the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2025) **123,** e70333 doi: 10.1111/tpj.70333

TECHNICAL ADVANCE

Triple labeling of metabolites for metabolome analysis (TLEMMA): a stable isotope labeling approach for metabolite identification and network reconstruction

Yonghui Dong[1,2,†], Liron Feldberg[3,†], Yi Wang[4], Uwe Heinig[1], Ilana Rogachev[1] and Asaph Aharoni[1,*]

[1]*Department of Plant and Environmental Sciences, Weizmann Institute of Science, Rehovot, Israel,*
[2]*Haixia Institute of Science and Technology, Fujian Agriculture and Forestry University, 350002 Fuzhou, China,*
[3]*Department of Analytical Chemistry, Israel Institute for Biological Research, Ness Ziona, Israel, and*
[4]*Xi'an Manareco New Materials Co., Ltd., Shaanxi, China*

Received 20 January 2025; revised 29 May 2025; accepted 26 June 2025.
*For correspondence (e-mail asaph.aharoni@weizmann.ac.il).
†These authors contributed equally to this work.

SUMMARY

Metabolite identification remains a significant challenge in mass spectrometry (MS)-based metabolomics research. To address this issue, we combined a triple-labeled precursor-based isotope tracing approach (TLEMMA) with high-resolution liquid chromatography-MS for metabolite identification and metabolic network construction. As a demonstration, we fed duckweed (*Spirodela polyrhiza*) with four forms of phenylalanine (Phe) including unlabeled Phe, Phe-$^{5}H_2$, Phe-$^{8}H_2$, and Phe-$^{13}C_9{}^{15}N_1$. The distinctive isotopic pattern obtained from MS spectra greatly facilitated data processing, enabling comprehensive extraction of all Phe-derived metabolites. Importantly, the labeling pattern allowed efficient metabolite identification by significantly reducing the number of structural and positional isomers. Using this approach, 47 phenylalanine-derived metabolites were putatively identified. To further evaluate the efficiency of metabolite identification in relation to the number of differently labeled precursors used, we compared the number of filtered candidates based solely on the labeling patterns obtained from unlabeled, single, dual, and triple isotope-labeled precursor tracing experiments. On average, TLEMMA eliminates the number of false candidates by 99.1% compared with unlabeled samples, 95% compared with single isotope-labeled samples, and 66.7% compared with dual isotope-labeled samples. This significant reduction in the number of false positives, along with the ability to identify previously unreported metabolites, demonstrates the power of TLEMMA in advancing the field of metabolomics and metabolic network reconstruction.

Keywords: metabolite identification, plant metabolomics, stable isotope labeling, technical advance.

INTRODUCTION

Liquid chromatography-mass spectrometry (LC–MS) has become a leading technique in metabolomics owing to its high sensitivity, high throughput, and broad molecular coverage (Wang et al., 2015). However, metabolite annotation using LC–MS is intrinsically difficult and is considered one of the major bottlenecks in metabolomics research (Dong et al., 2024; Kim et al., 2006; Shahaf et al., 2016). Unlike proteins or peptides, metabolites cannot be sequenced by MS; instead, each individual metabolite must undergo explicit structural elucidation based on accurate mass-to-charge ratio (*m/z*), retention time (RT), and MS/MS information (Chaleckis et al., 2019).

The first step towards metabolite identification is to determine the molecular formula (Nagao et al., 2014). While ultra-high-resolution MS has significantly improved our capability to assign molecular formulas, high resolving power and mass accuracy alone are typically insufficient for unambiguous formula determination, particularly for high molecular weight metabolites (mass >500 Da) (Chaleckis et al., 2019; Kind & Fiehn, 2006). Even when the molecular formula is correctly assigned, structural

elucidation remains challenging due to the presence of numerous isomers, which are particularly frequent in plants (Chokkathukalam et al., 2014; Zhao & Tian, 2022). For a given molecular formula, depending on the number and type of elements in the chemical composition, several to billions of isomers can theoretically exist (Kind & Fiehn, 2006). For instance, a simple formula like $C_{100}H_{202}$ could have more than 5.9×10^{39} possible structural isomers (Paton & Goodman, 2007).

Stable isotope labeling (SIL) has significantly advanced metabolite identification (Dong et al., 2024; Mitchell et al., 2024). Stable isotopes (e.g., ^{2}H, ^{13}C, and ^{15}N) have the same number of protons and nearly identical physicochemical properties as their common elements (e.g., ^{1}H, ^{12}C, ^{14}N) but differ in masses due to variations in the number of neutrons (Dong et al., 2024; Freund & Hegeman, 2017; Yu et al., 2023). As such, the mass shift between co-eluting labeled and unlabeled metabolites can be used to infer the number of labeled atoms, facilitating molecular formula assignment. Notably, the use of multi-isotope labeling such as the combination of uniform ^{13}C, ^{15}N, and ^{34}S labeling dramatically reduces the number of possible molecular formulas by providing additional constraints on the number of C, N, and S atoms in the formula (Giavalisco et al., 2011; Hegeman et al., 2007).

In addition to determining molecular formulae, SIL can also assist with metabolite structure elucidation. Two primary labeling approaches are widely employed: uniform labeling and partial positional labeling, each offering distinct advantages for metabolite identification. Uniform labeling involves substituting all atoms of a specific element (e.g., carbon) with stable isotopes (e.g., ^{13}C) in a precursor. This approach is particularly valuable for high-throughput screening, as it probes all potential metabolic outcomes of a labeled precursor in parallel (Chokkathukalam et al., 2014; Fernández-García et al., 2020; Nakabayashi & Saito, 2020; Simpson, Olson, et al., 2021; Simpson, Wunderlich, et al., 2021). For example, to elucidate quinone reaction products in wine, researchers spiked white and red wine samples with an equimolar mixture of unlabeled and uniformly ^{13}C-labeled quinones. By analyzing the characteristic M + n6 isotopologue pattern (reflecting six labeled carbons per quinone molecule), they identified 225 quinone-derived metabolites (Ji et al., 2022). Alternatively, partial positional labeling, where specific atoms within a precursor are selectively substituted with stable isotopes, provides a more powerful approach for metabolite structure elucidation, as it leverages prior knowledge of the precursor's structure to narrow the search space for derived metabolites (Dong et al., 2024). Recent applications of this approach include using ring-[$^{13C}_{6}$]-labeled phenylalanine to identify phenylpropanoids in Arabidopsis (Simpson, Olson, et al., 2021; Simpson, Wunderlich, et al., 2021) and employing both ring-[$^{13}C_6$]-labeled phenylalanine and tyrosine to detect phenylalanine-derived and tyrosine-derived metabolites in sorghum tissues (Simpson, Olson, et al., 2021; Simpson, Wunderlich, et al., 2021; Tuinstra et al., 2022). To improve "structural resolution," distinct labeling schemes can be applied to different substructures of a precursor. This strategy provides unambiguous "fingerprints" for precursor-derived metabolites, thus significantly improving confidence in identification. A notable implementation is Dual Labeling of Metabolites for Metabolome Analysis (DLEMMA). In a proof-of-concept study, *Arabidopsis* leaves were fed with two forms of differently labeled tryptophan—[Trp-$^{2}H_5$ (all five hydrogen atoms attached to the indolic ring were labeled) and Trp-$^{13}C_{11}{}^{15}N_2$ (all carbon and nitrogen atoms were labeled)]. This dual-labeling strategy identified 17 tryptophan-derived metabolites, demonstrating its capability in resolving structure isomers (Feldberg et al., 2009). In addition, this approach has been extended to mass spectrometry imaging for the identification of spatially localized tyrosine-derived metabolites (Feldberg et al., 2018).

