analytical chemistry

pubs.acs.org/ac Article

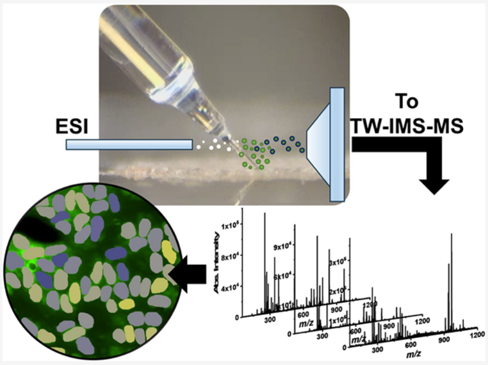
High-Throughput f-LAESI-IMS-MS for Mapping Biological Nitrogen Fixation One Cell at a Time

Marjan Dolatmoradi, Sylwia A. Stopka, Chloe Corning, Gary Stacey, and Akos Vertes*

Cite This: *Anal. Chem.* 2023, 95, 17741−17749

Read Online

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information



ABSTRACT: For the characterization of the metabolic heterogeneity of cell populations, high-throughput single-cell analysis platforms are needed. In this study, we utilized mass spectrometry (MS) enhanced with ion mobility separation (IMS) and coupled with an automated sampling platform, fiber-based laser ablation electrospray ionization (f-LAESI), for in situ high-throughput single-cell metabolomics in soybean (*Glycine max*) root nodules. By fully automating the in situ sampling platform, an overall sampling rate of 804 cells/h was achieved for high numbers (>500) of tissue-embedded plant cells. This is an improvement by a factor of 13 compared to the previous f-LAESI-MS configuration. By introducing IMS, the molecular coverage improved, and structural isomers were separated on a millisecond time scale. The enhanced f-LAESI-IMS-MS platform produced 259 sample-related peaks/cell, almost twice as much as the 131 sample-related peaks/cell produced by f-LAESI-MS without IMS. Using the upgraded system, two types of metabolic heterogeneity characterization methods became possible. For unimodal metabolite abundance distributions, the metabolic noise reported on the metabolite level variations within the cell population. For bimodal distributions, the presence of metabolically distinct subpopulations was established. Discovering these latent cellular phenotypes could be linked to the presence of different cell states, e.g., proliferating bacteria in partially occupied plant cells and quiescent bacteroids in fully occupied cells in biological nitrogen fixation, or spatial heterogeneity due to altered local environments.

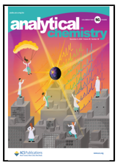
■ INTRODUCTION

Seemingly similar cells in a tissue, serving the same function, exhibit cellular heterogeneity caused by various internal and external factors, including deletion or overexpression of genes, their stage in the cell cycle, disease, or environmental perturbations.[1,2] Highly sensitive single-cell transcriptomics techniques are well positioned to explore the biochemical differences at the gene expression level. However, they only partially capture the functional characteristics of cellular processes linked to the phenotype. Single-cell metabolomics allows for identifying differences in metabolic processes resulting in functionally distinct cells.[3−5] Statistical analysis of metabolite abundances for representative cell populations can reveal latent cellular subtypes, distinct metabolic states, and rare cell types.[6−8] To obtain statistically meaningful data, large cell numbers have to be analyzed (hundreds to thousands of cells) that can only be achieved by high-throughput techniques.

Mass spectrometry (MS) is a prime tool for single-cell metabolomics due to its high sensitivity, broad molecular coverage, and quantitation capabilities.[6] For example, secondary ion MS (SIMS), matrix-free and matrix-assisted laser desorption ionization (LDI), and laser ablation electrospray ionization (LAESI) MS-based platforms with limits of detection in the range of 800 zmol to 1 fmol have shown promising results in these applications.[9] For single-cell measurements, rapid, low perturbation, and high-precision cell selection and manipulation are required.[10] For example, for free-floating cells microfluidic manipulation with electrospray ionization (ESI) MS, e.g., for proteins nanoPOTS and SCoPE MS, can be implemented.[11,12] For tissue-embedded cells, matrix-assisted laser desorption ionization (MALDI) MS enables metabolomics at cellular resolution.[13]

As plant cells are typically larger in size than animal cells, early laser sampling-based single-cell metabolomics focused on selected plant species, including *Allium cepa* (onion), *Arabidopsis thaliana*, *Egeria densa*, *Elodea canadensis*, and soybean (*Glycine max*) root nodule cells infected by green fluorescent protein (GFP)-tagged bacteria (*Bradyrhizobium japonicum*).[14−16] Capillary microsampling ESI-MS was used to study the physiological characteristics of single human hepatocytes (HepG2/C3A).[17] The throughput in these studies and in competing approaches in the literature ranges from ~2 cells/h for manual capillary microsampling ESI-IMS-MS, via

Received: August 15, 2023
Revised: October 24, 2023
Accepted: November 3, 2023
Published: November 21, 2023



https://doi.org/10.1021/acs.analchem.3c03651

62 cells/h by f-LAESI-MS using partial automation (motorized positioning in the X and Y directions only), to 1440 cells/h for single-cell printer liquid vortex capture (SCP-LVC) MS.[17−19] Recent advances in MALDI-MSI allow for an effective sampling rate of ∼30 pixels/s resulting in a more than 10-fold increase in imaging throughput. This improvement is achieved by using continuous stage movement (e.g., 5 mm/s) and uninterrupted high-repetition-rate laser pulses (up to 5 kHz) from a frequency-tripled solid-state laser.[20]

In an earlier report, we described fiber-based LAESI (f-LAESI) sampling for MS combined with a microscope that allows for the selection, ablation, and in situ analysis of tissue-embedded single plant cells.[18] To achieve a partially automated analysis of these cells, a motorized X−Y−Z translation stage and a dual-channel microscope were used. This platform was well suited for single-cell analysis owing both to low detection limits for metabolites and minimal need for sample preparation.[14]

Metabolite identification typically relies on a retention time from a separation method, accurate mass, tandem MS, and database search. Apart from capillary electrophoresis, chromatographic methods are not suited for single-cell samples.[21] Due to the severely limited sample size, tandem MS is not always possible. To help with the distinction of interfering compounds with close to identical accurate masses, ion mobility separation (IMS) can be helpful. This technique enables rapid (millisecond time scale) separation of isobaric species, enhances molecular coverage, and improves confidence in metabolite identification.[17,22,23] The earliest example of the combined application of IMS and MS in a single-cell study was looking at adenylate energy charge changes upon metabolic modulator exposure and lipid turnover rates in hepatocytes.[17] More recently, traveling wave ion mobility separation (TW-IMS)-MS with an integrated microfluidic device was used for the single-cell metabolic profiling of three cancer cell types. As a result, 73 additional sample-related peaks were detected, among which 19 distinct lipid species were differentiated based on collision cross section (CCS) value filtering.[24] In another study, IMS was used for the structural assignment of metabolites that were indistinguishable using accurate mass measurement alone. Using drift tube ion mobility separation (DT-IMS)-MS coupled to a LAESI ion source, single *A. cepa* epidermal cells were analyzed. For example, they identified cellotriose as the most likely structure for a detected trisaccharide among the six potential structural isomers. In another case, two isobaric structures, caranine (^{DT}CCS = 144.22 Å^2) and crinine/vittatine (^{DT}CCS = 153.61 Å^2) stereoisomers, were distinguished based on their drift time (DT).[23]

Metabolomic studies of plant development and growth and understanding the molecular mechanism of beneficial symbiosis, such as infections by soil bacteria, play a crucial role in crop improvement strategies. Particular attention has been given to plant growth-promoting bacteria (PGPB), in particular to nitrogen-fixing rhizobia (e.g., *B. japonicum*) interacting with legumes. Their symbiotic relationship results in the formation of root nodules, within which the exchange of nutrients between host and symbiont occurs.[25] This biological nitrogen fixation (BNF) is an important nitrogen supply route that benefits plant growth, resilience, and tolerance to abiotic stresses.

In this study, we utilized MS enhanced with IMS and coupled with a fully automated f-LAESI system for high-throughput analysis of $n > 1000$ single cells and to explore metabolic heterogeneity in soybean root nodules in their native environment. This provides statistically significant data on cellular heterogeneity in BNF, captures the shape of metabolite abundance distributions, and detects distinct metabolic states.

■ METHODS

Some of the methods (chemicals, sample preparation, bimodal microscopy, and part of f-LAESI sampling) have been described in earlier publications and in the Supporting Information.[18,26,27]

Image Processing and Automated Sampling. Image acquisition, processing, and positioning of the automated XYZ translation stage (MLS203, Thorlabs and MMP Series one, Mad City Laboratories) were performed using the MetaMorph software (Meta Series Software 7.10.2). Full automation of the stage was based on an earlier design with automated XY positioning. By adding automated positioning of the sample stage in the Z direction, the overall throughput of the system improved by a factor of ∼13. To maintain high spatial resolution and cover a large field of view (FOV), multiple adjacent image tiles were captured and stitched together. This enabled the targeting of a sufficient number of cells for statistical analysis. To maintain image quality throughout the process, autofocus was engaged for the objective. A processing pipeline was created for image segmentation, isolation of individual cells as objects, and the generation of their (x_i, y_i) centroid coordinates, where i enumerates the analyzed cells. As cells of a certain morphology were of interest, morphometric analysis was performed to find cells based on their area and/or shape factor in the image. For example, for infected nodule cells, 300−900 μm^2 area and 0.6−0.7 shape factor were applied as filtering criteria. To obtain accurate cell centroid coordinates (x_i, y_i), translation stage calibration was implemented prior to image acquisition.

The single-cell sampling platform and journals, created to drive the sample stage following the (x_i, y_i) centroid coordinates and along the Z axis, are used to position the selected cells sufficiently close to the etched fiber tip for ablation. The journal started with selecting a region of interest (ROI) within the FOV and with the identification of the cell selected for ablation. This task was followed by accurate translation to the cell centroid locations at 110 μm below the focal plane. After the stage was moved along the Z axis for focusing, bright-field/fluorescence FOV images were acquired before and after the ablation with a 500 ms delay ensuring that the stage was stabilized.

f-LAESI Sampling. The principle and implementation of f-LAESI-MS for single-cell measurements have been described in an earlier publication.[14] Briefly, 2.94 μm wavelength laser radiation produced by a Nd:YAG laser-driven optical parametric oscillator (IR Opolette 100TM, Opotek, Carlsbad, CA) was coupled into the distal end of a GeO_2-based glass optical fiber (250 μm core diameter, HP Fiber, Infrared Fiber Systems, Inc., Silver Spring, MD) using a 50 mm focal length plano-convex CaF_2 lens. The proximal end of the optical fiber stripped of polyimide coatings was placed in a 3% nitric acid solution to etch a tip with a 15 μm diameter, commensurate with the size of a single infected soybean root nodule cell. For high-throughput work, the fully automated XYZ translation stage presented each cell centroid on the (x_i, y_i) target list to the fiber tip for efficient ablation.[18]

TW-IMS-MS and Data Analysis. Mass spectra were collected in positive- and negative-ion modes on a TW-IMS-MS instrument (Synapt G2-S, Waters Co., Milford, MA) in the m/z 50−1200 range. The mass analyzer was calibrated by 0.5 mM sodium formate solution in 9:1 2-propanol/water (v/v), whereas for CCS calibration, poly-DL-alanine oligomers with residue numbers ranging from 3 to 14 were used. The generated calibration file and $^{TW}CCS_{N2}$ values for the standard were automatically generated using DriftScope 2.8 (Waters Co., Milford, MA). Accurate masses, CCS values, isotope distribution patterns, and tandem MS spectra were used for metabolite identification. The typical mass accuracy for our Synapt G2-S in the relevant mass range is <5 mDa. To achieve a better limit of detection necessary for single-cell measurements, the instrument was operated in sensitivity mode. As a consequence, in some cases, overlapping peaks degraded the mass accuracy.

Single-cell spectra were acquired by MassLynx (version 4.1, Waters Co., Milford, MA) and processed by mMass for deisotoping and extraction of sample-related peaks. To find potential matches for newly detected metabolites, PlantCyc (https://www.plantcyc.org/data/search) and BioCyc (https://biocyc.org/data/search) databases were searched. An in-house CCS library for metabolites, PNNL (https://metabolomics.pnnl.gov/) and AllCCS (http://allccs.zhulab.cn/) databases, the latter containing ∼2 million measured CCS values for small molecules and ∼12 million predicted CCSs, were queried for matches with the measured values. For the tentative identification of the ions, tandem MS was also performed with collision-induced dissociation (CID) in an Ar background gas. To utilize the acquired tandem MS data, the measured fragmentation patterns were searched against the MoNa database (https://mona.fiehnlab.ucdavis.edu/, last accessed June 27, 2023).

