analytical chemistry

Article

pubs.acs.org/ac

Development of a GC/Quadrupole-Orbitrap Mass Spectrometer, Part II: New Approaches for Discovery Metabolomics

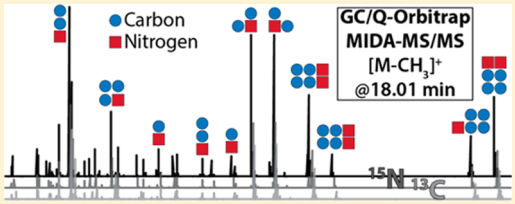
Amelia C. Peterson,[†] Allison J. Balloon,[†] Michael S. Westphall,[§] and Joshua J. Coon*[,†,‡,§]

[†]Departments of Chemistry and [‡]Biomolecular Chemistry, University of Wisconsin−Madison, Madison, Wisconsin 53706, United States

[§]Genome Center of Wisconsin, University of Wisconsin−Madison, Madison, Wisconsin 53706, United States

Supporting Information

ABSTRACT: Identification of unknown peaks in gas chromatography/mass spectrometry (GC/MS)-based discovery metabolomics is challenging, and remains necessary to permit discovery of novel or unexpected metabolites that may elucidate disease processes and/or further our understanding of how genotypes relate to phenotypes. Here, we introduce two new technologies and an analytical workflow that can facilitate the identification of unknown peaks. First, we report on a GC/Quadrupole-Orbitrap mass spectrometer that provides high mass accuracy, high resolution, and high sensitivity analyte detection. Second, with an "intelligent" data-dependent algorithm, termed molecular-ion directed acquisition (MIDA), we maximize the information content generated from unsupervised tandem MS (MS/MS) and selected ion monitoring (SIM) by directing the MS to target the ions of greatest information content, that is, the most-intact ionic species. We combine these technologies with ^{13}C- and ^{15}N-metabolic labeling, multiple derivatization and ionization types, and heuristic filtering of candidate elemental compositions to achieve (1) MS/MS spectra of nearly all intact ion species for structural elucidation, (2) knowledge of carbon and nitrogen atom content for every ion in MS and MS/MS spectra, (3) relative quantification between alternatively labeled samples, and (4) unambiguous annotation of elemental composition.



The overarching goal of metabolomics is to characterize all low-molecular-weight metabolites present in a biological system. Approaches for mass spectrometry (MS)-based metabolomics can be of two varieties: targeted and discovery. Targeted approaches permit absolute quantification of a limited set of known metabolites using internal reference standards, which simultaneously confirm endogenous metabolite identity. Discovery approaches attempt comprehensive analysis of the metabolome through unbiased investigation.[1] An advantage to the discovery approach is that it permits detection of novel compounds, which could elucidate a link between genotype and phenotype, thereby providing disease biomarkers.[2] That said, interpretation of mass spectra without a reference spectrum is not routine; despite their centrality to the metabolomics experiment, spectral interpretation and identification remain the most challenging aspects of the analysis.[3]

The challenges hampering spectral interpretation and identification in metabolomics are many fold. First, the targets of a metabolomic analysis are often chemically indistinct from the reagents used to prepare the metabolomic extract.[4−6] Second, with gas chromatography (GC)/MS multiple MS peaks per analyte may be present due to incomplete derivatization. At the same time, there may be multiple analytes per MS peak due to degradation and side reactions.[4] In a typical GC/MS study, only 5−15% of mass spectral features are assigned metabolomic identity.[7] Identifying the metabolic features is necessary, not only to improve analytical depth[8] and drive understanding of the metabolome, but also to prevent erroneous biological conclusions based on nonmetabolic signals.[4] Third, even for features of metabolic origin with quality mass spectra, true unknowns and novel derivatives of known analytes cannot be annotated via database searching.[9−12]

In vivo stable isotope incorporation techniques show promise in addressing many of these challenges.[2] Several groups have recently used metabolic labeling with stable isotopes (generally, ^{13}C and ^{15}N) in discovery, mostly liquid chromatography (LC)/MS-based, studies.[6,13−22] The stable isotope labeling (SIL) approach offers a means of discriminating between true metabolite signals and spurious background.[17−19] If MS/MS is employed, spectral interpretation and structural elucidation is greatly aided by partial knowledge of the elemental formula of each fragment ion. Lastly, SIL provides an internal standard for every metabolite to assess recovery[21] and provide relative quantification.[13,16,20,22]

Another strategy to reduce candidate elemental compositions is through filtering using a set of heuristic rules developed by Kind and Fiehn.[11] The Seven Golden Rules apply a number of chemical rules, elemental ratios, and elemental probabilities, and evaluate the accuracy of the experimental mass isotopomer abundance distribution.[10] Together with a MS having high mass and isotopomer ratio accuracy, these rules allow

Received: March 26, 2014
Accepted: August 28, 2014
Published: August 28, 2014

   dx.doi.org/10.1021/ac5014755 | *Anal. Chem.* 2014, 86, 10044−10051

assignment of the correct formula 98% of the time for compounds present in a database.[11]

In this report, we introduce two new technologies that hold promise to further facilitate unambiguous assignment of elemental composition in discovery, GC/MS-based metabolomic analyses. The first, our newly introduced GC/Quadrupole-Orbitrap mass spectrometer (see accompanying article[23]), enables highly flexible GC/MS analysis with high mass accuracy, resolution, sensitivity, and scan speed. The second, an "intelligent" data-dependent acquisition paradigm for small molecule discovery, termed "molecular-ion directed acquisition" or MIDA, maximizes the information content from data-dependent MS/MS and SIM by directing the instrument to sample the ions of greatest information content. Using polar metabolites from *Arabidopsis thaliana* extracts as a model system, we combine these two technologies with ^{13}C- and ^{15}N-*in vivo* metabolic labeling to enable MS and MS/MS-level annotation and relative quantification, the use of multiple derivatization and ionization conditions, and heuristic rules-based filtering of molecular formulas. We demonstrate the unsupervised, consistent acquisition of structurally rich MS/MS spectra for intact ion species from nearly all MS features over multiple analyses, knowledge of the carbon and nitrogen content in all MS and MS/MS peaks, and finally, unambiguous assignment of elemental compositions for all queried features.