Despite its great potential in structure identification, SIL is primarily used to determine molecular formula. Moreover, the extent to which SIL can be applied for metabolite structural elucidation has not yet been rigorously explored. In this study, we developed a triple-labeled precursor-based isotope tracing approach for high-confidence metabolite identification and metabolic network construction. In addition, we examined and compared the power of single-, dual-, and triple-labeled precursor-based isotope tracing for metabolite identification. A comprehensive and detailed description of the analytical and data processing workflow is provided to facilitate the application of this approach to different precursors and organisms of interest.

RESULTS AND DISCUSSION

The Triple Labeling of Metabolites for Metabolome Analysis concept

We describe an efficient analytical approach, namely Triple Labeling of Metabolites for Metabolome Analysis (TLEMMA) for metabolite identification. The workflow of this approach is outlined in Figure 1, in which TLEMMA was applied to study the metabolic pathway of the aromatic amino acid phenylalanine (Phe) precursor in duckweed (*Spirodela polyrhiza*). Five groups of independently grown duckweed were fed with unlabeled and differently labeled phenylalanine, including (1) control: samples without feeding; (2) samples fed with unlabeled phenylalanine; (3) samples fed with labeled I phenylalanine (Phe^{5}-H_2); (4) samples fed with labeled II phenylalanine (Phe^{8}-H_2); and (5) samples fed with a labeled III phenylalanine (Phe-$^{13}C_9{}^{15}N_5$) (Figure 1a). The five groups were then extracted and analyzed, respectively, using high-resolution

LC–MS in both positive and negative ion modes (Figure 1b). Principal component analysis (PCA) was first applied to evaluate the group differences after isotope feeding. The PCA score plots demonstrate clear separation of the five groups in both positive and negative ion modes (Figure S1), indicating effective incorporation of labeled atoms. The selection and clustering of the precursor-derived metabolites were based on two criteria: (1) If a metabolite is of precursor origin, a mass shift between the co-eluting unlabeled and labeled peaks will be observed. If no mass shift is detected, this metabolite cannot be assigned as a precursor-derived metabolite. (2) If a metabolite is of precursor origin, it should only contain the same type of labeled atoms as its fed precursor. For instance, metabolites labeled with ^{13}C and/or ^{15}N can only be expected in Phe-$^{13}C_9{}^{15}N_5$ fed group, but not in Phe-$^{2}H_5$ or Phe-$^{2}H_8$ group. The above criteria are based on the isotopologue behavior as demonstrated in Figure 1c, that the same metabolite derived from the differently labeled fed precursors appears as four co-eluting peaks (one unlabeled and three differently labeled metabolites). Only the mass spectra indicate that the four samples are derived from different labeling regimes. The recently developed R package Miso has incorporated the two criteria for automated comprehensive detection of the isotopologue groups and efficient elimination of the false positives (Dong et al., 2019). Compared to the non-isotope-based metabolomics approach, TLEMMA allows not only discrimination of precursor-derived ions but also facilitates metabolite identification. The number of labeled atoms can be calculated by subtracting the integer mass of the unlabeled peak from the integer mass of the co-eluting labeled peak, therefore providing additional constraint for molecular formula determination. Most importantly, the different labeling patterns of the same metabolite derived from the different labeled precursors are of great value for metabolite structural elucidation. Once the phenylalanine-derived metabolites were identified, a detailed metabolic network can then be constructed based on their labeling patterns and known biochemical reactions (Figure 1d).

Isotope effect evaluation

It is generally accepted that stable isotope-labeled metabolites co-elute with their unlabeled forms in LC–MS runs. While this is typically the case for ^{13}C-labeled metabolites, deuterium (^{2}H) labeling can sometimes lead to a substantial retention time shift (Kato et al., 2000a; Triebl & Wenk, 2018), with deuterium-labeled molecules being eluted first. This phenomenon is commonly known as the isotope effect. The most likely explanation is that deuteriums have a stronger binding with carbon atoms than hydrogens, thus introducing differences in physiochemical properties (such as polarity) among isotopologues (Iyer et al., 2004; Triebl & Wenk, 2018; Wieling, 2002). It has been reported that the degree of retention time shift is determined by the number of deuterium atoms and the size of the labeled compound. Small compounds labeled with more deuterium atoms tend to have a more pronounced retention time shift (Kachlicki et al., 2016a). In addition, the pH of LC–MS eluents was also found to affect the retention time shift. For instance, when a neutral eluent was applied [mobile phase: ammonium acetate (20 mM, pH 7.0), methanol, acetonitrile], a retention time shift of about 1.2 min was observed between unlabeled and 10 deuterium-labeled compounds; while with an acidic solvent system [mobile phase: ammonium formate (5 mM, pH 3.1), acetonitrile], unlabeled and 10 deuterium-labeled compounds were found to co-elute (Kato et al., 2000b).

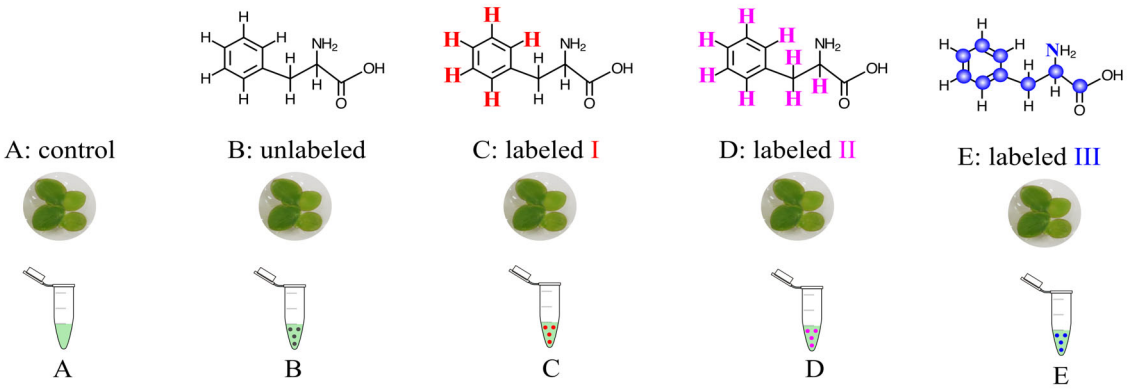
According to our design (Figure 1), up to 8 deuterium atoms could be introduced in phenylalanine-derived metabolites (polymers are not considered here as their formation is quite rare), which may lead to a significant retention time shift. As such, this shift must be carefully estimated and considered during data processing in order to select the complete precursor-derived metabolite and eliminate false positives. By inspecting the raw spectra, we have manually identified 5 isotopologue clusters in which the labeled metabolites from experiment group D (fed with Phe-$^{2}H_8$, Figure 1) were labeled with 8 deuterium atoms. The retention time shifts between the 8 deuterium-labeled metabolites and their respective unlabeled forms were calculated and plotted (Figure 2a). The representative LC chromatograms of metabolite *m/z* 293.1141 and its 8-deuterium-labeled form *m/z* 301.1632 were shown in Figure 2b. The maximum retention time shift is 0.13 min, demonstrating that the shift is negligible. The small retention time shift could be attributed to the fact that an acidic solvent system [mobile phase: formic acid (0.1%, pH ~2.7)-acetonitrile] was used in our LC–MS system.