Statistical Analysis. A total of 1614 root nodule cells were analyzed in seven acquisitions from seven tissue sections (section 1: 643 cells, section 2: 150 cells, section 3: 192 cells, section 4: 289 cells, section 5: 165 cells, section 6: 101 cells, and section 7: 74 cells). After background subtraction and deisotoping, $S/N > 3$ was required for peak detection throughout the study. The absolute ion intensities of the sample-related peaks in each acquisition were extracted. Prior to normalization, a quality control was applied to the spectra. The metric for quality control was the sum of the sample-related peak intensities. Spectra with a sum lower than a threshold, i.e., 5% of the sum for the strongest spectrum in the acquisition, were rejected. For spectral features missing in some of the spectra, e.g., because the related metabolite concentration was below the limit of detection, a missing value imputation strategy was selected. According to the literature, for metabolomics data, the most appropriate approach was found to be the quantile regression imputation of left-censored data (QRILC) in the missing not at random (MNAR) category.[28] This approach was implemented for our data using a web-based tool, MetImp, developed for MS-based metabolic profiling (https://metabolomics.cc.hawaii.edu/software/MetImp/). To minimize the distortion of the abundance distributions, a 20% limit was established for the number of missing peak intensity values for each ion.

The objective of ion intensity normalization is to factor out many of the experimental variables not related to biological variations (e.g., variations in the electrospray current) but to retain biological differences in metabolite levels. For ion intensity normalization, the following three approaches were tested: (a) normalization using the base peak intensity, (b) normalization based on the intensity of an ion with low variance for the cell population (e.g., disaccharide), and (c) normalization by the sum of the sample-related peak intensities. Comparison of these approaches for our data revealed the latter to be the most reliable and least dependent on the intensities of individual spectral features. This conclusion is reasonable because of the three options, the sum of the sample-related peak intensities is least dependent on individual ion intensity variations.

To visualize the abundance distributions of the normalized ion intensities, histograms and violin plots were constructed using the Origin data analysis package (OriginPro 2019, OriginLab Co., Northampton, MA). To capture the metabolite abundance variations for a unimodal distribution, metabolic noise values, η_i^2, defined as $\eta_i^2 = \sigma_i^2/\mu_i^2$, were calculated, where μ_i is the mean and σ_i is the standard deviation of the ion intensities for the ith metabolite. The distributions were approximated using normal and log-normal models. For the metabolites with bimodal abundance distributions, nonlinear deconvolution, based on normal and/or log-normal models, was used to separate the subpopulations.

RESULTS AND DISCUSSION

High-Throughput Single-Cell Analysis. To explore metabolic heterogeneity in the root nodule cells, the enhanced f-LAESI-IMS-MS system is employed. The main difference between the optimized and previous f-LAESI designs is the newly added ability to automate the Z stage movement with MetaMorph software. It is critical to bring the fiber tip sufficiently close to the cell surface for efficient cell ablation without piercing the cell wall. To optimize the distance between the fiber tip and the sample surface on the micrometer scale, a feedback loop from the autofocusing was implemented. By parking the fiber tip slightly above the focal point and bringing the cell surface into focus, a reproducible tip-to-surface distance can be achieved. To improve the uniformity of cell ablation, we controlled the humidity in the sample region throughout the experiment.

Between ablating cells i, in position (x_i, y_i), and $i + 1$, in position $(x_i + 1, y_i + 1)$, the stage was lowered by 110 μm to prevent breaking of the fiber tip. Then, the sample was lifted to bring the new cell under the fiber tip for ablation. To confirm the success of targeting single cells, image acquisition before and after each ablation was also performed. Mass spectra were collected after the ablation, and the entire process was repeated until all selected cells were analyzed. The entire process, including stage movement, cell ablation, image acquisition, collecting the mobiligram, and mass spectrum, took ∼4 s/cell, resulting in an anticipated ∼900 cells/h throughput.

Spectra were recorded from 1614 cells. After the application of quality control, 1074 spectra remained to study metabolic heterogeneity. After applying missing value imputation and removing outliers detected for each metabolite, the following number of spectra remained: 1074 for hexose phosphate and disaccharide, 1061 for phosphatidic acid (PA), 1041 for gluconic acid, 1000 for soyasaponin ag, and 1000 for PIP. To verify that the effect of missing value imputation on the abundance distributions is negligible, the shapes of the histogram for hexose phosphate were compared with and without imputation in Figure S1. It was clear from this comparison that the missing value imputation did not affect

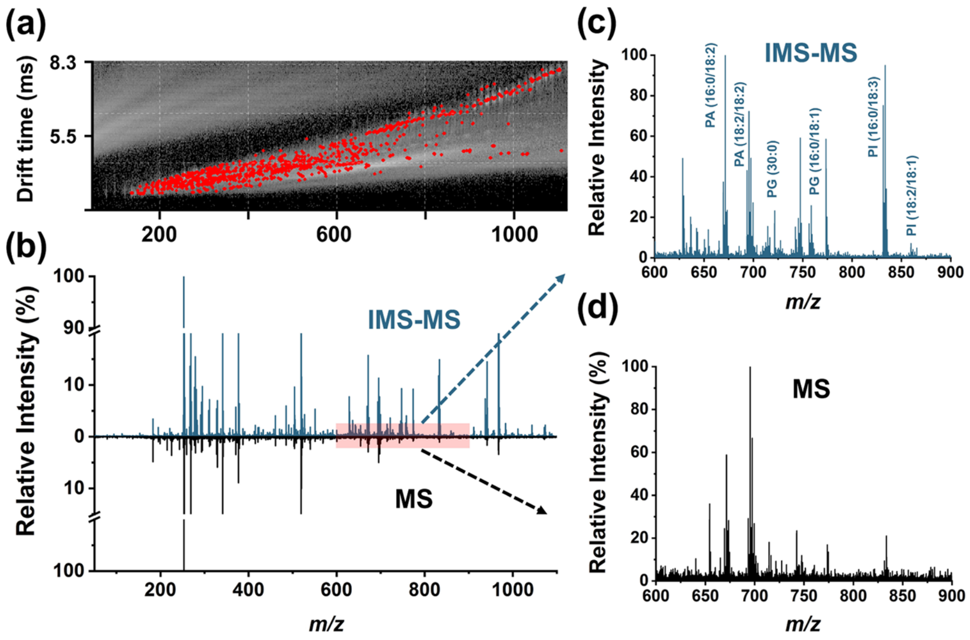


Figure 1. (a) Mobiligram from negative-ion mode by f-LAESI-IMS-MS of a single soybean root nodule cell. (b) Corresponding mass spectrum with IMS (blue) and an f-LAESI-MS mass spectrum without engaging IMS (black). The top (blue) spectrum, taken by f-LAESI-IMS-MS, showed improved molecular coverage with 259 sample-related peaks, whereas the bottom (black) mass spectrum, produced by f-LAESI-MS, exhibited only 131 peaks. (c, d) Mass spectra with and without IMS, respectively, zoomed to the *m*/*z* 600−900 range. They illustrate the enhancement in lipid molecular coverage using IMS, including improved S/N for PAs, PGs, and PIs in (c).

the bimodality of the distribution and did not result in an artifact. As a safeguard, the number of imputed missing values was always kept below 20% of the number of analyzed cells.

A realistic measure of the throughput can be calculated by dividing the number of analyzed cells in a single acquisition, e.g., 643 cells, by the corresponding analysis time, i.e., 48 min, resulting in a throughput of 804 cells/h (see Figure S2). Compared to the previously reported 62 cells/h throughput for high cell numbers (1084 cells were analyzed in ∼17.5 h in that study),[18] the acquisition rate was improved by a factor of ∼13. In addition, by automating the vertical stage movement, vibrations introduced by manual operation were eliminated, and the possibility of systematic errors due to co-sampling adjacent cells was minimized. Further throughput enhancements resulted from removing unnecessary delays in stage movement. In comparison, single-cell metabolomics of cultured adherent cells, performed by microscopy-guided MALDI-MSI, demonstrated >1000 cells/h throughput.[29] Although this is slightly higher throughput than the >800 cells/h demonstrated in the present study, f-LAESI-IMS-MS is performed on frozen tissue sections with very little sample preparation and under in situ conditions, while the MALDI-MSI technique requires cell dehydration, the elaborate application of a matrix and vacuum environment. For 30 000 rat cerebellar cells dissociated from tissue, MALDI-MSI was used and a high sampling rate was achieved.[30]

Enhanced Molecular Coverage in Infected Cells by IMS. In single-cell metabolomics of plant cells, the sampled subcellular volume is in the low nanoliter to subnanoliter range.[18] For such small sample sizes, analysis by MS is expected to provide incomplete molecular coverage. This can be demonstrated by comparing the number of spectral features for different ablation spot sizes. In LAESI-MS using conventional focusing with an ablation spot diameter of ∼300 μm (ablating multiple cells), 797 spectral features were observed, whereas with optical fiber-based ablation resulting in an ∼30 μm ablation spot (ablating a single cell), 125 spectral features were detected.[22,26] Past experience demonstrated that introducing IMS improves the signal-to-noise ratio (S/N) and enhances molecular coverage in metabolomics.[31] Thus, single-cell f-LAESI-IMS-MS measurements were performed with a TW-IMS device. The results were represented in mobiligrams that map ion intensities on a grayscale as a function of drift time (DT) and *m*/*z* values (see, e.g., Figure 1a). The drift times for the detected ions separated in the nitrogen buffer gas were converted into $^{\mathrm{TW}}\mathrm{CCS}_{\mathrm{N2}}$ values. Comparison of these values with CCS libraries facilitated metabolite assignment.[32]

As shown in Figure 1b, by incorporation of IMS, the number of sample-related spectral features is significantly increased. Careful analysis of single-cell mass spectra indicated that with IMS 259, spectral features were detected with an S/N > 3 cutoff (top spectrum in Figure 1b), whereas only 131 peaks were present in the mass spectra without IMS (bottom spectrum in Figure 1b). The spectra were collected on multiple pairs of cells. Despite the metabolite abundance variations observed within the cell population, a similar number of sample-related peaks were observed for the mass spectra that passed the quality control step. For a clear comparison in Figure 1, spectra with absolute intensities >10^5 were presented. For selected spectral features, tentatively assigned metabolites, measured accurate *m*/*z* and CCS values, and metabolomics standards initiative (MSI) identification levels are listed in Table S1 (see the Supporting Information).[33] According to the MSI, the most rigorous (level 1) identification can be achieved by, e.g., measuring accurate masses and tandem mass spectra. Two examples of tandem MS (at 30 eV collision energy) for *m*/*z* 253.051 identified as dihydroxyflavone and for *m*/*z* 695.464 identified as (PA(18:2/18:2)) are shown in Figure S3.

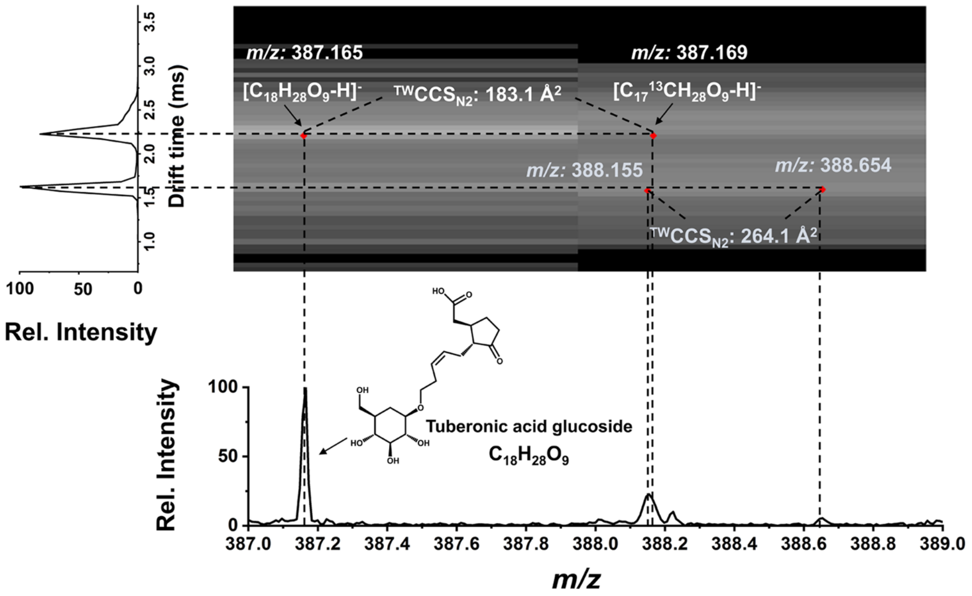


Figure 2. Detection of a doubly charged species in a single-cell measurement at *m/z* 388.155 obscured by ^{13}C isotopic peak at *m/z* 388.169 of another metabolite using IMS-MS. The extracted drift time profile for *m/z* 388.16 exhibits two peaks at 1.6 ± 0.1 and 2.2 ± 0.2 ms. The peak with $^{TW}CCS_{N2}$ = 183.1 Å^2 was identified as the ^{13}C isotope peak of tuberonic acid glucoside detected at *m/z* 387.165, whereas the peak with $^{TW}CCS_{N2}$ = 264.1 Å^2 appeared to be a yet unidentified doubly charged species followed by its ^{13}C isotopic peak at *m/z* 388.657.