EXPERIMENTAL SECTION

Molecular-Ion Directed Acquisition (MIDA). The GC/Quadrupole-Orbitrap's Python-based firmware was modified to enable MIDA. Prior to an experiment, the algorithm was informed by the user of (1) the ionization type (methane positive chemical ionization (PCI) or electron ionization (EI)), (2) the sample derivatization reagent (*N*-Methyl-*N*-(trimethylsilyl) trifluoroacetamide (MSTFA) or *N*-*tert*-Butyldimethylsilyl-*N*-methyltrifluoroacetamide (MTBSTFA)), (3) the mass tolerance to be used by the algorithm, (4) the minimum signal-to-noise of mass spectral peaks to be considered as the initial peak in a spectral pattern, (5) the member of the pattern to subsequently target by MS/MS or SIM, (6) the number of targets per MS spectrum to target for MS/MS or SIM, and (7) the duration of time to exclude targets from MS/MS or SIM analysis. MIDA was developed for the following combinations of ionization and derivatization: (1) methane PCI with *tert*-butyldimethylsilyl (*t*BDMS) derivatization, (2) methane PCI with trimethylsilyl (TMS) derivatization, and (3) EI with *t*BDMS derivatization.

Typical parameters employed for MIDA were as follows: mass tolerance of ±10 ppm; minimum S/N of 100; $[M + H]^+$ or $[M - t\text{-butyl}]^+$ target ion for methane PCI and EI, respectively; 1 target per MS spectrum; and no dynamic exclusion of targets (0 s exclusion). MIDA MS/MS scans were acquired with an isolation width of 5 Th, normalized collision energy of 25 eV, resolution of 17 500 fwhm, AGC target of 5×10^5, maximum injection time of 100 ms, and scan range of 65−850 Th. MIDA SIM scans used the same parameters as MS/MS scans except that the isolation width was 20 Th in order to capture the entire isotopic envelope of interest.

MIDA templates (*vide infra*) scored with empirically developed relationships based on a dot-product of the m/z and intensity of each member of the template. Heavily weighting in the m/z-domain was necessary to promote the selection of the correct series of ions. The scores for templates utilizing methane PCI and EI are given by the following eqs 1 and 2:

$$\text{score}_{\text{PCI}} = \sum_{j=1}^{n} \frac{I_j}{2} M_j^{\,3} \tag{1}$$

$$\text{score}_{\text{EI}} = \sum_{j=1}^{n} \frac{I_j}{2} M_j^{\,5} \tag{2}$$

where I_j and M_j are the S/N and m/z, respectively, of the jth member of n total template members.

Data Analysis. Manual curation of chromatograms and mass spectra, isotopic distribution simulations, and calculation of elemental compositions were performed with Xcalibur Qual Browser 2.3.23 (Thermo Fisher Scientific, San Jose, CA). Candidate elemental compositions were generated within a mass tolerance of ±5−10 ppm using the element constraints, $C_{0-150}H_{0-150}O_{0-50}N_{0-50}S_{0-50}P_{0-50}Si_{0-10}$, and subsequently filtered using the Seven Golden Rules[11] program (available at http://fiehnlab.ucdavis.edu/projects/Seven_Golden_Rules/), which was modified for *t*BDMS-derivatized compounds when applicable. Isotopomer abundance error tolerance was set to 15% if all three isotopes were present and higher if all isotopes were not present. Filtering by the number of carbon and nitrogen present was performed on the final list of candidates produced. To assess the algorithm accuracy rate for determining the correct target ion in a given spectrum, a set of 100 high-scoring spectra were manually validated by two independent reviewers for confirmation that the correct target was chosen.

Relative quantification of $^{13}C^{14}N/^{12}C^{14}N$ samples was performed on the $[M - t\text{-butyl}]^+$ (EI) or $[M + H]^+/[M - t\text{-butyl}]^+$ (CI) isotopomer cluster(s) for 28 randomly selected compounds using a "consensus spectrum" comprising the isotopomer intensities of each member of the cluster, averaged over the first half of the peak elution profile (to accommodate isotope swing[24]). The approximate contribution of each species to the isotopomer cluster was determined by the method of least-squares for overdetermined systems (Figure S3 in the Supporting Information).[25] Note, in this work, the straightforward use of only the peak heights of the ^{12}C- and ^{13}C-monoisotopic peaks for relative quantitation, as previously reported,[16] was not possible for several reasons. Since the incorporation of ^{13}C in the *Arabidopsis thaliana* model was not 100%, every $^{13}C^{14}N$-isotopomer cluster was considered to possibly also contain species having 1 or 2 fewer ^{13}C than the fully incorporated species; the contributions of these species were summed to produce the full contribution of the $^{13}C^{14}N$ sample relative to the $^{12}C^{14}N$ sample and then normalized to the ratio obtained in the 1:1 mixture analysis for that compound. Additionally, since the native metabolites in this study have only from 1 to 11 carbon atoms, a given $^{12}C^{14}N$ or $^{13}C^{14}N$ monoisotopic peak might also contain contributions from other ions having 1 or 2 hydrogen atoms less (mostly applicable in CI). We have anticipated in our analysis that the obtained ion clusters for each pair could overlap and contain as many as five different ionic species in EI and six different species in CI.

Further experimental details, including sample preparation and GC/MS conditions, are available in the Supporting Information.

dx.doi.org/10.1021/ac5014755 | *Anal. Chem.* 2014, 86, 10044−10051

RESULTS AND DISCUSSION

Molecular-Ion Directed Acquisition (MIDA). We report here an "intelligent," data-dependent acquisition (DDA) method for directing real-time tandem MS events (MIDA). In traditional, intensity-based DDA, an intensity filter is used to determine the targets of subsequent MS/MS scans.[3,26] In GC/MS, where significant fragmentation occurs upon ionization, triggering MS/MS events based on the most-intense species in the spectrum often results in the fragmentation of low-m/z, low-information content ions. As a result, most of the use of MS/MS in GC/MS analysis relies on targeted/scheduled methods, like selected reaction monitoring (SRM), which are not amenable to discovery applications.[3] To maximize the information content from MS/MS,[3] the instrument should be directed to preferred analytical targets, such as the molecular/pseudomolecular ion of each analyte. The spectral processing algorithm employed by MIDA directs the instrument to these ions by exploiting the expected adducts that form during the methane PCI process, as well as the characteristic fragmentation patterns of commonly employed derivatization reagents. From these patterns of ions, we have developed templates collectively comprising the mass differences resulting from fragmentation of, and adduction to, the molecular ion species for the following combinations of ionization and derivatization: (1) methane PCI with *t*BDMS derivatization, (2) methane PCI with TMS derivatization, and (3) EI with *t*BDMS derivatization.