Metabolite identification and network construction

Stable isotope tracing allows discrimination of precursor-derived MS peaks by comparing the mass shift between co-eluting peaks from unlabeled and labeled groups (Doppler et al., 2019; Yu et al., 2023). A mass shift of 0 indicates that the MS signal is not derived from the precursor. However, it is practically impossible to manually calculate the mass shifts of all co-eluting peaks and identify precursor-derived isotopologue clusters in highly complex biological samples. We have recently developed an R package, namely Miso, which automates the selection of the complete repertoire of isotopologue clusters (Dong et al., 2019). Although Miso was originally developed for dual-labeled precursor-based isotope tracing, it can also be directly applied to single- or multiple-labeled precursor-based isotope tracing experiments. Using Miso, 70 and 131

(a) Experiment Design

1 Stable Isotope Feeding : phenylalanine as precursor



(b) Data Acquisition

1 High resolution LC-MS measurements (postive and negative ion mode)

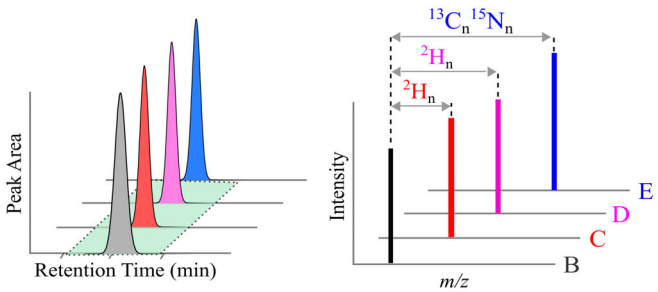
2 LC-MS/MS analysis (MS^e postive and negative ion mode)

(c) Data Processing

1 Pre-processing: Convert raw data into mzXML format (MSconvert)

2 Peak picking and Alignment: XCMS

3 MS signal filtering: Filter features according to retention time and m/z values (Miso)



4 Clustering: match feature groups over all samples, and generate isotopologue clusters (Miso)

Unlabeled			Labeled I				Labeled II				Labeled III			
m/z	Int.	RT	*m/z*	Int.	RT	Label	*m/z*	Int.	RT	Label	*m/z*	Int.	RT	Label
164.071	$5x10^6$	2.15	169.106	$5x10^6$	2.07	5	172.161	$4x10^6$	2.06	8	174.099	$8x10^6$	2.13	10
341.086	$2x10^5$	3.90	344.105	$7x10^5$	3.85	3	346.116	$4x10^5$	3.84	5	350.117	$1x10^6$	3.89	9
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

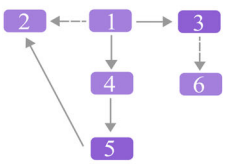
(d) Metabolite Identification and Network Construction

1 Molecular fomula determination

2 Substructure search according to labeling pattern

3 MS/MS filtering and confirmation

4 Metabolic network construction



each number represents one metabolite

Figure 1. Schematic representation of the TLEMMA approach.
(a) Duckweed plants were fed for 24 h with four forms of phenylalanine: unlabeled, labeled I (Phe-^{2}H5), labeled II (Phe-$^{2}H_8$), and labeled III (Phe-$^{13}C_9{}^{15}N_1$).
(b) The extracts were analyzed using high-resolution LC–MS in both positive and negative ion modes.
(c) The raw data were converted into mzML format and pre-processed with the XCMS R package. The resulting XCMS set object was then subjected to the Miso R package to detect isotopologue clusters.
(d) Metabolite were identified based on their accurate *m/z* values, labeling pattern, and MS/MS fragments. A metabolic network was then constructed using the identified metabolites.

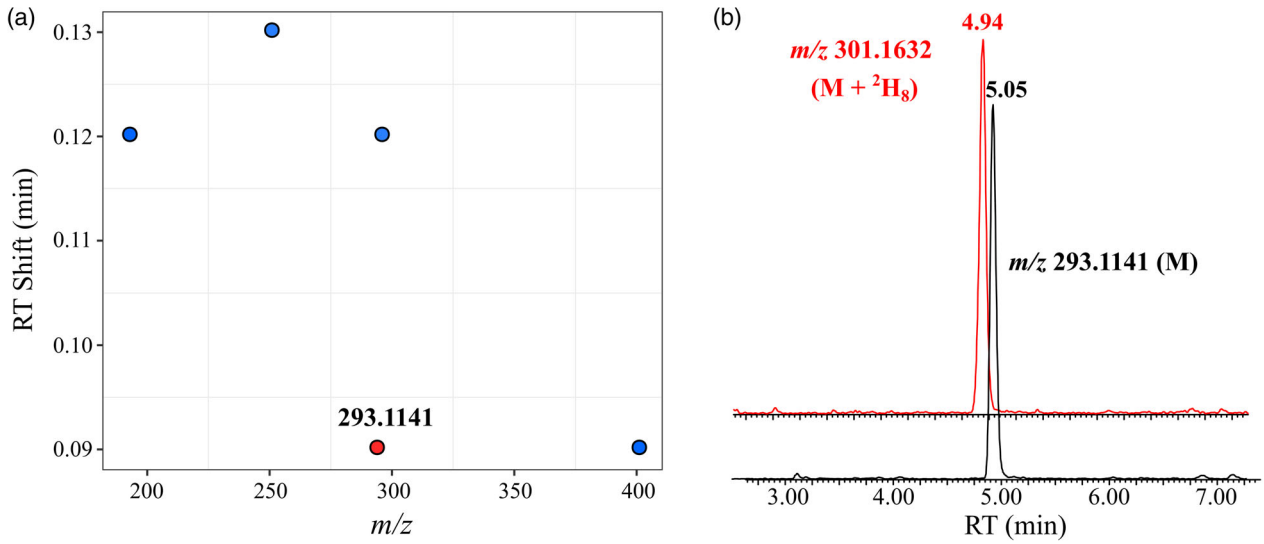


Figure 2. Evaluation of deuterium isotope effects.
(a) Five eight-deuterium-labeled metabolites (M + $^{2}H_8$) were manually identified from the label II phenylalanine-fed group (Phe-$^{2}H_8$). The retention time shifts between the eight-deuterium-labeled metabolites and their respective unlabeled forms (from the unlabeled phenylalanine-fed group) were calculated and plotted.
(b) Representative ion chromatograms of metabolite *m/z* 293.1141 and its eight-deuterium-labeled form *m/z* 301.1632.

isotopologue clusters were identified from positive and negative ion modes, respectively.

Next, each isotopologue cluster was identified according to its labeling patterns from TLEMMA analysis. The principle of TLEMMA-based metabolite identification is illustrated in Figure 3. One isotopologue cluster was detected in negative ion mode with the unlabeled and labeled *m/z* values being 293.1138 (Group B, unlabeled), 298.1437 (Group C, labeled I), 301.1638 (Group D, labeled II), and 303.1407 (Group E, labeled III), respectively (Figure 3a). This isotopologue cluster was also detected in the positive ion mode, which confirmed the labeling pattern in each experimental group (Figure 3a). The labeling pattern indicated that this metabolite was labeled with five ^{2}H from Group C, eight ^{2}H from Group D, and nine ^{13}C and one ^{15}N from E. The molecular formula of the unlabeled metabolite from this isotopologue cluster was first searched within a 5-ppm window using the i-FIT™ algorithm from Waters Masslynx software (version 4.1). When more than one molecular formula is found, the calculated number of labeled atoms based on the labeling pattern can help narrow down the number of the possible molecular formulas (e.g., the formula should contain at least eight H, nine C, and one N atoms in this case). In this example, only one molecular formula was found, and its neutral formula $C_{14}H_{18}N_2O_5$ was used for the database search. This formula returned 9021 possible candidates in SciFinder (https://scifinder.cas.org). As the metabolite was labeled with five deuterium atoms in experiment group C, it must contain an aromatic ring with five hydrogen atoms attached. This substructure-based searching, using the structure information from experiment group C, reduced the number of candidates from 9021 to 642. Similarly, incorporating the structure information from groups D and E reduced the candidate list to 136 and 254, respectively. When the structure information from groups C, D, and E were combined, the number of candidates decreased significantly to 93, 32 of which had been reported in biological samples. Based on MS/MS data, this metabolite was ultimately identified as gamma-glutamylphenylalanine (Figure 3a).

