Flux analysis in central carbon metabolism in plants:
$^{13}$C NMR experiments and analysis

by

Ganesh Sriram

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Chemical Engineering

Program of Study Committee:
Jacqueline V. Shanks, Major Professor
Julie A. Dickerson
Surya K. Mallapragada
Martin H. Spalding
R. Dennis Vigil

Iowa State University
Ames, Iowa
2004

Copyright © Ganesh Sriram, 2004. All rights reserved.
This is to certify that the doctoral dissertation of
Ganesh Sriram
has met the thesis requirements of Iowa State University

Signature was redacted for privacy.

Major Professor

Signature was redacted for privacy.

For the Major Program
Dedicated to
My Parents
# Table of Contents

List of figures viii
List of tables xi
Acknowledgments xii
Abstract xiv

1. Introduction and literature review. Plant metabolic engineering, $^{13}$C metabolic flux analysis
   - Metabolic flux analysis 3
   - Literature minireview: $^{13}$C metabolic flux analysis 4
   - Organization of dissertation 9
   - References 9

2. Improvements in metabolic flux analysis using carbon bond labeling experiments: bondomer balancing and Boolean function mapping 16
   - Abstract 16
   - Introduction 17
   - Methods 21
     - Analytical solutions for bondomer distributions 21
     - Modeling methodology 23
     - Identifiability analysis 33
     - Use of Boolean function mapping to simulate bondomer distributions 34
   - Results 38
     - Flux identifiability 39
   - Discussion 41
     - Analytical solutions for bondomer abundances through bondomer grouping 41
     - Reaction reversibility effects in the glycolysis and pentose phosphate pathways 43
     - Boolean function mapping 44
   - Summary 46
   - Acknowledgments 47
   - References 48
   - Appendix A: Nomenclature 53
     - Metabolites 53
     - Fluxes, reactions 54
   - Figure captions 60
3. Quantification of compartmented metabolic fluxes in developing soybean embryos by employing biosynthetically directed fractional $^{13}$C labeling, 2-D $^{13}$C, $^1$H NMR and comprehensive isotopomer balancing

Abstract 64
Introduction 65
Results 68
2-D NMR $^{13}$C, $^1$H spectra of sink metabolites, cross peak assignments 69
Fine structures of peaks, quantification of isotopomer abundances 70
Labeling states of precursor metabolites by retrobiosynthetic analysis 71
Isotopic equilibration of metabolites between compartments 72
The carbon in plastidic pyruvate originates largely from sucrose 74
Extracellular fluxes, fluxes contributing to biomass accumulation 74
Metabolic network model 75
Metabolic fluxes 76
Discussion 79
Conclusions and future directions 86
Materials and methods 87
Soybean cotyledon culture 87
Protein extraction, hydrolysis, amino acid quantification, and NMR sample preparation 87
Extracellular fluxes and fluxes contributing to biomass 88
NMR spectroscopy 88
Flux evaluation methodology and computer program 89
Acknowledgments 90
Literature cited 90
Figure captions 98
Supplementary material I: Multiplet intensities From 2-D $^{13}$C, $^1$H HSQC 110
Supplementary material II: Biomass composition fluxes 116
Supplementary material III: Accuracy of 2-D $^{13}$C, $^1$H HSQC experiment 118
Supplementary material IV: Activity diagrams, explanations, and mathematical concepts 120
Selection of feasible fluxes in two- (and higher-) dimensional flux parameter space 128
Generation and solution of isotopomer balances 131
Calculation of $\chi^2$ error 136
Statistical analysis of evaluated fluxes 137
Supplementary material V: Effectiveness of simulated annealing in finding global optima 139
2-dimensional trigonometric function 139
33-dimensional flux analysis problem 140
Supplementary material VI: Annotated input files to NMR2Flux 143

4. Metabolic flux analysis of Catharanthus roseus hairy roots by using biosynthetically directed fractional $^{13}$C-labeling and bondomer balancing 144
<table>
<thead>
<tr>
<th>5. Flux identifiability-based optimal design of $^{13}$C labeling experiment for <em>Glycine max</em> (soybean) embryo metabolism</th>
<th>160</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>160</td>
</tr>
<tr>
<td>Introduction</td>
<td>161</td>
</tr>
<tr>
<td>Methods</td>
<td>163</td>
</tr>
<tr>
<td>Results and discussion</td>
<td>165</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>168</td>
</tr>
<tr>
<td>References</td>
<td>168</td>
</tr>
<tr>
<td>Figure captions</td>
<td>169</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>6. Levulinic acid in soybean protein hydrolysate: Identification, 2-D [$^{13}$C, $^1$H] NMR study, and implications for metabolic investigations</th>
<th>181</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>181</td>
</tr>
<tr>
<td>Introduction</td>
<td>182</td>
</tr>
<tr>
<td>Materials and methods</td>
<td>183</td>
</tr>
<tr>
<td>Chemicals</td>
<td>183</td>
</tr>
<tr>
<td>Soybean cotyledon culture and protein extraction</td>
<td>183</td>
</tr>
<tr>
<td>Hydrolysis of protein and sugars</td>
<td>184</td>
</tr>
<tr>
<td>NMR spectroscopy</td>
<td>184</td>
</tr>
<tr>
<td>J-scaled spectra</td>
<td>185</td>
</tr>
<tr>
<td>Results and discussion</td>
<td>185</td>
</tr>
<tr>
<td>Soybean embryo protein hydrolysate contains degradation products of glucose and mannose</td>
<td>185</td>
</tr>
<tr>
<td>Identity of degradation products: Tracing of degradation product peaks to glucose carbon atoms</td>
<td>187</td>
</tr>
<tr>
<td>Identity of degradation products: J-coupling constants</td>
<td>188</td>
</tr>
<tr>
<td>Hexose degradation products are LVA and HMF</td>
<td>188</td>
</tr>
<tr>
<td>Carbon atoms C$_3$ and C$_5$ of LVA acid exhibit long-range coupling</td>
<td>189</td>
</tr>
<tr>
<td>J-scaled [$^{13}$C, $^1$H] spectra of LVA</td>
<td>190</td>
</tr>
</tbody>
</table>
List of figures

Figure 2.1. Glycolytic and pentose phosphate pathways. 61
Figure 2.2. Enumeration of bondomers for case IV, when hxi, tktA, tal, and tktB are reversible. 62
Figure 2.2. Enumeration of bondomers for case IV, when hxi, tktA, tal, and tktB are reversible. 62
Figure 2.3. Effect of reaction reversibility on bondomer abundance. 63
Figure 3.1. Isotopomers ('isotope isomers') of a metabolite. 102
Figure 3.2. 2-D [\(^{13}\)C, \(^{1}\)H] HSQC spectrum of protein hydrolysate. 103
Figure 3.3. Expanded views of [\(^{13}\)C, \(^{1}\)H] HSQC spectrum. 104
Figure 3.4. Comparison of corresponding multiplet intensities of metabolites from cytosol and plastid. 105
Figure 3.6. Metabolic fluxes evaluated for soybean embryos cultured on sucrose (10% w/w U-\(^{13}\)C) and glutamine. 107
Figure 3.7. Comparison of experimental and simulated multiplet intensities. 109
Diagram 1. Overview of flux evaluation procedure. 122
Diagram 2. Overview of NMR2Flux. 123
Diagram 3. SIMULATED_ANNEALING routine (function) in NMR2Flux, and its communication with the functions FEASIBLE_FLUX and CHI_SQUARED. 124
Diagram 4. Calculation of \(\chi^2\). 126
Diagram 5. Calculation of fluxes from flux parameters, and examination of the feasibility of the parameter set. 127
Figure 4.1. Expanded views of [\(^{13}\)C, \(^{1}\)H] HSQC spectrum of *Catharanthus roseus* hairy roots. 156
Figure 4.2. Metabolic flux map template consisting of primary and intermediate metabolic pathways in *Catharanthus roseus* hairy roots growing on (5% w/w U-\(^{13}\)C) sucrose. 157
Figure 4.3. Metabolic fluxes evaluated for *Catharanthus roseus* hairy roots growing on (5% w/w U-\(^{13}\)C) sucrose. 158
Figure 4.4. Comparison of experimental and simulated multiplet intensities. 159
Figure 5.1. Metabolic flux map for the plant system considered in this work (developing *Glycine max* (soybean) embryos cultured on sucrose and glutamine). 173
Figure 5.2. Optimal experiment design for *Glycine max* (soybean) embryos growing on sucrose and glutamine, case 1. 174

Figure 5.3. Optimal experiment design for *Glycine max* (soybean) embryos growing on sucrose and glutamine, case 2. 175

Figure 5.4. Optimal experiment design for *Glycine max* (soybean) embryos growing on sucrose and glutamine, case 3. 176

Figure 5.5. Comparison of statistical qualities and flux standard deviations in reference experiments and optimal experiments in cases 1, 2, 3. 177

Figure 5.6. Optimal experiment design for *Glycine max* (soybean) embryos growing on sucrose and glutamine, case 4. 178

Figure 5.7. Optimal experiment design for *Glycine max* (soybean) embryos growing on sucrose and glutamine, case 5. 179

Figure 5.8. Comparison of standard deviations of fluxes for reference experiments [11% U-\textsuperscript{13}C Suc, 0% U-\textsuperscript{13}C Gln; hollow bars] and optimal experiments [5% U-\textsuperscript{13}C Suc, 96% U-\textsuperscript{13}C Gln; filled bars] in case 5. 180

Figure 6.1(A). Aliphatic region of 2-D [\textsuperscript{13}C, \textsuperscript{1}H] HSQC spectrum of soybean protein hydrolysate. 200

Figure 6.1(B). Aliphatic region of 2-D [\textsuperscript{13}C, \textsuperscript{1}H] HSQC spectrum of U-\textsuperscript{13}C glucose hydrolysate. 201

Figure 6.2. Aliphatic region of 2-D [\textsuperscript{13}C, \textsuperscript{1}H] HSQC spectrum of 3-\textsuperscript{13}C glucose hydrolysate; 3-\textsuperscript{13}C glucose was hydrolyzed under the same conditions employed for protein. 202

Figure 6.3(A). Expanded views of cross peak B (carbon atom C\textsubscript{2} of levulinic acid, corresponding to carbon atom C\textsubscript{3} of glucose) in 2-D [\textsuperscript{13}C, \textsuperscript{1}H] HSQC spectrum of 2-\textsuperscript{13}C glucose hydrolysate. 203

Figure 6.3(B). Expanded views of cross peak B (carbon atom C\textsubscript{2} of levulinic acid, corresponding to carbon atom C\textsubscript{3} of glucose) in 2-D [\textsuperscript{13}C, \textsuperscript{1}H] HSQC spectrum of 4-\textsuperscript{13}C glucose hydrolysate. 204

Figure 6.4. Mechanism for the formation of 5-hydroxymethyl furfural, levulinic acid, and formic acid from glucose. 205

Figure 6.5. Expanded view of atom C\textsubscript{3} of levulinic acid (cross peak D in Figure 6.1) in 2-D [\textsuperscript{13}C, \textsuperscript{1}H] HSQC spectrum of U-\textsuperscript{13}C glucose hydrolysate. 206

Figure 6.6. (A) Expanded view of carbon atom C\textsubscript{2} of levulinic acid in 2-D [\textsuperscript{13}C, \textsuperscript{1}H] HSQC spectrum of soybean protein hydrolysate without J-scaling. 207

Figure 6.6. (B) Expanded view of carbon atom C\textsubscript{2} of levulinic acid in 2-D [\textsuperscript{13}C, \textsuperscript{1}H] HSQC spectrum of soybean protein hydrolysate with J-scaling. 208

Figure 6.7. Comparison of expected and measured multiplet intensities of selected atoms of levulinic acid (LVA) and 5-hydroxymethyl furfural (HMF), prepared
by hydrolyzing an isotopic mixture of glucose (5 mol% U-\textsuperscript{13}C, 95 mol% natural abundance).

**Figure 6.8.** Multiplet intensities of carbon atom C\textsubscript{2} of levulinic acid (LVA) (corresponding to carbon atom C\textsubscript{3} of glucose), evaluated from 2-D \textsuperscript{13}C, \textsuperscript{1}H\textsuperscript{M} HSQC spectra.

**Figure 7.1.** Proteinogenic amino acid proportions in *C. roseus* hairy roots. Hollow bars indicate 21-day-old roots, and filled bars indicate 28-day-old roots.

**Figure 7.2.** 1-D \textsuperscript{13}C NMR spectrum of water extract of 28 d-old *C. roseus* hairy roots grown on U-\textsuperscript{13}C sucrose.

**Figure 7.A-1.** Protein hydrolysis kinetics for bovine serum albumin.

**Figure 7.A-2.** Amino acid analysis of BSA hydrolysate, obtained by the outlined hydrolysis method.
List of tables

Table 2.1. Spreadsheet representation of metabolic reactions for processing by Boolean function mapping. 55

Table 2.2. Bondomer abundances in Case II, when only hxi is irreversible. 56

Table 2.3. Abundances of bondomers and measurable groups of bondomers in Case III, when hxi, tktA and tktB are irreversible. 58

Table I-1. Relative multiplet intensities of carbon nuclei, measured from 2-D HSQC spectrum of protein hydrolysate; protein was isolated from soybean cotyledons. 110

Table 6.1. NMR constants (chemical shifts, $\sigma$ and J-coupling constants, $J_{CC}$) for levulinic acid atoms. 199

Table 7.1. Proportions of biomass constituents in $C.\ roseus$ hairy roots. 222
Acknowledgments

I am very grateful to my research adviser, Dr. Jacqueline V. Shanks, who provided me with the motivation, freedom, facilities, support, and inspiration to enable me to produce this piece of research. She also contributed much toward my professional development. I feel privileged to have been mentored by her, and any amount of thanks will not be adequate for what I have gained from her association. I will miss the wonderful time I had as her graduate student.

I would like to thank Drs. Julie A. Dickerson (Electrical and Computer Engineering), Surya K. Mallapragada (Chemical Engineering), Martin H. Spalding (Genetics, Development, and Cell Biology), and R. Dennis Vigil (Chemical Engineering) for serving on my Program of Study committee. I am very thankful to Drs. Martin Spalding, Mark E. Westgate (Agronomy), Eve S. Wurtele (Genetics, Development, and Cell Biology), Ka-Yiu San (Chemical Engineering and Bioengineering, Rice University) and Ramon Gonzalez (Chemical Engineering) for several valuable suggestions on my work – and for kindly taking out large chunks of time from their busy schedules to discuss my research. I wish to thank Dr. Maria I. Klapa (Chemical Engineering, University of Maryland; whose research on $^{13}$C metabolic flux analysis inspired mine) for motivating my work on many occasions.

The assistance provided by Dr. Bruce Fulton (NMR facility; Biochemistry, Biophysics, and Molecular Biology) in programming the NMR pulse sequences and for training me on the Bruker NMR spectrometer is gratefully acknowledged. I am also thankful to Drs. Amy H. Andreotti and Kristine Brazin (Biochemistry, Biophysics, and Molecular Biology) for supplying some useful $^{13}$C-labeled protein; Dr. Louisa B. Tabatabai (Protein facility; Biochemistry, Biophysics, and Molecular Biology) for providing useful consultation on protein hydrolysis; Drs. Joan M. Peterson and Ruilian Zhou (Agronomy) for performing the soybean embryo culture and training me on protein and starch extraction, and David Gross
(Woodson-Tenent Laboratories, Des Moines, IA) for assistance with fiber analysis. I would like to thank my fellow graduate students and research group members: Murali Subramanian (for training me in plant tissue culture), Vidya Iyer (for the many interesting research discussions we had), Omar González-Rivera (for helping with *Catharanthus roseus* biomass composition and protein hydrolysis), Guy Sander and Chase Beisel (for taking up the baton in maintaining the genetically modified hairy root lines), Andrea Flack and Yingjun Wang (Martin Spalding group, for assistance with protein extraction). Also, thanks are due to my groupmates (past and present): Erik Hughes, Sarah Rollo, Madhuresh Choudhary, Christine Westgate, Erin Denefe, and everyone from the Ramon Gonzalez research group.

I thank Drs. Donna Kienzler (Graduate College) and Brent Shanks (Chemical Engineering) for guiding me during the Prepare Future Faculty (PFF) program. I admired, enjoyed, and highly benefited from the informative mentoring sessions with Dr. Brent Shanks.

I express thanks to United Engineering Foundation (now Engineering Conferences International), the Samuel Roberts Noble Foundation, and American Chemical Society, BIOT Division for fellowships, awards, or grants provided during the period of this research. This research was funded by the National Research Foundation, the Plant Sciences Institute (Iowa State University), and Iowa Soybean Promotion Board.

I wish to thank Kathye Law and DeAnn Pitman (Chemical Engineering office) for all their assistance, and for being such wonderful people to interact with.

I also wish to thank the friends I made in Ames, for making my stay so memorable: Rohini Ramaswamy, Nidhi Bhandari, Shashi Tenneti, Indu Jagannathan, Mamatha Balasubramanian, Sipho Ndlela, Rong Fan, Matt Kipper, Brian Anderson; and Chandrika Pai for widening my musical knowledge and abilities.

I thank my parents for their love and blessings. I owe everything to them.
Abstract

Metabolic flux analysis (MFA) is a fundamental diagnostic tool of metabolic engineering. The use of $^{13}$C labeling followed by NMR (nuclear magnetic resonance) analysis is a powerful method of flux quantification, that is gaining popularity in metabolic engineering. Flux quantification in plants has received very limited attention compared to that in microorganisms and mammalian cells. This is principally because plant biochemistry is more complex, and exhibits extensive subcellular compartmentation. This research concentrated on improvements in $^{13}$C labeling-based flux analysis methods, particularly oriented toward application to plants; and on performing systemic flux analyses in two model plant systems: *Glycine max* (soybean) embryos (agriculturally and nutritionally significant plant system), and *Catharanthus roseus* hairy roots (pharmaceutically significant plant system).

A concept called 'bondomer' was introduced, to simplify flux evaluation from labeling experiments that employ U-$^{13}$C (uniformly $^{13}$C-labeled) substrates. It was shown that the enumeration of feasible bondomers in a metabolic network reduced the number of bondomer balances 3-fold. This resulted in a simplified system that offers the advantages of faster computation, and easier flux identifiability analysis. Furthermore, a computational technique called Boolean function mapping was introduced to aid in the formulation of bondomer or isotopomer balances. This technique is more efficient than previously reported techniques toward the same purpose, and its efficiency is expected to be valuable while analyzing metabolic networks or reaction steps that are not completely known (which is fairly common in plant metabolism).

$^{13}$C labeling experiments were performed by feeding a mixture of U-$^{13}$C sucrose, naturally abundant sucrose, and glutamine to developing soybean (*Glycine max*) embryos. 2-D [$^{13}$C, $^1$H] NMR spectra of seed storage protein and starch hydrolysates were acquired, and yielded a labeling data set consisting of 155 $^{13}$C isotopomer abundances. A computer program,
NMR2Flux, was developed to automatically calculate fluxes from this data. This program accepts a user-defined metabolic network model, and incorporates recent mathematical advances toward accurate and efficient flux evaluation. Fluxes were calculated and statistical analysis was performed to obtain standard deviations. A high flux was found through the oxidative pentose phosphate pathway (19.99 ± 4.39 µmol d\(^{-1}\) cotyledon\(^{-1}\), or 104.2 carbon mol ± 23.0 carbon mol per 100 carbon mol of sucrose uptake). Separate transketolase and transaldolase fluxes could be distinguished in the plastid and the cytosol, and those in the plastid were found to be at least 6-fold higher. The backflux from triose to hexose phosphate was also found to be substantial in the plastid (21.72 ± 5.00 µmol d\(^{-1}\) cotyledon\(^{-1}\), or 113.2 carbon mol ± 26.0 carbon mol per 100 carbon mol of sucrose uptake). Forward and backward directions of anaplerotic fluxes could be distinguished. The glyoxylate shunt flux was found to be negligible.

In \textit{C. roseus} hairy roots, all fluxes primary/intermediary metabolic pathways such as glycolysis, pentose phosphate, TCA cycle, and the anaplerotic and glyoxylate shunts (except the GABA shunt) were well-identified. A rather high cycling was observed between the anaplerotic fluxes PEP \(\rightarrow\) OAA (cytosol) and Mal \(\rightarrow\) Pyr (mitochondrion).

Investigations on flux identifiability were carried out for the soybean embryo system. Using these, optimal labeling experiments were designed, that utilize judicious combinations of labeled varieties of two substrates (sucrose and glutamine), to maximize the statistical quality of the evaluated fluxes. Combinations of U-\(^{13}\)C sucrose and 1-\(^{13}\)C sucrose were examined for their effect on statistical quality of the glycolysis and pentose phosphate pathway flux information provided by them. Also, combinations of U-\(^{13}\)C sucrose and U-\(^{13}\)C glutamine were examined for their effect on flux identifiability.

In the 2-D \(^{13}\)C, \(^1\)H spectra of protein isolated from soybean embryos, four intense peaks were observed that could not be assigned to the any proteinogenic amino acids. These peaks
were shown to correspond to degradation products of glucose and mannose, and it was inferred that these arise from the acid degradation of the large amounts of glycosylating sugars (glucosamine and mannose) associated with soybean embryo protein. Further, the degradation products were identified as levulinic acid (4-oxo-pentanoic acid, major product) and 5-hydroxymethyl furfural (minor product). A 2-D NMR study was conducted on them, and it was shown that the metabolic information in the degradation products can be used toward metabolic flux or pathway analysis.

In addition, the elemental make-up and composition of the biomass of *Catharanthus roseus* hairy roots is reported. This information is crucial toward $^{13}$C metabolic flux analysis. The observed elemental composition was expressible by the formula C$_{12.0}$H$_{22.7}$N$_{0.4}$O$_{7.1}$. The most significant components in the hairy roots were soluble sugars (38.0% ± 1.0%) and lignin (15.4% ± 2.5%). We were able to account for 89.2% (± 9.7%) of the biomass.

The generic flux analysis tools developed in this work are expected to serve as a quantitative tool for metabolic studies and phenotype comparisons, and can be extended to other plant systems. Flux data, in synergy with transcript, protein and metabolite profiles, can contribute to a systemic understanding of plant metabolism and lead to the development of models for the same.
1. Introduction and literature review. Plant metabolic engineering, $^{13}$C metabolic flux analysis

Metabolic engineering is the science concerned with the directed modification of cellular metabolism in order to derive new or improved products from an organism. This field emerged about a decade ago (Bailey, 1991; Stephanopoulos and Vallino, 1991) as a distinct branch of genetic engineering (Lee and Papoutsakis, 1999).

However, metabolic engineering is distinct from genetic engineering, in that it lays emphasis on the engineering aspect. Akin to any engineering endeavor, metabolic engineering is comprised of separate synthesis and analysis components. The synthesis component involves insertion, deletion, and/or modification of metabolic pathways using recombinant DNA and other molecular biological technologies. The analysis component engages tools such as metabolic flux analysis (MFA) and metabolic control analysis (MCA), either to analyze the metabolic impact of pathway modifications, or to speculate or suggest new ones (Stephanopoulos, 2002).

Plants are profitable metabolic engineering targets because they produce an overwhelming assortment of metabolites (Morgan et al., 1999) with immense food, feed, and pharmaceutical applications. Indeed, research on the metabolic engineering of plants has been rapidly gaining pace in recent years, as exemplified by the nearly exponential increase in the number of publications per year on this subject (Hanson and Shanks, 2002).

Genetically modified plant products made an appearance in American and European markets in the mid-1990s, and ever since, numerous products derived from transgenic plants are being tested in food, feed, and industrial markets (Kleter et al., 2001; Mazur et al., 1999). In addition to their immense natural collection of biosynthetic pathways, plants also offer a major advantage as a metabolic engineering target: they can be grown in the field in huge quantities (Hanson and Shanks, 2002). As a matter of fact, genetically modified plants are
being cultivated over a land area of over 28 million hectares worldwide (Abelson and Hines, 1999). Not surprisingly therefore, it has been forecast that the potential of plant metabolic engineering is larger than that of all other types of metabolic engineering combined (Hanson and Shanks, 2002).

This research is concerned with the analysis aspect of metabolic engineering applied to plants. Its objectives were the development of tools for comprehensive metabolic flux analysis of two model plant systems: *Glycine max* (soybean) embryos and *Catharanthus roseus* hairy roots. *Glycine max* (soybean) embryos are of high nutritional relevance, as they are important sources of protein and oil. Therefore, the investigation of metabolic flux distribution in the primary metabolic pathways in them can yield information about possible metabolic engineering targets for amplifying protein or oil production. *Catharanthus roseus* is a valuable source of terpenoid indole alkaloids, some of which possess important therapeutic properties, e.g. vindoline and vinblastine (anti-neoplastic); serpentine and ajmalicine (anti-hypertension) (Bhadra and Shanks, 1997; Rijhwani and Shanks, 1998; Morgan et al., 1998; Morgan and Shanks, 1999). Hairy root cultures of the plant – generated through *Agrobacterium* infection – are biochemically and genetically stable model systems (Shanks and Bhadra, 1997). This makes the indole alkaloid production pathway in *C. roseus* an attractive metabolic engineering target. Recently, the J. Shanks group and their collaborators have recently engineered several hairy root lines each designed to overexpress one or more different enzyme(s) in this pathway (Hughes et al., 2003a; b). Furthermore, since *C. roseus* is an extensively studied plant system, the metabolic engineering of this plant around the tryptophan production pathway may help develop model plants with higher levels of tryptophan, an essential amino acid. Thus, metabolic engineering of *C. roseus* hairy roots is consequential, both for pharmaceutical and agricultural reasons. The development of a metabolic flux analysis tool for this system is expected to play a crucial role in this process.
Metabolic flux analysis

Metabolic flux analysis (MFA) has acquired a central place in metabolic engineering as a powerful diagnostic tool. It is important to note that fluxes, and not metabolite concentrations, are the variables which are useful for assessing the impact of genetic modifications. Fluxes describe what cells ‘do’, rather than how they ‘look’ (Stephanopoulos, 2002), and therefore reflect the ultimate effect of a genetic or enzymatic modification.

The objective of MFA is the quantification of all steady state intracellular fluxes in a metabolic network (Stephanopoulos and Vallino, 1991; Stephanopoulos, 1994; 1999; Wiechert, 2001). Metabolic flux maps, resulting from these analyses, provide a measure of the extent of engagement of various pathways in cellular metabolism (Stephanopoulos et al., 1998) and thus provide a detailed quantitative description of the metabolic state of the cell. More importantly, they provide a means to compare genetic variants, thus enabling the assessment of the metabolic impact of a genetic modification, and the suggestion of further modifications (Nielsen, 1998).

The most basic approach to metabolic flux analysis is stoichiometric MFA (Edwards et al., 1999; Varma and Palsson, 1994). This is performed by writing balances for intracellular metabolites based on the stoichiometry of the metabolic reactions, which results in a system of linear equations. In solving this system of equations, the rates of change in concentration of extracellular metabolites (i.e. substrates, products) are employed in order to resolve some or all of the degrees of freedom (Stephanopoulos et al., 1998). Stoichiometric MFA is mathematically simple, and the laboratory analysis involved can be performed rather inexpensively.

However, a number of shortfalls of such a pure stoichiometric approach have been recognized (Wiechert, 2001). For instance, in large, topologically complex or compartmented metabolic networks (such as those of many eukaryotes), there is a lack of measurable
extracellular metabolites that are related to the intracellular flux distribution (Schmidt et al., 1999). This is caused by the presence of parallel metabolic pathways, metabolic cycles and reversible or bidirectional reactions (Wiechert, 2001). These limitations can be overcome by supplementing stoichiometric MFA with data from $^{13}$C labeling experiments. This approach to MFA, called $^{13}$C metabolic flux analysis, involves the feeding of isotopically labeled ($^{13}$C) substrates, and subsequent analysis of the resulting labeled metabolites by nuclear magnetic resonance (NMR) spectroscopy or mass spectrometry (MS). The flux ratios at branchpoints in the network are reflected in the labeling pattern of the metabolites. If the flux ratios can be evaluated from the labeling pattern, they can be used to supplement the metabolite balances in stoichiometric MFA, thus compensating for the lack of measurements therein (Wiechert, 2001). While $^{13}$C MFA alleviates the drawbacks of stoichiometric MFA, it is much more expensive (owing to the cost of $^{13}$C substrates) and is also mathematically complex. A minireview on $^{13}$C MFA follows in the next section.

**Literature minireview: $^{13}$C metabolic flux analysis**

Stoichiometric MFA alone may not suffice to unambiguously evaluate all fluxes in a metabolic network, especially when the network contains parallel pathways, metabolic cycles, bidirectional (reversible) reactions, or split pathways where cofactors are not balanced. In such cases, $^{13}$C MFA – a label-based technique that makes possible the indirect measurement of internal fluxes in a cell – mitigates the situation.

Label-based MFA techniques are based upon the feeding of a label ($^{13}$C or $^{14}$C) to a biological system. These are unusual isotopes of carbon, and will be referred to as labels in the rest of this dissertation. A labeled carbon atom should be understood to be $^{13}$C, while an unlabeled carbon atom, to be $^{12}$C. In $^{13}$C MFA, the labeled substrate fed to the biological system is diluted with unlabeled substrate. Sometimes, a single substrate in which certain
carbon atoms are labeled (and the others unlabeled), for example 1-\textsuperscript{13}C glucose, may serve as a combination of labeled and unlabeled substrates. The main components of the ‘design’ of a carbon labeling experiment (CLE) are selection of the type of labeled substrate and the relative extents of labeled and unlabeled substrate in the feed (Stephanopoulos \textit{et al}., 1998).

The cells are then cultured at a metabolic and isotopic steady state. Absolute steady state is achieved in a chemostat or a carefully controlled and monitored batch experiment (Wiechert, 2001). During the CLE, the cells manufacture metabolites from both the labeled and unlabeled metabolites. Depending on the topology of the metabolic network and the values of the metabolic fluxes therein, the metabolite atoms will exhibit different labeling patterns. These labeling patterns can be measured through \textsuperscript{1}H, \textsuperscript{13}C, or [\textsuperscript{13}C, \textsuperscript{1}H] nuclear magnetic resonance (NMR) or gas chromatography-mass spectrometry (GC-MS or simply MS) analysis, and the values of the fluxes can be back-calculated from them.

Label-based MFA was initiated with \textsuperscript{14}C-labeling experiments in the early 1980s. For example, Blum and Stein (1982) have applied such a technique to \textit{Tetrahymena} cells. A concept called ‘isotopomer’ was introduced in the late 1980s, mainly through pioneering work by Malloy \textit{et al}. (1988). This work was applied to the citric acid (TCA) cycle in the heart, and the measurements were made using \textsuperscript{13}C NMR. At that time, it was only possible to obtain NMR data for one metabolite, \textit{viz}. glutamate. This limited the amount of analysis in this work. The first application of MS to MFA was by Katz \textit{et al}. (1993). Again, the number of measured (labeled) metabolites was limited. Zupke and Stephanopoulos (1994) then showed that \textsuperscript{13}C enrichments on carbon atoms could be measured using \textsuperscript{1}H NMR. They also introduced the elegant concept of atom mapping matrices (AMMs) which could be used to iteratively simulate \textsuperscript{13}C enrichments of metabolites in any general CLE. The first extensive labeling data set with more than 25 \textsuperscript{1}H NMR measurements was reported by Marx \textit{et al}. (1996). The well-timed papers by Wiechert \textit{et al}. (1997a; b) presented an elaborate treatment
of flux evaluation from enrichment measurements obtained from $^1$H NMR experiments. They also extensively discussed statistical analysis of the evaluated fluxes. Follstad and Stephanopoulos (1998) analyzed the effects of reversible reactions on the distribution of the $^{13}$C label, using the AMM technique.

Around the same time, the concept of isotopomer [introduced by Malloy et al. (1988)] was gaining increasing recognition in the CLE community. The term isotopomer is coined from isotope and isomer, and isotopomers refer to isomers of a molecule that differ in the labeling state ($^{12}$C and $^{13}$C) of their individual carbon atoms. It is easy to infer that a metabolite with \( n \) carbon atoms will have \( 2^n \) isotopomers. Isotopomer distributions of a metabolite are represented by vectors called isotopomer distribution vectors (IDVs) which contain the fractions of each isotopomer of the metabolite. While the erstwhile $^1$H NMR-based experiments only allowed the measurement of \( n \) enrichments per metabolite molecule, the measurement of isotopomers (through judiciously planned $^{13}$C NMR experiments) could provide up to \( 2^n \) measurements per metabolite, thus enormously increasing the amount of raw data to evaluate fluxes from. Schmidt et al. (1997) extended the prevalent concept of AMMs to isotopomer-mapping matrices, or IMMs. This made possible the iterative simulation of all isotopomers in a metabolic network. Schmidt et al. introduced a SIMULINK module that performed isotopomer balances to accomplish this. Klapa et al. (1999) illustrated the technique of isotopomer enumeration and balancing, wherein only physically realizable isotopomers were balanced. This technique is advantageous in experiments using specifically labeled substrates (e.g. 1-$^{13}$C glucose or 2-$^{13}$C pyruvate) in large percentages (90% and above), so that the natural abundance of $^{13}$C becomes negligible in the analysis. They applied this technique to the TCA cycle in mammalian cells, and developed analytical formulas that could be used to evaluate fluxes from NMR data, and to detect the presence or absence of the
glyoxylate shunt. A sequel paper discussed practical applications of this work (Park et al., 1999).

The applications of MFA to compare different phenotypes was on the rise in the late 1990s and early 2000s, e.g. the application by de Graaf et al. (1999) for Zymomonas mobilis. Wiechert et al. (1999) then introduced the concept of cumomer (cumulative isotopomer), which could be exploited to obtain analytical solutions to isotopomer balances, thus providing relief from the earlier iterative method which was computationally expensive, particularly for systems involving near-equilibrium reversible reactions. In a sequel paper, Möllney et al. (1999) analyzed the identifiability of fluxes from carbon labeling experiments, and detailed procedures for optimal design of CLEs, taking into account various errors that arise in the raw labeling data. These procedures were incorporated into a CLE evaluation software released in 2001 (Wiechert et al., 2001). Petersen et al. (2001) used this software along with experimental design, and were able to successfully demonstrate a significant futile cycle in Corynebacterium glutamicum. van Winden et al. (2001) reported experimental developments such as measurement of long-range $^{13}$C-$^{13}$C couplings and formulas to enable deconvolution of NMR spectra obtained in carbon labeling experiments.

The research efforts mentioned above all focused on specifically labeled substrates, wherein only certain carbon atoms were labeled. These could only be used to perform MFA on portions of metabolic networks. They also required large percentages of the labeled substrate in the feed (close to 100%) to obtain meaningful data. Meanwhile, another type of experiment, employing a U-$^{13}$C substrate (uniformly labeled substrate, which is fully $^{13}$C-labeled at all carbon atom positions) diluted with unlabeled substrate, was becoming popular. These experiments afforded a holistic evaluation of fluxes throughout a metabolic network, and required relatively small percentages of the labeled substrate (5–10%). Such an experiment was first introduced by Szyperski (1995; 1998) under the name ‘carbon Bond
Labeling Experiment’ (BLE). The mathematical analysis of these experiments began with semi-quantitative analysis by Szyperski (1995), to local metabolic flux analysis (MFA around a node, called MetaFoR) by Szyperski et al. (1999) and Sauer et al. (1999). Global flux analysis in such experiments using iterative solutions to isotopomer balances, was introduced by Schmidt et al. (1999). Both these global approaches used the isotopomer concept, which proves to be computationally expensive for modeling BLEs. A newly introduced concept, called ‘bondomer’, is better suited for processing of BLE data.

The concept of bondomers was introduced in this work (Sriram and Shanks, 2001; 2004; Chapter 2) and by van Winden et al. (2002). We have introduced bondomer enumeration and balancing, while van Winden et al. (2002) have introduced the concept of cumulative bondomer balancing to enable numerical solutions to BLEs. They have also shown that for BLEs, bondomer balancing is computationally economical as compared to isotopomer balancing. BLEs have been envisioned to be a promising strategy for evaluating metabolic fluxes in eukaryotic systems such as plants (Shachar-Hill, 2002; Roscher et al., 2000), and analyses of a qualitative nature have been reported recently (Glawischnig et al., 2002; Glawischnig et al., 2000, 2001; Bacher et al., 1999).

In this project, we developed $^{13}$C-based approaches to perform MFA on the two plant systems: *Glycine max* embryos and *Catharanthus roseus* hairy roots. $^{13}$C MFA was performed on the plant systems by employing isotopomer balancing (*Glycine max*) or bondomer balancing (*Catharanthus roseus*). The remainder of this dissertation consists of descriptions of the theoretical improvements contributed, the development of the labeling experiments, their mathematical treatment, the results obtained from them, and supplementary projects undertaken to support them.
Organization of dissertation

This chapter presented a literature minireview of $^{13}$C metabolic flux analysis. The rest of the dissertation is organized as described below.

Certain mathematical innovations were introduced in this research to efficiently model carbon bond labeling experiments and convert the data obtained from them into metabolic fluxes. These are the terms 'bond integrity' and 'bondomer', and a numerical solution method for bondomer balancing, 'Boolean function mapping'. These are described in Chapter 2.

Chapter 3 describes the $^{13}$C MFA performed on soybean embryos, and the details of the isotopomer balancing-based flux evaluation. Chapter 4 describes the $^{13}$C MFA performed on C. roseus hairy roots, where bondomer balancing-based flux evaluation (described in Chapter 2) was used to evaluate the fluxes.

Chapter 5 describes studies on flux identifiability and optimal experimental design for the soybean embryo system. Chapter 6 describes the identification and NMR study of hexose hydrolysis products (levulinic acid and 5-hydroxymethyl furfural) discovered in soybean protein hydrolysate, and suggestions for their use in metabolic flux analysis. Chapter 7 describes the determination of the biomass constituents of Catharanthus roseus hairy roots, an important requirement for MFA.

Chapter 8 summarizes the conclusions of this work and outlines future scope.

References


2. Improvements in metabolic flux analysis using carbon bond labeling experiments: bondomer balancing and Boolean function mapping


Ganesh Sriram§ and Jacqueline V. Shanks*

Department of Chemical Engineering, Iowa State University, Ames, IA 50011

Abstract

The biosynthetically directed fractional $^{13}$C labeling method for metabolic flux evaluation relies on performing a 2-D [$^{13}$C, $^1$H] NMR experiment on extracts from organisms cultured on a uniformly labeled carbon substrate. This article focuses on improvements in the interpretation of data obtained from such an experiment by employing the concept of bondomers. Bondomers take into account the natural abundance of $^{13}$C; therefore many bondomers in a real network are zero, and can be precluded a priori – thus resulting in fewer balances. Using this method, we obtained a set of linear equations which can be solved to obtain analytical formulas for NMR-measurable quantities in terms of fluxes in glycolysis and the pentose phosphate pathways. For a specific case of this network with four degrees of freedom, a priori identifiability of the fluxes was shown possible for any set of fluxes. For a more general case with five degrees of freedom, the fluxes were shown identifiable for a representative set of fluxes. Minimal sets of measurements which best identify the fluxes are

---

§Primary researcher and author.

*Principal investigator and corresponding author.

Copyright © Elsevier Inc. Reproduced with permission.
listed. Furthermore, we have delineated Boolean function mapping, a new method to iteratively simulate bondomer abundances or efficiently convert carbon skeleton rearrangement information to mapping matrices. The efficiency of this method is expected to be valuable while analyzing metabolic networks which are not completely known (such as in plant metabolism) or while implementing iterative bondomer balancing methods.

**Introduction**

Metabolic flux analysis (MFA) has acquired a central place in metabolic engineering as a powerful diagnostic tool. The objective of MFA is the quantification of all steady state intracellular fluxes in a metabolic network (Stephanopoulos and Vallino, 1991; Stephanopoulos, 1994; 1999; Wiechert, 2001). Metabolic flux maps resulting from these analyses provide a measure of the extent of contribution of various pathways in cellular metabolism (Stephanopoulos et al., 1998) and thus provide a detailed quantitative description of the metabolic state of the cell. More importantly, they provide a means to compare genetic variants, thus enabling the assessment of the metabolic impact of a genetic modification, and enabling the suggestion of further modifications (Bailey, 1990; Nielsen, 1998).

The most basic approach to metabolic flux analysis is stoichiometric MFA (Edwards et al., 1999; Varma and Palsson, 1994). This simple method, however, is known to have a number of pitfalls caused by the presence of parallel metabolic pathways, metabolic cycles and reversible or bidirectional reactions (Wiechert et al., 2001). Its limitations can be overcome by supplementing stoichiometric MFA with data from $^{13}$C labeling experiments, where the flux ratios at branchpoints in the network are reflected in the $^{13}$C labeling pattern of the metabolites, thus providing additional constraints to the stoichiometric equations and compensating for the lack of measurements therein (Stephanopoulos, 2002; Wiechert, 2001).
A large number of $^{13}$C labeling experiments reported in the literature have focused on feeding a specifically labeled carbon substrate to the organism of interest, and quantifying $^{13}$C enrichments or isotopomer distributions in metabolite carbon atoms. Recently, a different type carbon labeling experiment introduced by Szyperski (1995) has become popular. This method involves feeding the organism with a mixture of a uniformly $^{13}$C-labeled (U-$^{13}$C) substrate and a naturally abundant version of the same substrate. In such an experiment, the metabolic flux information is contained in the extent of scalar coupling between adjacent carbon atoms, measurable through 2-D $[^{13}$C, $^1$H] HSQC or COSY NMR experiments (Szyperski, 1995; 1998). Such a $^{13}$C labeling experiment is referred to either as biosynthetically directed fractional $^{13}$C labeling or carbon-carbon bond labeling experiment (Szyperski, 1995). Throughout this article, the term ‘bond labeling experiment’ (BLE) will be used to refer to this experiment. BLEs have been envisioned to be a promising strategy for evaluating metabolic fluxes in eukaryotic systems such as plants (Shachar-Hill, 2002; Roscher et al., 2000) and analyses of a qualitative nature have been reported recently (Glawischnig et al., 2002; Glawischnig et al., 2001, 2000; Bacher et al., 1999). This motivates research into methods for fast and efficient analysis of NMR data from BLEs.

