers have started applying ^{13}C isotope-assisted metabolic flux analysis tool to study the influence of nitrogen source on plant central carbon metabolism at a molecular level (Allen and Young, 2013; Masakapalli et al., 2013). For example, Masakapalli *et al*. (2013) found that by replacing ammonium with nitrate in the culturing media, the oxidative pentose phosphate pathway flux increased by 50% in Arabidopsis cells. Allen and Young's study (2013) showed that reducing the ratio of carbon and nitrogen from 37:1 to 13:1 leads to a 90% increase in downstream TCA cycle flux, which then increase the contribution of malic enzyme to pyruvate production. Here we probed the influence of different carbon and nitrogen supply on central carbon metabolism in poplar suspension cells

via steady state ^{13}C-MFA and transcriptomics data. We rationally designed a two-dimensional nutrient concentration gradients of carbon and nitrogen: (i) high carbon-high nitrogen (CN), (ii) high carbon-low nitrogen (Cn), (iii) low carbon-high nitrogen (cN) and (iv) low carbon-low nitrogen (cn). Micro-scale growth characteristics provide a first hint of how cells response differently to carbon and nitrogen starvations. Then we used ^{13}C isotopic labeling data and metabolic flux analysis, together with microarray data, to confirm these hypotheses. The altered metabolic flux distributions and gene expressions revealed the cells' capability to adapt low N environment by improving nitrogen use efficiency.

4.2. Materials and Methods

4.2.1. Poplar cell suspension culture and growth rate measurement

Throughout this work we used Poplar (*Populus. tremula* × *Populus. alba* clone 717-1B4) cell suspensions maintained in 125 mL Erlenmeyer flasks shaken at 125 rpm. The flasks were placed at 20 °C under continuous light from cool-white fluorescent light (Ecolux® Technology Plant and Aquarium F40T12 bulbs, photosynthetically active radiation of 200-300 μmol m^{-2} s^{-1}, incident radiation on flasks 28.3 ± 2.6 μmol m^{-2} s^{-1}). We subculture the cell suspensions every 7 days by transferring 0.6 g fresh cells from the previous growth cycle into 30 mL Murashige and Skoog medium (Phytotechnology Laboratories, Shawnee Mission, KS) media with glucose as the principal carbon source. As quantified in our previous publication (Nargund et al., 2014), the cells also consumed additional, minor carbon sources including anaplerotically fixed CO_2 (the cells did not photosynthetically fix

extracellular CO_2) and initially present unlabeled biomass that chiefly manifested as an influx of acetyl-CoA into the metabolic network. We realized the four environmental conditions by growing poplar cell suspensions under two initial concentrations of glucose (high: 20 g L^{-1}, low: 10 g L^{-1}) combined with two initial concentrations of their only nitrogen sources, NH_4NO_3 and KNO_3 (high: 1,650 mg L^{-1} NH_4NO_3 and 1,900 mg L^{-1} KNO_3, low: 264 mg L^{-1} NH_4NO_3 and 304 mg L^{-1} KNO_3). This gave four conditions in all of which the ratio of NH_4NO_3 and KNO_3 concentrations was constant: (i) high carbon-high nitrogen (CN), (ii) high carbon-low nitrogen (Cn), (iii) low carbon-high nitrogen (cN) and (iv) low carbon-low nitrogen (cn). We chose the aforementioned concentrations iteratively such that the cells were capable of both (i) growing continuously for 6 d to 7 d under any of the four conditions and (ii) exhibiting tangibly different cell growth characteristics from the CN condition. We acclimated cells on these four conditions for several months before performing ^{13}C MFA and microarray analysis.

For growth rate and extracellular measurements, we grew 12 flasks of cells in parallel under each environmental condition. On 2 d, 4 d, 6 d and 7 d, three biological replicate flasks for each condition were harvested by vacuum filtration through glass microfiber filter paper (Whatman, Piscataway, NJ). Fresh weights of the cell pellets were obtained by weighing them immediately after the filtration; dry weights are obtained by weighing the cell pellets after lyophilization for extended time (over 1 d). We measured residual glucose in the media by a glucose analyzer (YSI Incorporated, Yellow Springs, OH), and measured residual ammonium and nitrate levels by a

nitrite/nitrate assay kit (Sigma-Aldrich, St. Louis, MO). The residual nutrient concentrations were also were performed on three biological replicates for each condition.

4.2.2. Concurrent ^{13}C MFA and transcriptome profiling

We grew 12 flasks of poplar cells under each of the four environmental conditions. Of these, nine flasks were used to perform three parallel ILEs in three biological replicates. In these flasks, the naturally abundant glucose in the medium was replaced with: (i) 98% U-^{13}C glucose, (ii) 98% 1-^{13}C glucose or (iii) a mixture of 30% U-^{13}C and 70% naturally abundant glucose (Cambridge Isotope Laboratories, Andover, MA). The three remaining shake flasks per environmental condition were used for microarray analysis.

All flasks were grown for three weeks (with inoculation into fresh, appropriately ^{13}C-labeled medium every 7 d) in accordance of the three-week labeling strategy, which we demonstrated previously on the CN condition (Nargund et al., 2014). Specifically, after 7 d of growth, 0.6 g fresh cells were inoculated into fresh media that contained the same type of labeled glucose. After a cumulative growth period of three weeks (three inoculations), the cells were harvested, vacuum-filtered and immediately quenched in liquid nitrogen. The frozen cell pellets were stored at –80 °C until further analysis.

4.2.3. Measurement of isotopomer abundances of intracellular metabolites by GC-MS

Detailed protocols for the extraction of soluble metabolites, quantification of isotopomer abundances by GC-MS and correction of the raw mass isotopomer abundance data for the natural abundances of hydrogen, nitrogen, oxygen, sulfur, silicon and non-metabolic carbon atoms described in our previous article (Nargund et al., 2014), and were exactly followed in this study.

4.2.4. Flux evaluation from isotopomer data

The flux evaluation procedure follows the protocols discussed in **Sec. 3.2.4.** Moreover, we carried out a bootstrap Monte-Carlo algorithm to calculate the standard deviations of the estimated fluxes. Briefly, the program perturbed the measured isotopomers within their standard deviations for 400 times. Then it evaluates the fluxes for each perturbed set of isotopomers separately using the optimization procedure described above. The standard deviations of the fluxes were then calculated based on these 400 evaluated flux sets.

4.2.5. RNA extraction and microarray data processing

To extract RNA, the harvested cells were chilled by liquid nitrogen and ground to fine powder. Each cell sample was transferred to 2 mL tubes added with 1~1.5 mL RLT buffer (QIAGEN, Venlo, Netherlands) that contains 0.01 g soluble polyvinylpyrrolidone (Sigma-Aldrich, St. Louis, MO) and 10 μg/mL of 2-

Mercaptoethanol (Sigma-Aldrich, St. Louis, MO). Then it was added with 0.4 mL 5 M potassium acetate (Sigma-Aldrich, St. Louis, MO) at pH 6.5. After gently vortexing and centrifuging the tubes at 12, 000 rpm, we transferred the supernatant to two new 2 mL clear tubes and placed them in a fully automated RNA extraction instrument, QIACube (QIAGEN, Venlo, Netherlands) loaded with extraction buffer and DNase (QIAGEN, Venlo, Netherlands). After the automated extraction, we collected the extracted RNA solution and checked the RNA quality by measuring the absorbance ratio at 260 nm over 280 nm using a NanoDrop spectrophotometer (Thermo Scientific, Wilmington, DE).

The 12 extracted RNA solution samples (3 biological replicates for each C, N condition), together with GeneChip® poplar genome arrays (Affymetrix) were processed by Biopolymer/Genomics Core Laboratory Services (University of Maryland, Baltimore) for microarray processing. The microarray raw data that contained 61, 413 probesets was processed by software ArrayStar 5.0 (DNASTAR, Inc., Madison, WI) using RMA normalization method. Besides, we performed an unsupervised clustering analysis, principle components analysis (PCA) to make an initial comparison of the microarray raw data among the four conditions. The correspondence between microarray probesets and gene IDs were obtained from Affymetrix poplar annotations, release 33. The functions of the genes were obtained from three major online sources: Poplarcyc 5.0, Uniprot and PLEXdb.

4.3. Results

4.3.1. Low nitrogen availability improves nitrogen utilization efficiency (NUE)

We characterized various macroscopically measurable metabolic properties of the cells under the four conditions, including cell growth, carbon and nitrogen source uptake rates, biomass composition as well as distribution of amino acids in protein. Cell growth measurements **(Fig. 4.1a)** showed that growth was reduced substantially under both low-C and low-N conditions, as consistent with previous studies (Noorhana, 2011; Truong et al., 2013). Besides, the high-N conditions exhibited faster initial growth than low-N conditions. Analysis of nutrient uptake rates showed that the yield of biomass on carbon (**Fig. 4.1c**) remained relatively unchanged across the four conditions, ranging from 1.3 mg biomass (mg carbon)$^{-1}$ to 1.5 mg biomass (mg carbon)$^{-1}$. By contrast, the yield of biomass on nitrogen (**Fig. 4.1d**) was significantly higher under the low-N conditions (21 mg biomass [mg nitrogen]$^{-1}$ for the Cn condition, 23 mg biomass [mg nitrogen]$^{-1}$ for the cn condition) than under the high-N conditions (11 mg biomass [mg nitrogen]$^{-1}$ for the cN condition and 15 mg biomass [mg nitrogen]$^{-1}$ for the cn condition). This evidences higher NUE under nitrogen-limited conditions irrespective of carbon limitation. Furthermore, a comparison of carbon and nitrogen consumption (**Fig. 4.b**) showed that under the low-N conditions, the cells were able to utilize less nitrogen while consuming the same amount of carbon source, suggesting an altered carbon-nitrogen balance. Besides, distributions of amino acids in protein were sensitive to nitrogen availability **(Fig. 4.1e)**. For example, Leu constituted more than 15% of protein under the high-N conditions, but constituted less than 4% of protein under the low-N conditions.

Similarly, the combined proportion of Glu and Gln (Glx) was 20% under the high-N conditions but fell to less than 10% under the low-N conditions.

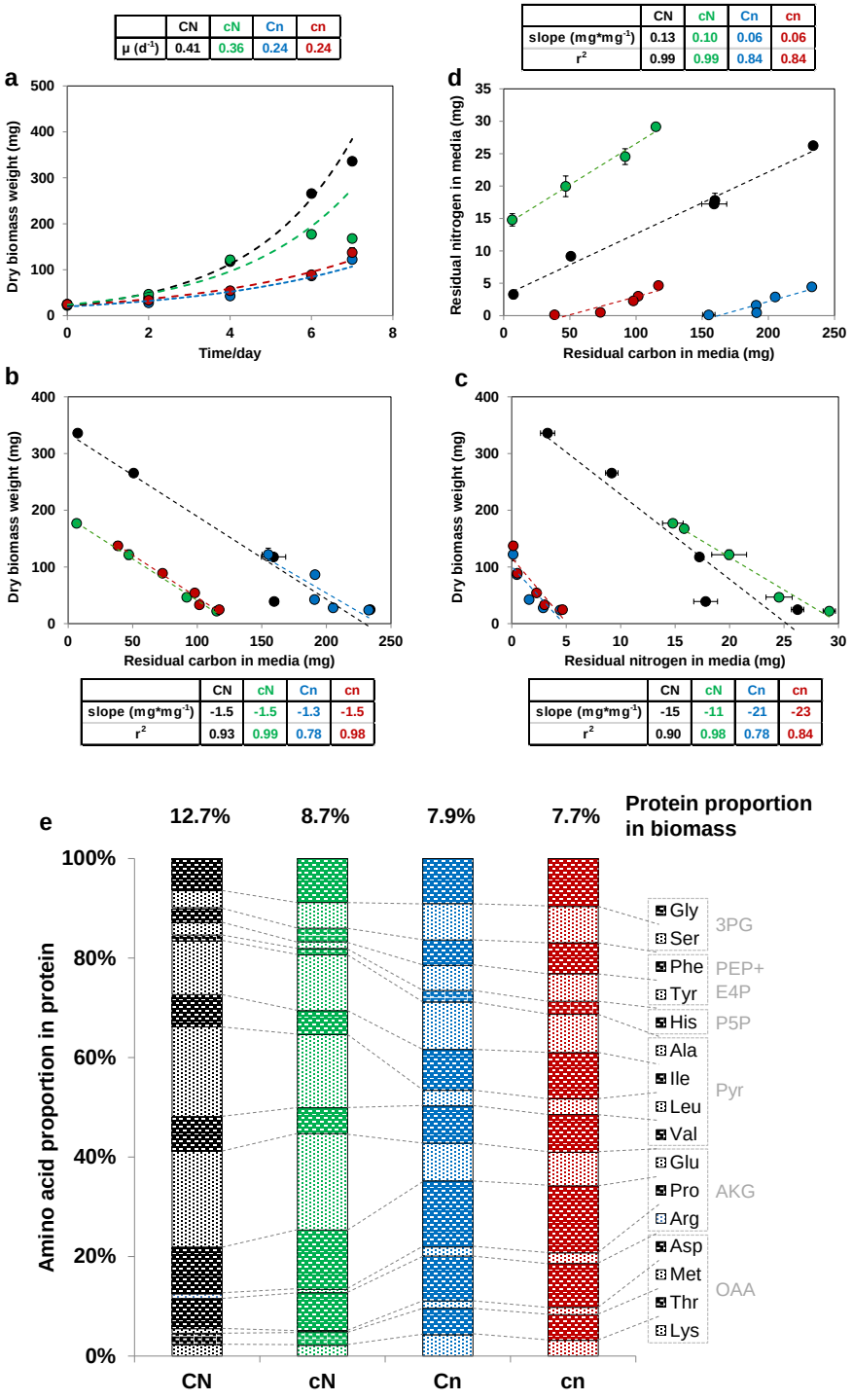


Figure 4.1. Effect of carbon and nitrogen limitation on growth rates, NUE and proteinogenic amino acid proportions in poplar cell suspensions.
The four environmental conditions of this study are distinguished by four colors: black represents CN; green represents cN, blue represents Cn and red represents cn. **(a)** Cell growth represented by dry biomass weight (mg) versus time (d); **(b)** yield of biomass on carbon represented by correlation between dry biomass weight (mg) and residual carbon in media (mg); **(c)** yield of biomass on nitrogen represented by correlation between dry biomass weight (mg) and residual nitrogen in media (mg). The slopes of the linear trendlines in **(b)** and **(c)**, which equal the biomass yield on carbon or nitrogen, respectively, are tabulated adjacent to the plots. **(d)** Correlation between residual nitrogen in media (mg) and residual carbon in media (mg). **(e)** Protein proportion in biomass and amino acid proportions in protein for the four conditions. The patterns in the stacked bar plot represent different amino acids, which are organized based on their metabolic precursors as indicated in the panel to the right.