Figure 1c,d depicts the *m/z* 600−900 region of the two spectra and highlights the difference in the number of lipid peaks, showing 34 peaks with and 15 peaks without IMS, respectively. Previous[22,26] and current tandem MS measurements on single-cell level (data not shown), in combination with the acquired CCS values, helped us identify some of the lipids as phosphatidic acids (PAs), phosphatidylglycerols (PGs), and phosphatidylinositols (PIs) (see Figure 1c). The importance of the heterogeneity of lipids at single-cell resolution has recently been demonstrated on a large scale.[30]

The benefit of improved S/N is demonstrated in Figure S4. According to the mobiligram (Figure S4a), there are five ionic species present in this region, four of them with indistinguishable drift times. The corresponding mass spectrum in the top part of Figure S4b presents the five ions with acceptable S/N values of greater than 3. Palmitoleate was assigned to the peak at *m/z* 253.224 with an Δ*m/z* of 6.3 mDa and an experimental ΔCCS of 5.5 Å^2. Although this metabolite had previously been found in the peribacteroid membrane surrounding nitrogen-fixing rhizobia,[34] this has never been detected with single-cell resolution, showing the significance of IMS for molecular coverage.

For the single-cell mass spectrum without IMS in the bottom part of Figure S4b, only a single peak at *m/z* 253.051 corresponding to the deprotonated form of dihydroxyisoflavone was observed. This metabolite had previously been reported as a key flavonoid present in root exudates and is demonstrated to be a critical infection signal.[35]

In addition, the feasibility of revealing spectral interferences by IMS-MS has been demonstrated. For example, in the mass spectrum in Figure 2, the monoisotopic peak of a metabolite ion is masked by the (M + 1) ^{13}C isotope peak of another metabolite. At *m/z* 388.169, where the ^{13}C isotope peak for *m/z* 387.165 appeared, two ions with $^{TW}CCS_{N2}$ values of 183.1 and 264.1 Å^2 were observed. Although the former is related to the metabolite detected at *m/z* 387.165 (tuberonic acid glucoside, predicted Ω = 190.1 Å^2), the latter allows for uncovering the presence of a doubly charged species. However, by searching *m/z* = 388.155 as $[M - 2H]^{2-}$ ionic species against the BioCyc and PlantCyc databases, no potential hits were found.

Peak annotation based on accurate mass alone in untargeted metabolomics is usually not possible, especially for those ions with marginal signal intensities that are not amenable to tandem MS. To facilitate ion identification, IMS can be incorporated to determine the CCS values for the unknown ions. For example, for the peak observed at *m/z* 252.078 (not shown), annotation as neopterin with a calculated *m/z* of 252.074 (Δ*m/z* = 4.0 mDa) can be supported by comparing the measured $^{TW}CCS_{N2}$ value of Ω = 146.9 Å^2 with the CCS reported for neopterin in the AllCCS database, Ω = 150.3 Å^2. The difference of 3.4 Å^2 is within the experimental uncertainty of the measurement. It is important to recognize that there are many other ions with CCSs very close to the measured values limiting the utility of IMS in peak annotations. Some additional help for annotations can be found in the literature. Neopterin, a catabolic product of guanosine triphosphate (GTP), had been previously detected by high-performance liquid chromatography (HPLC) in leguminous plants and bacteria including *B. japonicum*. It is thought to be released in the host upon infection by intracellular living bacteria.[36]

Another application of IMS in single-cell metabolomics is its potential to resolve structural isomers that appear as a single peak in the mass spectrum. For example, in the mobiligram seen in Figure S5a, two ions registered at *m/z* 449.101 with differing cross sections of CCS_1 = 189.6 Å^2 and CCS_2 = 197.9 Å^2 were detected (see Figure S5a). In the AllCCS database, there is one experimental CCS value in this range at CCS_{ast} = 198.3 Å^2 with *m/z* 449.108 (Δ*m/z* = 7.0 mDa) corresponding to astilbin (a flavanone glycoside) with an elemental formula of $C_{21}H_{22}O_{11}$.[37] Compatible computationally predicted CCS values in the database include its two structural isomers, glucopyranosyl tetrahydroxyflavanone with a CCS_{gpthf} of 194.7 Å^2, and tetrahydroxyflavanone glucoside with a CCS_{thfg} of

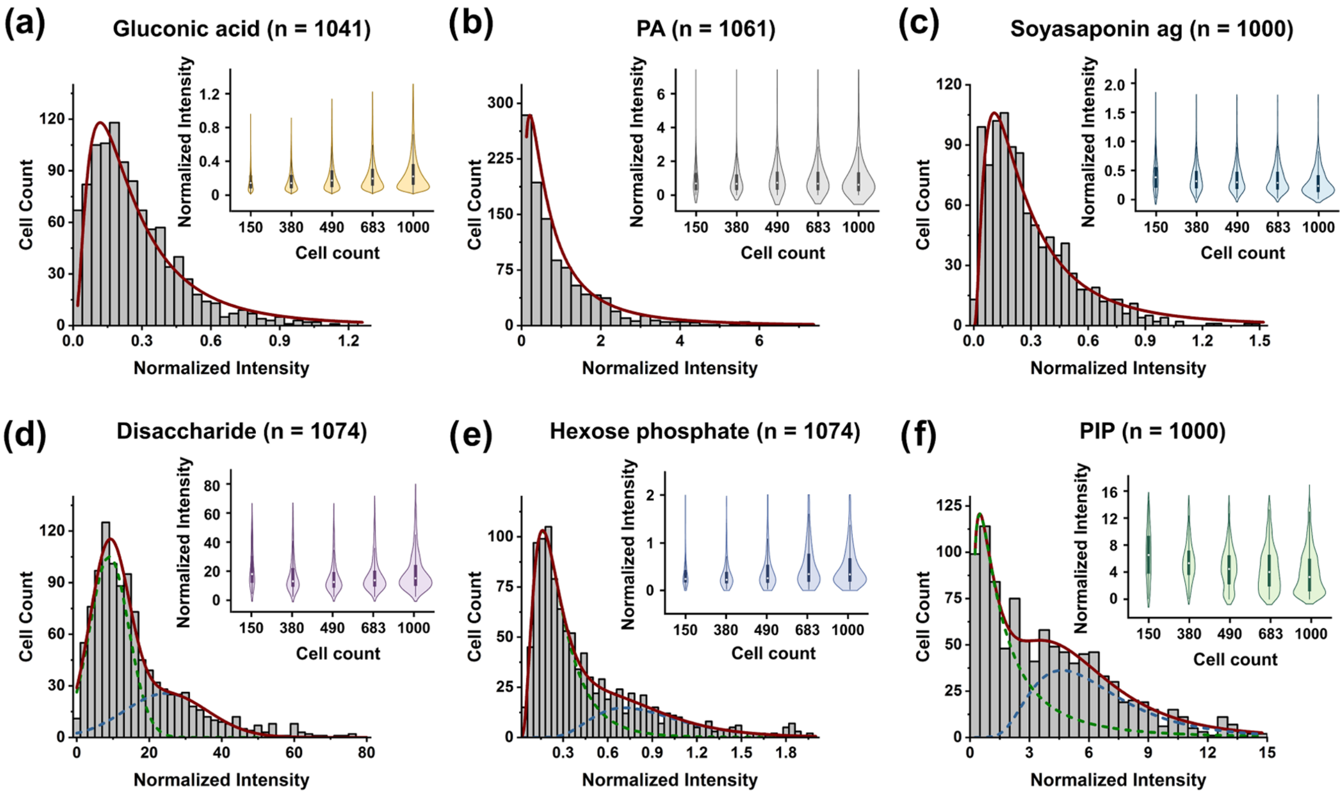


Figure 3. Distribution of normalized metabolite ion abundances for $n > 1000$ soybean root nodule cells along with violin plots for various cell numbers. While the distributions in the first row are log-normal, the distributions in the second row are bimodal. As it is demonstrated by the violin plots, raising the number of analyzed cells from 150 to 1000 increases the chance of detecting the second subpopulation.

199.9 Å^2. Based on past experience with determining CCS values for standards, we estimate that the accuracy of our CCS data is within ±5 Å^2. Although the mobiligram in Figure S5a shows the presence of two structural isomers, it appears that the measured CCS values and the CCS values in the database are too close to assign structures to them. Going forward, increasing the accuracy of the IMS might enable structural assignments.

For additional characterization of the two detected structures, tandem MS with CID was performed at 30 eV for the m/z 449.101 precursor ion, and fragments with m/z 287.079, 269.069, 259.059, and 243.088 were observed (see Figure S5b). The fragment with m/z 287.079 might be explained by the cleavage of the glycosidic bond resulting in fragments Y_0 or Z_0 for tetrahydroxyflavanone glucoside or astilbin molecular structures, respectively.[38] Breaking the C−C bond connecting the glucose to the aglycone in glucopyranosyl tetrahydroxyflavanone is less likely than breaking the C−O bond in the other two structures. The rest of the detected ions are formed by the loss of H_2O, CO, and CO_2 from the Y_0 or Z_0 ions. Similar fragmentation patterns of negative flavanone aglycone ions were observed in the literature.[39] Our observations for the accurate masses of the precursor and fragment ions were also consistent with reference data for the fragmentation of 3,5-dihydroxy-2-(4-hydroxyphenyl)-7-[3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxy-2,3-dihydrochromen-4-one at 20 eV collision energy provided in the MoNa database (https://mona.fiehnlab.ucdavis.edu/spectra/display/VF-NPL-QTOF003274, last accessed June 27, 2023).

In summary, the advantages of IMS in single-cell measurement include (1) reducing the complexity of the mass spectrum by decreasing the background noise and increasing the signal-to-noise ratio for the ions of interest, (2) separating structural isomers and close to isobaric species that have different CCSs, and (3) diminishing the impact of spectral interferences caused by species that are highly abundant and have similar m/z values to low-abundance ions. These three advantages helped to improve molecular coverage and detect low-abundance species.

Cellular Heterogeneity of Metabolite Abundances. Previous assessment of the cell numbers required to accurately capture the shape of metabolite abundance distributions for a system with two subpopulations revealed that at least $n > 200$ was necessary. Increasing the number of analyzed cells to $n \approx 1000$ further improved the fidelity of the data.[18] In this study, $n > 1000$ soybean root nodule cells were analyzed to capture the metabolic heterogeneity within cell populations in the infection zone.

Fluorescence images of the tissue sections were segmented to identify the infected cells, and morphometric analysis was used to select the cells targeted for analysis in seven separate acquisitions. The histograms of normalized abundance distributions for six metabolites, gluconic acid, PA (18:2/18:2), soyasaponin ag, disaccharide, hexose phosphate, and phosphatidylinositol phosphate (PIP) (20:1/18:2), are shown in Figure 3. While the first three metabolites exhibited log-normal distributions with correlation coefficients of $r^2 > 0.99$, the latter three indicated bimodal distributions suggesting the presence of two subpopulations. Due to the extended tail present in log-normal distributions, the median value, μ^*, and the shape parameter, σ^*, instead of the mean, μ, and standard deviation, σ, values were used. For the three metabolites with single log-normal ion abundance distributions, the median and shape parameters were $\mu^* = 0.23 \pm 0.02$ and $\sigma^* = 0.82 \pm 0.08$ for gluconic acid, $\mu^* = 0.72 \pm 0.02$ and $\sigma^* = 1.32 \pm 0.01$ for PA, and $\mu^* = 0.25 \pm 0.001$ and $\sigma^* = 0.98 \pm 0.003$ for soyasaponin ag.