Papers describing the conversion of NMR data from a carbon labeling experiment to metabolic fluxes have been based on the central concept of isotopomer balancing. Both analytical approaches (Klapa et al., 1999; Park et al., 1999; Rontein et al., 2002) and numerical approaches (Zupke and Stephanopoulos, 1994; Schmidt et al., 1997, 1999; Wiechert et al., 1997a, b, 1999; Möllney et al., 1999, Dauner et al., 2002) have been treated in detail. A generic software for flux evaluation based on the concept of isotopomer balancing, is also available (Wiechert, 2001). While the isotopomer concept is highly suited for modeling carbon labeling experiments involving a specifically labeled substrate or substrates, BLEs are better modeled using a newly introduced concept, called ‘bondomer’.
So far, only three groups have reported metabolic flux analysis from BLE data by using the bondomer (or similar) concept, instead of the isotopomer concept. An approach aiming at local metabolic flux ratio analysis, called MetaFoR, was introduced by Szyperski and co-workers (Szyperski et al., 1999). This approach relates fluxes with the abundances of intact carbon fragments in metabolites (calculable from 2-D $^{13}$C, $^1$H NMR data). MetaFoR is suitable only for local flux analysis around a node, but not for global flux analysis. The bondomer concept (Sriram and Shanks, 2001; van Winden et al., 2002) is a developed version of the intact fragment concept used in MetaFoR, and rigorous bondomer balancing, analogous to isotopomer balancing, can accomplish global flux analysis. In Sriram and Shanks (2001), bondomer enumeration and balancing have been introduced on an illustrative metabolic network (glycolysis and pentose phosphate pathway with irreversible reactions only). In van Winden et al. (2002) the method of cumulative bondomer balancing [analogous to cumomer (cumulative isotopomer) balancing] was introduced. More importantly it was shown that for BLEs, bondomer balancing is computationally economical as compared to isotopomer balancing (van Winden et al., 2002).

In this work we present an improved analytical solution technique and a new numerical technique toward the processing of BLE data using the bondomer concept. In writing bondomer balance equations, we have made use of an interesting property of bondomers— that natural abundance of $^{13}$C (ca. 1%) is automatically factored into the definition of a bondomer. Therefore, not every single bondomer of each metabolite is formed in a realistic metabolic network. Such non-realized bondomers can be precluded from the analysis a priori, resulting in substantially fewer bondomer balances than if all cumulative bondomers would be balanced. This approach is useful in that it results in lower computational time with respect to isotopomer/cumomer balancing or the balancing of all cumulative bondomers. This approach also simplifies the system state so that partial differentiation of the bondomers with
respect to the fluxes (and consequent calculation of sensitivities) is simplified. Such a simplification is not possible for the case of isotoperomers or if all cumulative bondomers are balanced.

This new approach is exemplified using glycolysis and the pentose phosphate pathway as an example. Next, we have used the analytical formulas obtained to draw conclusions about identifiability of fluxes from the NMR data. For a specific case of the glycolysis/pentose phosphate pathway network with four degrees of freedom, we have shown that a priori identifiability of fluxes from the measured data is possible analytically. For a more general case with five degrees of freedom, we have listed minimal sets of bondomer abundances which can be used to evaluate all fluxes and reaction reversibilities in the network.

The advantage of deriving explicit, analytical formulas is that they are straightforward and can be applied directly to flux determination, especially by users unfamiliar with the mathematical formalism of numerical models for carbon labeling experiments (Wiechert, 2001). Moreover, analytical solutions, being explicit, can provide a more intuitive interpretation of the data from a labeling experiment, which is typically lost in the complexity of an implicit numerical model (Dauner et al., 2000). In addition, it has been recognized that analytical formulas provide a time-conserving method to calculate sensitivities of the NMR multiplet intensities to fluxes, when such calculations are required in iterative optimization routines of numerical methods (Wiechert, 2001). Although elementary analytical formulae for BLEs have been reported before (Szoperski, 1995; Sauer et al., 1997), we have employed a more rigorous and complete enumeration and balancing of bondomers in arriving at the analytical formulae. The analytical solution part of this work employs an approach analogous to that of Klapa et al. (1999), who employed isotoperomer enumeration and balancing for a different pathway – the TCA cycle – to model carbon labeling experiments involving specifically labeled substrates. We have also accounted for
the effect of reversible reactions in elaborate detail. This makes our model more realistic, since reversible reactions are more a rule than an exception in biochemical systems.

In the second part of this work, we delineate Boolean function mapping, hitherto unreported in the evaluation of fluxes from BLEs and from carbon labeling experiments in general. Boolean function mapping converts a reactant bondomer to a product bondomer through a sequence of Boolean or arithmetic operations. It can be used to convert carbon skeletal rearrangements to bondomer mapping matrices or to iteratively calculate bondomer abundances on a need basis. This is shown to be an efficient algorithm. Its efficiency is realized in the case of unknown metabolic networks (quite common in plants) where numerous a posteriori modifications of an initially assumed network may be required before arriving at a network that matches the experimental data well. Boolean function mapping can also be incorporated in flux evaluation software based on iterative bondomer balancing, leading to an improvement in simulation time.

**Methods**

**Analytical solutions for bondomer distributions**

**Definitions**

Before we proceed to describe the development of the model, we recapitulate the definitions of the terms ‘bond integrity’ and ‘bondomer’. These definitions conform to those used by Sriram and Shanks (2001), and van Winden (2002). In this article, we also introduce a numbering scheme for bondomers, which will be of use in the Boolean function mapping technique to be described later.

**Bond integrity.** The bond integrity, $b$, of a covalent bond between two carbon atoms in a metabolite molecule is a property indicating if those two carbon atoms originated from the
same substrate molecule or not. For example, consider an experiment where glucose is the sole substrate and a 2-carbon metabolite (1 2) is formed. Then, if the two carbons 1 and 2 originated from the same substrate, the bond integrity of the (1-2) carbon-carbon bond is defined to be 1, and the bond is called an ‘intact’ or ‘unbroken’ bond. If the carbons 1 and 2 originated from separate substrate molecules and were biosynthetically reassembled together, the bond integrity of the bond between them is defined zero, and the bond is called a ‘biosynthetic’ or ‘broken’ bond. Thus for a single molecule, \( b_i \) takes the binary values 1 and 0. In this paper we have used the notation ‘1-2’ to denote an intact bond and ‘1 2’ to denote a biosynthetic bond.

**Bondomer.** Bondomers (*bond-isomers*) are molecules of the same metabolite, whose corresponding carbon-carbon bonds do not have the same bond integrities. For example, assuming glucose to be the substrate the G6P (glucose-6-phosphate) bondomer 1-2-3-4-5-6 represents a molecule formed from an intact molecule of glucose while the following bondomers: 1-2-3 4-5-6 (four unbroken bonds, one biosynthetic bond) and 1-2 3 4-5-6 (two broken bonds three unbroken bonds) represent molecules emanating from multiple source molecules of glucose. In this paper, the term ‘intact’ or ‘unbroken’ used for a bondomer denotes that all bonds in that bondomer are intact (i.e. \( b_i = 1 \) for all \( i \)) (such as 1-2-3-4-5-6 above).

**Bondomer numbering.** The number of bondomers for a given metabolite may sometimes be quite large, and it is therefore necessary to systematically assign numbers to them. We have used the following procedure: For a metabolite M having an \( n \)-carbon skeleton, \( 12345...n \), we number the C–C bonds in the molecule from left to right, and let \( b_i \) be the integrity of the \( i \)th bond. Thus,

\[
b_1 \ b_2 \ b_3 \ b_4 \ b_5 \ ... \ b_{n-1} \ n
\]
The bondomer number of this bondomer is then calculated as

\[ k = \sum_{i=1}^{n-1} 2^{i-1} b_i + 1 \]  

(1)

which is the decimal representation of the binary number given by the bond integrities, reading from right to left, plus 1. This number is always an integer, since in a single molecule the individual bond integrities are either 0 or 1. The abundance of this bondomer pool is then denoted by \( M_k \). It should be noted that the ‘1’ is added to the bondomer number solely for convenience, so that the lowest numbered bondomer is \( M_1 \).

To convert raw NMR data to bondomer abundances, expressions derived by Szyperski (1995) and extended by van Winden et al. (2002) may be used. It should be noted that \( P_n \), the probability of \(^{13}\)C labeling from natural abundance, is included in these formulae. Therefore natural abundance need not be separately considered in the flux analysis. Another interpretation of this statement is as follows: when a U-\(^{13}\)C substrate (e.g. glucose) is fed to an organism, the input bondomer mixture contains a single bondomer (1-2-3-4-5-6) whose relative abundance is 1.0, while the abundances of all other bondomers in the input bondomer vector are exactly zero. This follows from the above definition of a bondomer. (On the contrary, the input isotopomer mixture contains all 64 isotopomers of glucose at non-zero abundances.) It follows, as shown below, that many bondomers of metabolites in the network also have zero abundances, and need not be considered in the bondomer balances. This results in a simpler system state with fewer bondomer balances.

**Modeling methodology**

The modeling methodology employed in this paper is as follows: we (1) elucidate all the key biochemical reactions in the relevant pathways and then write metabolite balances for each metabolite, (2) recognize the carbon skeleton rearrangements in the biochemical reactions of interest, (3) enumerate all the bondomers of various metabolites formed in the reaction
network; and prepare an exhaustive list of them, (4) write balance equations for the bondomers, group bondomers appropriately to convert those balances to linear equations, and solve these to obtain analytical solutions for the bondomer abundances. These steps have been delineated below.

As mentioned above, the abundances of many bondomers in a realistic network could be zero. The bondomer enumeration [step (3) above] generates a list of non-zero bondomers. All bondomers which do not appear in this list can be eliminated from the bondomer balances a priori. Here it should be mentioned that though the technique of bondomer enumeration per se is not new, its use to eliminate zero bondomers and reduce the number of balance equations (as done in this paper) is new.

**Pathways and metabolic fluxes**

In this paper, we focus on glycolysis and pentose phosphate pathway (PPP), as mentioned in ‘Introduction’. These pathways are ubiquitous in all organisms, and are shown in their general form in Figure 2.1. The description of these pathways can be found in standard biochemistry texts. It may be noted that some of the reactions in the glycolytic and pentose phosphate pathways have been lumped. However the lumping does not affect the results, since all reactions that have been neglected in the process of lumping do not involve a carbon skeleton rearrangement, and are therefore redundant in this analysis. An assumption in this formulation is that downstream metabolism [e.g. the tricarboxylic acid (TCA) cycle or anaplerotic reactions] has negligible effect on this network, and that phosphoenolpyruvate is formed only from glycolysis and PPP. This assumption is valid for many simple systems (e.g. *E. coli* growing on glucose) and was also verified for two model plant systems (Sriram and Shanks, unpublished data).

To enumerate the bondomers formed in the pathways under consideration, it is necessary to know the carbon skeleton rearrangements involved in the respective reactions. This
information for the glycolytic and pentose phosphate pathways is available in the literature [for example, see Follstad and Stephanopoulos (1998)], and is not dealt with here.

Figure 2.1 also shows the flux distribution in this metabolic network. All fluxes are relative to the input glucose flux, which has been taken to be 1. The flux into the pentose phosphate pathway has been designated as $3z$, and all other fluxes can be calculated as a function of $z$, using metabolite balances and a metabolic steady state assumption, as described below.

**Metabolite balances**

Metabolite balances for each metabolite can be written in the form: accumulation = input − output (Stephanopoulos et al., 1998). This results in the following equations:

\[
\frac{d}{dt}[G6P] = v_{pts} - v_{kci} - v_{pgi} \tag{2}
\]

\[
\frac{d}{dt}[P5P] = v_{pgi} - 2v_{ckc} - v_{akf} \tag{3}
\]

\[
\frac{d}{dt}[S7P] = v_{ckc} - v_{tal} \tag{4}
\]

\[
\frac{d}{dt}[E4P] = v_{tal} - v_{akf} \tag{5}
\]

\[
\frac{d}{dt}[F6P] = v_{hti} + v_{tal} + v_{akf} - v_{pft} \tag{6}
\]

\[
\frac{d}{dt}[T3P] = 2v_{pft} + v_{ckc} - v_{tal} + v_{akf} - v_{gap} \tag{7}
\]

where $v$ denotes the flux of the reaction indicated by its subscript.

Further, we can assume that the metabolites are in quasi-steady state, which is justified because of high enzyme turnovers in biochemical systems (Stephanopoulos et al. 1998). As a consequence, the accumulation rates of all metabolites [left-hand sides of Eqs. (2)-(8)] can be equated to 0. Since $v_{pts}$ has been taken to be 1 and $v_{pgi}$ has been taken to be equal to $3z$, Eq. (2), along with the quasi-steady state assumption gives
\[ v_{hti} = 1 - 3z \]  
(8)

Also, from Eqs. (3), (4) and (5), we obtain
\[ v_{hti} = v_{tal} = v_{ktB} = z \]  
(9)

Using Eq. (6) and (7) then gives
\[ v_{pfr} = 1 - z \]  
\[ v_{gap} = 2 - z \]  
(10)

This completes the evaluation of all metabolic fluxes in the network under consideration.

**Specific cases of the network**

We have considered four different cases of the network with respect to reaction
reversibilities: (I) all reactions are irreversible, (II) only \textit{hxi} is reversible, (III) \textit{hxi}, \textit{tktA} and \textit{tktB} is reversible, and (IV) \textit{hxi}, \textit{tktA}, \textit{tktB}, and \textit{tal} are reversible. Case I is a is an idealized, simple example, while case IV is the most general case accommodating all thermodynamically reversible reactions. Cases II and III are special situations with restricted applicability, and case III is known to be especially applicable to many plant systems (Hatzfeld and Stitt, 1990; Hill and ap Rees, 1994). Only case IV, the most general case, is explained below, while the results for the other cases are given.

**Bondomer enumeration**

The bondomers formed in this case are depicted in Figure 2.2. Starting with an intact glucose molecule (GLU 32), we obtain \( G_{32} (1-2-3-4-5-6) \) through \textit{pts}, and \( P_{16} (1-2-3-4-5) \) through \textit{pgI}, \( S_{62} (1-2 3-4-5-6-7) \) and \( T_{4} (1-2-3) \) through \textit{tktA}, \( F_{26} (1-2 3 4-5-6) \) and \( E_{5} (1-2-3-4) \) through \textit{tal}, \( F_{30} (1-2 3-4-5-6) \) and \( T_{4} (1-2-3) \) through \textit{tktB}. Further, we also obtain \( F_{32} (1-2-3-4-5-6) \) through \textit{hxi} as well as \( T_{4} (1-2-3) \) and \( T_{5} (1-2-3) \) through \textit{pfk}.
Thus far, the bondomers that have been formed are the same as those in the case where all reactions are irreversible. However, since \( hxi \) is reversible, \( F_{26} \) and \( F_{30} \) formed in the previous cycle will yield \( G_{26} \) and \( G_{30} \) through the backward \( hxi \) reaction. In the next cycle through the pentose phosphate pathway, we obtain \( P_{13} \, (1 \, 2 \, 3 \rightarrow 4 \rightarrow 5) \) and \( P_{15} \, (1 \, 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 6) \) from \( G_{26} \) and \( G_{30} \) respectively through \( pgl \). Now there are 3 bondomers in the P5P pool (\( P_{13}, P_{15} \) and \( P_{16} \)), and various combinations of them will yield \( S_{61} \, (1 \, 2 \, 3 \rightarrow 4 \rightarrow 5 \rightarrow 6 \rightarrow 7) \), \( S_{68} \, (1 \, 2 \, 3 \, 4 \rightarrow 5 \rightarrow 6 \rightarrow 7) \), \( S_{57} \, (1 \, 2 \, 3 \, 4 \rightarrow 5 \rightarrow 6 \rightarrow 7) \), \( S_{50} \, (1 \, 2 \, 3 \, 4 \rightarrow 5 \rightarrow 6 \rightarrow 7) \) and \( S_{49} \, (1 \, 2 \, 3 \, 4 \rightarrow 5 \rightarrow 6 \rightarrow 7) \) through \( tktA \), as shown in Figure 2.2. Further bondomers formed along the pentose phosphate pathway are \( E_{7}, F_{17}, F_{18} \) and \( F_{25} \) (\( tal \)), \( F_{25}, F_{26}, F_{29} \) (\( tktB \)). The \( pfk \) reaction then yields the additional bondomer \( T_{1} \, (1 \, 2 \, 3) \) for T3P.

The backward \( hxi \) reaction will now yield the G6P bondomers \( G_{17}, G_{18}, G_{25} \) and \( G_{29} \) from the corresponding F6P bondomers. In the next cycle, \( G_{17} \) and \( G_{18} \) generate \( P_{9} \, (1 \, 2 \, 3 \rightarrow 4 \rightarrow 5) \) through \( pgl \), which results in \( S_{33} \) and \( S_{34} \) (\( tktA \)) and \( E_{5}, F_{1}, F_{2} \) (\( tal \)).

In the next course through the pentose phosphate pathway, \( F_{1} \) and \( F_{2} \) will yield \( G_{1} \) and \( G_{2} \) through the backward \( hxi \) reaction. These will subsequently form \( P_{1} \, (1 \, 2 \, 3 \rightarrow 4 \rightarrow 5) \) through \( pgl \), \( S_{1}, S_{2} \) (\( tktA \)) and \( E_{1} \) (\( tal \)). Proceeding in this manner, and after considering the reversibilities of the \( tktA, tal \) and \( tktB \) reactions, one can obtain the set of all bondomers formed in this metabolic network. The complete bondomer enumeration is shown in Figure 2.2, and the exhaustive set of bondomers for this case is as follows: \( \{G_{1}, G_{2}, G_{4}, G_{17}, G_{18}, G_{20}, G_{25}, G_{26}, G_{28}, G_{29}, G_{30}, G_{32}\} \) for G6P, a parallel set for F6P, \( \{T_{1}, T_{3}, T_{4}\} \) for T3P, \( \{P_{1}, P_{2}, P_{9}, P_{10}, P_{13}, P_{14}, P_{15}, P_{16}\} \) for P5P, \( \{S_{1}, S_{2}, S_{4}, S_{5}, S_{6}, S_{33}, S_{34}, S_{36}, S_{57}, S_{38}, S_{49}, S_{50}, S_{52}, S_{53}, S_{54}, S_{57}, S_{58}, S_{60}, S_{61}, S_{62}\} \) for S7P and \( \{E_{1}, E_{5}, E_{7}, E_{9}\} \) for E4P.

It should be noted that the number of bondomers for cases I–III are less than that for case IV: in case I only 6 bondomers are realized, while 40 bondomers are realized in case II and 49 bondomers in case III.
Bondomer balances

The steady state bondomer balances for case IV can be directly written, and they are as follows.

E4P:

\[
\theta = v_{tal}^+ \begin{bmatrix} S_1 + S_2 + S_4 + S_5 + S_6 \\ S_{33} + S_{34} + S_{36} + S_{37} + S_{38} \\ S_{49} + S_{50} + S_{32} + S_{33} + S_{34} \\ S_{57} + S_{58} + S_{60} + S_{61} + S_{62} \end{bmatrix} + v_{skib}^- \begin{bmatrix} F_1 + F_2 + F_4 \\ F_{17} + F_{18} + F_{20} \\ F_{25} + F_{26} + F_{28} \\ F_{29} + F_{30} + F_{32} \end{bmatrix} - (v_{skib}^+ + v_{tal}^-) \begin{bmatrix} E_1 \\ E_5 \\ E_7 \\ E_8 \end{bmatrix}
\]  

(11)

G6P:

\[(v_{psl} + v_{hsl}^+)_G_i = v_{hsl}^- F_i \quad i \neq 32\]

\[(v_{psl} + v_{hsl}^+)_G_i = v_{hsl}^- F_i + v_{psl}^+ \quad i = 32\]

(12)

T3P:

\[
\theta = v_{phk} \begin{bmatrix} F_1 + F_2 + F_4 + F_a \\ F_{17} + F_{18} + F_{20} + F_b \\ F_{26} + F_{25} + F_{28} + F_{29} + F_{30} + F_{32} + F_c \end{bmatrix} + (v_{sk4}^+ + v_{skib}^+) \begin{bmatrix} P_1 + P_2 \\ P_9 + P_{10} \\ P_{13} + P_{14} + P_{15} + P_{16} \end{bmatrix}
\]  

(13)

\[-(v_{psp} + v_{sk4}^- + v_{skib}^- + v_{tal}^+)_T_i \]

\[
\begin{bmatrix} T_1 \\ T_3 \\ T_4 \end{bmatrix} + v_{tal}^- \begin{bmatrix} F_1 + F_2 + F_4 \\ F_{17} + F_{18} + F_{20} \\ F_{26} + F_{25} + F_{28} + F_{29} + F_{30} + F_{32} \end{bmatrix}
\]
P5P:

\[
\begin{bmatrix}
G_1 + G_2 \\
G_4 \\
G_{17} + G_{18} \\
G_{20} \\
G_{25} + G_{26} \\
G_{28} \\
G_{29} + G_{30} \\
G_{32}
\end{bmatrix} + v_{\text{delta}}^{+} \frac{v_{\text{delta}}^{+}}{v_{\text{na}}^{+} + v_{\text{delta}}^{+}} P_{10} + v_{\text{delta}}^{-} \frac{v_{\text{delta}}^{-}}{v_{\text{na}}^{-} + v_{\text{delta}}^{-}} E_{7} + v_{\text{delta}}^{-} E_{8} + (S_b + S_c)T_4
\]

\[(14)\]

F6P:

\[
\begin{bmatrix}
G_1 & S_a T_1 \\
G_2 & S_a T_1 \\
G_4 & S_a T_1 \\
G_{17} & S_a T_3 \\
G_{18} & S_a T_3 \\
G_{20} & S_a T_5 \\
G_{25} & S_a T_5 \\
G_{26} & S_a T_4 \\
G_{28} & S_a T_4 \\
G_{29} & S_a T_4 \\
G_{30} & S_a T_4 \\
G_{32} & 0
\end{bmatrix} + v_{\text{shift}}^{+} P_{\text{odd}} E_7 - (v_{\text{shift}}^{-} + v_{\text{shift}}^{+} + v_{\text{shift}}^{+}) F_{20}
\]

\[(15)\]
In these equations, the superscripts ‘+’ and ‘−’ refer to forward and reverse reactions, respectively. For example, \( v_{hxi}^{+} \) and \( v_{hxi}^{-} \) are the fluxes of the forward and backward \( hxi \) reaction,

\[
G6P \xleftarrow{\text{hxi}} F6P
\]

If we assume the rate of the backward reaction to be \( r \), then we have \( v_{hxi}^{+} = 1-3z+r \) and \( v_{hxi}^{-} = r \). Similarly, the forward and backward fluxes of the other reactions can be substituted from their values in Eqs. (8)-(10) [net fluxes] or Figure 2.1 [reversible fluxes]. \( P_{\text{odd}} \) and \( P_{\text{even}} \) refer to P5P bondomers with odd (\( 1 \text{----} 2 \ldots \)) and even (\( 1-2\ldots \)) subscripts, respectively.

\[
\begin{align*}
0 &= v_{\text{atkl}}^{+} + v_{\text{atkl}}^{-} - (v_{\text{atkl}}^{+} + v_{\text{atkl}}^{-}) \\
&= E_{1}F_{a} \quad S_{1} \\
&= E_{1}F_{b} \quad S_{2} \\
&= E_{1}F_{c} \quad S_{3} \\
&= E_{2}F_{a} \quad S_{4} \\
&= E_{2}F_{b} \quad S_{5} \\
&= E_{2}F_{c} \quad S_{6} \\
&= E_{3}F_{a} \quad S_{7} \\
&= E_{3}F_{b} \quad S_{8} \\
&= E_{3}F_{c} \quad S_{9} \\
&= E_{4}F_{a} \quad S_{10} \\
&= E_{4}F_{b} \quad S_{11} \\
&= E_{4}F_{c} \quad S_{12} \\
&= E_{5}F_{a} \quad S_{13} \\
&= E_{5}F_{b} \quad S_{14} \\
&= E_{5}F_{c} \quad S_{15} \\
&= E_{6}F_{a} \quad S_{16} \\
&= E_{6}F_{b} \quad S_{17} \\
&= E_{6}F_{c} \quad S_{18} \\
&= E_{7}F_{a} \quad S_{19} \\
&= E_{7}F_{b} \quad S_{20} \\
&= E_{7}F_{c} \quad S_{21} \\
&= E_{8}F_{a} \quad S_{22} \\
&= E_{8}F_{b} \quad S_{23} \\
&= E_{8}F_{c} \quad S_{24} \\
&= E_{9}F_{a} \quad S_{25} \\
&= E_{9}F_{b} \quad S_{26} \\
&= E_{9}F_{c} \quad S_{27} \\
&= E_{10}F_{a} \quad S_{28} \\
&= E_{10}F_{b} \quad S_{29} \\
&= E_{10}F_{c} \quad S_{30} \\
&= E_{11}F_{a} \quad S_{31} \\
&= E_{11}F_{b} \quad S_{32} \\
&= E_{11}F_{c} \quad S_{33} \\
&= E_{12}F_{a} \quad S_{34} \\
&= E_{12}F_{b} \quad S_{35} \\
&= E_{12}F_{c} \quad S_{36} \\
&= E_{13}F_{a} \quad S_{37} \\
&= E_{13}F_{b} \quad S_{38} \\
&= E_{13}F_{c} \quad S_{39} \\
&= E_{14}F_{a} \quad S_{40} \\
&= E_{14}F_{b} \quad S_{41} \\
&= E_{14}F_{c} \quad S_{42} \\
&= E_{15}F_{a} \quad S_{43} \\
&= E_{15}F_{b} \quad S_{44} \\
&= E_{15}F_{c} \quad S_{45} \\
&= E_{16}F_{a} \quad S_{46} \\
&= E_{16}F_{b} \quad S_{47} \\
&= E_{16}F_{c} \quad S_{48} \\
&= E_{17}F_{a} \quad S_{49} \\
&= E_{17}F_{b} \quad S_{50} \\
&= E_{17}F_{c} \quad S_{51} \\
&= E_{18}F_{a} \quad S_{52} \\
&= E_{18}F_{b} \quad S_{53} \\
&= E_{18}F_{c} \quad S_{54} \\
&= E_{19}F_{a} \quad S_{55} \\
&= E_{19}F_{b} \quad S_{56} \\
&= E_{19}F_{c} \quad S_{57} \\
&= E_{20}F_{a} \quad S_{58} \\
&= E_{20}F_{b} \quad S_{59} \\
&= E_{20}F_{c} \quad S_{60} \\
&= E_{21}F_{a} \quad S_{61} \\
&= E_{21}F_{b} \quad S_{62} \\
&= E_{21}F_{c} \quad S_{63}
\end{align*}
\]
definitions of the other bondomer groups introduced in these equations are as follows:

\[ S_a = S_1 + S_3 + S_4 + S_5 + S_7 + S_6_1, \]
\[ S_b = S_2 + S_6 + S_4 + S_3 + S_7 + S_6_2, \]
\[ S_c = S_4 + S_6 + S_5 + S_6_0, \]
\[ F_a = F_1 + F_7 + F_25 + F_29, \]
\[ F_b = F_2 + F_18 + F_26 + F_30, \]
\[ F_c = F_4 + F_20 + F_28 + F_32. \]

It may be noted that the S7P balance, Eq. (16), has been employed to eliminate S7P
terminals from the second term of Eq. (14). Similarly, Eq. (16) can be used to eliminate the
S7P terms from the E4P balance, Eq. (11), giving

\[
0 = v_{t_1}^+ \frac{v_{t_2}^+}{v_{t_1}^-} = v_{t_1}^+ - v_{t_2}^- + v_{t_1}^+ - v_{t_2}^- + \]
\[ + v_{t_3}^+ \frac{v_{t_4}^+}{v_{t_3}^-} = v_{t_3}^+ - v_{t_4}^- + \]
\[ + v_{t_5}^+ \frac{v_{t_6}^+}{v_{t_5}^-} = v_{t_5}^+ - v_{t_6}^- + \]
\[ + v_{t_7}^+ \frac{v_{t_8}^+}{v_{t_7}^-} = v_{t_7}^+ - v_{t_8}^- + \]
\[ + v_{t_9}^+ \frac{v_{t_10}^+}{v_{t_9}^-} = v_{t_9}^+ - v_{t_10}^- + \]
\[ + v_{t_11}^+ \frac{v_{t_12}^+}{v_{t_11}^-} = v_{t_11}^+ - v_{t_12}^- + \]
\[ + v_{t_13}^+ \frac{v_{t_14}^+}{v_{t_13}^-} = v_{t_13}^+ - v_{t_14}^- + \]
\[ + v_{t_15}^+ \frac{v_{t_16}^+}{v_{t_15}^-} = v_{t_15}^+ - v_{t_16}^- + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
\[ + \]
For example, Eq. (18) was obtained by grouping the rows containing $P_{odd}$ and $P_{even}$ in the P5P balance, Eq. (14). The definitions of $G_a$, $G_b$ and $G_c$ are analogous to those of $F_a$, $F_b$ and $F_c$. Eqs. (18), (19), (20) and (21) are linear equations which can be solved to obtain expressions for the bondomer groups $P_{odd}$, $S_a$, $S_b$, etc., which can then be substituted in Eqs. (12), (13), (14), (15) and (17) to obtain expressions for all bondomers.

Thus the nonlinear bondomer balances have been reduced to two sets of linear equations to be solved in series. This is analogous to the concept of cumulative bondomers (van Winden et al., 2002), although we have used a top-down approach where the bondomer balances are written first, the multiplicative terms ($P_{odd}$, $S_a$, $S_b$, etc.) have been balanced next. This makes use of only those bondomers which are physically realized in this network, leading to a substantially lesser number of equations and a simplification of the system state.

To make the solutions compact we have used the following definition for the reversibility extent, $e$, of the hxi reaction

$$ e = \frac{r}{1 - z + r} \quad \text{(22)} $$

as well as the reversibility extents $e_A$ and $e_B$ for the tktA and tktB reactions,

$$ e_A = \frac{r_A + r_B}{r_A + r_B + 3z} \quad \text{as well as} $$

$$ e_B = \frac{r_B}{r_B + z} \quad \text{(23)} $$

The above procedure for enumerating realistic bondomers, writing bondomer balances and converting them to linear equations can be generalized to the following:

1. Enumerate all bondomers generated in the metabolic network, after repeated passes through the pathways. Prepare a list of these non-zero abundance bondomers.
2. Write balance equations for these bondomers. Call this equation set \((E_1)\). In the example above these correspond to Eqs. (11)–(16).

3. Identify fragments of bondomers that cause bilinearities. In the example these correspond to \(P_a, P_b, S_a\), etc.

4. Add appropriate balances from \(E_1\) above, to generate balances for the fragments identified in step 3 above. Call this equation set \((E_2)\). In the example this corresponds to Eqs. (18)–(21).

5. If the set of equations in step 4 is not linear, repeat steps (3) and (4) till the equations of step 4 are linear.

6. A set of equations \((E_1), (E_2), \ldots (E_n)\) is now available. Solve these starting with \((E_n)\) and proceeding backwards on to \((E_1)\).

Although this approach follows the logic as the pioneering cumomer balancing technique (Wiechert et al., 1999) or the balancing of all cumulative bondomers (van Winden et al., 2002), this approach utilizes a top-down approach where the equations for the highest bondomers are written first, and only realistic bondomers and bondomer fragments are balanced. For a real system in which only a subset of all bondomers are realized, this results in fewer balances. Also the fragments in step 3 are analogous to cumulative bondomers – as mentioned before, the method above considers only the ones that are realized in the network.

**Identifiability analysis**

Identifiability analyses were conducted to determine structural identifiability of the fluxes from available experimental data, for cases III and IV. Here, only structural identifiability is performed, i.e. the measurement errors are assumed to be zero (Isermann and Wiechert, 2003). The results indicate whether at all the fluxes in the network are identifiable from the
available measurement data. Statistical identifiability – the other aspect of identifiability –
takes into account measurement errors, and is out of the scope of this paper.

To test identifiability, the Jacobian matrix of the measurable groups of bondomers with
respect to the free flux parameters \( z, r, r_d, r_b, r_l \) was calculated, and its rank was
determined. This could be done analytically for case III and was done numerically for case IV.

**Use of Boolean function mapping to simulate bondomer distributions**

In this section, we introduce Boolean function mapping as a novel method of simulating
bondomer distributions. This method involves modeling of carbon skeletal rearrangement
steps as Boolean or arithmetic operations on the decimal representation of a bondomer [Eq.
(1)], thus enabling rapid and automatic calculation of the bondomer numbers of the products
which will be formed from a given set of reactant bondomers through a given reaction. The
theory and implementation of this method are described in the following.

The Boolean function mapping method is based upon the fact that all reactions in a metabolic
network can be represented as reactions between two reactants \( (R_1, R_2) \) to give two products
\( (P_1, P_2) \), i.e. they can be represented as ‘bi-bi’ reactions (Wiechert and Murzel, 2001). Here,
only the reactant and product molecules undergoing carbon skeletal rearrangements are
considered, as catalyzing molecules do not affect the \(^{13}\text{C} \) label or bondomer distribution.
Such a reaction can be represented as \( R_1 + R_2 \rightarrow P_1 + P_2 \). In the most general case, each
reactant or product molecule consists of up to two fragments \( (R_1^L, R_1^R, P_1^L, P_1^R) \); the
superscripts \( L \) and \( R \) denoting ‘left’ and ‘right’ fragments), which intermingle to form the
product molecules. They can thus be represented by the schema

\[
R_1 + R_2 \rightarrow R_1^L - R_1^R + R_2^L - R_2^R \rightarrow P_1^L - P_1^R + P_2^L - P_2^R \rightarrow P_1 + P_2
\]
We have classified the steps in this schema into four different moves: fragmentation, reversal, transposition and condensation. Each of these, as well as the Boolean functions used to model them, is explained below. The metabolic reaction referred to in the explanations is the transaldolase reaction, whose carbon skeletal rearrangement is given in the literature [e.g. Follstad and Stephanopoulos (1998)].

**Fragmentation.** This is the step in which a reactant molecule splits into two fragments, in the manner $R_i \rightarrow R_i^L + R_i^R$. For example, the transaldolase reaction ($S7P + T3P \rightarrow E4P + F6P$), $R_i$ is $S7P$, and $R_i^L$ is the 1-2-3 fragment of $S7P$. Since $T3P$ ($R_2$) does not undergo fragmentation in this reaction, $R_2^L$ is null, and $R_2^R$ is the entire $T3P$ molecule (1-2-3). We have designated the point of fragmentation, or the bond number at which fragmentation occurs, as $X$. Therefore, for $S7P$ in the transaldolase reaction, $X = 3$.

The bondomer number of the reactant can be converted to the bondomer numbers of its fragments, by integer division. Thus

$$R_i^L = R_i \% 2^{X-1} \text{ if } X > 0 \text{ else } R_i^L = 0$$

and

$$R_i^R = R_i / 2^X$$

where the "%" indicates the modulus operation (or the remainder after integer division), and the "/" indicates integer division.

**Reversal.** This step involves the complete reversal of a reactant fragment. This is an occasional step, and is encountered in reactions such as $F6P \rightarrow T3P$. Reversal is achieved by a bit-reversal of the reacting fragment. Bit-reversals can be achieved by straightforward algorithms, requiring no additional storage. Press et al., (1992) provide an example of the same.
Transposition. This step involves a shift in position of the bonds of a reacting fragment, when it becomes a product fragment. For example, in the transaldolase reaction, the T3P molecule becomes the right fragment of F6P, and the first and second bonds of T3P become the fourth and fifth bonds of the product molecule. In other words, they are shifted by three positions.

Transposition of a reactant fragment can accomplished by a bitshift operation, which is available in standard programming languages such as C or C++. A shift to a higher bond number (such as the example in the previous paragraph) is accomplished by a left shift (represented in C by "<<"), while a shift to a lower bond number is accomplished by a right shift (represented in C by ">>").

Condensation. This is the final step, in which the product molecule fragments combine to form entire molecules, \( P_i^L + P_i^R \rightarrow P_i \). Product molecule bondomer numbers can be obtained simply by adding the bondomer numbers of the their fragments,

\[ P_i = P_i^L + P_i^R \]

It should be noted that at the '1' which was added for convenience to the bondomer numbers in Eq. (1), is not needed in the Boolean function mapping operations. Therefore, a '1' should be subtracted from the reactant bondomer numbers before Boolean function mapping, and added to the product bondomer numbers after the mapping operation. This will become clear in the following example.

Example

The sequence of Boolean operations described above is illustrated here with an example. Consider the transaldolase reaction, S7P + T3P \( \rightarrow \) E4P + F6P. Starting with an arbitrary pair of reactant bondomers, say \( S_{61} (1 \ 2 \ 3-4-5-6-7) \) and \( T_3 (1 \ 2-3) \), we find from inspection of the carbon skeletal rearrangement (Figure 2.2), that the reaction should yield \( E_8 (1-2-3-4) \)
and $F_{17} (1 \ 2 \ 3 \ 4 \ 5-6)$. On performing the Boolean operations on the reactant bondomers, as shown below, we arrive at those exact product bondomers, shown below.

We first subtract 1 from the reactant bondomer number, obtaining $S_{60}$ and $T_2$.

**Fragmentation:** The left and right fragments of the reactant bondomers are as follows.

- $R_1$: $R_1^L = 60 \% \ 2^Y = 60 \% \ 2^2 = 0$, $R_1^R = 60 / 2^Y = 60 / 2^3 = 7$;
- $R_2$: $R_2^L = 0$, $R_2^R = 2 / 2^0 = 2 / 1 = 2$.

**Reversal:** None of the fragments in this example are reversed, so this step is not necessary here.

**Transposition:** The $R^R$ fragment is the only one that undergoes transposition; it is left shifted by three bits. The left shift operation converts $R_2^R$ to $R_2^R << 3 = 16$.

At this point, we can assign the reactant fragments to the product fragments,

- $P_1^L = R_1^R = 7$, $P_1^R = \text{null}$; $P_2^L = R_1^L = 0$, $P_2^R = R_2^R = 16$

**Condensation:** In this final step the product fragments can be added, giving $P_1 = 7 + 0 = 7$ (i.e. $E_7$) and $P_2 = 0 + 16 = 16$ (i.e. $F_{16}$). At this point we can add 1 to the product bondomer numbers to give $E_8$ and $F_{17}$, as expected.

Thus it was possible to use Boolean function mapping to calculate the correct product bondomers rapidly and automatically. The only input needed by the routine implementing the mapping is the carbon skeletal rearrangement and product fragment assignments, which can be written in a compact and clear form as shown in Table 2.1.

In the numerical simulation of BLEs, Boolean function mapping could be employed in two ways: (i) to automatically generate bondomer mapping matrices from carbon skeletal rearrangement data, which can be then used in procedures such as to simulate bondomer
distributions by the method of cumulative bondomer balancing, or (ii) to simulate bondomer distributions numerically, in an iterative fashion and on a need-basis.

In (ii) above, Boolean function mapping is proposed to be incorporated into an iterative solution of BLEs instead of BMMs. Iterative solution schemes (e.g. Schmidt et al., 1999) are still popular in $^{13}$C MFA research (Siddiquee et al., 2003; Jiao et al., 2003) due to their ease of implementation. In such schemes, a set of initial bondomer abundances are assumed, and the bondomer vector corresponding to each metabolite is updated at each successive step by pre-multiplying it with the BMM corresponding to the carbon-skeleton transition [analogous to the procedure of Schmidt et al. (1997)]. The pre-multiplication by the BMM is needed to trace the product bondomers that result from the reactant bondomers at each step. However, this tracing could also be achieved more efficiently by Boolean function mapping, resulting in a faster iterative procedure as shown in Discussion.

**Results**

The solutions of the bondomer balances for various cases are as follows. For case I, where the number of bondomers realized is only six, the T3P abundances are

$$T_3 = \frac{2z}{2-z}$$  \hspace{1cm} (24)

and

$$T_4 = \frac{2-3z}{2-z}$$  \hspace{1cm} (25)

For cases II and III, the solution is depicted in Table 2.2 (case II) and Table 2.3 (case III). Table 2.2 also shows a special case of $hxi$ reversibility: one where $hxi$ is completely reversible ($e = 1$ or $r \to \infty$). This is a common physiological scenario when the G6P bondomers are in complete equilibrium with the corresponding F6P bondomers, and Eq. (12)
reads \( G_i = F_i \forall i \). The explicit formulas for individual bondomers in case III as well as bondomers or measurable groups of bondomers in case IV are too lengthy to present in a table in symbolic form. These are available in a MATLAB® workspace file which may be obtained from the authors on request. They can also be obtained by solving the presented sets of linear equations symbolically.

**Flux identifiability**

Based on the analytical solution obtained for the abundances of various bondomers, we conducted analysis on the identifiability of fluxes in the network from all measurable bondomers or groups of bondomers. The analytical solution permitted evaluation of the Jacobian matrix in symbolic form for all four cases considered. The analysis of cases III and IV – which are applicable to practical situations – is presented here.

