Molecular BioSystems

RSC Publishing

PAPER

View Article Online
View Journal | View Issue

Cite this: *Mol. BioSyst.*, 2013, **9**, 99–112

Designer labels for plant metabolism: statistical design of isotope labeling experiments for improved quantification of flux in complex plant metabolic networks†

Shilpa Nargund and Ganesh Sriram*

Metabolic fluxes are powerful indicators of cell physiology and can be estimated by isotope-assisted metabolic flux analysis (MFA). The complexity of the compartmented metabolic networks of plants has constrained the application of isotope-assisted MFA to them, principally because of poor identifiability of fluxes from the measured isotope labeling patterns. However, flux identifiability can be significantly improved by *a priori* design of isotope labeling experiments (ILEs). This computational design involves evaluating the effect of different isotope label and isotopomer measurement combinations on flux identifiability, and thereby identifying optimal labels and measurements toward evaluating the fluxes of interest with the highest confidence. This article reports ILE designs for two major, compartmented plant metabolic pathways – the pentose phosphate pathway (PPP) and γ-aminobutyric acid (GABA) shunt. Together, these pathways represent common motifs in plant metabolism including duplication of pathways in different subcellular compartments, reversible reactions and cyclic carbon flow. To compare various ILE designs, we employed statistical *A*- and *D*-optimality criteria. Our computations showed that 1,2-^{13}C Glc is a powerful and robust label for the plant PPPs, given currently popular isotopomer measurement techniques (single quadrupole mass spectrometry [MS] and 2-D nuclear magnetic resonance [NMR]). Further analysis revealed that this label can estimate several PPP fluxes better than the popular label 1-^{13}C Glc. Furthermore, the concurrent measurement of the isotopomers of hexose and pentose moieties synthesized exclusively in the cytosol or the plastid compartments (measurable through intracellular glucose or sucrose, starch, RNA ribose and histidine) considerably improves the identifiability of PPP fluxes in the individual compartments. Additionally, MS-derived isotopomer measurements outperform NMR-derived measurements in identifying PPP fluxes. The potency of 1,2-^{13}C Glc can be improved substantially by combining it with other labels (*e.g.* 3-^{13}C Glc, 1-^{13}C Glc and U-^{13}C Glc) in parallel ILEs. For the GABA shunt, we calculated that 100% 2-^{13}C Ala and 100% U-^{13}C Gln constitute the best labels. We anticipate that the ILE designs presented in this article can enhance the quality of flux estimates in these two complex plant pathways. In the future, these ILE designs can be further improved by leveraging recent analytical and computational developments in isotope-assisted MFA.

Received 28th June 2012,
Accepted 9th October 2012

DOI: 10.1039/c2mb25253h

www.rsc.org/molecularbiosystems

1. Introduction

Metabolic flux analysis (MFA), a powerful technique to quantify cellular physiology,[1] involves the system-wide quantification of carbon traffic through cellular biochemical reactions. Metabolic fluxes represent a substantial portion of the "action" occurring in a cell or tissue.[2] Therefore, they are as important as other indicators of phenotype such as transcript levels, protein levels and enzyme activities.[3] Metabolic flux maps are instrumental toward a comprehensive understanding of metabolism.[3–8] By enabling visualization of carbon traffic in metabolic pathways,[9–18] these maps can suggest strategic metabolic engineering targets,[8,19] identify unknown[20,21] or apparently futile metabolic pathways[22] and potentially contribute toward building predictive models of metabolism.[3,23–25]

Department of Chemical and Biomolecular Engineering, University of Maryland, 1208D, Chemical and Nuclear Engineering Building 090, College Park, MD 20742, USA. E-mail: gsriram@umd.edu; Fax: +1 301 405 0523; Tel: +1 301 405 1261

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c2mb25253h

Isotope-assisted MFA is a powerful method of quantifying fluxes, especially in sophisticated metabolic networks such as those of plants. In this method the biological system of interest is fed a designed mixture of labeled (*e.g.* ^{13}C) and unlabeled (*e.g.* ^{12}C) stable isotopes in an isotope labeling experiment (ILE). Fluxes are iteratively evaluated from the ensuing isotopic labeling patterns (such as isotope isomers [isotopomers]) of metabolites and biomass components by computational techniques such as metabolic network modeling, isotopomer balancing, and global optimization.[1,13,26] This task is nontrivial because flux evaluation is a challenging parameter estimation problem in an extensive parameter space,[27] wherein the fluxes are parameters that have to be estimated from isotopomer abundances and other measurements.

Flux identifiability, the confidence with which a flux can be estimated from the information contained in isotope labeling patterns, is a valuable measure of the quality of information obtained from an ILE. Flux identifiability depends on properties of the ILE, and this dependence can be examined on two levels. On one level (structural identifiability), flux identifiability depends on (a) the layout or topology (stoichiometry and carbon atom rearrangements) of the metabolic network under investigation, (b) which labeled carbon sources are supplied and which atoms of these carbon sources are labeled and (c) which metabolites are analyzed and which isotopomers of these metabolites are measured. On another level (statistical identifiability), flux identifiability also depends on (d) the values of the fluxes in the network, (e) the relative proportions of the supplied labeled carbon sources and (f) the measurement errors of the labeling patterns.[28] Whereas the metabolic network layout and flux values (a and d) are beyond the control of the investigator, all the other factors (b, c, e, f) can be chosen judiciously to enhance flux identifiability. However, doing so is a difficult task because the optimal choices are usually not obvious and have to be determined by sophisticated mathematical procedures that compare different ILE designs on the basis of their statistical quality. Insufficient flux identifiability is an acute problem in plant metabolic networks due to their complexity, which arises due to the duplication of pathways in multiple intracellular compartments with different fluxes in each compartment, the existence of many bypasses and cyclic pathways, myriad interconnections between metabolic subnetworks and incompletely known biochemistry.

This work is motivated by the success of previous investigations on microorganisms as well as *in silico* work on mammalian cell and plant embryo metabolism,[28–33] which showed that the *a priori* design of ILEs can lead to a multifold increase in the information obtained from the ILEs.[16,28,29,34] In this article, we report the *a priori* design of ILEs for quantifying fluxes in two important plant metabolic pathways: the pentose phosphate pathway (PPP) and the γ-aminobutyric acid (GABA) shunt. The PPP (Fig. 1a) is a central pathway in plant metabolism and carries substantial carbon flux.[35] Although the PPP wastes a sixth of the carbon that passes through it as CO_2,[36] it regenerates the reductant NADPH (which is necessary for synthesizing fatty acids and certain amino acids) and provides carbon skeletons for the synthesis of nucleotides, phenylpropanoids, lignin and some amino acids.[35] Exceptionally, many plant cells contain the PPP in both the cytosol and the plastid compartments, likely with different carbon traffic patterns in each compartment as dictated by metabolic demand.[35] Due to this duplication and the exchange of PPP metabolites between the duplicated pathways, using ILEs to quantify flux through plant PPPs is challenging. Additional properties of the PPPs that complicate this problem include the presence of many reactions featuring substantial carbon skeleton rearrangement and the reversibility of many reactions that could occasionally cause a cyclic flow of carbon. Furthermore, the complexity of the PPPs also raises several open questions: (i) whether the compartmentation of the PPPs is standard across plants, (ii) how the distribution of carbon traffic in the PPPs varies with environmental conditions such as light, temperature and nitrogen source availability, and (iii) how plants regulate carbon traffic through the PPPs at the level of gene expression. These questions necessitate the development of a tool to precisely quantify carbon traffic in the PPPs.

Although previous investigations have addressed these questions to an extent, there are gaps in knowledge and a clear picture of PPP flux distribution across compartments is lacking. In the pioneering MFA work of Dieuaide-Noubhani *et al.*,[37] the metabolic redistribution of 1-^{14}C and 2-^{14}C glucose (Glc) into sucrose, free Glc and starch in maize root tip cells suggested that the PPP was mostly active in the plastid. Subsequent work by Shachar-Hill and co-workers[5,38,39] as well as Schwender and co-workers on *Brassica napus* embryos, maize embryos and maize endosperm used models in which the oxidative branch of PPP (from glucose-6-phosphate [G6P] to the pentose phosphates) was present in both the cytosol and the plastid, whereas the non-oxidative branch (the rest of the PPP) was present only in the plastid. However, other evidence points to the possibility that the PPP operates in both compartments. For example, Krook *et al.*[40] fed 1-^{13}C Glc to carrot suspension cells and compared the labeling in sucrose (synthesized from cytosolic hexose phosphates[41]) and starch (synthesized from plastidic hexose phosphates[42]). From this, they qualitatively inferred that the PPP is present in both the cytosol and the plastid. Sriram *et al.*,[9] in their work on soybean embryos, observed substantially different isotopomer abundances in hydrolysis products of sugars associated with glycosylated protein (derived from cytosolic hexose phosphates) and starch, and fitted their labeling data to a model that contained the oxidative and non-oxidative PPPs in both the cytosol and the plastid. Allen and co-workers conducted ILEs on soybean embryos and analyzed the labeling patterns of amino acids belonging to the large and small subunits of ribulose-1,5-bisphosphate carboxylase/oxygenase, which are synthesized in the plastid and cytosol respectively. This novel method revealed isotopic differences between amino acids originating in the cytosol and the plastid and also found evidence of flux through the PPP. Together, these results suggest that the allocation of flux to the PPPs in the different compartments may vary.[43] In a recent elaborate study, Masakapalli *et al.*[44] fed 1-^{13}C, 2-^{13}C and U-^{13}C Glc to *Arabidopsis thaliana* suspension cells and examined three