4.3.2. Differences in mass isotopomer abundances suggest flux variations in oxidative pentose phosphate pathway, TCA cycle and acetyl-CoA reflux

To resolve intracellular fluxes in cells under the four environmental conditions, we performed three-week ILEs as described in Experimental Procedures. Briefly, cells growing under each condition were fed (i) 98% U-^{13}C glucose, (ii) 98% 1-^{13}C glucose or (iii) a mixture of 30% U-^{13}C and 70% naturally abundant glucose in separate, parallel ILEs. The resulting isotopomer dataset consisted of 400 amino acid mass isotopomers for each condition **(Table S8-S10)**. A comparison of the isotopomer data across the four conditions showed that the largest number of significantly different isotopomer abundances lay between the CN and the cn condition (85 isotopomer abundances with difference > 0.03, $p < 0.05$ and 29 isotopomer abundances with difference > 0.06, $p < 0.05$) **(Fig. 4.2a)**. In mass isotopomer space, a difference of 0.06 (on a scale from 0 to 1) is substantial, considering that many recent ^{13}C MFA studies reported mass isotopomer differences less than 0.06 even between genetically variant strains that show significant flux differences (e.g. Kind *et al.*, 2013;

Wasylenko and Stephanopoulos, 2014; He *et al.*, 2014; Lonien and Schwender, 2009).

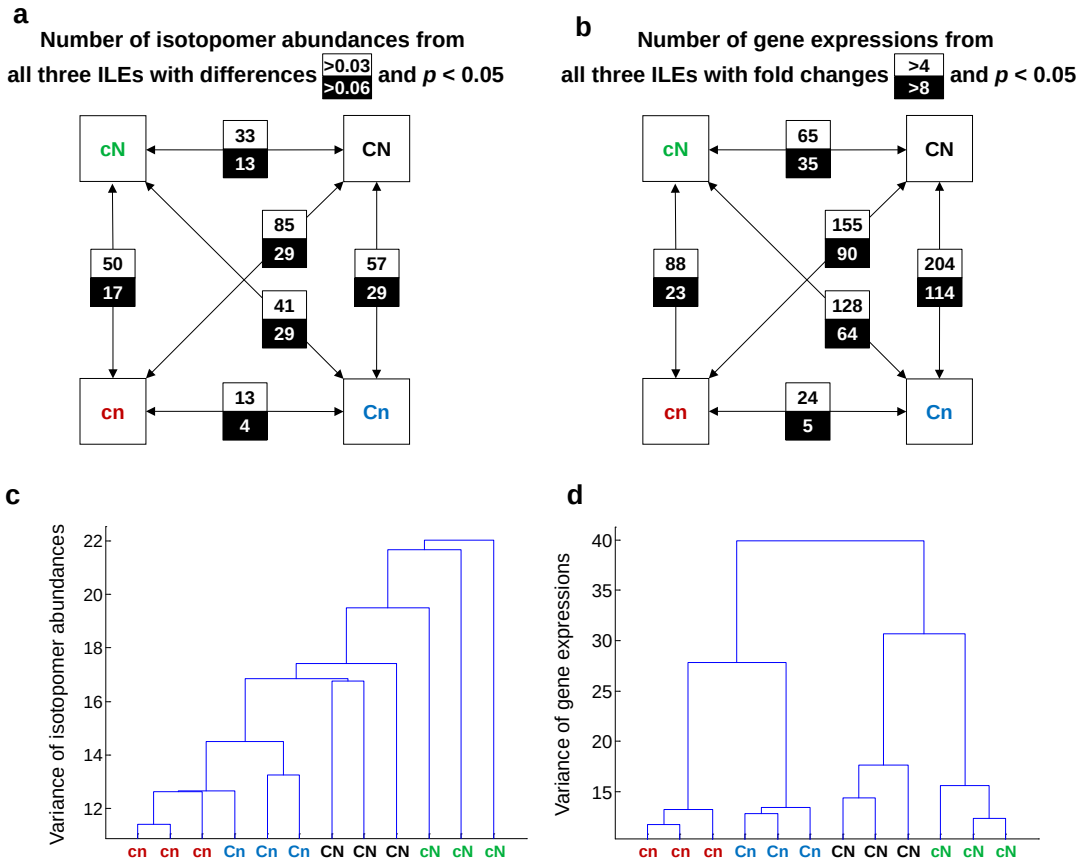


Figure 4.2. Unsupervised comparison of isotopomer data and clustering analysis of the microarray data across the four conditions.
We counted the number of isotopomers abundances (a) and transcriptomics levels (b) that are different between each pair among the four carbon-nitrogen conditions. The threshold is: difference > 0.075 with p value < 0.05 for isotopomer abundances and fold change > 8 with p value < 0.05 for transcriptomics levels. The two high nitrogen conditions are listed on the top and the two high carbon conditions are listed on the right. Besides, we performed unsupervised clustering analysis for both the isotopomer dataset (c) and the gene expressions dataset (d). The four conditions are listed along horizontal axis with three biological replicates each.

For example, Lonien and Schwender *et al.* (2009) observed four-fold flux alterations in Arabidopsis stains resulting from a transcription factor mutation, while the maximum difference of mass isotopomer abundances was ≈ 0.06. Thus the

differences in mass isotopomer abundances across conditions observed in this study suggest non-trivial flux differences.

The MIDs of certain amino acid fragments revealed straightforward information on flux alterations. For example, in the 1-^{13}C ILE, 53% ± 1% of the His[2345] fragment (i.e., the His fragment consisting of carbons C2-C5, same notation hereafter) and the His[23456] fragment were unlabeled under the Cn condition whereas under the CN condition, 61% ± 1% of the His[2345] and His[23456] fragments were unlabeled **(Fig. 4.3a)**. His originates from pentose-5-phosphate, which in turn is produced from glucose either via the oxidative or non-oxidative pentose phosphate pathway (PPP). The entire His molecule synthesized via the oxidative PPP does not contain the first carbon atom of glucose and would therefore be completely unlabeled in the 1-^{13}C glucose ILE. Under low-N conditions, since close to 50% of the His was unlabeled, it suggested that the oxidative and non-oxidative PPP contributed equally to its production. Whereas under high-N conditions, with 60% of the His molecules being unlabeled, the oxidative PPP might be the dominant route for His production. However, the isotopomer abundances of His are influenced not only by the flux through the oxidative PPP, but also by reaction reversibilities and cyclic flow in this and connected pathways (e.g. glycolysis). Therefore, quantification of flux differences between the four conditions requires full isotopomer modeling and statistical analysis.

Similar analysis applies to MIDs of Thr and Glu MS fragments from the 98% U-^{13}C glucose ILE. In the 98% U-^{13}C glucose ILE, we observed lower M+n (i.e. mass

isotopomer abundance with the highest mass and n refers to the number of carbon atoms in the amino acid fragment) and hence lower net ^{13}C enrichments in Thr fragments and Glu fragments under low-N conditions **(Fig. 4.3b & 4.3c)**. Both amino acids are derived from TCA cycle intermediates, with Thr originating from oxaloacetate and Glu from α-ketoglutarate. Consequently, the observation of lower net ^{13}C enrichments pointed to greater influxes of two unlabeled carbon sources in 98% U-13C glucose ILE under low nitrogen conditions: CO_2 uptake through anaplerotic reactions or unlabeled acetyl-CoA reflux through lipid degradation.

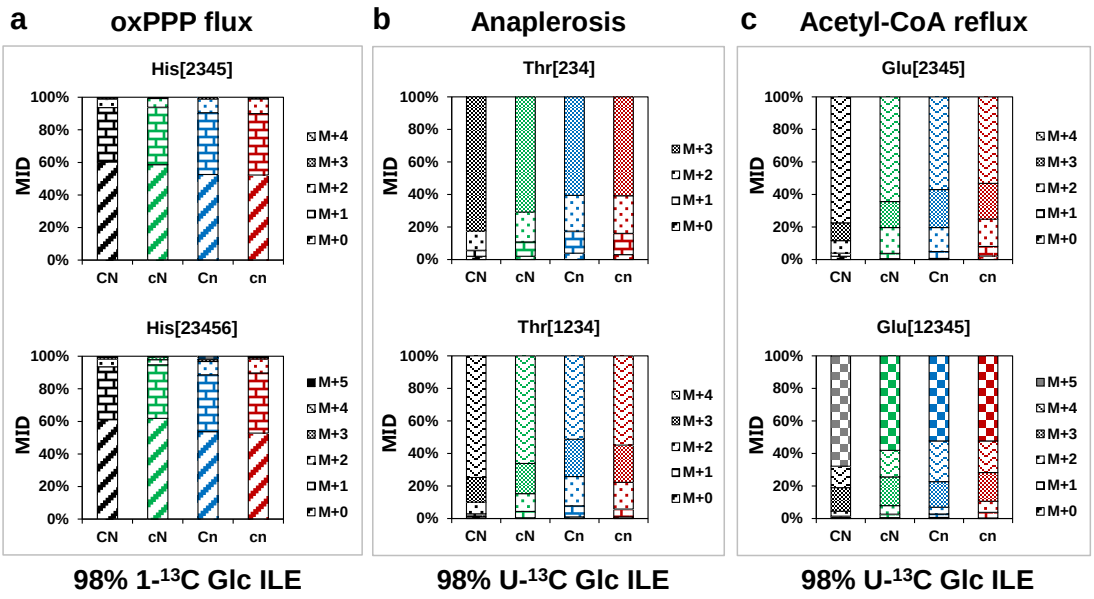


Figure 4.3. Mass isotopomer distributions (MIDs) under conditions of different carbon and nitrogen availability.
MIDs of fragments of three amino acids **(a)** His (from a 98% 1-^{13}C glucose ILE), **(b)** Thr (from a 98% U-^{13}C glucose ILE) and **(c)** Glu (from a 98% U-^{13}C glucose ILE) under the four conditions CN, cN, Cn and cn. The conditions are distinguished by four colors: black represents CN; green represents cN, blue represents Cn and red represents cn. Different MIDs from M+0 to M+n are distinguished by patterned filled bars, from bottom to top.

4.3.3. Lower oxPPP flux, higher anaplerotic flux and higher acetyl-CoA reflux were observed due to lack of nitrogen source

Our metabolic network model, similar to the one used in our previous study (Nargund, et al., 2014), consisted of glycolysis, PPP, TCA cycle, anaplerotic reactions and reflux of acetyl-CoA. The inclusion of acetyl-CoA reflux was supported by measured Glu isotopomer abundances (refer to **Sec. 4.3.2**). Glucose and anaplerotically fixed CO_2 are two carbon sources. Three intracellular compartments were included: cytosol, mitochondrion and plastid, with glycolysis and pentose phosphate pathways duplicated in both cytosol and plastid. The complete metabolic network model used for MFA is displayed in **Fig. 4.4** and **Table S11**. We applied a compartmentalized model because: (i) the different measured MIDs between Ser[123] and glycerol[123] under the CN condition (**Fig. 4.5**); (ii) the much lower sum of squared residuals (SSR) achieved using a compartmentalized model (216) compared with using a single-compartment model (1137) under the CN condition **(Fig. 4.5)**.