Normal distributions were used to deconvolute the two components present in the disaccharide distribution (Figure 3d). Nonlinear fitting resulted in mean values of $\mu_1 = 8.96 \pm 0.01$ and $\mu_2 = 24.62 \pm 0.20$, and standard deviations of $\sigma_1 = 5.38 \pm 0.02$ and $\sigma_2 = 11.43 \pm 0.13$ for the two components, respectively, with $r^2 > 0.99$. For the deconvolution of the hexose phosphate bimodal distribution, two log-normal distributions were selected as the underlying model because, compared to using normal distributions, a higher r^2 value ($r^2 = 0.98$) was achieved. Performing the deconvolution using a nonlinear fitting algorithm resulted in $\mu_1^* = 0.23 \pm 0.004$ and $\mu_2^* = 0.85 \pm 0.04$, and $\sigma_1^* = 0.60 \pm 0.01$ and $\sigma_2^* = 0.43 \pm 0.05$ for the underlying components. Similarly, for PIP, two log-normal distributions were used for modeling resulting in a good fit with $r^2 = 0.96$. The parameters for the two components were $\mu_1 = 1.52 \pm 0.12$, $\mu_2 = 5.80 \pm 0.28$, $\sigma_1 = 1.06 \pm 0.05$, and $\sigma_2 = 0.48 \pm 0.04$. In all three cases, fitting the data with a single log-normal distribution resulted in lower r^2 correlation coefficients ($r^2 = 0.95$ for disaccharide, $r^2 = 0.94$ for hexose phosphate, and $r^2 = 0.84$ for PIP) and significantly higher mean square weighted deviation (MSWD) values. In addition to the ion abundance distributions, violin plots have been generated, and they are shown as insets in Figure 3a−f. As the number of analyzed cells was increased from $n = 150$ to 1000, the violin plots for disaccharide, hexose phosphate, and PIP revealed the presence of two components.

Based on this analysis, the presence of two metabolically distinct subpopulations was established in the analyzed root nodule cells. They are characterized with lower and higher levels of disaccharide, hexose phosphate, and PIP in the cells, resulting from differences in metabolic fluxes and pathway activities.

In a recent study, by integrating single-nucleus and spatial transcriptomics the expression patterns of specifically expressed genes in infected cells of soybean root nodules have been investigated.[40] The comparison of gene expression levels, e.g., of symbiotic nitrogen fixation genes, between mature nodules (21 dpi) and immature nodules (12 dpi), suggested the presence of two transcriptionally distinct subcell types within the infection zone. Their findings, however, did not demonstrate differences at the metabolite level.

Using scanning electron microscopy imaging, our previous study demonstrated that at 21 dpi, there were two types of infected nodule cells that were either partially or fully occupied by rhizobia.[26] Single-cell metabolomics results indicated that certain metabolite (e.g., NAD^+) abundances exhibited bimodal distributions corresponding to two metabolic states. We hypothesized that such states could be explained by the existence of proliferating bacteria in the partially occupied cells and by the presence of nondividing quiescent bacteroids in the fully occupied cells.

Our results indicated that hexose phosphate levels also exhibited a bimodal abundance distribution. Hexose phosphates have a crucial role in the glycolysis pathway and metabolic processes that contribute to nitrogen fixation.[41] One can attribute the high and low hexose phosphate levels to the presence of infected cells that are partially and fully occupied by rhizobia, respectively. Similar observations were made for the bimodal disaccharide and PIP abundance distributions. Disaccharide levels as an intermediate energy reserve must be elevated to support the growth of proliferating bacteria, whereas quiescent bacteria require lower levels of this metabolite. Bimodal abundance distributions for lipids associated with plasma membranes can also be rationalized in correlation with proliferating and quiescent rhizobia.

For unimodal abundance distributions, metabolic heterogeneity can be characterized by the amplitude of the metabolic noise. To determine the biological component of the noise, first the technical noise (η_t^2) had to be measured. For 2.0 μL aliquots of 500 μM sucrose solution, f-LAESI-MS analysis was carried out under the conditions used for single-cell analysis. Intensities of the relevant ions from 25 replicates yielded $\mu_t = 34.49$ and $\sigma_t = 2.81$, resulting in $\eta_t^2 = 0.08$. Metabolic noise (η_m^2) values were determined by correcting the measured noise (η^2) for the technical noise ($\eta_m^2 = \eta^2 - \eta_t^2$) and were found to be $\eta_m^2 = 0.51$ for gluconic acid ($\mu_m = 0.26$), $\eta_m^2 = 1.23$ for PA ($\mu_m = 0.92$), and $\eta_m^2 = 0.53$ for soyasaponin ag ($\mu_m = 0.29$). Two of these values were close to the previously reported metabolic noise for the same species (i.e., $\eta_m^2 = 0.67$ for gluconic acid and $\eta_m^2 = 1.88$ for PA).[26]

For metabolites with bimodal distributions, deconvolution resulted in two metabolic noise values for the two subpopulations. For the subpopulations for disaccharide, $\eta_{m_1}^2 = 0.21$ ($\mu_{m_1} = 8.96$) and $\eta_{m_2}^2 = 0.12$ ($\mu_{m_2} = 24.62$); for hexose phosphate, $\eta_{m_1}^2 = 0.27$ ($\mu_{m_1} = 0.28$) and $\eta_{m_2}^2 = 0.04$ ($\mu_{m_2} = 0.85$); and for PIP, $\eta_{m_1}^2 = 0.95$ ($\mu_{m_1} = 2.14$) and $\eta_{m_2}^2 = 0.11$ ($\mu_{m_2} = 6.11$) were observed, where m_1 and m_2 stand for the subpopulations with the lower and higher mean abundance values, respectively. This data shows that higher mean values correspond to somewhat lower metabolic noise similar to earlier observations for metabolic noise.[18,26]

Overall, these results illustrate the ability of fully automated f-LAESI-IMS-MS to capture cellular heterogeneity directly from tissue-embedded single cells with enhanced throughput. Variations in metabolite abundances for $n > 1000$ infected root nodule cells were determined, and their heterogeneity was characterized. For unimodal distributions, a new metric, metabolic noise, was introduced. This quantity can be linked to the tightness of the control over the concentration of a given metabolite in the cell. Bimodal abundance distributions can be linked to the presence of two cellular subpopulations corresponding to low and high mean concentrations for a given metabolite.

■ CONCLUSIONS

Using a fully automated f-LAESI ion source, optimized for high-throughput in situ single-cell metabolomics, in combination with IMS-MS, metabolic heterogeneity of infected cells within soybean root nodules was investigated. The addition of an automated *Z* stage to the sampling platform with *XY* automation enabled an ∼13 times higher sampling rate (804 cells/h), resulting in increased analyzed cell numbers ($n > 1000$), beneficial for statistical analysis. The improved statistics allowed better characterization of cellular metabolic heterogeneity, including the identification of metabolic noise, as well as detection of distinct metabolic states and the corresponding hidden phenotypes. Although with MALDI-MSI a somewhat higher sampling rate (>1000 cells/h) has been achieved,[29] the required elaborate sample preparation, matrix application, and vacuum environment hinder in situ studies by this technique.

With this new system, we were able to gauge metabolic heterogeneity and show its relationship with biological nitrogen fixation. Unimodal and bimodal metabolite abundance distributions were captured and characterized by metabolic noise values for the former and deconvoluted distributions for the two subpopulations for the latter. We hypothesized that the metabolic states corresponding to these two subpopulations were defined by the proliferating bacteria in partially occupied plant cells and the quiescent bacteria in fully occupied cells. The findings about heterogeneity and the generated hypothesis capture new biology not available using conventional bulk measurements.

To enhance the molecular coverage observed in single-cell metabolomics, we utilized TW-IMS. Incorporation of this separation step reduced spectral interferences, allowed the distinction of structural isomers, and enhanced the S/N ratio, overall, resulting in a 2 times higher molecular coverage. The general utility of this approach is limited by the relatively low ~50 resolving power exhibited by the TW-IMS stage in our Synapt G2-S. The recent introduction of structures for lossless ion manipulation (SLIM) IMS technique with an estimated separation power up to ~1860 promises significant advances in the differentiation of isobaric ions yielding higher molecular coverage in single-cell metabolomics.[42]

Current limitations of the f-LAESI-IMS-MS approach include the inability to analyze single cells significantly smaller than 30 μm in diameter (e.g., most mammalian cell types) due to sensitivity constraints. In large part, this limitation is the result of atmospheric pressure sampling. Competing methods based on MALDI-MS are not burdened with this challenge, as for that technique, the sampling and ionization are performed in vacuum. Moving forward, improved ion collection at the atmospheric pressure LAESI interface and increased sensitivity of the IMS-MS system will enable the targeting of smaller cell sizes and broaden the application of in situ single-cell f-LAESI for the analysis of mammalian cells.

In spatial metabolomics, the location of the analyzed cells within the tissue is preserved. This added information enhances our understanding of the cellular interactions in the tissue. By acquisition of centroid coordinates for the analyzed cells from the optical image along with the metabolite abundances, f-LAESI-IMS-MS is a candidate for spatial metabolomics studies. This feature confers special benefits for exploring the dynamics of small-molecule transport in plant tissues, including their movement between the root and the infection zone within the nodule in BNF.

Emerging techniques for single-cell metabolomics, in combination with single-cell proteomics and/or transcriptomics, can form the basis of single-cell multiomics approaches. Ideally, the different modalities of this analysis are performed in the same cell. As most single-cell methods are destructive, this capability is currently not available. It is, however, possible to analyze highly similar cell populations by different single-cell techniques and explore potential correlations. To perform multiomics experiments on a single cell requires the development of reliable subcellular sampling. Such methods can be based on capillary microsampling, laser capture microdissection, or carefully controlled laser-ablation-based sampling approaches. Ultimately, simultaneous identification of transcript expression levels and protein and metabolite abundances from the same cell can contribute to the system-level understanding of cellular function.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.3c03651.

Effect of missing-data imputation, ion chromatogram for single-cell analysis, tandem mass spectra, improved S/N due to IMS, detection of structural isomers by IMS-MS, and ion annotation table (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Akos Vertes − *Department of Chemistry, The George Washington University, Washington, District of Columbia 20052, United States;* orcid.org/0000-0001-5186-5352; Phone: +1 (202) 994-2717; Email: vertes@gwu.edu; Fax: +1 (202) 994-5873

Authors

Marjan Dolatmoradi − *Department of Chemistry, The George Washington University, Washington, District of Columbia 20052, United States*

Sylwia A. Stopka − *Department of Chemistry, The George Washington University, Washington, District of Columbia 20052, United States; Department of Neurosurgery, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, United States;* orcid.org/0000-0003-3761-6899

Chloe Corning − *Department of Chemistry, The George Washington University, Washington, District of Columbia 20052, United States*

Gary Stacey − *Divisions of Plant Sciences and Biochemistry, Christopher S. Bond Life Sciences Center, University of Missouri, Columbia, Missouri 65211, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.3c03651

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The research reported in this publication was in part supported by the U.S. National Science Foundation, Plant Genome Program, Division of Integrative Organismal Systems, Award Number: IoS-1734145. Additional support was provided for M.D. as a Graduate Research Assistant by the Columbian College of Arts and Sciences of The George Washington University.

■ REFERENCES

(1) Paszek, P.; Ryan, S.; Ashall, L.; Sillitoe, K.; Harper, C. V.; Spiller, D. G.; Rand, D. A.; White, M. R. H. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 11644−11649.

(2) Zhang, L. W.; Vertes, A. *Angew. Chem., Int. Ed.* **2018**, *57*, 4466−4477.

(3) Taniguchi, Y.; Choi, P. J.; Li, G. W.; Chen, H. Y.; Babu, M.; Hearn, J.; Emili, A.; Xie, X. S. *Science* **2010**, *329*, 533−538.

(4) Rubakhin, S. S.; Romanova, E. V.; Nemes, P.; Sweedler, J. V. *Nat. Methods* **2011**, *8*, S20−S29.

(5) Strzelecka, P. M.; Ranzoni, A. M.; Cvejic, A. *Dis. Model. Mech.* **2018**, *11*, No. dmm036525, DOI: 10.1242/dmm.036525.