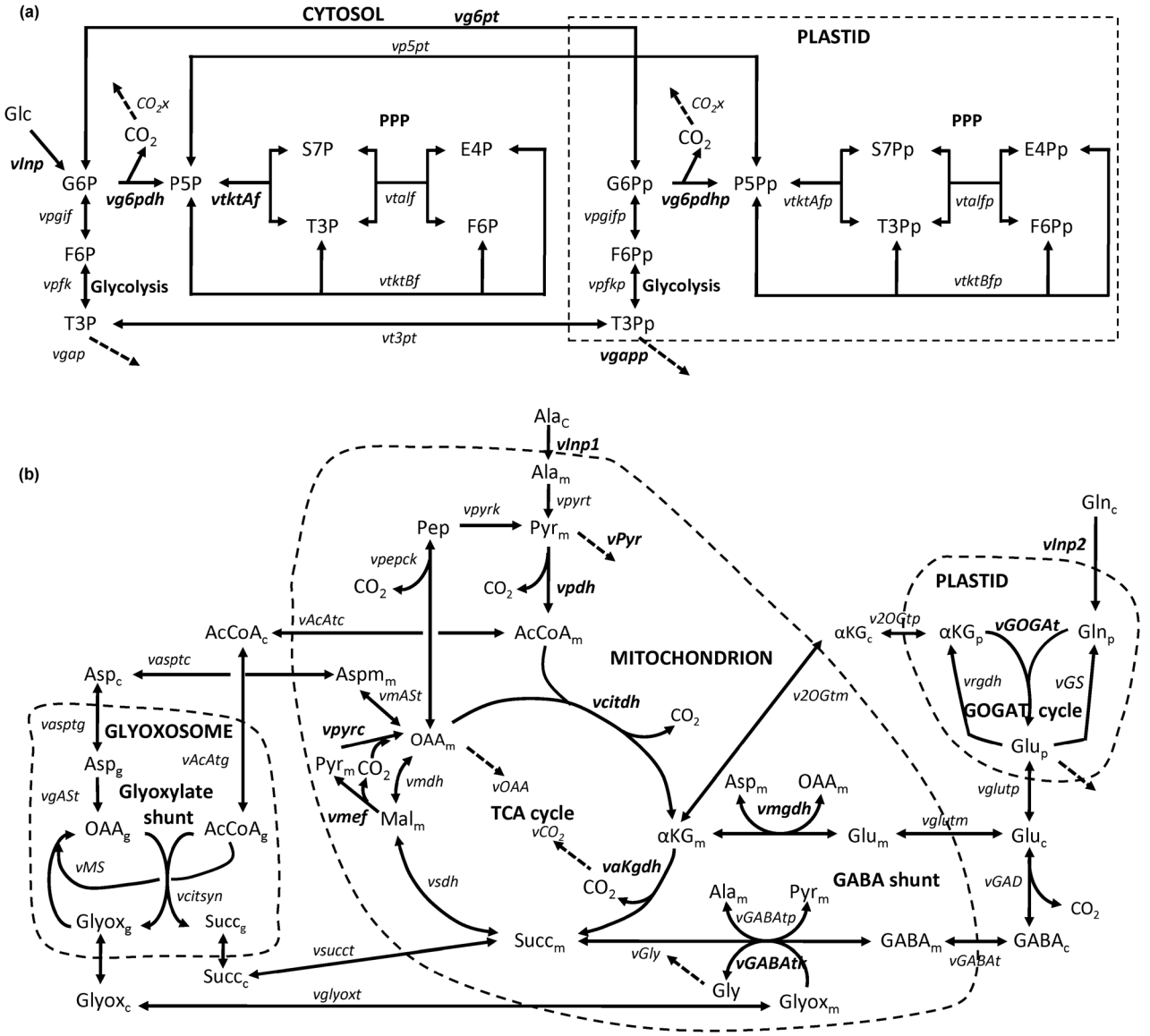

Fig. 1 (a) PPP metabolic model. Glycolysis and PPP duplicated in cytosol and plastid contain 13 metabolites and 32 fluxes (19 net fluxes and 13 reversibility extents). Flux balances allow dependent fluxes to be expressed as linear combinations of directly measurable (*vInp and vgapp*) and free fluxes (*vg6pdh, vtktAf, vg6pdhp* and *vg6pt)*. Dashed arrows indicate metabolites leaving the system. The plastidic compartment is enclosed by a dashed box; suffix 'p' indicates metabolites in the plastidic compartment. Abbreviations: E4P, erythrose-4-phosphate; F6P, fructose-6-phosphate; Glc, glucose; G6P, glucose-6-phosphate; P5P, pentose phosphates; S7P, sedoheptulose-7-phosphate; T3P, triose phosphates. (b) GABA shunt model. The TCA cycle in the mitochondrion, the glyoxylate shunt in the glyoxysome, the GOGAT cycle in the plastid and the GABA shunt across the mitochondrion and the cytosol together comprise a network with 31 metabolites and 34 net fluxes. The measured fluxes are *vInp1, vInp2* and *vPyr;* the free fluxes *are vPdh, vCitdh, vaKgdh, vMgdh, vMef, vPyrc, vGABAtk* and *vGogat*. The intracellular compartments are separated from each by dash lined regions. Dashed arrows indicate metabolites leaving the system. Suffixes 'm', 'p', 'c' and 'g' indicate metabolites in the mitochondrial, plastidic, cytosolic and glyoxysome compartments respectively. Abbreviations: AcCoA, acetyl CoA; Ala, alanine; Asp, aspartate; GABA, γ-aminobutyric acid; Gln, glutamine; Gly, glycine; Glu, glutamate; Glyox, glyoxylate; αKG, alpha-ketoglutarate; Mal, malate; OAA, oxaloacetate; Pep, phosphoenolpyruvate; Pyr, pyruvate; Succ, succinate.

metabolic models that could best account for the concomitant labeling patterns. These three models differed in the subcellular localization of the PPP – one model contained the oxidative and non-oxidative reactions of PPP only in the plastid, another model contained the oxidative reactions in cytosol and plastid with the non-oxidative reactions restricted to the plastid and the third model contained the oxidative and non-oxidative reactions of PPP in both the plastid and the cytosol. Surprisingly, Masakapalli *et al.* observed that all three models explained the data equally well. This illustrates the difficulty in identifying PPP fluxes accurately in plant metabolic networks. Our study aims to address these issues by determining the best isotopic labels and critical isotopomer measurements that can help obtain additional information that will help quantify the PPP fluxes more accurately. Although elaborate and comprehensive MFA studies on PPP in plants exist,[9,10,18,39,44,45] there has been little focus on the design of ILEs involving non-trivial isotope labels, especially for the PPP. We anticipate that the PPP flux estimates of previous studies can be

significantly improved by employing the isotope labeling strategies proposed in this article.

The GABA shunt (Fig. 1b) is a highly interconnected pathway that acts as a crosslink between carbon and nitrogen metabolism.[46] This pathway involves the conversion of glutamate to succinate *via* the non-protein amino acid GABA instead of *via* the tricarboxylic acid (TCA) cycle; therefore, the GABA shunt is a bypass of the TCA cycle. Although GABA is known to play various crucial roles in plants (integration of carbon and nitrogen metabolism,[47] defense against insect attack[48] and pollen tube development[49]) and animals (neurotransmitter[46,50,51]) little is known about carbon flow through the GABA shunt[46] relative to that through the TCA cycle. Researchers have hypothesized that the GABA shunt is a metabolic highway that carries significant carbon flux during normal conditions and even greater flux when a plant faces stress.[46] If this hypothesis is true, then the GABA shunt is one of the first major pathways taken by nitrogen after it enters primary metabolism. This article also reports the design of judicious combinations of labeled carbon sources fed in ILEs that will help test this hypothesis through MFA. Previous isotope-assisted MFA studies of plant metabolic networks[9,10,12,15,18,52] were designed to investigate most central carbon metabolic pathways but did not focus on the GABA shunt.

In this article we identify isotope labels and label combinations that improve the identifiability of important fluxes in PPP and GABA shunt pathways. We also identify the biomass components that contribute maximum labeling information toward flux identifiability. Additionally, we compare the usefulness of labeling information obtained from the two commonly used isotopomer measurement techniques – mass spectrometry (MS) and nuclear magnetic resonance (NMR).

2. Methods

2.1 Metabolic network models for the PPP and the GABA shunt

We modeled metabolic networks by using steady state flux balance equations of the form:

$$\boldsymbol{S}\cdot\boldsymbol{v} = \boldsymbol{0}$$

where $\boldsymbol{v}$ is a vector containing all fluxes and $\boldsymbol{S}$ is a stoichiometric matrix that represents metabolite balance in terms of the fluxes. An outcome of this relationship is that several fluxes in the network (''dependent'' fluxes) are expressible as linear combinations of a smaller set of parameters, which includes: (i) a set of fluxes termed ''free'' fluxes,[53] (ii) the few fluxes that are directly measurable (*e.g.* carbon source uptake), (iii) reversibility extents, relevant to pairs of reversible reactions and (iv) ''scrambling extents'' which, for pairs of reactions that have identical stoichiometries but different carbon atom rearrangements, are ratios indicating how the net flux is split across the two carbon atom rearrangements. SI (ESI†) describes a method for determining a feasible set of free fluxes.

Our model of the PPP (Fig. 1a; SII, ESI†) is based on reaction stoichiometries and carbon atom rearrangements from the Kyoto Encyclopedia of Genes and Genomes (KEGG)[54] as well as previous studies of the PPP.[35] This model comprises glycolysis and the PPP, each duplicated in the cytosol and the plastid compartments. The sole carbon source in the model is Glc (taken up through the flux *vInp*). Carbon exits the network either as triose phosphates from the plastid (flux *vgapp*) or as CO_2 (flux vCO_2x). We lumped the three carbon atom-metabolites dihydroxyacetone phosphate, glyceraldehyde-3-phosphate, phosphoenolpyruvate and pyruvate into a single metabolite named triose-3-phosphates (T3P);[55] we also lumped the five-carbon atom metabolites ribose 5-phosphate, ribulose-5-phosphate and xylulose 5-phosphate into a single metabolite named pentose-5-phosphates (P5P). The absence of carbon atom rearrangements between these metabolites and their relatively rapid equilibration[55] justify this lumping. We modeled the intercompartmental transport of metabolites, such as the reversible transport of G6P, P5P and T3P between the cytosol to the plastid, as bidirectional fluxes. Overall, the metabolic model contains 13 metabolites and 32 fluxes. Of these, two fluxes (*vInp* and *vgapp*) are measurable, and four (*vg6pdh, vtktAf, vg6pdhp, vg6pt*) are free fluxes. Additionally the model contains 13 reversibility extents. Our model of the GABA shunt (Fig. 1b; SII, ESI†) is also based on reaction stoichiometries and carbon atom rearrangements from KEGG and on previous studies on this pathway.[46,56] The model comprises the TCA cycle in the mitochondrion, the glutamine (Gln)-α-ketoglutarate aminotransferase (GOGAT) cycle in the plastid, the glyoxylate shunt in the glyoxysome and the GABA shunt that spans the mitochondrion and the cytosol. We lumped the metabolites fumarate and malate into a single pool[55] and modeled the intercompartmental transport of metabolites as bidirectional fluxes. The network has 31 metabolites and 65 fluxes of which three fluxes (*vInp1*, *vInp2* and *vPyr*) are measurable and eight (*vpdh, vcitdh, vpyrc, vmef, vGOGAt, vGABAtk, vmgdh, vaKgdh*) are free fluxes. Additionally the model contains 22 reversibility extents and one scrambling extent. The only carbon sources in the model are Ala and Gln (taken up through the fluxes *vInp1* and *vInp2* respectively). The metabolites pyruvate, CO_2, plastidic glutamate, mitochondrial oxaloacetate and mitochondrial glycine exit the metabolic network.

2.2 Simulations of ILEs by cumomer balancing

We simulated the isotopomer abundances of metabolites that are measurable by MS and NMR (SIII, ESI† lists these isotopomers) by using cumomer balancing[57] with stoichiometrically feasible flux values. Cumomers (*cum*ulative isotop*omers*) are defined as sums of specific isotopomers. This transformation enables conversion of nonlinear isotopomer balance equations to cascades of linear cumomer balance equations. Cumomers can be easily transformed back into isotopomers.[57] Cumomer balancing provides identical results to the more recent technique of elementary metabolite unit (EMU) balancing, but may require longer simulation times.