For each of the four environmental conditions, we evaluated metabolic fluxes from the entire dataset of around 400 isotopomer abundances from the three ILEs for that condition. The sum of residuals (SSR) between measured and calculated mass isotopomer abundances fall within the acceptable ranges for all the four conditions (achieved SSR: CN – 216, cN – 285, Cn – 331, cn – 395; acceptable SSR: CN – 389, cN – 375, Cn – 373, cn – 396).The flux estimations **(Table S12)** showed several differences in the fluxes through the glycolytic and PPP between the cytosol and plastid. Under all four conditions, the glycolytic flux (pgif) was greater than the

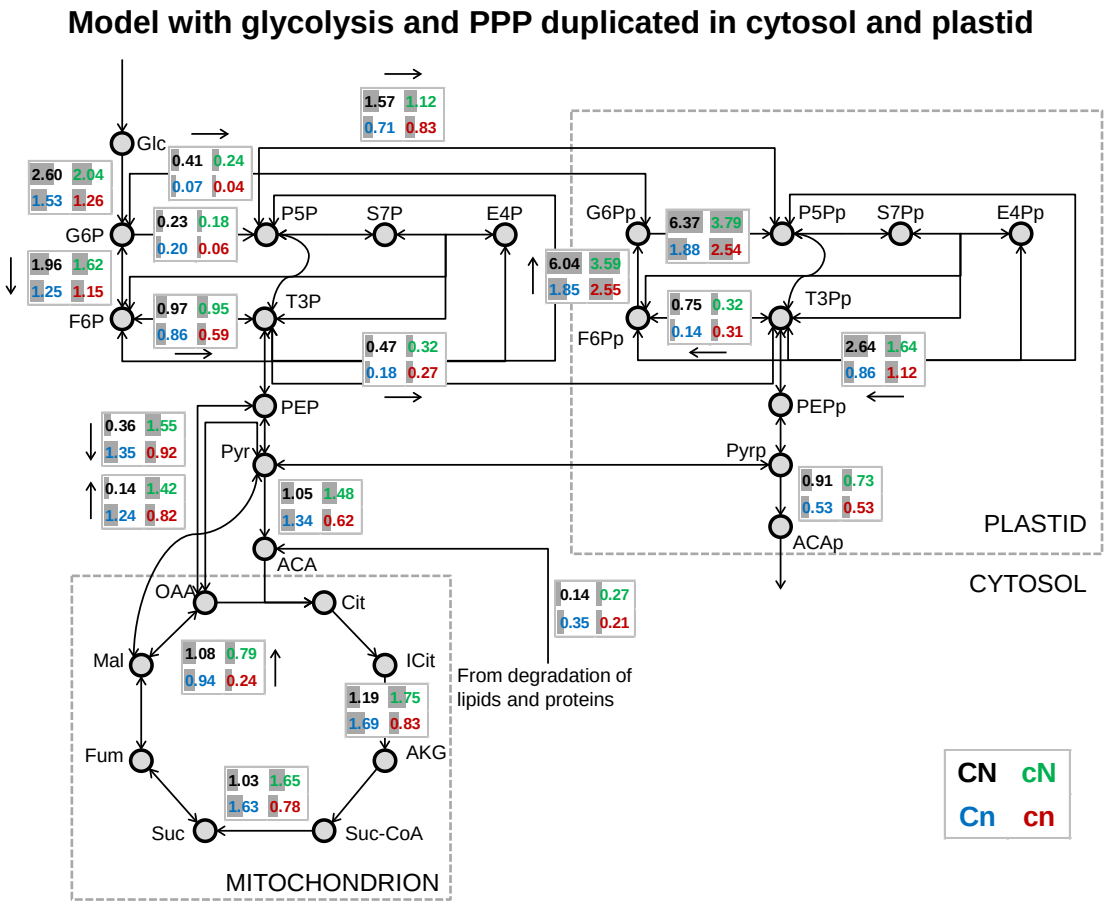
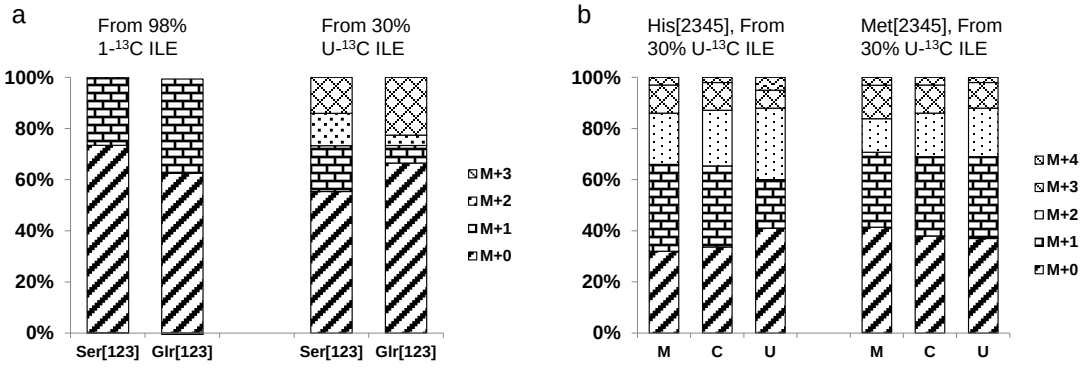


Figure 4.4. Metabolic flux map: low-N conditions are characterized by lower oxPPP flux and higher net anaplerotic flux.
This is the flux map for a three-compartmented (cytosol, mitochondrion and plastid) metabolic network model, with glycolysis and PPP duplicated in both cytosol and plastid. Circles represent metabolites and arrows represent chemical reactions. Flux values for the four conditions CN, cN, Cn and cn are labeled in boxes adjacent to the arrows. The lengths of the grey bars overlaying the numbers represent the scales of the flux values. For bidirectional reactions, net fluxes are indicated.

oxPPP flux (g6pdh) in the cytosol. However, in the plastid there was a very high flux through the oxPPP (g6pdhp) which was realized by the reversal of plastidic glycolysis (pgifp). More than 90% of the total PPP fluxes were found to occur in the plastid. While comparing the overall PPP fluxes (a combination of g6pdh and g6pdhp) across all four conditions, we found lower oxPPP fluxes due to deficiency of carbon or

nitrogen, especially of nitrogen (CN: 6.60 ± 0.89 μmol day^{-1} [mg biomass]$^{-1}$, cN: 3.97 ± 1.0 μmol day^{-1} [mg biomass]$^{-1}$, Cn: 2.08 ± 0.65 μmol day^{-1} [mg biomass]$^{-1}$, cn: 2.60 ± 0.62 μmol day^{-1} [mg biomass]$^{-1}$). The This is consistent with previous studies on soybean embryos showing that higher carbon and nitrogen ratios resulted in lower PPP flux (Allen and Young, 2013).



		Total	His[2345]	Met[2345]
Acceptable SSR in 28% U-^{13}C ILE		164	11	11
Achieved SSR in 28% U-^{13}C ILE	Compart-mented model	82	13	28
	Unicompart-mented model	793	395	66

Figure 4.5. The compartmented model was justified by comparisons between Ser and Glr MIDs and comparison between calculated isotopomer and measured isotopomers.
(a) compared the MIDs of Ser[123] and Glr[123]. MIDs of M+0 through M+3 were represented by lengths of the bars, from bottom to top. **(b)** shows the comparison between measured MIDs (represented by M) and calculated MIDs using multi-compartmented model (represented by C) and uni-compartmented model (represented by U), for two example amino acid fragments: His[2345] and Met[2345] (under CN condition in 30% U-13C ILE). The table below shows the acceptable and achieved sum of squared residues (SSR) by using the two models. This comparison confirms that the calculated isotopomer data using the multi-compartmented model agree better to measured isotopomer data compared to using the uni-compartmented model.

Besides, we observed a much higher forward flux of anaplerotic reactions under low-C and low-N conditions (CN: 0.36 ± 0.23 μmol day^{-1} [mg biomass]$^{-1}$, cN: 1.55 ± 0.27 μmol day^{-1} [mg biomass]$^{-1}$, Cn: 1.35 ± 0.22 μmol day^{-1} [mg biomass]$^{-1}$, cn: 0.92 ± 0.11 μmol day^{-1} [mg biomass]$^{-1}$). Significant influx of unlabeled acetyl-CoA were also observed under low-N conditions (CN: 0.14 ± 0.03 μmol day^{-1} [mg biomass]$^{-1}$, cN: 0.27 ± 0.04 μmol day^{-1} [mg biomass]$^{-1}$, Cn: 0.35 ± 0.04 μmol day^{-1} [mg biomass]$^{-1}$, cn: 0.21 ± 0.04 μmol day^{-1} [mg biomass]$^{-1}$). In our previous article (Nargund et al., 2014), we concluded that the poplar cells were able to obtain additional carbon sources apart from glucose: assimilation of CO_2 into malate and oxaloacetate, and recycled acetyl-CoA produced from degradation of lipids and various amino acids. Here a comparison across the four conditions further indicated that carbon or nitrogen deficiency led to higher acetyl-CoA reflux.

Moreover, we calculated the relative flux ratios at the G6P and PEP-Pyr (i.e. an artificial metabolite that combines PEP and Pyr pools) nodes **(Fig. 4.6)**. At the G6P node, the reversal of pgi (phosphoglucose isomerase) reaction accounted for about 50% of the oxPPP flux for the two low-C conditions. However, a comparison between the two high-C conditions (CN vs. Cn) revealed that when nitrogen source was abundant, pgi reaction was significantly more reversed to maintain a high oxPPP flux. At the PEP-Pyr node, the comparison of flux ratio results exactly mirrors the comparison of absolute flux values: a substantial proportion of carbon flow was redirected to anaplerotic reaction under the deficiency of either carbon or nitrogen supply.

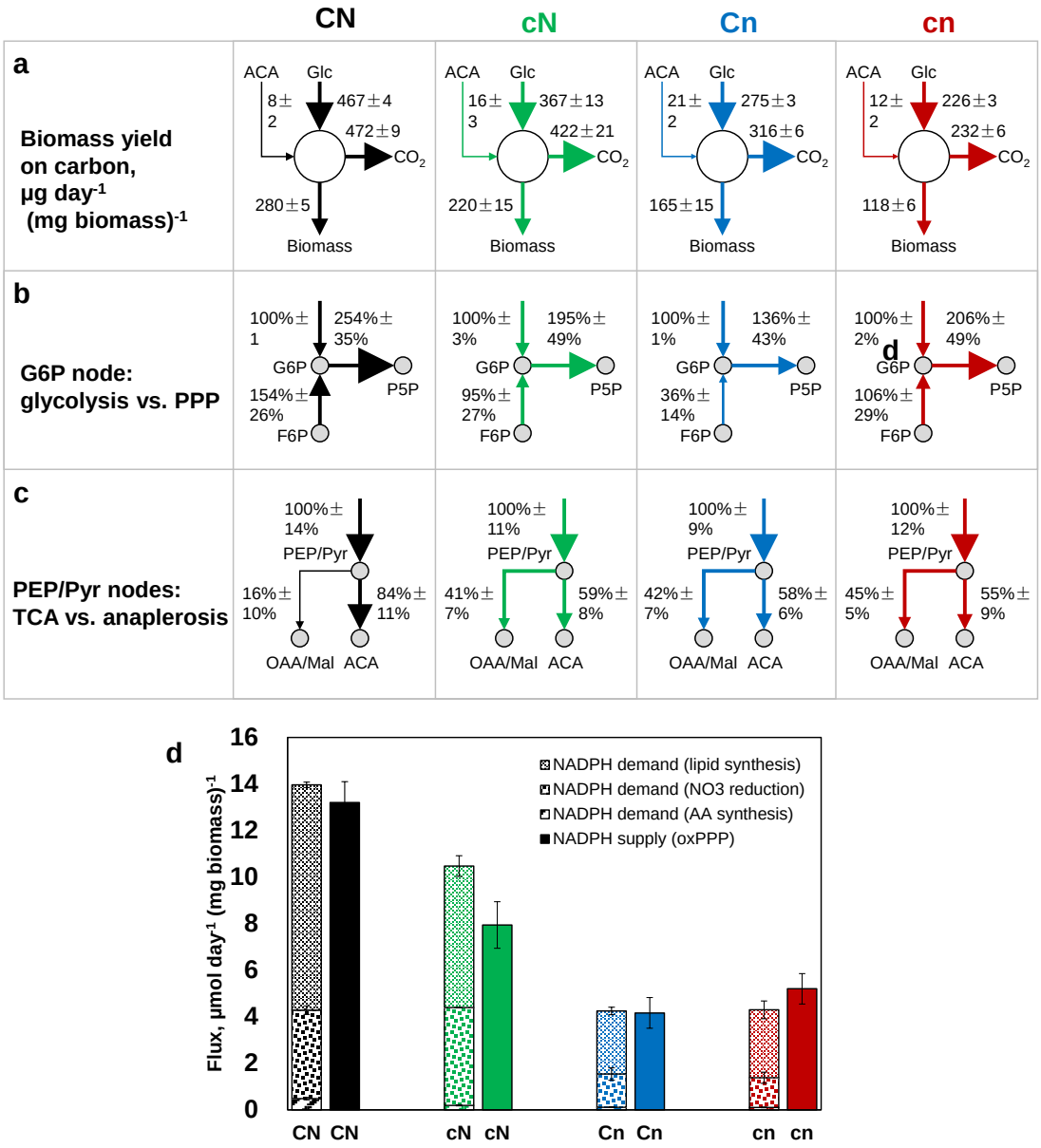


Figure 4.6. Metabolic flux map highlights: carbon partitioning and reductant balance.

Carbon partitioning as calculated from the flux data shown in **Fig. 4.4**: **(a)** overall partitioning of carbon from glucose (supplied carbon source) and acetyl-CoA (from reflux of proteins and lipids) between biomass and CO2 (all units: µg day-1 [mg biomass]-1), **(b)** partitioning between glycolysis and oxPPP at G6P branchpoint, **(c)** partitioning between TCA cycle and anaplerosis at consolidated PEP/Pyr branchpoint; arrow thicknesses are proportional to the relative fluxes. **(d)** NADPH supply through oxPPP (as calculated from the flux data shown in **Fig. 4.3**) and consumption through amino acid synthesis, lipid synthesis and nitrate reduction (as estimated from biomass measurements). NAPDH productions are represented by solid filled bars while NADPH consumptions are represented by patterned filled bars.

4.3.4. Gene expression was regulated more by nitrogen supply than by carbon supply

To learn more insights of how poplar cells adapt to carbon or nitrogen starvations to increase NUE, we performed microarray measurements of the poplar whole genome and an extensive comparison of the gene expression levels of 873 selected metabolic genes from the four environmental conditions. Consistent with the comparison of isotopomer abundances, lower nitrogen supply leads to more differentially expressed genes than lower carbon supply [**Fig. 4.2(b)**]. Moreover, an unsupervised hierarchical clustering study of the metabolic genes via principal component analysis method also showed similar trend. The dendrogram [**Fig. 4.2(d)**] clearly clustered the three biological replicates of each condition at the first level (variance around 5 ~ 15). The two low-N conditions and the two high-N conditions then clustered respectively at the second level (variance around 25 ~ 30).