By applying the same principle, TLEMMA can also be used to elucidate the structures of MS/MS fragments, which can then be used to assist in metabolite

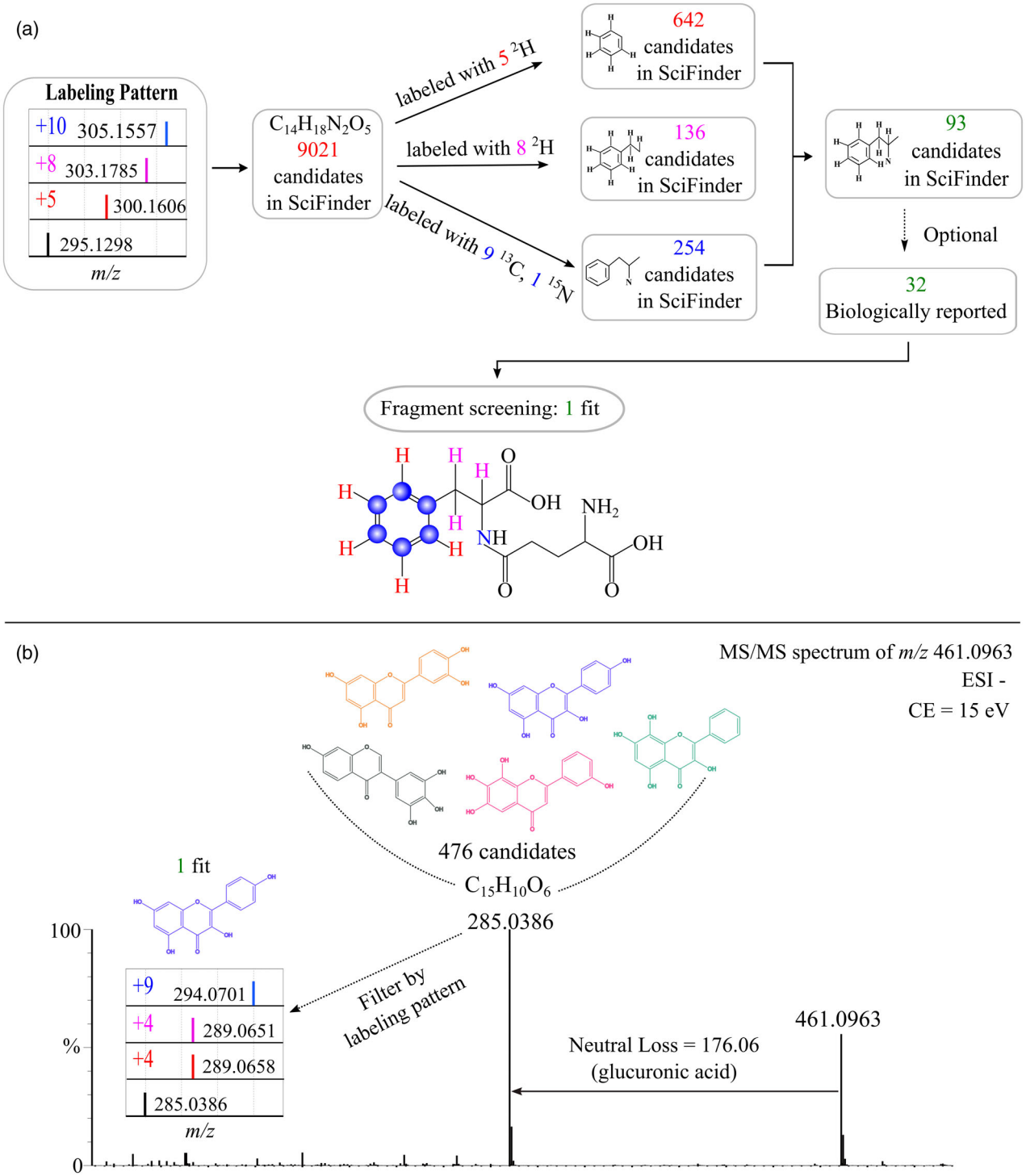


Figure 3. Representative workflow for metabolite identification using TLEMMA.
(a) Elemental composition is first determined based on accurate *m/z* value and natural isotope pattern from the unlabeled group. Next, sub-structural information inferred from the labeling pattern of each labeled group is used to narrow down the number of candidates in a database (i.e., SciFinder). MS/MS data are then used to further refine the structure and screen the candidates.
(b) In addition to identifying the precursor structure, TLEMMA can also be applied to infer the structures of the fragments, which are then used to elucidate the structure of the parent metabolite. For example, the fragment ion *m/z* 285.0495 is identified as kaempferol using TLEMMA, and considering the neutral loss, its parent metabolite is identified as kaempferol glucuronide.

identification. For instance, fragmentation trees of gamma-glutamylphenylalanine in both positive and negative ion modes were constructed using these identified MS fragments (Figure S2). In addition, comparing the MS/MS spectra of labeled and unlabeled counterparts provides valuable information for metabolite structural elucidation. Common fragments correspond to the unlabeled portions of the molecule, while differences in revealed the labeled portions. This approach is particularly useful for identifying metabolites with limited fragments. For instance, an isotopologue cluster was detected in the negative ion mode, with the unlabeled *m/z* value being 461.0963 (Figure 3b). Only one major fragment was observed for this metabolite, with *m/z* values being 285.0386 (Group B, unlabeled), 289.0658 (Group C, labeled I), 289.0651 (Group D, labeled II) and 294.0701 (Group D, labeled III), respectively. A molecular formula search of this fragment in SciFinder revealed 476 candidates, including several common flavones reported in plants, such as kaempferol, luteolin, fisetin, isoscutellarein, and orobol. As this ion was not further fragmented, we were unable to identify it solely based on its *m/z* value. However, with the assistance of the labeling patterns, this fragment was identified as kaempferol. The common neutral loss of 176.06 indicates the loss of a glucuronic acid from the parent metabolite (Figure 3b). Therefore, the parent metabolite was ultimately identified as kaempferol glucuronide (Level 2 identification).

Using TLEMMA, 47 phenylalanine-derived metabolites were putatively identified in duckweed (Table S1). Some of these metabolites were used for metabolic network construction, providing insights into the associated metabolic pathways (Figure 4).

Discrimination of positional isomers

The biological activities of many phenylalanine-derived metabolites are often regulated by the conjugation of sugar moieties (Dixon & Paiva, 1995). The sugar moieties can be located at different positions on the carbon skeleton, forming positional isomers. Determining the exact position of the linkage and distinguishing positional isomers are generally challenging using conventional MS-based approaches, as positional isomers often exhibit nearly identical fragmentation patterns (Kachlicki et al., 2016b; Yang et al., 2016). For the same reason, stable isotope tracing-based approaches are not always able to determine the definitive chemical structures of the query compounds. Nevertheless, stable isotope tracing can at least eliminate some false positional isomers based on the labeling pattern information, as demonstrated in the example shown in Figure 5. An isotopologue cluster with the labeling pattern of M + $^{2}H_3$ (Group C), M + $^{2}H_5$ (Group D), and M + $^{13}C_9$ (Group E), and an unlabeled *m/z* 355.1035 at the negative ion mode, was detected. The metabolite was identified as ferulic acid hexose. The labeling pattern suggests that the hexose can be only attached at the C4 position (Figure 5a), since the labeling pattern would change if the hexose were attached at any other positions. For instance, if the hexose were attached at the C8 position, the labeling pattern would shift to M + $^{2}H_3$ (Group C), M + $^{2}H_4$ (Group D), and M + $^{13}C_9$ (Group E), which does not match the experimental result (Figure 5b).