2.3 Isotopomer measurements simulated during ILE design

For the identifiability analysis we simulated a comprehensive list of isotopomers ($\boldsymbol{I}_{\mathrm{sim}}$) of metabolites that are known to be measurable by MS and NMR (Tables 1 and 2 list the corresponding metabolites; SIII, ESI† lists all measurable isotopomers). Usually

Table 1 Isotopomer measurements in I_{sim} in the PPP model. The isotopomer abundances of amino acids and carbohydrates whose metabolic precursors are known to be exclusively or predominantly synthesized in the cytosol and the plastid were included in I_{sim}

Analytical technique	Metabolites measured
MS	Ala_m, Glc, Gly_m, His_m, Phe_m, ribose, Ser_m, starch, Tyr_m, Val_m
1-D or 2-D NMR	Ala_n, Gly_n, His_n, LVAgc, LVAgp, LVArc, Phe_n, Ser_n, Tyr_n, Val_n

Subscripts 'm' and 'n' indicate isotopomer measurements by MS and NMR, respectively, for the same metabolite. Abbreviations: Glc, glucose; LVAgc, levulinic acid obtained by hydrolysis of cytosolic glucose; LVAgp, levulinic acid obtained by hydrolysis of plastidic glucose; LVArc, levulinic acid obtained by hydrolysis of cytosolic (RNA) ribose.

Table 2 Isotopomer measurements in I_{sim} in the GABA model. The isotopomer abundances of amino acids whose metabolic precursors are known to be synthesized in specific compartments were included in I_{sim}

Analytical technique	Metabolites measured
MS	Ala_m, Arg_m, Asp_m, Gly_m, Glu_m, Ile_m, Leu_m, Lys_m, Pro_m, Ser_m, Thr_m, Val_m
1-D or 2-D NMR	Ala_n, Arg_n, Asp_n, Glu_n, Ile_n, Leu_n, Lys_n, Met_n, Pro_n, Ser_n, Thr_n, Val_n

Subscripts 'm' and 'n' indicate isotopomer measurements by MS and NMR, respectively, for the same metabolite.

researchers use either MS or NMR to measure isotopomers; however, since these techniques often provide complementary labeling information for a given metabolite,[58] we simulated measurements from both techniques. To differentiate between the fluxes of pathways duplicated in the cytosol and plastid, our PPP model incorporated several metabolites that are known to be synthesized exclusively in one of these two compartments. The subcellular compartmental origins of particular metabolites are well established[59] whereas those of others may be determined by finding the localizations of the enzymes that catalyze their formation reactions.[60–65] Metabolites predominantly or exclusively synthesized in the cytosol include soluble Glc, ribose (from RNA) and Ala, whereas those with a plastidic origin include Val, His, Phe and Tyr, starch[36,59] as well as Gly and Ser. Our GABA shunt included the compartment-specific metabolites Ala, Gly, Ser Val, Ile, Pro, Thr, Asp, Glu, Lys and Arg.

2.4 Statistical flux identifiability

The mathematical techniques of quantifying identifiability of fluxes in a metabolic network have been established previously[28,29,66] and are discussed briefly here. The covariance of fluxes with respect to noisy isotopomer measurements is an indicator of flux identifiability, and the diagonal elements of the covariance matrix (**Cov**) represent variances of the corresponding fluxes.[67] The premise of identifiability analysis is that the **Cov** can be computed without prior knowledge of the true flux values. *A priori* identifiability analysis thus necessitates use of guessed values of free fluxes required to compute the matrix **Cov** which is given by the inverse of the Hessian (***H***) of the chi-square function (χ^2) between I_{sim} and experimentally measured isotopomers (I_{meas}, this term vanishes thus allowing *a priori* analysis, SIV, ESI†),[67]

$$\mathbf{Cov}(f,m) = [H(\chi^2)]^{-1}$$

$$H(\chi^2) = \frac{\partial^2 \chi^2}{\partial f_k \partial f_i}$$

where *k and l* are counters that go over all free fluxes. Comparing covariance matrices obtained using different ILE designs amounts to comparing their statistical flux identifiability. Previous studies have used scalar statistical criteria such as *A*- and *D*-optimality criteria to compare covariance matrices.[28,29] The *A*-optimality criterion is defined as the trace of the covariance matrix and the *D*-optimality criterion is the determinant of the covariance matrix.

$$D_{crit} = \det(\mathbf{Cov})$$

$$A_{crit} = \frac{\mathrm{tr}(\mathbf{Cov})}{n},$$

where n is the number of rows or columns in **Cov**.

A_{crit} signifies the arithmetic mean of the variances whereas $D_{crit}^{1/n}$ signifies the geometric mean of the variances. Since the criteria are proportional to flux variances, high identifiability corresponds to small *A*- and *D*-criteria. This work uses the *A*-criterion as a measure of identifiability since it has certain advantages over the *D*-criterion that were highlighted by Libourel *et al.*[29] Briefly, the difference between the arithmetic mean (AM) and the geometric mean (GM) of the variances of fluxes is greater for the *D*-optimal ILE designs.[29] This follows from the AM-GM inequality. This means that the *D*-optimality criterion may lead to a needle shaped confidence region of fluxes, *i.e.* a *D*-optimal ILE design may resolve all but one flux with an acceptable confidence interval. In this work we verified this claim for 5 pairs of optimal *A*- and *D*-designs (data not shown).

Because the *A*-criterion is inversely proportional to flux identifiability, we use the term ''information yield'' (IY), the square root of inverse of the *A*-criterion, as a metric to compare different ILE designs:

$$\mathrm{IY} = \frac{1}{\sqrt{A_{crit}}}$$

We implemented all isotopomer simulations and IY calculations on our flux evaluation computer program NMR2Flux+.[9,22] NMR2Flux+ uses cumomer balancing[57] (Section 2.2) to simulate

isotopomer abundances from a given set of fluxes, and uses the global optimization algorithm simulated annealing[68] to evaluate fluxes from a given set of isotopomer abundances.

3. Results and discussion

3.1 Plant PPP fluxes are best identified with 100% 1,2-^{13}C Glc (with our set of isotopomer measurements)

The choice of appropriately labeled carbon source is paramount in ILE design because it crucially determines both the amount of information obtainable from the experiment and the cost of the experiment. Therefore, our first objective was to determine which of the commercially available labels of Glc provide the maximal information toward identifying fluxes in the plant PPPs. Toward this we computed IY for ILEs that employ each of the eight commercially available Glc labels (ESI†) and naturally abundant Glc, mixed in different proportions. This analysis (Fig. 2) revealed that the PPP fluxes are best identified with 100% 1,2-^{13}C Glc (IY = 21.9 [arbitrary units]), followed by 100% 3-^{13}C Glc (IY = 19.6) and the popularly used[16,22,44] label 100% 1-^{13}C Glc (IY = 18.2). For all labels except U-^{13}C Glc, IY increases with the proportion of labeled Glc, implying that dilution of these labels with naturally abundant Glc reduces the information available from them. The exception, U-^{13}C Glc, is explained by the fact that 100% U-^{13}C Glc completely labels all carbon atoms of intracellular metabolites with ^{13}C, thus resulting in no differential distribution of the label by different pathways. Therefore, it is essential to dilute this label with naturally abundant Glc. Additionally, U-^{13}C Glc generates isotopomers that are suitable for measurement by NMR;[10,69] therefore, it is not surprising that this label is often employed at proportion as low as 5%[10] and 10%.[70]

Furthermore, we analyzed the performances of 55 Glc labels that, to our knowledge, are commercially unavailable except through custom synthesis (all isotopomers of Glc except those listed in SV, ESI†). Nine of these 56 Glc labels performed better than 1,2-^{13}C Glc, of which the three best labels were 100% 3,4,5,6-^{13}C Glc (IY = 23.7), closely followed by 100% 1,2,4-^{13}C (IY = 23.3) and 3,5,6-^{13}C Glc (IY = 22.9) (Table 3).

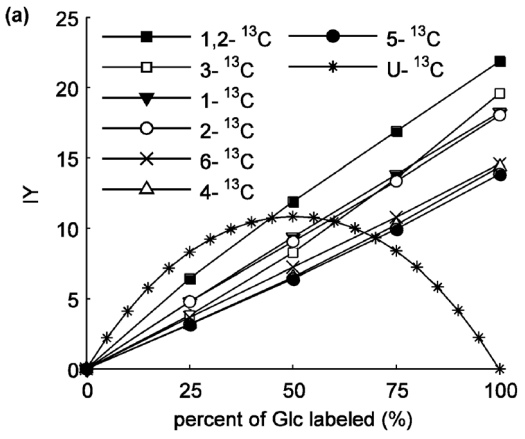

Fig. 2 100% 1,2-^{13}C Glc outperforms other commercially available Glc labels in estimating PPP fluxes. This plot depicts simulated IY against extent of labeling of Glc. The values on the horizontal axis indicate the percentage of the label in the supplied Glc; the rest of the supplied Glc is naturally abundant.

Table 3 Performances of commercially unavailable Glc labels for the plant PPP network. This list contains the top 10 best-performing Glc labels. All labels are at 100% of total Glc. Nine Glc labels perform better than 1,2-^{13}C Glc, which is the best performing amongst the commercially available Glc labels

Atom(s) of Glc labeled ^{13}C	IY
3,4,5,6	23.7
1,2,4	23.3
3,5,6	23.2
1,4,5,6	22.9
2,3	22.6
3,4,6	22.6
1,2,5	22.3
2,4,5,6	22.0
2,3,6	21.9
1,2,4,5,6	21.7

3.2 *In silico* ILEs corroborate the superiority of 1,2-^{13}C Glc over 1-^{13}C Glc in identifying PPP fluxes

The results presented above lead to the question: how does an IY value translate into actual flux identifiability – does the slightly higher IY of the best commercial label 1,2-^{13}C Glc (IY = 21.9) over the popularly used label 1-^{13}C Glc (IY = 18.24) imply that 1,2-^{13}C Glc is significantly better in identifying PPP fluxes? To answer this question, we performed *in silico* ILEs as follows: from arbitrarily chosen values of the four free fluxes in the PPP model, we simulated isotopomer abundances resulting from ILEs employing either 100% 1,2-^{13}C Glc or 100% 1-^{13}C Glc. We then treated these simulated isotopomer abundances as surrogate experimental measurements and allowed NMR2Flux+ to evaluate, by minimizing χ^2 through global optimization, a set of fluxes that best accounted for these surrogate measurements. Repeating this flux evaluation several (478) times from random initial points, we selected the evaluations that converged to χ^2 values less than 20 (corresponding to a confidence level of 99.96% for four degrees of freedom). This resulted in distributions for each flux (Fig. 3 and 4), which we compared to the initially chosen (“true”) flux values from which we had simulated the isotopomer abundances.