(6) Dolatmoradi, M.; Samarah, L. Z.; Vertes, A. *Anal. Sens.* **2022**, *2*, No. e202100032, DOI: 10.1002/anse.202100032.

(7) Armbrecht, L.; Dittrich, P. S. *Anal. Chem.* **2017**, *89*, 2−21.

(8) Oikawa, A.; Saito, K. *Plant J.* **2012**, *70*, 30−38.

(9) Taylor, M. J.; Lukowski, J. K.; Anderton, C. R. *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 872−894.

(10) Kumar, R.; Ghosh, M.; Kumar, S.; Prasad, M. *Front. Microbiol.* **2020**, *11*, No. 1152, DOI: 10.3389/fmicb.2020.01152.

(11) Zhu, Y.; Clair, G.; Chrisler, W. B.; Shen, Y. F.; Zhao, R.; Shukla, A. K.; Moore, R. J.; Misra, R. S.; Pryhuber, G. S.; Smith, R. D.; Ansong, C.; Kelly, R. T. *Angew. Chem., Int. Ed.* **2018**, *57*, 12370−12374.

(12) Budnik, B.; Levy, E.; Harmange, G.; Slavov, N. *Genome Biol.* **2018**, *19*, No. 161, DOI: 10.1186/s13059-018-1547-5.

(13) Ong, T. H.; Kissick, D. J.; Jansson, E. T.; Comi, T. J.; Romanova, E. V.; Rubakhin, S. S.; Sweedler, J. V. *Anal. Chem.* **2015**, *87*, 7036−7042.

(14) Shrestha, B.; Vertes, A. *Anal. Chem.* **2009**, *81*, 8265−8271.

(15) Zhang, L. W.; Foreman, D. P.; Grant, P. A.; Shrestha, B.; Moody, S. A.; Villiers, F.; Kwake, J. M.; Vertes, A. *Analyst* **2014**, *139*, 5079−5085, DOI: 10.1039/C4AN01018C.

(16) Stopka, S. A.; Khattar, R.; Agtuca, B. J.; Anderton, C. R.; Pasa-Tolic, L.; Stacey, G.; Vertes, A. *Front. Plant Sci.* **2018**, *9*, No. 1646, DOI: 10.3389/fpls.2018.01646.

(17) Zhang, L. W.; Vertes, A. *Anal. Chem.* **2015**, *87*, 10397−10405.

(18) Stopka, S. A.; Wood, E. A.; Khattar, R.; Agtuca, B. J.; Abdelmoula, W. M.; Agar, N. Y. R.; Stacey, G.; Vertes, A. *Anal. Chem.* **2021**, *93*, 9677−9687.

(19) Cahill, J. F.; Riba, J.; Kertesz, V. *Anal. Chem.* **2019**, *91*, 6118−6126.

(20) Spraggins, J. M.; Caprioli, R. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 1022−1031.

(21) Onjiko, R. M.; Portero, E. P.; Moody, S. A.; Nemes, P. *Anal. Chem.* **2017**, *89*, 7069−7076.

(22) Stopka, S. A.; Agtuca, B. J.; Koppenaal, D. W.; Pasa-Tolic, L.; Stacey, G.; Vertes, A.; Anderton, C. R. *Plant J.* **2017**, *91*, 340−354.

(23) Taylor, M. J.; Mattson, S.; Liyu, A.; Stopka, S. A.; Ibrahim, Y. M.; Vertes, A.; Anderton, C. R. *Metabolites* **2021**, *11*, No. 200, DOI: 10.3390/metabo11040200.

(24) Zhang, L. C.; Xu, T. F.; Zhang, J. T.; Wong, S. C. C.; Ritchie, M.; Hou, H. W.; Wang, Y. L. *Anal. Chem.* **2021**, *93*, 10462−10468.

(25) de Souza, R.; Ambrosini, A.; Passaglia, L. M. P. *Genet. Mol. Biol.* **2015**, *38*, 401−419, DOI: 10.1590/S1415-475738420150053.

(26) Samarah, L. Z.; Khattar, R.; Tran, T. H.; Stopka, S. A.; Brantner, C. A.; Parlanti, P.; Velickovic, D.; Shaw, J. B.; Agtuca, B. J.; Stacey, G.; Pasa-Tolic, L.; Tolic, N.; Anderton, C. R.; Vertes, A. *Anal. Chem.* **2020**, *92*, 7289−7298.

(27) Velickovic, D.; Agtuca, B. J.; Stopka, S. A.; Vertes, A.; Koppenaal, D. W.; Paga-Tolic, L.; Stacey, G.; Anderton, C. R. *ISME J.* **2018**, *12*, 2335−2338, DOI: 10.1038/s41396-018-0188-8.

(28) Wei, R. M.; Wang, J. Y.; Su, M. M.; Jia, E.; Chen, S. Q.; Chen, T. L.; Ni, Y. *Sci. Rep.* **2018**, *8*, No. 663, DOI: 10.1038/s41598-017-19120-0.

(29) Rappez, L.; Stadler, M.; Triana, S.; Gathungu, R. M.; Ovchinnikova, K.; Phapale, P.; Heikenwalder, M.; Alexandrov, T. *Nat. Methods* **2021**, *18*, 799−805.

(30) Neumann, E. K.; Ellis, J. F.; Triplett, A. E.; Rubakhin, S. S.; Sweedler, J. V. *Anal. Chem.* **2019**, *91*, 7871−7878.

(31) Shrestha, B.; Vertes, A. *Anal. Chem.* **2014**, *86*, 4308−4315.

(32) Stiving, A. Q.; Jones, B. J.; Ujma, J.; Giles, K.; Wysocki, V. H. *Anal. Chem.* **2020**, *92*, 4475−4483.

(33) Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W. M.; Fiehn, O.; Goodacre, R.; Griffin, J. L.; Hankemeier, T.; Hardy, N.; Harnly, J.; Higashi, R.; Kopka, J.; Lane, A. N.; Lindon, J. C.; Marriott, P.; Nicholls, A. W.; Reily, M. D.; Thaden, J. J.; Viant, M. R. *Metabolomics* **2007**, *3*, 211−221.

(34) Gaude, N.; Tippmann, H.; Flemetakis, E.; Katinakis, P.; Udvardi, M.; Dormann, P. *J. Biol. Chem.* **2004**, *279*, 34624−34630.

(35) Liu, C. W.; Murray, J. D. *Plants* **2016**, *5*, No. 33, DOI: 10.3390/plants5030033.

(36) Kohashi, M. *J. Biochem.* **1980**, *87*, 1581−1586.

(37) Zheng, X. Y.; Aly, N. A.; Zhou, Y. X.; Dupuis, K. T.; Bilbao, A.; Paurus, V. L.; Orton, D. J.; Wilson, R.; Payne, S. H.; Smith, R. D.; Baker, E. S. *Chem. Sci.* **2017**, *8*, 7724−7736.

(38) Domon, B.; Costello, C. E. *Glycoconjugate J.* **1988**, *5*, 397−409.

(39) Fabre, N.; Rustan, I.; de Hoffmann, E.; Quetin-Leclercq, J. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 707−715.

(40) Liu, Z. J.; Kong, X. Y.; Long, Y. P.; Liu, S. R.; Zhang, H.; Jia, J. B.; Cui, W. H.; Zhang, Z. M.; Song, X. W.; Qiu, L. J.; Zhai, J. X.; Yan, Z. *Nat. Plants* **2023**, *9*, 515−524.

(41) Yao, Y. B.; Yuan, H. M.; Wu, G. W.; Yan, J.; Zhao, D. S.; Chen, S.; Kang, Q. H.; Ma, C. M.; Gong, Z. P. *Symbiosis* **2022**, *88*, 21−35.

(42) Deng, L. L.; Webb, I. K.; Garimella, S. V. B.; Hamid, A. M.; Zheng, X. Y.; Norheim, R. V.; Prost, S. A.; Anderson, G. A.; Sandoval, J. A.; Baker, E. S.; Ibrahim, Y. M.; Smith, R. D. *Anal. Chem.* **2017**, *89*, 4628−4634.

Supplementary Information

High-throughput f-LAESI-IMS-MS for Mapping Biological Nitrogen Fixation One Cell at a Time

Marjan Dolatmoradi[1]*, Sylwia A. Stopka*[1, 2]*, Chloe Corning*[1]*, Gary Stacey*[3]*, and Akos Vertes*[1]*

[1] Department of Chemistry, The George Washington University, Washington, District of Columbia 20052, United States
[2] Department of Neurosurgery, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, United States
[3] Divisions of Plant Sciences and Biochemistry, Christopher S. Bond Life Sciences Center, University of Missouri, Columbia, Missouri 65211, United States

*Corresponding author
E-mail: vertes@gwu.edu Phone: +1 (202) 994-2717. Fax: +1 (202) 994-5873

Table of Contents

Experimental section

Chemicals. Methanol (A452-4), chloroform (C607-4), water (W5-4) at HPLC grade were purchased from Fisher Scientific (Waltham, MA, USA). For mass calibration, 0.5 mM sodium formate was made from sodium hydroxide (43617, Sigma-Aldrich) and formic acid (5.33002, Sigma-Aldrich) in 90:10 (v/v) 2-propanol (1.02781, Sigma-Aldrich) water mixture. Poly-DL-alanine (P9003, Sigma-Aldrich) was used for the calibration of the IMS system.

Sample Preparation. The process of culturing wild-type *B. japonicum* strain USDA 110 has been described in a previous publication.[1] Briefly, the bacterial cultures were incubated for 3 days at 30 °C in HEPES-MES (HM) medium, until the measured optical density at $\lambda = 600$ nm reached 0.8 (corresponding to 108 cells/mL). The rhizobia were pelleted by centrifugation at 3000 rpm for 10 minutes, separated from the supernatant and re-suspended in sterile deionized water. Sterilized *G. max* seeds of "Williams 82" were planted in a mixture of 3:1 vermiculite:perlite and were infected by the *B. japonicum* suspension using 500 µL per soybean seed. Pots were transferred to a growth chamber (Percival E36HO, Percival Scientific, Perry, IA) set to 30 °C with a 16 h light/8 h dark cycle. Every other day during the growth, Broughton and Dilworth (B&D) medium was added to the tray to reach the original level.[2] Twenty one days post-infection (dpi), intact nodules, attached to the primary root, were harvested, flash frozen, and stored at 80 °C. Prior to analysis, the nodules were embedded in 2.5% carboxymethylcellulose (CMC) (C4888, Sigma-Aldrich, St. Louis, MO), cryosectioned to 30 µm thickness (CM1800, Leica Microsystems Inc., Nussloch, Germany), and placed on a microscope slide. Then the slide was transferred to a Peltier stage mounted on the automated XYZ translation stage of the dual channel microscope.

Bimodal microscopy. Details for the dual-channel microscope designed for simultaneous brightfield and fluorescence imaging has been described in a previous publication.[3] Briefly, to excite GFP-fluorescence in the specimen, light from a 470 nm blue LED light source (MF469-35, Thorlabs, Newton, NJ) was directed through a 469 nm excitation filter and reflected by a dichroic mirror (MD498, Thorlabs, Newton, NJ) toward the objective. A cold white LED source (MCWHL5, Thorlabs, Newton, NJ) was used for brightfield illumination via a beam splitter (BSS10R, Thorlabs, Newton, NJ) coated for 400-700 nm. All generated wavelengths were focused to the specimen using Mitutoyo 5×, 10×, or 20× magnification long working distance objectives (MY5X-802, MY10X-803, MY20X 804, Thorlabs, Newton, NJ). Two monochromatic 4 megapixel scientific-grade CCD cameras (4070M-GE, Thorlabs, Newton, NJ) were used to capture the fluorescence and brightfield images.

f-LAESI sampling. Delivering 5 laser shots with ~0.9 ± 0.1 mJ laser energy (measured before coupling it into the fiber) to a single root nodule cell resulted in the creation of an ablation plume. To achieve LAESI the plume was intercepted and ionized by an electrospray. The spray solution in positive and negative ion modes consisted of 1:1 methanol/water (v/v) acidified with 0.1% glacial acetic acid, and 2:1 methanol/chloroform (v/v), respectively. They were supplied by a syringe pump (Physio 22, Harvard Apparatus, Holliston, MA, USA) at 400 nL/min through a stainless-steel emitter (MT320-100-5-5, i.d. 100 µm, New Objective, Woburn, MA, USA). Stable electrospray was achieved by supplying +3100 V in positive ion mode and −2200 V in negative ion mode by a high voltage power supply (PS350, Stanford Research Systems, Sunnyvale, CA, USA). The optimal distances between the emitter positioned on axis with the mass spectrometer orifice and the fiber tip and between the emitter tip and mass spectrometer inlet were found to be 0.5 cm and ~0.7 cm, respectively. To prevent water condensation from the ambient environment or the drying of the sample during extended analysis times, working at an optimum relative humidity in the sample region was essential. For this, the sample was surrounded by a home-built environmental chamber that was flushed with dry nitrogen gas. The relative humidity inside the chamber was measured by a portable humidity meter (Catalog No. 11-661-18, Fisher Scientific, Waltham, MA, USA) and the flow rate of the dry nitrogen was adjusted to achieve the desired humidity value (e.g., 40 % relative humidity for a nodule section sample).