In addition to sugar moieties, positions of other substituents can also be restricted based on information deduced from the labeling pattern. However, it is important to note that the ability to discriminate positional isomers is limited to those substituents whose positions, when changed, alter the labeling pattern.

Comparing the power of metabolite identification by the use of single-, dual-, or triple-labeled precursor-based isotope tracing

To understand the power of metabolite identification associated with the number of differently labeled precursors, we compared the number of candidates filtered by the labeling patterns from single-, dual-, and triple-labeled precursor-based isotope tracing. According to our experiment design (Figure 1), there are three single-labeled precursor-based isotope tracing (SLEMMA) groups: Group C, Group D, and Group E; three dual-labeled precursor-based isotope tracing (DLEMMA) groups: that is, Group C + Group D, Group C + Group E, and Group D + Group E; and one triple-labeled precursor-based isotope tracing group: Group C + Group D + Group E. However, substructure search based only on the labeling patterns from Group D can be very cumbersome. For example, 17 possible substructures can be deduced from the M + $^{2}H_5$ labeling pattern in Group D (Figure S3), and each must be searched individually against a database for metabolite structure identification. Therefore, only Group C and Group E were selected to represent SLEMMA, Group C + Group E to represent DLEMMA, and Group C + Group D + Group E to represent TLEMMA.

Three representative examples are shown in Figure 6. The number of metabolite candidates for a given molecular formula is significantly reduced when additional structural constraints are applied. For instance, a database search of the molecular formula $C_{10}H_{11}NO_3$ in SciFinder revealed 6767 compounds with this elemental composition. However, when structural constraints were applied, only 467, 341, 57, and 10 candidates remained for S TEMMA1, SLEMMA2, DLEMMA, and TLEMMA, respectively (Figure 6a). Similar results were observed for other examples (Figure 6b,c). For simplicity, stereoisomers were counted as filtered candidates but were considered the same metabolite during subsequent identification. To generalize these findings, we calculated the mean number of selected candidates across all studied metabolites, as shown in Figure 6d. On average, 78.8, 93.1, and 96.6% of

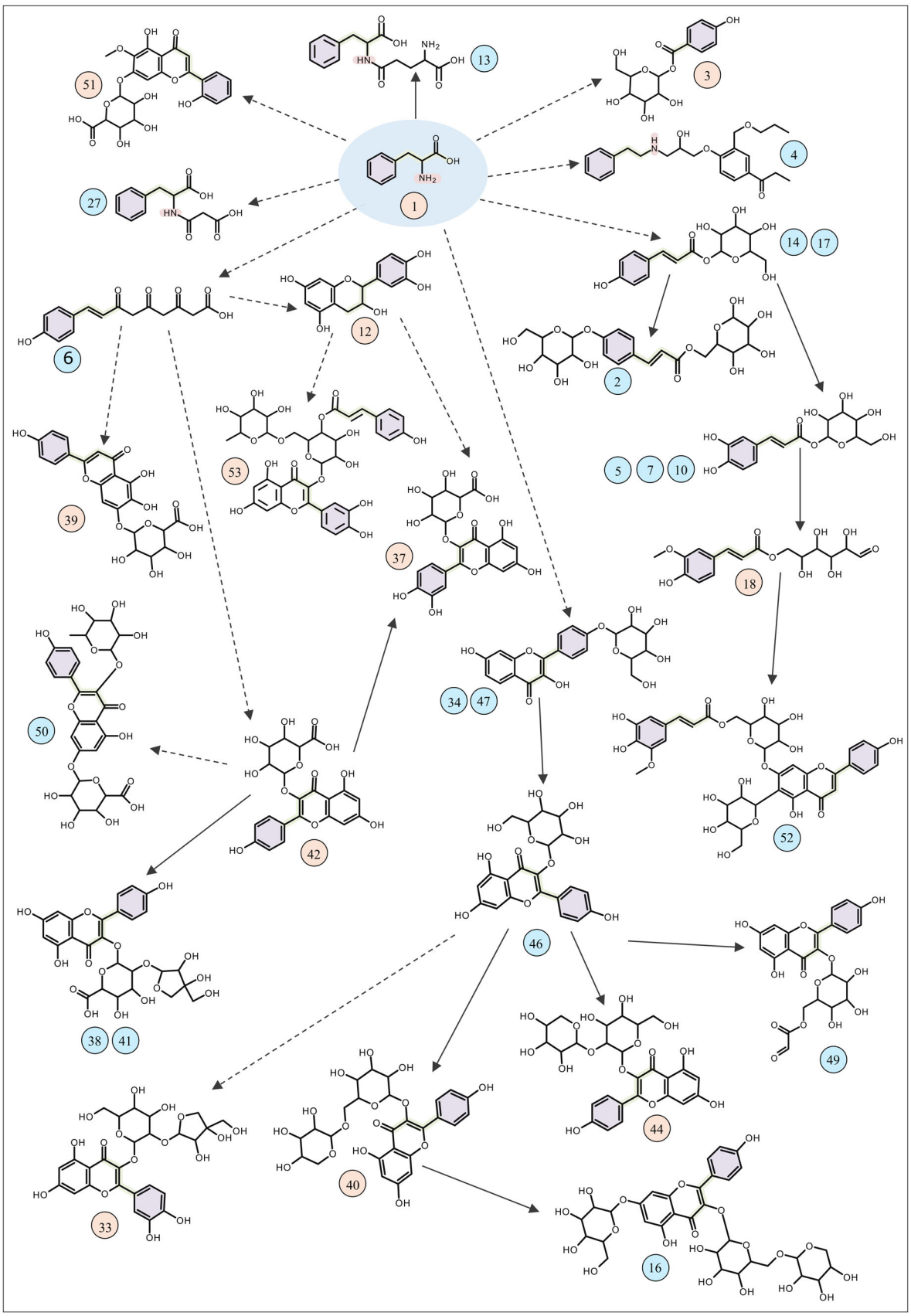


Figure 4. Putative phenylalanine-derived metabolic network in duckweed.
The pathway was constructed based on possible biochemical reactions. Dashed and solid arrows represent multiple and single biochemical reactions in the pathway, respectively. The metabolite numbers correspond to those in Table S1. Metabolites identified with Level 1 identification are indicated by numbers with a pink background, while metabolites identified at other levels are marked with a blue background.

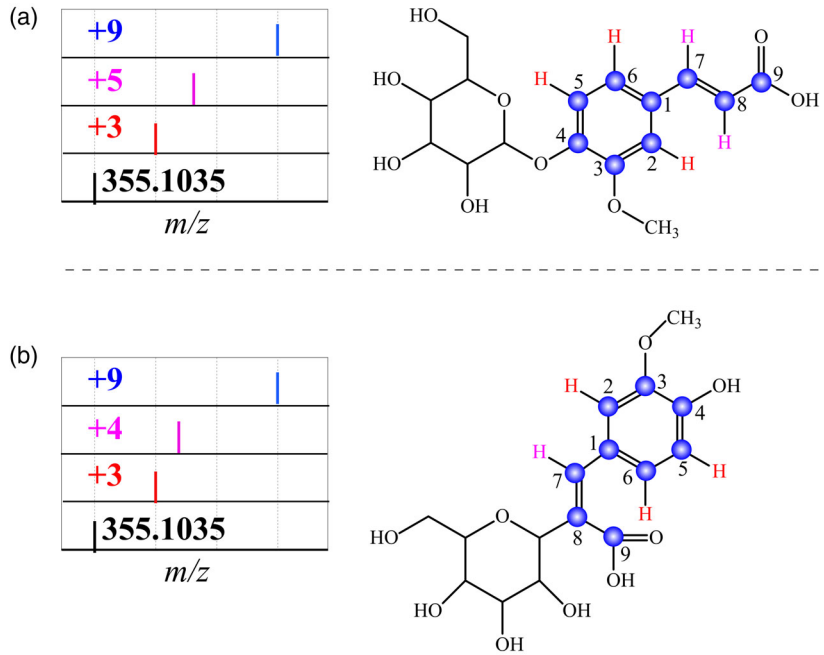


Figure 5. Representative example of using triple-labeled precursor-based isotope tracing for the identification of positional isomers.
An isotopic cluster with the labeling pattern of M + $^{3}H_2$, M + $^{5}H_2$, and M + $^{13}C_9$, and an unlabeled *m/z* 355.1035 in negative ion mode was detected. The metabolite was identified as ferulic acid hexose. The labeling pattern indicates that the hexose is attached at the C4 position, as attachment at other positions (e.g., like C2, C5, C6, C7, or C8) would alter the labeling pattern (a). For example, attachment at the C8 position would result in a pattern of M + $^{3}H_2$, M + $^{4}H_2$, and M + $^{13}C_9$, which does not match the experimental result (b).