Interestingly, some fluxes had bimodal distributions, *e.g. vt3pt* (Fig. 3a), *vpfk*, *vpgifp*, *vpfkp*, *vg6pdhp and vg6pt* (not shown), whereas other fluxes had unimodal distributions, *e.g. vtktAf* (Fig. 4b). The bimodal distributions exhibited a major peak close to the true flux value and a minor peak far away from it. For instance, the major peak in the distribution of the flux *vt3pt* represented 385 out of 478 (~81%) flux evaluations and was centered around the true flux value of 0.2, whereas the minor peak that represented the remaining 19% of the flux evaluations was centered away from the true flux value (Fig. 3a). However, the points on the minor peak corresponded to χ^2 values between 7 and 20; therefore, using a stringent cutoff of $\chi^2 < 7$ completely eliminated the minor peak (Fig. 3b). Applying a $\chi^2 < 7$ cutoff to eight other fluxes that initially showed a bimodal distribution eliminated their minor peaks and retained the major peaks centered around the true flux

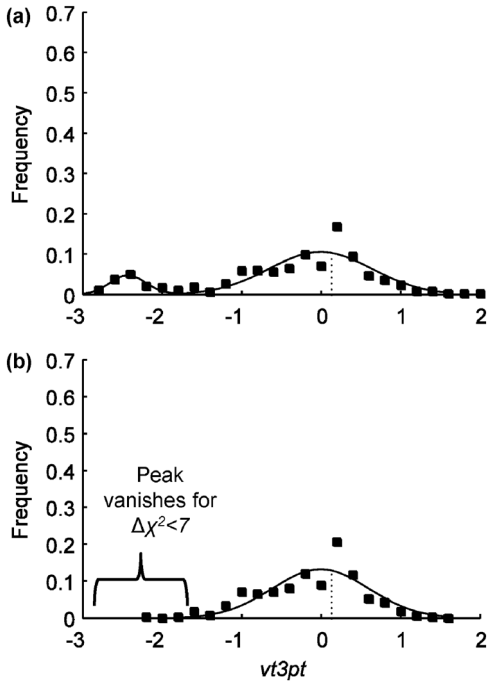

Fig. 3 Enforcing stricter criteria for the χ^2 goodness-of-fit function can improve flux estimates and eliminate bimodal distributions. We simulated isotopomer abundances for ILEs employing 100% 1,2-^{13}C Glc, treated these simulated isotopomer abundances as surrogate experimental measurements and then, by minimizing χ^2 through global optimization, evaluated a set of fluxes that best accounted for these surrogate measurements. We repeated this flux evaluation 478 times from random initial points to obtain a flux distribution. Enforcing a goodness-of-fit criterion of $\chi^2 = 20$ (corresponding to a confidence level of 99.96% for four degrees of freedom) gave bimodal distributions for some fluxes (the distribution of the flux *vt3pt* is shown here); enforcing the stricter criterion of $\chi^2 = 7$ eliminated the minor peak and retained the major peak around the ''true'' flux value from which we had originally simulated the isotopomer abundances. Therefore, enforcing a stricter goodness-of-fit fitting criterion can significantly improve flux estimates.

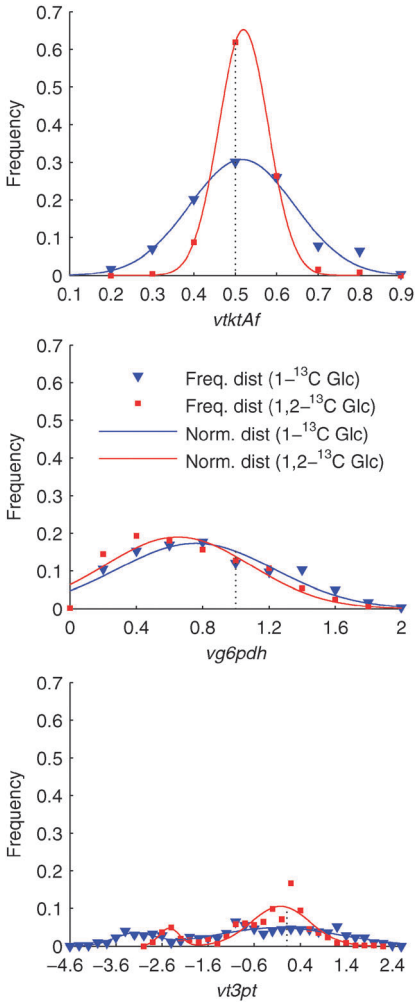

Fig. 4 *In silico* ILEs with 100% 1,2-^{13}C Glc estimate several fluxes better than 100% 1-^{13}C Glc. We obtained flux distributions as explained in the caption of Fig. 3 and the text. The flux (a) *vtktAf* is more identifiable by the *in silico* ILE with 100% 1,2-^{13}C Glc (red symbols and line) than with 100% 1-^{13}C Glc (blue symbols and line), as the distribution corresponding to 1,2-^{13}C Glc clusters is closer to the true flux (dotted line). Both labels identify the flux (b) *vg6pdh* to nearly the same extent. The flux (c) *vt3pt* is identifiable only with 100% 1,2-^{13}C Glc. This illustrates the superiority of 1,2-^{13}C Glc over 1-^{13}C Glc in estimating fluxes in the compartmented plant PPPs.

values. This suggests that accurate identification of PPP fluxes requires a stringent χ^2 cutoff value heuristically learned from *a priori* simulations. In the distributions discussed in the rest of Section 3.2 (Fig. 4), we only consider flux sets corresponding to $\chi^2 < 7$.

These flux distributions showed that while both 1,2-^{13}C Glc and 1-^{13}C Glc identified certain fluxes equally well (*e.g.* the cytosolic oxidative PPP flux *vg6pdh*), 1,2-^{13}C Glc identified certain fluxes significantly better than 1-^{13}C Glc (*e.g.* the non-oxidative PPP flux *vtktAf*) (Fig. 4a and b). Furthermore, 1,2-^{13}C Glc was able to reasonably identify certain fluxes that were not at all identified by 1-^{13}C Glc (*e.g.* the intercompartmental T3P transport flux *vt3pt*) (Fig. 4c). Apart from corroborating the identifiability results of Section 3.1, this outcome demonstrates that relatively small increases in IY could translate into significant differences in flux identifiability. Overall, 1,2-^{13}C Glc identified as many as 10 out of the 19 fluxes in the model very close to their ''true'' values, including glycolytic fluxes in the cytosol (*vpgif*, *vpfk*) and the plastid (*vpgifb*) and the oxidative PPP flux in the cytosol (*vg6pdh*) (SV, ESI†). However, 1,2-^{13}C Glc does not identify all fluxes satisfactorily – certain fluxes, including glycolytic fluxes in the plastid (*vpfkp*), the oxidative PPP fluxes in the plastid (*vg6pdhp*) and the fluxes of the intercompartmental G6P and P5P transporters (*vg6pt*, *vp5pt*) (SV, ESI†) were not well identified by this label. Therefore we examined if combining this label with other labels would increase flux identifiability (Section 3.7).

3.3 Is 1,2-^{13}C Glc always the best choice for the plant PPPs?

Two recent investigations that have focused on designing labels for mammalian PPPs serve as a benchmark for our work. In a study that focused on the mammalian PPP (unicompartmental model, reactions of the PPP assumed irreversible),[71] Crown and Antoniewicz identified 100% 2,3,4,5,6-^{13}C Glc and its complement 100% 1-^{13}C Glc as the best Glc labels for this pathway. In another study focused on mammalian primary metabolic

pathways including the PPP (unicompartmental model, reversible reactions included),[31] Metallo *et al.* identified 1,2-^{13}C Glc as the best among 11 Glc labels for estimating PPP and glycolysis fluxes. Interestingly, our work converged the result of Metallo *et al.* – we identified 1,2-^{13}C Glc as the best commercially available label and its complement 3,4,5,6-^{13}C Glc as the best commercially unavailable label for the plant PPPs. This similarity is despite several major differences between the metabolic network models and isotopomer measurements considered by Metallo *et al.* and our study – our PPP model contains compartmentalized duplicates of the PPP and glycolysis as well as intercompartmental transport reactions, a hallmark of plant cells. Additionally, we have considered a larger number and variety of readout metabolites, including essential amino acids not synthesized by mammalian cells, sugars and nucleic acids as well as two complementary isotopomer measurement techniques (NMR and MS). Metallo *et al.* rationalized the superiority of 1,2-^{13}C Glc by showing that if Glc were labeled at the C-2 atom, it would enrich a larger number of carbon atoms of PPP metabolites than if it were labeled at other carbon atoms (*e.g.* C-4). This is due to the repeated breakage and re-formation of the C-1–C-2, C-2–C-3 and C-3–C-4 bonds of Glc in the reversible reactions of the PPP; conversely, the bonds between C-4–C-5 and C-5–C-6 remain largely intact.[31,35]

To advance this line of reasoning, we examined how many isotopomer abundances change significantly (by >0.01 units on a scale ranging from 0 to 1 units) when the two most important fluxes in the PPP network – the oxidative PPP fluxes in the cytosol (*vg6pdh*) and the plastid (*vg6pdhp*) – are each perturbed by 20%. This calculation showed that 100% 1,2-^{13}C Glc surpasses both 100% 1-^{13}C Glc and 30% U-^{13}C Glc in both the sum of isotopomer abundance changes (Fig. 5a) and the number of altered isotopomers (Fig. 5b). However, this result applies to the set of MS- and NMR-derived isotopomer measurements considered in our model (SIII, ESI†). Although this is a large set of isotopomers, MS and NMR can only measure a subset of all 2^n isotopomers of an n-carbon metabolite. Unexpectedly, we found that if all 2^n isotopomers of each biomass component could be measured (instead of our subset of MS- or NMR-derived measurements), 30% U-^{13}C Glc surpasses the other two labels in both the sum of isotopomer abundance changes (Fig. 5c) and the number of altered isotopomers (Fig. 5d). Therefore, 1,2-^{13}C Glc is the best label for the plant PPPs with respect to the subset of isotopomers that can be measured by the currently popular versions of MS (single quadrupole) and NMR (2-D [^{13}C, ^{1}H] HSQC[52,72] or [^{1}H, ^{1}H] TOCSY[52]). However, the measurement of all possible isotopomers of all PPP metabolites may potentially result in a new experimental design. Quantifying the abundances of all isotopomers of the five-, six- and seven-carbon metabolites of the PPP may require significant advancement of the measurement techniques. However, recent developments such as tandem MS can measure all 2^n isotopomers of four-carbon metabolites such as Asp.[73,74] Therefore, it is reasonable to expect that further improvements may make it possible to measure a large fraction of the 2^n isotopomers of each PPP metabolite.