For case III, assuming that bondomers can be experimentally measured from proteinogenic amino acids and polysaccharides originating from G6P (starch, trehal), the set of bondomers or bondomer groups which can be measured is \( \{E_1 + E_5, E_7, E_8, P_1 + P_2 + P_9 + P_{10}, P_{13} + P_{14}, P_{15}, P_{16}, T_1 + T_3, T_4, G_1 + G_2 + G_4 + G_17 + G_{18} + G_{20}, G_{23} + G_{26} + G_{28}, G_{29} + G_{30}, G_{32}\} \), i.e. a total of 13 measurements. Since the sum of some of the bondomers is 1 (e.g. \( \Sigma P_i = 1 \) the number of linearly independent bondomer measurements reduces to 9. The Jacobian of the above measurement vector with respect to the four free fluxes in this case \( (z, r, r_A, r_B) \) was obtained symbolically using MATLAB®. The rank of this matrix could also be calculated symbolically, and was determined to be 4, signifying that the four fluxes in case III are identifiable from the above measurements for any set of fluxes. Thus the a priori identifiability was proved for case III.

In case IV, we have the following measurements: \( \{E_1, E_5, E_7, E_8, P_1, P_2, P_9, P_{10}, P_{13}, P_{14}, P_{15}, P_{16}, P_{odd}, P_{even}, T_1, T_3, T_4, G_1 + G_2 + G_4, G_{17} + G_{18} + G_{20}, G_{23} + G_{26} + G_{28}, G_{29} + G_{30}, G_{32}, G_a, G_b, G_c\} \), i.e. a total of 25 measurements. Here, measurements from 1-bond and long-range C-C
couplings that have been reported to be possible this far, have been considered. Again, since the sum of some of the bondomers is 1 (e.g. $P_{odd} + P_{even} = 1$) the number of linearly independent bondomer measurements reduces to 19. The Jacobian for case IV was too enormous to process further, a problem also encountered by van Winden et al. (2002). We therefore substituted the values of the fluxes $z = 0.20$, $r = 0.40$, $r_A = 3.80$, $r_L = 0.33$, and $r_B = 0.625$, to further process the Jacobian and obtain representative results regarding flux identifiability.

Since case IV contains only 5 free fluxes ($z$, $r$, $r_A$, $r_B$, $r_L$), the measurement of any 5 independent bondomers (or groups of bondomers) should theoretically suffice to solve for the 5 free fluxes. There are $19C_5 = 11,628$ such combinations available, from the 19 independent measurements. Numerically, however, a group of 5 measurements would provide a good estimate of the 5 free fluxes, only if the condition number of the corresponding 5×5 Jacobian submatrix is less than the reciprocal of the precision to which the measurements can be made. We found that 65% of the 11,628 combinations satisfied this criterion. Going by the condition number of the Jacobian matrix, the best 5 sets of measurements were found to be as follows: 

$$\{E_5, P_9, P_{15}, P_{16}, G_{25}+G_{26}+G_{28}\}; \{E_5, P_{10}, P_{15}, P_{16}, G_{25}+G_{26}+G_{28}\}; \{E_5, P_{15}, P_{16}, G_{25}+G_{26}+G_{28}, G_a\}; \{E_5, P_2, P_{15}, P_{16}, G_{25}+G_{26}+G_{28}\}; \{E_5, P_{15}, P_{16}, G_{25}+G_{26}+G_{28}, G_{29}+G_{30}\}.$$ 

The condition numbers of the corresponding Jacobian submatrices were not very different from the condition number of the entire Jacobian matrix where all 19 independent measurements are considered. We also noted bondomers measurable from proteinogenic amino acids alone ($E_i$, $T_i$, $P_i$) or bondomers measurable from polysaccharides alone ($G_i$) are sufficient to give good estimates of fluxes. When a combination of proteinogenic amino acid and polysaccharide data is used, the condition number improves, meaning that using data from polysaccharides reinforces the metabolic flux information contained in the proteinogenic amino acids.
Discussion

Analytical solutions for bondomer abundances through bondomer grouping

In this work, we presented analytical solutions for bondomer abundances in the glycolysis and pentose phosphate pathways, to aid in the evaluation of fluxes from BLEs. The solution method uses the technique of bondomer enumeration and balancing of bondomers whose abundances are not identically zero. We employed the interesting property of bondomers that natural abundance is automatically factored into them, which implies that not all bondomers are generated in a realistic network. By precluding bondomers which are identically zero in the metabolic network of interest, one obtains a system of equations containing fewer bondomer balances. For example, in case IV, the total number of equations needed was 50, [11 equations for the bondomer groups $P_{odd}$, $P_{even}$, etc. – Eqs. (18)–(21); and 39 equations for the individual bondomers – Eqs. (11)–(15) and (17)]. In the cumulative bondomer balancing method described previously in the literature, balances would have to be written for all bondomers in the network whether they are physically realized or not. This would have resulted in 156 equations, which is more than three times the number that was solved here. In the case of isotopomers every isotopomer in the network would have had non-zero abundance (because of natural abundance). This would have resulted in 312 equations having to be written for a complete solution. While the cumulative bondomer concept (van Winden et al., 2002) illustrated the strength of the bondomer concept (and reduced the number of equations by half), we show that a further improvement that can be achieved (a further reduction from 156 to 39 equations) by the bondomer enumeration and zero bondomer elimination-based analytical solution method presented here.

The significance of solving for all bondomers is that this would enable the user to make maximal use of NMR data from 1-bond C–C couplings, long-range C–C couplings, as well as mass spectroscopy data. If only 1 bond C–C couplings are considered, the number of
equations due to our method reduces to 33, since only linear combinations of bondomers need to be solved for. This is also a 30% reduction compared to the 43 equations solved using the cumulative bondomer method where all cumulative bondomers were balanced (van Winden et al., 2002).

A systematic approach for writing realistic bondomer balances is also presented. This is a top-down approach where the equations for the highest order bondomers are written first. This is the chief difference between the presented method and the cumomer and cumulative bondomer balancing techniques described earlier in the literature, which are bottom-up methods. The two methods (top-down or bottom-up) would give exactly the same equations in case of isotopomers, since all isotopomers are realized in a metabolic network due to natural abundance of $^{13}$C. On the contrary, the top-down approach is useful for bondomer balancing since it results in much fewer balances, as shown in the example.

This reduction of the number of bondomer balances can be expected to result in lower computation times. This is because matrix factorization – which is the most time-consuming step in BLE or CLE simulation – is an $O(N^3)$ process ($N$ being the number of rows or columns). Therefore, if the number of equations reduces by a factor $f$, the computation time will reduce by a factor greater than $f$. The time needed to solve the 50 linear balances described in this paper was $3.6 \times 10^{-4}$ s (Pentium computer, 3.06 GHz, 512 MB RAM, Red Hat Linux). For other networks such as *Catharanthus roseus* central carbon metabolism (Sriram and Shanks, in preparation), the time needed was of the same order of magnitude. Because the linear balances have to be set up only once, the increased computational efficiency of this method would be amplified in iterative flux evaluation procedures needing hundreds or even thousands of bondomer abundance simulations as part of a global optimization routine. When we used the linear balances described in this paper in such an
optimization routine, we were able to evaluate fluxes and reversibility extents in an average time of 20 s (same computer system as above).

Also, many researchers use Monte Carlo techniques to obtain a probability distribution of the fluxes (e.g. Schmidt et al., 1999). Such techniques (e.g. Press et al., 1992) need very little knowledge of analytical statistics and are relatively straightforward to implement, although time-consuming as noted by Möllney et al. (1999) and Araúzo-Bravo and Shimizu (2003). The above reduction in time per simulation should make it relatively easier to implement Monte Carlo techniques for statistical analysis.

Further, the simplification of the system state due to fewer balances brings identifiability analysis closer to reach. For the specific case of the metabolic network considered in case III which had four degrees of freedom, a priori identifiability of fluxes could be proved, irrespective of the value of the fluxes chosen. This is the first example of a practical network with four degrees of freedom (three of them being reaction reversibilities), where a priori identifiability analysis could be proved. A similar analysis was not possible for case IV. Nevertheless, a semi-numerical procedure was implemented and identifiability was shown for a representative set of fluxes. Minimal sets of measurements which best identify the free fluxes were listed. It was also shown that information from amino acids alone is sufficient to identify the fluxes, and information from sugars (G6P) reinforces the flux information. This has implications on the development of a high-throughput flux analysis, as the use of only amino acid data implies that only one extraction/hydrolysis needs to be performed, which considerably reduces the experimental effort.

**Reaction reversibility effects in the glycolysis and pentose phosphate pathways**

An end-user of a BLE is primarily interested in net fluxes only. Although the NMR spectrum is a non-trivial imprint of the net fluxes, the reversibilities of reactions in the network can affect the relationship between net fluxes and NMR-measurable quantities. This study also
sheds light on how reaction reversibilities affect the net-flux-bondomer abundance relationship. For example, it was seen that the formula for $T_4$ bondomer when no reactions are assumed reversible, is

$$T_4 = \frac{2 - 3z}{2 - z}$$  \hspace{1cm} (26)

However, when the $hxi$ reaction is assumed reversible, the expression expands to

$$T_4 = \frac{2 + 4ez - 3z}{(1 + 2ez)(2 - z)}$$  \hspace{1cm} (27)

For example when the value of $z$ is 0.30, the abundance $T_4$ can vary from 0.65 for $e = 0$, to 0.85 for $e = 1$, a 30% change. Further, when the $tktA$ and $tktB$ reactions are also assumed reversible, the expression changes to

$$T_4 = \frac{5z - 4ez(1 - z - e_b) - (2 + 3z^2)(1 - e_b) - 3ze_b(2 + ez)}{(2 - z)[(1 - z - e_b)(1 + 2ez) + ze_b(2 + ez)]}$$  \hspace{1cm} (28)

and this changes to an even lengthier expression when the $tal$ reaction is also reversible (not shown). This depicts how the reversibility of a reaction can affect the correlation between fluxes and NMR data. Figure 2.3 depicts a graphical representation of the reversibility effect of the $tktB$ reaction – the bondomer $T_4$ is shown as a function of $z$, for values of $e_b$ from 0 through 1.

**Boolean function mapping**

In this work, we have introduced Boolean function mapping as a quick method to obtain bondomer mapping matrices (BMMs) from compact representations of carbon skeletal rearrangements. This method has been shown to be efficient compared with previously published methods.
This is helpful in the case where a metabolic network is not completely known, but is intended to be deciphered using $^{13}$C NMR data. For instance, significant chunks of pathways in plant metabolism are unknown, and many new pathways are being discovered (e.g. Schwender et al., 1997). The development of $^{13}$C MFA techniques for such systems will necessarily require numerous a posteriori modifications of an initially assumed metabolic network, till one is obtained that matches the experimental data well. Thus, new reactions and their carbon skeletal rearrangement have to be introduced many times. This is essentially a process of pathway discovery using $^{13}$C NMR data. A systematic process of such discovery would involve a combinatorial optimization which would conduct an a posteriori modification of an assumed metabolic network, till one is obtained that best fits the $^{13}$C NMR data. Tools such as Boolean function mapping presented here, and metabolic pathway synthesis (Mavrovouniotis et al., 1992a,b; Stephanopoulos et al., 1998) would be highly useful in such research.

The Boolean function mapping method can also be incorporated in iterative solutions to BLE (analogous to those proposed by Schmidt et al. (1997) for CLEs) which are still popular because of their ease of implementation (Siddiquee et al., 2003; Jiao et al., 2003). Here, Boolean function mapping can replace BMMs in tracing which product bondomers are formed from the given reactant bondomers, at each step of the iteration. Since not all bondomers of a given molecule are physically realized (as explained before) one can simulate and balance only the bondomers that are actually realized in an experiment. Additionally, since all metabolic reactions can be represented as bi-bi reactions where only one reactant undergoes fragmentation, this method results in fewer operations than the previous technique of Schmidt et al. (1997). This saves considerable computational time in comparison to the BMM method where all possible bondomers have to be balanced. For example, in a particular case (the tktA reaction) we found only 11 S7P bondomers were
realized, out of a possible 64. The Boolean function mapping method to simulate this reaction required only 1536 operations, compared to a BMM multiplication, which needed 4094 steps. In most simulations of practical metabolic networks, we found that the number of operations was of the order of 100. This improved method is hence computationally economical. While simulating bondomer distributions in *Catharanthus roseus* hairy roots (Sriram and Shanks, in preparation) the simulation time was 0.48 s (3.06 GHz processor, 512 MB RAM, Red Hat Linux). This time is for a method employing iterative solution of the balance equations where Boolean function mapping is used instead of mapping matrices in each iteration step. When we compared Boolean function mapping-based iteration with mapping matrix-based iteration on the same computer we noticed that the Boolean function mapping method, owing to its efficiency, performed better by a factor of at least 3.

**Summary**

In this paper, we have introduced analytical and numerical improvements in the interpretation of data obtained from a bond labeling experiment by employing the concept of bondomers. Because natural abundance of $^{13}$C is automatically factored into the definition of a bondomer, a number of bondomers in real networks are identically zero. These can be eliminated a priori, thus resulting in fewer balance equations. The technique of bondomer enumeration and balancing was shown to yield a compact set of linear equations which can be solved to obtain analytical formulas for NMR-measurable quantities in terms of fluxes. This has been demonstrated on an exemplar metabolic network rich in reversible bi-bi (two substrate-two product) reactions. Based on the analytical solution obtained, the identifiability of fluxes for two practical cases of glycolysis/pentose phosphate pathway was analyzed. For one case with four degrees of freedom (case III), a priori identifiability was proved irrespective of the set of fluxes. For the more general case with five degrees of freedom (case
IV), identifiability was proved using a representative set of fluxes, and minimal sets of bondomer abundances that would provide good estimates of the fluxes were listed. Further, Boolean function mapping, a hitherto unreported numerical innovation useful toward flux evaluation, has been delineated. This method has shown to be efficient compared to previously described methods for the same purpose, the efficiency arising out of the fact that many bondomers in a real network have zero abundances. Two applications of this method are proposed: (i) calculating bondomer-mapping matrices, and (ii) use in an iterative solution for BLEs.

We have been able to extend the analytical solution method shown in this paper to other networks, notably the central carbon metabolic network of soybean embryos (Sriram and Shanks, in preparation) and *Catharanthus roseus* (Sriram and Shanks, in preparation). Further, we are also extending the analysis to two-substrate BLEs and metabolic networks involving subcellular compartments. Although it may be argued that the analytical solution method is network-specific, it consumes only a fraction of the time needed for preparing an NMR sample and conducting an NMR experiment. The development of a bondomer-based flux evaluation software specifically for BLEs, will also be discussed in our future publications.

**Acknowledgments**

This material is based upon work supported by the National Science Foundation under grants BES 9906978 and BES 0224600, and the Plant Sciences Institute at Iowa State University.
References


**Appendix A: Nomenclature**

**Metabolites**

Aro aromatic ring of phenylalanine and tyrosine  
E erythrose-4-phosphate  
E4P erythrose-4-phosphate  
F fructose-6-phosphate  
F6P fructose-6-phosphate  
G glucose-6-phosphate  
G6P glucose-6-phosphate  
GLU glucose  
S sedoheptulose-7-phosphate  
S7P sedoheptulose-7-phosphate  
T triose-3-phosphate (glyceraldehyde-3-phosphate)  
T3P triose-3-phosphate (glyceraldehyde-3-phosphate)
### Fluxes, reactions

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Enzyme Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>gap</td>
<td>glyceraldehyde phosphate dehydrogenase</td>
</tr>
<tr>
<td>hxi</td>
<td>hexose isomerase</td>
</tr>
<tr>
<td>pfk</td>
<td>phosphofructokinase</td>
</tr>
<tr>
<td>pgl</td>
<td>phosphogluconate</td>
</tr>
<tr>
<td>pts</td>
<td>phosphotransferase</td>
</tr>
<tr>
<td>pyk</td>
<td>pyruvate kinase</td>
</tr>
<tr>
<td>tal</td>
<td>transaldolase</td>
</tr>
<tr>
<td>tktA, tktB</td>
<td>transketolase</td>
</tr>
</tbody>
</table>
Table 2.1. Spreadsheet representation of metabolic reactions for processing by Boolean function mapping. The columns $R_1$, $R_2$, $P_1$, $P_2$ and $X$ denote the reactants, products, and the bond number at which fragmentation occurs, $X$. The columns $P_1^L$, $P_1^R$, $P_2^L$, $P_2^R$ denote which reactant fragments are allocated as which product fragments. It may be noted that when a metabolite does not fragment (i.e. $X = 0$, e.g. G6P in the $hxif$ reaction) the unbroken molecule is the same as its right fragment (and not left fragment). It may also be noted that the $pts$ reaction does not have any internal metabolites as reactants, and the $gap$ reaction does not have any internal metabolites as products. The reversal of a fragment is indicated by appending the lower-case letter ‘$r$’ as the third character of an entry, e.g. $pfk$ reaction.

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Reactants</th>
<th>Products</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R_1$</td>
<td>$X$</td>
</tr>
<tr>
<td>$pts$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$hxif$</td>
<td>G6P</td>
<td>0</td>
</tr>
<tr>
<td>$hxib$</td>
<td>F6P</td>
<td>0</td>
</tr>
<tr>
<td>$pfk$</td>
<td>F6P</td>
<td>3</td>
</tr>
<tr>
<td>$pgl$</td>
<td>G6P</td>
<td>1</td>
</tr>
<tr>
<td>$tktAf$</td>
<td>P5P</td>
<td>2</td>
</tr>
<tr>
<td>$tktAb$</td>
<td>S7P</td>
<td>2</td>
</tr>
<tr>
<td>$talf$</td>
<td>S7P</td>
<td>3</td>
</tr>
<tr>
<td>$talb$</td>
<td>F6P</td>
<td>3</td>
</tr>
</tbody>
</table>
Table 2.2. Bondomer abundances in Case II, when only hxi is irreversible. Only physically realizable bondomers are shown; those not shown have an abundance of 0. The abundances of (groups of) G6P bondomers (precursors of oligo- and polysaccharides) can be calculated from the relationships: \( \{G_1 + G_2 = P_1; G_{17} + G_{18} = P_9; G_{25} + G_{26} = P_{13}; G_{29} + G_{30} = P_{15}; G_{32} = P_{16}\} \)

<table>
<thead>
<tr>
<th>Bondomer</th>
<th>Abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( hxi )</td>
</tr>
<tr>
<td></td>
<td>partially reversible</td>
</tr>
<tr>
<td></td>
<td>( 0 &lt; e &lt; 1 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bondomer</th>
<th>Abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T_1 )</td>
<td>( \frac{4ez}{(2-z)(1+2ez)^2} )</td>
</tr>
<tr>
<td>( T_3 )</td>
<td>( \frac{2ez}{(2-z)(1+2ez)^2} )</td>
</tr>
<tr>
<td>( T_4 )</td>
<td>( \frac{2+4ez-3z}{(1+2ez)(2-z)} )</td>
</tr>
<tr>
<td>( P_1 )</td>
<td>( \frac{4z^3ez^2}{(2-z)(1+ez)(1+2ez)^2} )</td>
</tr>
<tr>
<td>( P_9 )</td>
<td>( \frac{2z^2}{(2-z)(1+ez)(1+2ez)^2} )</td>
</tr>
<tr>
<td>( P_{13} )</td>
<td>( \frac{ez(2+4ez-3z)}{(2-z)(1+2ez)(1+ez)} )</td>
</tr>
<tr>
<td>( P_{15} )</td>
<td>( \frac{ez}{(1+2ez)(1+ez)} )</td>
</tr>
<tr>
<td>( P_{16} )</td>
<td>( \frac{1}{1+2ez} )</td>
</tr>
</tbody>
</table>
Table 2.2. (Continued.)

<table>
<thead>
<tr>
<th>$E_1$</th>
<th>1 2 3 4</th>
<th>$\frac{4ez^2(1-ez)}{(2-z)(1+2ez)^2} \cdot \frac{ez}{1+ez}$</th>
<th>$\frac{4z^2(1-z)}{(2-z)(1+2z)^2} \cdot \frac{z}{1+z}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_8$</td>
<td>1 2 3-4</td>
<td>$\frac{2ez(1-ez)}{(2-z)(1+2ez)^2} \cdot \frac{z}{1+ez}$</td>
<td>$\frac{2z(1-z)}{(2-z)(1+2z)^2} \cdot \frac{z}{1+z}$</td>
</tr>
<tr>
<td>$E_7$</td>
<td>1 2-3-4</td>
<td>$\frac{ez(2+4ez-3z)}{(2-z)(1+2ez)(1+ez)}$</td>
<td>$\frac{(2+z)z}{(2-z)(1+2z)(1+2z)}$</td>
</tr>
<tr>
<td>$E_8$</td>
<td>1-2-3-4</td>
<td>$\frac{1}{(1+ez)}$</td>
<td>$\frac{1}{1+z}$</td>
</tr>
</tbody>
</table>
Table 2.3. Abundances of bondomers and measurable groups of bondomers in Case III, when hxi, tktA and tktB are irreversible. Only physically realizable bondomers are shown; those not shown have an abundance of 0. Some bondomer abundances are also expressed in terms of $p (=T_1+T_3)$ and $q (=T_4)$.

<table>
<thead>
<tr>
<th>Bondomer</th>
<th>Abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_1^+$</td>
<td>$1 \ 2 \ 3$</td>
</tr>
<tr>
<td>$T_3$</td>
<td>$1 \ 2 \ -3$</td>
</tr>
<tr>
<td>$T_4$</td>
<td>$1-2-3$</td>
</tr>
<tr>
<td>$P_{1+}$</td>
<td>$1 \ 2 \ 3 \ 4 \ 5$</td>
</tr>
<tr>
<td>$P_{2+}$</td>
<td>$1-2 \ 3 \ 4 \ 5$</td>
</tr>
<tr>
<td>$P_{9+}$</td>
<td>$1 \ 2 \ 3 \ 4-5$</td>
</tr>
<tr>
<td>$P_{10}$</td>
<td>$1-2 \ 3 \ 4-5$</td>
</tr>
<tr>
<td>$P_{13+}$</td>
<td>$1 \ 2 \ 3-4-5$</td>
</tr>
<tr>
<td>$P_{14}$</td>
<td>$1-2 \ 3-4-5$</td>
</tr>
</tbody>
</table>
Table 2.3. (Continued.)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_1^+$</td>
<td>$z(1 - ez)(2 - e_B){(1 - z - e_B)[ez(1 + e_A) + e_A] + ze_B(1 + 2e_A)}$</td>
<td>$(2 - z){(1 - z - e_B)(1 + 2ez) + ze_B(2 + ez)}(1 + ez + eze_A)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$E_5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$E_7$</td>
<td>$4ez(1 - z - e_B) + (2 + 3z^2)(1 - e_B) + 3ze_B(2 + ez) - 5z$</td>
<td>$\frac{1 - z}{(2 - z){(1 - z - e_B)(1 + 2ez) + ze_B(2 + ez)}(1 + ez + eze_A)}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$E_8$</td>
<td>${ze_B[e(1 + e_A) - (2e_A + 1)] + e_Ae_B + (1 - e_A)(1 - z)}$</td>
<td>$\frac{1 - ez + eze_A}{(1 - z)(1 + ez + eze_A)}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$G_1^+$</td>
<td>$\frac{pez(1 + e_A)}{1 + ez + eze_A}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$G_2^+$</td>
<td>$\frac{ez^2(1 - z)(1 - ez)(2 - e_B)(1 + e_A)}{(2 - z){(1 - z - e_B)(1 + 2ez) + ze_B(2 + ez)}(1 + ez + eze_A)}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$G_{17}^+$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$G_{18}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$G_{25}^+$</td>
<td>$\frac{qez(1 + e_A)}{1 + ez + eze_A}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$G_{26}$</td>
<td>$\frac{ez[5z - 4ez(1 - z - e_B) - (2 + 3z^2)(1 - e_B) - 3ze_B(2 + ez)](1 + e_A)}{(2 - z){(1 - z - e_B)(1 + 2ez) + ze_B(2 + ez)}(1 + ez + eze_A)}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$G_{29}^+$</td>
<td>$\frac{ez{e(1 + e_A) - (2e_A + 1)} + e_Ae_B + (1 - e_A)(1 - z)}$</td>
<td>$\frac{[1 - z - e_B)(1 + 2ez) + ze_B(2 + ez)](1 + ez + eze_A)}{[(1 - z - e_B)(1 + 2ez) + ze_B(2 + ez)]}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$G_{30}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$G_{32}$</td>
<td>$\frac{(1 - z) - e_B(2z - 1)}{[(1 - z - e_B)(1 + 2ez) + ze_B(2 + ez)]}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure captions

Figure 2.1. Glycolytic and pentose phosphate pathways. Metabolites are represented within white ovals, and gray ovals connected to metabolites show compounds (amino acids, nucleotides) for which that metabolite is a precursor. Reactions are indicated using gray arrows, and black arrows alongside the reaction are labeled using values of the flux(es) in the forward (and reverse, if applicable), direction(s). Some metabolites (F6P, T3P) appear at more than one place in the figure, to avoid unnecessary intersections of lines. Such multiple appearances are connected by dotted lines.

Figure 2.2. Enumeration of bondomers for case IV, when hxi, tktA, tal, and tktB are reversible. Arrows representing reactions are color-coded as shown in legend. Forward and reverse reactions are represented using solid and dashed arrows, respectively. Arrows have rounded ends near the reactant molecule, and pointed ends near the product molecule. Some bondomers appear at more than one place in the figure, to avoid unnecessary intersections of lines. Small vertical or horizontal bars are provided next to some bondomers, to clarify the grouping of the lines connecting the bondomers, when not obvious. In the pfk reaction, dense gray lines indicate the formation of 2 product molecules of T3P, while faint gray lines indicate the formation of a single T3P molecule. The formation of T3P bondomers from F6P through the talb reaction (F_{1,2,4} \to T_1, F_{17,18,20} \to T_3, F_{25,26,28,29,30,32} \to T_4) is omitted, for clarity.

Figure 2.3. Effect of reaction reversibility on bondomer abundance. The abundance of the bondomer $T_4$ is shown as a function of the pentose phosphate pathway free flux variable, $z$, for $e = 0.5$, $r_1 = 0$ and various values of the reversibility $e_B$. This bondomer is independent of the reversibility $e_A$. 
Figure 2.1. Glycolytic and pentose phosphate pathways.
Figure 2.2. Enumeration of bondomers for case IV, when hxi, tktA, tal, and tktB are reversible.
Figure 2.3. Effect of reaction reversibility on bondomer abundance.
3. Quantification of compartmented metabolic fluxes in developing soybean embryos by employing biosynthetically directed fractional $^{13}$C labeling, 2-D [$^{13}$C, $^1$H] NMR and comprehensive isotopomer balancing

A paper under revision in Plant Physiology.

Ganesh Sriram§, D. Bruce Fulton*, Vidya V. Iyer§, Joan Marie Peterson&; Ruilian Zhou&, Mark E. Westgate§, Martin H. Spalding§, Jacqueline V. Shanks*

Departments of Chemical Engineering (G. S., V. V. I., J. V. S.), Biochemistry, Biophysics and Molecular Biology (D. B. F.), Agronomy (J. M. P., R. Z., M. E. W.), Genetics, Developmental and Cell Biology (M. H. S.); Iowa State University, Ames, IA 50011

Abstract

Metabolic flux quantification in plants is instrumental in the detailed understanding of metabolism, but is difficult to perform on a systemic level. Toward this aim, we report the development and application of a computer-aided metabolic flux analysis tool that enables

---

1 Financial sources: National Science Foundation (BES-0224600); Plant Sciences Institute, Iowa State University; Iowa Soybean Promotion Board.

§ Primary researcher and author.

* NMR facility manager.

§ Graduate student in J. V. Shanks lab, assisted with NMR analysis.

& Research scientists in M. E. Westgate lab, performed soybean embryo culture.

‡ Co-principal investigator.

* Principal investigator and corresponding author.
the concurrent evaluation of fluxes in several primary metabolic pathways. Labeling experiments were performed by feeding a mixture of U-\textsuperscript{13}C sucrose, naturally abundant sucrose, and glutamine to developing soybean (\textit{Glycine max}) embryos. 2-D \textsuperscript{13}C, \textsuperscript{1}H NMR spectra of seed storage protein and starch hydrolysates were acquired, and yielded a labeling data set consisting of 155 \textsuperscript{13}C isotopomer abundances. We developed a computer program to automatically calculate fluxes from this data. This program accepts a user-defined metabolic network model, and incorporates recent mathematical advances toward accurate and efficient flux evaluation. Fluxes were calculated and statistical analysis was performed to obtain standard deviations. A high flux was found through the oxidative pentose phosphate pathway (19.99 ± 4.39 μmol d\textsuperscript{-1} cotyledon\textsuperscript{-1}, or 104.2 carbon mol ± 23.0 carbon mol per 100 carbon mol of sucrose uptake). Separate transketolase and transaldolase fluxes could be distinguished in the plastid and the cytosol, and those in the plastid were found to be at least 6-fold higher. The backflux from triose to hexose phosphate was also found to be substantial in the plastid (21.72 ± 5.00 μmol d\textsuperscript{-1} cotyledon\textsuperscript{-1}, or 113.2 carbon mol ± 26.0 carbon mol per 100 carbon mol of sucrose uptake). Forward and backward directions of anaplerotic fluxes could be distinguished. The glyoxylate shunt flux was found to be negligible. Such a generic flux analysis tool can serve as a quantitative tool for metabolic studies and phenotype comparisons, and can be extended to other plant systems.

\textbf{Keywords:} Metabolic flux analysis, soybean embryo, \textit{Glycine max}, compartmented metabolism, \textsuperscript{13}C labeling, NMR, isotopomer.

\textbf{Introduction}

The evaluation of metabolic flux is instrumental in understanding carbon partitioning in plant metabolism. Since fluxes provide a quantitative depiction of carbon flow through competing metabolic pathways (Ratcliffe and Shachar-Hill, 2001), they are an important physiological
characteristic akin to levels of transcripts, proteins, and metabolites (Sauer, 2004). Flux measurements and comparisons of fluxes between phenotypes can provide insights toward selection of appropriate metabolic engineering targets (Stephanopoulos, 1999; Glawischnig et al., 2002) and toward the construction of predictive models of plant metabolism (Ratcliffe and Shachar-Hill, 2001), the necessity for which has been emphasized recently (Girke et al., 2003; Katagiri, 2003; Raikhel and Coruzzi, 2003).

Although the importance of flux measurement in plants has often been stressed (Roscher et al., 2000; Ratcliffe and Shachar-Hill, 2001; Shachar-Hill, 2002; Sweetlove et al., 2003), it has received rather limited attention in plant science as compared to profiling technologies for transcript, protein and metabolite levels (Kruger and von Schaewen, 2003; Sweetlove et al., 2003). This is principally due to the fact that fluxes have to be quantified by back-calculating them from their effect on the distribution of a labeled substrate, and such calculation requires a detailed mathematical model if it has to be accurate. Mathematical models relating labeling data to fluxes are often non-trivial, particularly in the case of compartmented metabolism inherent in plants. Consequently, flux measurement technology in plants remains under-developed (Ratcliffe and Shachar-Hill, 2001; Sweetlove et al., 2003).

Not surprisingly, most papers that reported labeling studies in plants have focused on the qualitative goal of inferring which pathways are in operation (identification of metabolic network topology), but not on the mathematically involved endeavor of evaluating how much carbon is processed by those pathways (quantification of flux). For example, Wheeler et al. (1998) delineated the pathway of ascorbic acid synthesis in higher plants from $^{14}$C labeling data, and Krook et al. (1998) showed that two separate oxidative pentose phosphate pathways (oxPPP) operate in the cytosol and the plastid, using $^{13}$C enrichment data from Daucus carota cell suspensions. More recently, a suite of articles by Eisenreich and co-workers (Bacher et al., 1999; Glawischnig et al., 2000; Glawischnig et al., 2001;
Glawischnig et al., 2002) reported the abundances of isotopomers of several isolated sink metabolites in maize kernels. Although these articles demonstrated advances in label measurement technology, the inferences from them were either qualitative or semi-quantitative.

Two recent pioneering research efforts have concentrated on quantification of fluxes in plants. In the first, Raymond and co-workers calculated fluxes through glycolysis, oxPPP, tricarboxylic acid (TCA) cycle and anaplerotic reactions, using $^{13}\text{C}$ atom enrichment data of metabolites isolated from tomato suspension cells (Rontein et al., 2002) and maize root tips (Dieuaide-Noubhani et al., 1995). However, $^{13}\text{C}$ isotopomer abundance data provide richer metabolic information than $^{13}\text{C}$ atom enrichments. Isotopomers are isomers of a metabolite which differ in the labeling state ($^{13}\text{C}$ or $^{12}\text{C}$) of their individual carbon atoms, as shown in Figure 3.1. In a system that was supplied with a mixture of U-$^{13}\text{C}$ labeled and U-$^{12}\text{C}$ labeled substrates, the isotopomer measurements can be used to track carbon-carbon connectivities in metabolites (Szyperski, 1995; 1998). Therefore, isotopomer abundance data facilitate statistically superior estimation of metabolic fluxes and reaction reversibilities than atom enrichment data (Wiechert et al., 1999). The second important development in flux quantification in plants was by Ohlrogge and co-workers, who used mass isotopomer data of metabolites separated from developing Brassica napus embryos to calculate fluxes (Schwender et al., 2003), although only the glycolysis and oxPPP were considered.

Despite these advances, flux measurement in plants is in its early stages. Systemic evaluation of fluxes from overdetermined isotopomer data sets, as well as detailed statistical analysis of the evaluated fluxes have not yet been implemented in plant metabolism to the extent of their application to prokaryotic metabolism. Secondly, the quantification of fluxes of parallel pathways in two compartments (e.g. cytosolic and plastidic oxPPP, mitochondrial and plastidic malic enzymes) has not been reported to date. Also, while the aforementioned
studies have isolated metabolites before collecting labeling data, this effort-intensive step is not necessary since two-dimensional (2-D) NMR can be used to resolve a mixture of several metabolites.

In this paper we report labeling studies and flux quantification in developing embryos of soybean (*Glycine max*), metabolizing sucrose and glutamine in liquid culture. Soybeans are important sources of protein, oil, and nutraceuticals, and developing embryos are important in vitro models systems to study them (Saravitz and Raper, 1995). There exists motivation to physiologically characterize this system to understand carbon partitioning between pathways and identify potential metabolic engineering targets. We acquired an overdetermined isotopomer abundance data set in a high-throughput fashion, by using two-dimensional (2-D) NMR. To convert the isotopomer abundances in this data set to fluxes, a computer tool was developed, which incorporated recent mathematical and statistical advances (Wiechert *et al.*, 1999; Wiechert and Murzel, 2001; Sriram and Shanks, 2001; 2004) in metabolic flux analysis theory. This tool is not specific to our metabolic network model and accepts user-defined metabolic models. Our results show that in the developing embryos, large amounts of carbon are shunted through the oxPPP, and through the gluconeogenic pathway from triose phosphate to fructose phosphate in the plastid. The activities of the anaplerotic pathways, glyoxylate shunt and γ-aminobutyric acid (GABA) shunt were also quantified. Moreover, we were able to distinguish between parallel pathways in separate compartments.

**Results**

We performed labeling experiments by culturing developing soybean embryos in liquid medium with sucrose (10% w/w U-13C, 90% w/w naturally abundant) and glutamine (naturally abundant) as the only carbon sources. This labeling technique is termed ‘biosynthetically directed fractional 13C labeling’ (Szyperski, 1995). After 6 d of culture, a
protein fraction and a starch fraction were extracted from the embryos and hydrolyzed. 2-D NMR experiments were performed on the respective hydrolysates, and heteronuclear $^{13}$C, $^1$H] type NMR spectra were acquired. These spectra were used to quantify isotopomorphic compositions of sink metabolites.

2-D NMR $[^{13}$C, $^1$H] spectra of sink metabolites, cross peak assignments

A $[^{13}$C, $^1$H] heteronuclear single quantum correlation (HSQC) spectrum of the seed protein hydrolysate is shown in Figure 3.2. The $^{13}$C axis (labeled Fl) on this spectrum spans the $^{13}$C chemical shift range 10 to 50 ppm. Cross peaks on this spectrum correspond to carbon atoms (that are attached to protons) of compounds in the protein hydrolysate. In the spectrum shown in Figure 3.2, we identified aliphatic carbon atoms of 16 amino acids, levulinic acid (LVA), and 5-hydroxymethyl furfural (HMF). Each carbon atom was identifiable by its unique $^{13}$C and $^1$H chemical shifts as well as distinctive coupling patterns and J-coupling constants ($J_{CC}$). Explanations of chemical shifts and $J_{CC}$ are provided in Harris (1983).

The amino acids identified in the spectrum resulted from degradation of the seed protein under the hydrolysis conditions employed (145 °C, vacuum, 6 N HCl), and are therefore proteinogenic amino acids synthesized in the embryos. The LVA and HMF peaks appear on the spectrum because soybean seed storage protein (most of the protein in the developing embryo) is highly glycosylated (Doyle et al., 1986), the attached sugars being predominantly mannose and glucosamine (Yamauchi and Yamagishi, 1979). Under the hydrolysis conditions employed, the hexose skeletons of mannose, glucose and glucosamine are converted to LVA and HMF (Sriram G, Iyer VV, Shanks JV, unpublished data).

To assign the cross peaks to carbon atoms, chemical shift values for the amino acids obtained from Wüthrich et al. (1976) and $J_{CC}$ values obtained from Krivdin and Kalabin (1989) were used. The assignments were also verified using supplementary 2-D and 3-D NMR spectra of the hydrolysate of a 100% $^{13}$C-labeled protein (data not shown). Chemical shifts and $J_{CC}$
values for the carbon atoms of LVA and HMF were obtained by analyzing $^{13}\text{C, } ^1\text{H}$ spectra of hydrolysates of glucose labeled at various positions (Sriram G, Iyer VV, and Shanks JV, unpublished data).

A second $^{13}\text{C, } ^1\text{H}$ HSQC spectrum of the seed storage protein was acquired, where the $^{13}\text{C}$ axis spanned the chemical shift range 90 to 160 ppm. Herein, the aromatic carbon atoms of Tyr, Phe and His were detected. A $^{13}\text{C, } ^1\text{H}$ spectrum of the starch hydrolysate was acquired, where the $^{13}\text{C}$ axis spanned the chemical shift range 10 to 50 ppm. This spectrum contained peaks corresponding to the aliphatic carbon atoms of LVA and HMF. This is expected since starch is a glucose polymer, and its hydrolysate should contain LVA and HMF for the reasons stated above.

**Fine structures of peaks, quantification of isotopomer abundances**

The cross peaks in the $^{13}\text{C, } ^1\text{H}$ spectrum displayed peak splitting along the $^{13}\text{C}$ dimension, due to $^{13}\text{C-}^{13}\text{C}$ scalar coupling. This is evident in expanded views of the cross peaks, e.g. Gly $\alpha$ and Asp $\beta$ (Figure 3). Detailed descriptions of scalar coupling, why it causes peak splitting, and the types of multiplet peaks resulting from peak splitting, are provided in Harris (1983) and Cavanagh et al. (1996). Briefly, such peak splitting indicates the presence of multiple isotopomers of the detected compounds. The Gly $\alpha$ peak exhibits a central singlet peak (s), and two doublet peaks (d) distributed on either side of the singlet (Figure 3A). The singlet in the Gly $\alpha$ fine structure represents a population of Gly isotopomers in which the $\alpha$ atom ($\alpha$) has a $^{13}\text{C}$ nucleus and the carboxyl atom (C) adjacent to it has a $^{12}\text{C}$ nucleus. Whereas, the doublet represents a population of Gly isotopomers in which the $\alpha$ atom and the carboxyl atom both have $^{13}\text{C}$ nuclei. Using boldface to represent $^{13}\text{C}$ atoms and regular font for $^{12}\text{C}$ atoms, the isotopomer population corresponding to the singlet may be represented as [Ca], and the one corresponding to the doublet as [Ca]. Likewise, the fine structure of the Asp $\beta$ cross peak shows a singlet (s), a doublet (d1), a doublet (d2) and a double doublet (dd)
(Figure 3B). In the above notation, the isotopomer population represented by the singlet is \([xαβγ]\), that represented by the doublet \(d1\) is \([xαβγ]\), that by the doublet \(d2\) is \([xαβγ]\), and that by the double doublet is \([xαβγ]\). Here 'x' stands for an undeterminable labeling state, i.e. the atom represented by ‘x’ (Asp carboxyl) cannot be detected from the Asp β fine structure.

These satellite peaks observed in the fine structure of a given cross peak are termed multiplets. The abundances of the isotopomer populations represented by the multiplets are directly proportional to the integrals of the respective multiplet peaks. We quantified peak integrals by various methods depending on the complexity of the fine structure, as described in Materials and Methods. The isotopomeric compositions of amino acids resulting from the quantification (a total of 155 relative isotopomer abundances) are listed in Supplementary Material I.

Labeling states of precursor metabolites by retrobiosynthetic analysis

To evaluate metabolic fluxes of reactions in primary metabolism, the isotopomer abundances of central metabolic precursors need to be calculated. These were determined from the labeling states of the sink metabolites by retrobiosynthetic reconstruction, following the approach of Szyperski (1995) and Glawischnig et al. (2001). For instance, Thr is metabolically synthesized from its precursor plastidic oxaloacetate (OAA\(^p\)), and the four carbon atoms of Thr (denoted as \([Cαβγ]\)), correspond to the carbon atoms of OAA\(^p\) (denoted as \([1234]\)). Since OAA\(^p\) is the only source of Thr, its isotopomeric composition should be calculable from that of Thr. For example the fractional abundance of the OAA isotopomer \([123x]\) (relative to the total OAA pool) should be equal to that of the Thr isotopomer \([Cαβx]\), or the intensity of the singlet in the Thr α peak (relative to the total Thr α signal). The precursor isotopomer structures corresponding to the quantified multiplets are shown in Supplementary Material I. In some cases, only sums of isotopomers can be assigned to a multiplet, rather than a single isotopomer. This occurs for cross peaks such as Tyr β, where
the doublet represents the sum of two Tyr isotopomers. It also occurs for cross peaks of sink metabolites that are synthesized from multiple metabolic precursors, such as Lys, which is synthesized from pyruvate and OAA.