the candidates were eliminated using structural constraints from SLEMMA, DLEMMA, and TLEMMA, respectively. Notably, TLEMMA was able to further eliminate an additional 83.7 and 50% of the false candidates from SLEMMA and DLEMMA, respectively.

It is critical to emphasize that the efficacy of metabolite identification using this approach does not strictly scale with the number of labeled precursors; instead, it hinges on the strategic design of labeling patterns. While prior knowledge of metabolic pathways can guide substructure selection, this approach is not limited to known metabolic modifications. The key lies in optimizing both precursor selection and labeling strategy. Precursor selection should align with the biological question in order to maximize the likelihood of detecting biologically relevant metabolites (Fernández-García et al., 2020). In terms of labeling, it is crucial to maximize the number of substructure coverages so that researchers can detect both anticipated and novel precursor-derived metabolites by tracing isotopic "fingerprints" and improve metabolite identification confidence by correlating the "fingerprints" with specific substructure information.

While effective in narrowing down possible structural isomers (Figure 6), TLEMMA alone is not sufficient for confident metabolite identification. Additional methods, such as spectral library matching, manual interpretation, and machine learning-based structural elucidation, are often needed to achieve high-confidence metabolite identification. Another limitation of this approach is its targeted nature, which confines identification to metabolites of specific pathways. Careful consideration must be given to the design of the labeling scheme. It is essential to ensure that the labeled precursor covers as many substructures as possible, while also allowing for clear distinction between different substructures through distinct labeling patterns. In addition, the precursor is suggested to be labeled with at least 4 atoms to avoid the disturbance of natural isotopic peaks. During precursor feeding, each sample group should be supplied with only one type of labeled precursor. This helps prevent the

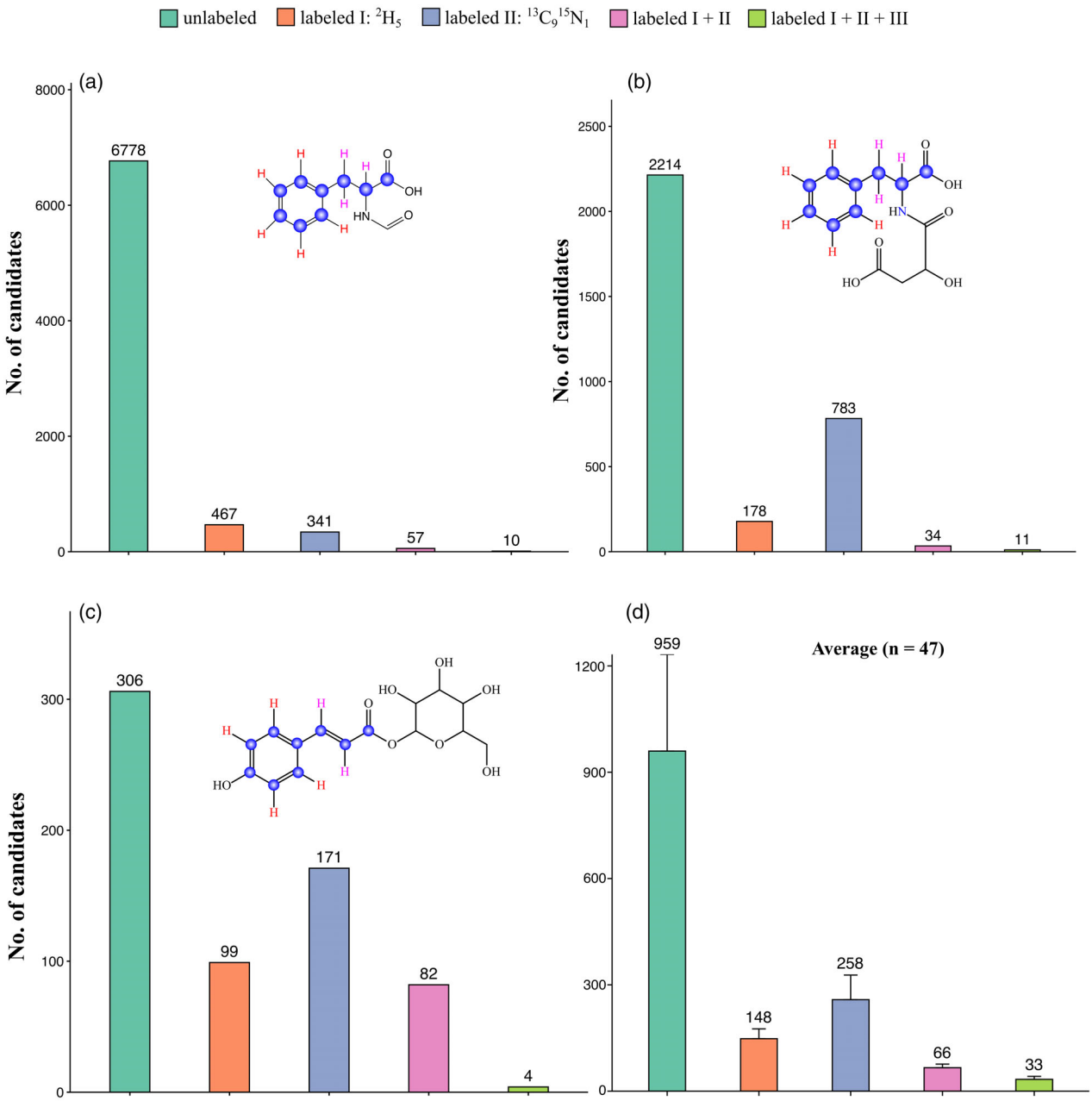


Figure 6. Evaluation of the power of metabolite identification by comparing the number of filtered candidates based on the labeling patterns of single-, dual-, and triple-labeled precursor-based isotope tracing.
(a–c) Three representative examples demonstrate the number of filtered candidates based on single-, dual-, and triple-labeled precursor-based isotope tracing.
(d) The average number of filtered candidates from 53 isotopologue clusters, based on single-, dual-, and triple-labeled precursor-based isotope tracing. The bar plot represents mean + SD.

formation of metabolic products derived from multiple precursor sources, which could complicate the mass spectra and hinder metabolite identification. Labeling efficiency is another critical factor, as it largely determines the labeling depth and the number of precursor-derived metabolites. It is suggested to optimize precursor concentration and feeding duration with unlabeled precursor to ensure sufficient labeling depth.