Members of each MIDA template have a set mass difference from a "template initiator" ion, the lowest m/z template member. To ensure the specificity of MIDA for its intended target, most members of a template are required to be present. However, since the analyte dictates the presence of certain ions, some template members are optional; this ensures adequate sensitivity of the algorithm. Any required template member can be targeted by the subsequent MS/MS analysis. For the combination of methane PCI and *t*BDMS, for example, the template has five ions, three required ($[M - C_4H_9]^+$, $[M - CH_3]^+$, and $[M + H]^+$), and two optional ($[M + C_2H_5]^+$ and $[M + C_3H_5]^+$). The initiator ion, $[M - C_4H_9]^+$, and $[M - CH_3]^+$ (Δm = 42.04695 Da) correspond to the loss of a *t*-butyl and methyl moiety, respectively, from the *t*BDMS groups derivatizing the molecule. The remaining three ions result from proton transfer and adduct formation reactions during methane PCI, with mass differences from the initiator of 58.07825, 86.10955, and 98.10955 Da. Since the two adduct ions can be of low abundance or absent, they are optional in this template (see Figure S1A in the Supporting Information).

Use of a template in the MIDA process is presented in the following example (see Figure S1B in the Supporting Information). Prior to the start of a GC/MS analysis, the user (1) selects the template by specifying the ionization method and sample derivatization type (e.g., PCI and *t*BDMS), (2) sets the member of the template to target (e.g., $[M - C_4H_9]^+$), and (3) establishes the mass error tolerance (e.g., ±10 ppm) and S/N threshold (e.g., 100) to enforce for matching the templates to the acquired MS spectra. Following the acquisition of a MS spectrum, the on-board instrument computer "scans" the MIDA template across the MS spectrum. At each potential "template initiator" ion (any ion having S/N > 100), the spectrum is queried for m/z values falling within ±10 ppm of each template member. If all required members are found, the template is considered "complete", and a dot product-based score is calculated based on the m/z and intensity of each template member. Next, all "completed" templates for the MS spectrum are stratified by score and m/z, and the user-specified target member, i.e. $[M - C_4H_9]^+$, of the highest-scoring template is then isolated and fragmented by the instrument for the subsequent MS/MS spectrum. The instrument then proceeds to target other templates, in order of decreasing score, if multiple data-dependent events are specified, or acquires the next MS spectrum. Because of the time constraints of gas chromatography, as well as its high separation efficiency, a single MS/MS event (top 1) without any dynamic exclusion of previously selected ions was found necessary. This process repeats throughout an analysis to yield MS and MS/MS data for nearly all eluting analytes present in the sample, with the targeted ion for each peak profiled over the entirety of its elution (Figure S1C in the Supporting Information).

To assess the accuracy of this approach for directing MS/MS, the target selected by the MIDA algorithm was confirmed by manual annotation of a set of 100 high-scoring spectra per analysis. For each template, two separate researchers graded three analyses and the accuracy results were averaged. This was necessary because only manual annotation is possible in the absence of library reference spectra. Using this technique, the MIDA algorithm had an accuracy rate of 93.6% and 91.3% for *t*BDMS derivatization with methane PCI and EI, respectively. The accuracy rate fell slightly, to 88.3%, for TMS derivatization with methane PCI.

An important consideration of any such "real-time" algorithm is the amount of overhead, or interscan time, required for execution. MIDA (utilizing 17 500 fwhm resolution for both MS and MS/MS scans) proceeds at a rate of 9.3 Hz (108 ms per scan), while regular DDA is approximately 16% faster, proceeding at 11 Hz (91 ms per scan). As a point of reference, MS-only acquisition runs at 13 Hz (77 ms), 16% and 28% faster than DDA and MIDA, respectively (Figure S1D in the Supporting Information).

Further details of the MIDA algorithm along with the corresponding pseudocode are available in the Supporting Information.

MIDA with Metabolic Labeling for Discovery Metabolomics and Structural Elucidation. To assess the utility of MIDA, the structural information gained by intelligent acquisition of MS/MS spectra and the advantages of metabolic labeling for assigning elemental compositions were combined to study polar metabolites from *A. thaliana*. Identical polar fraction extracts of *A. thaliana* grown under four metabolic labeling conditions were analyzed: natural abundance ($^{12}C^{14}N$), ^{13}C-enriched ($^{13}C^{14}N$), ^{15}N-enriched ($^{12}C^{15}N$), and both ^{13}C- and ^{15}N-enriched ($^{13}C^{15}N$). Following derivatization, samples were analyzed using MS and MIDA-MS/MS, as depicted in Figure 1. Spectra from either MS or MIDA-MS/MS analysis allow assignment of the number of nitrogens and carbons comprising the underivatized molecule (Figure 1A). Using MIDA-MS/MS, in Figure 1B,C, the algorithm's recognition of fragmentation/adduction patterns, rather than specific m/z or intensity values, permits the acquisition of MS/MS spectra of the $[M - CH_3]^+$ of each species in four separate analyses. Comparison of these MS/MS spectra allows immediate assignment of the number of nitrogen and carbon atoms present in each fragment ion. This information, along with high mass accuracy m/z measurements, can reduce the number of potential elemental composition candidates for a given peak

