SUPPORTING INFORMATION

for

Title

Development of a GC/quadrupole-Orbitrap mass spectrometer, Part II: New approaches for discovery metabolomics

Authors
Amelia C. Peterson¹, Allison J. Balloon¹, Michael S. Westphall³, and Joshua J. Coon¹,²,³*

Departments of Chemistry¹ and Biomolecular Chemistry², University of Wisconsin – Madison, Madison, WI, USA

Genome Center of Wisconsin³, University of Wisconsin – Madison, Madison, WI, USA

*Address correspondence to: phone 608-263-1718, fax 608-890-0167, email jcoon@chem.wisc.edu.
# TABLE OF CONTENTS

1. Supplemental Experimental Methods .................................................................................................................... 3  
2. Molecular Ion-Directed Acquisition (MIDA) ....................................................................................................... 5  
3. Pseudocode for Molecular Ion-Directed Acquisition (MIDA) logic ............................................................... 10  
4. Supplemental Figures ........................................................................................................................................ 12  
   4.1. Figure S1 .................................................................................................................................................... 12  
   4.2. Figure S2 .................................................................................................................................................... 13  
   4.3. Figure S3 .................................................................................................................................................... 14  
   4.4. Figure S4 .................................................................................................................................................... 18  
   4.5. Figure S5 .................................................................................................................................................... 19  
5. Supplemental Information References ................................................................................................................. 20
1. Supplemental Experimental Methods

**Reagents.** Unless otherwise specified, all reagents and standards were purchased from Sigma-Aldrich (St. Louis, MO). Methanol, water, pyridine, and hexane (GC or LC/MS grades) were purchased from Fisher Scientific (Fair Lawn, NJ). Compressed gases (methane, helium, and nitrogen) were ultra-high purity grade and purchased from Airgas (Madison, WI).

**Stable isotope metabolic labeling, plant growth, and tissue harvest.** Stable isotope-labeled growth media variants were prepared by adding the following, per liter, to 1X Murashige and Skoog M059 solution: 1.5mL 1 M CaCl$_2$, 0.75 mL 1 M MgSO$_4$, 0.825 g NH$_4$NO$_3$, 0.96 g KNO$_3$, 0.5 g MES, and 10 g D$_6$glucose. For $^{15}$N-labeled plants, NH$_4$NO$_3$ and KNO$_3$ were substituted with $^{15}$NH$_4$$^{15}$NO$_3$ and K$^{15}$NO$_3$. For $^{13}$C-labeled plants, D$_6$glucose was substituted with $^{13}$C-labeled D$_6$glucose. For $^{15}$N/$^{13}$C-labeled plants, both substitutions were made. Stable isotope-labeled salts and D$_6$glucose were obtained from Cambridge Isotope Laboratories (Andover, MA).

Wild-type *Arabidopsis thaliana* Columbia seeds (Lehle Seeds) were grown in liquid culture. Briefly, 30 mg of sterilized seeds were added to autoclaved, cooled medium in magenta cubes (Nalgene, Rochester, NY), and stratified in the dark at 4 °C for 48 h, after which the magenta cubes were placed at RT under continuous light, shaking at 100 rpm, for 12 d. Tissue was harvested by removing the plant tissue from the magenta cube with forceps, spinning vigorously for 5 s, flash freezing in liquid nitrogen, and macerating in a pre-frozen mortar and pestle. The tissue powders ($^{12}$C$_{14}$N, $^{12}$C$_{15}$N, $^{13}$C$_{14}$N, and $^{13}$C$_{15}$N) were immediately collected in separate, pre-frozen 50 mL conical tubes and stored at -80 °C.

**Extract and standard preparation.** Aliquots of ~250-350 mg homogenized powder were extracted as detailed by Fiehn$^{24}$. Briefly, each aliquot of macerate was combined with cold, degassed extraction solution (CHCl$_3$/CH$_3$OH/H$_2$O, 1:2.5:1 v/v/v), and 200 ng each of internal standards glycerol-d$_8$ (Isotec, Miamisburg, OH), benzoic-d$_5$ acid (Isotec), DL-alanine-2,3,3,3-d$_4$, and glycine-$^{13}$C$_2$. The samples were incubated with agitation at 4 °C followed by sonication in ice. Solids were pelleted by centrifugation, and the supernatant was extracted with H$_2$O. The aqueous polar and organic lipophilic
phases were separated by centrifugation. The upper polar phase was dried under reduced pressure and subsequently subjected to methoxyamination (20 mg/mL O-methoxyamine HCl in pyridine, 90 min, 30 °C) and silylation with either N-(t-butyldimethylsilyl)-N-methyltrifluoroacetamide (MTBSTFA + 1% TBDMCS; 30 min, 70 °C) or N-Methyl-(trimethylsilyl) trifluoroacetamide (MSTFA 1% TMCS; 30 min, 37 °C) (Thermo Scientific, Bellafonte, PA). Samples were incubated for 2 h at RT prior to GC/MS analysis. Authentic standards were prepared by the same methoxyamination and silylation protocols.