Figures

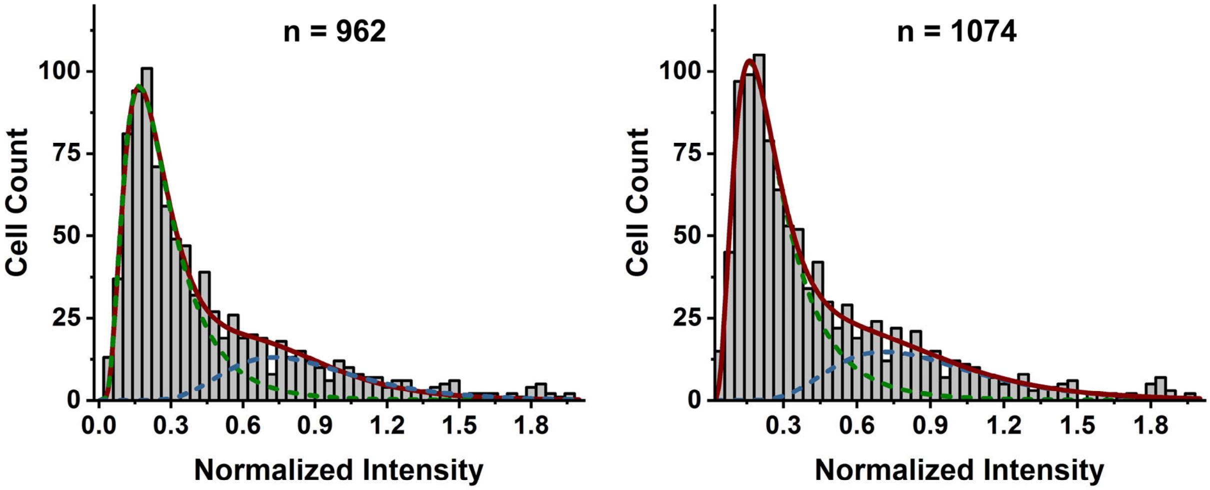


Figure S1. Effect of missing value imputation on normalized hexose phosphate abundance distributions. Comparing deconvolutions for the distributions without (left panel) and with (right panel) missing value imputation resulted in high quality nonlinear fits with $r^2 > 0.99$ correlation coefficients for both cases.

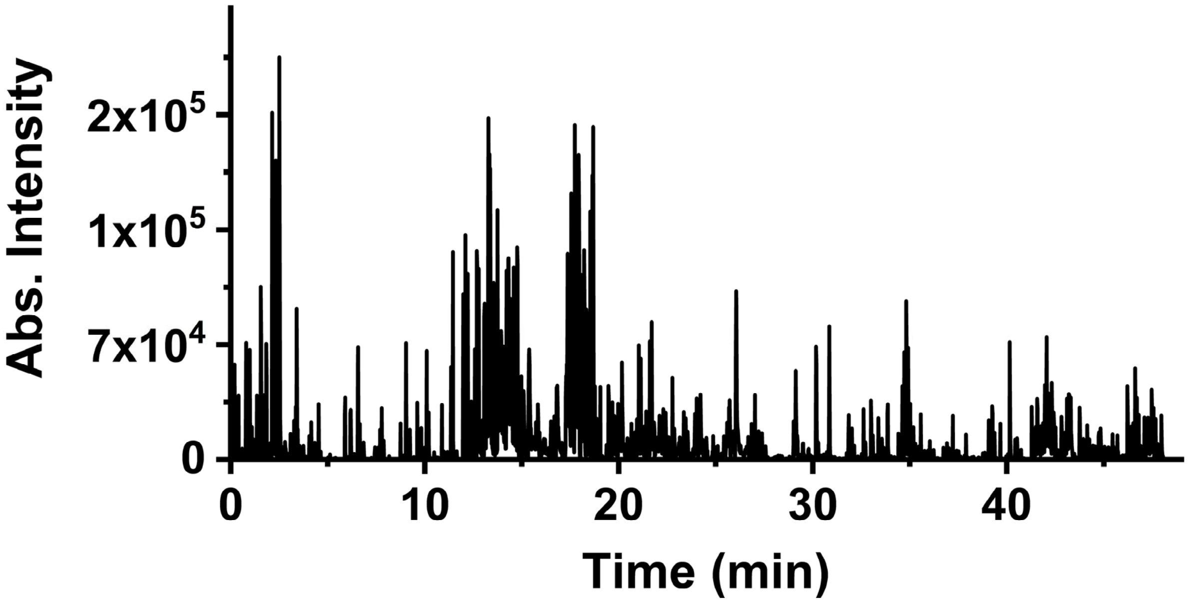


Figure S2. Ion chromatogram of disaccharide species (*m/z* 341. 109) obtained from the ablation of 643 single soybean root nodule cells in 48 minutes (804 cells/h) using negative ion f-LAESI-IMS-MS.

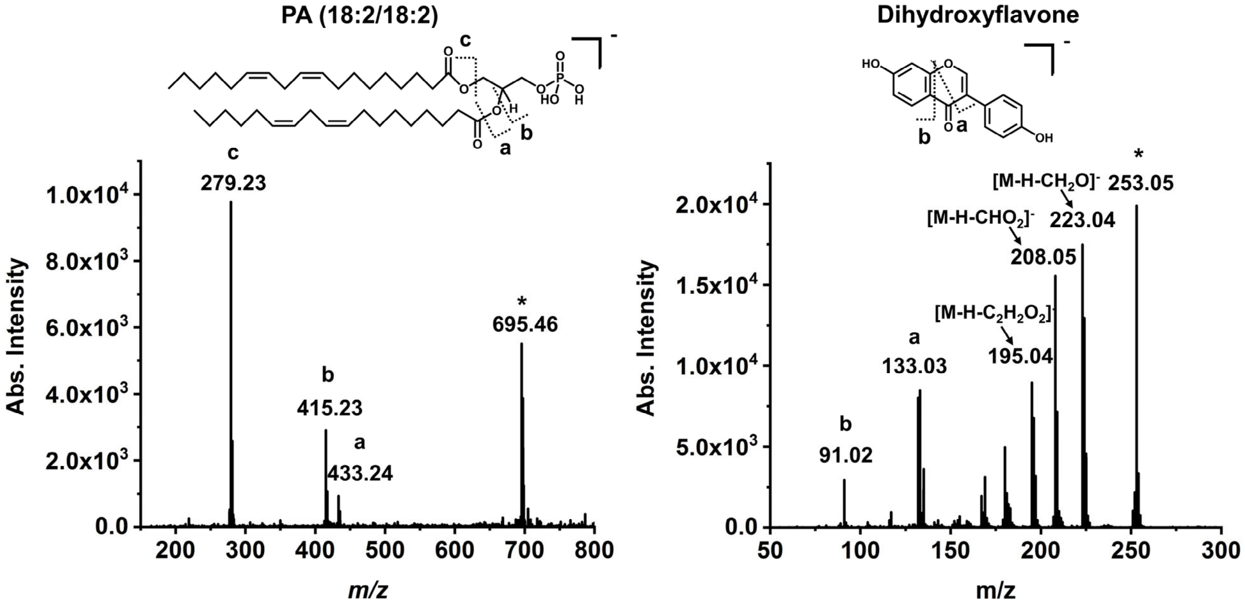


Figure S3. Tandem MS at 30 eV collision energy for *m/z* 695.464 (PA(18:2/18:2)) (left panel) and *m/z* 253.051 (Dihydroxyflavone) (right panel).

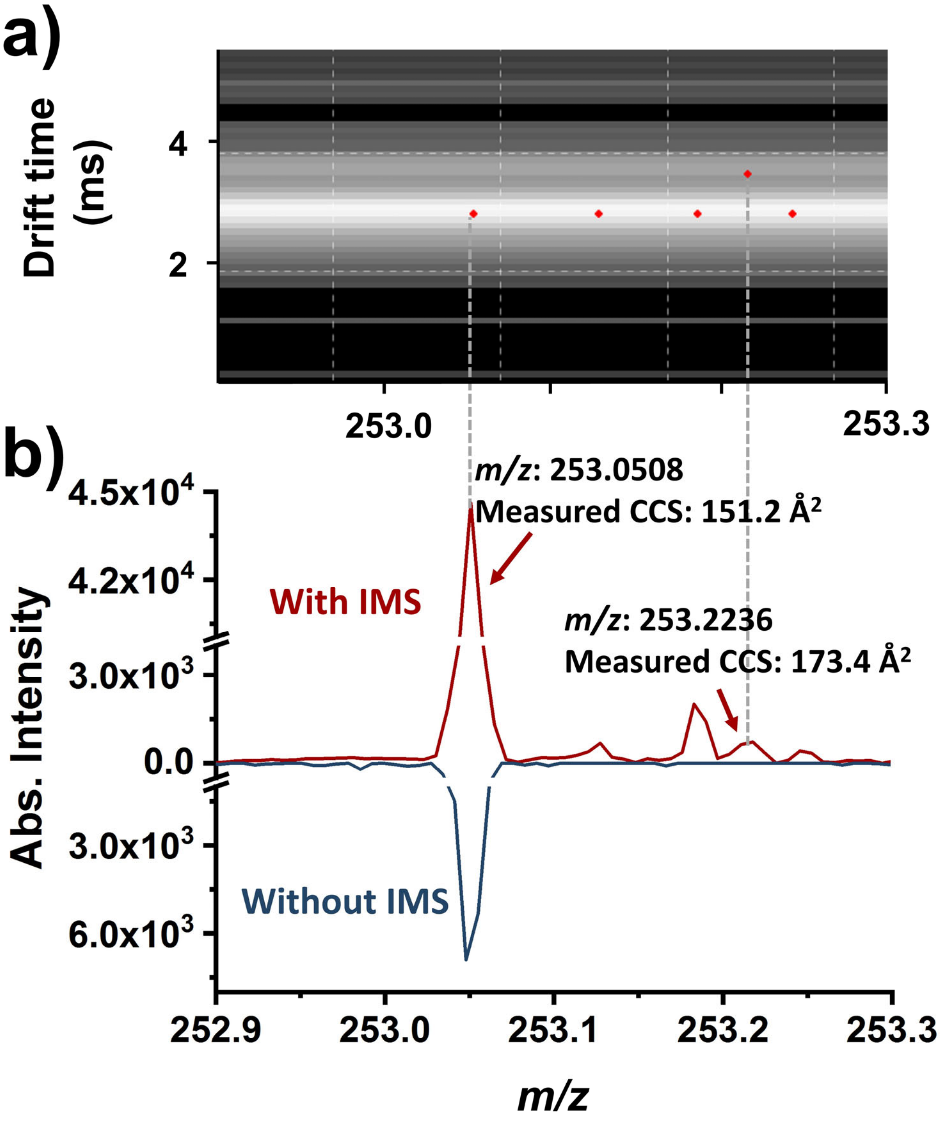


Figure S4. Improved S/N in single-cell analysis due to IMS. (a) Segment of a mobiligram obtained from f-LAESI-IMS-MS analysis of a single soybean root nodule cell, representing drift time for ions detected in negative ion mode (red dots) vs. *m/z* values. In panel (b) the top mass spectrum for f-LAESI-IMS-MS revealed five ionic species in the *m/z* 252.9 to 253.3 range compared to a single ion type detected for f-LAESI-MS (bottom spectrum). The latter, observed at *m/z* 253.051, corresponded to deprotonated dihydroxy isoflavone. For the top spectrum, the ion at *m/z* 253.224 (S/N = 4.6) was found to be palmitoleate with a calculated *m/z* 253.217. The measured CCS of 173.4 Å^2 was comparable to the CCS of 167.9 Å^2 for palmitoleate in the AllCCS database.