dx.doi.org/10.1021/ac5014755 | *Anal. Chem.* 2014, 86, 10044−10051

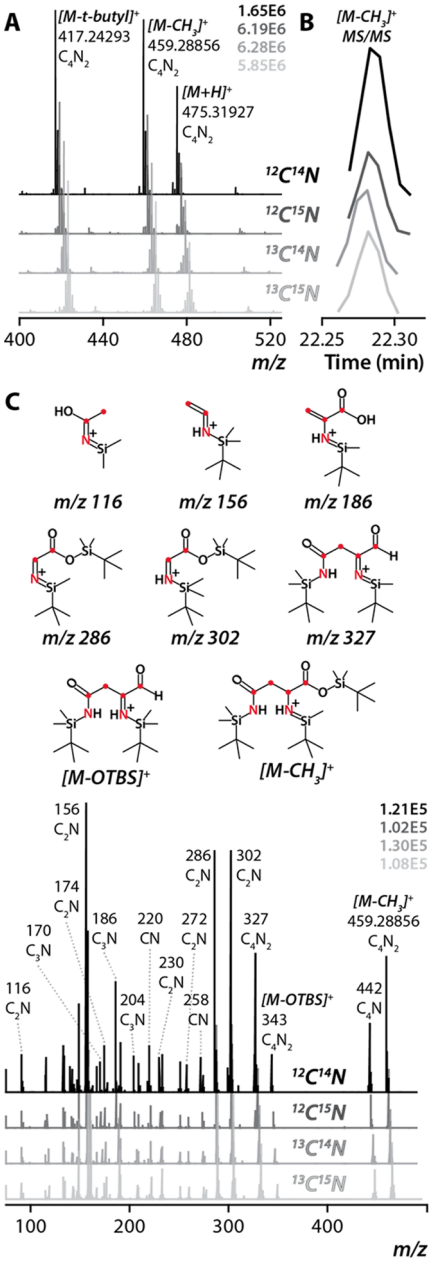


Figure 1. Typical MIDA-MS/MS with metabolic labeling data. (A) Partial MS spectrum showing $[M - t\text{-butyl}]^+$, $[M - CH_3]^+$, $[M + H]^+$, and methane PCI adduct ions (unlabeled) for asparagine − 3TBS under four different metabolic labeling states. Comparison of the m/z shifts of similar ions between the four states allows assignment of the number of carbons and nitrogens in each ion. (B) Profiles of the MIDA-MS/MS of $[M-CH_3]^+$ over the entire elution profile of asparagine − 3TBS for all four labeling states. (C) MS/MS of asparagine − 3TBS $[M-CH_3]^+$ ion from all four labeling states. Comparison of the m/z shifts of similar ions between the four spectra allows assignment of the number of carbon and nitrogen in each fragment ion. Above, structures proposed using the knowledge of the number of carbons and nitrogens (shown as red circles or red letters, respectively) for eight ions in the MS/MS spectrum. The MS/MS spectrum confirms the structure of asparagine − 3TBS.

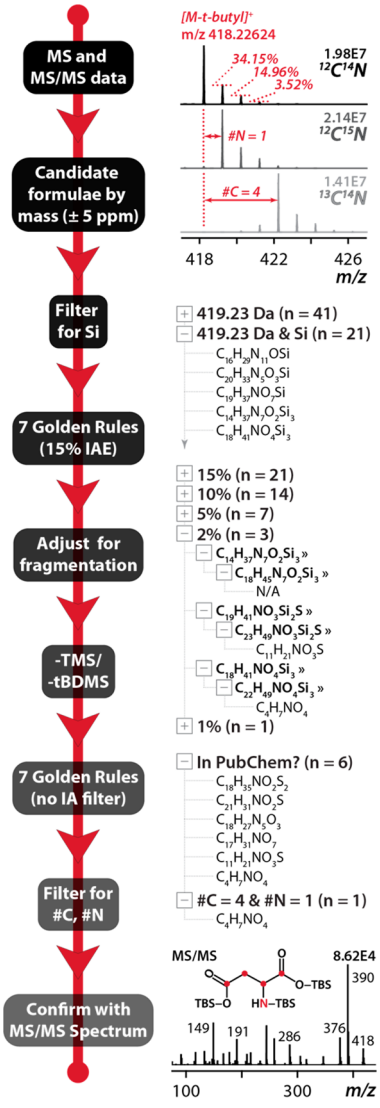


Figure 2. Workflow for spectral annotation and structural confirmation. From top to bottom, first, the ion type selected by MIDA for MS/MS, the m/z, the abundance of the first−third isotopomers, and the number of carbons and nitrogens present are noted from the MS spectrum. Candidate formulas are then generated within ±5 ppm tolerance of the neutralized mass of the ion and filtered for Si to result in a list of 21 formulas (out of 41). Candidates are submitted for filtering by the Seven Golden Rules with a 15% isotopomer abundance error (IAE) threshold. All 21 formulas meet the 15% threshold, 14 meet a 10% threshold, 7 at 5%, 3 at 2%, and 1 at 1%. All formulas meeting the 15% threshold are made intact by addition of C_4H_8, a *t*-butyl group (shown in the second level under the formulas meeting the 2% IAE threshold). Silylation groups are removed from the intact formulas, as shown in the third level under the formulas meeting the 2% IAE threshold. The desilylated formulas are refiltered by the Seven Golden Rules. The six formulas present in PubChem are further filtered by the number of nitrogen and carbon present in the analyte (four carbons and one nitrogen) to yield a single formula, $C_4H_7NO_4$, which is confirmed using the MS/MS spectrum, and tentatively identified as aspartate − 3TBS.

and facilitate its structural characterization, as shown by the suggested structures for eight peaks in the spectrum in Figure 1C.

The workflow illustrated in Figure 2 was developed to generate unique elemental composition assignments and tentative identifications for each putative metabolite. A list of candidate elemental compositions within ±5 ppm of the MIDA-targeted ion was generated using lax constraints on the number of allowable carbon, nitrogen, oxygen, hydrogen, sulfur, phosphorus, and silicon atoms. This initial list was filtered for

dx.doi.org/10.1021/ac5014755 | *Anal. Chem.* 2014, 86, 10044−10051

Table 1. Selected Compounds Tentatively Identified by Our Workflow in Methane PCI with TMS Derivatization Analyses Showing the Reduction of Candidates with Each Filtering Step (See Figure 2)