For relative quantitation studies, $^{12}$C$^{14}$N and $^{13}$C$^{14}$N samples were first mixed in a 1:1 (v/v) ratio to assess the overall concentration difference between the two samples. Using this information, the samples were mixed again in a concentration-adjusted 1:1 ratio. Serial dilutions of the $^{13}$C$^{14}$N sample into the $^{12}$C$^{14}$N sample were performed from the 1:1 mix to generate 1:2, 1:5, 1:10, and 1:20 $^{13}$C$^{14}$N:$^{12}$C$^{14}$N mixes.

**Gas chromatography.** GC/MS experiments were performed on a Trace GC Ultra gas chromatograph (Thermo Electron, Milan, Italy) equipped with a CTC Analytics PAL autosampler (Zwingen, Switzerland). Samples (1 µL) were injected at a split ratio of 1:20, 1:50, or 1:100, and injector temperature of 230 °C. Compounds were separated on a 30 m x 0.25 mm (ID) x 0.25 µm (df) Crossbond 5% diphenyl/95% dimethyl polysiloxane column (Restek Rxi-5Sil MS, Bellefonte, PA) with He carrier gas (1 mL/min) using one of the following oven programs: (1) 5 min isothermal at 50 °C, 80 °C/min to 80 °C, 10 °C/min to 275 °C, 80 °C/min to 320 °C, and 10 min isothermal at 320 °C; or (2) 5 min isothermal at 70 °C, 5 °C/min to 320 °C, and 5 min isothermal at 320 °C. The GC was interfaced to the MS via a transfer line held at 250 °C.

**Mass spectrometry.** Experiments were performed on a GC/Quadrupole-Orbitrap mass spectrometer developed in-house through collaboration with Thermo Fisher Scientific (Bremen, Germany). The design, construction, and characterization of this instrument are described in detail in the accompanying article$^{23}$. Briefly, a standalone Orbitrap MS (Exactive, Thermo Fisher Scientific) was coupled to a single-quadrupole GC/MS through an adapter manifold mounted in place of the Orbitrap instrument’s collision cell, connecting the GC/MS’s quadrupole to the Orbitrap’s e-trap via two flatapoles.
and split lens for ion gating. The atmospheric pressure inlet and associated ion optics of the Orbitrap MS were removed, and the HCD cell and single-quadrupole’s detector were fitted on the opposite side of the c-trap. Instrument firmware (based on Thermo Q Exactive firmware version 2.0 Build 146201) and electronics were subsequently extensively modified to permit simultaneous control of both component instruments via a single data system.

The GC/Quadrupole-Orbitrap’s ionization source was configured for either electron ionization (EI) or methane positive chemical ionization (PCI) (electron energy 70 eV, source temperature 200 °C, 2 mL/min methane reagent gas, when applicable). Unless otherwise noted, full scan MS mass spectra were acquired at a resolution (m/Δm) of 17,500 (relative to m/z 200), an automatic gain control (AGC) target of 5E5, maximum injection time of 100 ms, and scan range of 70-850 Th. The instrument was routinely tuned and calibrated using FC-43 calibrant. Mass accuracy was calibrated approximately every 24 h.

2. Molecular Ion-Directed Acquisition (MIDA)

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 PCI or EI), 2) the sample derivatization reagent (MSTFA or 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, and 7) the duration of time to exclude targets from MS/MS or SIM analysis.