We further evaluated what specific genes responded to different carbon and nitrogen availability. Specifically, we searched the biological functions of all the 873 metabolic genes via PoplarCyc and UniProt (Apweiler et al., 2004). **Fig. 4.7** shows a comparison between the expressions of genes encoding chemical reactions in central carbon metabolism and the corresponding evaluated fluxes. All fold changes were calculated in $\log_2$ scale with respect to the CN condition. We found that many of these transcriptomic observations agreed well with our flux estimations, such as downregulation of one *G6pDH* (cN, -2.4 ± 0.2; Cn, -3.9 ± 0.3; cn, -4.1 ± 0.1) and *PyrK* (cN, -1.2 ± 0.2; Cn, -1.9 ± 0.2; cn, -2.1 ± 0.2) under low-N conditions. Besides,

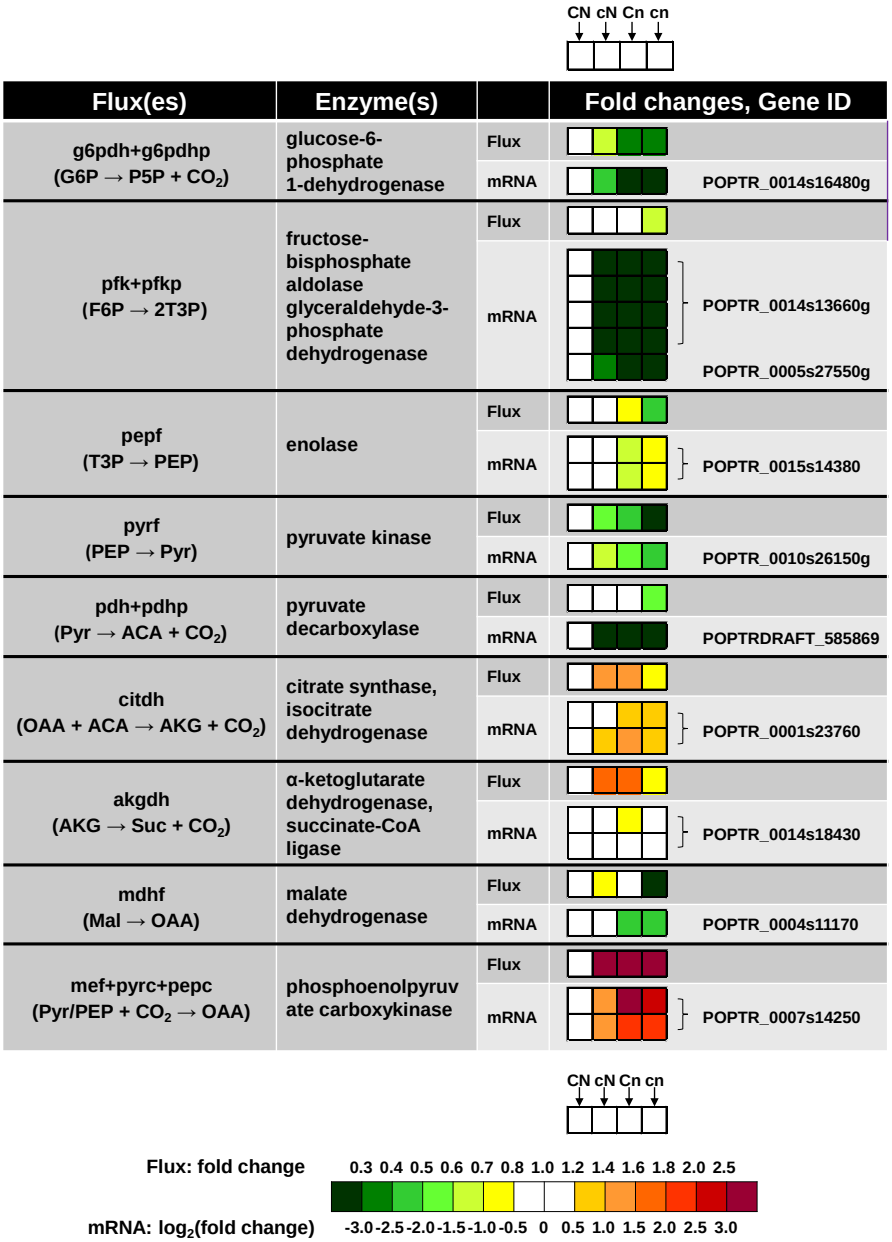


Figure 4.7. Comparison between evaluated fluxes and transcript measurements suggests that many flux changes were transcriptionally regulated.
For nine important metabolic reactions, we compared flux changes with changes in the expression level of the genes encoding the enzymes catalyzing the reactions. Flux names, chemical reaction stoichiometry, enzyme names, gene ensemble IDs as well as fold changes of both fluxes and gene expression levels are listed. Fold changes (relative to the CN condition) are expressed by color scales. Separate color scales are used for flux and gene expression fold changes as shown in the color bar below.

we also observed an upregulation of *PepC* under low-N conditions (cN, 0.1 ± 0.2; Cn, 3.2 ± 0.2; cn, 2.8 ± 0.2), consistent to the higher anaplerotic fluxes predicted by MFA (**Section 4.3.3**).

Despite of these observed consistencies between transcriptomics data and metabolic fluxes, we also found a few inconsistencies. Two examples are the universally underexpressed *GapDH* (glyceraldehyde-3-phosphate dehydrogenase) and underexpressed *PDH* (pyruvate decarboxylase) under low-C or low-N conditions (cN, -3.9 ± 0.1; Cn, -5.5 ± 0.1; cn, -5.6 ± 0.4), while evaluated fluxes show very small difference among all the four conditions.

Moreover, we investigated the expressions of genes that encode enzymes involved in pathways beyond central carbon metabolism. **Fig. 4.8** displays the genes categorized in six groups we are particularly interested in. Since we focus on studying the influence of low nitrogen supply on transcriptome, here we will discuss the comparison between the Cn and the CN condition only. Under nitrogen deficiency, we observed significantly altered transcriptome in central carbon metabolism **(Fig. 4.8a)**, as discussed previously. More differentially expressed genes in primary metabolism include one *AcT* (acyl-CoA thioesterase) and two *AcLs* (acyl-CoA ligase) that were both upregulated under the Cn condition (*AcT*: Cn, 4.4 ± 0.1; *AcL*: Cn, 1.9 ± 0.1 and 1.8 ± 0.1). The transcriptomics data also shed light on the influence of nitrogen availability on synthesis of amino acids. **Fig. 4.8b** showed the genes controlling the synthesis of four amino acids: Gln, Glu, Cys and Gly. Most of them

remain robust against changes of carbon or nitrogen supply. Additionally, the transcriptomics data indicated that productions of many secondary metabolites were also influenced by nitrogen starvation. For example, substantial downregulation of three inositol-3-phosphate synthase genes (**Fig. 4.8c**, *I3pS*: Cn, -7.5 ± 0.1, -3.4 ± 0.5 and -4.4 ± 0.2) were observed; the photosystem and light harvesting genes were also downregulated universally at low nitrogen supply (**Fig. 4.8d**, *LHCB, PSA, PSOE*, etc.: Cn, fold changes ranging from -8.3 ± 0.5 to -0.5 ± 0.1); Five lignin metabolic genes and three phenylalanine metabolic genes were significantly upregulated (**Fig. 4.8e**, *CCAR*: Cn, 4.8 ± 0.2; *CA3MT*: Cn, 1.7 ± 0.1 and 1.8 ± 0.1; *LnCA*: Cn, 5.2 ± 0.1, 5.4 ± 0.1 and 4.4 ± 0.1; *PheAL*: Cn, 4.6 ± 0.2, 2.0 ± 0.7, 1.6 ± 0.2; *APDH*: Cn, 2.2 ± 0.2, 2.4 ± 0.1, 2.1 ± 0.4, 1.5 ± 0.6); four cytochrome P450 defense response genes also showed substantial upregulation (**Fig. 4.8f**, *CYP450D*: Cn, 4.4 ± 0.7, 3.4 ± 0.1, 4.9 ± 0.1 and 4.7 ± 0.2), while two cytochrome P450 oxidation-reduction genes showed substantial downregulation (CYP450OR: Cn, -5.8 ± 0.2 and -5.7 ± 0.3). Overall, these results are consistent with Gutierrez *et al.*'s (2007) study, which concluded that nitrate deprivation resulted in reduced mRNA levels of genes encoding photosynthesis and increased mRNA levels of genes encoding many secondary metabolites.

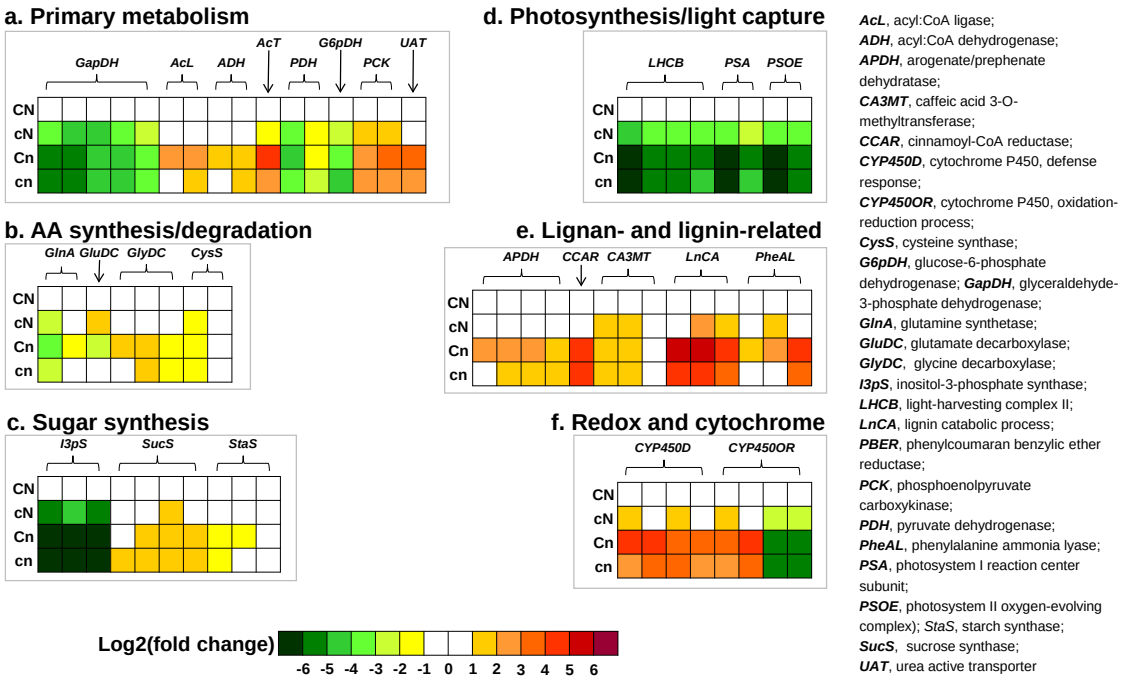


Figure 4.8. Selected families of differentially expressed genes.
Amongst the 873 differentially expressed genes encoding enzymes related to metabolism, these plots depict the expression levels of ~50 genes whose products catalyze reactions in six representative metabolic subsystems including **(a)** central carbon metabolism, **(b)** amino acid synthesis/degradation, **(c)** lignan- and lignin-related, **(d)** redox and cytochrome, **(e)** sugar synthesis and **(f)** photosynthesis/light capture. In each panel, rows correspond to one of the four conditions CN, cN, Cn and cn, whereas columns correspond to individual genes. Expression levels are represented on a $\log_2$ scale using a heat map with a spectrum ranging from red (positive fold change) to green (negative fold change). Gene name abbreviations are shown in the figure and they follow the HUGO Gene Nomenclature Committee guidelines: http://www.genenames.org/hgnc-guidelines#genenames.

4.4. Discussion

4.4.1. Minimization of metabolite reflux and achievement of isotopic steady state

In our previous study (Nargund et al., 2014), we studied the poplar cell suspensions growing for one week under the CN condition as the model system and demonstrated that these cells were characterized by mixing and reflux between

initially present naturally abundant ("unlabeled") biomass and newly synthesized metabolites and biomass. Different from that work which aimed at studying "reflux" phenomenon, here we designed the ILEs such as the influence of "reflux" on isotopomer measurements was minimized. Firstly, we were performing three week ILEs that washed out most of the initially present biomass. Besides, we were measuring amino acids in soluble metabolite pools that were less affected by "reflux" compared to proteinogenic amino acids (Nargund et al., 2014). In fact, the total ^{13}C enrichments of most amino acids measured in this work were substantially higher than those we reported previously, indicating that a much smaller proportion of initially present, unlabeled biomass was left at cell harvesting after three weeks of growth. Nevertheless, we still observed a few isotopomer signatures (e.g. MIDs of Glu[12345]: M+2, 8% ± 1%; M+3, 11% ± 1%) that could be explained by the acetyl-CoA reflux (Nargund et al., 2014), the only "reflux" considered in this study.

Moreover, our experimental design also ensured that the isotopic steady state was approximately achieved. To verify this, we measured the MIDs of intracellular soluble amino acids obtained from cells harvested after each week of growth throughout the three week growth period. Here we selected the cells growing under the CN condition in 30% U-^{13}C ILE as an example. The result (**Fig. 4.9, both panels**) showed that after two weeks of growth, the total ^{13}C enrichments of most amino acids were already very close to 30%. This comparison justified our application of a steady state model to estimate fluxes.