ion	*m/z*	no. C	no. N	MS/MS[a]	no. by mass[b]	no. Pub Chem[c]	no. N and no. C[d]	native formula[e]	proposed ID[f]	verified[g]	mass error (ppm)	avg IAE (%)
$[M - CH_3]^+$	174.0584	3	0	×	1	1	1	$C_3H_4O_3$	pyruvic acid TMS MOX	×	4.5	1.7
$[M - CH_3]^+$	234.1161	0	0	×	10	1	1	H_3NO	hydroxylamine 3TMS	–	4.3	1.2
$[M - CH_3]^+$	218.1028	3	1	×	2	1	1	$C_3H_7NO_2$	alanine 2TMS	×	4.8	1.2
$[M + H]^+$	235.0818	2	0	–	9	2	1	$C_2H_2O_4$	oxalic acid 2TMS	×	−0.8	8.0
$[M + H]^+$	176.0741	2	0	–	3	0	1	$C_2H_2O_3$	glyoxylic acid TMS MOX	×	−1.8	3.1
$[M - CH_3]^+$	220.0822	1	0	×	2	0	1	CH_2O_3	carbonic acid 2TMS MOX	–	4.0	1.5
$[M + H]^+$	262.1655	5	1	×	6	2	1	$C_5H_{11}NO_2$	valine 2TMS	×	−0.6	1.3
$[M - CH_3]^+$	189.0873	1	2	×	5	3	1	CH_4N_2O	urea 2TMS	×	6.0	2.0
$[M - CH_3]^+$	179.0524	7	0	×	3	2	1	$C_7H_6O_2$	benzoic acid TMS	×	5.4	1.6
$[M - CH_3]^+$	262.1471	2	1	×	4	2	1	C_2H_7NO	ethanolamine 3TMS	×	4.9	1.6
$[M + H]^+$	315.1026	0	0	×	29	7	1	H_3O_4P	phosphate 3TMS	×	0.5	0.7
$[M - CH_3]^+$	260.1499	6	1	×	4	2	1	$C_6H_{13}NO_2$	leucine 2TMS	×	3.4	4.7
$[M - CH_3]^+$	293.1419	3	0	×	9	3	1	$C_3H_8O_3$	glycerol 3TMS	×	3.7	2.0
$[M + H]^+$	276.1814	6	1	×	6	2	1	$C_6H_{13}NO_2$	isoleucine 2TMS	×	−1.7	2.3
$[M + H]^+$	196.0790	6	1	–	2	2	1	$C_6H_5NO_2$	nicotinic acid TMS	×	−0.9	2.5
$[M - CH_3]^+$	244.1182	5	1	×	3	1	1	$C_5H_9NO_2$	proline 2TMS	×	5.0	1.2
$[M - CH_3]^+$	245.0660	4	0	–	7	2	1	$C_4H_4O_4$	maleic acid 2TMS	×	4.5	2.0
$[M - CH_3]^+$	276.1262	2	1	×	17	8	1	$C_2H_5N0_2$	glycine 3TMS	×	5.3	5.7
$[M - CH_3]^+$	247.0818	4	0	×	7	2	1	$C_4H_6O_4$	succinic acid 2TMS	×	3.9	1.2
$[M - CH_3]^+$	307.1209	3	0	×	18	4	1	$C_3H_6O_4$	glyceric acid 3TMS	–	4.5	0.8
$[M - CH_3]^+$	245.0657	4	0	×	19	5	1	$C_4H_4O_4$	fumaric acid 2TMS	×	5.5	1.4
$[M - CH_3]^+$	306.1372	3	1	×	10	3	1	$C_3H_7NO_3$	serine 3TMS	×	3.4	1.2
$[M - CH_3]^+$	320.1530	4	1	×	13	3	1	$C_4H_9NO_3$	threonine 3TMS	×	2.7	1.1
$[M - CH_3]^+$	320.1534	4	1	×	27	7	1	$C_4H_9NO_3$	allothreonine 3TMS	–	1.5	8.6
$[M - CH_3]^+$	349.1316	5	0	–	32	9	1	$C_5H_8O_5$	citramalic acid 3TMS	–	3.5	13.4
$[M - CH_3]^+$	335.1159	4	0	×	29	7	1	$C_4H_6O_5$	malate 3TMS	×	3.9	0.9
$[M - CH_3]^+$	267.0865	7	0	–	10	1	1	$C_7H_6O_3$	hydroxybenzoic acid 2TMS	–	4.8	7.4
$[M + H]^+$	350.1633	4	1	×	23	7	1	$C_4H_7N0_4$	asparti acid 3TMS	–	0.3	0.5
$[M + H]^+$	274.1290	5	1	×	7	2	1	$C_5H_7NO_3$	pyroglutamic acid 2TMS	×	−0.4	0.9
$[M - CH_3]^+$	332.1531	5	1	–	15	4	1	$C_5H_9NO_3$	hydroxyproline 3TMS	×	2.4	10.2
$[M - CH_3]^+$	304.1574	4	1	×	21	7	1	$C_4H_9NO_2$	4-aminobutyric acid 3TMS	×	5.1	1.0
$[M - CH_3]^+$	322.1135	3	1	–	44	11	1	$C_3H_7NO_2S$	cysteine 3TMS	×	6.0	15.9
$[M - CH_3]^+$	409.1717	4	0	–	66	18	1	$C_4H_8O_5$	threonic acid 4TMS	–	1.6	1.8
$[M - CH_3]^+$	333.1844	5	2	×	14	5	1	$C_5H_{12}N_2O_2$	ornithine 3TMS	–	3.3	0.4
$[M - CH_3]^+$	348.1473	5	1	×	24	7		$C_5H_9NO_4$	glutamic acid 3TMS	×	4.4	0.9
$[M + H]^+$	310.1649	9	1	×	10	2	1	$C_9H_{11}NO_2$	phenylalanine 2TMS	×	1.5	1.1
$[M - CH_3]^+$	333.1483	4	2	×	20	7	1	$C_4H_0N_2O_3$	asparagine 3TMS	×	2.5	0.7
$[M - CH_3]^+$	419.2037	5	2	×	56	13	1	$C_5H_{10}N_2O3$	glutamine 4TMS	–	1.5	0.4
$[M - CH_3]^+$	361.2346	4	2	–	15	6	1	$C_4H_{12}N_2$	putrescine 4TMS	–	1.7	6.7
$[M + H]^+$	436.1638	6	0	×	189	1	1	$C_6H_6O_7$	2-oxalosuccinic acid 3TMS MOX	–	0.0	1.7
$[M - CH_3]^+$	347.1630	5	2	×	47	12	1	$C_5H_{10}O_5$	glutamine 3TMS	×	5.2	0.8
$[M - CH_3]^+$	447.1869	7	0	–	104	22	1	$C_7H_{10}O_5$	shikimic acid 4TMS	–	2.5	5.5
$[M - CH_3]^+$	465.1604	6	0	×	157	23	1	$C_6H_8O_7$	citrate 4TMS	×	3.8	0.5
$[M - CH_3]^+$	358.1801	6	3	×	17	6	1	$C_6H_{11}N_3O_2$	arginine[-NH3] 3TMS	–	2.0	2.3
$[M - CH_3]^+$	422.1499	6	0	×	75	1	1	$C_6H_8O_8$	2-(Glycoloyloxy)succinic acid 3TMS MOX	–	−1.7	1.4
$[M + H]^+$	229.1163	10	2	–	6	1	1	$C_{10}H_8N_2$	beta-indole-3-acetonitrileTMS	–	−3.1	4.5
$[M - CH_3]^+$	431.1779	4	4	×	86	19	1	$C_4H_6N_4O_3$	allantoin 4TMS	–	2.9	1.0
$[M - CH_3]^+$	356.1639	6	3	–	23	7	1	$C_6H_9N_3O_2$	histidine 3TMS	–	3.5	6.2
$[M + H]^+$	363.2314	6	2	×	17	5	1	$C_9H_{14}N_2O_2$	lysine 3TMS	×	−0.2	5.0
$[M - CH_3]^+$	382.1687	9	1	×	35	7	1	$C_9H_{11}NO_3$	tyrosine 3TMS	–	2.2	1.7
$[M - CH_3]^+$	449.1660	6	0	×	127	22	1	$C_6H_8O_6$	ascorbic acid 4TMS	–	2.9	4.3
$[M - CH_3]^+$	435.1870	6	0	–	90	19	1	$C_6H_{10}O_5$	1,6-anhydroglucose 4TMS	–	2.3	0.9
$[M + H]^+$	613.3080	6	0	–	384	45	1	$C_6H_{12}O_6$	inositol 6TMS	–	−0.2	7.8