3.4 Superior performance of 1,2-^{13}C Glc is largely independent of PPP flux values

A central premise of statistical flux identifiability analysis is that the information contained in an ILE is generally independent of the values of the fluxes. Therefore, the best labels determined by assuming one set of flux values should also be the best labels for any stoichiometrically feasible set of flux values. To examine if this was the case, we repeated the analysis described in Section 3.1 and Fig. 2 for 36 randomly chosen combinations of free flux values that spanned the stoichiometric range of the metabolic network. This analysis (Fig. 6) revealed that 100% 1,2-^{13}C Glc was the best-performing label in 32 of the 36 (89%) combinations of flux values and that 100% 3-^{13}C was the second best Glc label in 20 out of 36 (56%) combinations. The consistent performance of 1,2-^{13}C Glc over a

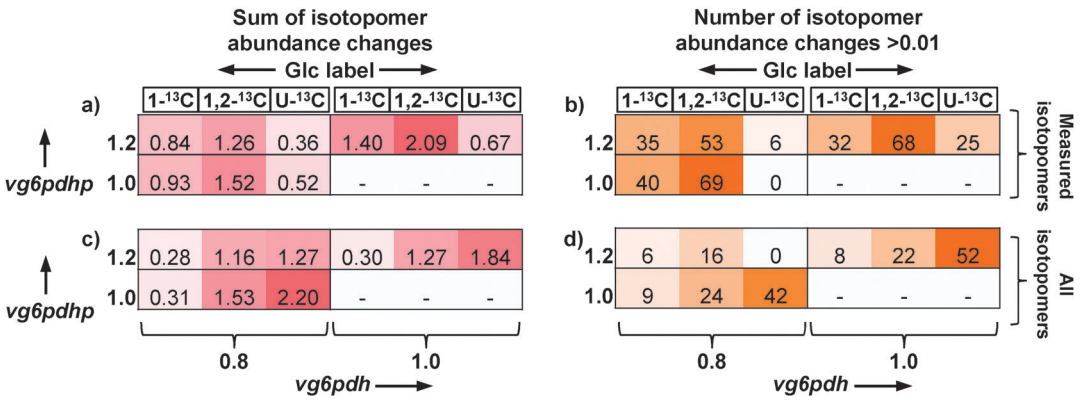

Fig. 5 MS- and NMR-derived isotopomer abundances are most sensitive to flux changes in the ILE with 100% 1,2-^{13}C Glc. To examine the sensitivities of isotopomers to flux changes, we simulated ILEs employing 100% 1-^{13}C, 100% 1,2-^{13}C and 30% U-^{13}C Glc, and specifically examined how the perturbation of the cytosolic oxidative PPP flux *vg6pdh* (1.0 → 0.8) and the plastidic oxidative PPP flux *vg6pdhp* (1.0 → 1.2) alters isotopomer abundances. (a) The sum of changes and (b) the number of changes >0.01 in the MS- and NMR-derived subset of isotopomer abundances indicate that 100% 1,2-^{13}C Glc renders the isotopomer abundances most sensitive to the given flux changes. Interestingly, (c) the sum of changes and the (d) number of changes >0.01 in all 2^n isotopomers of all PPP metabolites indicate that 30% U-^{13}C Glc renders isotopomer abundances most sensitive to the given flux changes.

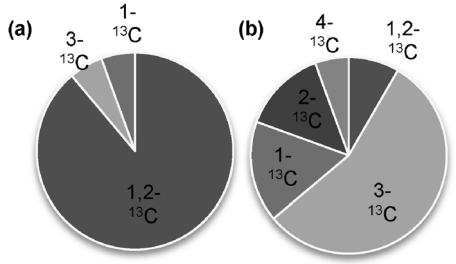

Fig. 6 Superior performance of the 1,2-^{13}C Glc label is consistent across different sets of free flux values. Calculation of IY for 36 ILEs with different sets of free flux values shows that (a) 100% 1,2-^{13}C Glc performs the best for 89% of the free flux combinations and (b) 100% 3-^{13}C Glc performs second-best for 56% of free flux combinations, thus validating the premise that the identifiability analysis is fairly independent of free flux values.

range of feasible flux values validates the premise stated above and suggests that this label should provide significant information for most plant PPP networks. Some of the isotopomer abundances in $\boldsymbol{I}_{\text{sim}}$ were absent in these simulations; nonetheless, this does not affect the performance of the Glc labels.[29]

3.5 Labeling information contained in hexose and pentose sugars is critical in elucidating PPP compartmentation

Traditionally, isotope-assisted MFA involves the measurement of labeling in proteinogenic amino acids derived from acid hydrolysis of a cell pellet or a protein extract.[69,70,75] This experimentally straightforward technique provides metabolic information from various parts of the central carbon metabolic network because the biosynthetic precursors of the amino acids are distributed throughout this network. However, this may be inadequate for the plant PPPs. The complex carbon rearrangements and intercompartmental transfer of sugars in the plant PPPs may not be reflected in the few amino acids that originate from PPP metabolites. The measurement of labeling in hexose and pentose sugars of the PPP may perhaps reveal more information, especially on compartmentation. For example, carbohydrates such as intracellular sucrose or Glc and sugars in glycosylated protein (mannose, glucosamine) reflect cytosolic hexose phosphates, whereas starch reflects plastidic G6P. Therefore, a comparative analysis of the isotopomers of these compounds can potentially reveal differences between cytosolic and plastidic G6P pools, as conceptualized by Roscher *et al.*[45] and experimentally demonstrated by Sriram *et al.*[9] Similarly, ribose in nucleic acids likely has a predominantly cytosolic origin; hence its isotope labeling may contrast with the pentose backbone of histidine, which has a plastidic origin. Consequently, comparing the isotopomers of ribose from nucleic acids and histidine from protein will reveal differences in the P5P pools of the cytosol and the plastid.

However, the extraction, processing and analysis of many of the aforementioned compounds are laborious and not surprisingly, their concurrent measurement for isotope-assisted MFA is rare and has not been reported for plants. Therefore, it is worthwhile to computationally analyze the incremental benefit of measuring the labeling in hexose and pentose sugars. Toward this, we simulated IY for ILEs with varying proportions of 1,2-^{13}C Glc by sequentially including in $\boldsymbol{I}_{\text{sim}}$ labeling measurements from the following biomass components: (i) only proteinogenic amino acids, (ii) intracellular glucose, (iii) starch and (iv) RNA ribose. Clearly, each successive labeling measurement substantially increases IY, the highest incremental benefit being in the case of RNA ribose (Fig 8a).

Further, to examine whether metabolic information from hexose and pentose sugars is more pronounced in MS or NMR measurements, we calculated IY for exclusively MS measurements (Fig. 7a) and exclusively NMR measurements (Fig. 7b). With NMR, one can measure hexose and pentose sugars through their five-carbon acid hydrolysis product levulinic acid (LVA):[72] LVA resulting from hydrolysis of glycosylated protein is LVAgc, that resulting from starch hydrolysis is LVAgp and that resulting from ribose hydrolysis is LVArc. Again, each successive labeling measurement provides significant additional information with the highest incremental benefit being in the case of ribose-derived LVA (Fig. 7a and b).

The concurrent measurement of hexose and pentose phosphate pools from the cytosol and the plastid increases IY likely because it disentangles the effects of the oxidative and the non-oxidative PPPs in the two compartments. In both the cytosol and the plastid, G6P loses its C-1 carbon atom to form P5P in the oxidative PPP. Therefore, reduced labeling of cytosolic or plastidic G6P in an ILE employing 1-^{13}C Glc or 1,2-^{13}C Glc can reveal the presence of the oxidative PPP in the corresponding compartment. However, the complex carbon rearrangements in the downstream non-oxidative PPP in both compartments can confuse this interpretation. For example, high fluxes through the cytosolic oxidative PPP and the plastidic non-oxidative PPP can superimpose over each other, making the compartmentation unidentifiable. However, many of the rearrangements in the non-oxidative PPP are captured in the pentose phosphates synthesized in the respective compartments. Therefore the addition of pentose phosphate isotopomer measurements from the cytosol and the plastid can provide information orthogonal to that contained in the hexose phosphates. This explains the superior performance of the combination of isotopomer measurements from intracellular glucose, starch, RNA and histidine.

3.6 MS outperforms NMR in identifying PPP fluxes

The simulations presented above also shed light on the relative efficacies of MS- and NMR-derived isotopomers in estimating PPP fluxes. MS and NMR measure different linear combinations of the isotopomers of a particular metabolite:[58] MS measures the mass isotopomer abundances of metabolite fragments, whereas 1-D NMR measures positional ^{13}C enrichments and 2-D NMR measures populations of isotopomers containing different sequences of ^{13}C-^{13}C or ^{12}C-^{13}C bonds. As per the results in Fig. 7, MS measurements yield substantially higher IY values than NMR measurements; therefore, MS is clearly better than NMR in measuring fluxes through the plant PPPs. It could be argued that this is due to the higher analytical sensitivity of MS – based on our previous experience and the literature,[9,58] errors in MS-derived isotopomer abundances are typically less

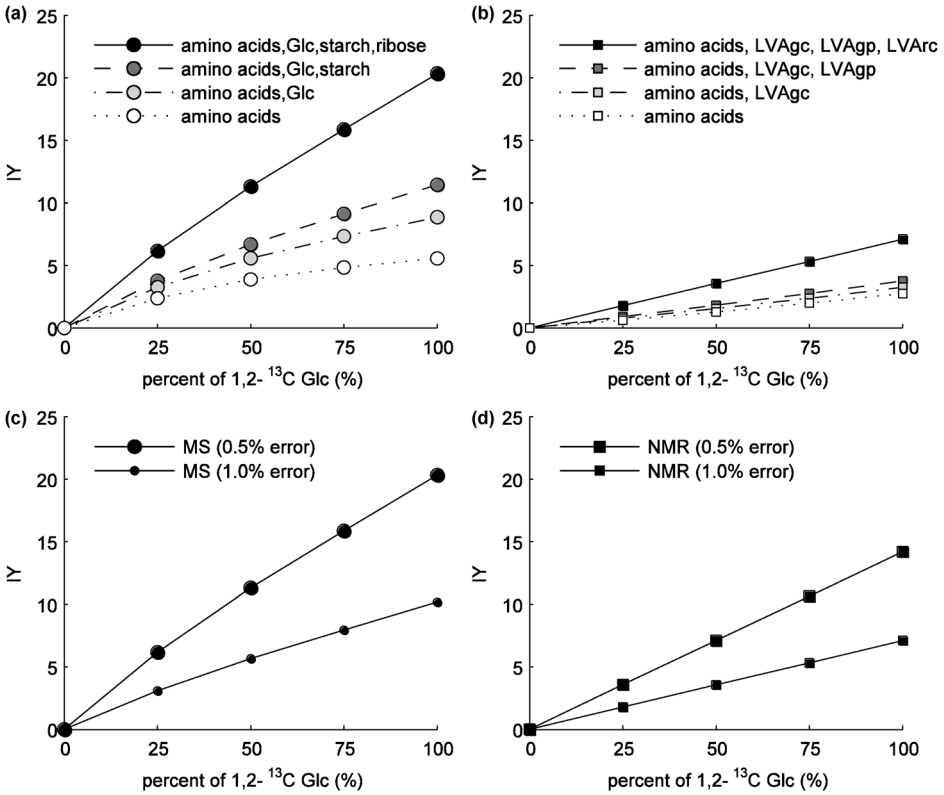

Fig. 7 Measurement of isotopomers of hexose and pentose metabolites with cytosolic and plastidic origins substantially improves identifiability. IYs of ILEs with different percents of 1,2-^{13}C Glc and including isotopomer measurements of different biomass components by (a) MS and (b) NMR show that ribose isotopomers contribute substantially to improving identifiability of compartmented PPP fluxes. Additionally, for the same metabolite and measurement error, the ability of the (c) MS to accurately identify the PPP fluxes is greater than that of the (d) NMR. Abbreviations: LVAgc, cytosolic glucose-derived levulinic acid, LVAgp, plastidic glucose (starch)-derived levulinic acid; LVArc, cytosolic ribose-derived levulinic acid.

than 0.005 (or 0.5%), whereas errors in NMR-derived abundances are of the order of 0.01 (or 1%). To examine whether MS measurements were superior solely due to their higher precision, we compared the MS and NMR measurements of the same metabolites at the same precision level. Fig. 7c and d depict that even if MS- and NMR-derived isotopomers had identical precision, MS gives higher IY values than NMR. This implies that, for the given metabolites whose labeling was measured, MS performs better than NMR in identifying plant PPP fluxes because the types of isotopomers it measures are more sensitive to PPP fluxes, and not only because it is a more precise technique. In other words, the MS measurements confer higher structural identifiability due to their increased sensitivity to PPP fluxes as well as higher statistical identifiability due to their higher precision. The superiority of MS over NMR is specific to the PPPs; the relative strengths of these techniques may compare differently for a network with a different topology.