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 tBDMS derivatization, 2) methane PCI with TMS derivatization, and 3) EI with tBDMS derivatization. Shown in Supplemental Figure S1A, members of each template have a set mass difference, as denoted by the arrows, from a “template initiator” ion, the lowest m/z
member of the template. To ensure the specificity of the MIDA algorithm for its intended target, most members of a template must be present (solid lines). However, since the analyte dictates the presence of certain ions, some template members are not required (dotted lines); this ensures adequate sensitivity of the algorithm. Any required member of the template can be targeted by the subsequent MS/MS analysis. For the combination of methane PCI and tBDMS, for example, the template has five ions, three required ([M - C₄H₉]⁺, [M - CH₃]⁺, and [M + H]⁺), and two optional ([M + C₂H₅]⁻ and [M + C₃H₇]⁻). The initiator ion, [M - C₄H₉]⁺, and [M - CH₃]⁺ (Δm = 42.04695 Da) correspond to the loss of a t-butyl and methyl moiety, respectively, from the tBDMS groups derivatizing the molecule. The remaining three ions result from proton transfer and adduct formation reactions during methane PCI:

\[ M + CH₅⁺ (or C₂H₅⁺) \rightarrow [M + H]⁺ + CH₄ (or C₂H₄) \]

and

\[ M + C₂H₅⁺ (or C₃H₇⁺) \rightarrow [M + C₂H₅]⁺ (or [M + C₃H₇]⁺) \]

Mass differences from the initiator are 58.07825, 86.10955, and 98.10955 Da, respectively. The two adduct ions can be of low abundance or absent, and thus are optional in this template. The species to which each template member corresponds are shown at the bottom of Supplemental Figure S1A.

Typical parameters employed for MIDA were as follows: mass tolerance of 10 ppm; minimum S/N of 100; [M + H]⁺ or [M - t-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, AGC target of 5E5, 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 and scan range were 20 Th in order to capture the entire isotopic envelope of interest.

An example of the MIDA process is presented in Supplemental Figure S1B. Prior to the start of a GC/MS analysis, the user: 1) selects the template by specifying the ionization method and sample derivatization type (here, PCI and tBDMS); 2) sets the member of the template to target ([M - C₄H₉]⁺);
and, 3) establishes the mass error tolerance (10 ppm) and S/N threshold (100) to enforce for matching the templates to the acquired MS spectra. Following the acquisition of MS spectrum #6327 at 22.28 min (step 1), the on-board instrument computer scans the selected template over the MS spectrum (step 2). At each potential “template initiator” 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 of the template are found, a dot product-based score (explained below) is calculated based on the m/z and intensity of each template member. In step 2, all of the completed templates for MS spectrum #6327 are stratified by score and m/z, with the highest-scoring template (initiating with m/z 417) shown to correctly correspond to the [M - C₄H₉]⁺ through [M + C₃H₅]⁺ ions of the current analyte (asparagine – 3TBS). In step 3, the instrument acquires MS/MS spectrum #6328 on the user-selected template member ([M - C₄H₉]⁺) of the highest scoring template match, on m/z 417. 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 (step 4). Due to 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 necessary. This process repeats throughout an analysis to yield MS and MS/MS data for nearly all eluting analytes present in the sample.

The section of a MIDA analysis shown in Supplemental Figure S1C illustrates that for each peak, the targeted ion of the template is profiled over the entirety of its elution (red traces; n.b., only a selection of MS/MS traces are shown for clarity).

The algorithm to score the templates was empirically developed, and is based on a dot-product of the m/z and intensity of each member of the template. Initially starting as a straight dot-product score, it was heavily weighted in the m/z-domain to promote the selection of the correct series of ions. The scores for templates utilizing methane PCI and EI are given by the following equations (3, 4):

\[
\text{score}_{\text{PCI}} = \sum_{j=1}^{n} \frac{1}{2} M_j^3
\]

(1)
where $I_j$ and $M_j$ are the S/N and $m/z$, respectively, of the $j^{th}$ member of $n$ total template members. The EI score was weighted more heavily in the $m/z$-domain, and multiple templates were developed (as seen in Supplemental Figure S1A, all of which are scanned over each MS spectrum) because EI spectra contain fewer ions that can direct the algorithm to the correct species in each spectrum. For EI with $t$BDMS derivatization, the algorithm is generally targeted at the $[M - C_4H_9]^+$ ion, which is characteristically the most abundant ion in the high-$m/z$ range of the spectrum (usually $>5\%$ of base peak$^1$), using the presence of a $[M - CH_3]^+$ ion and, occasionally, the molecular ion. A template defined by two peaks does not provide high specificity within a spectrum, especially with the rather common mass difference of 42 Da ($C_3H_6$). Thus, to ensure accuracy, a greater bias in the $m/z$-domain and additional templates with specificity to certain classes of analytes were required.