The multiplet intensities of the sink metabolites provide an overdetermined data set for the calculation of the isotopomeric compositions of their precursor. This is because usually, multiple sink metabolites are synthesized from the same metabolic precursor. For example, plastidic phosphoenolpyruvate (PEP\textsuperscript{p}) is a metabolic precursor to both Phe and Tyr. The abundance of the PEP\textsuperscript{p} isotopomer [123] determined from the Phe \( \alpha \) singlet was 0.205 ± 0.012, while that determined from the Tyr \( \alpha \) singlet was 0.208 ± 0.017. Both values are in close agreement. Good consistency was noted for all amino acids synthesized from the same precursor. Leu \( \delta^1 \) was found to be the only exception. When the abundance of the isotopomer [x23] of plastidic pyruvate (Pyr\textsuperscript{p}) was determined from the singlets of three amino acids synthesized from Pyr\textsuperscript{p} (Leu \( \delta^1 \), Val \( \gamma^1 \), and Ile \( \gamma^2 \)), the values obtained from Val \( \gamma^1 \) (0.235 ± 0.002) and Ile \( \gamma^2 \) (0.242 ± 0.002) were in agreement with each other and with the abundance of the same isotopomer [x23] of PEP\textsuperscript{p}, the immediate precursor of Pyr. However, the value obtained from Leu \( \delta^1 \) (0.357 ± 0.005) is significantly different.

**Isotopic equilibration of metabolites between compartments**

We observed that the isotopomeric compositions of two hexose nucleotide pools – one located in the cytosol and the other in the plastid – were dissimilar. These were obtained from the LVA and HMF peaks of the protein and starch hydrolysates. Starch, a glucose polymer, is synthesized from plastidic ADP-glucose, which is in isotopic equilibrium with the hexose phosphates in the plastid. Therefore, the isotopomeric composition of the glucose monomer of starch (obtained from its hydrolysis products LVA and HMF) reflects the isotopomeric composition of the plastidic hexose phosphate pool. On the contrary, the hexose sugars attached to glycosylated protein are synthesized from nucleotide sugars UDP-glucose.
or GDP-mannose (Faik et al., 2000; Baldwin et al., 2001), the synthesis of which occurs in the cytosol (Coates et al., 1980). Therefore, the isotopomeric composition of the cytosolic hexose nucleotide pool is determinable from the LVA and HMF peaks in the protein hydrolysate spectrum. Figure 3.4A shows a comparison of the isotopomer abundances of the cytosolic and plastidic hexose nucleotide pools determined thus. Clearly, most of the abundances are significantly different. This suggests the presence of separate, non-equilibrating metabolic pathways in the cytosol and the plastid.

However, the isotopomeric compositions of the triose phosphates in cytosolic and plastidic compartments were not significantly different. These were obtained by comparing the multiplet intensities of Ala α, Phe α, and Tyr α. Phe and Tyr are synthesized from plastidic PEP and therefore reflect the isotopomeric composition of the plastidic triose phosphates. Ala is synthesized both in the cytosol and in the plastid (Ireland and Lea, 1999); therefore its isotopomeric composition represents a combination of those of the triose phosphates in both compartments. From Figure 3.4B it is evident that corresponding isotopomer abundances are similar. However, the multiplet intensities of Ser α (doublet d1 and double doublet) are different. This suggests the involvement of Ser in reactions in which the triose phosphate pool does not participate.

Interestingly, the multiplet intensities of His α (which represents the isotopomers of the [xx345] fragment of plastidic pentose phosphate, P5P) bear similarity with the multiplets of Ala α, Phe α, and Tyr α. This suggests rapid equilibration between P5P and the triose phosphate pool (T3P) represented by Ala α, Phe α, and Tyr α. P5P and T3P are interconverted by the transketolase reaction, a highly reversible reaction of the non-oxidative limb of the oxPPP (Lea and Leegood, 1999). The presence of transketolase in the plastid is therefore implied.
The carbon in plastidic pyruvate originates largely from sucrose

We found that the carbon in plastidic pyruvate (Pyr\textsuperscript{p}) originates almost entirely from sucrose, and not from other external carbon substrates glutamine or CO\textsubscript{2}. The carbon in any metabolite synthesized in the embryos could be a mixture of the carbon from three available external carbon sources – sucrose, glutamine (both present in the liquid medium), and CO\textsubscript{2} (through photosynthetic fixation). To determine the source of the carbon in Pyr\textsuperscript{p}, we determined the \textsuperscript{13}C enrichment of atom 3 of plastidic pyruvate (Pyr\textsuperscript{p}), and compared it with the \textsuperscript{13}C enrichment of the three external carbon sources. The doublet intensities of Leu $\delta^2$ and Val $\gamma^2$ provide the \textsuperscript{13}C enrichment of atom 3 of plastidic pyruvate (Pyr\textsuperscript{p}) synthesized \textit{de novo} since the beginning of the labeling experiment (Szyperski, 1995). In our data, these intensities were: Leu $\delta^2$(d) = 0.123 ± 0.001, and Val $\gamma^2$(d) = 0.118 ± 0.002.

Thus the enrichment of Pyr\textsuperscript{p} synthesized during the labeling experiment, was almost identical to the enrichment of the substrate sucrose, and very different from the enrichments of the substrates Gln or CO\textsubscript{2}. (Sucrose had a substantial \textsuperscript{13}C enrichment in the range 0.11 to 0.12, i.e. 0.10 from $^{\text{U-13}}$C Suc and approximately 0.011 from naturally \textsuperscript{13}C abundance; while glutamine and CO\textsubscript{2} had a small natural \textsuperscript{13}C abundance of 0.011.) The observation that the enrichment of Pyr\textsuperscript{p} is identical to that of sucrose, suggests that the carbon in Pyr\textsuperscript{p} originates entirely from sucrose. Therefore, glutamine or external CO\textsubscript{2} fixation by photosynthesis make small or negligible contributions to the carbon in Pyr\textsuperscript{p}, since a substantial contribution from either of these carbon sources to Pyr\textsuperscript{p} would have resulted in a \textsuperscript{13}C enrichment considerably lower than 0.11.

**Extracellular fluxes, fluxes contributing to biomass accumulation**

The measurements of extracellular fluxes and fluxes toward biomass synthesis were as follows. The average rate of biomass accumulation in the developing soybean embryos was 2.3 g d\textsuperscript{-1} cotyledon\textsuperscript{-1}. The sucrose consumption was \textit{9.59}$\times$10\textsuperscript{-6} µmol d\textsuperscript{-1} cotyledon\textsuperscript{-1}. The
contents of biomass, including protein, oil, starch, and proteinogenic amino acid proportions, are listed in Supplementary Material II. The proportions of amino acids in the protein compared well with published values for soybean embryo seed storage protein (Bewley and Black, 1994).

**Metabolic network model**

The calculation of metabolic fluxes from labeling data requires a model of the metabolic network. Our model is shown in Figure 3.5. The model includes all principal pathways of primary metabolism [glycolysis, oxPPP, TCA cycle, anaplerotic shunts, glyoxylate shunt and γ-aminobutyric acid (GABA) shunt], and the biosynthetic pathways that convert the primary metabolic precursors to sink metabolites. Also, it includes three metabolic compartments: cytosol, plastid and mitochondrion. The pathways in the model were assigned to specific compartments based on information in the current literature (see references, below). Some pathways could not be unequivocally assigned to a single compartment, since they are known to operate separately in multiple compartments. Thus, we included separate glycolysis and oxPPP pathways in the cytosol and plastid, as well as separate malic enzyme (Mal → Pyr) fluxes in the plastid and mitochondrion.

The sources of information for the (primary metabolic and biosynthetic) pathways in the model were the recent literature on soybean embryo or higher plant biochemistry (Breitkreuz and Shelp, 1995; Chollet *et al.*, 1996; Lam *et al.*, 1996; Casati *et al.*, 1999; Herman and Weaver, 1999; Singh, 1999; Drincovich *et al.*, 2001; Jeanneau *et al.*, 2002), plant biochemistry texts (Dey and Harborne, 1997, Lea and Leagood, 1999) and the online catalog Soybase (2004). These sources also provided information on the precursors of the sink metabolites were also obtained from the above sources. Stoichiometries and carbon atom rearrangements for the reactions were obtained from the KEGG encyclopedia (KEGG, 2004).
The reactions in the model were assumed reversible unless information on irreversibility was available. All reversible reactions were modeled as two fluxes (see Supplementary Material IV). The reaction from succinate to malate in the TCA cycle can lead to an inversion of the labeling pattern, owing to the fact that succinate is a symmetrical molecule while malate is not (Schmidt et al., 1999). To account for this fact, this reaction was modeled as two fluxes, one that conserves the carbon skeleton, and another that inverts the same.

The metabolic model also incorporated the observations reported in the previous sections. Specifically, the photosynthetic reactions (Calvin cycle) were not included in our model because carbon assimilation by external CO₂ fixation was found to be negligible. Because differences were observed between the isotopomeric compositions of the cytosolic and plastidic hexose nucleotide pools, we assumed separate glycolysis pathways and oxPPPs to operate in those compartments, with the cytosolic and plastidic G6P pools acting as precursors to the respective hexose nucleotide pools. To account for the difference in the isotopomeric compositions of Ser and T3P, we incorporated a reversible reaction between Ser and Gly, which is known to occur during the catabolism of Ser in heterotrophic plant tissues (Singh, 1999).

**Metabolic fluxes**

Fluxes in the above metabolic network were calculated from the measured isotopomeric abundances, extracellular fluxes and biomass composition, by using a flux evaluation mathematical routine that incorporated isotopomeric balancing and global optimization, as described in Materials and Methods and Supplementary Material IV. The objective of this routine was to calculate a set of stoichiometrically feasible fluxes that best accounts for the isotopomeric and extracellular flux measurements. The flux evaluation routine was implemented by a computer program, NMR2Flux.
To ensure that the iteratively evaluated flux solution is unique, flux evaluations were repeated at least 300 times from arbitrary starting points. Standard deviations for the fluxes and reversibility extents were computed from a statistical analysis. Herein, the errors in the measured multiplet intensities were used to perform 300 Monte Carlo estimations of fluxes (Press et al., 1992). This generated probability distributions of each flux. Standard deviations were calculated by fitting a normal distribution through these flux distributions. Two such distributions (for two anaplerotic fluxes) are shown in Supplementary Material IV.

The evaluated fluxes are listed in Figure 3.6, and also depicted in Figure 3.5 (arrow widths in this figure are directly proportional to flux). Figure 3.7 depicts the agreement between experimental multiplet intensities and those simulated from the evaluated fluxes. It can be seen that the evaluated fluxes explain the labeling data well. Only a few outlier points can be observed, of which the Leu δ1 intensities (shown with filled circles) were the largest contributors to the cumulative $\chi^2$ error between the simulated and experimental intensities.

Standard deviations for the fluxes and reversibility extents were computed from a statistical analysis. Herein, the errors in the measured multiplet intensities were used to perform 300 Monte Carlo estimations of fluxes (Press et al., 1992). This generated probability distributions of each flux. Standard deviations were calculated by fitting a normal distribution through these flux distributions. Two such distributions (for two anaplerotic fluxes) are shown in Supplementary Material IV.

The flux into the oxPPP was found to be $9.10 \pm 3.85 \text{ µmol d}^{-1}\text{ cotyledon}^{-1}$ in the cytosol, and $10.90 \pm 4.97 \text{ µmol d}^{-1}\text{ cotyledon}^{-1}$ in the plastid, the total flux being $19.99 \pm 4.39 \text{ µmol d}^{-1}\text{ cotyledon}^{-1}$. On a carbon mole basis, this is $104.2 \text{ carbon mol} \pm 23.0 \text{ carbon mol per 100 carbon mol of sucrose uptake.}$ Also, the flux of the hexose phosphate isomerase reaction is in the direction fructose-6-phosphate (F6P) $\rightarrow$ glucose-6-phosphate (G6P) in both the cytosol and the plastid. This indicates that the glycolysis and oxPPP are operating in a cyclic manner,
with the reverse hexose isomerase reaction feeding the oxPPP. Further, we were able to
distinguish between the fluxes through the reversible non-oxidative limbs of the oxPPP in the
cytosol and the plastid, which are catalyzed by transketolase and transaldolase. These fluxes
were observed to be substantial in the plastid (5.45 ± 1.50 μmol d⁻¹ cotyledon⁻¹) compared to
the cytosol (0.61 ± 0.25 μmol d⁻¹ cotyledon⁻¹). The flux from T3P to F6P was observed to be
21.72 ± 5.00 μmol d⁻¹ cotyledon⁻¹ in the plastid, and 0.31 ± 0.36 μmol d⁻¹ cotyledon⁻¹ in the
cytosol. The anaplerotic flux in the cytosol from PEP to OAA was observed to be 2.12 ± 0.31
μmol d⁻¹ cotyledon⁻¹, while the reverse flux from malate (Mal) to Pyr was 0.90 ± 0.41 μmol d⁻¹
cotyledon⁻¹ in the mitochondrion, and 0.57 ± 0.25 μmol d⁻¹ cotyledon⁻¹ in the plastid. The
glyoxylate shunt flux was 0.47 ± 0.03 μmol d⁻¹ cotyledon⁻¹, and therefore small compared to
the TCA cycle flux (through citrate synthase and aconitase) of 5.11 ± 1.33 μmol d⁻¹
cotyledon⁻¹. Most of the carbon in the TCA cycle appeared to be shunted through GABA
(confidence limits: 5.50 to 9.86 μmol d⁻¹ cotyledon⁻¹) rather than succinate thiokinase
(confidence limits: 0.0 to 2.56 μmol d⁻¹ cotyledon⁻¹). Also, the exchange fluxes between Gln,
Glu and α-ketoglutarate were found to be high.

Since our flux evaluation program NMR2Flux allowed alterations in the metabolic network
model easily, many modifications to the initially assumed model were examined for their
ability to accurately account for the labeling data. If a modification explained the data
significantly better, it was accepted. Such a posteriori changes to the model are the inclusion
of the flux from T3P to F6P (which resulted in a 50% reduction in the χ² error), and the
inclusion of an efflux into the medium from the succinate and malate nodes in the TCA
cycle. We included a reversible reaction between Ser and Gly, to account for the observed
isotopomeric difference between Ser and other glycolytic 3-carbon units. A photorespiratory
pathway was not required to account for this difference. The uptake of carbon from
glutamine had to be included to better account for the multiplet intensities of the amino acids
of the glutamate family, and Asp/Asn. Also, an efflux from the malate and succinate nodes of the TCA cycle was included. A contribution to Gly synthesis from glyoxylate (GOx) was included, the calculated flux of which (0.46 ± 0.24 μmol d⁻¹ cotyledon⁻¹) is comparable to that of the synthesis of Gly from 3-phosphoglycerate (3PG) (0.45 ± 0.24 μmol d⁻¹ cotyledon⁻¹).

Discussion

The use of steady state isotope labeling methods is a reliable technique to quantify fluxes in metabolic pathways. Its use in plant metabolism began with the measurement of ¹³C enrichments in tissues supplied with labeled substrate. The measurement of isotopomers (Glawischnig et al., 2000) was an important advance, since isotopomer abundances provide information on carbon-carbon connectivities (Szyperski, 1995; 1998), and hence an overdetermination of the labeling state compared to enrichments. However, the interpretation of observed isotopomer abundances has been largely qualitative or semi-quantitative.

In the present work, we calculated fluxes from overdetermined isotopomer abundance data. Fluxes are not explicit mathematical functions of the labeling data, therefore their evaluation from large isotopomer data sets is not trivial. Flux evaluation from isotopomer data requires isotopomer balancing and global optimization, which we employed in this work. To facilitate a quick and efficient flux evaluation, we incorporated recent developments in efficient flux analysis such as automated isotopomer balancing and global optimization using simulated annealing. A computer program, NMR2Flux, that incorporated the above methods was written to automatically evaluate fluxes from isotopomer data and a model of the metabolic network. All mathematical details are explained in Supplementary Material IV and V. To the extent of our knowledge, such a comprehensive flux analysis has been performed on a plant system for the first time.
Furthermore, the experimental methodology employed in this paper has the potential to become a high-throughput one, because the employment of 2-D NMR obviated the need to physically separate the sink metabolites to measure their isotopomer abundances. Thus, the experimental load was considerably reduced. In previous research that reported $^{13}\text{C}$ labeling measurements of several metabolites from plants (Glawischnig et al., 2000; Rontein et al., 2002; Schwender et al., 2003), the detected metabolites were separated by chromatography. Besides, 2-D $[^{13}\text{C}, ^1\text{H}]$ NMR enables the measurement of $^{13}\text{C}$ isotopomers with the high sensitivity of $^1\text{H}$ NMR rather than the relatively low sensitivity of $^{13}\text{C}$ NMR (Szyperski, 1998; Wiechert et al., 1999). The use of 2-D NMR is therefore crucial in the large-scale application of flux evaluation in plants.

The attainment of isotopic steady state is essential for the calculation of fluxes from the labeling data. For the in vitro soybean embryo culture employed here, the residence time of sucrose in the cells is approximately 9.4 h, as calculated from the uptake of sucrose ($1.179 \times 10^{-6}$ mol h$^{-1}$) and the free space sucrose concentration documented for developing soybean embryos (37 mM: Lichtner and Spanswick, 1981). Isotopic steady state is attained within 5 residence times (i.e. $< 48$ h or 2 d), which is less than the 6 d labeling period employed here. Therefore, the labeling data in this work can be assumed to be at steady state.

We observed that the measured isotopomer abundances of Leu $\delta^1$ did not agree with the simulated abundances of its precursor, Pyr$^0$, but other amino acids synthesized from Pyr$^0$ (Val $^\gamma^1$, Ile $^\gamma^2$) agreed well. Also, the Leu $\delta^1$ abundances were the largest contributors to the cumulative $\chi^2$ error between simulated and experimental abundances. We also found this anomaly in isotopomer data from soybean embryos cultured at other temperatures (Iyer VV, Sriram G, Shanks JV, unpublished data), and from Catharanthus roseus hairy roots (Sriram G, Shanks JV, unpublished data). This was not an artifact of the protein hydrolysis or NMR, since it has neither been reported in isotopomer abundance data from hydrolysates of protein
from prokaryotes (e.g. Szyperski, 1995) nor observed in our data from hydrolysates of 100% 
\(^{13}\)C-labeled prokaryotic protein. Also, this anomaly was not observed in other Leu atoms.

This anomaly suggests that the assumption of Pyr\(^p\) as a precursor to Leu \(\delta^1\), or the assumed
‘linear’ metabolic pathway from Pyr\(^p\) to Leu \(\delta^1\) is incorrect. However, there exists substantial
evidence that Pyr\(^p\) is the precursor of Leu \(\delta^1\) (Singh 1999). Also, the observed isotopomer
abundances Val \(\gamma^1\), Ile \(\gamma^2\), agreed with the simulated abundances of Pyr\(^p\). Therefore, we
conclude that the pathway between 2-oxoisovalerate and Leu (the only part of the pathway of branched amino acid synthesis from Pyr\(^p\) not shared by Val or Ile) may have other precursors
feeding into it, and needs further investigation.

We observed that the hexose nucleotide pools in the cytosol and plastid were not in isotopic
equilibrium. This result is supported by the finding that in \(D.\ carota\) cells, the \(^{13}\)C
enrichments of the carbon atoms of sucrose (synthesized from the cytosolic hexose phosphate
pool) and starch (synthesized from the plastidic hexose phosphate pool) were significantly
different throughout the period of labeling study (Krook et al., 1998). However, Keeling
(1991), Rontein et al. (2002) and Schwender et al. (2003) reported nearly similar
enrichments for sucrose and starch, which contrasts with our result. Therefore, the
equilibration of hexoses between the two compartments may be a function of metabolic
demand, and no general conclusion can be made.

On the other hand, our data showed that the T3P pools in the cytosol and plastid have the
same isotopomic composition and were not distinguishable. This has been observed
previously by Rontein et al. (2002) and Schwender et al. (2003). One possibility that
explains this result is that the two pools may be exchanging rapidly, i.e. they are in
equilibrium. Another possibility that could account for this result is the absence of enolase in
the plastid. Enolase catalyzes the conversion of 3-phosphoglycerate (3PG) to PEP, and its
absence has been reported in chloroplasts as well as non-photosynthetic plastids of various
species (Fischer et al., 1997). If it is absent in the plastid, the PEP (and/or pyruvate) required for plastidic biosynthesis may have to be manufactured in the cytosol. This would necessitate the export of T3P from the plastid to the cytosol (by the T3P/phosphate transporter) and the import of PEP in the opposite direction (by the PEP/phosphate transporter). This model suggests a single "lower" glycolytic pathway in many plastids.

Our data showed negligible photosynthetic carbon assimilation, although the cotyledons were green during culture. This agrees with $^{13}$C label data from Brassica napus embryos (Schwender and Ohlrogge, 2002), and also with studies by Chao et al. (1995) who found that developing soybean embryos that retained chlorophyll expressed the photosynthetic mRNAs Lhcβ and RbcS in insignificant amounts during the ‘filling’ period (period of storage protein accumulation). Also, studies by Eastmond and Rawsthorne (1998) concluded that the accumulation of storage products by embryos was largely heterotrophic with a minor photosynthetic contribution. Furthermore, we observed that a photorespiratory pathway was not necessary to account for the observed differences between the isotopomer abundances of Ser and other amino acids that reflect the T3P pool. Together, these results signify that green embryos are chloroplast-containing tissue functioning heterotrophically, and lacking any significant photosynthetic or related function.

We detected a substantial flux through the oxPPPs in the cytosol and plastid. In plants, the function of the oxPPP is believed to be two-fold: provision of reductant (particularly NADPH) in the plastid and hexose metabolism in the cytosol. In the plastid, the NADPH generated in the oxPPP is used for lipid and protein synthesis, nitrogen or glutamine assimilation, and combating oxidative stress (Hauschild and von Schaewen, 2003). Since NADPH does not cross membranes, the plastid requirement of NADPH must be generated within the same compartment. We determined the plastidic NADPH requirement and availability in our system, based on the evaluated fluxes, to be $22.28 \pm 0.22 \mu$mol d$^{-1}$
cotyledon\(^{-1}\). This includes the amounts needed for de novo amino acid and fatty acid synthesis, and for glutamine assimilation. The availability of NADPH from the plastidic oxPPP is 21.80 (± 9.93) μmol d\(^{-1}\) cotyledon\(^{-1}\). Thus, the plastidic oxPPP provides 98% (± 45%) of the NADPH requirement. This points to the oxPPP as a substantial contributor to the NADPH pool in the plastid. This is in concordance with previous studies that have found that the plastidic oxPPP is stimulated in response to high demands for NADPH (Hauschild and von Schaewen, 2003) and is coupled to glutamate assimilation (Esposito et al., 2003).

However, our value of NADPH availability from the oxPPP is higher than that calculated by Schwender et al. (2003) for developing B. napus embryos, who estimated the oxPPP to provide 22% to 45% of the plastidic requirement. This difference could be explained by the fact that B. napus embryos predominantly synthesize lipids, whereas G. max embryos synthesize both protein and lipids. The demand for reductant could be met differently in these systems. Further, the high glutamate assimilation in our system may have stimulated the oxPPP.

We were able to distinguish between the fluxes through the reversible non-oxidative limbs of the oxPPP in the cytosol and the plastid. These fluxes are catalyzed by transketolase and transaldolase, and were observed by us to be substantial in the plastid, and small or negligible in the cytosol. Ireland and Dennis (1980) have detected these enzymes in the plastid and cytosol of soybean nodules, indicating that the soybean genome may contain genes encoding plastidic and cytosolic isoforms of these enzymes. However the results reported here indicate that in our system, they are either not expressed or are not sufficiently active in the cytosol.

The compartmentation of these enzymes in plants has been subject to investigation recently, but still remains an open question. For example, Debnam and Emes (1999) found that in spinach and pea, transketolase or transaldolase activity was confined only to the plastid, whereas in tobacco these enzymes were found in both compartments. Further, Eicks et al.
(2002) found that all putative transketolase and transaldolase genes in Arabidopsis have a plastid-targeting sequence, and have suggested that exclusive confinement of transketolase and transaldolase to the plastid may be the case in higher plants. However, *Craterostigma plantagineum* contains both plastidic and cytosolic genes for these enzymes (Bernacchia et al., 1995), and is an exception. Nevertheless, the emerging general picture is one where transketolase and transaldolase operate only in the plastid (Kruger and von Schaewen, 2003). Interestingly, our results are consistent with this model. Since the function of transketolase and transaldolase enzymes is to convert the pentose phosphates formed in the oxPPP to hexose and triose phosphates and return them to glycolysis for further catabolism, it is therefore natural to see high fluxes through them in the plastid in our system, where the pentose phosphates formed in the oxPPP (during NADPH generation, see above) may have no major role.

One possible criticism of our result indicating separate transketolase and transaldolase fluxes in the cytosol and plastid, is that the cytosolic nucleotide-diphosphate sugars acting as precursors for protein glycosylation may not be equilibrating isotopically with the cytosolic hexose phosphate pool (and therefore not derived from cytosolic G6P as assumed in our model), but may be directly derived from sucrose synthase with little connection to the hexose phosphate pool. However, since large exchange fluxes have been reported between sucrose and the hexose phosphate pool (Dieuaide-Noubhani et al., 1995; Rontein et al., 2002), this criticism may not hold. Secondly, the fluxes calculated for cytosolic and plastidic transketolase and transaldolase, using the assumed model are consistent with the currently emerging model (substantial transketolase and transaldolase fluxes in the plastid and zero or negligible fluxes of these enzymes in the cytosol; see above) of the pentose phosphate pathway in plants (Kruger and von Schaewen, 2003).
We found that a flux from T3P to F6P had to be included in our model to account for the observed isotopomeric composition of starch. The oxPPP alone could not account for observed labeling pattern. A high T3P → F6P flux was detected in the plastid while a negligible flux was detected in cytosol. The F6P → T3P conversion catalyzed by phosphofructokinase is irreversible, and the only plastidic enzyme responsible a flux in the opposite direction is fructose-1,6-bisphosphatase. Interestingly, this enzyme is usually associated with photosynthetic plastids where it converts the photosynthate to starch for storage. Its existence has been ruled out in heterotrophic tissues, although it has been detected in pea embryos (Entwistle and ap Rees, 1990). Together with this result, our finding suggests an atypical role for plastidic fructose-1,6-bisphosphatase in embryo metabolism.

Plants contain anaplerotic enzymes catalyzing both directions of the PEP/Pyr → OAA/Mal conversion, and the plastidic Mal → Pyr conversion is thought to provide plastidic Pyr and/or NADPH toward biosynthesis. However, we found little cycling between these reactions, and the function of the anaplerotic fluxes appears to be replenishment of the small amount of OAA lost from the TCA cycle to Asp and Asn biosynthesis. We observed a rather small flux from Mal → Pyr in the plastid. This indicates that in our system, Mal does not substantially contribute carbon toward biosynthesis or NADPH availability.

The GABA shunt has been detected in soybean cotyledons previously (Breitkreuz and Shelp, 1995), and we found that it appears to be preferred over succinate thiokinase reaction of the TCA cycle. We also observed negligible flux through the glyoxylate shunt. This is natural to expect, since the glyoxylate enzymes in embryos are turned on only at the start of germination (Reynolds and Smith, 1995) and during leaf senescence. The glyoxylate shunt metabolizes acetate units resulting from the breakdown of lipids, and in a tissue that primarily accumulates lipids, its activity is anticipated to be insignificant.
Conclusions and future directions

In this work we performed $^{13}$C labeling experiments on developing soybean embryos and obtained exhaustive labeling data from sink metabolites by 2-D NMR. It was possible to quantify carbon partitioning through several metabolic processes including glycolysis, oxPPP, gluconeogenesis, anaplerotic pathways, TCA cycle, and the glyoxylate and GABA shunts. Furthermore we were also able to distinguish between fluxes in different compartments, based on labeling data of sink metabolites synthesized in separate compartments. The experimental methodology employed in this work has the potential to become a high-throughput one. Further reduction of the sample size and duration of labeling period should be possible, and could be optimized. The computer program developed to calculate fluxes from the labeling data is generic. We expect both these features to increase the applicability of flux analysis in plants.

As demonstrated here, flux analysis can provide insights on physiology and function. Comparison of fluxes between genetic or environmental variants can provide valuable information about the effects of genetic or environmental manipulations on the physiology. This is particularly relevant in the context of the recent upsurge in plant metabolic engineering (Hanson and Shanks, 2002). Together with high-throughput data on transcripts, proteins, and metabolites, systemic flux data can provide the basis for understanding the functioning of plants from a systems biology perspective (Sweetlove et al., 2003). Work is underway in our laboratory to evaluate fluxes in soybean embryos grown in different environments, and in another plant system. Using recent theoretical developments on flux identifiability (Isermann and Wiechert, 2003) we are also working toward designing labeling experiments on plant systems with judicious combinations of labeled substrates so as to increase the flux information available from them.
Materials and methods

Soybean cotyledon culture

Soybean (Glycine max cv. Evans) was grown in a growth chamber at 27 °C/20 °C and 14 h photoperiod. 18 days after flowering, pods were harvested from the central section of the main stem and embryos isolated for in vitro culture. Three cotyledons were selected for uniform initial size (100 to 120 mg fresh weight) and cultured aseptically, in 20 mL liquid medium containing 146 mM sucrose (10% w/w U-13C, 90% w/w commercial, with a natural 13C abundance of 1.1%), 37 mM glutamine (commercial, natural 13C abundance of 1.1%) as the only carbon sources. This labeling technique is termed ‘biosynthetically directed fractional 13C labeling’ (Szyperski, 1995). U-13C sucrose was purchased from Isotec (Miamisburg, OH). The in vitro culture was maintained at 25°C, 100 rpm and approximately 100 μE m⁻² s⁻¹ light intensity. After 6 d of culture, cotyledons were harvested, rinsed with non-labeled medium, and lyophilized at -50°C and 133 ×10⁻³ mbar for 72 h. The lyophilized embryos were finely ground for further processing.

Protein extraction, hydrolysis, amino acid quantification, and NMR sample preparation

Protein was extracted from ground samples in 100 mM phosphate buffer (pH 7.2), at 4 °C for 15 min. The extract was repeated four times, and the consolidated supernatant was assayed for protein using the Bradford test (Bio-Rad, Hercules, CA).

Protein hydrolysis was performed in hydrolysis tubes (Pierce Endogen, Rockford, IL), to which 6N hydrochloric acid was added in the ratio 0.5 mL HCl:400 μg protein. The hydrolysis tube was evacuated, flushed with nitrogen to remove residual oxygen, and re-evacuated. Hydrolysis was performed at 150°C for 4 h. The acid in the hydrolysate was evaporated in a Rapidvap evaporator (Labconco, Kansas City, MO). The residue was re-dissolved in 2 mL deionized water, lyophilized for 72 h, and dissolved in 500 μL D₂O in an
NMR tube. The pH of the NMR sample was adjusted to 0.5 using DCI. Amino acids in the sample were quantified by HPLC, after derivatization with phenylisothiocyanate to produce phenylthiocarbamyl amino acid derivatives, which were eluted by a reverse phase C18 silica column, with detection at 254 nm.

**Extracellular fluxes and fluxes contributing to biomass**

Biomass growth was quantified by measuring embryo fresh weight. Lipids were extracted in hexane at 45°C and quantified by weight. Sucrose consumption was measured by using HPLC.

**NMR spectroscopy**

2-D \[^{13}C, ^1H\] heteronuclear single quantum correlation (HSQC) NMR spectra (Bodenhausen and Ruben, 1980) were collected on a Bruker Avance DRX 500 MHz spectrometer, at 298 K. The reference to zero ppm was set using the methyl signal of dimethylsilapentanesulfonate (Sigma, St. Louis, MO) as an internal standard. The resonance frequency of \(^{13}C\) and \(^1H\) were 125.7 MHz and 499.9 MHz respectively. The spectral width was 5482.26 Hz along the \(^1H\) (F2) dimension and 5028.05 Hz along the \(^{13}C\) (F1) dimension. Peak aliasing was used in order to minimize the sweep width along the F1 dimension. The number of complex data points was 1024 \((^1H) \times 900\) \((^{13}C)\). A modification of the INEPT pulse sequence was used for acquiring HSQC spectra (Bodenhausen and Ruben, 1980). The number of scans was generally set to 16. Assignment of amino acid peaks on the HSQC spectrum was verified using 2D \[^1H, ^1H\] total correlation (TOCSY) and 3D \[^{13}C, ^1H, ^1H\] TOCSY spectra (Braunschweiler and Ernst, 1983), which were acquired with a 100% labeled protein sample. While acquiring TOCSY spectra, the DIPSI-2 sequence (Shaka et al., 1988) was used for isotropic mixing, with a mixing time of 76 ms.
The software Xwinnmr™ (Bruker) was used to acquire all spectra, and the software NMRView™ (Johnson and Blevins, 1994; available at http://onemoonscientific.com/nmrview, last accessed on 20 July 2004) was used to quantify non-overlapping multiplets on the HSQC spectrum. To quantify overlapping multiplets (α-amino acid and LVA peaks), which could not be processed with NMRView, a peak deconvolution software was written. This software was based on a spectral model proposed by van Winden et al. (2001). Additionally, 2-D spectra were obtained that were J-scaled along the F1 dimension, by incorporating pulse sequences described by Willker et al. (1997) and Brown (1984) into the HSQC pulse sequence. J-scaling increases multiplet separation by an even integral factor J and eliminates multiplet overlap. J-scaling factors of 6 or 8 were employed.

**Flux evaluation methodology and computer program**

Fluxes were evaluated from isotopomer data by using isotopomer balancing and a global routine. All mathematical details are presented in Supplementary Material IV and V. The objective of this flux evaluation procedure is to guess a set of stoichiometrically feasible fluxes (per the metabolic network supplied by the user) that best accounts for the measured isotopomer abundances. The computer program that evaluates fluxes, NMR2Flux, is implemented in the programming language C, on the Red Hat Linux operating system.

Upon request, all novel materials described in this paper (flux evaluation program NMR2Flux, modified NMR pulse sequences, and peak deconvolution software) will be made available in a timely manner for non-commercial research purposes, subject to the requisite permission from any third-party owners of all or parts of the material. Obtaining any permissions will be the responsibility of the requestor.
Acknowledgments

The authors wish to thank Dr. Eve Wurtele (Department of Botany, Iowa State University) for helpful discussions on this work and for her comments on the manuscript. Dr. Amy Andreotti (Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University) for a gift of 100% $^{13}$C-labeled protein; and Dr. Louisa B. Tabatabai (Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University) for consultations on protein hydrolysis.

Literature cited


Bodenhausen G, Ruben DJ (1980) Natural abundance nitrogen-15 NMR by enhanced


storage proteins of *Glycine max* and *Phaseolus vulgaris*. J Biol Chem 261: 9228-9238


Metabolic flux in developing maize kernels. Plant Physiol 125: 1178-1186


Ireland RJ, Dennis DT (1980) Isozymes of the glycolytic and pentose-phosphate pathways in storage tissues of different oilseeds. Planta 149: 476-479


Wüthrich K (1976) NMR in biological research: peptides and proteins. North Holland, Amsterdam

Figure captions

Figure 3.1. Isotopomers (‘isotope isomers’) of a metabolite. Five isomers (i, ii, iii, iv, v) of the 6-carbon metabolite glucose-6-phosphate (G6P) are shown. These isomers differ in the labeling state ($^{13}$C or $^{12}$C) of their individual carbon atoms. Here, $^{13}$C atoms are depicted as filled circles, and $^{12}$C atoms are depicted as hollow circles. Isotopomers are notated using boldface for $^{13}$C, regular font for $^{12}$C, as illustrated on the right of each isotopomer. A metabolite with $n$ carbon atoms has $2^n$ isotopomers, but only $n$ positional enrichment measurements. Therefore isotopomer measurements usually provide superior information about the labeling state of a metabolite, compared to positional $^{13}$C enrichments.

Figure 3.2. 2-D [$^{13}$C, $^1$H] HSQC spectrum of protein hydrolysate. Protein was isolated from soybean cotyledons cultured on sucrose (10% w/w U-$^{13}$C) and glutamine. Cross peaks represent carbon atoms of hydrolysate constituents (proteinogenic amino acids, HMF, LVA). The names of some amino acid nuclei are omitted for clarity.

Figure 3.3. Expanded views of [$^{13}$C, $^1$H] HSQC spectrum: (A) Gly α and (B) Asp β cross peaks. 1-D slices are shown alongside. The multiplet peaks are s: singlet; d, d1, d2: doublet; dd: double doublet.

Figure 3.4. (A) Comparison of corresponding multiplet intensities of levulinic acid (LVA) and 5-hydroxymethylfurfural (HMF) from [$^{13}$C, $^1$H] spectra of protein (hollow bars) and starch (filled bars) hydrolysates. Each peak represents an isotopomer of the cytosolic (protein hydrolysate) or the plastidic (starch hydrolysate) hexose nucleotide pool. (See Results for details.) These precursor isotopomers are shown next to the x-axis, following the notation introduced in Results. Starch hydrolysate multiplets whose multiplet intensities are significantly different from the corresponding protein hydrolysate multiplets, are marked with ‘*’. (B) Comparison of corresponding multiplet intensities of Ala α, Phe α, Tyr α, Ser
Figure 3.5. Metabolic flux map template consisting of primary and intermediate metabolic pathways in developing *Glycine max* (soybean) cotyledons cultured on sucrose (10% w/w U-\[^{13}\text{C}\]) and glutamine. Fluxes are proportional to arrow widths. ( Fluxes less than 0.4 \(\mu\text{mol \text{d}^{-1}}\) cotyledon\(^{-1}\) are shown with dashed lines.) Arrows indicate the direction of net flux. A complete numerical listing of estimated flux values and reaction reversibilities is provided in Figure 3.5. Intracellular metabolites are shown in white ovals, and gray ovals show sink metabolites (proteinogenic amino acids, polysaccharides, etc.) Metabolites taken up from/secreted into the medium are shown in blue ovals. Metabolic pathways are color-coded as follows: dark red (glycolysis and sucrose metabolism), pale blue (pentose phosphate pathway), orange (TCA cycle), blue-gray (pyruvate dehydrogenase link), mauve (anaplerotic fluxes), dark yellow (glyoxylate shunt), green (glutamate metabolism, GABA shunt, and associated intercompartmental transport fluxes), gray (fluxes towards biomass synthesis, and black (all intercompartmental transport fluxes except those involved in Glu metabolism and GABA shunt). Abbreviations for intracellular metabolites are as follows: Sue (sucrose), G6P (glucose-6-phosphate), F6P (fructose-6-phosphate), T3P (triose-3-phosphate), P5P (pentose-5-phosphate), S7P (sedoheptulose-7-phosphate), E4P (erythrose-4-phosphate), 3PG (3-phosphoglycerate), PEP (phosphoenolpyruvate), Pyr (pyruvate), acetCoA (acetyl CoA), iCit (isocitrate), aKG (\(\alpha\)-ketoglutarate), Scn (succinate), Mal (malate), OAA (oxaloacetate), GABA (\(\gamma\)-aminobutyric acid), SSA (succinic semi-aldehyde), GOx (glyoxylate). Sink metabolite abbreviations are as follows: PSac (polysaccharides), Nuc (carbon skeleton of nucleotides), Sta (starch). Asp and Asn are denoted together as Asx. Glu and Gln are denoted together as Glx. F6P and T3P appear at two different locations each in the cytosol and
plastid, to avoid confusing intersections of lines. Each flux is assigned a short name based on the name of the gene encoding one of the metabolic reactions represented by it. Intracellular metabolites and fluxes with a superscript are located in specific subcellular compartments – c: cytosol, p: plastid, m: mitochondrion. If a flux has no superscript, its compartmentation could not be unambiguously determined (such as gap, eno and pyk, and some fluxes toward biosynthesis).

Figure 3.6. Metabolic fluxes evaluated for soybean embryos cultured on sucrose (10% w/w U-13C) and glutamine. Absolute fluxes are expressed in μmol d⁻¹ cotyledon⁻¹. Each flux is assigned a short name (usually based on the name of the gene encoding one of the metabolic reactions represented by the flux). Abbreviations for metabolites and color coding for fluxes is the same as in Figure 3.4. Fluxes grouped by ‘{’ are those of forward and backward reactions catalyzed by the same enzyme (The forward and backward fluxes F6P → T3P and T3P → F6P are not grouped thus, since they are catalyzed by different enzymes.) For grouped fluxes, the net flux and reaction reversibility are shown. Net fluxes are in the direction of the reaction with the subscript ‘f’. Subscripts on reaction names: f: forward reaction, b: backward reaction. Superscripts: c: cytosol, p: plastid, m: mitochondrion. If a flux has no superscript, its compartmentation could not be unambiguously determined. Reversibilities are reported as percentages, with 0%=irreversible, 100%=reaction at equilibrium. irrev.: reaction assumed irreversible/known to be thermodynamically irreversible; n.d.: reversibility of the reaction could not be determined as it was found to have negligible effect on the isotopomer abundances; SD: standard deviation of flux or reversibility.

Figure 3.7. Comparison of experimental and simulated multiplet intensities, depicting how closely the evaluated fluxes account for the labeling data (multiplet intensities or isotopomer abundances). The x-axis represents experimental multiplet intensities measured from [13C,
$^1$H] spectra; the y-axis represents relative intensities that were simulated by the computer program NMR2Flux, corresponding to the evaluated fluxes (Figure 3.5). Multiplet intensities are shown as fraction of total signal. The thick diagonal line is the 45° diagonal, on which the error between measurement and simulation is zero. The thin lines enclose 90% of all data points (all points with error ≤ 0.0434). The singlet and doublet intensities of Leu δ$^1$ are shown as filled circles.
### Figure 3.1. Isotopomers (‘isotope isomers’) of a metabolite.