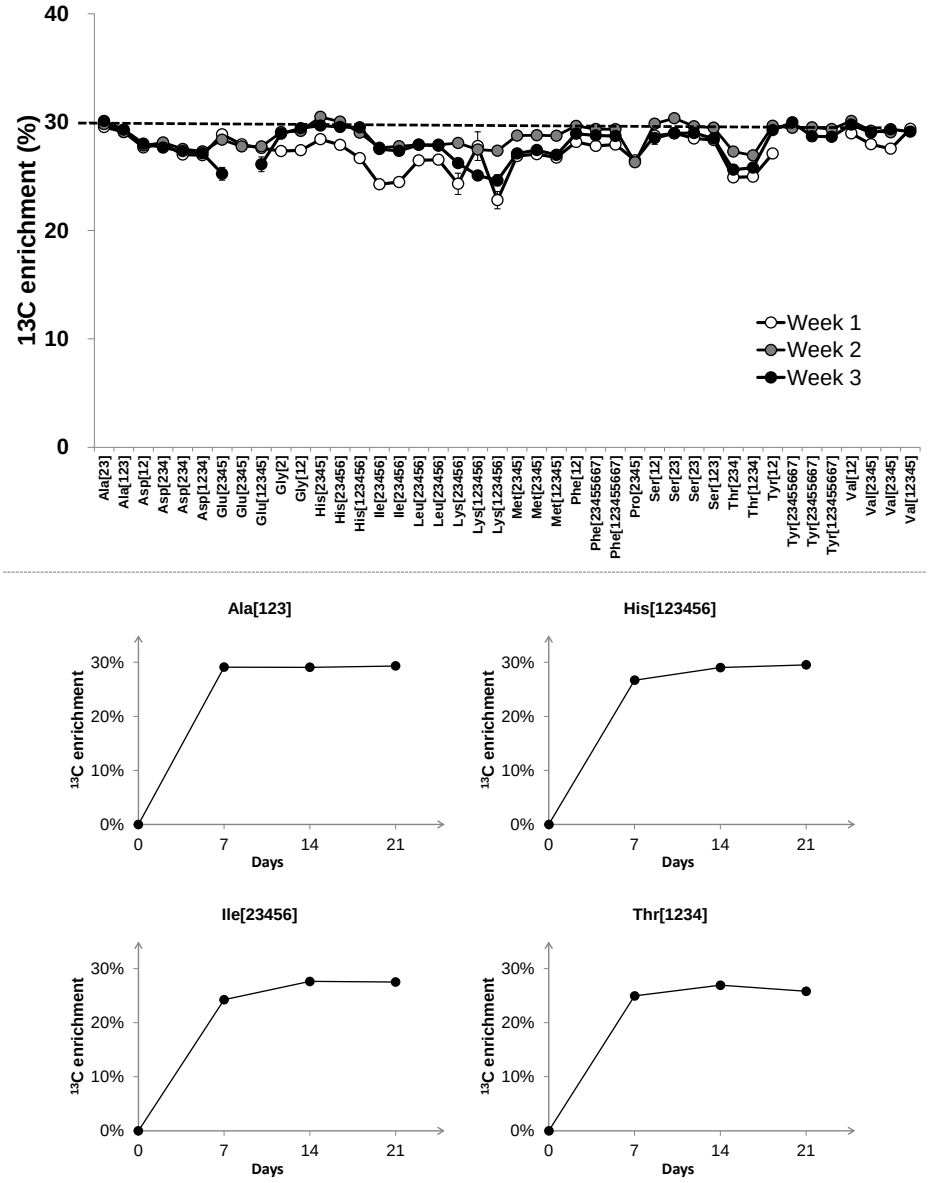


Figure 4.9. The comparisons of MIDs and ^{13}C enrichments of soluble amino acids from the cells harvested after each week of growth verified the achievement of isotopic steady state.

(a), the comparison of ^{13}C enrichments of all the amino acid fragments between from the cells harvested after 7 days, 14 days and 21 days. **(b)**, the development of ^{13}C enrichments of four example amino acids throughout the three weeks of growth. This comparison revealed that the MIDs and ^{13}C enrichments of soluble amino acids from cells of two weeks old and three weeks old were very close, indicating the approximate achievement of isotopic steady state.

4.4.2. Isotopomer abundances and transcriptomics data supported the duplication of glycolysis and PPP in poplar cell suspensions

The duplication of central carbon metabolism in multiple cellular compartments have been studied extensively for other organisms. For example, Masakapalli, *et al.* (2010) concluded that labeling data alone was not sufficient to distinguish the cytosolic and plastidic PPP. Another more recent study on soybean embryo metabolism came to the similar conclusion (Allen and Young, 2013). However, other evidence such as the enzyme localization information suggested that glycolysis and PPP are very likely to be present in both cytosol and plastid (Kruger and von Schaewen, 2003). Here, the compartmentation in poplar cells was supported by multiple observations. One example is the comparison of MIDs between Ser[123] and glycerol[123] fragments (**Fig. 4.5a**) under CN condition. Both Ser and glycerol (Glr) originate from triose phosphate, whereas they exhibit more than 10% difference in MIDs from 98% 1-^{13}C glucose ILE and 30% U-^{13}C glucose ILE. This observation justifies the duplicated T3P pools in cytosol and plastid. Besides, we also performed MFA using a unicompartmented model in which cytosolic pathways and plastidic pathways were not distinguished. Calculated isotopomer data using unicompartmented model showed significantly more inferior agreement with measured isotopomer data than using the multi-compartmented model (sum of squared residues [SSR] under CN condition: multi-compartmented model, 216 vs. unicompartmented model, 1137). **Fig. 4.5b** displays the comparisons between measured and calculated isotopomers (entire calculated isotopomer data is not shown)

obtained from 30% U-^{13}C glucose ILE for two amino acid fragments, His[2345] and Met[2345], as examples.

To further substantiate our compartmented model, we searched the compartmental information for all the *G6pDH* genes of poplar from PLEXdb (Wise et al., 2007) database. Among a total of four *G6pDH* genes (eight probe sets), one plastidic *G6pDH* gene showed substantial underexpression under low N conditions, while all cytosolic *G6pDH* genes showed similar expressions across all the C-N conditions. This was consistent with the evaluated flux results, which revealed that plastidic g6pdh flux accounted for the major difference in total g6pdh flux.

4.4.3. Some flux changes including oxPPP and anaplerotic reactions are likely regulated at the transcriptional level

Under low nitrogen conditions, we have observed some metabolic flux responses similar to those reported by Allen and Young (2013) on soybean embryo, including reduced flux of plastidic oxPPP pathway, increased flux of anaplerotic reactions and fructose bisphosphate aldolase reactions. However, some flux results are different from those reported by Allen and Young (2013), such as a more active upstream TCA cycle under low nitrogen supply. Nevertheless, most of our observed flux changes agreed well with our microarray data, as discussed in **Sec. 4.3.4**, suggesting that these reactions may be controlled at the transcriptional level. Nevertheless, for a few reactions, gene expression data and flux estimation results are deviant, or even at opposite trends **(Fig. 4.7)**. The explanations for this might be that

the enzymes encoded by the genes are present in different organelles or that these chemical reactions are controlled more by post transcriptional intracellular processes like phosphorylation (Tovar-Mendez et al., 2003).

4.4.4. Lower NADPH production through oxPPP is consistent with lower NADPH requirement under low N conditions

It has been shown that altering culturing conditions can significantly redirect carbon flow from Glycolysis to PPP (Ayar-Kayali, 2010; Cadière et al., 2011; Masakapalli et al., 2014, 2013). Here interestingly, we even observed substantially reversed flux of plastidic phosphoglucose isomerase (pgifp, **Sec. 4.3.3**), which is the consequence of the huge oxidative PPP (g6pdh, **Sec. 4.3.3**) flux. This revealed that the poplar cells produce pyruvate, the feedstock of the downstream TCA cycle, primarily through PPP. Our flux estimations showed that cells under high-N conditions exhibit higher oxidative PPP flux and thus more reversed pgi reaction. The first step of PPP, g6pdh reaction, produces two NADPH at the cost of losing one carbon atom in the form of CO_2. One major consumer of NADPH or NADH is the nitrate reduction process. The conversion of one mole of nitrate to ammonia costs one mole of NADPH or NADH, and six moles of reduced ferredoxin (i.e. four moles of NADPH equivalent in total). The operation of the plastidic oxidative PPP has been shown to be directly linked with the assimilation of inorganic nitrogen (Bussel 2013). A knockdown of the plastidic oxidative PPP enzyme, 6-phosphogluconolactonase 3, in Arabidopsis was found to lower the contents of amino acids and lead to an N-starved profile in the mutants compared to wild type plants (Bussel 2013). Another

major consumer of NADPH is the synthesis of fatty acids, since the addition of two carbon atoms to a fatty acid chain costs two NADPH. Estimated fluxes showed that upon nitrogen deficiency, poplar cells preserve the carbon atoms for synthesizing more ATP through TCA cycle, while producing less reducing power in plastid. This is consistent with the similar CO_2 outflux with respect to glucose uptake rate across the four conditions (**Fig. 4.6a**), and also explains similar biomass yield on carbon source showed in **Fig. 4.1b**. Quantitatively, lower NADPH production via oxPPP corresponds well to lower NADPH consumption by nitrate reduction, lipid synthesis and amino acid synthesis under low-N conditions **(Fig. 4.6d)**. Besides, for all the four C-N conditions, NADPH production through oxPPP and NADPH consumption by the above biological processes were approximately balanced.

4.4.5. Photosynthetic and light-harvesting proteins may serve as nitrogen sinks

The gene expression data hold substantial information of secondary metabolism under low nitrogen availability. One interesting observation is the lower activities of photosynthetic and light harvesting genes under low nitrogen availability. Previous reports on various species showed that the expression levels of many genes could be affected by nitrogen starvation. For example, Bi, *et al.* (2007) found that among entire genome 44 genes encoding photosystem I and II family proteins in Arabidopsis leaves were down-regulated under nitrogen starvation (e.g. *At1g03600* and *At4g28660* both show -3.0 fold change under N limitation). Another study also concluded that most photosynthesis related genes were repressed (e.g. *GRMZM2G461279* and *GRMZM2G429955* both show -1.1 fold change under N

limitation, in $\log_2$ scale) in nitrogen deficient maize leaves (Schluter et al., 2012). Martin et al studied the effects of varying C/N availability on the growth of Arabidopsis seedlings, and found that the levels of the photosynthetic genes dropped significantly under N starvation conditions (Martin 2002). Our poplar cells are non-photosynthetic, but we observed much more significant downregulation of photosystem genes under low nitrogen supply. This observation also put forward the possibility of increasing photosynthetic productivity in other photosynthetic tissues or organisms by increasing nitrogen supply.

More importantly, our observation indicated that these photosynthetic and light-harvesting proteins served as nitrogen sinks in poplar cells when nitrogen supply is sufficient. This hypothesis was supported by the substantial larger reflux of acetyl-CoA under Cn condition (0.35 ± 0.04 μmol day^{-1} [mg biomass]$^{-1}$), as compared with CN condition (0.14 ± 0.03 μmol day^{-1} [mg biomass]$^{-1}$), because acetyl-CoA can be produced from degradation of lipids and various amino acids including Glu, Thr, Val and Tyr. In fact, one previous report showed that 16 photosystems genes were significantly under-expressed under nitrogen starvation in rice roots, which are also non-photosynthetic tissues (Cai et al., 2012).

Therefore, based on the comparisons of growth measurements, intracellular fluxes and transcriptomics data across the four conditions, we are proposing a mechanism that the poplar cell suspensions may apply to improve NUE under low nitrogen availability **(Fig. 4.10)**. When nitrogen supply is abundant, oxPPP is more

active, producing more reducing power (NADPH) to incorporate the inorganic nitrogen source into various storage proteins. When nitrogen supply is deficient, oxPPP flux is reduced, consistent with the lower demand of the reducing power to assimilate and reduce NO_3^- and thus less storage proteins are synthesized. Meanwhile, a larger proportion of carbon flow is redirected to the downstream TCA cycle, maintaining a robust ATP production for major cellular activities including biomass synthesis.

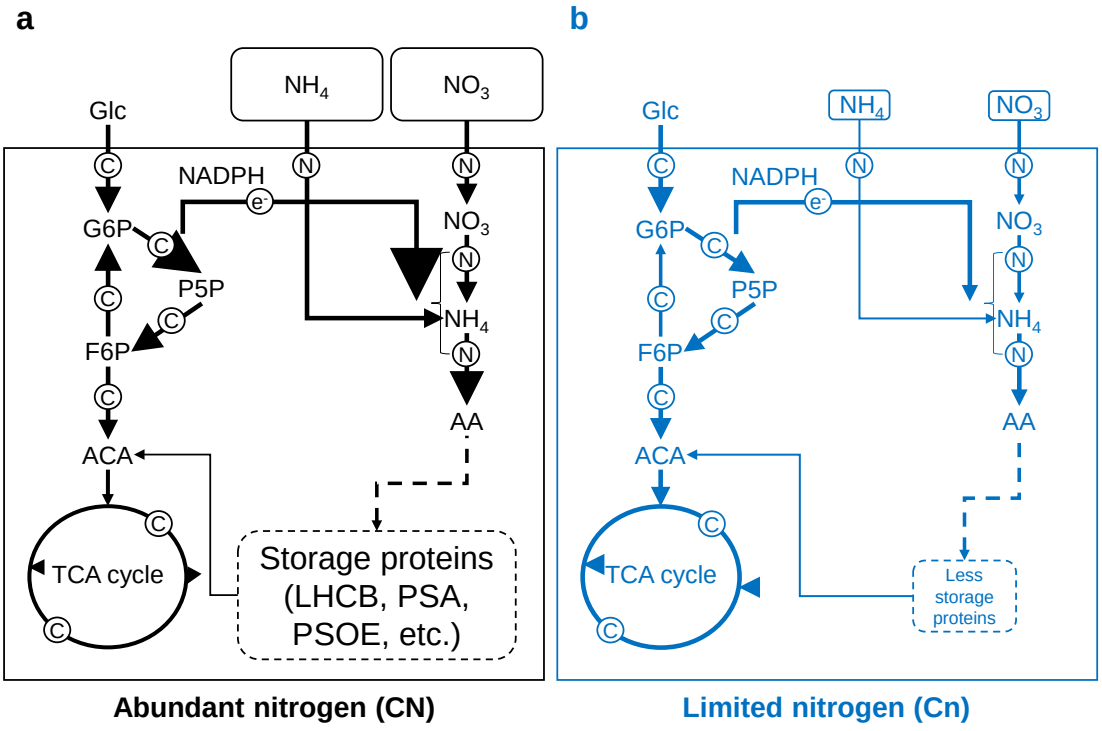


Figure 4.10. Lower NAPDH production, higher TCA cycle flux and (hypothesized) reduced storage protein synthesis contribute to improved NUE under low nitrogen supply.
These diagrams summarize the results for the (a) CN and (b) Cn conditions to propose a mechanism for the higher NUE observed under the low-N conditions. Glucose uptake rates are normalized in both panels and all carbon flows are relative fluxes to glucose uptake fluxes. Arrow thicknesses are subject to different scales between carbon flows (represented by ⓒ), nitrogen flows (Ⓝ) and NADPH flow

(⊖). The areas of the NH4 and NO3 boxes are proportional to their concentrations in media. Dashed lines represents that these processes or explanations are hypothesized.