To assess the accuracy of the algorithm, we employed a “crowd-sourcing” technique. An offline version of the MIDA algorithm analyzed a MIDA analysis, post-acquisition, and submitted a selection of approximately 100 spectra per analysis to the user to manually annotate as correct or incorrect. For each template, two separate researchers graded three analyses and the accuracy results were averaged. This was necessary because only manual annotation is available 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.

Supplemental Figure S1D demonstrates that the attainable duty cycle with MIDA is sufficient to properly capture the elution profile of two close-eluting species (black), as well as sample both analytes’ molecular ion species with MS/MS (light and dark red). In the inset, details of the scan times and overall scan rates are given for the first, less intense, analyte (light red). Each MS/MS scan averaged about 173
ms (about 65 ms of which were attributable to additional ion accumulation time), while each MS scan averaged about 74 ms in length (ion accumulation times greater than the length of the mass analysis were negligible). Overall, this corresponded to a duty cycle of 8.1 Hz. For the more abundant second analyte (dark red), the time required for MIDA decreases.
3. Pseudocode for Molecular Ion-Directed Acquisition (MIDA) logic

```plaintext
if last scan was MS1 scan {
    Get list of mass differences specified by the pattern(s) selected by the user;
    Get all peaks from MS1 spectrum with S/N > 3;
    Sort peaks by ascending m/z;
    for each pattern to match against the spectrum {
        for each peak in the spectrum {
            Get the current peak;
            if the current peak's S/N > user's S/N threshold {
                Populate lists with minimum values and maximum values for each mass "bin" specified by the members of the pattern given the current peak's m/z and user-selected mass error tolerance;
                Save the minimum and maximum value of the first member of the current pattern;
                //The next member of the pattern must fall between the current minimum and maximum
                Add the current peak to temporary list of matched pattern members;
            } else {
                // The peak is inside a bin
                Add the peak in the bin to the list of temporary pattern members;
            }
        }
    }
    Calculate the initial score {
        if the ionization type is EI
            score = (intensity / 2) * (current peak's m/z ^ 5);
        else
            score = (intensity / 2) * (current peak's m/z ^ 3);
    }
    for each peak with an m/z > current peak m/z {
        Get the peak;
        if the peak's m/z < current minimum {
            // We haven't reached the bin yet
            Continue to the next peak in the list;
        } else {
            if the peak's m/z > current maximum {
                // We jumped over the bin
                if the current bin number is < required number of bins {
                    // The pattern is incomplete
                    Discard the temporary list of matched pattern members;
                    Break out of loop;
                } else {
                    // We still have enough members to make a pattern, so check the next bin
                    Increment the bin;
                    if there are still bins left in the pattern {
                        Update the current minimum and maximum;
                        Continue to next peak in the list;
                    } else {
                        // There are no more bins, save the temporary list
                        Add the temporary list of peaks in pattern to a master list of patterns for this spectrum if the score of this pattern is not already in the master list;
                        Break out of loop;
                    }
                }
            } else {
                // The peak is inside a bin
                Add the peak in the bin to the list of temporary pattern members;
            }
        }
    }
    Update score {
        if the ionization type is EI
            score += (intensity / 2) * (current peak's m/z ^ 5);
        else
            score += (intensity / 2) * (current peak's m/z ^ 3);
    }
}
```
Increment the bin;
if there are no bins left in the pattern after incrementing
{
    // There are no more bins, save the temporary list
    Add the temporary list of peaks in a pattern to a master
    list of patterns for this spectrum if the score of this
    pattern is not already in the master list;
    Break out of loop;
} else
{
    Update the current minimum and maximum;
}
}

if we break out of the loop AND there are still bins left from the last peak
AND the minimum number of required bins has been met
{
    // Special case of pattern falling a very end of the peak list, save the temporary list
    Add the temporary list of peaks in a pattern to a master list of patterns for this spectrum if the
    score of this pattern is not already in the master list;
}
for each temporary pattern added to the master list of patterns
{
    Add the target peak of the pattern (the one selected for MS2 or SIM as specified
    by the user) to a list of targets;
}
}

Sort the list of targets by the score of the pattern from which the target was derived;
for each data-dependent event (e.g. ddTop1 = 1, ddTop5 = 5)
{
    if there are target peaks found by the preceding algorithm
    {
        Set up a new SIM or MS2 scan using the highest scoring target peak m/z;
    } else
    {
        // No targets were found by the algorithm, do normal data-dependent mode
        Set up a new MS2 scan using the most intense peak in the spectrum;
    }
}
Add the new scan as the next scan in the instrument scan queue;
4. Supplemental Figures