CONCLUSION

Metabolite identification is a major bottleneck in MS-based metabolomic studies. Stable isotope labeling has long been used as a key tool for unambiguous molecular formula determination and, more recently, for structural elucidation. Here, we present a TLEMMA for efficient metabolite identification. This approach not only narrows

down the number of molecular formulas but also significantly reduces the number of structural isomers for a given molecular formula. In some cases, it is also able to constrain the number of positional isomers. As a proof of concept, we applied this approach by feeding duckweed with three differently labeled forms of phenylalanine as tracing precursors: Phe-$^{2}H_5$, Phe-$^{2}H_8$, and Phe-$^{13}C_9{}^{15}N_1$. Using this method, 47 phenylalanine-derived metabolites were identified. To explore the power of metabolite identification associated with the number of differently labeled precursors, we compare the number of filtered candidates based solely on the labeling patterns from unlabeled, single-, dual-, and triple-isotope-labeled precursor tracing experiments. On average, TLEMMA eliminates 96.6, 83.7, and 50% of the false candidates from unlabeled, SLEMMA, and DLEMMA, respectively.

EXPERIMENTAL PROCEDURES

Materials

Phenylalanine (>98%) was purchased from Sigma-Aldrich (Rehovot, Israel), and phenylalanine-$^{2}H_5$ (^{2}H >98%), phenylalanine-$^{2}H_8$ (^{2}H >98%), and phenylalanine-$^{13}C9^{15}N_1$ (^{13}C >99%, ^{15}N >99%) were from Cambridge Isotope Laboratories (Andover, MA, USA). All other reagents were purchased from Sigma-Aldrich Chemicals.

Duckweed fronds (*Spirodela polyrhiza*) were propagated at 26°C, 30–50 μmol m^{-2} s^{-1} cool white-fluorescent light, and 5% CO_2 in 250 ml Erlenmeyer flasks containing 80 ml half-strength Hutner's solution (SP medium). Twenty-five plants were precultivated for 4 weeks, and the medium was replaced with fresh medium each week. Two-frond colonies, obtained from this precultivation, were fed with a phenylalanine precursor for 24 hours. The feeding experiment consisted of 6 groups, with each group containing 5 two-frond colonies: Group A (SP medium), Group B (5 mg ml^{-1}, unlabeled Phe), Group C (5 mg ml^{-1}, Phe-$^{2}H_5$, Label I), Group D (5 mg ml^{-1}, Phe-$^{2}H_8$, Label II), and Group E (5 mg ml^{-1}, Phe-$^{13}C_9{}^{15}N_1$, Label III) (Figure 1).

Duckweed fronds were gently washed twice in Milli Q water and dried with Kimwipes. Subsequently, they were snap-frozen in liquid nitrogen and ground into fine powder using an analytical mill (IKA; A11 basic). Next, 100 mg of the frozen fine powder was extracted with 80% methanol:water (v/v) containing 0.1% formic acid. The mixture was briefly vortexed and then sonicated for 20 min. Finally, the extracts were centrifuged for 15 min at 14 000 ***g*** and filtered through a 0.22 μm PTFE filter (Acrodisc® CR 13 mm; PALL Life Sciences, Germany) prior to LC–MS analysis.

LC–MS analysis

Chromatographic experiments were performed on an ultra-high-performance liquid chromatography with quadrupole time-of-flight mass spectrometry (UPLC-QTOF-MS) system (HDMS Synapt, Waters, UK). Separation was performed on a Waters ACQUITY UPLC BEH C18 column (100 × 2.1 mm, 1.7 μm) with the column temperature set at 35°C. The mobile phase consisted of 0.1% formic acid in acetonitrile/water (5:95, v/v, phase A) and 0.1% formic acid (v/v) in acetonitrile (phase B). The LC method was as follows: the linear gradient was from 100 to 72% phase A over 22 min, from 72 to 0% phase A over 14 min, then held at 100% phase B for a further 2 min, and then returned to the initial conditions (100% phase A) for 0.5 min and conditioned at 100% phase A for 1.5 min. A divert valve (Rheodyne) excluded the first 1 min and last 2 min of gradient from the injection. Data acquisition was performed in the MS^E mode with an energy ramp that records accurate mass precursor and fragment ion information from every detectable component in a sample. The MS^E mode rapidly alternates between two functions: the first acquiring low-energy accurate mass precursor ion spectra and the second acquiring elevated-energy accurate mass fragment ion spectra. The collision energy was set to 4 eV for the low-energy function and to 10–30 eV ramp for the high-energy function in the positive ion mode (15–35 eV in the negative ion mode). Scan time for each function was set to 0.25 s. The MS^E mode with energy ramp was used to add as much of the specific fragmentation patterns of each metabolite as possible in one continuous run. LC–MS/MS experiments were run using the following settings: capillary spray at 1.0 kV; cone voltage at 27 eV. Argon was used as the collision gas for collision-induced dissociation (CID). In addition, a mixture of 14 standard metabolites, injected after every 10 samples, was used for quality control (Table S1).

Data processing and metabolite identification

The raw instrument data were first converted into mzML format using ProteoWizard MSConvert (Kessner et al., 2008). Peak picking was performed using XCMS (Smith et al., 2006) under the R environment. The resulting XCMS set objects were then directly subjected to the R package Miso (Dong et al., 2019) for isotopologue cluster detection. Metabolite identification was performed using the SciFinder Database (https://scifinder.cas.org). To decrease the number of false positives, we further searched those tentatively identified metabolites by matching accurate mass, retention time, and fragment information against our in-house database Weizmass (V2) (Shahaf et al., 2016). The metabolite identification confidence level for each identified metabolite in this study was assigned following the modified four-level classification scheme from the Metabolomics Standards Initiative (Sumner et al., 2007): Level 1. Validated identification based on reference standard, with retention time, accurate mass, and MS/MS matching. Level 2. Putative identification based on matching fragmentation data to metabolite MS/MS libraries. Level 3. Tentative structures match precursor m/z to a metabolite database. Level 4. Unique molecular formula determined based on isotope abundance distribution, charge state, and adduct ion determination.

ACKNOWLEDGEMENTS

This work was supported by the Israel Ministry of Science and Technology (grant number 3-14297). A.A. is the incumbent of the Peter J. Cohn Professorial Chair.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Principal component analysis (PCA) of LC-MS data obtained from five duckweed sample groups.

Figure S2. Manually constructed fragmentation trees of gamma-Glutamylphenylalanine.

Figure S3. Possible substructures deduced from Group D with the labeling pattern M + $^{2}H_{5}$.

Table S1. Phenylalanine dereived metaboliteds identied in Lemma using triple labeling.

REFERENCES

Chaleckis, R., **Meister, I.**, **Zhang, P.** & **Wheelock, C.E.** (2019) Challenges, progress and promises of metabolite annotation for LC–MS-based metabolomics. *Current Opinion in Biotechnology*, **55**, 44–50.

Chokkathukalam, A., **Kim, D.-H.**, **Barrett, M.P.**, **Breitling, R.** & **Creek, D.J.** (2014) Stable isotope-labeling studies in metabolomics: new insights into structure and dynamics of metabolic networks. *Bioanalysis*, **6**, 511–524.

Dixon, R.A. & **Paiva, N.L.** (1995) Stress-induced phenylpropanoid metabolism. *Plant Cell*, **7**, 1085–1097.

Dong, Y., **Feldberg, L.** & **Aharoni, A.** (2019) Miso: an R package for multiple isotope labeling assisted metabolomics data analysis. *Bioinformatics*, **35**, 3524–3526.

Dong, Y., **Feldberg, L.**, **Aharoni, A.** & **Heinig, U.** (2024) Metabolite annotation through stable isotope labeling. *TrAC Trends in Analytical Chemistry*, **181**, 118037.