<table>
<thead>
<tr>
<th>Isotopomers</th>
<th>Atom number</th>
<th>Isotopomer notation</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>1 2 3 4 5 6</td>
<td>G6P[123456]</td>
</tr>
<tr>
<td>ii</td>
<td>●●●●●●</td>
<td>G6P[123456]</td>
</tr>
<tr>
<td>iii</td>
<td>●●●●●●</td>
<td>G6P[123456]</td>
</tr>
<tr>
<td>iv</td>
<td>1 2 3 4 5 6</td>
<td>G6P[123456]</td>
</tr>
<tr>
<td>v</td>
<td>●●●●●●</td>
<td>G6P[123456]</td>
</tr>
</tbody>
</table>
Figure 3.2. 2-D $[^{13}\text{C}, ^{1}\text{H}]$ HSQC spectrum of protein hydrolysate.
Figure 3.3. Expanded views of $^{13}\text{C}, ^1\text{H}$ HSQC spectrum.
Figure 3.4. Comparison of corresponding multiplet intensities of metabolites from cytosol and plastid.
Figure 3.5. Metabolic flux map template consisting of primary and intermediate metabolic pathways in developing *Glycine max* (soybean) cotyledons cultured on sucrose (10% w/w U-$^{13}$C) and glutamine.
Figure 3.6. Metabolic fluxes evaluated for soybean embryos cultured on sucrose (10% w/w U-$^{13}$C) and glutamine.
### Anaplerotic reactions

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Direction</th>
<th>Flux (μmol/gDW·h)</th>
<th>Irreversibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEP + CO₂ → OAA²</td>
<td>2.12</td>
<td>2.18</td>
<td>irrev.</td>
</tr>
<tr>
<td>Malᵐ → Pyrᵐ + CO₂</td>
<td>0.90</td>
<td>0.41</td>
<td>irrev.</td>
</tr>
<tr>
<td>Malᵢ → Pyrᵢ + CO₂</td>
<td>0.57</td>
<td>0.25</td>
<td>irrev.</td>
</tr>
</tbody>
</table>

### GABA shunt

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Direction</th>
<th>Flux (μmol/gDW·h)</th>
<th>Irreversibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gluᵃ → GABAᵃ + CO₂</td>
<td>7.68</td>
<td>2.18</td>
<td>irrev.</td>
</tr>
<tr>
<td>GABAᵃ → GABAᵐ</td>
<td>7.68</td>
<td>2.18</td>
<td>n. d.</td>
</tr>
<tr>
<td>GABAᵐ → SSAᵐ</td>
<td>7.68</td>
<td>2.18</td>
<td>49.4</td>
</tr>
<tr>
<td>SSAᵐ → Scnᵐ</td>
<td>7.68</td>
<td>2.18</td>
<td>n. d.</td>
</tr>
</tbody>
</table>

### Substrate entry

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Direction</th>
<th>Flux (μmol/gDW·h)</th>
<th>Irreversibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sucᵃ → G6P² + F6P²</td>
<td>9.59</td>
<td>0.10</td>
<td>n. d.</td>
</tr>
<tr>
<td>Glnʰ → Gln³</td>
<td>4.76</td>
<td>0.23</td>
<td>23.6</td>
</tr>
<tr>
<td>Glu³ → Gluᵐ</td>
<td>0.00</td>
<td>0.00</td>
<td>3.4</td>
</tr>
</tbody>
</table>

### Glutamate assimilation

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Direction</th>
<th>Flux (μmol/gDW·h)</th>
<th>Irreversibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gin¹ → Gin³</td>
<td>4.76</td>
<td>0.23</td>
<td>n. d.</td>
</tr>
<tr>
<td>aKGl¹ → aKGl³</td>
<td>4.76</td>
<td>0.23</td>
<td>n. d.</td>
</tr>
<tr>
<td>aKGl³ → aKGl⁵</td>
<td>4.76</td>
<td>0.23</td>
<td>n. d.</td>
</tr>
<tr>
<td>gogatp → aKGl³ + Gin³ → Glu³ + Glu⁵</td>
<td>88.52</td>
<td>43.83</td>
<td>irrev.</td>
</tr>
<tr>
<td>gdhfp → Glu³ → aKGl³</td>
<td>83.94</td>
<td>43.83</td>
<td>99.0</td>
</tr>
<tr>
<td>gdhdp → aKGl³ → Glu³</td>
<td>0.00</td>
<td>0.00</td>
<td>99.0</td>
</tr>
<tr>
<td>asf → Glu³</td>
<td>83.68</td>
<td>43.83</td>
<td>61.0</td>
</tr>
<tr>
<td>asb → Glu³</td>
<td>0.00</td>
<td>0.00</td>
<td>28.1</td>
</tr>
</tbody>
</table>

### Malate shuttle

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Direction</th>
<th>Flux (μmol/gDW·h)</th>
<th>Irreversibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>malT₁ → Malʰ → Mal↑</td>
<td>-0.80</td>
<td>0.44</td>
<td>96.5</td>
</tr>
<tr>
<td>malT₂ → Malʰ → Mal↑</td>
<td>1.32</td>
<td>0.44</td>
<td>n. d.</td>
</tr>
<tr>
<td>mdh → OAA³</td>
<td>2.12</td>
<td>0.31</td>
<td>n. d.</td>
</tr>
<tr>
<td>mdh → OAA⁵</td>
<td>0.75</td>
<td>&lt;0.08</td>
<td>n. d.</td>
</tr>
</tbody>
</table>

### Biosynthesis of Ser and Gly

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Direction</th>
<th>Flux (μmol/gDW·h)</th>
<th>Irreversibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>3PG → Ser</td>
<td>1.17</td>
<td>&lt;0.08</td>
<td>irrev.</td>
</tr>
<tr>
<td>Ser → Gly + Cl⁻</td>
<td>0.45</td>
<td>0.24</td>
<td>46.4</td>
</tr>
<tr>
<td>Gly + Cl⁻ → Ser</td>
<td>0.00</td>
<td>0.00</td>
<td>1.2</td>
</tr>
<tr>
<td>GOx → Gly</td>
<td>0.46</td>
<td>0.24</td>
<td>irrev.</td>
</tr>
</tbody>
</table>

### Biosynthesis of Ser and Gly (other than those related to Ser and Gly) and effluxes into medium

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Direction</th>
<th>Flux (μmol/gDW·h)</th>
<th>Irreversibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO₂ → resp</td>
<td>49.81</td>
<td>2.56</td>
<td>irrev.</td>
</tr>
<tr>
<td>G6P → biomass</td>
<td>1.99</td>
<td>0.29</td>
<td>irrev.</td>
</tr>
<tr>
<td>G6P → biomass</td>
<td>0.83</td>
<td>0.27</td>
<td>irrev.</td>
</tr>
<tr>
<td>F6P → biomass</td>
<td>0.03</td>
<td>0.03</td>
<td>irrev.</td>
</tr>
<tr>
<td>P5P → biomass</td>
<td>2.08</td>
<td>0.16</td>
<td>irrev.</td>
</tr>
<tr>
<td>P5P → biomass</td>
<td>0.12</td>
<td>&lt;0.08</td>
<td>irrev.</td>
</tr>
<tr>
<td>E4P → biomass</td>
<td>0.37</td>
<td>&lt;0.08</td>
<td>irrev.</td>
</tr>
<tr>
<td>E4P → biomass</td>
<td>0.74</td>
<td>&lt;0.08</td>
<td>irrev.</td>
</tr>
<tr>
<td>ACA² → biomass</td>
<td>10.95</td>
<td>&lt;0.08</td>
<td>irrev.</td>
</tr>
<tr>
<td>ACA² → biomass</td>
<td>0.53</td>
<td>&lt;0.08</td>
<td>irrev.</td>
</tr>
<tr>
<td>Pyr → biomass</td>
<td>2.76</td>
<td>&lt;0.08</td>
<td>irrev.</td>
</tr>
<tr>
<td>OAA → biomass</td>
<td>1.50</td>
<td>&lt;0.08</td>
<td>irrev.</td>
</tr>
<tr>
<td>Glu → biomass</td>
<td>1.59</td>
<td>&lt;0.08</td>
<td>irrev.</td>
</tr>
<tr>
<td>Pyr → Pyr</td>
<td>0.17</td>
<td>0.12</td>
<td>irrev.</td>
</tr>
<tr>
<td>Malᵢ → Malᵢ</td>
<td>2.32</td>
<td>0.53</td>
<td>irrev.</td>
</tr>
<tr>
<td>Scnᵢ → Scnᵢ</td>
<td>0.00</td>
<td>0.00</td>
<td>irrev.</td>
</tr>
</tbody>
</table>

**Figure 3.6.** (Continued.)
Figure 3.7. Comparison of experimental and simulated multiplet intensities.
Supplementary material I: Multiplet intensities From 2-D \([^{13}\text{C}, ^1\text{H}]\)

HSQC

Table I-1. Relative multiplet intensities of carbon nuclei, measured from 2-D HSQC spectrum of protein hydrolysate; protein was isolated from soybean cotyledons. The detected cross peaks correspond to carbon atoms of hydrolysis products of sink metabolites (proteinogenic amino acids, HMF and LVA). Cross peaks are grouped according to the precursors of central carbon metabolism that they originate from. The sink metabolite precursors are shown and are color-coded for elucidatory purposes. Their abbreviations are explained in the caption of Figure 3.5. A superscript on a precursor indicates that the precursor is from a specific subcellular compartment (c: cytosol, p: plastid, m: mitochondrion); the absence of a subscript indicates that the sink metabolite originates from a precursor from more than one compartment.

Multiplets are abbreviated as follows. s: singlet, d: doublet, dd: double doublet, t: triplet, qd: quadruple doublet. The isotopomer structure (of the precursor molecule) corresponding to the multiplet is shown in the right-most column, per the following convention: boldface denotes \(^{13}\text{C}\) atoms, regular font denotes \(^{12}\text{C}\) atoms. Only atoms of the precursor whose labeling is discernible from the amino acid atom, are numbered; other atoms are indicated as ‘x’. Some amino acid atoms emanate from multiple molecules of the same precursor, or from multiple precursors. The isotopomer structures of all such precursor(s) are indicated in those cases. When more than one multiplet of the same type are present, they are numbered (e.g. d1, d2).

Relative intensities of multiplets are depicted as fractions of the total \(^{13}\text{C}\) signal for an atom, and were measured as described in Materials and Methods. The standard deviations (SD) of the intensities represent the larger of the errors from the signal-to-noise ratio of the HSQC spectrum, or differences between intensities evaluated from duplicate HSQC spectra.
Table I-1. (Continued.)

<table>
<thead>
<tr>
<th>Hydrolysate</th>
<th>Cross peak (multiplet)</th>
<th>Intensity</th>
<th>SD Precursor</th>
<th>Isotopomer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein</td>
<td>Gly α (s)</td>
<td>0.458</td>
<td>0.005 T3P/GOx</td>
<td>[12x]</td>
</tr>
<tr>
<td>Protein</td>
<td>Gly α (d)</td>
<td>0.542</td>
<td>0.005 T3P/GOx</td>
<td>[12x]</td>
</tr>
<tr>
<td>Protein</td>
<td>Ser α (s)</td>
<td>0.272</td>
<td>0.008 Ser</td>
<td>[123]</td>
</tr>
<tr>
<td>Protein</td>
<td>Ser α (d1)</td>
<td>0.211</td>
<td>0.008 Ser</td>
<td>[123]</td>
</tr>
<tr>
<td>Protein</td>
<td>Ser α (d2)</td>
<td>0.127</td>
<td>0.008 Ser</td>
<td>[123]</td>
</tr>
<tr>
<td>Protein</td>
<td>Ser α (dd)</td>
<td>0.390</td>
<td>0.008 Ser</td>
<td>[123]</td>
</tr>
<tr>
<td>Protein</td>
<td>Ser β (s)</td>
<td>0.497</td>
<td>0.015 Ser</td>
<td>[x23]</td>
</tr>
<tr>
<td>Protein</td>
<td>Ser β (d)</td>
<td>0.503</td>
<td>0.015 Ser</td>
<td>[x23]</td>
</tr>
<tr>
<td>Protein</td>
<td>Ala α (s)</td>
<td>0.196</td>
<td>0.010 Pyr</td>
<td>[123]</td>
</tr>
<tr>
<td>Protein</td>
<td>Ala α (d1)</td>
<td>0.023</td>
<td>0.010 Pyr</td>
<td>[123]</td>
</tr>
<tr>
<td>Protein</td>
<td>Ala α (d2)</td>
<td>0.206</td>
<td>0.010 Pyr</td>
<td>[123]</td>
</tr>
<tr>
<td>Protein</td>
<td>Ala α (dd)</td>
<td>0.573</td>
<td>0.010 Pyr</td>
<td>[123]</td>
</tr>
<tr>
<td>Protein</td>
<td>Ala β (s)</td>
<td>0.244</td>
<td>0.002 Pyr</td>
<td>[x23]</td>
</tr>
<tr>
<td>Protein</td>
<td>Ala β (d)</td>
<td>0.756</td>
<td>0.002 Pyr</td>
<td>[x23]</td>
</tr>
<tr>
<td>Protein</td>
<td>Ile γ2 (s)</td>
<td>0.242</td>
<td>0.002 Pyr</td>
<td>[x23]</td>
</tr>
<tr>
<td>Protein</td>
<td>Ile γ2 (d)</td>
<td>0.759</td>
<td>0.002 Pyr</td>
<td>[x23]</td>
</tr>
<tr>
<td>Protein</td>
<td>Leu δ1 (s)</td>
<td>0.357</td>
<td>0.005 Pyr</td>
<td>[x23]</td>
</tr>
<tr>
<td>Protein</td>
<td>Leu δ1 (d)</td>
<td>0.643</td>
<td>0.005 Pyr</td>
<td>[x23]</td>
</tr>
<tr>
<td>Protein</td>
<td>Val α (s)</td>
<td>0.354</td>
<td>0.027 Pyr</td>
<td>[12x][x2x]</td>
</tr>
<tr>
<td>Protein</td>
<td>Val α (d1)</td>
<td>0.558</td>
<td>0.027 Pyr</td>
<td>[12x][x2x]</td>
</tr>
<tr>
<td>Protein</td>
<td>Val α (d2)</td>
<td>0.053</td>
<td>0.027 Pyr</td>
<td>[12x][x2x]</td>
</tr>
<tr>
<td>Protein</td>
<td>Val α (dd)</td>
<td>0.034</td>
<td>0.027 Pyr</td>
<td>[12x][x2x]</td>
</tr>
<tr>
<td>Protein</td>
<td>Val γ1 (s)</td>
<td>0.235</td>
<td>0.002 Pyr</td>
<td>[x23]</td>
</tr>
<tr>
<td>Protein</td>
<td>Val γ1 (d)</td>
<td>0.765</td>
<td>0.002 Pyr</td>
<td>[x23]</td>
</tr>
<tr>
<td>Protein</td>
<td>Val γ2 (s)</td>
<td>0.882</td>
<td>0.002 Pyr</td>
<td>[x2x][x3x]</td>
</tr>
<tr>
<td>Protein</td>
<td>Val γ2 (d)</td>
<td>0.118</td>
<td>0.002 Pyr</td>
<td>[x2x][x3x]</td>
</tr>
<tr>
<td>Protein</td>
<td>Leu δ2 (s)</td>
<td>0.877</td>
<td>0.001 Pyr</td>
<td>[x2x][x3x]</td>
</tr>
<tr>
<td>Protein</td>
<td>Leu δ2 (d)</td>
<td>0.123</td>
<td>0.001 Pyr</td>
<td>[x2x][x3x]</td>
</tr>
<tr>
<td>Protein</td>
<td>Phe α (s)</td>
<td>0.205</td>
<td>0.012 PEP</td>
<td>[123]</td>
</tr>
<tr>
<td>Protein</td>
<td>Phe α (d1)</td>
<td>0.035</td>
<td>0.012 PEP</td>
<td>[123]</td>
</tr>
<tr>
<td>Protein</td>
<td>Phe α (d2)</td>
<td>0.151</td>
<td>0.012 PEP</td>
<td>[123]</td>
</tr>
<tr>
<td>Protein</td>
<td>Phe α (dd)</td>
<td>0.608</td>
<td>0.012 PEP</td>
<td>[123]</td>
</tr>
<tr>
<td>Protein</td>
<td>Phe β (s)</td>
<td>0.161</td>
<td>0.011 PEP</td>
<td>[x23][x2x]</td>
</tr>
<tr>
<td>Protein</td>
<td>Phe β (d1)</td>
<td>0.750</td>
<td>0.011 PEP</td>
<td>[x23][x2x]</td>
</tr>
<tr>
<td>Protein</td>
<td>Phe β (d2)</td>
<td>0.014</td>
<td>0.011 PEP</td>
<td>[x23][x2x]</td>
</tr>
</tbody>
</table>
Table I-1. (Continued.)

<table>
<thead>
<tr>
<th>Hydrolysate</th>
<th>Cross peak (multiplet)</th>
<th>Intensity</th>
<th>SD</th>
<th>Precursor</th>
<th>Isotopomer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein</td>
<td>Phe β (dd)</td>
<td>0.074</td>
<td>0.011</td>
<td>PEP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x23]/[x2x]</td>
</tr>
<tr>
<td>Protein</td>
<td>Tyr α (s)</td>
<td>0.208</td>
<td>0.017</td>
<td>PEP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[123]</td>
</tr>
<tr>
<td>Protein</td>
<td>Tyr α (d1)</td>
<td>0.000</td>
<td>0.000</td>
<td>PEP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[123]</td>
</tr>
<tr>
<td>Protein</td>
<td>Tyr α (d2)</td>
<td>0.180</td>
<td>0.017</td>
<td>PEP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[123]</td>
</tr>
<tr>
<td>Protein</td>
<td>Tyr α (dd)</td>
<td>0.612</td>
<td>0.017</td>
<td>PEP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[123]</td>
</tr>
<tr>
<td>Protein</td>
<td>Tyr β (s)</td>
<td>0.155</td>
<td>0.028</td>
<td>PEP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x23]/[x2x]</td>
</tr>
<tr>
<td>Protein</td>
<td>Tyr β (d)</td>
<td>0.737</td>
<td>0.028</td>
<td>PEP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x23]/[x2x] + [x23]/[x2x]</td>
</tr>
<tr>
<td>Protein</td>
<td>Tyr β (l)</td>
<td>0.109</td>
<td>0.028</td>
<td>PEP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x23]/[x2x]</td>
</tr>
<tr>
<td>Protein</td>
<td>Leu α (s)</td>
<td>0.225</td>
<td>0.009</td>
<td>ACoAP/Pyr&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[12]/[x2x]</td>
</tr>
<tr>
<td>Protein</td>
<td>Leu α (d1)</td>
<td>0.709</td>
<td>0.009</td>
<td>ACoAP/Pyr&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[12]/[x2x]</td>
</tr>
<tr>
<td>Protein</td>
<td>Leu α (d2)</td>
<td>0.032</td>
<td>0.009</td>
<td>ACoAP/Pyr&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[12]/[x2x]</td>
</tr>
<tr>
<td>Protein</td>
<td>Leu α (dd)</td>
<td>0.033</td>
<td>0.009</td>
<td>ACoAP/Pyr&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[12]/[x2x]</td>
</tr>
<tr>
<td>Protein</td>
<td>Leu β (s)</td>
<td>0.815</td>
<td>0.007</td>
<td>ACoAP/Pyr&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x2]/[x2x]/[x2x]</td>
</tr>
<tr>
<td>Protein</td>
<td>Leu β (d)</td>
<td>0.185</td>
<td>0.007</td>
<td>ACoAP/Pyr&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x2]/[x2x]/[x2x] + [x2]/[x2x]/[x2x]</td>
</tr>
<tr>
<td>Protein</td>
<td>Leu β (l)</td>
<td>0.000</td>
<td>0.007</td>
<td>ACoAP/Pyr&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x2]/[x2x]/[x2x]</td>
</tr>
<tr>
<td>Protein</td>
<td>His α (s)</td>
<td>0.242</td>
<td>0.025</td>
<td>P5PP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xx345]</td>
</tr>
<tr>
<td>Protein</td>
<td>His α (d1)</td>
<td>0.000</td>
<td>0.000</td>
<td>P5PP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xx345]</td>
</tr>
<tr>
<td>Protein</td>
<td>His α (d2)</td>
<td>0.162</td>
<td>0.025</td>
<td>P5PP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xx345]</td>
</tr>
<tr>
<td>Protein</td>
<td>His α (dd)</td>
<td>0.596</td>
<td>0.025</td>
<td>P5PP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xx345]</td>
</tr>
<tr>
<td>Protein</td>
<td>His β (s)</td>
<td>0.321</td>
<td>0.007</td>
<td>P5PP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x234x]</td>
</tr>
<tr>
<td>Protein</td>
<td>His β (d1)</td>
<td>0.019</td>
<td>0.008</td>
<td>P5PP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x234x]</td>
</tr>
<tr>
<td>Protein</td>
<td>His β (d2)</td>
<td>0.551</td>
<td>0.008</td>
<td>P5PP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x234x]</td>
</tr>
<tr>
<td>Protein</td>
<td>His β (dd)</td>
<td>0.108</td>
<td>0.007</td>
<td>P5PP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x234x]</td>
</tr>
<tr>
<td>Protein</td>
<td>His δ2 (s)</td>
<td>0.513</td>
<td>0.008</td>
<td>P5PP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[12xx]</td>
</tr>
<tr>
<td>Protein</td>
<td>His δ2 (d)</td>
<td>0.487</td>
<td>0.008</td>
<td>P5PP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[12xx]</td>
</tr>
<tr>
<td>Protein</td>
<td>Tyr δ1 (s)</td>
<td>0.214</td>
<td>0.008</td>
<td>PEP&lt;sup&gt;p&lt;/sup&gt;/E4PP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x23]/[1xx]</td>
</tr>
<tr>
<td>Protein</td>
<td>Tyr δ2 (s)</td>
<td>0.214</td>
<td>0.008</td>
<td>PEP&lt;sup&gt;p&lt;/sup&gt;/E4PP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x23]/[1xx]</td>
</tr>
<tr>
<td>Protein</td>
<td>Tyr δ1 (d)</td>
<td>0.692</td>
<td>0.008</td>
<td>PEP&lt;sup&gt;p&lt;/sup&gt;/E4PP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x23]/[1xx] + [x23]/[1xx]</td>
</tr>
<tr>
<td>Protein</td>
<td>Tyr δ2 (d)</td>
<td>0.692</td>
<td>0.008</td>
<td>PEP&lt;sup&gt;p&lt;/sup&gt;/E4PP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x23]/[1xx] + [x23]/[1xx]</td>
</tr>
<tr>
<td>Protein</td>
<td>Tyr ε1 (s)</td>
<td>0.094</td>
<td>0.008</td>
<td>PEP&lt;sup&gt;p&lt;/sup&gt;/E4PP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x23]/[1xx]</td>
</tr>
<tr>
<td>Protein</td>
<td>Tyr ε2 (s)</td>
<td>0.094</td>
<td>0.008</td>
<td>PEP&lt;sup&gt;p&lt;/sup&gt;/E4PP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x23]/[1xx]</td>
</tr>
<tr>
<td>Protein</td>
<td>Tyr ε1 (d)</td>
<td>0.462</td>
<td>0.006</td>
<td>PEP&lt;sup&gt;p&lt;/sup&gt;/E4PP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xx3]/[12xx]</td>
</tr>
<tr>
<td>Protein</td>
<td>Tyr ε2 (d)</td>
<td>0.462</td>
<td>0.006</td>
<td>PEP&lt;sup&gt;p&lt;/sup&gt;/E4PP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xx3]/[12xx]</td>
</tr>
<tr>
<td>Protein</td>
<td>Tyr ε1 (d)</td>
<td>0.190</td>
<td>0.006</td>
<td>PEP&lt;sup&gt;p&lt;/sup&gt;/E4PP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xx3]/[12xx] + [xx3]/[12xx]</td>
</tr>
<tr>
<td>Protein</td>
<td>Tyr ε2 (d)</td>
<td>0.190</td>
<td>0.006</td>
<td>PEP&lt;sup&gt;p&lt;/sup&gt;/E4PP&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xx3]/[12xx] + [xx3]/[12xx]</td>
</tr>
<tr>
<td>Hydrolysate</td>
<td>Cross peak (multiplet)</td>
<td>Intensity</td>
<td>SD</td>
<td>Precursor</td>
<td>Isotopomer</td>
</tr>
<tr>
<td>------------</td>
<td>-----------------------</td>
<td>-----------</td>
<td>----</td>
<td>-----------</td>
<td>------------</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Tyr ε1 (t)</td>
<td>0.348</td>
<td>0.006</td>
<td>PEP/E4P</td>
<td>[xx3]; [12xx]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Tyr ε2 (t)</td>
<td>0.000</td>
<td>0.000</td>
<td>PEP/E4P</td>
<td>[x234]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Arg β (s)</td>
<td>0.830</td>
<td>0.053</td>
<td>Glu&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x234x]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Arg β (d)</td>
<td>0.170</td>
<td>0.053</td>
<td>Glu&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x234x] + [x234x]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Arg δ (s)</td>
<td>0.663</td>
<td>0.012</td>
<td>Glu&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xxx45]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Arg δ (d)</td>
<td>0.337</td>
<td>0.012</td>
<td>Glu&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xxx45]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Glu β (s)</td>
<td>0.821</td>
<td>0.017</td>
<td>Glu&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x234x]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Glu β (d)</td>
<td>0.179</td>
<td>0.017</td>
<td>Glu&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x234x] + [x234x]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Glu γ (s)</td>
<td>0.694</td>
<td>0.009</td>
<td>Glu&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xx345]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Glu γ (d1)</td>
<td>0.016</td>
<td>0.009</td>
<td>Glu&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xx345]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Glu γ (d2)</td>
<td>0.290</td>
<td>0.009</td>
<td>Glu&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xx345]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Glu γ (dd)</td>
<td>0.000</td>
<td>0.000</td>
<td>Glu&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xx345]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Pro γ (s)</td>
<td>0.639</td>
<td>0.029</td>
<td>Glu&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xx345]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Pro γ (d)</td>
<td>0.361</td>
<td>0.029</td>
<td>Glu&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xx345] + [xx345]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Pro γ (t)</td>
<td>0.000</td>
<td>0.000</td>
<td>Glu&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xx345]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Pro δ (s)</td>
<td>0.641</td>
<td>0.036</td>
<td>Glu&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xxx45]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Pro δ (d)</td>
<td>0.359</td>
<td>0.036</td>
<td>Glu&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xxx45]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Asp α (s)</td>
<td>0.433</td>
<td>0.010</td>
<td>OAA&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[123x]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Asp α (d1)</td>
<td>0.087</td>
<td>0.010</td>
<td>OAA&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[123x]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Asp α (d2)</td>
<td>0.283</td>
<td>0.010</td>
<td>OAA&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[123x]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Asp α (dd)</td>
<td>0.197</td>
<td>0.010</td>
<td>OAA&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[123x]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Asp β (s)</td>
<td>0.406</td>
<td>0.010</td>
<td>OAA&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x234]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Asp β (d1)</td>
<td>0.343</td>
<td>0.010</td>
<td>OAA&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x234]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Asp β (d2)</td>
<td>0.082</td>
<td>0.010</td>
<td>OAA&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x234]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Asp β (dd)</td>
<td>0.169</td>
<td>0.010</td>
<td>OAA&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x234]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Ile α (s)</td>
<td>0.596</td>
<td>0.013</td>
<td>OAA&lt;sup&gt;p&lt;/sup&gt;/Pyr&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[12xx]; [x2x]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Ile α (d1)</td>
<td>0.287</td>
<td>0.013</td>
<td>OAA&lt;sup&gt;p&lt;/sup&gt;/Pyr&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[12xx]; [x2x]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Ile α (d2)</td>
<td>0.060</td>
<td>0.013</td>
<td>OAA&lt;sup&gt;p&lt;/sup&gt;/Pyr&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[12xx]; [x2x]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Ile α (dd)</td>
<td>0.058</td>
<td>0.013</td>
<td>OAA&lt;sup&gt;p&lt;/sup&gt;/Pyr&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[12xx]; [x2x]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Ile γ1 (s)</td>
<td>0.660</td>
<td>0.032</td>
<td>Pyr&lt;sup&gt;p&lt;/sup&gt;/OAA&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x2x]; [xx34]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Ile γ1 (d)</td>
<td>0.340</td>
<td>0.032</td>
<td>Pyr&lt;sup&gt;p&lt;/sup&gt;/OAA&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x2x]; [xx34] + [x2x]; [xx34]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Ile γ1 (t)</td>
<td>0.000</td>
<td>0.000</td>
<td>Pyr&lt;sup&gt;p&lt;/sup&gt;/OAA&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[x2x]; [xx34]</td>
</tr>
<tr>
<td><em>Protein</em></td>
<td>Ile δ (s)</td>
<td>0.732</td>
<td>0.007</td>
<td>OAA&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xx34]</td>
</tr>
</tbody>
</table>
Table I-1. (Continued.)

<table>
<thead>
<tr>
<th>Hydrolysate</th>
<th>Cross peak (multiplet)</th>
<th>Intensity</th>
<th>SD</th>
<th>Precursor</th>
<th>Isotopomer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein</td>
<td>Ile δ (d)</td>
<td>0.268</td>
<td>0.007 OAA&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xx34]</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Met γ (s)</td>
<td>0.730</td>
<td>0.033 OAA&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xx34]</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Met γ (d)</td>
<td>0.270</td>
<td>0.033 OAA&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xx34]</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Thr α (s)</td>
<td>0.403</td>
<td>0.050 OAA&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[123x]</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Thr α (d1)</td>
<td>0.122</td>
<td>0.050 OAA&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[123x]</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Thr α (d2)</td>
<td>0.225</td>
<td>0.050 OAA&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[123x]</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Thr y2 (s)</td>
<td>0.709</td>
<td>0.006 OAA&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xx34]</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Thr y2 (d)</td>
<td>0.291</td>
<td>0.006 OAA&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xx34]</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Lys γ (s)</td>
<td>0.655</td>
<td>0.023 OAA&lt;sup&gt;p&lt;/sup&gt;/Pyr&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xx34]-[xx3]</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Lys γ (d)</td>
<td>0.345</td>
<td>0.023 OAA&lt;sup&gt;p&lt;/sup&gt;/Pyr&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xx34]-[xx3]</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Lys γ (t)</td>
<td>0.000</td>
<td>0.000 OAA&lt;sup&gt;p&lt;/sup&gt;/Pyr&lt;sup&gt;p&lt;/sup&gt;</td>
<td>[xx34]-[xx3]</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Lys β (s)</td>
<td>0.235</td>
<td>0.013 OAA&lt;sup&gt;p&lt;/sup&gt;/Pyr&lt;sup&gt;p&lt;/sup&gt;</td>
<td>½({xx234} + [xx23]-[xx34])</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Lys β (d)</td>
<td>0.660</td>
<td>0.013 OAA&lt;sup&gt;p&lt;/sup&gt;/Pyr&lt;sup&gt;p&lt;/sup&gt;</td>
<td>½({xx234} + [xx23]-[xx34])</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Lys β (t)</td>
<td>0.106</td>
<td>0.013 OAA&lt;sup&gt;p&lt;/sup&gt;/Pyr&lt;sup&gt;p&lt;/sup&gt;</td>
<td>½({xx234} + [xx23]-[xx34])</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Lys δ (s)</td>
<td>0.250</td>
<td>0.008 OAA&lt;sup&gt;p&lt;/sup&gt;/Pyr&lt;sup&gt;p&lt;/sup&gt;</td>
<td>½({xx234} + [xx23]-[xx34])</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Lys δ (d)</td>
<td>0.642</td>
<td>0.008 OAA&lt;sup&gt;p&lt;/sup&gt;/Pyr&lt;sup&gt;p&lt;/sup&gt;</td>
<td>½({xx234} + [xx23]-[xx34])</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Lys ε (s)</td>
<td>0.108</td>
<td>0.008 OAA&lt;sup&gt;p&lt;/sup&gt;/Pyr&lt;sup&gt;p&lt;/sup&gt;</td>
<td>½({xx234} + [xx23]-[xx34])</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Lys ε (d)</td>
<td>0.318</td>
<td>0.005 OAA&lt;sup&gt;p&lt;/sup&gt;/Pyr&lt;sup&gt;p&lt;/sup&gt;</td>
<td>½({xx234} + [xx23]-[xx34])</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>HMF 1 (s)</td>
<td>0.271</td>
<td>&lt;0.010 G6P&lt;sup&gt;p&lt;/sup&gt;/F6P&lt;sup&gt;e&lt;/sup&gt;</td>
<td>[12xxxx]</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>HMF 1 (d)</td>
<td>0.729</td>
<td>&lt;0.010 G6P&lt;sup&gt;p&lt;/sup&gt;/F6P&lt;sup&gt;e&lt;/sup&gt;</td>
<td>[12xxxx]</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>LVA 3 (s)</td>
<td>0.207</td>
<td>&lt;0.005 G6P&lt;sup&gt;p&lt;/sup&gt;/F6P&lt;sup&gt;e&lt;/sup&gt;</td>
<td>[xx234xx]</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>LVA 3 (d1)</td>
<td>0.085</td>
<td>&lt;0.005 G6P&lt;sup&gt;p&lt;/sup&gt;/F6P&lt;sup&gt;e&lt;/sup&gt;</td>
<td>[xx234xx]</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>LVA 3 (d2)</td>
<td>0.029</td>
<td>&lt;0.005 G6P&lt;sup&gt;p&lt;/sup&gt;/F6P&lt;sup&gt;e&lt;/sup&gt;</td>
<td>[xx234xx]</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>LVA 3 (dd)</td>
<td>0.678</td>
<td>&lt;0.005 G6P&lt;sup&gt;p&lt;/sup&gt;/F6P&lt;sup&gt;e&lt;/sup&gt;</td>
<td>[xx234xx]</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>LVA 4 (s)</td>
<td>0.116</td>
<td>&lt;0.005 G6P&lt;sup&gt;p&lt;/sup&gt;/F6P&lt;sup&gt;e&lt;/sup&gt;</td>
<td>[xx3456]</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>LVA 4 (d1)</td>
<td>0.000</td>
<td>&lt;0.005 G6P&lt;sup&gt;p&lt;/sup&gt;/F6P&lt;sup&gt;e&lt;/sup&gt;</td>
<td>[xx3456]</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>LVA 4 (d2)</td>
<td>0.059</td>
<td>&lt;0.005 G6P&lt;sup&gt;p&lt;/sup&gt;/F6P&lt;sup&gt;e&lt;/sup&gt;</td>
<td>[xx3456]</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>LVA 4 (d3)</td>
<td>0.007</td>
<td>&lt;0.005 G6P&lt;sup&gt;p&lt;/sup&gt;/F6P&lt;sup&gt;e&lt;/sup&gt;</td>
<td>[xx3456]</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>LVA 4 (dd1)</td>
<td>0.001</td>
<td>&lt;0.005 G6P&lt;sup&gt;p&lt;/sup&gt;/F6P&lt;sup&gt;e&lt;/sup&gt;</td>
<td>[xx3456]</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>LVA 4 (dd2)</td>
<td>0.006</td>
<td>&lt;0.005 G6P&lt;sup&gt;p&lt;/sup&gt;/F6P&lt;sup&gt;e&lt;/sup&gt;</td>
<td>[xx3456]</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>LVA 4 (dd3)</td>
<td>0.074</td>
<td>&lt;0.005 G6P&lt;sup&gt;p&lt;/sup&gt;/F6P&lt;sup&gt;e&lt;/sup&gt;</td>
<td>[xx3456]</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>LVA 4 (q)</td>
<td>0.736</td>
<td>&lt;0.005 G6P&lt;sup&gt;p&lt;/sup&gt;/F6P&lt;sup&gt;e&lt;/sup&gt;</td>
<td>[xx3456]</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>LVA 6 (s)</td>
<td>0.132</td>
<td>&lt;0.005 G6P&lt;sup&gt;p&lt;/sup&gt;/F6P&lt;sup&gt;e&lt;/sup&gt;</td>
<td>[xx3456]</td>
<td></td>
</tr>
</tbody>
</table>
Table I-1. (Continued.)

<table>
<thead>
<tr>
<th>Hydrolysate</th>
<th>Cross peak (multiplet)</th>
<th>Intensity</th>
<th>SD Precursor</th>
<th>Isotopomer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein</td>
<td>LVA 6 (d1)</td>
<td>0.020</td>
<td>&lt; 0.005</td>
<td>G6P^2/F6P^c</td>
</tr>
<tr>
<td>Protein</td>
<td>LVA 6 (d2)</td>
<td>0.000</td>
<td>&lt; 0.005</td>
<td>G6P^2/F6P^c</td>
</tr>
<tr>
<td>Protein</td>
<td>LVA 6 (dd)</td>
<td>0.848</td>
<td>&lt; 0.005</td>
<td>G6P^2/F6P^c</td>
</tr>
<tr>
<td>Starch</td>
<td>HMF 1 (s)</td>
<td>0.348</td>
<td>&lt; 0.010</td>
<td>G6p^6</td>
</tr>
<tr>
<td>Starch</td>
<td>HMF 1 (d)</td>
<td>0.653</td>
<td>&lt; 0.010</td>
<td>G6p^6</td>
</tr>
<tr>
<td>Starch</td>
<td>LVA 3 (s)</td>
<td>0.459</td>
<td>&lt; 0.005</td>
<td>G6p^6</td>
</tr>
<tr>
<td>Starch</td>
<td>LVA 3 (d1)</td>
<td>0.245</td>
<td>&lt; 0.005</td>
<td>G6p^6</td>
</tr>
<tr>
<td>Starch</td>
<td>LVA 3 (d2)</td>
<td>0.050</td>
<td>&lt; 0.005</td>
<td>G6p^6</td>
</tr>
<tr>
<td>Starch</td>
<td>LVA 3 (dd)</td>
<td>0.245</td>
<td>&lt; 0.005</td>
<td>G6p^6</td>
</tr>
<tr>
<td>Starch</td>
<td>LVA 3 (dl)</td>
<td>0.250</td>
<td>&lt; 0.005</td>
<td>G6p^6</td>
</tr>
<tr>
<td>Starch</td>
<td>LVA 4 (d1)</td>
<td>0.015</td>
<td>&lt; 0.005</td>
<td>G6p^6</td>
</tr>
<tr>
<td>Starch</td>
<td>LVA 4 (d2)</td>
<td>0.005</td>
<td>&lt; 0.005</td>
<td>G6p^6</td>
</tr>
<tr>
<td>Starch</td>
<td>LVA 4 (d3)</td>
<td>0.000</td>
<td>&lt; 0.005</td>
<td>G6p^6</td>
</tr>
<tr>
<td>Starch</td>
<td>LVA 4 (dd1)</td>
<td>0.063</td>
<td>&lt; 0.005</td>
<td>G6p^6</td>
</tr>
<tr>
<td>Starch</td>
<td>LVA 4 (dd2)</td>
<td>0.000</td>
<td>&lt; 0.005</td>
<td>G6p^6</td>
</tr>
<tr>
<td>Starch</td>
<td>LVA 4 (dd3)</td>
<td>0.460</td>
<td>&lt; 0.005</td>
<td>G6p^6</td>
</tr>
<tr>
<td>Starch</td>
<td>LVA 4 (qd)</td>
<td>0.204</td>
<td>&lt; 0.005</td>
<td>G6p^6</td>
</tr>
<tr>
<td>Starch</td>
<td>LVA 6 (s)</td>
<td>0.167</td>
<td>&lt; 0.005</td>
<td>G6p^6</td>
</tr>
<tr>
<td>Starch</td>
<td>LVA 6 (d1)</td>
<td>0.139</td>
<td>&lt; 0.005</td>
<td>G6p^6</td>
</tr>
<tr>
<td>Starch</td>
<td>LVA 6 (d2)</td>
<td>0.029</td>
<td>&lt; 0.005</td>
<td>G6p^6</td>
</tr>
<tr>
<td>Starch</td>
<td>LVA 6 (dd)</td>
<td>0.665</td>
<td>&lt; 0.005</td>
<td>G6p^6</td>
</tr>
</tbody>
</table>
Supplementary material II: Biomass composition fluxes

1. Proportions of biomass constituents in developing soybean embryos (measured as described in Materials and Methods).

<table>
<thead>
<tr>
<th>Component</th>
<th>Percentage (w/w) in dry weight</th>
<th>How determined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein</td>
<td>31.7</td>
<td>measured</td>
</tr>
<tr>
<td>Oil</td>
<td>15.6</td>
<td>measured</td>
</tr>
<tr>
<td>Starch</td>
<td>6.3</td>
<td>measured</td>
</tr>
<tr>
<td>Ash + moisture</td>
<td>13.0</td>
<td>from previous work in our laboratory</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>33.4</td>
<td>remainder of biomass</td>
</tr>
</tbody>
</table>

2. Amino acid composition of seed protein. Amino acids were measured by HPLC. Composition is shown as molar percentage of protein hydrolysate.

<table>
<thead>
<tr>
<th>AA</th>
<th>Molar %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>7.20</td>
</tr>
<tr>
<td>Arg</td>
<td>9.80</td>
</tr>
<tr>
<td>Asp+Asn</td>
<td>12.55</td>
</tr>
<tr>
<td>Glu+Gln</td>
<td>18.61</td>
</tr>
<tr>
<td>Gly</td>
<td>6.12</td>
</tr>
<tr>
<td>His</td>
<td>1.86</td>
</tr>
<tr>
<td>Ile</td>
<td>4.19</td>
</tr>
<tr>
<td>Leu</td>
<td>7.48</td>
</tr>
<tr>
<td>Lys</td>
<td>5.78</td>
</tr>
<tr>
<td>Met</td>
<td>1.23</td>
</tr>
<tr>
<td>Phe</td>
<td>3.77</td>
</tr>
<tr>
<td>Pro</td>
<td>4.98</td>
</tr>
<tr>
<td>Ser</td>
<td>5.48</td>
</tr>
<tr>
<td>Thr</td>
<td>3.54</td>
</tr>
<tr>
<td>Tyr</td>
<td>1.95</td>
</tr>
<tr>
<td>Val</td>
<td>5.46</td>
</tr>
</tbody>
</table>

3. Fatty acid composition of oil was obtained from data reported by Dey and Harborne (1997) for soybean seed oil.
4. The proportions of sugar monomers (glucose, fructose, etc.) in carbohydrates content was estimated from values reported by Mullin and Xu (2000) for soybean cotyledons. These authors report no significant variation in the carbohydrate content across soybean varieties.