4.1 Figure S1

**Figure S1.** (A) Templates for each supported combination of ionization and sample derivatization type. The mass differences between the template initiator ion (left-most ion), and each template member are shown by the black arrows. Required members of the template are denoted by solid lines, optional members by dotted lines. For each ionization/derivatization type, the associated template(s) are scanned over each scan to identify the analyte species designated below. Any required member of a template can be isolated for subsequent MS/MS or SIM. (B) The MIDA process includes four steps: 1) acquisition of an MS scan; 2) analysis of the MS spectrum using the appropriate ionization/derivatization template, and scoring of the matching templates to identify the best scoring template (boxed in red); 3) acquisition of a MS/MS or SIM scan on the target member of the highest scoring template, [M – t-butyl]+ at m/z 417; and 4) acquisition of further MS/MS or SIM scans in order of decreasing template match scores (not shown), or acquisition of the next MS scan. (C) Section of a MIDA-MS/MS analysis showing the MS profile (black), and full elution profiles of MS/MS scans (red) of the [M – t-butyl]+ of selected analytes. (D) Detail of the starred peak in C showing the MS profile (black) and MS/MS profiles for the [M – t-butyl]+ of two species (light and dark red). The inset shows the elapsed time for MS/MS (light red), and MS (black) scans over the first analyte. The overall acquisition rate is 8.1 Hz.
4.2 Figure S2

Figure S2. (A) Boxplots of mass error in parts-per-million and isotopomer abundance error (IAE) in percent for all annotated mass spectral features (n = 81). ‘M+1’, ‘M+2’, and ‘M+3’ represent the IAE for the first, second, and third isotopomers of a given monoisotopic peak, while ‘Avg’ is the average IAE for all isotopomers. (B) Relationship of mass error and IAE with the spectral signal-to-noise (S/N) of the annotated feature. Mass accuracy is independent of spectral S/N, while IAE increases for low S/N features.
4.3 Figure S3
Figure S3. Shown above in each panel are extracted ion chromatograms of the base peak (within ±5 ppm mass tolerance), and below, EI mass spectra for 31 of the identified metabolite from Arabidopsis (red) and their corresponding authentic standard (black).
**Figure S4.** Process for determining the approximate contribution of each species in an experimental isotopomer distribution using the method of least squares for over-determined systems. **(A)** An experimental distribution containing 4 distinct species ($n$) has ions at 9 $m/z$ values ($m$). The abundance at each $m/z$ value makes up the $m$-dimensional solution matrix, $b$. Each of the 4 species has a theoretical abundance distribution given by their elemental formulae. The theoretical abundance of the $n$ species at all $m$ $m/z$ values makes up the $m \times n$-dimensional matrix $A$. **(B)** Equation relating the matrix of theoretical distributions for the 4 species and the matrix of experimentally measured abundances at the 9 $m/z$ values. The combination of these 4 theoretical distributions in some proportion, denoted by the matrix $x$, yields the experimental distribution. **(C)** The formula for the least squares approximation of the matrix $x$. Solving for $x$ yields a matrix describing the approximate relative contribution of each of the $n$ species in the experimental cluster.
4.5 Figure S5

Figure S5. (A) Accuracy and precision of quantification for dilutions of the $^{13}$C$^{14}$N sample into the $^{12}$C$^{14}$N sample relative to a 1:1 mix from 28 features extracted from methane PCI full scan data. The target ratio at each dilution is denoted by a dotted grey line. (B) Effect of abundance of the $^{12}$C$^{14}$-[M - C$^{4}$H$^{8}$]$^{+}$ ion of the extracted $^{12}$C$^{14}$N: $^{13}$C$^{14}$N pair on the accuracy and precision of quantification (target dilution ratios for each dilution, i.e., 1.5, 3.0, 5.5, and 10.5, are shown as horizontal dotted lines). The accuracy and precision of quantification decrease (manifesting as over-estimated dilution ratios) for features having low abundance. Data from 28 features extracted from EI full scan analyses.
5. Supplemental References

(1) Fiehn, O.; Kopka, J.; Trethewey, R. N.; Willmitzer, L. *Anal. Chem.* **2000**, *72*, 3573-3580.