3.7 Performance of Glc labels in pairs and triads of ILEs in estimating PPP fluxes

We quantified the performance of ILEs that simultaneously employed pairs of commercially available Glc labels (SV, ESI†) at different proportions (data not shown). This analysis revealed that there was no merit in using mixtures of Glc labels, because in general, an ILE employing a mixture of two labels had a lower IY than ILEs that individually employed each label constituting the mixture. For instance, an ILE employing a mixture of 50% 1,2-^{13}C Glc and 50% 1-^{13}C Glc (IY = 17.5) had a lower IY than one employing 100% 1,2-^{13}C (IY = 21.9; Fig. 2) or 100% 1-^{13}C Glc (IY = 18.2; Fig. 2). This is because a mixture of two labels suffers from the dilution effect that occurs when a single label is diluted with a naturally abundant version of the carbon source (Fig. 2). Isotopomers resulting from one of the two labels mask those of the other, thus diminishing the information obtained from the label mixture.

Nevertheless, the deployment of two Glc labels in two or more parallel ILEs followed by flux evaluation from the combined measurements of both the ILEs should enhance flux identifiability. This will leverage the unique information offered by the two labels without allowing one to mask the other.[12,29] To explore this possibility, we examined the performance of pairs and triads of commercially available Glc labels when used in parallel ILEs. All labels performed relatively better when paired with 100% 1,2-^{13}C Glc and relatively worse when paired with 50% U-^{13}C Glc (Fig. 8a). The pair {1,2-^{13}C, 3-^{13}C Glc} performed the best (Fig. 8a), with the performance of each label increasing with increase in its proportion and reaching a maximum at 100% proportion of both labels (Fig. 8b). The IY of the combination {100% 1,2-^{13}C, 100% 3-^{13}C Glc} was 29.5 (Fig. 8a), significantly higher than that of just 100% 1,2-^{13}C Glc (IY = 21.9; Fig. 2).

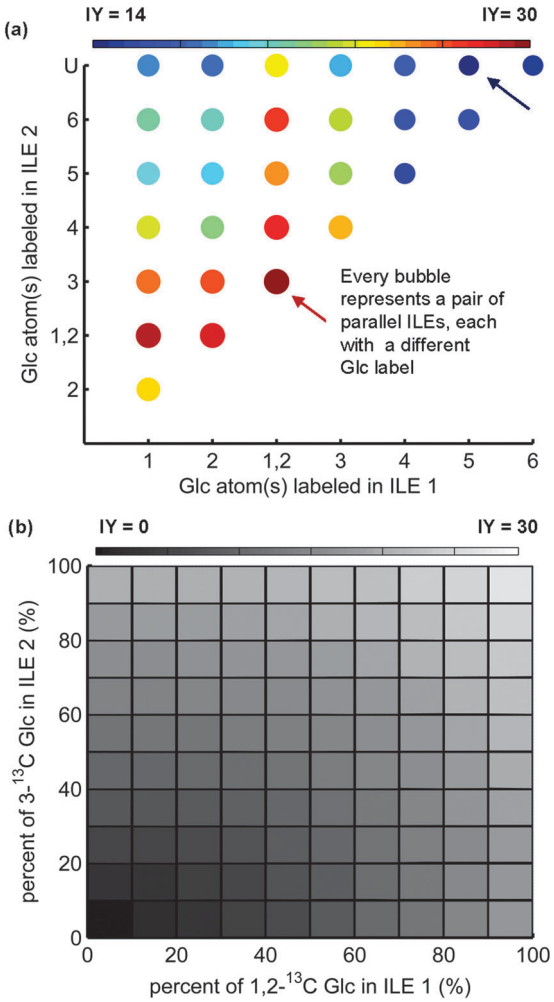

Fig. 8 Identifiability of PPP fluxes improves upon pooling isotopomer abundances from two parallel ILEs. (a) The axes list the 100% Glc labels (50% in the case of U-^{13}C) in two parallel ILEs – ILE1 and ILE2. Each bubble represents a pair of parallel ILEs whose IY was obtained by pooling the isotopomer measurements from these ILEs. Both the sizes of the bubbles and their color (as indicated in the color bar) are independently proportional to IY. Pooling isotopomer measurements from ILEs with 100% 1,2-^{13}C Glc and 3-^{13}C Glc (IY = 29.5, highlighted by a red arrow) is the most advantageous and better than the single ILE with 100% 1,2-^{13}C Glc (IY = 21.9). (b) IY (lighter shades of gray correspond to increased IY as shown in the color bar) of pooled isotopomer measurements from ILEs with 1,2-^{13}C and 3-^{13}C increases with increase in their proportions, *i.e.* dilution of any of the two labels with naturally abundant Glc is undesirable.

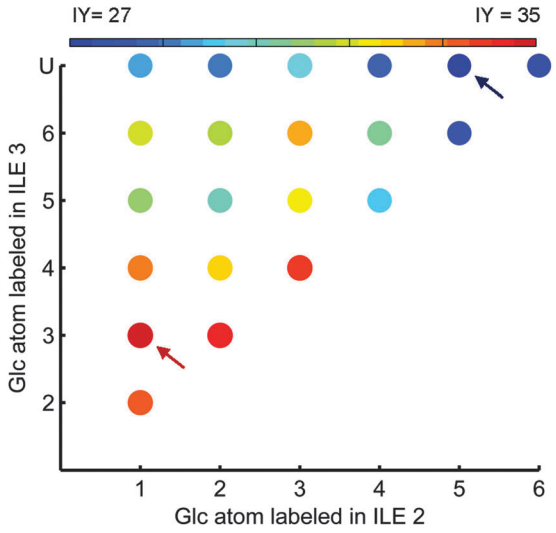

Fig. 9 Performances of Glc labels for triads of ILEs. This bubble plot is similar to the one in Fig. 8a. After locking 100% 1,2-^{13}C Glc as one label, we determined IYs for triads of ILEs employing this label and two other Glc labels. Pooling measurements from ILEs using 100% 1,2-^{13}C Glc, 100% 3-^{13}C Glc and 100% 1-^{13}C Glc, respectively, is the most advantageous, with IY = 34.8.

Identifying optimal triads of ILEs was a more difficult problem due to the large number of possible triads. To circumvent this difficulty, we locked 100% 1,2-^{13}C Glc as one of the labels in the triad (due to its superior performance established in Section 3.1–3.4). Then we explored two other labels that could be used with 1,2-^{13}C Glc in parallel experiments. The triad {100% 1,2-^{13}C Glc, 100% 1-^{13}C, 100% 3-^{13}C Glc} performed the best with IY = 34.8 (Fig. 9), significantly higher than the best pair {1,2-^{13}C, 3-^{13}C Glc} (IY = 29.5; Fig. 8a) or single label 1,2-^{13}C (IY = 21.9; Fig. 1).

We investigated the best labels or label combinations for the plant PPPs when exclusively MS-derived or exclusively NMR-derived isotopomer measurements are available (Table 4; all labels at 100% proportion). 1,2-^{13}C, 3-^{13}C and 1-^{13}C Glc consistently featured amongst the top labels, 1,2-^{13}C Glc being more prominent when MS measurements are available and 3-^{13}C Glc being more prominent when NMR measurements are available. This suggests that more or less the same set of labels is optimal when either MS or NMR measurements are available, although the use of MS leads to substantially greater flux identifiability.

3.8 Performance of labeled carbon sources for the GABA shunt

We chose Ala and Gln as the carbon sources for the GABA shunt network because they enter this network *via* completely different routes – Ala enters through the mitochondrial TCA cycle and Gln through the plastidic GOGAT cycle. Therefore, their differential labeling can potentially provide significant flux information for this network. We evaluated IY for commonly available Ala and Gln labels (SV, ESI†) at different proportions. Table 5 lists the five best performing labels of Ala and Gln, of which the combination {100% 2-^{13}C Ala, 100% U-^{13}C Gln} has the highest IY (= 71.8) toward estimating GABA shunt fluxes. The combination {100% 2-^{13}C Ala, 75% U-^{13}C Gln} has nearly the same IY (= 70.3) as 75% 3-^{13}C Gln and may serve as a good substitute especially as it offers high identifiability at less than 100% proportion, *i.e.* at reduced experimental cost.

The fluxes in the GABA network are more interlinked than those in the PPP network due to the presence of many cyclic pathways. This makes it difficult to track the fates of different carbon atoms. Additionally, the fates of labeled carbon atoms originating from Ala and Gln cannot be distinguished from each other once they are assimilated into succinate (Succ$_m$). Thus speculating on the reasons for the relative merits of different Ala and Gln labels is not easy. Nevertheless, simulations

Table 4 Performance of Glc labels when I_{sim} comprised solely MS- or solely NMR-derived isotopomer measurements. This list contains five of the best performing Glc labels and their IYs for single, pairs and triads of ILEs. MS measurements have higher IYs when compared to corresponding NMR measurements

	Best five single ILEs		Best five paired ILEs			Best five triads of ILEs			
Measured by	Label	IY	Label 1	Label 2	IY	Label 1	Label 2	Label 3	IY
MS	1,2	20.3	1,2	1	27.3	1,2	1	3	32.6
MS	1	17.7	1,2	3	27.0	1,2	1	2	32.5
MS	3	17.7	1,2	2	26.6	1,2	1	4	32.3
MS	2	16.8	1,2	4	26.4	1,2	3	2	32.0
MS	6	13.7	1,2	6	25.5	1,2	3	4	32.0
NMR	3	7.5	3	1,2	10.6	1,2	3	4	11.9
NMR	1,2	7.1	3	4	9.4	1,2	3	2	11.8
NMR	2	5.2	3	2	9.3	1,2	3	6	11.5
NMR	4	4.0	4	1,2	9.0	1,2	3	1	11.0
NMR	5	3.7	3	6	8.9	1,2	3	5	10.9

Table 5 Five best-performing Ala and Gln labels for the GABA shunt network. An ILE with 100% 2-^{13}C Ala and 100% U-^{13}C Gln corresponds to the best IY

Ala label	Percent Ala	Gln label	Percent Gln	IY
2	100	U	100	71.8
2	100	U	75	70.3
2	100	U	50	68.1
2	100	3	75	65.8
2	100	3	100	65.6

showed that an ILE employing 100% 2-^{13}C Ala and 100% U-^{13}C Gln generated a greater number of isotopomers in the GABA network than other ILEs (data not shown), which partially explains the superior performance of this label combination.