References


Supplementary material III: Accuracy of 2-D $^{13}\text{C}, ^1\text{H}$ HSQC experiment

The isotopomer abundances reported in this paper are crucially dependent on the accuracy of the 2-D $^{13}\text{C}, ^1\text{H}$ HSQC experiment used to measure them. We verified that the $^{13}\text{C}, ^1\text{H}$ experiment employed by us can accurately measure isotopomer abundances, by performing it on two samples containing known quantities of three commercial isotopomers of Ala (a representative amino acid).

The final mole fractions of different Ala isotopomers in the samples A and B, are shown in Table III-1. These mole fractions were chosen so that the $^{13}\text{C}$ enrichment of Ala in both samples is less than 10%, and the total Ala concentration in the samples A and B was approximately equal to the average concentration of any amino acid or levulinic acid in our samples of soybean protein or starch hydrolysate. Thus we emulated the conditions of the experiment reported in Materials and Methods. Also, acquisition parameters of the 2-D $^{13}\text{C}, ^1\text{H}$ HSQC experiment (on samples A and B) and the quantification of relative intensities were exactly as in Materials and Methods.

Table III-1. Mole fractions of labeled and natural abundance Ala isotopomers in samples A and B.

<table>
<thead>
<tr>
<th>Type of labeled Ala (purity %)</th>
<th>Sample A (mol fraction)</th>
<th>Sample B (mol fraction)</th>
</tr>
</thead>
<tbody>
<tr>
<td>natural abundance Ala</td>
<td>0.8982</td>
<td>0.9491</td>
</tr>
<tr>
<td>$2,3^{13}\text{C}_2$ Ala (99.0%)</td>
<td>0.0499</td>
<td>0.0214</td>
</tr>
<tr>
<td>U-$^{13}\text{C}$ Ala (98.5%)</td>
<td>0.0518</td>
<td>0.0305</td>
</tr>
</tbody>
</table>

Figure III-1 shows a comparison of measured versus actual relative intensities of multiplets of Ala (or Ala isotopomer abundances). The actual abundances of the Ala isotopomers in
samples A and B were calculated from the mol fraction of the Ala varieties in the mixture, and from the (known) natural abundance of $^{13}$C and the (known) purity of the isotopic compounds. Standard formulae from the literature (Szyperski, 1995) were used to calculate these abundances. Clearly, the 2-D $[^{13}$C, $^1$H] HSQC experiment, under the conditions in this paper, can accurately measure relative multiplet intensities (or isotopomer abundances).

**Figure III-1.** Experiment to verify accuracy of 2-D $[^{13}$C, $^1$H] HSQC experiment. Samples A and B were prepared by mixing different commercially available isotopomers of Ala, in the proportions shown in Table III-1. A 2-D $[^{13}$C, $^1$H] HSQC experiment was performed on these samples, using the same acquisition parameters as described in Materials and Methods.
Supplementary material IV: Activity diagrams, explanations, and mathematical concepts

This document contains activity diagrams for the flux evaluation procedure and the computer program NMR2Flux, and explanations of the relevant mathematical concepts. All activity diagrams are prepared as per the UML (Universal Modeling Language) convention.

Diagram 1 depicts an overview of the flux evaluation procedure in this work. An experimenter performs the labeling experiment, harvests the biomass, prepares an NMR sample, acquires 2-D $^{13}\text{C}, ^{1}\text{H}$ NMR spectra, and quantifies multiplet intensities, as explained in Materials and Methods. The multiplet intensities are stored in an input file (Input file 3).

Subsequently, a user of NMR2Flux (who could be the same as the experimenter) assumes a metabolic network. This network is based on the known biochemistry of the plant system being studied and on indications from the NMR data. The NMR2Flux user prepares three input files (Input files 1, 2, 4) containing information about the metabolic network.

NMR2Flux currently accepts input files in CSV (comma-separated values) format. Input files can be prepared with spreadsheet programs (e.g. Microsoft Excel) and saved in this format. Supplementary Material VI contains the Input files used in this work.

NMR2Flux processes the input files and evaluates fluxes iteratively $N$ (usually 300 or greater) times. Each flux evaluation is begun from an arbitrary starting set of fluxes. The user then supplies the $N$ sets of evaluated fluxes, to obtain means and standard deviations for all fluxes. Figure IV-3 depicts examples of such fluxes evaluated by NMR2Flux, and normal distributions fitted through them.
Diagram 2 provides an overview of NMR2Flux. The program reads the supplied input files, stores the information in indexed vectors and matrices, and allocates vectors to store fluxes and cumulative isotopomers. The function FEASIBLE_FLUX (Diagram 5) is called, to examine if the number of flux parameters supplied by the user is correct. If this is incorrect, the user is prompted to correct the parameters. If the number of parameters is correct, NMR2Flux evaluates $N$ flux evaluations (runs), and reports the $N$ sets of evaluated fluxes.

To transform the errors in the measured multiplet intensities to errors (or standard deviations) in the fluxes, a Bootstrap Monte Carlo simulation is employed (Press et al., 1992), wherein the measurements are perturbed (by their known standard error) after each run. A 'perturb flag', if set, enables this feature.

Diagram 3 depicts the simulated annealing routine. During each flux evaluation, this routine evaluates a set of fluxes that minimizes the $\chi^2$ error between the measured multiplet intensities and those corresponding to the evaluated flux.

An initial set of flux parameters is chosen and verified for stoichiometric feasibility (by the function FEASIBLE_FLUX). If feasible, a $\chi^2$ error ($\chi^2_{\text{current}}$) corresponding to the fluxes is calculated (Diagram 4). A next candidate set of flux parameters is selected. After its feasibility is examined by FEASIBLE_FLUX, the $\chi^2$ error ($\chi^2_{\text{new}}$) corresponding to the next candidate set of fluxes is calculated.

The next candidate set is accepted or rejected as per the Metropolis acceptance criterion (Press et al., 1992):

- If $\chi^2_{\text{new}} \leq \chi^2_{\text{current}}$, accept the next candidate point.
- If $\chi^2_{\text{new}} > \chi^2_{\text{current}}$:
  - Generate a uniform random number $r$ in the range $[0,1]$.
  - If $\chi^2_{\text{new}} > \chi^2_{\text{current}}$, accept the new point if $(\chi^2_{\text{new}} - \chi^2_{\text{current}}) < T \ln(r)$. 
Diagram 1. Overview of flux evaluation procedure.
Diagram 2. Overview of NMR2Flux.
Diagram 3. SIMULATED_ANNEALING routine (function) in NMR2Flux, and its communication with the functions FEASIBLE_FLUX and CHI_SQUARED.
If the new point is accepted, it becomes the current optimum. If the point is rejected, the existing optimum is retained and a new candidate point is searched for. The acceptance criterion ensures that better points are always accepted, but worse points are occasionally accepted. This is a characteristic of simulated annealing (Press et al. 1992), that enables it to reach the global optimum of challenging functions with multiple local optima. Here, $T$ is an artificial temperature.

The manipulation of $T$, called the cooling schedule, is as follows:

- Start with a high $T$. In this work, this value was $2.0 \times 10^6$.
- After $M$ new candidate points or $S$ accepted new points (whichever occurs first), reduce the temperature by a factor $e$. In this work, the values $M=100$, $S=20$, and $e=0.99$ were used.

Further, the stepsize of each flux parameter is reduced as follows:

- Start with a large stepsize for each flux parameter. These are provided by the user in Input file 2.
- After $M_1$ new candidate points or $S_1$ accepted new points (whichever occurs first), reduce the stepsize by a factor $e_1$, provided:
  - The derivative of $x^2$ along the direction of that flux parameter is higher than the current value of $T$.

In this work, the values $M_1=50$, $S_1=10$, and $e_1=0.50$ were used. The condition on stepsize reduction ensures that the stepsize reduction of a parameter is roughly proportional to the sensitivity of the objective function to that parameter.

Finally, the simulated annealing routine is stopped if the stopping criterion $T < T_{\text{stop}}$ is reached. The value of $T_{\text{stop}}$ used by in this work was 0.5.

**Diagram Description:**
- **SIMULATE_ISOTOPOMERS**
  - Order = 1
  - Formulate balances for cumomers of current order in matrix form
  - Balances are linear, solve by U-decomposition
  - Cumulative isotopomer abundances
  - Current order $\geq O_{max}$ or next order
- **SIMULATE_SPECTRUM**
  - Convert (cumulative) isotopomer abundances to sink metabolite multiplet intensities (using information from Input file 4)
  - Simulated multiplet intensities
- **CHI_SQUARED**
  - Fluxes
  - Experimental multiplet intensities from Input file 4
  - $\chi^2$ from simulated and experimental multiplet intensities
Allocate matrices $G_x$ and $G_m$ and fill in stoichiometric coefficients.

1st call to this function not 1st call

$G_x$ is singular $G_x$ is nonsingular

Allocate vector $v_x$ and fill in flux measurements and current flux parameters.

Solve for $v_x$ using $v_x = G_x^{-1}G_m v_m$

1 or more irreversible reactions has/have negative net flux

all irreversible reactions have nonnegative net flux

Calculate reversible and scrambling fluxes.

all fluxes

Diagram 5. Calculation of fluxes from flux parameters, and examination of the feasibility of the parameter set.
Diagram 4 depicts the calculation of $\chi^2$. (Cumulative) isotopomer balancing is explained later in this document.

Diagram 5 depicts the calculation of fluxes from flux parameters, and examination of the feasibility of the parameter set. The matrices $G_X$, $G_M$, and the vectors $v$, $v_M$ are explained in Stephanopoulos et al. (1998). Flux parameters are feasible only if the matrix $G_X$ is nonsingular.

Selection of feasible fluxes in two- (and higher-) dimensional flux parameter space

Flux parameter space. Figure IV-1 depicts the flux parameter space in a metabolic network with two flux parameters, $v_1$ and $v_2$. Any combination $[v_1,v_2]$ may not be feasible, as $v_1$ and $v_2$ are stoichiometrically constrained. In the figure, the stoichiometric constraints (i.e. metabolite balances) are denoted by dark red lines. The allowed values of $[v_1,v_2]$ lie within the convex polygonal space contained by these lines.

Figure IV-1. Flux parameter space.
Choosing an initial random, feasible parameter combination. To choose an initial random, feasible combination \([v_1, v_2]\) we follow the following procedure, similar to Locatelli (2002).

Choose a rectangular sampling region (shown with a dark black border), which accommodates the entire convex feasible region. Such a sampling region can easily usually be found by trial and error, and without much effort.

Generate a candidate point \([v'_1, v'_2]\) within this rectangular region, by using uniform random distributions along \(v_1\) and \(v_2\).

If this candidate point is not in the feasible (white) region, but in the infeasible (blue) region, reject it and generate another candidate point. Repeat this process until a point that lies in the feasible region, is found.

For movement from one set of flux parameters to another (during simulated annealing), the same procedure is followed. However, the sampling region (whose sides are the \textit{stepsizes} along each dimension) is progressively shrunk, as per the stepsize reduction schedule explained above.

Notes

1. In higher dimensional parameter spaces, the stoichiometric constraints are planes or hyperplanes rather than lines, and the feasible parameter space is a convex polyhedron rather than a convex polygon. However, the same logic can be applied.

2. If a flux parameter \(v_i\) denotes the reversibility of a pair of reversible reactions or the scrambling extent of a pair of parallel reactions (see definitions below), then the only constraint on it is \(v_i \in [0,1]\).
Definitions: Net flux, reversibility. Consider the two reactions $A \leftrightarrow B$. This is a system of two reactions $A \rightarrow B$ and $B \rightarrow A$. In many cases, these could be catalyzed by the same enzyme [e.g. hexose isomerase (hxi) catalyzes $\text{G6P} \leftrightarrow \text{F6P}$].

Let the reaction with the higher flux (say, $A \rightarrow B$) be designated as the forward reaction, and the other reaction ($B \rightarrow A$) as the backward reaction. Let the fluxes of the forward and reverse reactions be $v_f$ and $v_b$ respectively. Then, the following definitions hold:

$$v_n = v_f - v_b$$
$$r = \frac{v_b}{v_f}$$

Here, $v_n$ is the net flux of the two-reaction system, and $r$ is its reversibility. Also, the forward reaction is termed as the reversibility partner of the backward reaction (see Input file 2). The reversibility indicates extent of approach to equilibrium ($r=0$ signifies an irreversible reaction, $r=1$ signifies a reaction at thermodynamic equilibrium). Note that $r$ will always lie between 0 and 1.

Definitions: Parallel fluxes, extent of scrambling. Consider two reactions $A \overset{1}{\rightarrow} B$ and $A \overset{2}{\rightarrow} B$; that have the same stoichiometry, but different carbon skeletal rearrangements. An example of such a pair of reactions is succinate (Sen) $\rightarrow$ malate (Mal). Since Sen is symmetrical but Mal is not, two possible isotopomers of Mal could result from one isotopomer of Sen.

Fluxes in such a pair are termed parallel reactions. Let $v_1$ and $v_2$ be the fluxes of the two parallel reactions. Then, the following definitions hold:

$$v_n = v_1 + v_2$$
$$s = \frac{v_2}{v_n}$$
Here, s is the *scrambling extent* of this reaction system. It is so named because the product B in one of the parallel reactions will be a scrambled version of the product B in the other reaction. Note that s will always lie between 0 and 1.

The above definitions can be extended to reaction systems with more than one reactant or product.

**Generation and solution of isotopomer balances**

The procedure for generation and solution of isotopomer balances (from an assumed set of fluxes) involved four steps, and was as follows.

**Step 1. Compilation of carbon atom rearrangement information**

Carbon atom rearrangements of all reactions in the network were obtained from KEGG (2004). For example, the stoichiometry and the carbon atom rearrangement of the transaldolase (tal^f) reaction (S7P + T3P → E4P + F6P) are shown in Figure IV-2A. These were stored in Input file 2 in a concise spreadsheet format (see Sriram and Shanks, 2004).

**Step 2. Mapping reactant isotopomers to product isotopomers**

The following is an example of mapping between reactant and product isotopomers in a metabolic reaction. Consider the isotopomers S7P^[123xxxx] and T3P^[123]. From the carbon atom rearrangement for this reaction (Figure IV-2), it can be deduced that when these participate in the transketolase reaction (tal^p), the F6P^p isotopomer F6P^p^[123456] is one of the products (see Figure IV-2B).

The routine Boolean function mapping (Sriram and Shanks, 2004) was used to automatically map reactant isotopomers to product isotopomers, using the carbon atom rearrangement

---

2 The notation for isotopomers is explained in Figure 3.1 and Results.
information in Input file 2. The implementation of Boolean function mapping is explained in Sriram and Shanks (2004). Subsequently, an indexed list of all reactant-product isotopomer combinations was prepared for every reaction in the network.

A  **Stoichiometry:**
\[ S7P + T3P \rightarrow E4P + F6P \]

**Carbon skeleton rearrangement:**

\[ \begin{array}{c}
S7P \quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \\
E4P \quad 1 \quad 2 \quad 3 \quad 4 \\
F6P \quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \\
T3P \quad 1 \quad 2 \quad 3 
\end{array} \]

B  **Formation of \( F6P_{123456} \):**

\[ \begin{array}{c}
S7P_{123456} \\
F6P_{123456} \\
T3P_{123} \\
E4P \quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \\
S7P \quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \\
T3P \quad 1 \quad 2 \quad 3 
\end{array} \]

**Figure IV-2.** (A) Stoichiometry and carbon skeleton rearrangement for the transaldolase (\( tal \)) reaction. (B) Formation of the F6P isotopomer \( [123456] \) by this reaction. \(^{13}\)C atoms are shown in white font within dark boxes, and \(^{12}\)C atoms are shown in dark font within hollow boxes.

**Step 3. Construction of isotopomer balances**

Isotopomer balances were constructed from the list of all reactant and product isotopomers for all reactions, compiled above.
For example, consider the isotopomer $F6P^{[123456]}$. As seen above, this is formed from the isotopomers $S7P^{[123xxxx]}$ and $T3P^{[123]}$ through the $tal^f$ reaction. This isotopomer can also be formed from $G6P^{[123456]}$ through the $hxi^f$ reaction, from $T3P^{[123]}$ and $T3P^{[123]}$ through the $f16bp^f$ reaction, or from $P5P^{[12xxx]}$ and $E4P^{[1234]}$ through the $tktB^f$ reaction. Further, this isotopomer can be depleted by all reactions involving $F6P$ as a reactant, i.e. $pfk^p$, $hxi^f$, $tal^f$, and $tktB^f$. From this information, the steady-state isotopomer balance for the isotopomer $F6P^{[123456]}$ can be written as follows:

$$0 = \Sigma \{\text{fluxes of formation}\} - \Sigma \{\text{fluxes of depletion}\}$$

or

$$0 = \left\{ v_{hxi^f} \cdot S7P^{[123xxxx]} \cdot T3P^{[123]} + v_{hxi^f} \cdot G6P^{[123456]} \right\}$$
$$+ \left\{ v_{f16bp^f} \cdot T3P^{[123]} \cdot T3P^{[123]} + v_{tal^f} \cdot P5P^{[12xxx]} \cdot E4P^{[1234]} \right\}$$
$$- \left\{ v_{pfk^p} \cdot F6P^{[123456]} + v_{hxi^f} \cdot F6P^{[123456]} + v_{tal^f} \cdot F6P^{[123456]} + v_{tktB^f} \cdot F6P^{[123456]} \right\}$$

(IV-1)

where $v$ denotes the value of the flux of the reaction indicated by its subscript.

Isotopomer balances can be written for all the isotopomers (564 in number) in the metabolic network along these lines.

**Step 4. Solution of isotopomer balances by exact isotopomer balancing**

Isotopomers appear as nonlinear terms in isotopomer balance equations (IV-1), therefore their solution is not trivial. We solved the isotopomer balances by transforming them to a cascade of linear equations. This transformation was accomplished by the cumulative

---

3 See Figure 3.6 for a listing of all reactions in the metabolic network of this paper.
isotopomer method (Wiechert et al., 1999; Wiechert and Murzel, 2001). This method is explained below, following two definitions central to it.

**Definition: Cumulative isotopomer.** Cumulative isotopomers are defined as isotopomers whose atoms are either labeled ($^{13}$C), or the labeling state is undetermined. As per the notation used in this paper, cumulative isotopomers have either a boldface or an x at each position. Examples of cumulative isotopomers are E4P[^xx3x], T3P[^12x], and G6P[^12x45x].

**Definition: Order (or weight) of a cumulative isotopomer.** The order of a cumulative isotopomer is the number atoms it that are definitely labeled. Thus, the order of E4P[^xx3x] is 1, the order of T3P[^12x] is 2, and the order of G6P[^12x45x] is 3.

The central idea of the cumulative isotopomer method, is that the nonlinear terms in a balance on a cumulative isotopomer of order $n$ are either fluxes or cumulative isotopomers of orders less than $n$. Thus, a balance cumulative isotopomer of order 1 contained only fluxes as nonlinear terms. For example:

Balance on order 1 cumulative isotopomer, F6P[^1xxxx]:

$$0 = \left\{ v_{\text{taf}} \cdot S7P[^1xxxxx] \right\} + v_{\text{haf}} \cdot G6P[^1xxxxx]$$

$$+ \left\{ v_{f16bp} \cdot T3P[^xx3] \right\} + v_{\text{thb}} \cdot P5P[^1xxxx]$$

$$- \left\{ v_{p6d} \cdot F6P[^1xxxx] \right\} + v_{\text{hld}} \cdot F6P[^1xxxx] + v_{\text{taf}} \cdot F6P[^1xxxx] + v_{\text{dab}} \cdot F6P[^1xxxx] \right\}$$

(IV-2)

Balance on order 1 cumulative isotopomer, F6P[^xxxx5x]:

$$0 = \left\{ v_{\text{taf}} \cdot T3P[^x2x] \right\} + v_{\text{haf}} \cdot G6P[^xxxx5x]$$

$$+ \left\{ v_{f16bp} \cdot T3P[^x2x] \right\} + v_{\text{thb}} \cdot E4P[^xx3]$$

$$- \left\{ v_{p6d} \cdot F6P[^xxxx5x] \right\} + v_{\text{hld}} \cdot F6P[^xxxx5x] + v_{\text{taf}} \cdot F6P[^xxxx5x] + v_{\text{dab}} \cdot F6P[^xxxx5x] \right\}$$

(IV-3)
Therefore, once a set of fluxes was assumed, the order 1 cumulative isotopomer balances were a set of linear equations. These were solved by using LU-decomposition (Press et al., 1992), which is a straightforward method that can be easily automated.

Further, the order 2 cumulative isotopomer balances only contained fluxes or order 1 cumulative isotopomers as nonlinear terms. For example:

Balance on order 2 cumulative isotopomer, $F_{6P}^{[23xxx]}$:

$$\begin{align*}
0 &= \left\{ v_{\text{tal}i} \cdot S7P^{[x23xxx]} + v_{\text{hax}i} \cdot G6P^{[x23xxx]} \\
&\quad + v_{f6p} \cdot T3P^{[12x]} + v_{dkg} \cdot P5P^{[x2xxx]} \cdot E4P^{[1xxx]} \right. \\
&\quad - \left. v_{red} \cdot F6P^{[x23xxx]} + v_{haxi} \cdot F6P^{[23xxx]} + v_{haxk} \cdot F6P^{[23xxx]} \right\}
\end{align*}$$

(IV-4)

Similarly, the order 3 cumulative isotopomer balances only contained fluxes, order 1 cumulative isotopomers or order 2 cumulative isotopomers as nonlinear terms. For example:

Balance on order 3 cumulative isotopomer, $F_{6P}^{[123xxx]}$:

$$\begin{align*}
0 &= \left\{ v_{\text{tal}i} \cdot S7P^{[123xxx]} + v_{\text{hax}i} \cdot G6P^{[123xxx]} \\
&\quad + v_{f6p} \cdot T3P^{[123]} + v_{dkg} \cdot P5P^{[12xxx]} \cdot E4P^{[1xxx]} \right. \\
&\quad - \left. v_{red} \cdot F6P^{[123xxx]} + v_{haxi} \cdot F6P^{[123xxx]} + v_{haxk} \cdot F6P^{[123xxx]} \right\}
\end{align*}$$

(IV-5)

Likewise, the order 6 cumulative isotopomer balances only contained fluxes, or lower order cumulative isotopomers as nonlinear terms. For example:

Balance on order 6 cumulative isotopomer, $F_{6P}^{[123456]}$:
For a given metabolic network, cumulative isotopomer balances have to be generated up to an order $O_{max}$, which is equal the number of atoms in the longest precursor metabolite. In this work, $O_{max}$ was 7, the number of atoms in $S7P^e$ or $S7P^p$.

Once all cumulative isotopomer balances were generated in this manner and solved, the abundances of the isotopomers were calculated in terms of the cumulative isotopomers. For example, the abundance of $F6P^{[123456]}$ in terms of cumulative isotopomers is:

$$F6P^{[123456]} = F6P^{[1245x]} - F6P^{[12x45]} - F6P^{[12345]} + F6P^{[123456]} \quad (IV-7)$$

The transformation of the $F6P^{[12345]}$ balance (Eq. IV-1) into a cumulative isotopomer balance cascade and their automatic solution through LU-decomposition thus enabled the automation of isotopomer balance solutions. Additionally, the cumulative isotopomer approach drastically shortened solution time (100-fold) as compared to iterative solution methods such as those of Schmidt et al. (1999).

**Calculation of $\chi^2$ error**

The final quantity that is minimized by simulated annealing is the $\chi^2$ error. This is calculated from the experimental and simulated multiplet intensities. Experimental multiplet intensities are supplied by the user, in Input file 3. Simulated multiplet intensities can be calculated from the isotopomer abundances evaluated as shown above, and using the information supplied by the user in Input file 4. The $\chi^2$ error is then computed as

$$\chi^2 = \sum_{j=1}^{p} \left[ I_j - I_{sj} \right]^2 \quad (IV-8)$$
Here, $I_j$ and $I_{xy}$ are the $j$-th simulated and experimental multiplet intensities, out of a total of $P$ multiplet intensities. In our work, $P$ was equal to 155.

**Statistical analysis of evaluated fluxes**

Standard deviations for the fluxes and reversibility extents were computed from a statistical analysis. Herein, the errors in the measured multiplet intensities were used to perform 300 Monte Carlo estimations of fluxes (Press et al., 1992). This generated probability distributions of each flux. Standard deviations were calculated by fitting a normal distribution through these flux distributions. Two such distributions (for two anaplerotic fluxes) are shown in Figure IV-3.

**Figure IV-3.** Statistical analysis of evaluated fluxes. Probability distributions of fluxes were obtained by taking into account the errors in the experimental relative intensities, and by using Monte Carlo simulations. Hollow circles represent the distribution for anaplerotic flux $\text{ppc}^c$ ($\text{PEP}^c \rightarrow \text{OAA}^c$), and filled circles represent the distribution for anaplerotic flux $\text{me}^m$. 
(Mal$^m \rightarrow$ Pyr$^m$). The solid lines represent normal distributions fitted through these data, using the software JMP$^\text{TM}$. Standard deviations for all fluxes were obtained from such normal distributions.
Supplementary material V: Effectiveness of simulated annealing in finding global optima

Simulated annealing is a well-documented technique to find global optima of multidimensional functions. This is supported by the following articles or book chapters: Bélisle (1992); Press et al. (1992) (section 10.9); Boender and Romeijn (1995); Zabinsky (1998); Locatelli (2000); Onbaşoğlu and Özdamar (2001); Locatelli (2002); Wood et al. (2002).

Before proceeding to use simulated annealing to find the global optimum of our flux evaluation problem, we ensured that this technique is able to find global optima of two problems that have multiple local optima, but a single, known global optimum. The first problem was a two-dimensional trigonometric function, and the second problem was a 33-dimensional flux evaluation problem similar to our problem.

In the cases that we analyzed, we found simulated annealing is always able to find the global optimum, provided the algorithm is run sufficiently slowly (in technical terms: provided the annealing is gradual enough). How slowly the algorithm should be run to ensure convergence to the global optimum, depends on the specific problem, and can be found by a user without much effort.

2-dimensional trigonometric function

The function

\[ f(x, y) = \frac{\cos[\pi(x - x_g)]\cos[2\pi(y - y_g)]}{1 + (x - x_g)^2 + (y - y_g)^2} \quad (V-1) \]

in two parameters (dimensions) \( x \) and \( y \) has infinite local minima, but a single global minimum. It can be shown analytically that the value of this function is always greater than
\( f(x_g, y_g) \), the value of the function at \((x_g, y_g)\). Therefore \((x_g, y_g)\) is the lowest possible value of the function, or the global minimum.

Figure V.1 shows a plot of this function with \(x_g = 2.46\), and \(y_g = 1.34\). The figure shows a clear view of the global minimum and the local minima surrounding it. (Note that the \( f \) axis has been inverted to afford this perspective.)

We used our implementation of simulated annealing to find the global optimum of this function. 1,000 separate simulated annealing runs were implemented, each from a different, random starting point. The starting points were in the range \((-5.0 < x < 5.0), (-5.0 < y < 5.0)\). In all of these 1000 runs, our program evaluated the optimum as \((x = 2.4600, y = 1.3400)\), with the error in \(x\) or \(y\) always being less than 0.00000050. Within each run, the number of iterations required to converge to the global optimum was between 30,000 and 70,000. The average time taken for each run was about \(3 \times 10^{-3}\) s.

Also, the global optimum was always evaluated, even when the range of \(x\) or \(y\) was expanded to \((-50.0 < x, y < 50.0)\) or \((-500.0 < x, y < 500.0)\).

Local algorithm optimization algorithms, such as the downhill simplex (Press et al., 1992) almost never converged at the global minimum.

**33-dimensional flux analysis problem**

We formulated a 33-dimensional flux analysis problem similar to the flux evaluation problem in this paper. Isotopomer abundances were simulated from a known set of fluxes. It was examined if NMR2Flux (that uses simulated annealing as its optimization algorithm) can evaluate the original set of fluxes without any knowledge of that set of fluxes.

We found that with the exceptions of the lower glycolytic pathway in the cytosol and plastid (from T3P to Pyr) and the reversibilities marked ‘n.d' in Figure 3.6, all fluxes were identified well within the standard deviations listed in Figure 3.6, in 200 separate flux evaluations.
starting from random initial points. This shows that the simulated annealing can handle flux evaluation problems such as the one discussed in this paper.

**Figure V.1.** Graph of the function $f$ in Eq. V-1. The values $x_g = 2.43, y_g = 1.33$ were used. The $f$ axis is inverted so as to afford a clear view of the global minimum, which is surrounded by several local minima.

**References**


Supplementary material VI: Annotated input files to NMR2Flux

This material contains the four input files to the computer program NMR2Flux. The files are annotated with explanations. They are in PDF format in the attached CD-ROM, and can be viewed with Adobe Acrobat Reader, available free of charge at http://www.adobe.com (Last accessed on 20 July 2004).
4. Metabolic flux analysis of *Catharanthus roseus* hairy roots by using biosynthetically directed fractional \(^{13}\)C-labeling and bondomer balancing

An early draft of a paper, to be expanded and submitted to *Metabolic Engineering*.

Ganesh Sriram\(^1\), D. Bruce Fulton\(^2\) and Jacqueline V. Shanks\(^1\)

Departments of \(^1\)Chemical Engineering; \(^2\)Biophysics, Biochemistry, and Molecular Biology;

Iowa State University, Ames, IA 50011

**Abstract**

Metabolic flux analysis is recognized as a powerful tool in metabolic engineering. We report the application of a carbon bond labeling experiment for evaluating metabolic fluxes in a plant system (*Catharanthus roseus* hairy roots). The experiment involves culturing the roots on a mixture of 5% (w/w) U-\(^{13}\)C sucrose and 95 % natural-abundance (predominantly \(^{12}\)C) sucrose, extracting the proteinogenic amino acids, and conducting a 2-D \(^{13}\)C, \(^1\)H HSQC NMR experiment on them. The multiplet intensities of various carbon nuclei from the NMR spectrum were analyzed and converted to metabolic fluxes by using bondomer balancing. We found that all fluxes primary/intermediary metabolic pathways such as glycolysis, pentose phosphate, TCA cycle, and the anaplerotic and glyoxylate shunts (except the GABA shunt) were well-identified. A rather high cycling was observed between the anaplerotic fluxes PEP

---

\(^1\) Primary researcher and author.

\(^2\) NMR facility manager.

\(^*\) Principal investigator and corresponding author.
→ OAA (cytosol) and Mal → Pyr (mitochondrion). This methodology is expected to be valuable in assessing the impact of genetic or environmental changes in the hairy roots.

**Introduction**

Metabolic flux analysis (MFA) is a fundamental diagnostic tool of metabolic engineering. Flux measurements and comparisons of fluxes between phenotypes can provide valuable insights toward the selection of appropriate metabolic engineering targets, the systematic elucidation of metabolic control (Stephanopoulos, 1999; 2002), and toward the construction of predictive models of metabolism (Stephanopoulos, 2002; Sweetlove *et al.*, 2003). The use of $^{13}$C labeling followed by NMR (nuclear magnetic resonance) or MS (mass spectroscopy) analysis is a powerful method of flux quantification, that is gaining popularity in metabolic engineering.

Flux quantification in plants has received very limited attention compared to that in microorganisms and mammalian cells (Sweetlove *et al.*, 2003). This is principally because plant biochemistry is more complex, and exhibits extensive subcellular compartmentation. This makes flux analysis difficult, because the detailed mathematical models needed to relate labeling data to fluxes are often highly nonlinear and non-trivial to solve, particularly in the case of compartmented metabolism inherent in plants.

* Catharanthus roseus is a source of high-value indole alkaloids, some of which possess therapeutic properties. Hairy roots of the plant are biochemically and genetically stable model systems of the plant. Since the production of indole alkaloids by the *C. roseus* plant or roots is very low, metabolic engineering of the indole alkaloid pathway in *C. roseus* is an attractive prospect. Recently, quite a few genetic variants of *C. roseus* hairy roots were generated in our laboratory (Hughes *et al.*, 2004). A metabolic flux analysis tool for *C. roseus* hairy roots, is expected to play a crucial role in assessing their metabolic performance.
In this work, we implemented a bond labeling experiment (Szyperski, 1995) on *C. roseus* hairy roots, and employed the bondomer concept (Sriram and Shanks, 2001; 2004a; van Winden et al., 2002) to interpret and evaluate fluxes from the labeling data.

**Materials and methods**

**Hairy root culture**

*C. roseus* LBE-6-1 hairy roots were cultured at 26°C in 50 mL liquid medium containing 30 gL\(^{-1}\) sucrose (5% w/w U-\(^{13}\)C) and 1.65 gL\(^{-1}\) Gamborg B-5 salts (Sigma), as per procedures outlined in Bhadra and Shanks (1997). Hairy roots were harvested after a growth period of 15 days. They were freeze-dried for 72 h and weighed. The hairy roots were then ground and stored at -80 °C for extraction and analysis of biomass components.

**Protein extraction, hydrolysis, amino acid quantification, and NMR sample preparation**

Protein was extracted from ground samples in 100 mM phosphate buffer (pH 7.2), at 4 °C for 15 min. The extract was repeated four times, and the consolidated supernatant was assayed for protein using the Bradford test (Bio-Rad, Hercules, CA).

Protein hydrolysis was performed in hydrolysis tubes (Pierce Endogen, Rockford, IL), to which 6N hydrochloric acid was added in the ratio 0.5 mL HCl:400 µg protein. Prior to hydrolysis, the hydrolysis tube was evacuated, flushed with nitrogen to remove residual oxygen, and re-evacuated. Hydrolysis was performed at 150°C for 4 h. The acid in the hydrolysate was evaporated in a Rapidvap evaporator (Labconco, Kansas City, MO). The residue was re-dissolved in 2 mL deionized water, lyophilized for 72 h, and dissolved in 500 µL D\(_2\)O in an NMR tube. The pH of the NMR sample was adjusted to 0.5 using DCI. Amino acids in the sample were quantified by HPLC, after derivatization with phenylisothiocyanate
to produce phenylthiocarbamyl amino acid derivatives, which were eluted by a reverse phase C\textsubscript{18} silica column, with detection at 254 nm.

**Extracellular fluxes and fluxes contributing to biomass**

Biomass growth was measured by dry weight measurements on \textsuperscript{12}C duplicates, and verification of the final biomass with the harvested \textsuperscript{13}C culture. Biomass composition was measured in another work (Chapter 7). Extracellular metabolites (sucrose, glucose, fructose, pyruvate, malate, succinate) were measured by HPLC. The HPLC system (Waters, Milford, MA) consisted of: two model 515 HPLC pumps, a model 717 auto-sampler, a model 2414 refractive index (RI) detector, a model 680 automated gradient controller, and a carbohydrate analysis column (Waters model 084038, 30×3.9 cm, pore size 125 Å). The mobile phase was a 75:25 mixture of acetonitrile and deionized water. The flow rate was maintained at 0.5 mL min\textsuperscript{-1}, and the pumps were used with an operating pressure of 2.6 to 4.0 MPa. Injection volumes were 15μL, and the time required to run a sample was 60 min. The temperature was maintained at 24 °C. Standards and samples were injected on the same day. The retention times were determined by injecting a 15μL sample. Data acquisition and analysis were performed with Empower™ software (Waters).

**NMR spectroscopy**

2-D \textsuperscript{13}C, \textsuperscript{1}H heteronuclear single quantum correlation (HSQC) NMR spectra (Bodenhausen and Ruben, 1980) were collected on a Bruker Avance DRX 500 MHz spectrometer, at 298 K. The reference to zero ppm was set using the methyl signal of dimethylsilapentanesulfonate (Sigma, St. Louis, MO) as an internal standard. The resonance frequency of \textsuperscript{13}C and \textsuperscript{1}H were 125.7 MHz and 499.9 MHz respectively. The spectral width was 5482.26 Hz along the \textsuperscript{1}H (F2) dimension and 5028.05 Hz along the \textsuperscript{13}C (F1) dimension. Peak aliasing was used in order to minimize the sweep width along the F1 dimension. The number of complex data points was 1024 (\textsuperscript{1}H) × 900 (\textsuperscript{13}C). A modification of the INEPT pulse sequence was used for
acquiring HSQC spectra (Bodenhausen and Ruben, 1980). The number of scans was generally set to 16. Assignment of amino acid peaks on the HSQC spectrum was verified using 2D \([\text{H}, \text{H}]\) total correlation (TOCSY) and 3D \([\text{C}, \text{H}, \text{C}]\) TOCSY spectra (Braunschweiler and Ernst, 1983), which were acquired with a 100% labeled protein sample. While acquiring TOCSY spectra, the DIPSI-2 sequence (Shaka et al., 1988) was used for isotropic mixing, with a mixing time of 76 ms.

The software Xwinnmr™ (Bruker) was used to acquire all spectra, and the software NMRView™ (Johnson and Blevins, 1994; available at http://onemoonscientific.com/nmrview, last accessed 20 July 2004) was used to quantify non-overlapping multiplets on the HSQC spectrum. To quantify overlapping multiplets (\(\alpha\)-amino acid and LVA peaks), which could not be processed with NMRView, a peak deconvolution software was written. This software was based on a spectral model proposed by van Winden et al. (2001). Additionally, 2-D spectra were obtained that were J-scaled along the F1 dimension, by incorporating pulse sequences described by Willker et al. (1997) and Brown (1984) into the HSQC pulse sequence. J-scaling increases multiplet separation by an even integral factor J and eliminates multiplet overlap. J-scaling factors of 6 or 8 were employed.

**Flux evaluation methodology and computer program**

We developed a computer program (NMR2Flux) that automatically evaluates fluxes from multiplet intensity data of the type reported in this paper. This program accepts any metabolic network model (in terms of reaction stoichiometries and carbon skeleton rearrangements) in a spreadsheet format (e.g. Table 1 of Sriram and Shanks, 2004a). The version of the program employed in this work uses bondomer balancing, which results in a large reduction of computation time, compared to isotopomer balancing (Sriram and Shanks,
NMR2Flux is implemented in the programming language C, on the Red Hat Linux operating system.

**Results and discussion**

Biomass growth, and sucrose consumption were measured as described in Materials and Methods. In the present study (15 d old or mid-exponential phase hairy roots), sucrose was the only metabolite detected in the medium. Sucrose hydrolysis products (glucose, fructose) and organic acids were below HPLC detection limits, although substantial amounts of glucose and fructose were found in the medium for 30 d old (late-exponential) hairy roots (data not shown). In an earlier study in our laboratory (Bhadra and Shanks, 1997), small amounts of glucose and fructose (< 1 g L$^{-1}$) were measured in the medium during the mid-exponential phase, and larger amounts (up to 10 g L$^{-1}$) were measured in late-exponential phase, which agree with the above results. Biomass composition (and hence, the fluxes contributing to biomass) were measured in another work (Chapter 7).

We acquired a 2-D [$^{13}$C, $^1$H] HSQC NMR spectrum of the protein hydrolysate from *Catharanthus roseus* hairy roots. The spectrum contained 50–60 proteinogenic amino acid and sugar hydrolysate peaks. Expanded views of selected peaks are shown in Figure 4.1. These peaks contain information about metabolic fluxes. As an instance, consider Ile γ$^2$ (Figure 4.1A). Singlet peaks occur due to $^{13}$C-$^{12}$C fragments in metabolite molecules, whereas doublet peaks occur due to $^{13}$C-$^{13}$C fragments. Since we are employing a mixture of [5% $^{13}$C, 95% $^{12}$C] sucrose as the sole carbon substrate, the $^{13}$C-$^{12}$C fragments are most likely formed by biosynthetic assembly of carbon fragments from two different substrate molecules ($^{13}$C and $^{12}$C), while the $^{13}$C-$^{13}$C fragments, are most likely intact fragments from a $^{13}$C sucrose molecule. This indicates that Ile (γ$^2$) nuclei have two distinct metabolic histories, most likely due to two different pathways that went into the making of the molecule. For
example, when the glycolytic and pentose phosphate pathways operate together, two different types of Ile molecules can be formed, which would give rise to the singlet-doublet pattern observed in Figure 1A. A quantification of relative intensities of the singlet and doublet peaks, followed by mathematical modeling, provided the branchpoint split ratios in a flux map. In this manner, various branchpoint split ratios were obtained, thus completing the flux map.

We also acquired a 2D [$^1\text{H}, ^1\text{H}$] TOCSY spectrum to quantify positional enrichments of the proteinogenic amino acids and sugar hydrolysates observed in the spectrum. The average enrichment was measured to be $0.0552 \pm 0.0033$, which agrees well with the value $0.0560$ calculated from the labeling extent of the substrate, natural $^{13}\text{C}$ abundance, and dilution by initially present hairy root biomass. This indicates a uniform uptake of the $^{13}\text{C}$ label, in the same $^{13}\text{C}:^{12}\text{C}$ proportion as in the medium.

Fluxes were calculated from the labeling data using bondomer balancing, and a model of central carbon metabolism in C. roseus hairy roots (Figure 4.2). External fluxes were measured, and biomass fluxes were calculated in a previous work (Chapter 7).

Figure 4.3 depicts the fluxes through the key pathways, in this system. Most fluxes are well-identified, with the exception of the GABA shunt. We have calculated in another work (Chapter 5) that large proportions of labeled glutamine may be needed to identify the GABA shunt satisfactorily. Figure 4.4 depicts how closely the evaluated fluxes explain the measured labeling data.