4.5. Conclusions

In this study we applied a fully compartmented model of central carbon metabolism for MFA. This model was justified not only by the measured isotopomer abundance data and statistical test of calculated isotopomer abundances, but also by the compartmental information of the genes encoding enzymes involved in pentose phosphate pathways. Estimated fluxes explained the macroscopic growth characteristics, whereas they were unable to directly reveal how the poplar cells improve NUE at low nitrogen supply. This is because the understanding of nitrogen recycling mechanism requires information including abundances of related proteins or expression levels of related genes, which cannot be obtained from MFA on central carbon metabolism. Transcriptome, on the other hand, is not capable of predicting the fluxes through intracellular pathways, because gene expressions may not necessarily be consistent with enzyme levels, enzyme activities or flux values. However, transcriptomics data covers a wide range of cellular activities, which provides more insights on how cells responded to nitrogen deficiency beyond central carbon metabolism. Here we successfully connected the fluxomics observations and transcriptomics observations to explain how poplar cells improve NUE under low nitrogen supply. We also showed the consistencies between fluxes and gene expressions (including *G6pDH* and *PepC*) indicating these pathways could be regulated at the transcriptional level.

Chapter 5: Summary and Future Outlooks

5.1. Summary

Isotope-assisted MFA is a technique that integrates experimental design and implementation, measurements of massive amount of labeling data, and simulating labeling data to calculate fluxes via powerful computational tools. This work aims at: (1) Increasing the computational efficiency and improving the accuracy of flux estimation by developing new modeling technique; (2) Applying MFA to an important case study – investigating the metabolic landscape influence of nutrient deficiency on poplar cells and understand the mechanism to improve NUE. The two approaches we proposed in the first chapter successfully improved the computational efficiency for instationary MFA. Comparisons of the total number of labeling variables that need to be balanced and the computational time required to solve the balance equations, shows that applying these new modeling approaches leads to a ~20 fold improvement of computational efficiency in some cases. This potentially allows MFA to be applied to large, complex metabolic networks, which is currently difficult.

Additionally, we developed a metabolic network modeling strategy, “reflux”, to account for the influence of “initial present” biomass on the accuracy of flux evaluation. Compared with previously reported modeling strategies, isotopomers calculated via this “reflux” model agrees considerably better with measurement data. Besides, the estimate flux results using this reflux model and the flux results from another methodology involving minimal reflux are consistent. With the inevitability

of cell cultures in “fed-batch” environment, we anticipate a widespread application of this “reflux” modeling method in steady state MFA.

We applied steady state MFA on cell suspensions from poplar tree, which has been shown to be capable of recycling its nitrogen reserves efficiently. We compared the fluxomics data and transcriptomics data obtained from cells growing under different carbon and nitrogen supply levels. Both ^{13}C MFA results and gene expression results show that central carbon metabolism are substantially affected by nitrogen deficiency. This includes a lower oxPPP flux and a higher activity of anaplerotic reactions. Expression levels of many genes encoding secondary metabolism further shed light on potential NUE improving mechanisms, which can be validated by proteiomics measurements.

5.2. Future outlooks

The “reflux” model discussed in this study is an isotopic steady-state model. However, a steady-state model of metabolite reflux is not exact as reflux of the initial biomass in a batch culture is strictly an instationary process. For example, the assumption that refluxed biomass components are naturally abundant throughout the ILE is only a first approximation. In reality, the aggregate (sum of initial and newly synthesized) biomass gets more and more ^{13}C-enriched during the course of the ILE; correspondingly, the refluxed biomass components get gradually enriched during the ILE. Here we used an isotopic steady state model as an approximation of dynamic MFA to simulate the reflux process and demonstrated its ability to explain the

anomalous isotopomer measurements. Nevertheless, it is still desirable to incorporate the metabolite reflux model into instationary isotope MFA (M. Antoniewicz et al., 2007; Nöh and Wiechert, 2006; Wiechert and Noh, 2005). By accomplishing this, we should expect more accurate metabolic pool size estimations as well as intracellular flux estimations. We hope the future studies of metabolite reflux in isotope MFA will incorporate these extensions.

To formulate a complete conclusion on what exact proteins the poplar cells recruit to store the redundant nitrogen and recycle these nitrogen reserves, more proteomics studies are desirable. Some previous work have reported the methodology to identify peptides in *Arabidopsis thaliana* via ^{15}N isotopic labeling experiments (Nelson et al., 2007) and to infer compartmental information of specific proteins via MFA applying ^{13}C peptide label experiments (Mandy et al., 2014). Besides, the application of genome scale models (e.g. Poolman et al., 2009) may also help revealing where the carbon or nitrogen flow is redirected. Future studies may include manipulating genes that are likely to be involved in NUE-improving mechanisms and testing its influence on nitrogen uptake rate and fluxes in central carbon metabolism. For example, our hypothesized mechanism will be strengthened if a knock-out or knock-down of *g6pdh* genes and light harvesting genes result in decreased nitrogen uptake rate and slowed cell growth. We anticipate that this work enables a more in-depth understanding of how poplar cells improves nitrogen source efficiency at the metabolic and genetic levels and moreover, shed light on potential strategies of genetic modifications to improve NUE in other organisms.

Appendices

Table S1. Mass isotopomer abundances of proteinogenic amino acids from three ILEs

MIDs of proteinogenic amino acids were measured in three ILEs performed for 7 d on poplar cell suspensions under light. The ILEs used 28% U-13C, 100% 1-13C and 98% U-13C glucose, respectively, as the sole organic carbon source. MIDs are reported as percentages and were obtained by correcting GC-MS spectral data for the natural isotopic abundances of O, H, N, S, P, Si and non-metabolic C atoms14. For each ILE, we report individual MID measurements of three biological replicates, an average and a standard deviation (SD). ND, not determined.

Table S2. Mass isotopomer abundances of intracellular metabolites from one-week ILE under light and dark environment.

MIDs of intracellular metabolites were measured in ILEs performed for 7 d on poplar cell suspensions grown under light and dark. The ILEs used 98% U-^{13}C glucose as the sole organic carbon source. MIDs are reported as percentages and were obtained by correcting GC-MS spectral data for the natural isotopic abundances of O, H, N, S, P, Si and non-metabolic C atoms(Sriram et al., 2008). For each condition, we report individual MID measurements of three biological replicates, an average and a standard deviation (SD).

Table S3. Fluxes evaluated through the four metabolic models

This table reports averages and standard deviations of flux distributions obtained by performing 280 bootstrap Monte Carlo simulations (Press et al., 2007) that accounted for the larger of the measurement and biological errors in the measured isotopomer abundances. The three models employed amino acid reflux (AA_in), glucose dilution (Gluc_dilu) and isotopomer correction (Iso_corr), respectively, to model the batch culture. **Tables S4, S5, S6** and **S7** list stoichiometries and carbon atom rearrangements for the reactions. SD, standard deviation.

Tables S4-S7. Metabolic network models: S4: uniform isotopomer correction model (Iso_corr[U]); S5: glucose dilution model (Gluc_dilu); S6: variable isotopomer correction model (Iso_corr[V]); S7: amino acid reflux model (AA_in)

In these three-compartment models, metabolites with a subscript "p" are plastidic, metabolites with a subscript "m" are mitochondrial and metabolites without a are cytosolic. Carbon atom rearrangements of reactions are indicated by the numbers in parenthesis; these numbers correspond to carbon atoms numbered according to IUPAC rules. Estimated net flux values, standard deviations and reaction reversibilities are shown for each reaction. For a bidirectional reaction: $A \underset{v_b}{\overset{v_f}{\rightleftharpoons}} B$, where v_f and v_b are the fluxes of the forward and backward reactions, respectively ($v_f \geq v_b$) , the net flux is $v_f - v_b$ and the reaction reversibility is v_b/v_f (expressed as a percentage). SD, standard deviation.

Table S8-S10. Mass isotopomer distribution (MID) measurements for three ILEs under different carbon-nitrogen conditions.

For each carbon-nitrogen condition, we measured the MIDs of soluble amino acids in three ILEs (**Table S8.**, 98% 1-^{13}C glucose; **Table S9.**, 28% U-^{13}C glucose and **Table S10.**, 98% U-^{13}C glucose). MIDs were presented after the correction for the natural abundances of N, O, H, S, Si, P and non-metabolic C atoms using a MATLAB program. There are three biological replicates in each ILE. Averages and standard deviations of MIDs, as well as averages and standard deviations of total ^{13}C enrichment of each MS fragment, are shown here.

Table S11. The compartmented metabolic network model used in MFA.

In this compartmented model, Column A displays the flux names that correspond to the flux values in **Table S12.** Column B displays the reaction stoichiometry and carbon atom rearrangements of each chemical reaction. The carbon atoms of the reactants are numbered based on IUPAC rules. Column C indicates if the reactions are reversible (reversible if with an “R”).

Table S12. Metabolic fluxes for each chemical reaction under each carbon-nitrogen condition.

The flux names are consistent with that in **Table S4.** Flux values and standard deviations were obtained via MFA on the MIDs measured from all three ILEs, as discussed in **Section 5.4**. The units of the flux values and the standard deviations are μmol day^{-1} $(mg\ biomass)^{-1}$. Negative flux values indicate that the net fluxes flow at the opposite direction to the default direction shown by the reaction stoichiometry in **Table S11**.

Bibliography

Ahn, W.S., Antoniewicz, M.R., 2011. Metabolic flux analysis of CHO cells at growth and non-growth phases using isotopic tracers and mass spectrometry. Metab. Eng. 13, 598–609. doi:10.1016/j.ymben.2011.07.002

Allen, A.E., Dupont, C.L., Obornik, M., Horak, A., Nunes-Nesi, A., McCrow, J.P., Zheng, H., Johnson, D.A., Hu, H., Fernie, A.R., Bowler, C., 2011. Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. Nature 473, 203–207. doi:10.1038/nature10074

Allen, D.K., Young, J.D., 2013. Carbon and nitrogen provisions alter the metabolic flux in developing soybean embryos. Plant Physiol. 161, 1458–1475. doi:10.1104/pp.112.203299

Amtmann, A., Armengaud, P., 2009. Effects of N, P, K and S on metabolism: new knowledge gained from multi-level analysis. Curr. Opin. Plant Biol. 12, 275–283. doi:10.1016/j.pbi.2009.04.014

Antoniewicz, M., Kraynie, D., Laffend, L., Gonzalezlergier, J., Kelleher, J., Stephanopoulos, G., 2007. 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. doi:10.1016/j.ymben.2007.01.003

Antoniewicz, M.R., Kelleher, J.K., Stephanopoulos, G., 2007. Elementary metabolite units (EMU): A novel framework for modeling isotopic distributions. Metab. Eng. 9, 68–86. doi:10.1016/j.ymben.2006.09.001

Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M.J., Natale, D.A., O'Donovan, C., Redaschi, N., Yeh, L.-S.L., 2004. UniProt: the Universal Protein knowledgebase. Nucleic Acids Res. 32, D115–119. doi:10.1093/nar/gkh131

Ayar-Kayali, H., 2010. Pentose phosphate pathway flux analysis for glycopeptide antibiotic vancomycin production during glucose-limited cultivation of *Amycolatopsis orientalis*. Prep. Biochem. Biotechnol. 41, 94–105. doi:10.1080/10826068.2010.535401

Bi, Y.-M., Wang, R.-L., Zhu, T., Rothstein, S.J., 2007. Global transcription profiling reveals differential responses to chronic nitrogen stress and putative nitrogen regulatory components in *Arabidopsis*. BMC Genomics 8, 281. doi:10.1186/1471-2164-8-281

Cadière, A., Ortiz-Julien, A., Camarasa, C., Dequin, S., 2011. Evolutionary engineered *Saccharomyces cerevisiae* wine yeast strains with increased in vivo flux through the pentose phosphate pathway. Metab. Eng. 13, 263–271. doi:10.1016/j.ymben.2011.01.008

Cai, H., Lu, Y., Xie, W., Zhu, T., Lian, X., 2012. Transcriptome response to nitrogen starvation in rice. J. Biosci. 37, 731–747. doi:10.1007/s12038-012-9242-2

Cantón, F.R., Suárez, M.F., Cánovas, F.M., 2005. Molecular aspects of nitrogen mobilization and recycling in trees. Photosynth. Res. 83, 265–278. doi:10.1007/s11120-004-9366-9