dx.doi.org/10.1021/ac5014755 | *Anal. Chem.* 2014, 86, 10044−10051

Table 1. continued

ion	m/z	no. C	no. N	MS/MS[a]	no. by mass[b]	no. Pub Chem[c]	no. N and no. C[d]	native formula[e]	proposed ID[f]	verified[g]	mass error (ppm)	avg IAE (%)
$[M - CH_3]^+$	441.1623	5	4	–	109	31	1	$C_5H_4N_4O_3$	uric acid 4TMS	–	2.7	3.2
$[M + H]^+$	421.2159	11	2	×	55	11	1	$C_{11}H_{12}N_2O_2$	tryptophan 3TMS	–	−0.4	5.6
$[M - CH_3]^+$	353.1238	11	0	×	41	9	1	$C_{11}H_{12}O_5$	sinapic acid 2TMS	–	2.1	1.4

[a]MIDA-MS/MS data was acquired on the $[M-CH_3]^+$ of the analyte. [b]Number of elemental formulas within ±5−10 ppm of the neutralized measured mass. [c]Number of elemental formulas present in PubChem after filtering by the Seven Golden Rules, accounting for fragmentation, removing TMS group, and refiltering by the Seven Golden Rules. [d]Number of elemental formulas remaining after constraining the formulas present in PubChem by the number of carbons and nitrogen in the native analyte. [e]Assigned elemental formula for the native analyte. [f]Proposed identification of the assigned elemental formula based on metabolites expected in *A. thaliana*. [g]Metabolites verified with authentic standards (see Figure S3 in the Supporting Information).

the presence of silicon (required given the sample preparation and the templates used by the MIDA algorithm) and then subjected to further attrition by the Seven Golden Rules[11] using a 15% isotopomer abundance error (IAE) threshold. The heuristic filters employ the LEWIS and SENIOR chemical rules, accuracy of isotopomer abundance patterns, elemental ratios, elemental ratio probabilities, the presence of derivatizable functional groups, and the presence of elemental formulas in the PubChem database (http://pubchem.ncbi.nlm.nih.gov/). The resultant list was filtered for compositions meeting the 15% IAE threshold, and then each elemental composition was adjusted to be fully intact; for example, in Figure 2 each remaining candidate with less than 15% IAE was made "intact" by adding C_4H_8 to each formula (to account for the loss of *t*-butyl, less one hydrogen which was already added to neutralize the initial ion). Thus, the composition $C_{18}H_{41}NO_4Si_3$ became $C_{22}H_{49}NO_4Si_3$. The intact candidates were then stripped of their *t*BDMS groups to reveal a list of native (but possible methoxyaminated) molecules ($C_{22}H_{49}NO_4Si_3$ with loss of 3 *t*BDMS became $C_4H_7NO_4$). After resubmission to the Seven Golden Rules (with a 100% IAE threshold), the remaining elemental compositions present in the PubChem database were filtered by the number of carbon and nitrogen atoms known to be present from the spectral data. If no matching compounds were found and methoxyamination was suspected (note, the $=N-O-CH_3$ moiety added by methoxyamination did not contain ^{13}C or ^{15}N and, thus, was "invisible" by our metabolic labeling method), the compounds were demethoxyaminated by subtraction of NCH_3 and refiltered for nitrogen and carbon. The remaining candidate was then confirmed by annotation of the associated MS/MS data.

Using this strategy, we have made confident elemental composition assignments and suggested plausible identifications for over 80 compounds, some of which we successfully identified as artifacts of the analysis (e.g., hydroxylamine and portions of the benzoic and carbonic acid populations). Table 1 shows a selection of the identified compounds from analyses using TMS derivatization and methane PCI. The mean mass error for annotated mass spectral features was 2.4 ppm (σ = 2.1 ppm, median = 2.5 ppm); the mean percent error for isotopomer abundances (including, first, second, and third isotopomer abundance errors) was 2.9% (σ = 2.9%, median = 1.7%) (Figure S2A in the Supporting Information). While isotopomer abundance accuracy was reduced for low abundance compounds, the mass errors were independent of abundance (Figure S2B in the Supporting Information).

Although the metabolic labeling approach leaves little room for ambiguity, 31 of the identified compounds were also subsequently validated by comparison to an authentic reference standard (Table 1). Spectral and chromatographic comparisons between the confirmed metabolites and standards are shown in Figure S3 in the Supporting Information. For all assignments but three, knowledge of the number of nitrogens and carbons present in the parent molecule permitted a unique result among compounds present in the PubChem database. In the first of these cases, the two remaining compositions for a derivatized $[M-CH_3]^+$ ion at m/z 348.14728, containing 5 carbon and 1 nitrogen atoms, were $C_5H_9NO_4$, corresponding to amino acid glutamic acid, and C_5HN, corresponding to 2,4-pentadiynenitrile, a compound thought to be formed in the atmosphere of Saturn's moon Titan or in the interstellar medium.[27] Occam's razor, the MS/MS spectrum, and the unlikelihood of 2,4-pentadiynenitrile's derivatization with four TMS groups make glutamic acid the clear choice in this case of ambiguity. Similarly, in the second case, analysis of several mass spectral features having a derivatized $[M-CH_3]^+$ ion at m/z 554.26355 containing 6 carbon and no nitrogen atoms resulted in two compositions remaining after filtering, $C_6H_{12}O_6$ − 5 TMS, 1 MOX (likely corresponding to several isomers of glucose) and $C_6H_4O_2$ − 6 TMS, 1 MOX, of which only the former was chemically possible. The third case was similarly unambiguous: $C_6H_{10}O_4$ − 3 TMS, an expected fragment of a di/trisaccharide, following cleavage of the glycosidic bond, versus C_6H_2 − 4 TMS, which is not chemically possible.