We observed that the isotopomer abundances of Leu $\delta^1$ [singlet (s): $0.474 \pm 0.002$] did not agree with the abundances other amino acids [Val $\gamma^1$ (s): $0.323 \pm 0.002$; Ile $\gamma^2$ (s): $0.326 \pm 0.003$] generated from the same metabolic precursor, plastidic pyruvate. We have observed this in soybean embryos also (Chapter 3), and have ruled out NMR or other experimental
artifacts. It is possible that the known biochemistry of Leu biosynthesis (assumed in our model) is incorrect, which would explain the result.

We observed rather small pentose phosphate pathway fluxes compared to our previous work with soybean embryos (Chapter 2). This result could be explained by the fact that soybean embryos synthesize large amounts of protein and lipids, both of which require the reductant NADPH provided by the pentose phosphate pathway. The NADPH-requiring processes in this system, namely the synthesis of lignin and proteins, did not constitute a substantial NADPH demand.

However, we observed a high cycling between the anaplerotic fluxes PEP → OAA (cytosol) and Mal → Pyr (mitochondrion). This is an apparently futile cycle, and does not directly contribute to biosynthesis. One likely explanation for this result is provision of reductant (NADH or NADPH) in the mitochondrion. The Mal → Pyr conversion, catalyzed by the malic enzyme, can provide NAD(P)H, so that the PEP$^c$ → OAA$^c$ → OAA$^m$ → Mal$^m$ → Pyr$^m$ pathway provides one mole of NAD(P)H, whereas the PEP$^c$ → Pyr$^c$ → Pyr$^m$ pathway does not provide any reductant. Therefore, the anaplerotic cycle may supplement the NADH provided by the TCA cycle for respiration. We observed that the respiration rate of the mid-exponential phase hairy roots was high. [The respiration rate was calculated from the rate of carbon dioxide production (obtained from a carbon balance), which consumed over 75% of the carbon from sucrose.] Such a high rate of respiration may require more NADH than can be provided by the TCA cycle, which explains the substantial anaplerotic cycling. From the evaluated fluxes, we calculated the NADH supply by the TCA cycle to be 6.254 ± 0.178 mmol d$^{-1}$ g biomass$^{-1}$, and that provided by the anaplerotic cycle to be 0.438 ± 0.248 mmol d$^{-1}$ g biomass$^{-1}$. 
Another explanation for the anaplerotic cycle could be that it replenishes the carbon lost from the αKG node of the TCA cycle toward nitrogen fixation. However, the total protein and amino acid content of the hairy roots was rather small to justify a substantial anaplerotic flux. We were unable to identify a plastidic flux from F6P to T3P, in this system. The values of the cumulative $\chi^2$ error did not change by when this reaction was included in our model. This is likely because labeling data from starch hydrolysate was not measured in this work. However, it is known that fructose-1,6-bisphosphatase, the enzyme responsible for this flux, does not exist in heterotrophic plastids such as our system. Embryos are considered and shown (Chapter 2) to be the only exceptions.

The methodology presented in this paper is expected to be valuable in assessing the impact of genetic or environmental changes in the hairy roots.

References


**Acknowledgments**

This material is based upon work supported by the National Science Foundation under grants BES 9906978 and BES 0224600, and the Plant Sciences Institute at Iowa State University.

**Figure captions**

**Figure 4.1.** Expanded views of $^{13}$C, $^1$H] HSQC spectrum of *Catharanthus roseus* hairy roots growing on (5% w/w U-$^{13}$C) sucrose. (A) (upper figure) Ile γ and (B) (lower figure) Asp α cross peaks. 1-D slices are shown alongside. The multiplet peaks are s: singlet; d, d1, d2: doublet; dd: double doublet.

**Figure 4.2.** Metabolic flux map template consisting of primary and intermediate metabolic pathways in *Catharanthus roseus* hairy roots growing on (5% w/w U-$^{13}$C) sucrose. Intracellular metabolites are shown in white ovals, and gray ovals show sink metabolites (proteinogenic amino acids, polysaccharides, etc.) Metabolic pathways are color-coded as follows: dark red (glycolysis and sucrose metabolism), pale blue (pentose phosphate pathway), orange (TCA cycle), blue-gray (pyruvate dehydrogenase link), mauve (anaplerotic fluxes), dark yellow (glyoxylate shunt), green (glutamate metabolism, GABA shunt, and associated intercompartmental transport fluxes), gray (fluxes towards biomass synthesis, and black (all intercompartmental transport fluxes except those involved in Gln metabolism and GABA shunt). Abbreviations for intracellular metabolites are as follows: Sue (sucrose), G6P
(glucose-6-phosphate), F6P (fructose-6-phosphate), T3P (triose-3-phosphate), P5P (pentose-5-phosphate), S7P (sedoheptulose-7-phosphate), E4P (erythrose-4-phosphate), 3PG (3-phosphoglycerate), PEP (phosphoenolpyruvate), Pyr (pyruvate), acetCoA (acetyl CoA), iCit (isocitrate), aKG (α-ketoglutarate), Scn (succinate), Mal (malate), OAA (oxaloacetate), GABA (γ-aminobutyric acid), SSA (succinic semi-aldehyde), GOx (glyoxylate). Sink metabolite abbreviations are as follows: PSac (polysaccharides), Nuc (carbon skeleton of nucleotides), Sta (starch). Asp and Asn are denoted together as Asx. Glu and Gln are denoted together as Glx. F6P and T3P appear at two different locations each in the cytosol and plastid, to avoid confusing intersections of lines. Each flux is assigned a short name based on the name of the gene encoding one of the metabolic reactions represented by it. Intracellular metabolites and fluxes with a superscript are located in specific subcellular compartments – c: cytosol, p: plastid, m: mitochondrion. If a flux has no superscript, its compartmentation could not be unambiguously determined (such as gap, eno and pyk, and some fluxes toward biosynthesis).

Figure 4.3. Metabolic fluxes evaluated for Catharanthus roseus hairy roots growing on (5% w/w U-13C) sucrose. Absolute fluxes are expressed in mmol d^{-1} g biomass^{-1}. Error bars represent standard deviations of the fluxes evaluated from 200 Monte Carlo simulations.

Figure 4.4. Comparison of experimental and simulated multiplet intensities. The x-axis represents experimental multiplet intensities measured from [13C, 1H] spectra (error bars represent standard errors measured from signal/noise ration of the spectra); the y-axis represents relative intensities that were simulated by the computer program NMR2Flux, corresponding to the evaluated fluxes of Figure 3 (error bars represent standard errors of the simulated intensities from 200 simulations). Multiplet intensities are shown as fraction of total signal.
Figure 4.1. Expanded views of $^{13}\text{C}, \text{^1H}$ HSQC spectrum of *Catharanthus roseus* hairy roots.
Figure 4.2. Metabolic flux map template consisting of primary and intermediate metabolic pathways in *Catharanthus roseus* hairy roots growing on (5% w/w U-13C) sucrose.
w/ W-13C (C) sucrose.

Figure 4.3. Metabolic fluxes evaluated for Calothamnus rosesi growing on (5%)
Figure 4.4. Comparison of experimental and simulated multiplet intensities.
5. Flux identifiability-based optimal design of $^{13}$C labeling experiment for Glycine max (soybean) embryo metabolism

An early draft of a paper, to be expanded and submitted to Biotechnology and Bioengineering.

Ganesh Sriram§ and Jacqueline V. Shanks*

Department of Chemical Engineering, Iowa State University, Ames, IA 50011

Abstract

Flux identifiability is an important question in $^{13}$C metabolic flux analysis, and there is motivation to develop labeling experiments so that the flux information contained by the label measurements is optimal. In this work, optimal $^{13}$C labeling experiments were designed based on statistical flux identifiability, for soybean embryo central carbon metabolism. The statistical quality of the labeling experiment was adjudged based on the $D$-optimality criterion. First, we examined combinations of U-$^{13}$C sucrose and 1-$^{13}$C sucrose for their effect on statistical quality of the glycolysis and pentose phosphate pathway flux information provided by them. The measurement of the $^{13}$C enrichment of CO$_2$ does not have any effect on the statistical quality, but the measurement of labeling data in the nucleotide hydrolysate improves the quality by 15 to 20%. In all measurement scenarios, the best combination of U-$^{13}$C and 1-$^{13}$C sucrose was [13% U-$^{13}$C, 86% 1-$^{13}$C]. This optimal experiment reduces the standard deviation of pentose phosphate pathway influxes by a factor of 3 to 4. We also

§ Primary researcher and author.
* Principal investigator and corresponding author.
examined combinations of U-$^{13}$C sucrose and U-$^{13}$C glutamine for their effect on flux identifiability. When only the glycolysis and pentose phosphate pathway fluxes are of interest, U-$^{13}$C glutamine caused a maximum increase in statistical quality by a factor 1.2. However, when all fluxes are of interest, the quality increased by a factor 4.7, at the optimal substrate combination [5% U-$^{13}$C sucrose, 96% U-$^{13}$C glutamine]. This combination reduced the standard deviations for the pentose phosphate pathway influxes and the flux into the GABA shunt, by about 50% compared to the reference experiment [11% U-$^{13}$C sucrose, 0% U-$^{13}$C glutamine]. We expect these results to facilitate better $^{13}$C labeling experimental design for the soybean embryo system, and for plant systems with similar metabolic pathways.

**Introduction**

Metabolic flux analysis is a cornerstone of metabolic engineering. Fluxes and flux maps provide a basis for distinguishing between phenotypes and for quantifying the impact of genetic or environmental modifications. The information gained from flux analysis can provide insights toward selection of appropriate metabolic engineering targets, or for the construction of predictive models of metabolism (Stephanopoulos, 1999). The use of $^{13}$C labeling followed by NMR (nuclear magnetic resonance) analysis is a powerful method of flux quantification (Stephanopoulos, 2002; Wiechert, 2001), and is gaining popularity in metabolic engineering.

Flux identifiability is an important question in $^{13}$C labeling-based flux analysis. This is because labeling data obtained from such experiments are nonlinear functions of the fluxes, and it is often not obvious if the fluxes in a metabolic network are indeed identifiable (or observable) from the data. Two aspects of flux identifiability are recognized – structural and statistical identifiability (Isermann and Wiechert, 2003). In structural identifiability, the
labeling data are assumed to be perfect or noise free, and it is determined whether or not the observed measurements correspond to a unique set of fluxes. van Winden *et al.* (2001) made some important advances toward structural identifiability from isotopomer data; however, this analysis could not be extended to large or physically realizable systems because of the highly nonlinear nature of the measurement sensitivity equations. Sriram and Shanks (2004a) showed that in carbon bond labeling experiments (Szyperski, 1995), the bondomer concept can be effectively employed to simplify the system state, and make structural identifiability possible for realistic systems. Isermann and Wiechert (2003) provided a comprehensive analysis of structural identifiability theory.

In statistical identifiability, the errors in the labeling data are considered, and it is determined if, in spite of these errors, the measurements correspond to a unique flux distribution, with reasonably small standard deviations for the fluxes. Moreover, since flux identifiability and flux standard deviations are substantially altered by varying the labeling state of the substrates fed to the system (Möllney *et al.*, 1999), there exists potential to design ‘optimal experiments’ which best identify the fluxes in a network (or which contain maximum information about the fluxes). This is usually a numerical endeavor, unlike structural identifiability which is usually an analytical endeavor.

This paper focuses on statistical identifiability-based optimal $^{13}$C experiment design for *Glycine max* (soybean) embryos. In a previous work (Sriram *et al.*, 2004b), we evaluated fluxes in three compartments from $^{13}$C isotopomer data obtained from soybean embryos cultured on sucrose (10% w/w $^{13}$C, 90 % w/w natural abundance), and glutamine (natural abundance). It was felt that some fluxes in this system could be better identified by suitable selection of the combinations and $^{13}$C-labeling extents of the substrates. The objective of this work was to design such optimal carbon substrate combinations for improving flux identifiability in soybean embryo metabolism.
Methods

A metric of statistical quality of a $^{13}$C labeling experiment is required to gauge the quality of the experiment, and to compare how an experiment with a certain choice of substrates performs in relation to another experiment with a different choice of substrates. Since fluxes are the outcomes of $^{13}$C labeling experiments and improvement (reduction) of the flux standard deviations is the objective of optimal experiment, it is reasonable to use a scalar based on the flux standard deviations as a quality metric. In this work, we employed the reciprocal of the geometric mean of the standard deviations of the fluxes of interest, as a quality metric.

The computational approach in this work is along the lines of Möllney et al. (1999). Let $v$, $x$, and $I$ denote vectors representing the fluxes, isotopomer abundances, and labeling data (multiplet intensities measured by NMR) of a metabolic network. It is well-known that the relationship between $v$ and $x$

$$x = f(v)$$  \hspace{1cm} (1)

is nonlinear (Wiechert and Murzel, 2001), whereas the relationship between $x$ and $I$ is linear, and can be expressed by a matrix multiplication (Möllney et al., 1999)

$$I = M \cdot x$$  \hspace{1cm} (2)

Further, all the fluxes in $v$ are not linearly independent but are stoichiometrically related. A linearly independent subset of the fluxes, the flux parameters (represented by the vector $p$) can be related to the fluxes as

$$v = G \cdot p$$  \hspace{1cm} (3)

Together, these relationships can be consolidated into a nonlinear relationship between the flux parameters and the labeling data.
The D-criterion (Takors et al., 1997; Möllney et al., 1999) is a popular statistical quality metric. It represents the volume of the confidence ellipsoid of the evaluated flux parameters, and is equal to the covariance matrix of the flux parameters:

\[ D = \text{Cov}(p) \]  

The standard deviation of a flux parameter can be calculated as:

\[ \text{SD}(p_i) = \sqrt{C_{ii}} \]

where \( C_{ii} \) is the \( i \)th diagonal element of \( \text{Cov}(p) \).

The covariance matrix can be evaluated by obtaining a distribution of fluxes by Monte Carlo simulations (Press et al., 1992) or by differentiating the relationship in Eq. (4). This differentiation can be done analytically (Wiechert et al., 1999) or numerically. In this work, numerical differentiation was employed. The covariance matrix was built from the partial derivatives \( \frac{\partial I}{\partial p_j} \), using standard statistical formulae (Press et al., 1992). Also, submatrices of the covariance matrix were considered rather than the entire covariance matrix, when the identifiability of certain fluxes is of interest.

The scalar \( D \) is the volume of the confidence ellipsoid of the evaluated fluxes. Its \( n \)th root \( \sqrt[n]{D} \) (where \( n \) is the number of flux parameters) is geometric mean of the standard deviations of the flux parameters. We employed the reciprocal of this geometric mean (designated \( I \)) as the quality metric of a labeling experiment. It can be inferred that if a labeling experiment is better than a reference experiment, then its value of \( I \) should be higher than that of the reference experiment.

All computations of \( I \) are reported with respect to a reference experiment, which is usually the experiment performed in our previous work (Sriram and Shanks, 2004b) with [11\% U-13C sucrose] as the only labeled substrate.
We examined combinations of U-\(^{13}\)C sucrose, 1-\(^{13}\)C sucrose, and U-\(^{13}\)C glutamine for their ability to provide an improved labeling data set, which would identify fluxes in the soybean embryo network better. Our previous work had employed multiplet intensity or positional enrichment measurements from protein and starch hydrolysates. We examined the effects of other measurements (CO\(_2\) enrichment; labeling data from nucleotide hydrolysate which reflect the precursor plastidic pentose phosphate) on flux identifiability. The values of fluxes at which identifiability was determined, were the ones evaluated in our previous work (Sriram and Shanks 2004b; see Figure 5.1).

Note: Throughout this paper, when substrate combinations are denoted as [11\% U-^{13}\)C sucrose, 0\% 1-^{13}\)C sucrose], it should be understood that the remainder of the substrate at natural abundance.

**Results and discussion**

We examined combinations of U-\(^{13}\)C sucrose and 1-\(^{13}\)C sucrose for their effect on flux identifiability, when only the glycolysis and pentose phosphate pathway fluxes are of interest. The determination of fluxes in this subset of metabolism is of much interest in the metabolic engineering community. Figure 5.2 depicts the statistical quality for various combinations, relative to the reference experiment [11\% U-^{13}\)C sucrose, 0\% 1-^{13}\)C sucrose], when the multiplet intensities and positional enrichments in protein and starch hydrolysates (case 1). Figure 5.3 depicts the corresponding statistical quality when CO\(_2\) enrichment is also measured (case 2). The measurement of CO\(_2\) enrichment improves the statistical quality only marginally (by about 0.05\%). A possible explanation for this observation is that the information from CO\(_2\) could be concealed in other measurements made from the protein and starch hydrolysates. Figure 5.4 depicts the statistical quality when labeling data from nucleotide hydrolysate is also measured (case 3). We observed that the information from the
nucleotide hydrolysate improves the information content by 15 to 20%. The best combination of U-$^{13}$C and 1-$^{13}$C sucrose for the cases 1, 2, and 3 was the same [13% U-$^{13}$C, 86% 1-$^{13}$C]. Also, within each case, when only U-$^{13}$C sucrose is used, the quality does not improve compared to the reference experiment.

Figure 5.5A shows the statistical quality of the reference experiment [11% U-$^{13}$C, 0% 1-$^{13}$C] and the optimal experiment for the cases 1, 2, and 3. Compared to the reference experiment of case 1, the optimal experiment of case 3 provides an improvement in the statistical quality by a factor 1.6. Figure 5.5B shows the improvements afforded by these substrate combinations, on the standard deviations of the cytosolic and plastidic pentose phosphate pathway fluxes. In all three cases, the optimal experiment reduces the standard deviation of these fluxes, by a factor of four (cases 1 and 2) or three (case 3). The measurements in case 3, however, reduce the standard deviations only by a factor 1.25, compared to corresponding experiments in case 1 or 2. The optimal experiment (which requires a rather high amount of label) has the disadvantage that it may be expensive to implement multiple times or on a large scale. Whereas, the case 3 measurements require an additional extraction and hydrolysis step, which may by time-consuming. The identifiability results presented here should facilitate intelligent choices of labels and/or measurements, for identifying the pentose phosphate pathway in soybean embryos, and for distinguishing between phenotypes that are likely to exhibit differences in the glycolysis and pentose phosphate pathway fluxes.

We also examined combinations of U-$^{13}$C sucrose and U-$^{13}$C glutamine for their effect on flux identifiability. Two cases were considered: when only the glycolysis and pentose phosphate pathway fluxes are of interest (case 4) and when all fluxes in the network are of interest (case 5). Figure 5.6 depicts the statistical quality of for various combinations of U-$^{13}$C sucrose and U-$^{13}$C glutamine, for case 4. Figure 5.7 depicts the corresponding
information for case 5. Statistical quality is shown relative to the reference experiment [11% U-$^{13}$C sucrose, 0% U-$^{13}$C glutamine].

In case 4, the quality does not exceed 1.3 for the best combination. Also, the contour lines are in general parallel to the U-$^{13}$C glutamine axis. These is can be explained by our observation that in the soybean embryo system, the carbon substrate glutamine does not substantially contribute to metabolites in the glycolysis and pentose phosphate pathways, which are predominantly synthesized from sucrose (Sriram et al., 2004b). The labeling state of glutamine, may however affect the enrichment of CO$_2$, if measured. From the results, it is clear that such an effect dominates only when the U-$^{13}$C glutamine:U-$^{13}$C sucrose ratio is high.

However, when all fluxes in the network are of interest, the fraction of labeled glutamine plays a significant role in flux identifiability. The optimal substrate combination with the best statistical quality, was found to be [5% U-$^{13}$C sucrose, 96% U-$^{13}$C glutamine]. Figure 5.8 shows the improvements in the flux standard deviations of this optimal experiment, compared to the reference experiment. The standard deviations for the pentose phosphate pathway influxes and the flux into the GABA shunt drop by about 50%. The results of case 5 should enable better planning of $^{13}$C experiments, in phenotype comparisons where all or most fluxes are likely to be different.

In the present work, we have shown that additional measurements and/or different combinations of labeled substrates can improve the identifiability of fluxes in soybean embryo metabolism, over our previous work (Sriram et al., 2004b). We expect these results to facilitate better $^{13}$C labeling experimental design for the soybean embryo system, and for plant systems with similar metabolic pathways.

A significant problem in statistical flux identifiability analysis is that such analysis requires values of the fluxes a priori. Thus determination of the best combination of substrates and
measurements is a cyclic process. In this work, we employed the values of the fluxes evaluated for that system using $^{13}$C experiments with non-optimal substrate combination. However, when deciding the flux identifiability of a system for which no $^{13}$C experiments have been performed, we recommend that approximate flux values obtained from stoichiometric flux-balance models be used.

**Acknowledgments**

This material is based upon work supported by the National Science Foundation under grants BES 9906978 and BES 0224600, and the Plant Sciences Institute at Iowa State University.

**References**


developing soybean embryos by employing biosynthetically directed fractional $^{13}$C labeling, 2-D [$^{13}$C, $^1$H] NMR and comprehensive isotopomer balancing. *Plant Physiol.* Under revision.


**Figure captions**

**Figure 5.1.** Metabolic flux map for the plant system considered in this work (developing *Glycine max* (soybean) embryos cultured on sucrose and glutamine). Arrow widths are directly proportional to fluxes. Subscripts on reaction names: f: forward reaction, b: backward reaction. Superscripts: c: cytosol, p: plastid, m: mitochondrion. If a flux has no
superscript, its compartmentation could not be unambiguously determined. Net fluxes are in the direction of the reaction with the subscript 'f'.

**Figure 5.2.** Optimal experiment design for *Glycine max* (soybean) embryos growing on sucrose and glutamine, case 1. In this case, the isotopomer abundances and positional enrichments from all amino acids and sugars in the protein and starch hydrolysates are measured, whereas those in CO₂ and nucleotide hydrolysates are not measured. Statistical quality is calculated for the glycolysis and pentose phosphate pathway fluxes (in the cytosol and plastid), for various extents of two labeled substrates, U⁻¹³C sucrose (y-axis) and 1⁻¹³C sucrose (x-axis) in the sucrose feed. Statistical quality is shown relative to the quality of the reference experiment [11% U⁻¹³C, 0% 1⁻¹³C] of this case, which is taken to be 1 unit.

**Figure 5.3.** Optimal experiment design for *Glycine max* (soybean) embryos growing on sucrose and glutamine, case 2. In this case, the isotopomer abundances and positional enrichments from all amino acids and sugars in the protein and starch hydrolysates, and CO₂, are measured, whereas those in nucleotide hydrolysates are not measured. Statistical quality is calculated for the glycolysis and pentose phosphate pathway fluxes (in the cytosol and plastid), for various extents of two labeled substrates, U⁻¹³C sucrose (y-axis) and 1⁻¹³C sucrose (x-axis) in the sucrose feed. Statistical quality is shown relative to that of the reference experiment [11% U⁻¹³C, 0% 1⁻¹³C] of case 1, which is taken to be 1 unit.

**Figure 5.4.** Optimal experiment design for *Glycine max* (soybean) embryos growing on sucrose and glutamine, case 3. In this case, the isotopomer abundances and positional enrichments from all amino acids and sugars in the protein and starch hydrolysates, CO₂, and nucleotide hydrolysates are measured. Statistical quality is calculated for the glycolysis and pentose phosphate pathway fluxes (in the cytosol and plastid), for various extents of two labeled substrates, U⁻¹³C sucrose (y-axis) and 1⁻¹³C sucrose (x-axis) in the sucrose feed.
Statistical quality is shown relative to that of the reference experiment [11% U-^{13}C, 0% 1-{^13}C] of case 1, which is taken to be 1 unit.

**Figure 5.5 (A)** Comparison of statistical qualities of reference experiments [11% U-^{13}C, 0% 1-{^13}C; hollow bars] and optimal experiments [13% U-^{13}C, 86% 1-{^13}C; filled bars] in cases 1, 2, 3. Statistical quality is shown relative to that of the reference experiment case 1, which is taken to be 1 unit. (B) Comparison of standard deviations of the pentose phosphate pathway influx in cytosol (pglc) and plastid (pglp) of reference experiments and optimal experiments in cases 1, 2, 3. Bars are explained in legend. Standard deviations are shown relative to that of the reference experiment of case 1, which is taken to be 1 unit.

**Figure 5.6.** Optimal experiment design for *Glycine max* (soybean) embryos growing on sucrose and glutamine, case 4. In this case, the isotopomer abundances and positional enrichments from all amino acids and sugars in the protein and starch hydrolysates, CO₂, and nucleotide hydrolysates are measured. Statistical quality is calculated for the glycolysis and pentose phosphate pathway fluxes (in the cytosol and plastid), for various extents of two labeled substrates, U-^{13}C sucrose (y-axis) and U-^{13}C glutamine (x-axis). Statistical quality is shown relative to that of the reference experiment [11% U-^{13}C Suc, 0% U-^{13}C Gln] of this case, which is taken to be 1 unit.

**Figure 5.7.** Optimal experiment design for *Glycine max* (soybean) embryos growing on sucrose and glutamine, case 5. In this case, the isotopomer abundances and positional enrichments from all amino acids and sugars in the protein and starch hydrolysates, CO₂, and nucleotide hydrolysates are measured. Statistical quality is calculated all fluxes, for various extents of two labeled substrates, U-^{13}C sucrose (y-axis) and U-^{13}C glutamine (x-axis). Statistical quality is shown relative to that of the reference experiment [11% U-^{13}C Suc, 0% U-^{13}C Gln] of this case, which is taken to be 1 unit.
Figure 5.8. Comparison of standard deviations of fluxes for reference experiments [11% U-$^{13}$C Suc, 0% U-$^{13}$C Gln; hollow bars] and optimal experiments [5% U-$^{13}$C Suc, 96% U-$^{13}$C Gln; filled bars] in case 5. The abbreviations for fluxes are: cytosolic pentose phosphate pathway influx ($pglc$), plastidic pentose phosphate pathway influx ($pglp$), anaplerotic flux phosphoenolpyruvate carboxylase (cytosol) from phosphoenolpyruvate $\rightarrow$ oxaloacetate ($ppcc$), anaplerotic flux malic enzyme (plastid) from malate $\rightarrow$ pyruvate ($mep$), glyoxylate shunt flux ($iclm$), GABA shunt flux ($gdhc$). Standard deviations are shown relative to that of the cytosolic pentose phosphate pathway influx ($pglc$), which is taken to be 1 unit.
Figure 5.1. Metabolic flux map for the plant system considered in this work (developing Glycine max (soybean) embryos cultured on sucrose and glutamine).
Figure 5.2. Optimal experiment design for *Glycine max* (soybean) embryos growing on sucrose and glutamine, case 1.
Figure 5.3. Optimal experiment design for *Glycine max* (soybean) embryos growing on sucrose and glutamine, case 2.
Figure 5.4. Optimal experiment design for *Glycine max* (soybean) embryos growing on sucrose and glutamine, case 3.
Figure 5.5 Comparison of statistical qualities and flux standard deviations in reference experiments and optimal experiments in cases 1, 2, 3.
Figure 5.6. Optimal experiment design for *Glycine max* (soybean) embryos growing on sucrose and glutamine, case 4.
Figure 5.7. Optimal experiment design for *Glycine max* (soybean) embryos growing on sucrose and glutamine, case 5.
Figure 5.8. Comparison of standard deviations of fluxes for reference experiments [11% U-$^{13}$C Suc, 0% U-$^{13}$C Gln; hollow bars] and optimal experiments [5% U-$^{13}$C Suc, 96% U-$^{13}$C Gln; filled bars] in case 5.
6. Levulinic acid in soybean protein hydrolysate: Identification, 2-D $^{13}$C, $^1$H NMR study, and implications for metabolic investigations

A paper to be submitted to Journal of Agricultural and Food Chemistry.

Ganesh Sriram¹, Vidya V. Iyer¹, D. Bruce Fulton², and Jacqueline V. Shanks³

Departments of Chemical Engineering, Biochemistry, Biophysics and Molecular Biology; Iowa State University, Ames, IA 50011

Abstract

The use of two-dimensional $[^{13}$C, $^1$H] NMR spectra of fractionally labeled proteinogenic amino acids for metabolic investigations is gaining popularity. In the 2-D $[^{13}$C, $^1$H] spectra of protein isolated from soybean embryos, we observed four intense peaks that could not be assigned to the any proteinogenic amino acids. We show that these peaks correspond to degradation products of glucose and mannose, and infer that these arise from the acid degradation of the large amounts of glycosylating sugars (glucosamine and mannose) associated with soybean embryo protein. Further, the degradation products were identified as levulinic acid (4-oxo-pentanoic acid, major product) and 5-hydroxymethyl furfural (minor product). We suggest that the peaks of the degradation products $[^{13}$C, $^1$H] be used to glean metabolic information about the glycosylating sugars. To this end, we demonstrated that the

¹ Primary researcher and author.
² Graduate student in J. V. Shanks lab, assisted with hydrolyses and NMR analyses.
³ NMR facility manager.
* Corresponding author.
isotopomeric compositions of the parent hexoses are identical to those of the degradation products. Furthermore, a 2-D NMR study of levulinic acid was conducted. This included the measurement of the J-coupling constants ($J_{CC-s}$) between the atoms of levulinic acid. A long-range coupling was observed between atoms C$_3$ and C$_5$ of levulinic acid. Also, 2-D [$^{13}$C, $^1$H] NMR spectra of levulinic acid that were J-scaled along the $^{13}$C dimension provided additional information compared to the corresponding spectra of the parent hexose. This renders the degradation products levulinic acid and 5-hydroxymethyl furfural promising indicators of metabolism. In particular, we show that a comparison of the isotopomer abundances of levulinic acid between [$^{13}$C, $^1$H] spectra of soybean protein hydrolysate and starch hydrolysate, can be used to distinguish separate glycolysis and pentose phosphate pathways in the cytosol and plastid compartments.

**Keywords:** Levulinic acid, soybean (*Glycine max*), protein hydrolysate, nuclear magnetic resonance (NMR), J-coupling, plant metabolism.

**Introduction**

The analysis of hydrolysates of fractionally $^{13}$C-labeled protein from bacteria or fungi by two-dimensional (2-D) nuclear magnetic resonance (NMR) is a popular method to obtain information on metabolic pathways and fluxes (Szyperski, 1995, 1998; Wiechert, 2001; Sauer, 2004). A similar analysis of plant protein hydrolysates holds promise as a powerful method to shed light on plant metabolism (Shachar-Hill, 2001); however, such an analysis of a plant protein hydrolysate has not been reported to date.

During such an analysis of fractionally $^{13}$C-labeled protein from developing soybean embryos, we observed some peaks on the 2-D [$^{13}$C, $^1$H] spectrum that could not be assigned to any amino acid present in the protein hydrolysate. The objective of this study was to
identify these peaks, and investigate the possibility of using any information contained in them, toward metabolic investigations.

We designed NMR experiments to show that the peaks are acid degradation products of glucose and mannose, and their sources are the large amounts of glycosylating sugars (mostly glucosamine and mannose) in soybean embryo protein. The unknown peaks were identified as belonging to levulinic acid and 5-hydroxymethyl furfural. We measured the hitherto unknown NMR coupling constants of levulinic acid, and explored characteristics of its NMR spectra. Furthermore, we showed that they preserve the metabolic information contained in the hexose they are produced from. Therefore, the metabolic information in them can be used toward pathway or flux analysis.

Levulinic acid has been reported by researchers as a degradation product of glucose. However, its detection in soybean protein hydrolysate, its study by 2-D \([^{13}\text{C}, ^1\text{H}]\) NMR and from a metabolic flux analysis perspective, as done in this paper, are novel.

**Materials and methods**

**Chemicals**

All labeled hexose sugars (glucose labeled at various positions, mannose), and levulinic acid were purchased from Sigma (St. Louis, MO).

**Soybean cotyledon culture and protein extraction**

Soybean (*Glycine max* cv. Evans) was grown in a growth chamber at 27 °C/20 °C and 14 h photoperiod. 18 days after flowering, pods were harvested from the central section of the main stem and embryos isolated for in vitro culture. Three cotyledons of uniform size were cultured aseptically, in 20 mL liquid medium containing 146 mM sucrose (10% w/w U-\(^{13}\text{C}\)), 37 mM glutamine as the only carbon sources. After 6 d of culture, cotyledons were
harvested, rinsed with non-labeled medium, and lyophilized at -50°C and 133 × 10⁻³ mbar for 72 h. The lyophilized embryos were finely ground for further processing.

Protein was extracted from ground samples in 100 mM phosphate buffer (pH 7.2), at 4 °C for 15 min. The extract was repeated four times, and the consolidated supernatant was assayed for protein using the Bradford test (Bio-Rad, Hercules, CA).

Hydrolysis of protein and sugars

Hydrolysis of soybean embryo protein and sugars was performed in hydrolysis tubes (Pierce Endogen, Rockford, IL), to which 6N hydrochloric acid was added in the ratio 0.5 mL HCl:400 μg protein (or sugar). To prevent oxidation during hydrolysis, the hydrolysis tube was evacuated, flushed with nitrogen, and re-evacuated. Hydrolysis was performed at 150°C for 4 h. The acid in the hydrolysate was evaporated in a Rapidvap evaporator (Labconco, Kansas City, MO). The residue was re-dissolved in 2 mL deionized water, lyophilized for 72 h, and constituted in 500 μL D₂O in an NMR tube. The pH of the NMR sample was adjusted to 0.5 using DCl.

Amino acids in the protein hydrolysate were quantified by HPLC, after derivatization with phenylisothiocyanate to produce phenylthiocarbamyl amino acid derivatives, which were eluted by a reverse phase C₁₈ silica column, with detection at 254 nm.

NMR spectroscopy

2-D [¹³C, ¹H] heteronuclear single quantum correlation (HSQC) NMR spectra (Bodenhausen and Ruben, 1980) were collected on a Bruker Avance DRX 500 MHz spectrometer, at 298 K. The reference to zero ppm was set using the methyl signal of dimethylsilapentanesulfonate (Sigma, St. Louis, MO) as an internal standard. The resonance frequency of ¹³C and ¹H were 125.7 MHz and 499.9 MHz respectively. The spectral width was 5482.26 Hz along the ¹H (F2) dimension and 5028.05 Hz along the ¹³C (F1) dimension. Peak aliasing was used in
order to minimize the sweep width along the F1 dimension. The number of complex data points was 1024 ($^1$H) $\times$ 900 ($^{13}$C). A modification of the INEPT pulse sequence was used for acquiring HSQC spectra (Bodenhausen and Ruben, 1980). The number of scans was generally set to 16. Assignment of amino acid peaks on the HSQC spectrum was verified using 2D [$^1$H, $^1$H] total correlation (TOCSY) and 3D [$^{13}$C, $^1$H, $^1$H] TOCSY spectra (Braunschweiler and Ernst, 1983). While acquiring TOCSY spectra, the DIPSI-2 sequence (Shaka et al., 1988) was used for isotropic mixing, with a mixing time of 76 ms.

The software Xwinmr™ (Bruker) was used to acquire all spectra, and the software NMRView™ (Johnson and Blevins, 1994; available at http://onemoonscientific.com/nmrview, last accessed on 20 July 2004) was used to quantify non-overlapping multiplets on the HSQC spectrum. To quantify overlapping multiplets (certain cross peaks of LVA and $\alpha$-amino acids) which could not be processed with NMRView, a peak deconvolution software was written. This software was based on a spectral model proposed by van Winden et al. (2001).

**J-scaled spectra**

2-D [$^{13}$C, $^1$H] spectra were obtained that were J-scaled along the F1 dimension, by incorporating pulse sequences described by Willker et al. (1997) and Brown (1984) into the HSQC pulse sequence. J-scaling increases multiplet separation by an even integral factor J and eliminates multiplet overlap. J-scaling factors of 6 or 8 were employed.

**Results and discussion**

*Soybean embryo protein hydrolysate contains degradation products of glucose and mannose*
We found that the soybean embryo protein hydrolysate contained compounds that can be generated by the acid degradation of glucose and mannose. Figure 6.1A depicts the aliphatic region of a 2-D $^{13}$C, $^1$H HSQC spectrum of the hydrolysate of water-soluble protein extracted from soybean embryos grown on sucrose (10% w/w U-$^{13}$C) and glutamine in liquid culture. The cross peaks in this spectrum correspond to the constituents of the protein hydrolysate. Most cross peaks were assigned to the carbon atoms of 16 proteinogenic amino acids that were independently detected in the protein hydrolysate by HPLC. Assignment of cross peaks to amino acid atoms was done from literature data on their $^{13}$C and $^1$H chemical shifts and J-coupling constants (Wüthrich, 1976; Krivdin and Kalabin, 1989); and was confirmed by performing 2-D [$^1$H, $^1$H] and 3-D [$^{13}$C, $^1$H, $^1$H] TOCSY experiments on the protein hydrolysate. However, four intense cross peaks (marked A, B, C, and D) could not be assigned to any amino acid atom.

This suggested the presence of substantial amounts of carbon compounds other than proteinogenic amino acids, in the protein. We hypothesized that these cross peaks may have arisen from the conversion of glycosylating sugars in the protein, under the hydrolysis conditions. Soybean seed storage protein (most of the protein in the developing embryo) is highly glycosylated (Doyle et al., 1986), the attached sugars being predominantly mannose and glucosamine (Yamauchi and Yamagishi, 1979). To test this hypothesis, we hydrolyzed U-$^{13}$C glucose under the similar conditions as the protein, and acquired a 2-D [$^{13}$C, $^1$H] spectrum of the glucose hydrolysate. This spectrum contained cross peaks at the same positions as the unassigned cross peaks in the protein hydrolysate (Figure 6.1B). Also, these peaks were detected in a 2-D [$^{13}$C, $^1$H] spectrum of U-$^{13}$C mannose hydrolysate (not shown).
Identity of degradation products: Tracing of degradation product peaks to glucose carbon atoms

As a first step to probe the identity of the hexose hydrolysis product(s), we traced the atoms represented by the peaks A, B, C, and D to carbon atoms of glucose; and also deduced the functional groups of these atoms. The tracing of peaks A–D to glucose carbons was done by acquiring 2-D $^{13}$C, $^1$H] spectra of hydrolysates of glucose labeled at specific positions. For example, in a $^{13}$C, $^1$H] spectrum of 3-$^{13}$C glucose hydrolysate, the most intense cross peak was B (Figure 6.3). Therefore, the cross peak B of a glucose hydrolysis product corresponds to carbon atom C$_3$ of glucose. Although the cross peaks A, B, and C were also detected on the 3-$^{13}$C glucose hydrolysate spectrum, they had intensities at least 20-fold lesser than cross peak B, and resulted from the natural abundance of other carbon atoms in 3-$^{13}$C glucose.

Likewise, we observed that cross peaks A, C, and D correspond to glucose carbon atoms C$_1$, C$_5$, and C$_4$ respectively.

Furthermore, the $^{13}$C and $^1$H chemical shifts of the peaks B, C, and D (Table 6.1) suggest that they correspond to methyl (CH$_3$) or methylene (CH$_2$) groups. The $^{13}$C chemical shift (62.0) and $^1$H chemical shift of peak A suggests that it represents an alcoholic (CH$_2$OH) functional group.

A hydrolysate of 2-$^{13}$C glucose did not display any peak in a $^{13}$C, $^1$H] spectrum. Its 1-D $^{13}$C NMR spectrum contained a peak near 180 ppm. Similar observations were also made for the hydrolysate of 5-$^{13}$C glucose. This suggests that the atoms in the degradation products corresponding to glucose C$_2$ and glucose C$_5$ are carbonyl (COOH) or formyl (C=O) functional groups.
Identity of degradation products: J-coupling constants

As the next step in identifying the hexose degradation product(s), we measured the J-coupling constants ($J_{CC}$s) between their carbon atoms. Coupling of a $^{13}$C atom with another $^{13}$C atom causes peak splitting, which enables measurement of such coupling (Harris, 1983). For instance, to measure the J-coupling constants of the atom corresponding to peak B, we examined its expanded views in $[^{13}$C, $^1$H] NMR spectra of hydrolysates of specifically labeled glucose. We observed that peak B displayed peak splitting in spectra of 2-13C and 4-13C glucose hydrolysates (Figure 6.3), whereas it displayed no peak splitting in the spectra of hydrolysates of glucose labeled at other positions. This showed that the atom corresponding to peak B is coupled to two other atoms, and enabled the measurement of its J-coupling constant, $J_{CC}$, with the coupled atoms. The value of the $J_{CC}$ between the atoms corresponding to peaks B and D was 37.5 Hz. Such a large value indicated that this is a one-bond coupling (Krivdin and Kalabin, 1989). Similarly, the atom corresponding to peak B also exhibited a one-bond coupling (54.6 Hz) with the atom derived from glucose $C_2$.

The atom corresponding to peak D was coupled to two other atoms by one bond couplings (> 35 Hz), indicating that its functional group cannot be COOH. Therefore, peak D represents a CO group. Also, peak C displayed a one-bond coupling (37.5 Hz) with only one atom – the atom derived from glucose $C_5$.

Hexose degradation products are LVA and HMF

We also observed that the peak A (CH$_2$OH) was less intense than the other peaks in a $^{13}$C or $^1$H spectrum. This suggests its concentration was not the same as that of the other peaks, and that two hexose degradation products (major and minor) were present in the hydrolysate.

From the above observations, we hypothesized that the hexose degradation products were COOH–CH$_2$–CH$_2$–CO–CH$_3$ (levulinic acid) and CH$_2$OH–? (a yet unknown compound).
We acquired 2-D $^{13}$C, $^1$H and 1-D $^{13}$C NMR spectra of commercial levulinic acid, and observed that it contained peaks with chemical shifts identical to those observed in the glucose hydrolysate (data not shown). This identified the major hexose degradation product as levulinic acid.