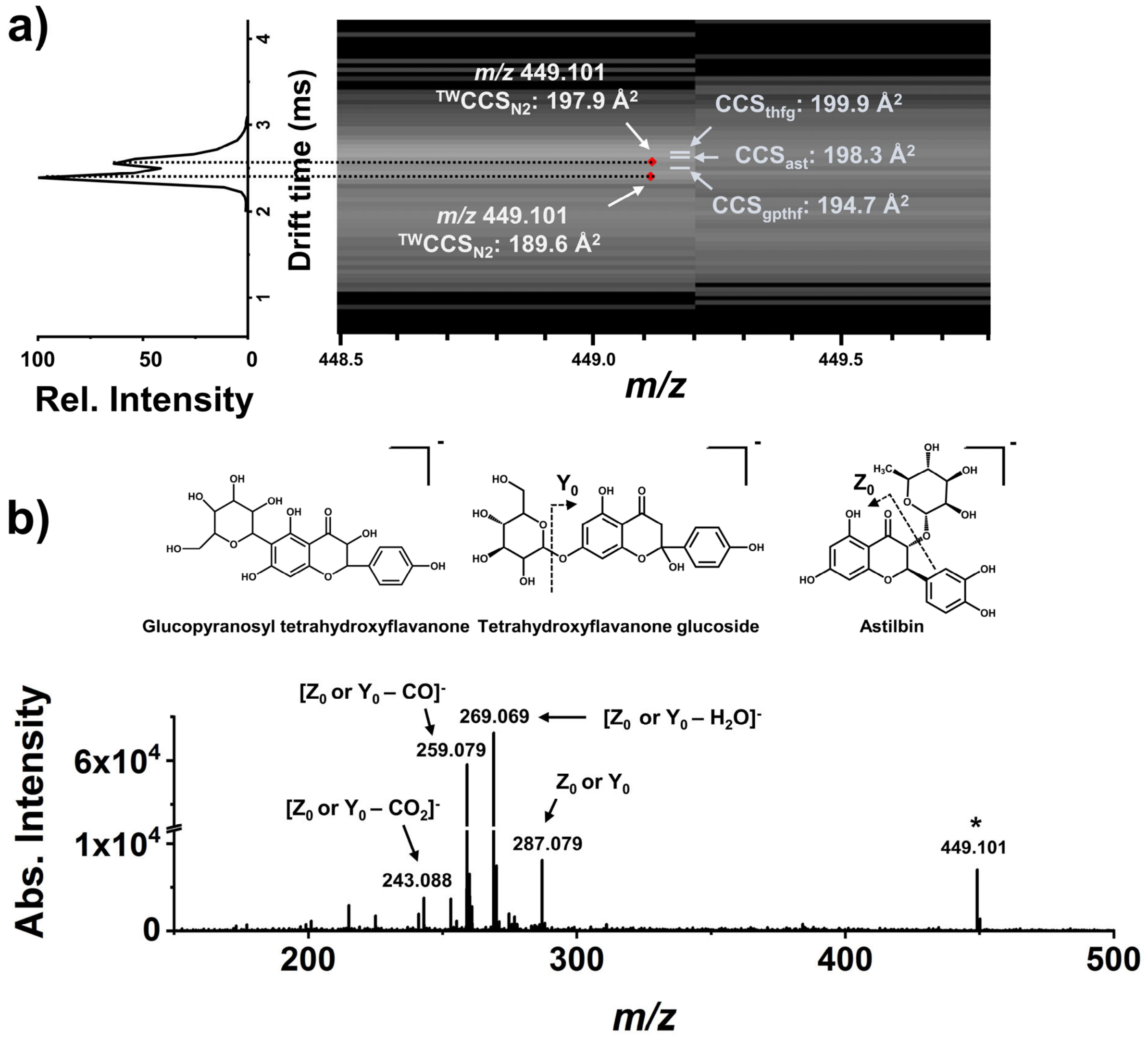


Figure S5. Detection of structural isomers by IMS-MS in single-cell measurements. **(a)** The mobiligram with an extracted drift time profile for *m/z* 449.101 demonstrated two peaks at 2.4±0.1 ms (CCS = 189.6 Å^2) and 2.6±0.2 ms (CCS = 197.9 Å^2). Previously, astilbin had been detected in soybean root nodules with a measured CCS of 198.3 Å^2.[4] For structural isomers glucopyranosyl tetrahydroxyflavanone and tetrahydroxyflavanone glucoside, the AllCCS database reported calculated CCS values of CCS = 194.7 Å^2 and CCS = 199.9 Å^2, respectively. The three reference values are indicated in the mobiligram by horizontal lines next to the measured points. **(b)** Tandem MS of the precursor ion with *m/z* 449.101. The *m/z* 287.079 fragment might be explained by the cleavage of the glycosidic bond resulting in fragments Y_0 or Z_0 for tetrahydroxyflavanone glucoside or astilbin, respectively.

Table S1. Tentative metabolite assignments for selected ionic species detected in infected soybean root nodule cells based on measured accurate masses, CCS values, and tandem MS. The fidelity of metabolite identification is characterized by the Metabolomics Standard Initiative (MSI) levels.[5]

Compound	Formula	Ion Type	Measured mass	Calculated mass	Δm (mDa)	Experimental CCS	Reference CCS	ΔCCS (Å2)	MSI level
5-Hydroxyisourate	$C_5H_4N_4O_4$	$[M-H]^-$	183.0164	183.016	0.4	130.4	131.2	-0.8	2
Gluconic acid	$C_6H_{12}O_7$	$[M-H]^-$	195.0471	195.0510	-3.9	131.9	132.9	-1.0	1
Acetylcarnitine	$C_9H_{17}NO_4$	$[M-H]^-$	202.1087	202.1085	0.2	147.7	152.2	-4.5	1
Homocitrate	$C_7H_{10}O_7$	$[M-H]^-$	205.0324	205.0354	-3.0	130.9	130.0	0.9	1
Glucarate	$C_6H_{10}O_8$	$[M-H]^-$	209.0306	209.0303	0.3	128.0	132.1	-4.1	1
Valylvaline	$C_{10}H_{20}N_2O_3$	$[M-H]^-$	215.1388	215.1401	-1.3	151.3	154.0	-2.7	2
Cystine	$C_6H_{12}N_2O_4S_2$	$[M-H_2O-H]^-$	221.0100	221.0055	4.5	135.6			3
Hydroxyjasmonic acid	$C_{12}H_{18}O_4$	$[M-H]^-$	225.1145	225.1132	1.3	150.5	154.6	-4.1	2
Hydroxyphenyl propionic acid sulfate	$C_9H_{10}O_6S$	$[M-H_2O-H]^-$	227.0063	227.0014	4.9	133.7			3
Glutamyl taurine	$C_7H_{14}N_2O_6S$	$[M-H_2O-H]^-$	235.0281	235.0389	-10.8	140.1			3
Glucosamine phosphate	$C_6H_{14}NO_8P$	$[M-H_2O-H]^-$	240.0379	240.0273	10.6	143.0	143.1	-0.1	2
Methylglutarylcarnitine	$C_{11}H_{19}NO_6$	$[M-H_2O-H]^-$	242.1002	242.1028	-2.6	136.0			3
Hydroxyferulic acid	$C_{10}H_{10}O_5$	$[M+Cl]^-$	245.0322	245.0222	10.0	138.7			3
Leucyl-aspartatic acid	$C_{10}H_{18}N_2O_5$	$[M-H]^-$	245.1211	245.1144	6.7	150.5	155.2	-4.7	1
Glutamylcysteine	$C_8H_{14}N_2O_5S$	$[M-H]^-$	249.0564	249.0550	1.4	147.2	151.7	-4.5	2
Neopterins	$C_9H_{11}N_5O_4$	$[M-H]^-$	252.078	252.0738	4.2	147.7	150.3	-2.6	2
Dihydroxyflavone	$C_{15}H_{10}O_4$	$[M-H]^-$	253.0508	253.0506	0.2	151.2	155.2	-4.0	1
Palmitoleate	$C_{16}H_{30}O_2$	$[M-H]^-$	253.2236	253.2168	6.8	173.4	167.9	5.5	2
Hexose Phosphate	$C_6H_{13}O_9P$	$[M-H]^-$	259.0211	259.0225	-1.4	141.0	147.3	-6.3	1
Glutamyl-iso/leucine	$C_{11}H_{20}N_2O_5$	$[M-H]^-$	259.1266	259.1294	-3.4	157.6	160.0	-2.4	1
Inosine	$C_{10}H_{12}N_4O_5$	$[M-H]^-$	267.0727	267.0735	-0.8	152.9	158.8	-5.9	2
Trihydroxyflavone	$C_{15}H_{10}O_5$	$[M-H]^-$	269.0451	269.0455	-0.4	156.1	161.5	-5.4	2
Phosphogluconic acid	$C_6H_{13}O_{10}P$	$[M-H]^-$	275.0262	275.0174	8.8	150.1	148.4	1.7	2
Linolenic Acid	$C_{18}H_{30}O_2$	$[M-H]^-$	277.2188	277.2173	1.5	180.3	175.8	4.5	2

Linoleic acid	$C_{18}H_{32}O_2$	$[M-H]^-$	279.2302	279.2330	-2.8	180.2	174.3	5.9	2
Oleic acid	$C_{18}H_{34}O_2$	$[M-H]^-$	281.2488	281.2486	0.2	180.5	175.2	5.3	2
Xanthosine	$C_{10}H_{12}N_4O_6$	$[M-H]^-$	283.0616	283.0684	-6.8	159.5	159.6	-0.1	2
Hydroxylinolenic acid	$C_{18}H_{30}O_3$	$[M-H]^-$	293.2143	293.2122	2.1	176.0	171.6	4.4	2
Hydroxylinoleic acid	$C_{18}H_{32}O_3$	$[M-H]^-$	295.2303	295.2278	2.5	177.4	178.3	-0.9	2
Dihydroxy dimethoxyisoflavanone	$C_{17}H_{16}O_6$	$[M-H_2O-H]^-$	297.0724	297.0763	-3.9	165.9			3
Hexahydroxyflavan	$C_{15}H_{14}O_7$	$[M-H]^-$	305.0703	305.0667	3.6	160.4	164.3	-3.9	2
Trimethoxyflavone	$C_{18}H_{16}O_5$	$[M-H]^-$	311.0902	311.0925	-2.3	165.4	169.7	-4.3	2
Dihydroxy dimethoxyisoflavanone	$C_{17}H_{16}O_6$	$[M-H]^-$	315.0844	315.0869	-2.5	166.5	173.0	-6.5	1
Homoglutathione	$C_{11}H_{19}N_3O_6S$	$[M-H]^-$	320.0936	320.0922	1.4	163.2	169.8	-6.6	1
Tetrahydroxyisoflavone	$C_{15}H_{10}O_6$	$[M+Cl]^-$	321.0139	321.0171	-3.2	153.0			2
Cytidine monophosphate	$C_9H_{14}N_3O_8P$	$[M-H]^-$	322.0429	322.044	-1.1	158.8	164.0	-5.2	2
Pyranosylhexose	$C_{12}H_{22}O_{10}$	$[M-H]^-$	325.1081	325.114	-5.9	164.0	167.7	-3.7	1
Hydroxyterpineol glucoside	$C_{16}H_{28}O_7$	$[M-H]^-$	331.1705	331.1762	-5.7	184.3	179.2	5.1	2
Hexosylglycerol phosphate	$C_9H_{19}O_{11}P$	$[M-H]^-$	333.0576	333.0592	-1.6	164.6	164.2	0.4	1
Ascorbic acid glucoside	$C_{12}H_{18}O_{11}$	$[M-H]^-$	337.0724	337.0776	-5.2	164.1	169.2	-5.1	1
Cyclodopa glucoside	$C_{15}H_{19}NO_9$	$[M-H_2O-H]^-$	338.0838	338.0876	-3.8	154.3			3
Disaccharide	$C_{12}H_{22}O_{11}$	$[M-H]^-$	341.1112	341.1089	2.3	164.5	168.7	-4.2	1
Cinnamoyl glucoside	$C_{15}H_{18}O_7$	$[M+Cl]^-$	345.0686	345.0747	-6.1	163.5			3
N-Acetylmuramoyl alanine	$C_{14}H_{22}N_2O_8$	$[M-H_2O-H]^-$	345.1335	345.1303	3.2	171.4			2
Deoxyadenosine monophosphate	$C_{10}H_{14}N_5O_6P$	$[M+Cl]^-$	366.0416	366.0376	4.0	161.4			3
Hydroxymethylglutathione	$C_{11}H_{15}N_3O_7S$	$[M+Cl]^-$	368.0421	368.0330	9.1	161.2			3
Dihydroferulic acid glucuronide	$C_{16}H_{19}O_{10}$	$[M-H]^-$	371.0995	371.0984	1.1	179.3	174.1	5.2	2
Dodecanedioylcarnitine	$C_{19}H_{35}NO_6$	$[M-H]^-$	372.2366	372.2392	-2.6	188.0	189.0	-1.0	2
Oleoyl glycine	$C_{20}H_{37}NO_3$	$[M+Cl]^-$	374.254	374.2467	7.3	188.5			3
Disaccharide	$C_{12}H_{22}O_{11}$	$[M+Cl]^-$	377.0824	377.0856	-3.2	169.2			2
Acetyl dihexose	$C_{14}H_{24}O_{12}$	$[M-H]^-$	383.1132	383.1195	-6.3	180.4	178.2	2.2	1
Sinapoylglucose	$C_{17}H_{22}O_{10}$	$[M-H]^-$	385.1169	385.1140	2.9	192.1	187.9	4.2	1
Tuberonic acid glucoside	$C_{18}H_{28}O_9$	$[M-H]^-$	387.1689	387.1661	2.8	182.8	186.6	-3.8	1