The success of the SIL strategy is mirrored by the results of the only other study utilizing this technology to aid identification of unknown spectral features in GC/MS-based metabolomics. Herebian and colleagues[22] studied the metabolome of *Corynebacterium glutamicum*, a species of bacteria used for the industrial-scale production of glutamic acid,[28] through both ^{13}C and ^{15}N metabolic labeling. Using this strategy, they classified several compounds, previously considered part of their *C. glutamicum* metabolite library, as artifacts introduced during sample preparation. Additionally, several hitherto unidentified MS peaks were elucidated using nitrogen and carbon atom counts, as well as knowledge of the number of methoximes and TMS groups present. However, unlike in this study where a unique elemental formula was obtained in nearly all cases, Herebian et al. could not arrive at a unique result in several cases.

Relative Quantification via Metabolic Labeling and MIDA-SIM. Besides attempting to catalog all components in a given sample, metabolomic studies also aim to provide comparative analyses.[2,29] With differential metabolic labeling, samples can be mixed and analyzed simultaneously, and quantitative information on each analyte can be gathered within the same analysis. This strategy obviates concerns of incomparability due to variations in analysis conditions or

dx.doi.org/10.1021/ac5014755 | *Anal. Chem.* 2014, 86, 10044−10051

instrument performance between separate analyses. Additionally, it permits quantification of numerous, natural-abundance samples against a common, metabolically labeled sample, enabling large-scale relative quantification experiments.[2]

To establish feasibility, the $^{13}C^{14}N$-labeled TBS-derivatized sample was serially diluted into the $^{12}C^{14}N$-labeled TBS-derivatized sample at five different ratios (1:1, 2:1, 5:1, 10:1, and 20:1 $^{12}C^{14}N/^{13}C^{14}N$) and the mixes analyzed with methane PCI and EI. For 28 compounds in each analysis, the $^{12}C/^{13}C$ ion pair cluster, corresponding to the $[M-C_4H_9]^+$ of each compound, was manually extracted. The method of least-squares for overdetermined systems[25] was then employed to estimate the relative contribution of each species present in the extracted ion cluster based on the theoretical isotopomer abundance distributions for each species in isolation (Figure S4 in the Supporting Information). The results of this experiment under EI full-scan conditions are shown in black in Figure 3 (for CI full-scan data, see Figure S5A in the Supporting Information).

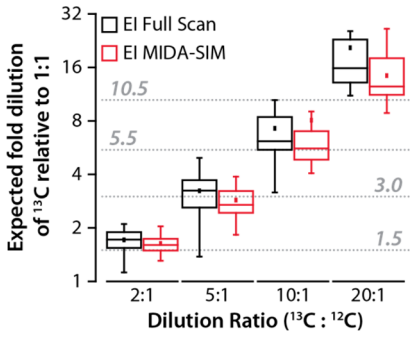


Figure 3. Relative quantification with MIDA-SIM. Accuracy and precision of quantification for dilution of the $^{13}C^{14}N$ sample into the $^{12}C^{14}N$ sample relative to a 1:1 mix. Data from 28 features extracted from EI full scan or EI MIDA-triggered SIM data are shown in black and red, respectively. The improvement of S/N with use of MIDA-SIM enhances quantification accuracy and precision. The target ratio at each dilution is denoted by a dotted gray line.

Note that with complex isotopic clusters, knowledge of the elemental formula is critical to performing relative quantitation. Determining the contribution of each species in the cluster requires that the theoretical isotopomer abundance distribution is known, which further requires knowledge of the elemental composition of the peak. Given that the $^{12}C/^{13}C$ pair also serves to assist assignment of elemental composition by signifying the number of carbon atoms present in the analyte, the two samples act as internal standards for each other. This approach is unlikely to have the accuracy of identification or quantification via an authentic internal reference standard due to incomplete incorporation (see methods). With this caveat, the data, while showing slight overestimation of mixing ratios especially at large dilution ratios (i.e., 20:1), provide sufficient accuracy and reproducibility to detect and estimate the relative abundance of analytes between two samples.

The accuracy of relative quantitation decreases with lower abundance analytes as the dilution ratio increases[30] (Figure S5B in the Supporting Information). In an analogous experiment, Giavalisco and colleagues[16] demonstrated slightly better quantitative accuracy and precision using liquid chromatography/Fourier transform ion cyclotron resonance-mass spectrometry (LC/FTICR-MS)-based relative quantification of various ratios of mixed $^{12}C^{14}N$ and $^{13}C^{15}N$-labeled *A. thaliana* extracts. One explanation for this discrepancy is that the abundance of the pseudomolecular ion will have greater S/N when soft ionization techniques (electrospray ionization (ESI)) are employed, which will yield better quantitative accuracy. One method of increasing S/N is to selectively enrich the population of interest in the gas-phase via selected ion monitoring (SIM). Using a wide 20 Th isolation window to capture the entire $^{12}C/^{13}C$ ion pair cluster, we modified the MIDA algorithm to trigger a SIM scan on the algorithm-selected $[M-C_4H_9]^+$ ion rather than perform MS/MS. We first quantified the S/N enhancement for isolated ions relative to the preceding full scan, finding the average enhancement over ~116k measurements to be 1.8-fold (±3.4-fold). In accordance with our hypothesis, relative quantification accuracy and precision also improved relative to full-scan quantification, as seen in red in Figure 3. These data indicate that gas-phase enrichment through the discovery MIDA-SIM approach can improve relative quantification by reducing some of the bias resulting from insufficient analyte signal. Furthermore, incorporation of a MIDA-SIM scan into the MIDA-MS/MS workflow described above ensures high-quality data for both identification and relative quantification purposes for nearly all analytes across multiple samples.

CONCLUSION

Because of the chemical diversity represented by the metabolome, unknown peak annotation and subsequent structural elucidation in discovery GC/MS-based metabolomics remain intractable issues. According to Fiehn, these gaps must be bridged if GC/MS is to realize its full potential within the metabolomics toolbox.[31] Herein, we have detailed the development and use of two technologies and an analysis workflow that help to address this need. Our newly introduced GC/Quadrupole-Orbitrap MS[23] provides high resolution, mass accuracy, and sensitivity MS data that permit the reliable use of strict filters for candidate elemental formulas. Additionally, stable-isotope labeling, in conjunction with our molecular-ion directed acquisition (MIDA) approach for MS/MS, guarantees not only information-rich MS/MS spectra for intact, or nearly intact, ionic species but also immediate readout of the number of carbon and nitrogen atoms present in each precursor and product ion species. Taken together, these data-driven approaches permit unambiguous assignment of elemental composition to all queried MS features in this study. While we did not employ the standard methods of chromatographic deconvolution, retention index correlation, and spectral database searching, our technology and analysis workflow are complementary to all existing approaches and can be easily incorporated into any standard workflow to further advance the tools available to the GC/MS-based discovery metabolomics community.