4. Summary and outlook

This article explores in substantial detail the design of ILEs toward quantifying fluxes through two important, complex, compartmented plant metabolic pathways. We determined optimal combinations of commercially available Glc labels for the PPP as well as Ala and Gln labels for the GABA shunt. In particular, we established that given currently popular isotopomer measurement techniques (single quadrupole MS and 1-D or 2-D NMR), 1,2-^{13}C Glc is a powerful and robust label for the plant PPPs. We also calculated that its potency can substantially be improved by combining it with other labels (*e.g.* 3-^{13}C Glc, 1-^{13}C Glc and U-^{13}C Glc) in parallel ILEs. We showed that measuring the labeling patterns of hexose and pentose moieties synthesized exclusively in the cytosol or the plastid is important toward evaluating fluxes in the individual compartments. Specifically, the concurrent measurement of RNA ribose, intracellular glucose or sucrose and starch, although laborious, adds critically to the information obtained from the ILE. Additionally, we showed that MS outperforms NMR in identifying fluxes in the PPPs. The label designs and measurements proposed in this study have not been simultaneously employed for plant MFA. End-users of this work should bear in mind that although our metabolic models are representative of typical scenarios in plant cells, our optimal ILE designs obtained are most directly applicable to the ILEs employing the carbon sources and isotopomer measurements that we have considered. Researchers employing other labeled carbon sources (*e.g.* sucrose instead of glucose for the PPP), significantly different isotopomer measurements (*e.g.* certain intracellular metabolites not considered by us) or techniques (*e.g.* tandem MS) should repeat our analyses with appropriate changes to the model. Furthermore, the choice of labeled substrates, usually the largest contributing factor to flux identifiability, is heavily influenced by the costs of the substrates. The results of our study are not cost-sensitive since changes in prices are dependent on factors beyond the control of investigators. Therefore, we have examined ILE designs involving both commercially available ''catalog'' labels and exotic, expensive labels that may be available only through custom synthesis. Ultimately, a balance of the aforementioned factors will enable an end-user to select an appropriate ILE. Currently, we are employing the label designs proposed in this study to investigate metabolic fluxes in *Arabidopsis thaliana* and poplar cell suspensions.

Designing ILEs is a rigorous computational process due to the variety of available labels and measurement possibilities that need to be optimized. Nevertheless, it offers valuable insights into performing an efficient ILE and ensures that maximum information is gained from the ILE. In the future, this work can be advanced by making use of recently reported improvements in isotope MFA on the analytical and computational fronts. On the analytical front, it is necessary to expand the spectrum of intracellular metabolites whose labeling is analyzed as well as to use high resolution instruments such as liquid chromatography-MS.[76] In this context, Antoniewicz and co-workers' tandem MS methodology[73,77] and novel NMR methods are likely to enable the measurement of a much larger subset of isotopomers than is currently possible. On the computational front, it is essential to use optimization algorithms to efficiently probe the multidimensional space of all available isotope labels for a given ILE. Recently, Stephanopoulos and co-workers applied a genetic algorithm,[31] while Palsson and co-workers applied Monte Carlo sampling toward this purpose.[78] Furthermore, Antoniewicz and co-workers have pioneered an EMU-based technique that rationally deduces the optimal labels for an ILE by tracing the number of different ways a product

isotopomer can be synthesized from given substrate isotopomers. Current implementations of this technique[32,71] have focused on illustrative or irreversible networks; in the future this technique may provide significant insights on complex networks such as the ones explored in this article. Finally, methods that integrate other omics studies such as transcriptomics and proteomics with MFA can augment and enhance the flux information available from isotope MFA.

Author contributions

SN and GS conceived this study and jointly made improvements to the computer program NMR2Flux+ to enable the necessary computation. SN performed the computation, wrote substantial portions of the manuscript and drafted a revised version after review. GS wrote and critically edited the manuscript. Both authors read and approved the final version of the manuscript prior to submission.

Acknowledgements

This work was partially funded by the National Science Foundation (award number IOS 0922650) as well as Department of Chemical and Biomolecular Engineering, University of Maryland and A. James Clark School of Engineering, University of Maryland (faculty startup grant to GS).

References

1 G. Stephanopoulos and D. E. Stafford, Metabolic engineering: a new frontier of chemical reaction engineering, *Chem. Eng. Sci.*, 2002, **57**, 2595–2602.
2 G. Stephanopoulos, *AIChE J.*, 2002, **48**, 920–926.
3 L. J. Sweetlove, R. L. Last and A. R. Fernie, Predictive metabolic engineering: a goal for systems biology, *Plant Physiol.*, 2003, **132**, 420–425.
4 L. Sweetlove, D. Fell and A. Fernie, Getting to grips with the plant metabolic network, *Biochem. J.*, 2008, **409**, 27–41.
5 J. Schwender, J. B. Ohlrogge and Y. Shachar-Hill, A flux model of glycolysis and the oxidative pentose phosphate pathway in developing *Brassica napus* embryos, *J. Biol. Chem.*, 2003, **278**, 29442–29453.
6 J. Schwender, J. Ohlrogge and Y. Shachar-Hill, Understanding flux in plant metabolic networks, *Curr. Opin. Plant Biol.*, 2004, **7**, 309–317.
7 A. R. Fernie, P. Geigenberger and M. Stitt, Flux an important, but neglected, component of functional genomics, *Curr. Opin. Plant Biol.*, 2005, **8**, 174–182.
8 J. Schwender, Metabolic flux analysis as a tool in metabolic engineering of plants, *Curr. Opin. Biotechnol.*, 2008, **19**, 131–137.
9 G. Sriram, *et al.*, Quantification of compartmented metabolic fluxes in developing soybean embryos by employing biosynthetically directed fractional ^{13}C labeling, two-dimensional [^{13}C, ^{1}H] nuclear magnetic resonance, and comprehensive isotopomer balancing, *Plant Physiol.*, 2004, **136**, 3043–3057.
10 G. Sriram, D. B. Fulton and J. V. Shanks, Flux quantification in central carbon metabolism of *Catharanthus roseus* hairy roots by ^{13}C labeling and comprehensive bondomer balancing, *Phytochemistry*, 2007, **68**, 2243–2257.
11 D. Rontein, M. Dieuaide-Noubhani, E. J. Dufourc, P. Raymond and D. Rolin, The metabolic architecture of plant cells. Stability of central metabolism and flexibility of anabolic pathways during the growth cycle of tomato cells, *J. Biol. Chem.*, 2002, **277**, 43948–43960.
12 J. Schwender, Y. Shachar-Hill and J. B. Ohlrogge, Mitochondrial metabolism in developing embryos of *Brassica napus*, *J. Biol. Chem.*, 2006, **281**, 34040–34047.
13 I. G. L. Libourel and Y. Shachar-Hill, Metabolic flux analysis in plants: From intelligent design to rational engineering, *Ann. Rev. Plant Biol.*, 2008, **59**, 625–650.
14 C. J. Baxter, *et al.*, The metabolic response of heterotrophic Arabidopsis cells to oxidative stress, *Plant Physiol.*, 2007, **143**, 312–325.
15 A. P. Alonso, F. D. Goffman, J. B. Ohlrogge and Y. Shachar-Hill, Carbon conversion efficiency and central metabolic fluxes in developing sunflower (Helianthus annuus L.) embryos, *Plant J.*, 2007, **52**, 296–308.
16 T. C. R. Williams, *et al.*, Metabolic network fluxes in heterotrophic Arabidopsis cells: Stability of the flux distribution under different oxygenation conditions, *Plant Physiol.*, 2008, **148**, 704–718.
17 D. K. Allen, I. G. Libourel and Y. Shachar-Hill, Metabolic flux analysis in plants: coping with complexity, *Plant, Cell Environ.*, 2009, **32**, 1241–1257.
18 D. K. Allen, J. B. Ohlrogge and Y. Shachar-Hill, The role of light in soybean seed filling metabolism, *Plant J.*, 2009, **58**, 220–234.
19 G. Stephanopoulos, Metabolic fluxes and metabolic engineering, *Metab. Eng.*, 1999, **1**, 1–11.
20 J. Schwender, M. Seemann, H. Lichtenthaler and M. Rohmer, Biosynthesis of isoprenoids (carotenoids, sterols, prenyl side-chains of chlorophylls and plastoquinone) via a novel pyruvate/glyceraldehyde 3-phosphate non-mevalonate pathway in the green alga *Scenedesmus obliquus*, *Biochem. J.*, 1996, **316**, 73–80.
21 J. Schwender, F. Goffman, J. B. Ohlrogge and Y. Shachar-Hill, Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds, *Nature*, 2004, **432**, 779–782.
22 G. Sriram, *et al.*, Global metabolic effects of glycerol kinase overexpression in rat hepatoma cells, *Mol. Genet. Metab.*, 2008, **93**, 145–159.
23 P. V. Minorsky, Achieving the in silico plant. Systems biology and the future of plant biological research, *Plant Physiol.*, 2003, **132**, 404–409.
24 R. G. Ratcliffe and Y. Shachar-Hill, Measuring multiple fluxes through plant metabolic networks, *Plant J.*, 2006, **45**, 490–511.
25 R. Rios-Estepa and B. M. Lange, Experimental and mathematical approaches to modeling plant metabolic networks, *Phytochemistry*, 2007, **68**, 2351–2374.
26 W. Wiechert, ^{13}C metabolic flux analysis, *Metab. Eng.*, 2001, **3**, 195–206.
27 M. R. Antoniewicz, J. K. Kelleher and G. Stephanopoulos, Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements, *Metab. Eng.*, 2006, **8**, 324–337.
28 M. Möllney, W. Wiechert, D. Kownatzki and A.A. de Graaf, Bidirectional reaction steps in metabolic networks: IV. Optimal design of isotopomer labeling experiments, *Biotechnol. Bioeng.*, 1999, **66**, 86–103.
29 I. G. L. Libourel, J. P. Gehan and Y. Shachar-Hill, Design of substrate label for steady state flux measurements in plant systems using the metabolic network of *Brassica napus* embryos, *Phytochemistry*, 2007, **68**, 2211–2221.
30 M. J. Araúzo-Bravo and K. Shimizu, An improved method for statistical analysis of metabolic flux analysis using isotopomer mapping matrices with analytical expressions, *J. Biotechnol.*, 2003, **105**, 117–133.
31 C. M. Metallo, J. L. Walther and G. Stephanopoulos, Evaluation of ^{13}C isotopic tracers for metabolic flux analysis in mammalian cells, *J. Biotechnol.*, 2009, **144**, 167–174.
32 S. B. Crown, W. S. Ahn and M. R. Antoniewicz, Rational design of ^{13}C-labeling experiments for metabolic flux analysis in mammalian cells, *BMC Syst. Biol.*, 2012, **6**, 43.
33 J. L. Walther, C. M. Metallo, J. Zhang and G. Stephanopoulos, Optimization of ^{13}C isotopic tracers for metabolic flux analysis in mammalian cells, *Metab. Eng.*, 2012, **14**, 162–171.
34 W. Wiechert, M. Mollney, S. Petersen and A. A. de Graaf, A universal framework for ^{13}C metabolic flux analysis, *Metab. Eng.*, 2001, **3**, 265–283.
35 N. J. Kruger and A. von Schaewen, The oxidative pentose phosphate pathway: structure and organisation, *Curr. Opin. Plant Biol.*, 2003, **6**, 236–246.
36 H.-W. Heldt, *Plant Biochemistry*, Academic Press, 2004.
37 M. Dieuaide-Noubhani, G. Raffard, P. Canioni, A. Pradet and P. Raymond, Quantification of compartmented metabolic fluxes in maize root tips using isotope distribution from ^{13}C- or ^{14}C-labeled glucose, *J. Biol. Chem.*, 1995, **270**, 13147–13159.