Caspi, R., Altman, T., Dreher, K., Fulcher, C.A., Subhraveti, P., Keseler, I.M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L.A., Ong, Q.,

Paley, S., Pujar, A., Shearer, A.G., Travers, M., Weerasinghe, D., Zhang, P., Karp, P.D., 2011. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. doi:10.1093/nar/gkr1014

Chen, X., Alonso, A.P., Allen, D.K., Reed, J.L., Shachar-Hill, Y., 2011. Synergy between ^{13}C-metabolic flux analysis and flux balance analysis for understanding metabolic adaption to anaerobiosis in *E. coli*. Metab. Eng. 13, 38–48. doi:10.1016/j.ymben.2010.11.004

Choi, J., Antoniewicz, M.R., 2011. Tandem mass spectrometry: a novel approach for metabolic flux analysis. Metab. Eng. 13, 225–233. doi:10.1016/j.ymben.2010.11.006

Crown, S.B., Ahn, W.S., Antoniewicz, M.R., 2012. Rational design of ^{13}C-labeling experiments for metabolic flux analysis in mammalian cells. BMC Syst. Biol. 6, 43. doi:10.1186/1752-0509-6-43

Crown, S.B., Antoniewicz, M.R., 2012a. Selection of tracers for ^{13}C-Metabolic Flux Analysis using Elementary Metabolite Units (EMU) basis vector methodology. Metab. Eng. 14, 150–161. doi:10.1016/j.ymben.2011.12.005

Crown, S.B., Antoniewicz, M.R., 2012b. Parallel labeling experiments and metabolic flux analysis: Past, present and future methodologies. Metab. Eng. 16C, 21–32. doi:10.1016/j.ymben.2012.11.010

Dauner, M., Bailey, J.E., Sauer, U., 2001. Metabolic flux analysis with a comprehensive isotopomer model inBacillus subtilis. Biotechnol. Bioeng. 76, 144–156. doi:10.1002/bit.1154

Escobar, M.A., Geisler, D.A., Rasmusson, A.G., 2006. Reorganization of the alternative pathways of the Arabidopsis respiratory chain by nitrogen supply: opposing effects of ammonium and nitrate. Plant J. Cell Mol. Biol. 45, 775–788. doi:10.1111/j.1365-313X.2005.02640.x

Fiehn, O., Kopka, J., Trethewey, R.N., Willmitzer, L., 2000. Identification of Uncommon Plant Metabolites Based on Calculation of Elemental Compositions Using Gas Chromatography and Quadrupole Mass Spectrometry. Anal. Chem. 72, 3573–3580. doi:10.1021/ac991142i

Gao, P., Xin, Z., Zheng, Z.-L., 2008. The OSU1/QUA2/TSD2-encoded putative methyltransferase is a critical modulator of carbon and nitrogen nutrient balance response in *Arabidopsis*. PloS One 3, e1387. doi:10.1371/journal.pone.0001387

Goudar, C., Biener, R., Boisart, C., Heidemann, R., Piret, J., de Graaf, A., Konstantinov, K., 2010. Metabolic flux analysis of CHO cells in perfusion culture by metabolite balancing and 2D [^{13}C, ^{1}H] COSY NMR spectroscopy. Metab. Eng. 12, 138–149. doi:10.1016/j.ymben.2009.10.007

Gutierrez, R.A., Lejay, L.V., Dean, A., Chiaromonte, F., Shasha, D.E., Coruzzi, G.M., 2007. Qualitative network models and genome-wide expression data define carbon/nitrogen-responsive molecular machines in *Arabidopsis*. GENOME Biol. 8. doi:10.1186/gb-2007-8-1-r7

Heldt, H.-W., Piechulla, B., 2010. Plant Biochemistry, 4th ed. Academic Press, Burlington, MA

He, L., Xiao, Y., Gebreselassie, N., Zhang, F., Antoniewiez, M.R., Tang, Y.J., Peng, L., 2014. Central metabolic responses to the overproduction of fatty acids in *Escherichia coli* based on ^{13}C-metabolic flux analysis. Biotechnol. Bioeng. 111, 575–585. doi:10.1002/bit.25124

Hockin, N.L., Mock, T., Mulholland, F., Kopriva, S., Malin, G., 2011. The Response of Diatom Central Carbon Metabolism to Nitrogen Starvation Is Different from That of Green Algae and Higher Plants. PLANT Physiol. 158, 299–312. doi:10.1104/pp.111.184333

Iyer, V.V., Sriram, G., Fulton, D.B., Zhou, R., Westgate, M.E., Shanks, J.V., 2008. Metabolic flux maps comparing the effect of temperature on protein and oil biosynthesis in developing soybean cotyledons. Plant Cell Environ. 31, 506–517. doi:10.1111/j.1365-3040.2008.01781.x

Kang, J., Turano, F.J., 2003. The putative glutamate receptor 1.1 (AtGLR1.1) functions as a regulator of carbon and nitrogen metabolism in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. 100, 6872–6877. doi:10.1073/pnas.1030961100

Kind, S., Becker, J., Wittmann, C., 2013. Increased lysine production by flux coupling of the tricarboxylic acid cycle and the lysine biosynthetic pathway—Metabolic engineering of the availability of succinyl-CoA in *Corynebacterium glutamicum*. Metab. Eng. 15, 184–195. doi:10.1016/j.ymben.2012.07.005

Klapa, M.I., Aon, J.C., Stephanopoulos, G., 2003. Ion-trap mass spectrometry used in combination with gas chromatography for high-resolution metabolic flux determination. Biotechniques 34, 832–6, 838, 840 passim.

Kruger, N.J., Huddleston, J.E., Le Lay, P., Brown, N.D., Ratcliffe, R.G., 2007. Network flux analysis: Impact of ^{13}C-substrates on metabolism in *Arabidopsis thaliana* cell suspension cultures. Phytochemistry 68, 2176–2188. doi:10.1016/j.phytochem.2007.03.033

Kruger, N.J., Masakapalli, S.K., Ratcliffe, R.G., 2011. Strategies for investigating the plant metabolic network with steady-state metabolic flux analysis: lessons from an *Arabidopsis* cell culture and other systems. J. Exp. Bot. 63, 2309–2323. doi:10.1093/jxb/err382

Kruger, N.J., Ratcliffe, R.G., 2009. Insights into plant metabolic networks from steady-state metabolic flux analysis. Biochimie 91, 697–702. doi:10.1016/j.biochi.2009.01.004

Kruger, N.J., von Schaewen, A., 2003. The oxidative pentose phosphate pathway: structure and organisation. Curr. Opin. Plant Biol. 6, 236–246. doi:10.1016/S1369-5266(03)00039-6

Lehninger, A.L., 2005. Lehninger principles of biochemistry, 4th ed. ed. W.H. Freeman, New York.

Leighty, R.W., Antoniewicz, M.R., 2012. Parallel labeling experiments with [U-^{13}C] glucose validate *E. coli* metabolic network model for ^{13}C metabolic flux analysis. Metab. Eng. 14, 533–541. doi:10.1016/j.ymben.2012.06.003

Little, D.Y., Rao, H., Oliva, S., Daniel-Vedele, F., Krapp, A., Malamy, J.E., 2005. The putative high-affinity nitrate transporter NRT2.1 represses lateral root initiation in response to nutritional cues. Proc. Natl. Acad. Sci. U. S. A. 102, 13693–13698. doi:10.1073/pnas.0504219102

Lonien, J., Schwender, J., 2009. Analysis of metabolic flux phenotypes for two *Arabidopsis thaliana* mutants with severe impairment in seed storage lipid synthesis. Plant Physiol 151, 1617–1634. doi:10.1104/pp.109.144121

Mandy, D.E., Goldford, J.E., Yang, H., Allen, D.K., Libourel, I.G.L., 2014. Metabolic flux analysis using ^{13} C peptide label measurements. Plant J. 77, 476–486. doi:10.1111/tpj.12390

Marschner, H. (Ed.), 2012. Marschner's mineral nutrition of higher plants, 3rd ed. ed. Elsevier/Academic Press, London ; Waltham, MA.

Masakapalli, S.K., Kruger, N.J., Ratcliffe, R.G., 2013. The metabolic flux phenotype of heterotrophic *Arabidopsis* cells reveals a complex response to changes in nitrogen supply. Plant J. 74, 569–582. doi:10.1111/tpj.12142

Masakapalli, S.K., Lay, P.L., Huddleston, J.E., Pollock, N.L., Kruger, N.J., Ratcliffe, R.G., 2010. Subcellular flux analysis of central metabolism in a heterotrophic *Arabidopsis thaliana* cell suspension using steady-state stable isotope labeling. Plant Physiol. 152, 602–609. doi:10.1104/pp.109.151316

Masakapalli, S.K., Ritala, A., Dong, L., van der Krol, A.R., Oksman-Caldentey, K.-M., Ratcliffe, R.G., Sweetlove, L.J., 2014. Metabolic flux phenotype of tobacco hairy roots engineered for increased geraniol production. Phytochemistry 99, 73–85. doi:10.1016/j.phytochem.2013.12.007

Masoudi-Nejad, A., Goto, S., Endo, T.R., Kanehisa, M., 2008. KEGG bioinformatics resource for plant genomics research, in: Edwards, D. (Ed.), Plant Bioinformatics, Methods in Molecular Biology. Humana Press, pp. 437–458.

Metallo, C.M., Walther, J.L., Stephanopoulos, G., 2009. Evaluation of ^{13}C isotopic tracers for metabolic flux analysis in mammalian cells. J. Biotechnol. 144, 167–174. doi:10.1016/j.jbiotec.2009.07.010

Nargund, S., Misra, A., Zhang, X., Coleman, G.D., Sriram, G., 2014. Flux and reflux: metabolite reflux in plant suspension cells and its implications for isotope-assisted metabolic flux analysis. Mol. Biosyst. 10, 1496. doi:10.1039/c3mb70348g

Nargund, S., Sriram, G., 2013. Designer labels for plant metabolism: statistical design of isotope labeling experiments for improved quantification of flux in complex plant metabolic networks. Mol. Biosyst. 9, 99. doi:10.1039/c2mb25253h

Nelson, C.J., Huttlin, E.L., Hegeman, A.D., Harms, A.C., Sussman, M.R., 2007. Implications of15N-metabolic labeling for automated peptide identification in *Arabidopsis thaliana*. PROTEOMICS 7, 1279–1292. doi:10.1002/pmic.200600832

Nöh, K., Wahl, A., Wiechert, W., 2006. Computational tools for isotopically instationary ^{13}C labeling experiments under metabolic steady state conditions. Metab. Eng. 8, 554–577. doi:10.1016/j.ymben.2006.05.006

Nöh, K., Wiechert, W., 2006. Experimental design principles for isotopically instationary ^{13}C labeling experiments. Biotechnol. Bioeng. 94, 234–251. doi:10.1002/bit.20803

Nöh, K., Wiechert, W., 2004. Parallel solution of cascaded ode systems applied to ^{13}C-labeling experiments, in: Bubak, M., Albada, G.D., Sloot, P.M.A., Dongarra, J. (Eds.), Computational Science - ICCS 2004. Springer Berlin

Heidelberg, Berlin, Heidelberg, pp. 594–597. doi:10.1007/978-3-540-24687-9_80

Noorhana, 2011. Effect of Nitrogen on Growth and Lipid Content of *Chlorella pyrenoidosa*. Am. J. Biochem. Biotechnol. 7, 124–129. doi:10.3844/ajbbsp.2011.124.129

Nunes-Nesi, A., Fernie, A.R., Stitt, M., 2010. Metabolic and signaling aspects underpinning the regulation of plant carbon nitrogen interactions. Mol. Plant 3, 973–996. doi:10.1093/mp/ssq049

Oliveira IC., Brenner E., Chiu J., Hsieh MH., Kouranov A., Lam HM., Shin MJ., Coruzzi G., 2001. Metabolite and light regulation of metabolism in plants: lessons from the study of a single biochemical pathway. Braz J Med Biol Res. 34, 567-575. doi:10.1590/S0100-879X2001000500003

Omasa, T., Furuichi, K., Iemura, T., Katakura, Y., Kishimoto, M., Suga, K., 2009. Enhanced antibody production following intermediate addition based on flux analysis in mammalian cell continuous culture. Bioprocess Biosyst. Eng. 33, 117–125. doi:10.1007/s00449-009-0351-8

Orth, J.D., Thiele, I., Palsson, B.Ø., 2010. What is flux balance analysis? Nat. Biotechnol. 28, 245–248. doi:10.1038/nbt.1614

Peterman, TK., Goodman, HM., 1991. The glutamine synthetase gene family of Arabidopsis thaliana: light-regulation and differential expression in leaves, roots and seeds. Mol Gen Genet. 230, 145-154. doi:10.1007/BF00290662

Poolman, M.G., Miguet, L., Sweetlove, L.J., Fell, D.A., 2009. A genome-scale metabolic model of *Arabidopsis thaliana* and some of its properties. Plant Physiol pp.109.141267. doi:10.1104/pp.109.141267

Pregitzer, K.S., Dickmann, D.I., Hendrick, R., Nguyen, P.V., 1990. Whole-tree carbon and nitrogen partitioning in young hybrid poplars. Tree Physiol. 7, 79–93. doi:10.1093/treephys/7.1-2-3-4.79

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 2007. Numerical Recipes 3rd Edition: The Art of Scientific Computing, 3rd ed. Cambridge University Press.

Quek, L.-E., Dietmair, S., Krömer, J.O., Nielsen, L.K., 2010. Metabolic flux analysis in mammalian cell culture. Metab. Eng. 12, 161–171. doi:10.1016/j.ymben.2009.09.002

Quek, L.-E., Wittmann, C., Nielsen, L.K., Krömer, J.O., 2009. OpenFLUX: efficient modelling software for ^{13}C-based metabolic flux analysis. Microb. Cell Factories 8, 25. doi:10.1186/1475-2859-8-25

Ramakrishna, R., Edwards, J.S., Mcculloch, A., Palsson, B.O., 2000. Flux-balance analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints. Am J Physiol Regul Integr Comp Physiol 280, 695–704.