ASSOCIATED CONTENT

Supporting Information

Additional figures and tables. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Phone: 608-263-1718. Fax: 608-890-0167. E-mail: jcoon@chem.wisc.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank A.J. Bureta for illustration support and Derek J. Bailey for access to the Python modules upon which MIDA was based. *Arabidopsis thaliana* samples were kindly provided by Benjamin B. Minkoff and Prof. Michael R. Sussman (UW-Madison, Dept. of Biochemistry). This work was funded by Thermo Fisher Scientific, the National Institutes of Health (Grant 1R01GM107199), and the National Science Foundation (Grant 1237936).

■ REFERENCES

(1) Fiehn, O. *Plant Mol. Biol.* **2002**, *48*, 155−171.
(2) Birkemeyer, C.; Luedemann, A.; Wagner, C.; Erban, A.; Kopka, J. *Trends Biotechnol.* **2005**, *23*, 28−33.
(3) Kind, T.; Fiehn, O. *Bioanal. Rev.* **2010**, *2*, 23−60.
(4) Xu, F.; Zou, L.; Ong, C. N. *TrAC, Trends Anal. Chem.* **2010**, *29*, 269−280.
(5) Jankevics, A.; Merlo, M. E.; de Vries, M.; Vonk, R. J.; Takano, E.; Breitling, R. *Metabolomics* **2012**, *8*, 29−36.
(6) Bueschl, C.; Krska, R.; Kluger, B.; Schuhmacher, R. *Anal. Bioanal. Chem.* **2013**, *405*, 27−33.
(7) Hummel, J.; Strehmel, N.; Selbig, J.; Walther, D.; Kopka, J. *Metabolomics* **2010**, *6*, 322−333.
(8) Scheltema, R. A.; Decuypere, S.; Dujardin, J. C.; Watson, D. G.; Jansen, R. C.; Breitling, R. *Bioanalysis* **2009**, *1*, 1551−1557.
(9) Kind, T.; Wohlgemuth, G.; Lee do, Y.; Lu, Y.; Palazoglu, M.; Shahbaz, S.; Fiehn, O. *Anal. Chem.* **2009**, *81*, 10038−10048.
(10) Kind, T.; Fiehn, O. *BMC Bioinf.* **2006**, *7*, 1−10.
(11) Kind, T.; Fiehn, O. *BMC Bioinf.* **2007**, *8*, 1−20.
(12) Fiehn, O.; Kopka, J.; Trethewey, R. N.; Willmitzer, L. *Anal. Chem.* **2000**, *72*, 3573−3580.
(13) Engelsberger, W. R.; Erban, A.; Kopka, J.; Schulze, W. X. *Plant Methods* **2006**, *2*, 14.
(14) Kluger, B.; Bueschl, C.; Lemmens, M.; Berthiller, F.; Haubl, G.; Jaunecker, G.; Adam, G.; Krska, R.; Schuhmacher, R. *Anal. Bioanal. Chem.* **2013**, *405*, 5031−5036.
(15) Giavalisco, P.; Li, Y.; Matthes, A.; Eckhardt, A.; Hubberten, H. M.; Hesse, H.; Segu, S.; Hummel, J.; Kohl, K.; Willmitzer, L. *Plant J.* **2011**, *68*, 364−376.
(16) Giavalisco, P.; Kohl, K.; Hummel, J.; Seiwert, B.; Willmitzer, L. *Anal. Chem.* **2009**, *81*, 6546−6551.
(17) Giavalisco, P.; Hummel, J.; Lisec, J.; Inostroza, A. C.; Catchpole, G.; Willmitzer, L. *Anal. Chem.* **2008**, *80*, 9417−9425.
(18) Hegeman, A. D.; Schulte, C. F.; Cui, Q.; Lewis, I. A.; Huttlin, E. L.; Eghbalnia, H.; Harms, A. C.; Ulrich, E. L.; Markley, J. L.; Sussman, M. R. *Anal. Chem.* **2007**, *79*, 6912−6921.
(19) Rodgers, R. P.; Blumer, E. N.; Hendrickson, C. L.; Marshall, A. G. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 835−840.
(20) Wu, L.; Mashego, M. R.; van Dam, J. C.; Proell, A. M.; Vinke, J. L.; Ras, C.; van Winden, W. A.; van Gulik, W. M.; Heijnen, J. J. *Anal. Biochem.* **2005**, *336*, 164−171.
(21) Mashego, M. R.; Wu, L.; Van Dam, J. C.; Ras, C.; Vinke, J. L.; Van Winden, W. A.; Van Gulik, W. M.; Heijnen, J. J. *Biotechnol. Bioeng.* **2004**, *85*, 620−628.
(22) Herebian, D.; Kupper, U.; Schomburg, D.; Marner, F. J. *Biol. Chem.* **2007**, *388*, 865−871.
(23) Peterson, A. C.; Hauschild, J.-P.; Quarmby, S. T.; Krumwiede, D.; Lange, O.; Lemke, R. A. S.; Grosse-Coosmann, F.; Horning, S.; Donohue, T. J.; Westphall, M. S.; Coon, J. J.; Griep-Raming, J. *Anal. Chem.* **2014**, DOI: 10.1021/ac5014767.
(24) Werner, R. A.; Brand, W. A. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 501−519.
(25) Anton, H. *Elementary Linear Algebra*; 9th ed.; John Wiley and Sons, Inc.: Hoboken, NJ, 2005.
(26) Mann, M.; Hendrickson, R. C.; Pandey, A. *Annu. Rev. Biochem.* **2001**, *70*, 437−473.
(27) Trolez, Y.; Guillemin, J. C. *Angew. Chem., Int. Ed. Engl.* **2005**, *44*, 7224−7226.
(28) Eggeling, L.; Bott, M. *Handbook of Corynebacterium Glutamicum*; CRC Press: Boca Raton, FL, 2005.
(29) Fiehn, O.; Kopka, J.; Dormann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L. *Nat. Biotechnol.* **2000**, *18*, 1157−1161.
(30) Carrillo, B.; Yanofsky, C.; Boismenu, D.; Latterich, M.; Kearney, R. E. In *Proceedings of the 54th Conference of the American Society of Mass Spectrometry and Allied Topics*, Seattle, WA, May 28−June 1, 2006.
(31) Fiehn, O. *TrAC, Trends Anal. Chem.* **2008**, *27*, 261−269.