38 A. P. Alonso, V. L. Dale and Y. Shachar-Hill, Understanding fatty acid synthesis in developing maize embryos using metabolic flux analysis, *Metab. Eng.*, 2010, **12**, 488–497.
39 A. P. Alonso, D. L. Val and Y. Shachar-Hill, Central metabolic fluxes in the endosperm of developing maize seeds and their implications for metabolic engineering, *Metab. Eng.*, 2011, **13**, 96–107.
40 J. Krook, D. Vreugdenhil, C. Dijkema and L. van der Plas, Sucrose and starch metabolism in carrot (*Daucus carota* L.) cell suspensions analysed by 13-labelling: indications for a cytosol and a plastid-localized oxidative pentose phosphate pathway, *J. Exp. Bot.*, 1998, **49**, 1917–1924.
41 H. Usuda and G. E. Edwards, Localization of Glycerate Kinase and Some Enzymes for Sucrose Synthesis in C3 and C4 Plants 1, *Plant Physiol.*, 1980, **65**, 1017–1022.
42 S. Streb, B. Egli, S. Eicke and S. C. Zeeman, The Debate on the Pathway of Starch Synthesis: A Closer Look at Low-Starch Mutants Lacking Plastidial Phosphoglucomutase Supports the Chloroplast-Localized Pathway, *Plant Physiol.*, 2009, **151**, 1769–1772.
43 D. K. Allen, R. W. Laclair, J. B. Ohlrogge and Y. Shachar-Hill, Isotope labelling of Rubisco subunits provides *in vivo* information on subcellular biosynthesis and exchange of amino acids between compartments, *Plant, Cell Environ.*, 2012, **35**, 1232–1244.
44 S. K. Masakapalli, P. Le Lay, J. E. Huddleston, N. L. Pollock, N. J. Kruger and R. G. Ratcliffe, Subcellular flux analysis of central metabolism in a heterotrophic *Arabidopsis thaliana* cell suspension using steady-state stable isotope labeling, *Plant Physiol.*, 2010, **152**, 602–619.
45 A. Roscher, N. J. Kruger and R. G. Ratcliffe, Strategies for metabolic flux analysis in plants using isotope labelling, *J. Biotechnol.*, 2000, **77**, 81–102.
46 A. Fait, H. Fromm, D. Walter, G. Galili and A. R. Fernie, Highway or byway: the metabolic role of the GABA shunt in plants, *Trends Plant Sci.*, 2008, **13**, 14–19.
47 N. Bouche and H. Fromm, GABA in plants: just a metabolite? *Trends Plant Sci.*, 2004, **9**, 110–115.
48 A. W. Bown, D. E. Hall and K. B. MacGregor, Insect footsteps on leaves stimulate the accumulation of 4-aminobutyrate and can be visualized through increased chlorophyll fluorescence and superoxide production, *Plant Physiol.*, 2002, **129**, 1430–1434.
49 R. Palanivelu, L. Brass, A. F. Edlund and D. Preuss, Pollen tube growth and guidance is regulated by POP2, an Arabidopsis gene that controls GABA levels, *Cell*, 2003, **114**, 47–59.
50 B. J. Shelp, A. W. Bown and D. Faure, Extracellular {gamma}-aminobutyrate mediates communication between plants and other organisms, *Plant Physiol.*, 2006, **142**, 1350–1352.
51 B. J. Shelp, A. W. Bown and M. D. McLean, Metabolism and functions of gamma-aminobutyric acid, *Trends Plant Sci.*, 1999, **4**, 446–452.
52 V. V. Iyer, *et al.*, Metabolic flux maps comparing the effect of temperature on protein and oil biosynthesis in developing soybean cotyledons, *Plant, Cell Environ.*, 2008, **31**, 506–517.
53 W. Wiechert and A. A. de Graaf, Bidirectional reaction steps in metabolic networks: I. Modeling and simulation of carbon isotope labeling experiments, *Biotechnol. Bioeng.*, 1997, **55**, 101–117.
54 A. Masoudi-Nejad, S. Goto, T. R. Endo and M. Kanehisa, KEGG bioinformatics resource for plant genomics research, *Plant Bioinformatics*, 2008, **406**, 437–458.
55 W. A. van Winden, J. J. Heijnen, P. J. T. Verheijen and J. Grievink, A priori analysis of metabolic flux identifiability from ^{13}C-labeling data, *Biotechnol. Bioeng.*, 2001, **74**, 505–516.
56 S. M. Clark, *et al.*, Biochemical characterization, mitochondrial localization, expression, and potential functions for an Arabidopsis {gamma}-aminobutyrate transaminase that utilizes both pyruvate and glyoxylate, *J. Exp. Bot.*, 2009, **60**, 1743–1757.
57 W. Wiechert, M. Möllney, N. Isermann, M. Wurzel and A. A. de Graaf, Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems, *Biotechnol. Bioeng.*, 1999, **66**, 69–85.
58 B. Christensen and J. Nielsen, Isotopomer analysis using GC-MS, *Metab. Eng.*, 1999, **1**, 282–290.
59 B. K. Singh, *Plant Amino Acids (Books in Soils, Plants, & the Environment)*, CRC, 1998.
60 T. Huang, *et al.*, Analysis and Prediction of the Metabolic Stability of Proteins Based on Their Sequential Features, Subcellular Locations and Interaction Networks, *PLoS ONE*, 2010, **5**, e10972.
61 Z.-C. Wu, X. Xiao and K.-C. Chou, iLoc-Plant: a multi-label classifier for predicting the subcellular localization of plant proteins with both single and multiple sites, *Mol. Biosyst.*, 2011, **7**, 3287–3297.
62 K.-C. Chou and H.-B. Shen, Plant-mPLoc: A Top-Down Strategy to Augment the Power for Predicting Plant Protein Subcellular Localization, *PLoS ONE*, 2010, **5**, e11335.
63 S. Briesemeister, J. Rahnenführer and O. Kohlbacher, Going from where to why—interpretable prediction of protein subcellular localization, *Bioinformatics*, 2010, **26**, 1232–1238.
64 O. Emanuelsson, H. Nielsen, S. Brunak and G. von Heijne, Predicting Subcellular Localization of Proteins Based on their N-terminal Amino Acid Sequence, *J. Mol. Biol.*, 2000, **300**, 1005–1016.
65 S.-M. Chi and D. Nam, WegoLoc: accurate prediction of protein subcellular localization using weighted Gene Ontology terms, *Bioinformatics*, 2012, **28**, 1028–1030.
66 W. Wiechert, C. Siefke, A. A. de Graaf and A. Marx, Bidirectional reaction steps in metabolic networks: II. Flux estimation and statistical analysis, *Biotechnol. Bioeng.*, 1997, **55**, 118–135.
67 W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, 1992.
68 M. Locatelli, *Simulated annealing algorithms for continuous global optimization. Handbook of Global Optimization*, Springer, 2002.
69 T. Szyperski, ^{13}C-NMR, MS and metabolic flux balancing in biotechnology research, *Q. Rev. Biophys.*, 1998, **31**, 41–106.
70 T. Szyperski, Biosynthetically directed fractional ^{13}C-labeling of proteinogenic amino acids. An efficient analytical tool to investigate intermediary metabolism, *Eur. J. Biochem.*, 1995, **232**, 433–448.
71 S. B. Crown and M. R. Antoniewicz, Selection of tracers for ^{13}C-Metabolic Flux Analysis using Elementary Metabolite Units (EMU) basis vector methodology, *Metab. Eng.*, 2012, **14**, 150–161.
72 G. Sriram, V. V. Iyer, D. B. Fulton and J. V. Shanks, Identification of hexose hydrolysis products in metabolic flux analytes: A case study of levulinic acid in plant protein hydrolysate, *Metab. Eng.*, 2007, **9**, 442–451.
73 J. Choi and M. R. Antoniewicz, Tandem mass spectrometry: A novel approach for metabolic flux analysis, *Metab. Eng.*, 2011, **13**, 225–233.
74 J. Choi, M. T. Grossbach and M. R. Antoniewicz, Measuring Complete Isotopomer Distribution of Aspartate Using Gas Chromatography/Tandem Mass Spectrometry, *Anal. Chem.*, 2012, **84**, 4628–4632.
75 K. Schmidt, J. Nielsen and J. Villadsen, Quantitative analysis of metabolic fluxes in Escherichia coli, using two-dimensional NMR spectroscopy and complete isotopomer models, *J.l Biotechnol.*, 1999, **71**, 175–189.
76 R. J. Kleijn, *et al.*, Metabolic flux analysis of a glycerol-overproducing Saccharomyces cerevisiae strain based on GC-MS, LC-MS and NMR-derived ^{13}C-labelling data, *FEMS Yeast Res.*, 2007, **7**, 216–231.
77 M. R. Antoniewicz, J. K. Kelleher and G. Stephanopoulos, Accurate Assessment of Amino Acid Mass Isotopomer Distributions for Metabolic Flux Analysis, *Anal. Chem.*, 2007, **79**, 7554–7559.
78 J. Schellenberger, *et al.*, Predicting outcomes of steady-state ^{13}C isotope tracing experiments using Monte Carlo sampling, *BMC Syst. Biol.*, 2012, **6**, 9.