Ranganathan, S., Tee, T.W., Chowdhury, A., Zomorrodi, A.R., Yoon, J.M., Fu, Y., Shanks, J.V., Maranas, C.D., 2012. An integrated computational and experimental study for overproducing fatty acids in *Escherichia coli*. Metab. Eng. 14, 687–704. doi:10.1016/j.ymben.2012.08.008

Raun, W., Johnson, G., 1999. Improving nitrogen use efficiency for cereal production. Agron. J. 91, 357–363. doi:10.2134/agronj1999.00021962009100030001x

Sahrawy, M., Avila, C., Chueca, A., Cánovas, F.M., López-Gorgé, J., 2004. Increased sucrose level and altered nitrogen metabolism in *Arabidopsis thaliana* transgenic plants expressing antisense chloroplastic fructose-1,6-bisphosphatase. J. Exp. Bot. 55, 2495–2503. doi:10.1093/jxb/erh257

Sannigrahi, P., Ragauskas, A.J., Tuskan, G.A., 2010. Poplar as a feedstock for biofuels: A review of compositional characteristics. Biofuels Bioprod. Biorefining 4, 209–226. doi:10.1002/bbb.206

Sauer, U., 2006. Metabolic networks in motion: ^{13}C-based flux analysis. Mol Syst Biol 2, 62. doi:10.1038/msb4100109

Schluter, U., Mascher, M., Colmsee, C., Scholz, U., Brautigam, A., Fahnenstich, H., Sonnewald, U., 2012. Maize source leaf adaptation to nitrogen deficiency affects not only nitrogen and carbon metabolism but also control of phosphate homeostasis. Plant Physiol. 160, 1384–1406. doi:10.1104/pp.112.204420

Schmidt, K., Carlsen, M., Nielsen, J., Villadsen, J., 1997. Modeling isotopomer distributions in biochemical networks using isotopomer mapping matrices. Biotechnol. Bioeng. 55, 831–840. doi:10.1002/(SICI)1097-0290(19970920)55:6<831::AID-BIT2>3.0.CO;2-H

Schwender, J., 2008. Metabolic flux analysis as a tool in metabolic engineering of plants. Curr. Opin. Biotechnol. 19, 131–137. doi:10.1016/j.copbio.2008.02.006

Schwender, J., Goffman, F., Ohlrogge, J.B., Shachar-Hill, Y., 2004. Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. Nature 432, 779–782. doi:10.1038/nature03145

Shen, T., Rui, B., Zhou, H., Zhang, X., Yi, Y., Wen, H., Zheng, H., Wu, J., Shi, Y., 2013. Metabolic flux ratio analysis and multi-objective optimization revealed a globally conserved and coordinated metabolic response of E. coli to paraquat-induced oxidative stress. Mol. Biosyst. 9, 121. doi:10.1039/c2mb25285f

Sriram, G., Fulton, D.B., Iyer, V.V., Peterson, J.M., Zhou, R., Westgate, M.E., Spalding, M.H., Shanks, J.V., 2004. Quantification of compartmented metabolic fluxes in developing soybean embryos by employing biosynthetically directed fractional ^{13}C labeling, two-dimensional [^{13}C, ^{1}H] nuclear magnetic resonance, and comprehensive isotopomer balancing. Plant Physiol. 136, 3043–3057. doi:10.1104/pp.104.050625

Sriram, G., Fulton, D.B., Shanks, J.V., 2007a. Flux quantification in central carbon metabolism of *Catharanthus roseus* hairy roots by ^{13}C labeling and comprehensive bondomer balancing. Phytochemistry 68, 2243–2257. doi:10.1016/j.phytochem.2007.04.009

Sriram, G., Iyer, V.V., Fulton, D.B., Shanks, J.V., 2007b. Identification of hexose hydrolysis products in metabolic flux analytes: A case study of levulinic acid in plant protein hydrolysate. Metab. Eng. 9, 442–451. doi:10.1016/j.ymben.2007.07.003

Sriram, G., Rahib, L., He, J.-S., Campos, A.E., Parr, L.S., Liao, J.C., Dipple, K.M., 2008. Global metabolic effects of glycerol kinase overexpression in rat hepatoma cells. Mol. Genet. Metab. 93, 145–159. doi:10.1016/j.ymgme.2007.09.008

Sriram, G., Shanks, J.V., 2004. Improvements in metabolic flux analysis using carbon bond labeling experiments: bondomer balancing and Boolean function mapping. Metab. Eng. 6, 116–132. doi:10.1016/j.ymben.2004.02.003

Srour, O., Young, J.D., Eldar, Y.C., 2011. Fluxomers: a new approach for ^{13}C metabolic flux analysis. BMC Syst. Biol. 5, 129. doi:10.1186/1752-0509-5-129

Stanturf, J.A., van Oosten, D.A., Netzer, M.D., Coleman, 2001. Ecology and silviculture of poplar plantations. In: Poplar Culture in North America.D. I. Dickmann, J. G. Isebrands, J. E. Eckenwalder, and J. Richardson eds. NRC Research Press, Ottawa, Ontario, Canada.

Stephanopoulos, G., 1999. Metabolic fluxes and metabolic engineering. Metab. Eng. 1, 1–11. doi:10.1006/mben.1998.0101

Stephanopoulos, G., Vallino, J., 1991. Network rigidity and metabolic engineering in metabolite overproduction. Science 252, 1675–1681. doi:10.1126/science.1904627

Szyperski, T., 1995. Biosynthetically directed fractional ^{13}C-labeling of proteinogenic amino acids. an efficient analytical tool to investigate intermediary metabolism. Eur. J. Biochem. 232, 433–448. doi:10.1111/j.1432-1033.1995.tb20829.x

Tovar-Mendez, A., Miernyk, J.A., Randall, D.D., 2003. Regulation of pyruvate dehydrogenase complex activity in plant cells. Eur. J. Biochem. 270, 1043–1049. doi:10.1046/j.1432-1033.2003.03469.x

Truong, Q., Koch, K., Yoon, J.M., Everard, J.D., Shanks, J.V., 2013. Influence of carbon to nitrogen ratios on soybean somatic embryo (cv. Jack) growth and composition. J. Exp. Bot. 64, 2985–2995. doi:10.1093/jxb/ert138

Tuskan, G.A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R.R., Bhalerao, R.P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapman, J., Chen, G.-L., Cooper, D., Coutinho, P.M., Couturier, J., Covert, S., Cronk, Q., Cunningham, R., Davis, J., Degroeve, S., Déjardin, A., dePamphilis, C., Detter, J., Dirks, B., Dubchak, I., Duplessis, S., Ehlting, J., Ellis, B., Gendler, K., Goodstein, D., Gribskov, M., Grimwood, J., Groover, A., Gunter, L., Hamberger, B., Heinze, B., Helariutta, Y., Henrissat, B., Holligan, D., Holt, R., Huang, W., Islam-Faridi, N., Jones, S., Jones-Rhoades, M., Jorgensen, R., Joshi, C., Kangasjärvi, J., Karlsson, J., Kelleher, C., Kirkpatrick, R., Kirst, M., Kohler, A., Kalluri, U., Larimer, F., Leebens-Mack, J., Leplé, J.-C., Locascio, P., Lou, Y., Lucas, S., Martin, F., Montanini, B., Napoli, C., Nelson, D.R., Nelson, C., Nieminen, K., Nilsson, O., Pereda, V., Peter, G., Philippe, R., Pilate, G., Poliakov, A., Razumovskaya, J., Richardson, P., Rinaldi, C., Ritland, K., Rouzé, P., Ryaboy, D., Schmutz, J., Schrader, J., Segerman, B., Shin, H., Siddiqui, A., Sterky, F., Terry, A., Tsai, C.-J., Uberbacher, E., Unneberg, P., Vahala, J., Wall, K., Wessler, S., Yang, G., Yin, T., Douglas, C., Marra, M., Sandberg, G., Peer, Y.V. de, Rokhsar, D., 2006. The Genome of Black Cottonwood, Populus trichocarpa (Torr. & Gray). Science 313, 1596–1604. doi:10.1126/science.1128691

Van Winden, W.A., Heijnen, J.J., Verheijen, P.J.T., 2002. Cumulative bondomers: A new concept in flux analysis from 2D [^{13}C, ^{1}H] COSY NMR data. Biotechnol. Bioeng. 80, 731–745. doi:10.1002/bit.10429

Varma, A., Palsson, B.O., 1994. Metabolic flux balancing: basic concepts, scientific and practical use. Bio/Technology 12, 994–998. doi:10.1038/nbt1094-994

Wahl, S.A., Nöh, K., Wiechert, W., 2008. ^{13}C labeling experiments at metabolic nonstationary conditions: an exploratory study. BMC Bioinformatics 9, 152. doi:10.1186/1471-2105-9-152

Wasylenko, T.M., Ahn, W.S., Stephanopoulos, G., 2015. The oxidative pentose phosphate pathway is the primary source of NADPH for lipid overproduction from Glucose in *Yarrowia lipolytica*. Metab. Eng. doi:10.1016/j.ymben.2015.02.007

Wasylenko, T.M., Stephanopoulos, G., 2014. Metabolomic and ^{13}C-metabolic flux analysis of a xylose-consuming *Saccharomyces cerevisiae* strain expressing xylose isomerase: xylose metabolic flux analysis. Biotechnol. Bioeng. 112, 3, 470-483. doi:10.1002/bit.25447

Wasylenko, T.M., Stephanopoulos, G., 2013. Kinetic isotope effects significantly influence intracellular metabolite ^{13} C labeling patterns and flux determination. Biotechnol. J. 8, 1080–1089. doi:10.1002/biot.201200276

Weitzel, M., Noh, K., Dalman, T., Niedenfuhr, S., Stute, B., Wiechert, W., 2013. 13CFLUX2--high-performance software suite for ^{13}C-metabolic flux analysis. Bioinformatics 29, 143–145. doi:10.1093/bioinformatics/bts646

Wiechert, W., 2001. ^{13}C metabolic flux analysis. Metab. Eng. 3, 195–206. doi:10.1006/mben.2001.0187

Wiechert, W., Möllney, M., Isermann, N., Wurzel, M., Graaf, A.A. de, 1999. Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems. Biotechnol. Bioeng. 66, 69–85. doi:10.1002/(SICI)1097-0290(1999)66:23.3.CO;2-Y

Wiechert, W., Nöh, K., 2013. Isotopically non-stationary metabolic flux analysis: complex yet highly informative. Curr. Opin. Biotechnol. 24, 979–986. doi:10.1016/j.copbio.2013.03.024

Wiechert, W., Noh, K., 2005. From stationary to instationary metabolic flux analysis. Adv. Biochem. Eng. Biotechnol. 92, 145–172. doi:10.1007/b98921

Wise, R.P., Caldo, R.A., Hong, L., Shen, L., Cannon, E., Dickerson, J.A., 2007. BarleyBase/PLEXdb. Methods Mol. Biol. Clifton NJ 406, 347–363.

Wittmann, C., Heinzle, E., 1999. Mass spectrometry for metabolic flux analysis. Biotechnol. Bioeng. 62, 739. doi:10.1002/(SICI)1097-0290(19990320)62:6<739::AID-BIT13>3.3.CO;2-5

Wu, B., Zhang, B., Feng, X., Rubens, J.R., Huang, R., Hicks, L.M., Pakrasi, H.B., Tang, Y.J., 2009. Alternative isoleucine synthesis pathway in cyanobacterial species. Microbiology 156, 596–602. doi:10.1099/mic.0.031799-0

Wu, C., Xiong, W., Dai, J., Wu, Q., 2015. Genome-based metabolic mapping and ^{13}C flux analysisreveal systematic properties of an oleaginous microalga *Chlorella protothecoides*. Plant Physiol. 167, 586–599. doi:10.1104/pp.114.250688

Wu, G., Truksa, M., Datla, N., Vrinten, P., Bauer, J., Zank, T., Cirpus, P., Heinz, E., Qiu, X., 2005. Stepwise engineering to produce high yields of very long-chain

polyunsaturated fatty acids in plants. Nat. Biotechnol. 23, 1013–1017. doi:10.1038/nbt1107

Young, J.D., Walther, J.L., Antoniewicz, M.R., Yoo, H., Stephanopoulos, G., 2008. An elementary metabolite unit (EMU) based method of isotopically nonstationary flux analysis. Biotechnol. Bioeng. 99, 686–699. doi:10.1002/bit.21632

Zhang, P., Foerster, H., Tissier, C.P., Mueller, L., Paley, S., Karp, P.D., Rhee, S.Y., 2005. MetaCyc and AraCyc. Metabolic pathway databases for plant research. Plant Physiol. 138, 27–37. doi:10.1104/pp.105.060376

Zhila, N.O., Kalacheva, G.S., Volova, T.G., 2005. Influence of nitrogen deficiency on biochemical composition of the green alga Botryococcus. J. Appl. Phycol. 17, 309–315. doi:10.1007/s10811-005-7212-x

Zhu, B., Coleman, G.D., 2001. Phytochrome-mediated photoperiod perception, shoot growth, glutamine, calcium, and protein phosphorylation influence the activity of the poplar bark storage protein gene promoter (bspA). Plant Physiol. 126, 342–351. doi:10.1104/pp.126.1.342

Zupke, C., Stephanopoulos, G., 1994. Modeling of isotope distributions and intracellular fluxes in metabolic networks using atom mapping matrixes. Biotechnol. Prog. 10, 489–498. doi:10.1021/bp00029a006