Levulinic acid is known to be produced in a sequence of reactions starting from glucose (Weenen and Tjan, 1999; Figure 6.4). 5-hydroxymethyl furfural (HMF), the penultimate compound in this sequence, contains a CH$_2$OH group whose chemical shifts and J$^{CC}$ (Kelly et al., 2003) are identical to those measured for peak A. The other carbon atoms of HMF are either aldehydic (CHO) or aromatic carbons. We acquired a $^{13}$C, $^1$H NMR spectrum of the soybean protein hydrolysate and glucose hydrolysates spanning the aromatic region of chemical shifts. Cross peaks located at the reported chemical shifts of HMF (Kelly et al., 2003) were observed in these spectra, thus confirming the presence of HMF as the minor hexose degradation product.

From the above results, it is reasonable to deduce that the hexose hydrolysis products are formed from the glycosylating sugars glucosamine and mannose, associated with soybean seed protein. The metabolic source of these sugars is the hexose nucleotide (UDP-glucose or UDP-mannose) pool in the cytosol (Faik et al., 2000; Baldwin et al., 2001; Coates et al., 1980). Fractionally $^{13}$C-labeled hexose sugars can contain significant information on primary metabolism, especially on the glycolysis and pentose phosphate pathways (Roscher et al., 2000). We explored some characteristics of 2-D $^{13}$C, $^1$H spectra of LVA, and explored if the degradation products LVA and HMF contain useable metabolic information.

**Carbon atoms C$_3$ and C$_5$ of LVA acid exhibit long-range coupling**

The J-coupling constants of all LVA atoms are listed in Table 6.1. To the extent of our knowledge, this information has not been reported before. An interesting observation was the long range coupling (13.6 Hz) between the carbon atoms C$_3$ and C$_5$ of LVA. This suggests
that these atoms are spatially proximate in the LVA molecule. A consequence of this long range coupling is the intricate fine structure observed in the peaks corresponding to these atoms. For example, an expanded view of the $[^{13}\text{C}, ^{1}\text{H}]$ spectrum of LVA atom C3 (generated from U-$^{13}\text{C}$ glucose) shows a triple doublet peak. This peak arises because the triplet peak (that corresponds to the fully $^{13}\text{C}$-labeled C2-C3-C4 fragment) is split due to long range coupling with the $^{13}\text{C}$-labeled C5 atom. Such multiplet structures can discriminate isotopomer fragments up to four carbon atoms long, such as the C2-C3-C4-C5 fragment here. This is not possible in a spectra of hexoses, which lack long-range couplings. Therefore $[^{13}\text{C}, ^{1}\text{H}]$ spectra of LVA atoms can provide additional isotopomer information, about compared to the spectra of their parent hexoses.

**J-scaled $[^{13}\text{C}, ^{1}\text{H}]$ spectra of LVA**

We investigated the effect of J-scaling on the $[^{13}\text{C}, ^{1}\text{H}]$ of LVA. J-scaling usually increases the multiplet separation by an even integral factor, J, and eliminates spectral overlap (Willker *et al.*, 1997). Figure 6.6 shows the effect of J-scaling on the C2 peak of LVA. J-scaling resulted in two sets of multiplet peaks that shared a common singlet. In one set (marked with subscript $JS$) the C1-C2 coupling was amplified by the J-scaling factor. Whereas, in the other set (unsubscripted), the C1-C2 coupling remained unamplified. In both cases, the C2-C3 coupling was amplified by the J-scaling factor. Both sets of multiplets shared a common singlet. The ratios of corresponding multiplets in the two sets ($d/d_{JS}$, $dd/dd_{JS}$) were nearly identical.

The exact reason for this observation is unknown, although it has been speculated that J-scaling may result extra resonances in strongly coupled systems (Brown, 1984). This observation was observed for the atom C3 as well, and is reproducible. In the case of intricate multiplet structures such as those of atom C3 (which, in fractionally $^{13}\text{C}$-labeled cases,
contained up to 19 multiplets) we were able to use the J-scaling technique to successfully deconvolute the spectra.

**LVA and HMF reflect the isotopomeric composition of the parent hexose**

If metabolic information from LVA and HMF is to be used to analyze pathways and fluxes, it is necessary that LVA and HMF reflect the isotopomeric compositions of the parent hexoses that they are generated from. We verified that this is the case, by hydrolyzing a mixture of glucose isotopes (5 mol% U-$^{13}$C glucose, 95 mol% natural abundance glucose) under the conditions described. The isotopomeric compositions of LVA and HMF in the hydrolysate were quantified from 2-D [$^{13}$C, $^1$H] spectra, as described in Materials and Methods. The isotopomer abundances of LVA and HMF were found to be equal to the abundances of the corresponding glucose isotopomers (Figure 6.7). This proves that the degradation reactions (Figure 6.4) are irreversible and do not cause back-mixing of label, which can potentially alter the isotopomeric compositions.

**Use of LVA and HMF to differentiate cytosolic and plastidic hexoses**

We compared the isotopomeric compositions of LVA and HMF generated by hydrolysis of soybean protein and soybean starch respectively, and found that they were significantly different. Figure 6.8 shows a comparison of the isotopomer abundances of atom C$_3$ of LVA obtained by hydrolyzing soybean protein and starch, which were sequentially isolated from soybean embryos cultured on sucrose (10% w/w U-$^{13}$C) and glutamine. LVA in the protein hydrolysate spectrum originates from hexoses associated with the protein which are cytosolic in origin. Whereas, LVA in the starch hydrolysate spectrum originates from the glucose monomers that constitute starch, which are plastidic in origin. The difference in their multiplet intensities (and hence isotopomer abundances) suggests different fluxes in the two compartments. The glycolysis and pentose phosphate pathway fluxes have the most
significant influence on the isotopomeric compositions of hexoses. Therefore, this result implies different fluxes of these pathways in the cytosol and the plastid.

Summary

In this work, we have identified unknown compounds in soybean protein hydrolysate as levulinic acid and 5-hydroxymethyl furfural. 2-D [$^{13}$C, $^1$H] NMR studies were conducted on levulinic acid. These included the measurement of the hitherto unreported J-coupling constants ($J_{CC}$) of this compound. The NMR studies revealed a long range coupling between carbon atoms C$_3$ and C$_5$ of LVA, which resulted in intricate multiplet structures such as a triple doublet for atom C$_4$ arising from a fully $^{13}$C-labeled C$_2$-C$_3$-C$_4$-C$_5$ fragment. We also observed that J-scaled [$^{13}$C, $^1$H] spectra of LVA exhibit two distinct sets of multiplets for atoms C$_2$ and C$_3$ – one with J-scaling and one without. Furthermore, we investigated the possibility of using the metabolic information from LVA toward metabolic flux analysis, since using the metabolic information from LVA in a protein hydrolysate obviates the necessity of separating the glycosylating sugars from the protein. We confirmed that LVA preserved the isotopomeric composition of its parent hexose. It can also provide additional isotopomeric information than the parent hexose, because of its long range couplings. This strategy was used to differentiate between cytosolic and plastidic glycolysis and pentose phosphate pathways in soybean.

Acknowledgment

The authors wish to acknowledge Dr. Mark E. Westgate (Department of Agronomy, Iowa State University) for soybean embryo culture and protein extraction, and for useful suggestions on this work.
Abbreviations used

HMF: 5-hydroxymethyl furfural

HSQC: heteronuclear single-quantum correlation (spectroscopy)

LVA: levulinic acid

Literature cited

Bodenhausen, G., Ruben, D. J. Natural abundance nitrogen-15 NMR by enhanced

Braunschweiler, L., Ernst, R. R. Correlation transfer by isotropic mixing: application to

Brown, L. R. Differential scaling along $\omega_1$ in COSY experiments. J. Magn. Reson. 1984, 57,
513-518.

seed storage proteins of Glycine max and Phaseolus vulgaris. J Biol. Chem. 1986, 261,
9228-9238.

Eller, F. J. Interference by methyl levulinate in determination of total fat in low-fat, high-
sugar products by as chromatographic fatty acid ester (GC-FAME) analysis. J AOAC Int.
1999, 82, 766-769.

Johnson, B. A., Blevins, R. A. NMRView: a computer program for the visualization and


Weenen, H., Tjan, S. B. 3-Deoxyglucosone as flavour precursor. In Trends in Flavour Research; Maarse, H. M., and van der Heij, D. G., Eds; Elsevier Science; 1999, 327-337.


**Financial sources**

This paper is based upon work funded by the National Science Foundation (BES-0224600); and the Plant Sciences Institute, Iowa State University.

**Table caption**

**Table 6.1.** NMR constants (chemical shifts, $\sigma$ and J-coupling constants, $J_{cc}$) for levulinic acid atoms. $J_{cc}$-s were measured by examining $^{13}$C spectra of hydrolysates of glucose labeled at specific positions (see Results and Figure 6.2). $J_{cc}$-s are reported as 0.0 Hz when they were below detectable limits (1 Hz). The last column indicates the carbon atom glucose corresponding to the levulinic acid carbon atom on its row.

**Figure Captions**

**Figure 6.1.** (A) Aliphatic region of 2-D [$^{13}$C, $^1$H] HSQC spectrum of soybean protein hydrolysate; protein was isolated from soybean embryos cultured on sucrose (10% w/w U-$^{13}$C) and glutamine. Most cross peaks were assigned to the carbon atoms of 16 proteinogenic amino acids, which were detected in the protein hydrolysate by HPLC. Four intense cross peaks (marked A, B, C, and D) could not be assigned to any amino acid atom. Assignment of cross peaks to amino acid atoms was done from literature data on their $^{13}$C and $^1$H chemical shifts and J-coupling constants; and was confirmed by performing 2-D [$^1$H, $^1$H] and 3-D [$^{13}$C, $^1$H, $^1$H] TOCSY experiments on the protein hydrolysate. The names of a few amino acid
nuclei are omitted for clarity. (B) The same region of 2-D [$^{13}$C, $^1$H] HSQC spectrum of U-$^{13}$C glucose hydrolysate; U-$^{13}$C glucose was hydrolyzed under the same conditions employed for protein. Only four cross-peaks were observed, and these were at the same positions as the unassigned cross peaks in the protein hydrolysate.

**Figure 6.2.** Aliphatic region of 2-D [$^{13}$C, $^1$H] HSQC spectrum of 3-$^{13}$C glucose hydrolysate; 3-$^{13}$C glucose was hydrolyzed under the same conditions employed for protein. The cross peak marked B was the most intense peak detected on this spectrum. Three other cross peaks were also observed on this spectrum (those corresponding to cross peaks A, C, and D in Figure 6.1), but they had intensities at least 20-fold lesser than cross peak B, and are barely visible at this resolution.

**Figure 6.3.** Expanded views of cross peak B (carbon atom C$_2$ of levulinic acid, corresponding to carbon atom C$_3$ of glucose) in 2-D [$^{13}$C, $^1$H] HSQC spectrum of (A) 2-$^{13}$C glucose hydrolysate, and (B) 4-$^{13}$C glucose hydrolysate. 1-D slices along the $^{13}$C dimension are shown beside the cross peak. In both cases, the cross peak B exhibits peak splitting, due to coupling with another $^{13}$C atom, thus enabling the measurement of its J-coupling constant, J$_{CC}$, with the coupled atom. The measured J$_{CC}$ values were (A) 54.6 Hz for the C$_2$-C$_3$ coupling, (B) 37.5 Hz for the C$_3$-C$_4$ coupling. All J$_{CC}$-s measured thus are reported in Table 6.1.

**Figure 6.4.** Mechanism for the formation of 5-hydroxymethyl furfural, levulinic acid, and formic acid from glucose.

**Figure 6.5.** Expanded view of atom C$_3$ of levulinic acid (cross peak D in Figure 6.1) in 2-D [$^{13}$C, $^1$H] HSQC spectrum of U-$^{13}$C glucose hydrolysate. The peak exhibits a double triplet, arising from three couplings: with atoms C$_2$ (37.5 Hz), C$_4$ (37.5 Hz), C$_5$ (13.6 Hz).
Figure 6.6. (A) Expanded view of carbon atom C\textsubscript{2} of levulinic acid in 2-D \textsuperscript{13}C, \textsuperscript{1}H HSQC spectrum of soybean protein hydrolysate without J-scaling. The multiplets shown are: singlet (s), doublet (d), J-scaled double doublet (dd). (B) Expanded view of the same peak in 2-D \textsuperscript{13}C, \textsuperscript{1}H HSQC spectrum of the same soybean protein hydrolysate; the HSQC was acquired with J-scaling (by a factor 6) along the \textsuperscript{13}C dimension. J-scaling usually amplifies J-coupling constants of coupled peaks by the J-scaling factor (6, in this case). Here, two sets of coupled multiplet peaks were observed. In one set (marked with subscript \textsubscript{iS}) the C\textsubscript{1}-C\textsubscript{2} coupling was amplified by the J-scaling factor. Whereas, in the other set (unsubscripted), the C\textsubscript{1}-C\textsubscript{2} coupling remained unamplified. In both cases, the C\textsubscript{2}-C\textsubscript{3} coupling was amplified by the J-scaling factor. Both sets of multiplets shared a common singlet. The ratios of corresponding multiplets in the two sets (d/d\textsubscript{iS}, dd/dd\textsubscript{iS}) were nearly identical.

Figure 6.7. Comparison of expected and measured multiplet intensities of selected atoms of levulinic acid (LVA) and 5-hydroxymethyl furfural (HMF), prepared by hydrolyzing an isotopic mixture of glucose (5 mol\% U-\textsuperscript{13}C, 95 mol\% natural abundance). Isotopomer abundances were measured as described in Materials and Methods. Expected multiplet intensities were calculated from the known isotopomeric composition of the hydrolyzed glucose, and assuming that isotopomeric composition is not altered by hydrolysis. The multiplets shown are: singlet (s), doublet (d), double doublet (dd), and triple doublet (td). All multiplet intensities are expressed as a fraction of the total signal of their cross peak. This result establishes that the isotopomeric composition of glucose can be back-calculated from that of LVA and HMF.

Figure 6.8. Multiplet intensities of carbon atom C\textsubscript{2} of levulinic acid (LVA) (corresponding to carbon atom C\textsubscript{3} of glucose), evaluated from 2-D \textsuperscript{13}C, \textsuperscript{1}H HSQC spectra of (A) soybean protein hydrolysate, and (B) soybean starch hydrolysate. All multiplet intensities are expressed as a fraction of the total signal of their cross peak. Protein and starch were
sequentially isolated from soybean embryos cultured on sucrose (10% w/w U-^{13}C) and glutamine. LVA in the protein hydrolysate spectrum originates from hexoses associated with the protein which are cytosolic in origin. Whereas, LVA in the starch hydrolysate spectrum originates from the glucose monomers that constitute starch, which are plastidic in origin. The difference in their multiplet intensities (and hence isotopomer abundances) suggests different glycolysis and pentose phosphate pathway fluxes in the cytosol and the plastid (see Results).
Table 6.1. NMR constants (chemical shifts, $\sigma$ and J-coupling constants, $J_{CC}$) for levulinic acid atoms.

<table>
<thead>
<tr>
<th>Atom (n)</th>
<th>Glucose atom</th>
<th>$\sigma_{HH}$ (ppm)</th>
<th>$\sigma_{13C}$ (ppm)</th>
<th>$C_1$-$C_n$</th>
<th>$C_2$-$C_n$</th>
<th>$C_3$-$C_n$</th>
<th>$C_4$-$C_n$</th>
<th>$C_5$-$C_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>$xx$</td>
<td>180.0</td>
<td>$xx$</td>
<td>54.6</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2.57</td>
<td>30.5</td>
<td>54.6</td>
<td>$xx$</td>
<td>37.5</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>2.86</td>
<td>40.5</td>
<td>0.0</td>
<td>37.5</td>
<td>$xx$</td>
<td>37.5</td>
<td>13.6</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>$xx$</td>
<td>180.0</td>
<td>0.0</td>
<td>0.0</td>
<td>37.5</td>
<td>$xx$</td>
<td>37.5</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>2.18</td>
<td>31.8</td>
<td>0.0</td>
<td>0.0</td>
<td>13.6</td>
<td>37.5</td>
<td>$xx$</td>
</tr>
</tbody>
</table>
Figure 6.1(A). Aliphatic region of 2-D $^{13}$C, $^1$H HSQC spectrum of soybean protein hydrolysate.
Figure 6.1(B). Aliphatic region of 2-D \([^{13}C, ^1H]\) HSQC spectrum of U-\(^{13}C\) glucose hydrolysate.
Figure 6.2. Aliphatic region of 2-D [$^{13}$C, $^1$H] HSQC spectrum of 3-$^{13}$C glucose hydrolysate; 3-$^{13}$C glucose was hydrolyzed under the same conditions employed for protein.
Figure 6.3(A). Expanded views of cross peak B (carbon atom C₂ of levulinic acid, corresponding to carbon atom C₃ of glucose) in 2-D $[^{13}\text{C}, ^{1}\text{H}]$ HSQC spectrum of 2-$^{13}$C glucose hydrolysate.
Figure 6.3(B). Expanded views of cross peak B (carbon atom C₂ of levulinic acid, corresponding to carbon atom C₃ of glucose) in 2-D $^{13}\text{C}, ^1\text{H}$ HSQC spectrum of 4-$^{13}\text{C}$ glucose hydrolysate.
**Figure 6.4.** Mechanism for the formation of 5-hydroxymethyl furfural, levulinic acid, and formic acid from glucose.
Figure 6.5. Expanded view of atom C$_3$ of levulinic acid (cross peak D in Figure 6.1) in 2-D [\textsuperscript{13}C, \textsuperscript{1}H] HSQC spectrum of U-\textsuperscript{13}C glucose hydrolysate.
**Figure 6.6.** (A) Expanded view of carbon atom C₂ of levulinic acid in 2-D \(^{13}\text{C}, ^1\text{H}\) HSQC spectrum of soybean protein hydrolysate without J-scaling.
Figure 6.6. (B) Expanded view of carbon atom C₂ of levulinic acid in 2-D $[^{13}\text{C}, ^{1}\text{H}]$ HSQC spectrum of soybean protein hydrolysate with J-scaling.
Figure 6.7. Comparison of expected and measured multiplet intensities of selected atoms of levulinic acid (LVA) and 5-hydroxymethyl furfural (HMF), prepared by hydrolyzing an isotopic mixture of glucose (5 mol% U-13C, 95 mol% natural abundance).
Figure 6.8. Multiplet intensities of carbon atom C₂ of levulinic acid (LVA) (corresponding to carbon atom C₃ of glucose), evaluated from 2-D $[^{13}\text{C}, ^{1}\text{H}]$ HSQC spectra.
7. Determination of biomass composition of *Catharanthus roseus* hairy roots for metabolic flux analysis

A paper to be submitted to *Biotechnology Progress*.

Ganesh Sriram§$, Omar González-Rivera§, and Jacqueline V. Shanks*

Department of Chemical Engineering, Iowa State University, Ames, IA 50011

Abstract

Metabolic flux analysis plays a key role in metabolic engineering, and information about biomass composition of the organism is indispensable to flux analysis. We report the elemental make-up and composition of the biomass of *Catharanthus roseus* hairy roots, which is a pharmaceutically significant plant system and an important metabolic engineering target. The observed elemental composition was: carbon (41.9%), hydrogen (6.6%), nitrogen (1.6%), oxygen (33.0%); expressible by the formula C\textsubscript{12.0}H\textsubscript{22.7}N\textsubscript{0.4}O\textsubscript{7.1}. In addition, the hairy roots contained ash (14.6% ± 0.4%) and moisture (2.2% ± 1.2%). The observed proportions of water-soluble biomass constituents were: water-soluble protein (3.2% ± 0.2%), starch (2.6% ± 0.5%), soluble sugars (38.0% ± 1.0%). Also, individual proteinogenic amino acids were quantified. The observed percentages of water-insoluble biomass constituents were: lipids (6.05% ± 2.45%), cellulose (1.5% ± 0.5%), hemicellulose (5.6% ± 1.5%), and lignin.

---

§ Primary researchers, these authors have contributed equally to this work.
§ Primary author.
* Principal investigator and corresponding author.
(15.4% ± 2.5%). We were able to account for 89.2% (± 9.7%) of the biomass, which enabled the precise determination of 12 external fluxes in a metabolic flux map.

**Introduction**

Metabolic engineering is the science of rational modification of organisms to confer improved, beneficial properties to them, and the development of analysis tools to measure the altered properties (Bailey, 1991). Metabolic flux analysis (MFA), the quantification of intracellular metabolic fluxes, is a central diagnostic tool in metabolic engineering (Stephanopoulos *et al.*, 1998; Stephanopoulos, 1999). Knowledge of biomass composition of the organism is crucial to MFA, as the evaluated fluxes may be sensitive to the biomass composition (Pramanik and Keasling, 1998). Furthermore, fluxes contributing to biomass are usually not identifiable by $^{13}$C label-based flux analyses, as they do not substantially affect the distribution of $^{13}$C labels. Therefore they have to be calculated from the biomass composition. Consequently, a quantification of biomass composition of an organism is imperative before MFA is performed.

In this work, we report the biomass composition of *Catharanthus roseus* hairy roots, an important metabolic engineering target and a promising system for the application of flux analysis. *C. roseus* is a source of high-value indole alkaloids, some of which possess therapeutic properties, [e.g. vindoline and vinblastine (anti-neoplastic), and serpentine and ajmalicine (anti-hypertension)] (Bhadra and Shanks, 1997). Hairy root cultures of the plant, generated through *Agrobacterium* infection, are biochemically and genetically stable model systems (Shanks and Bhadra, 1997). Because of the low quantities of alkaloids produced by wild-type plants or hairy roots, the alkaloid production pathway in *C. roseus* is an important metabolic engineering target. Additionally, since *C. roseus* is an extensively studied plant system, metabolic engineering of this plant around the tryptophan production pathway may
help develop model plants with higher levels of tryptophan, an essential amino acid. Therefore, metabolic engineering of *C. roseus* hairy roots is significant, both for pharmaceutical and agricultural reasons, and metabolic flux analysis is expected to play a crucial role in this process. The determination of the biomass composition of *C. roseus* hairy roots is therefore necessary.

Although the proportions of polymers constituting biomass in various organs of intact plants has been listed (Poorter and Villar, 1997), very little information related to biomass composition of this system or of other plant hairy root systems, is available in the literature. Kwok and Doran (1995) reported the elemental composition of certain hairy roots (other than *C. roseus*). However, no information on the biomass composition (elemental constituents or proportions of individual biomass composition) is available in the literature, to the extent of our knowledge. The objective of this study, was therefore, to quantify these in *C. roseus* hairy roots.

**Materials and methods**

**Hairy root culture.** *C. roseus* LBE-6-1 hairy roots were cultured at 26°C in 50 mL liquid medium containing 30 gL⁻¹ sucrose and 1.65 gL⁻¹ Gamborg B-5 salts (Sigma), as per procedures outlined in Bhadra and Shanks (1997). Hairy roots were harvested after a growth period of 21 or 28 days. They were weighed, freeze-dried for 72 h and weighed again. The hairy roots were then ground for stored for extraction and analysis of biomass components.

**Elemental composition.** Elemental composition of the hairy roots was determined by combustion in a CHN analyzer (LECO). The percentages of carbon, hydrogen and nitrogen were directly determined by the analyzer through off-gas analysis. The percentage of oxygen was determined by subtracting ash and moisture contents.
Ash and moisture contents. Ash content was determined by combustion of freeze-dried hairy roots in a muffle furnace, on a pre-weighed evaporating dish. Freeze-dried hairy roots contain some moisture, and this was determined gravimetrically (Miller, 1998).

Lipid extraction. Lipids were extracted by contacting freeze-dried, ground hairy roots with hexane in the ratio 18.4 mg roots: 1 mL hexane. The extractions were carried out at 45°C for 1 h each and were performed five times. The supernatant contained the extracted lipids, which were quantified by weighing.

Protein and soluble sugar extraction. Water-soluble protein was extracted by washing freeze-dried and ground hairy roots in phosphate buffer (pH 7.2), at 8°C for 15 min. The washes were repeated four times, and the consolidated supernatant was assayed for protein. Soluble sugars were similarly extracted from a separate aliquot of the hairy root sample, by contacting five times with distilled, de-ionized water. The extract was freeze-dried for 36 h and reconstituted in D₂O for NMR analysis.

Quantification of protein and starch. Protein was assayed using the Bradford test (Bio-Rad, Hercules, CA). Starch was assayed iodometrically, with spectroscopic quantification of the starch-iodine complex at 565 nm.

Protein hydrolysis, amino acid quantification. Protein hydrolysis was performed in hydrolysis tubes (Pierce Endogen), which are pressure resistant test tubes provided with an arm that enables evacuation. For hydrolysis, 6N hydrochloric acid was added to protein in the ratio 1 mL 6N HCl: 400 µg protein. The hydrolysis tube was then evacuated of air, flushed with nitrogen (to remove any residual oxygen) and evacuated again. The effect of nitrogen flushing to evacuate residual oxygen is examined in Appendix I. The tube was then placed in a hot block at 150°C for 3 h. Optimization of hydrolysis time is discussed in Appendix I. The acid in the hydrolysate was evaporated in a Rapidvap evaporator (Labconco, Kansas City, MO). The residue was reconstituted in 2 mL deionized water and freeze-dried for 72 h. This
was then reconstituted in deionized water and analyzed for amino acids. Amino acids were quantified by HPLC, after derivatization with phenylisothiocyanate to produce phenylthiocarbamyl amino acid derivatives, which were eluted by a reverse phase C18 silica column, with detection at 254 nm.

**NMR spectroscopy of soluble sugars.** 1-D NMR spectra of soluble sugars were collected on a Bruker Avance DRX 500 MHz spectrometer at 298 K. The reference to zero ppm was set using the methyl signal of dimethylsilapentanesulfonate (Sigma) as an internal ¹H standard. A known amount of 2,3-¹³C labeled alanine (Sigma) was added as an internal ¹³C standard. The resonance frequency of ¹³C was 125.7 MHz. The spectral width was 5028.05 Hz along the ¹³C dimension. The number of data points was 2048 (¹³C), and the number of scans was 4096. Assignment of peaks was performed using 2-D [¹H, ¹H] correlation spectra (COSY) and 2-D [¹³C, ¹H] HSQC spectra (Bodenhausen and Ruben, 1980, Braunschweiler and Ernst, 1983). The software Xwinnmr™ (Bruker) was used to acquire and quantify all spectra.

**Quantification of lignin, cellulose, hemicellulose.** Lignin was assayed after extraction by sulfuric acid (Ankom, 2004). Cellulose, hemicellulose and lignin were quantified together as fiber (neutral detergent and acid detergent). Hemicellulose was quantified as the difference between neutral detergent and acid detergent fiber, and cellulose was quantified as the difference between acid detergent fiber and lignin.

**Results and discussion**

**Elemental composition**

We determined the elemental composition of the hairy roots to be: carbon (41.9%), hydrogen (6.6%), nitrogen (1.6%), oxygen (33.0%). This is also expressible by the formula C₁₂₀H₂₂₇N₀·₄O₇·₇. Also, the freeze-dried hairy roots contained moisture some moisture (2.2% ± 1.2%) and substantial ash (14.6% ± 0.4%). The dry weight per flask of 28 d-old hairy roots,
was 493.5 mg ± 6.5 mg, and the amount of ash in this amount of roots (72.34 mg ± 1.685 mg) was close to the amount of minerals added to the liquid media (77.5 mg), indicating that most minerals (93%) were consumed by 28 d.

**Proportions of protein, proteinogenic amino acids, lipids, starch, fiber**

The observed proportion of protein in the biomass was 3.2% (± 0.2%). This number represents only water-soluble proteins. Water-insoluble proteins (associated with the cell-wall or membranes) are expected to constitute a smaller percentage of the biomass. We also quantified the proportions of individual amino acids in the protein. Figure 7.1 depicts these proportions in 21 d- and 28 d-old hairy roots. Interestingly, the composition of proteinogenic amino acids does not significantly vary during this period.

We determined the proportion of hexane-extractable lipids to be 6.05% (± 2.45%), in 28 d-old hairy roots. Starch was detected in the hairy roots iodometrically, and its proportion in 28 d-old hairy roots was (2.6% ± 0.5%).

The observed percentages of fiber constituents were in the biomass were: cellulose (1.5% ± 0.5%), hemicellulose (5.6% ± 1.5%), and lignin (15.4% ± 2.5%). This proportion of lignin is rather high. Lignin production is one of the responses of *Agrobacterium* interaction (which generates the hairy roots) and the function of lignin overproduction is believed to be cell wall reinforcement (Rijhwani and Shanks, 1996).

**Proportions of soluble sugars**

The aforementioned constituents accounted for approximately half the dry weight of the hairy roots. We found that most of the remainder was soluble sugars or small polymers of sugars. To determine the remainder of the biomass constituents, we analyzed a water extract of the hairy roots by NMR spectroscopy. Hairy roots were grown on sucrose (5% w/w U-13C, 95% w/w natural abundance), and water-soluble components were extracted in distilled, de-
ionized water. Figure 7.2 shows a 1-D $^{13}$C NMR spectrum of the water extract. The peaks marked ‘Ala $\alpha$’ and ‘Ala $\beta$’ are from 2,3-$^{13}$C alanine, which was added to the extract as an internal standard. All other peaks were quantified relative to their intensity.

The peaks in the range 60 to 100 ppm accounted for 94.9% of the total carbon in the extract. Their chemical shifts indicated that they are sugars (e.g. glucose, fructose). To further investigate their identity, a 2-D [$^{13}$C, $^1$H] HSQC (heteronuclear single quantum correlation spectroscopy) spectrum and a 2-D [$^1$H, $^1$H] COSY (correlation spectroscopy) spectrum of the extract were acquired (not shown). Both spectra revealed that these peaks had $^1$H chemical shifts in the range 3 to 5 ppm, which is characteristic of small sugars. Also, both spectra largely resembled those of glucose. Therefore, most of the carbon in the water extract was soluble sugars (probably glucose, glucans or small polymers of glucose).

The amount of carbon corresponding to the peaks identified as soluble sugars, was 15.2 mol% of the original biomass (from which the water extract was obtained). Using a carbon proportion of 40% (valid for many small sugars including the likely candidate glucose) we estimated that the soluble sugars constituted 38.04% (± 1.0%) of the biomass.

The quantified proportions of the biomass constituents are listed in Table 7.1. We were able to account for 89.2% of the biomass (lower limit: 78.9%, upper limit: 99.4%). We suspect the small remainder of biomass to be chiefly free amino acids, alkaloids, and other secondary metabolites.

Interestingly, we did not find any large pools of free amino acids. Some species of hairy roots are reported to contain large amounts of opines (compounds consisting of sugars and amino acids, e.g. agropine = glucose + glutamine). Opines are thought of as a result of Agrobacterium infection. Any opines in our system were below detectable levels, as the amino acids in them were not detected by NMR.
Uses of determining biomass composition

Using the estimated composition of the *C. roseus* hairy root biomass, we were able to precisely calculate 12 biomass fluxes in a metabolic map of this system. These fluxes were the fluxes of the following precursors towards biomass: cytosolic and plastidic hexose phosphates, cytosolic and plastidic pentose phosphates, plastidic erythrose-4-phosphate, plastidic phosphoenolpyruvate, phosphate cytosolic and plastidic pyruvate, cytosolic triose phosphate, plastidic acetyl CoA, α-ketoglutarate, and oxaloacetate. These fluxes contributing to biomass are usually not identifiable by $^{13}$C label-based flux analyses, as they do not significantly affect the distribution of a label. Therefore, the quantification of biomass constituents is a necessary prerequisite for $^{13}$C label-based flux analysis.

Another important function of the biomass composition is that it indicates the fate of the carbon taken up by the hairy roots. The fate of the acquired carbon helps in the calculation of the construction cost (Poorter and Villar, 1997) i.e. optimal amounts of substrate needed when the roots are produced on a large scale, e.g. in a bioreactor.

Finally, the fate of the carbon, as indicated by the biomass composition indicates possible metabolic engineering targets. For example, most of the soluble sugars that we detected in our system appear to be deposited for storage, and there exists potential for metabolic engineering of this system to divert some of this carbon toward more valuable products (such as alkaloids) which are produced by the roots in small quantities.

Appendix I. Protein hydrolysis optimization

In this study, we optimized two parameters related to protein hydrolysis: the hydrolysis time and the introduction of nitrogen flushing to completely expel oxygen from the hydrolysis tube. The optimization studies were performed with bovine serum albumin (BSA), chosen as a representative protein. Figure 7.A-1 shows a plot of residual protein concentration versus
time of hydrolysis for BSA. The hydrolysis is effectively complete within 20 min, and any hydrolysis time above this would assuredly hydrolyze all protein. Also, we observed that 2 h and 3 h hydrolysis experiments gave the same results (Figure 7.A-2A).

We found that it was necessary to completely evacuate oxygen from the hydrolysis tube by nitrogen flushing and evacuation. This was necessary to prevent the oxidation of the amino acids methionine and tyrosine during hydrolysis. Figure 7.A-2A depicts the results of a hydrolysis that was performed without oxygen evacuation. Methionine and tyrosine are conspicuously absent. Figure 7.A-2B shows the results of a hydrolysis performed after flushing nitrogen into the hydrolysis tube and evacuating before hydrolysis. It is clear that methionine and tyrosine are recovered in this case.

Acknowledgments

The authors acknowledge D. Bruce Fulton (Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University) for assistance with NMR analysis; Louisa B. Tabatabai (Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University) for consultation on protein hydrolysis; and David Gross (Woodson Tenent Laboratories, Des Moines, IA) for analysis of fiber constituents. This material is based upon work supported by the National Science Foundation under grants BES 9906978 and BES 0224600, and the Plant Sciences Institute at Iowa State University.

References


<table>
<thead>
<tr>
<th>Component</th>
<th>Proportion SD (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(% DW)</td>
</tr>
<tr>
<td>Soluble sugars</td>
<td>38.04</td>
</tr>
<tr>
<td>Lignin</td>
<td>15.40</td>
</tr>
<tr>
<td>Ash</td>
<td>14.65</td>
</tr>
<tr>
<td>Lipids</td>
<td>6.05</td>
</tr>
<tr>
<td>Hemicellulose</td>
<td>5.60</td>
</tr>
<tr>
<td>Protein</td>
<td>3.15</td>
</tr>
<tr>
<td>Iodometric starch</td>
<td>2.58</td>
</tr>
<tr>
<td>Unbound moisture</td>
<td>2.20</td>
</tr>
<tr>
<td>Cellulose</td>
<td>1.50</td>
</tr>
</tbody>
</table>

Table 7.1. Proportions of biomass constituents in *C. roseus* hairy roots.
Figure 7.1. Proteinogenic amino acid proportions in C. roseus hairy roots. Hollow bars indicate 21-day-old roots, and filled bars indicate 28-day-old roots.
Figure 7.2. 1-D $^{13}$C NMR spectrum of water extract of 28 d-old $C.\ roseus$ hairy roots grown on U-$^{13}$C sucrose.

2,3-$^{13}$C alanine was used as an internal standard. The intensities of the Ala $\alpha$ and $\beta$ peaks (which were identical) were used as a unit to determine the intensities of all other peaks.
Figure 7.A-1. Protein hydrolysis kinetics for bovine serum albumin.
Figure 7.A-2. Amino acid analysis of BSA hydrolysate, obtained by the outlined hydrolysis method.

(A) (upper figure) Without nitrogen flushing and oxygen evacuation, (B) (lower figure) with nitrogen flushing and oxygen evacuation. Hashed bars indicate the published amino acid composition of BSA (theoretical), the other two types of bars indicate the results obtained in this work (experimental): filled bars denoting a 2-h hydrolysis experiment, and hollow bars denoting a 3-h hydrolysis experiment. with nitrogen flushing and oxygen evacuation.
8. Conclusions and future directions

This work reported theoretical and experimental advances with respect to $^{13}$C metabolic flux analysis of plant systems.

Theoretical improvements

The theoretical improvements bondomer, bondomer balancing, and Boolean function mapping were introduced in the analysis of $^{13}$C labeling data. The bondomer concept was introduced as an alternative to the isotopomer concept, for modeling a class of $^{13}$C labeling experiments in which the U-$^{13}$C and natural abundance forms of a single carbon substrate are employed. For this class of experiments, the bondomer concept is computationally efficient compared to the isotopomer concept. In particular, it was shown that because the natural abundance of $^{13}$C is automatically factored into the definition of a bondomer, a number of bondomers in real metabolic networks have zero abundance, and need not be included in the bondomer balances. In real networks such as the glycolysis/pentose phosphate pathway, this reduces the number bondomer balances by a factor of 2 to 3, thus increasing the computational efficiency. Also, the bondomer concept simplifies mathematical descriptions of the labeling state, making a priori structural identifiability easier. It may be possible to extend bondomer theory to two- or multiple-substrate scenarios.

The novel algorithm Boolean function mapping was introduced to facilitate the processing of carbon skeleton rearrangement information. This algorithm was shown to be more efficient compared to existing techniques, and its efficiency is expected to be valuable in the analysis of metabolic networks which are not completely known, as is common in plant metabolism.

Comprehensive $^{13}$C flux analysis of plant systems

$^{13}$C labeling experiments were performed on developing soybean embryos and Catharanthus roseus hairy roots. We obtained exhaustive labeling data from sink metabolites by 2-D NMR.
It was possible to quantify carbon partitioning through several metabolic processes including glycolysis, oxPPP, gluconeogenesis, anaplerotic pathways, TCA cycle, and the glyoxylate shunt. Furthermore, we were also able to distinguish between fluxes in different compartments, based on labeling data of sink metabolites synthesized in separate compartments. The fluxes of the compartmented pentose phosphate pathway evaluated in this work agreed with the emerging model of this pathway in higher plants, and with ‘omic’ (genomic and proteomic) data from other higher plants.

The experimental methodology employed in this work has the potential to become a high-throughput one. Further reduction of the sample size and duration of labeling period should be possible, and could be optimized. The computer program developed to calculate fluxes from the labeling data is generic. We expect both these features to increase the applicability of flux analysis in plants.

**Limitations and caveats**

In this work, the $^{13}$C labeling methodology has been demonstrated to be a powerful physiological diagnostic tool for plants. However, it has certain limitations. Currently, only plants growing in liquid media can be handled. Further, the attainment of isotopic and metabolic steady state are important requirements in this analysis. This limits the types of experimental plant systems to which such analysis can be applied. Also, if the experimental plants contain several organs or cell types, experiments need to be designed to decouple the effect of each cell type on the observed labeling patterns.

Logistically, the sensitivity of the measuring instrument (NMR or MS) and the cost of labeled substrates may also be factors that determine what types of systems may be explored using $^{13}$C flux analysis.

However, the $^{13}$C labeling methodology is useful to build and verify models and hypotheses.
Immediate future scope. Analysis of environmental and genetic variants

The analysis of environmental and genetic variants of plants is an important utility of $^{13}$C flux analysis. As demonstrated here, flux analysis can provide insights on physiology and function. Comparison of fluxes between genetic or environmental variants can provide valuable information about the effects of genetic or environmental manipulations on the physiology. This is particularly relevant in the context of the recent upsurge in plant metabolic engineering.

Currently, the flux analysis of both environmental and genetic variants of plants are in progress in our laboratory. One current research project focuses on evaluating fluxes in soybean embryos grown or cultured at five different temperatures. The temperatures have been found to have a significant effect on the biomass growth. This research project will examine whether or not the intracellular carbon flow is affected by temperature.

Another research project focuses on evaluating fluxes in genetic variants of *Catharanthus roseus* hairy roots generated by our collaborators. The genetic modifications are in secondary metabolism. They have noticeable effects on phenotype, and can be expected to have significant impact on central carbon metabolism. This research project will investigate the impact of the genetic modifications on the central carbon metabolism of *C. roseus* hairy roots.

Long-term future scope. Pathway identification and discovery

It may be possible to identify unknown parts of metabolic networks by using $^{13}$C labeling. Two examples were demonstrated in this work. A flux from T3P to F6P (attributed to fructose-1,6-bisphosphatase) was identified in the plastid in soybean embryos. It was necessary to insert this reaction into the metabolic network model to explain certain observations in the labeling data, particularly a substantial biosynthetic bond between carbon
atoms C\textsubscript{3} and C\textsubscript{4} of starch glucose. Secondly, the anomaly related to Leu \(^{\delta^1}\) possibly conceals an unknown pathway, and remains an open question.

A systematic \(^{13}\text{C}\) labeling-based pathway discovery tool could be envisaged, wherein metabolic pathways that fit the observed labeling data are intelligently guessed. Such a tool would utilize algorithms for optimization in discrete space, for metabolic pathway synthesis, and perhaps artificial intelligence.

**Long-term future scope. Plant systems biology**

The field of plant systems biology has become very popular recently (see the June 2003 issue of *Plant Physiology*, which was devoted to this subject.) past. Plant systems biology is the study of a plant at the level of a whole system, rather than the reductionist approach of studying at the level of individual genes or enzymes. Therefore, systems biology studies the connections and interactions between various components of a plant, and endeavors to create an *in silico* model of a plant. This requires large amounts of data that simultaneously describe various aspects of a plant. High-throughput data on fluxes, transcripts, proteins, and metabolites, can contribute immensely to this area. Together, these can provide the basis for understanding the functioning of plants from a systems perspective, and aid toward the long-term vision of constructing *in silico* models of plants.
Biographical sketch

Ganesh Sriram was born to Renuka and P. G. Sriram in Irinjalakuda, India, on 27 March 1976. He received his Bachelor of Technology and Master of Technology degrees in Chemical Engineering, in 1997 and 1999 respectively, from Indian Institute of Technology Bombay, Mumbai, India. He was awarded the Best Innovative Master’s Level Project award from the Indian National Academy of Engineering in 1999, the United Engineering Foundation Conference Fellowship for Metabolic Engineering III in 2000, and the Best Presentation Award at the Third International Congress for Metabolomics in 2004. He has served as a Graduate Research Assistant and Teaching Assistant at Indian Institute of Technology Bombay, and Iowa State University.