Doppler, M., **Bueschl, C.**, **Kluger, B.**, **Koutnik, A.**, **Lemmens, M.**, **Buerstmayr, H.** ***et al.*** (2019) Stable isotope–assisted plant metabolomics: combination of global and tracer-based labeling for enhanced untargeted profiling and compound annotation. *Frontiers in Plant Science*, **10**, 1366.

Feldberg, L., **Dong, Y.**, **Heinig, U.**, **Rogachev, I.** & **Aharoni, A.** (2018) DLEMMA-MS-imaging for identification of spatially localized metabolites and metabolic network map reconstruction. *Analytical Chemistry*, **90**, 10231–10238.

Feldberg, L., **Venger, I.**, **Malitsky, S.**, **Rogachev, I.** & **Aharoni, A.** (2009) Dual labeling of metabolites for metabolome analysis (DLEMMA): a new approach for the identification and relative quantification of metabolites by means of dual isotope labeling and liquid chromatography–mass spectrometry. *Analytical Chemistry*, **81**, 9257–9266.

Fernández-García, J., **Altea-Manzano, P.**, **Pranzini, E.** & **Fendt, S.-M.** (2020) Stable isotopes for tracing mammalian-cell metabolism in vivo. *Trends in Biochemical Sciences*, **45**, 185–201.

Freund, D.M. & **Hegeman, A.D.** (2017) Recent advances in stable isotope-enabled mass spectrometry-based plant metabolomics. *Current Opinion in Biotechnology*, **43**, 41–48.

Giavalisco, P., **Li, Y.**, **Matthes, A.**, **Eckhardt, A.**, **Hubberten, H.-M.**, **Hesse, H.** ***et al.*** (2011) Elemental formula annotation of polar and lipophilic metabolites using ^{13}C, ^{15}N and ^{34}S isotope labelling, in combination with high-resolution mass spectrometry. *The Plant Journal*, **68**, 364–376.

Hegeman, A.D., **Schulte, C.F.**, **Cui, Q.**, **Lewis, I.A.**, **Huttlin, E.L.**, **Eghbalnia, H.** ***et al.*** (2007) Stable isotope assisted assignment of elemental compositions for metabolomics. *Analytical Chemistry*, **79**, 6912–6921.

Iyer, S.S., **Zhang, Z.-P.**, **Kellogg, G.E.** & **Karnes, H.T.** (2004) Evaluation of deuterium isotope effects in Normal-phase LC-MS-MS separations using a molecular modeling approach. *Journal of Chromatographic Science*, **42**, 383–387.

Ji, J., **Liu, X.**, **Hu, X.**, **Chen, F.**, **Bueschl, C.**, **Schuhmacher, R.** ***et al.*** (2022) A novel method combining stable isotopic labeling and high-resolution mass spectrometry to trace the quinone reaction products in wines. *Food Chemistry*, **383**, 132448.

Kachlicki, P., **Piasecka, A.**, **Stobiecki, M.** & **Marczak, Ł.** (2016) Structural characterization of flavonoid glycoconjugates and their derivatives with mass spectrometric techniques. *Molecules*, **21**, 1494.

Kato, K., **Jingu, S.**, **Ogawa, N.** & **Higuchi, S.** (2000) Determination of pibutidine metabolites in human plasma by LC-MS/MS. *Journal of Pharmaceutical and Biomedical Analysis*, **24**, 237–249.

Kessner, D., **Chambers, M.**, **Burke, R.**, **Agus, D.** & **Mallick, P.** (2008) ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics*, **24**, 2534–2536.

Kim, S., **Rodgers, R.P.** & **Marshall, A.G.** (2006) Truly "exact" mass: elemental composition can be determined uniquely from molecular mass measurement at ~0.1 mDa accuracy for molecules up to ~500 Da. *International Journal of Mass Spectrometry*, **251**, 260–265.

Kind, T. & **Fiehn, O.** (2006) Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinformatics*, **7**, 234.

Mitchell, J.M., **Chi, Y.**, **Zheng, S.**, **Thapa, M.**, **Wang, E.** & **Li, S.** (2024) Annotation of metabolites in stable isotope tracing untargeted metabolomics via Khipu-web. *Journal of the American Society for Mass Spectrometry*, **35**, 2824–2835.

Nagao, T., **Yukihira, D.**, **Fujimura, Y.**, **Saito, K.**, **Takahashi, K.**, **Miura, D.** ***et al.*** (2014) Power of isotopic fine structure for unambiguous determination of metabolite elemental compositions: in silico evaluation and metabolomic application. *Analytica Chimica Acta*, **813**, 70–76.

Nakabayashi, R. & **Saito, K.** (2020) Higher dimensional metabolomics using stable isotope labeling for identifying the missing specialized metabolism in plants. *Current Opinion in Plant Biology*, **55**, 84–92.

Paton, R.S. & **Goodman, J.M.** (2007) Exploration of the accessible chemical space of acyclic alkanes. *Journal of Chemical Information and Modeling*, **47**, 2124–2132.

Shahaf, N., **Rogachev, I.**, **Heinig, U.**, **Meir, S.**, **Malitsky, S.**, **Battat, M.** ***et al.*** (2016) The WEIZMASS spectral library for high-confidence metabolite identification. *Nature Communications*, **7**, 12423.

Simpson, J.P., **Olson, J.**, **Dilkes, B.** & **Chapple, C.** (2021) Identification of the tyrosine- and phenylalanine-derived soluble metabolomes of sorghum. *Frontiers in Plant Science*, **12**, 714164.

Simpson, J.P., **Wunderlich, C.**, **Li, X.**, **Svedin, E.**, **Dilkes, B.** & **Chapple, C.** (2021) Metabolic source isotopic pair labeling and genome-wide association are complementary tools for the identification of metabolite–gene associations in plants. *The Plant Cell*, **33**, 492–510.

Smith, C.A., **Want, E.J.**, **O'Maille, G.**, **Abagyan, R.** & **Siuzdak, G.** (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry*, **78**, 779–787.

Sumner, L.W., **Amberg, A.**, **Barrett, D.**, **Beale, M.H.**, **Beger, R.**, **Daykin, C.A.** ***et al.*** (2007) Proposed minimum reporting standards for chemical analysis chemical analysis working group (CAWG) metabolomics standards initiative (MSI). *Metabolomics*, **3**, 211–221.

Triebl, A. & **Wenk, M.R.** (2018) Analytical considerations of stable isotope labelling in Lipidomics. *Biomolecules*, **8**, 151.

Tuinstra, C.R., **Luo, Z.**, **Simpson, J.**, **Dilkes, B.** & **Chapple, C.** (2022) LC/MS coupled to PODIUM establishes the amino acid metabolomes of *Sorghum bicolor*. *The FASEB Journal*, **36**, fasebj.2022.36.S1.R2711.

Wang, Y., **Liu, S.**, **Hu, Y.**, **Li, P.** & **Wan, J.-B.** (2015) Current state of the art of mass spectrometry-based metabolomics studies – a review focusing on wide coverage, high throughput and easy identification. *RSC Advances*, **5**, 78728–78737.

Wieling, J. (2002) LC-MS-MS experiences with internal standards. *Chromatographia*, **55**, S107–S113.

Yang, H., **Shi, L.**, **Zhuang, X.**, **Su, R.**, **Wan, D.**, **Song, F.** ***et al.*** (2016) Identification of structurally closely related monosaccharide and disaccharide isomers by PMP labeling in conjunction with IM-MS/MS. *Scientific Reports*, **6**, 28079.

Yu, D., **Zhou, L.**, **Liu, X.** & **Xu, G.** (2023) Stable isotope-resolved metabolomics based on mass spectrometry: methods and their applications. *TrAC Trends in Analytical Chemistry*, **160**, 116985.

Zhao, M. & **Tian, J.** (2022) Recent advances on the structure identification of components in food. *eFood*, **3**, e51.