Uridine diphosphate	$C_9H_{14}N_2O_{12}P_2$	$[M-H]^-$	402.991	402.9949	-3.9	169.1	172.8	-3.7	1
Galactinol dihydrate	$C_{12}H_{26}O_{13}$	$[M+Cl]^-$	413.1051	413.1067	-1.6	182.8			3
Daidzin	$C_{21}H_{20}O_9$	$[M-H]^-$	415.0967	415.1035	-6.8	202.6	203.0	-0.4	2
Dihexose phosphate	$C_{12}H_{23}O_{14}P$	$[M-H]^-$	421.0733	421.0753	-2.0	175.5	180.4	-4.9	1
Indolylmethyl glucosinolate	$C_{16}H_{20}N_2O_9S_2$	$[M-H_2O-H]^-$	429.0451	429.0426	2.5	177.4			3
Acetyldaidzin	$C_{23}H_{22}O_{10}$	$[M-H_2O-H]^-$	439.0953	439.1029	-7.6				3
Dihydrophaseic acid glucoside	$C_{21}H_{32}O_{10}$	$[M-H]^-$	443.2056	443.1923	13.3	196.9	203.4	-6.5	2
Tetrahydroxyflavanone glucoside	$C_{21}H_{22}O_{11}$	$[M-H]^-$	449.1014	449.1083	-6.9	197.9	199.9	-2.0	2
Dihydroxy dimethoxyisoflavone glucoside	$C_{23}H_{24}O_{11}$	$[M-H]^-$	475.1261	475.1240	2.1	205.9	214.2	-8.3	2
Trisaccharide	$C_{18}H_{32}O_{16}$	$[M-H]^-$	503.1654	503.1618	3.6	205.8	210.1	-4.3	1
Hydroxy-pentamethoxyflavanone rhamnoside	$C_{26}H_{32}O_{12}$	$[M+Cl]^-$	517.163	517.1651	-2.1	199.7			3
Crocetin monoglucosyl ester	$C_{22}H_{34}O_{12}$	$[M+Cl]^-$	525.1779	525.1739	4.0	223.5			3
Cyclic ADP-ribose	$C_{15}H_{21}N_5O_{13}P_2$	$[M-H]^-$	540.0605	540.0538	6.7	207.4	202.2	5.2	3
Trihydroxy tetramethoxyflavone glucoside	$C_{25}H_{28}O_{14}$	$[M-H]^-$	551.1436	551.1406	3.0	210.2	217.9	-7.7	2
dTDP-hexose	$C_{16}H_{26}N_2O_{16}P_2$	$[M-H]^-$	563.0646	563.0685	-3.9	202.7	208.5	-5.8	1
UDP-hexose	$C_{15}H_{24}N_2O_{17}P_2$	$[M-H]^-$	565.0466	565.0483	-1.7	207.1	207.5	-0.4	1
Hexaprenyl diphosphate	$C_{30}H_{52}O_7P_2$	$[M-H]^-$	585.3112	585.3116	-0.4	233.2	228.3	4.9	2
UDP-acetyl-hexosamine	$C_{17}H_{27}N_3O_{17}P_2$	$[M-H]^-$	606.0724	606.0743	-1.9	221.0	223.1	-2.1	2
Tetrasaccharide	$C_{24}H_{42}O_{21}$	$[M-H]^-$	665.2097	665.2146	-4.9	228.9	225.1	3.8	3
PA(16:0/18:3)	$C_{37}H_{67}O_8P$	$[M-H]^-$	669.4532	669.4501	3.1	250.0	255.0	-5.0	2
PA(16:0/18:2)	$C_{37}H_{69}O_8P$	$[M-H]^-$	671.4669	671.4657	1.2	251.2	256.7	-5.5	2
PA(16:0/18:1)	$C_{37}H_{71}O_8P$	$[M-H]^-$	673.4836	673.4814	2.2	253.1	259.3	-6.2	2
PA(18:2/18:3)	$C_{39}H_{67}O_8P$	$[M-H]^-$	693.4493	693.4501	-0.8	252.2	257.4	-5.2	2
PA(18:2/18:2)	$C_{39}H_{69}O_8P$	$[M-H]^-$	695.4642	695.4657	-1.5	254.0	258.6	-4.6	1
PA(18:2/18:1)	$C_{39}H_{71}O_8P$	$[M-H]^-$	697.4821	697.4813	0.8	255.8	260.9	-5.1	2
PA(18:1/18:1)	$C_{39}H_{73}O_8 P$	$[M-H]^-$	699.4913	699.4970	-5.7	257.9	263.0	-5.1	2

PE (16:1/18:1)	$C_{39}H_{74}NO_8P$	$[M-H]^-$	714.5005	714.5074	-6.9	260.3	265.9	-5.6	2
PE (16:0/18:1)	$C_{39}H_{76}NO_8P$	$[M-H]^-$	716.5225	716.5230	-0.5	262.2	265.7	-3.5	2
PG(16:0/16:0)	$C_{38}H_{75}O_{10}P$	$[M-H]^-$	721.5016	721.5025	-0.9	264.9	270.9	-6.0	2
PE(18:2/18:2)	$C_{41}H_{74}NO_8P$	$[M-H]^-$	738.512	738.5079	4.1	269.5	268.8	0.7	2
PE(18:2/18:1)	$C_{41}H_{76}NO_8P$	$[M-H]^-$	740.52	740.5236	-3.6	271.2	270.7	0.5	2
PE (18:2/18:0)	$C_{41}H_{78}NO_8P$	$[M-H]^-$	742.5306	742.5382	-7.6	268.1	272.6	-4.5	2
PG(16:0/18:2)	$C_{40}H_{75}O_{10}P$	$[M-H]^-$	745.4982	745.5025	-4.3	267.5	273.1	-5.6	2
PG(16:0/18:1)	$C_{40}H_{77}O_{10}P$	$[M-H]^-$	747.5156	747.5182	-2.6	270.1	275.2	-5.1	2
PG(18:1/18:1)	$C_{42}H_{79}O_{10}P$	$[M-H]^-$	773.5284	773.5338	-5.4	275.7	282.7	-7.0	2
Soyasaponin III	$C_{42}H_{68}O_{14}$	$[M-H]^-$	795.4574	795.4513	6.1	267.2	274.6	-7.4	3
PI(16:0/18:3)	$C_{43}H_{77}O_{13}P$	$[M-H]^-$	831.5062	831.5029	3.3	283.3	286.3	-3.0	2
PI(16:0/18:2)	$C_{43}H_{79}O_{13}P$	$[M-H]^-$	833.5233	833.5186	4.7	284.3	286.5	-2.2	2
PI(16:0/18:1)	$C_{43}H_{81}O_{13}P$	$[M-H]^-$	835.5301	835.5331	-3.0	285.7	288.0	-2.3	2
PI(18:2/18:2)	$C_{45}H_{79}O_{13}P$	$[M-H]^-$	857.5087	857.518	-9.3	285.5	290.2	-4.7	2
PI(18:2/18:1)	$C_{45}H_{81}O_{13}P$	$[M-H]^-$	859.5314	859.5337	-2.3	291.1	291.5	-0.4	2
PI(18:2/18:0)	$C_{45}H_{83}O_{13}P$	$[M-H]^-$	861.5436	861.5493	-5.7	288.8	292.9	-4.1	2
Soyasaponin II	$C_{47}H_{76}O_{17}$	$[M-H]^-$	911.4927	911.5010	-8.3	301.6	293.7	7.9	2
Tragopogonsaponin M	$C_{51}H_{76}O_{16}$	$[M-H_2O-H]^-$	925.5021	925.4949	7.2	304.5	303.6	0.9	3
Araliasaponin I	$C_{47}H_{76}O_{18}$	$[M-H]^-$	927.4969	927.4959	1.0	301.4	298.4	3.0	3
Dehydrosoyasaponin I	$C_{48}H_{76}O_{18}$	$[M-H]^-$	939.4977	939.4959	1.8	310.4	299.9	10.5	2
Soyasaponin I	$C_{48}H_{78}O_{18}$	$[M-H]^-$	941.5076	941.5115	-3.9	311.5	297.9	13.6	2
PI(22:5/22:5)	$C_{53}H_{81}O_{14}P$	$[M-H_2O-H]^-$	953.5175	953.5180	-0.5	313.3			3
Soyasaponin B	$C_{48}H_{76}O_{19}$	$[M-H]^-$	955.4882	955.4903	-2.1	312.6	301.9	10.7	2
Soyasaponin V	$C_{48}H_{78}O_{19}$	$[M-H]^-$	957.5016	957.5065	-4.9	313.3	298.6	14.7	2
PIP(20:1/18:2)	$C_{47}H_{86}O_{16}P_2$	$[M-H]^-$	967.5322	967.5318	0.4	320.3	312.1	8.2	2
PIP(18:1-O/20:1)	$C_{47}H_{86}O_{17}P_2$	$[M-H]^-$	983.5256	983.5267	-1.1	322.5	316.1	6.4	3
Maltohexaose	$C_{36}H_{62}O_{31}$	$[M-H]^-$	989.3164	989.3202	-3.8	268.5	280.0	-11.5	3
PIP(18:1-2OH/20:2)	$C_{47}H_{86}O_{18}P_2$	$[M-H]^-$	999.512	999.5217	-9.7	325.6	319.8	5.8	3
Triterpenoid saponin pentaglycoside	$C_{51}H_{80}O_{20}$	$[M-H]^-$	1011.5109	1011.5170	-6.1	346.7			3
Pisumsaponin I	$C_{51}H_{80}O_{21}$	$[M-H]^-$	1027.5151	1027.5119	3.2	308.4	316.1	-7.7	3

Medicoside I	$C_{52}H_{84}O_{22}$	$[M-H_2O-H]^-$	1041.5215	1041.527	-5.5	313.7			3
Caraganiside A	$C_{52}H_{84}O_{21}$	$[M-H]^-$	1043.5389	1043.5433	-4.4	316.0	320.2	-4.2	3
Matesaponin 3	$C_{53}H_{86}O_{22}$	$[M-H_2O-H]^-$	1055.542	1055.5427	-0.7	317.4			3
PIP(22:6-OH/22:2)	$C_{53}H_{88}O_{17}P_2$	$[M-H]^-$	1057.5444	1057.5424	2.0	321.0	326.5	-5.5	3
Medicoside I	$C_{52}H_{84}O_{22}$	$[M-H]^-$	1059.5344	1059.5381	-3.7	321.3	315.5	5.8	3
Soyasapogenol A tetraglycoside	$C_{53}H_{84}O_{22}$	$[M-H_2O-H]^-$	1071.5431	1071.5382	4.9	321.1			3
Matesaponin 3	$C_{53}H_{88}O_{18}P_2$	$[M-H]^-$	1073.5463	1073.5538	-7.5	326.0	321.2	4.8	3
Soyasaponin ag	$C_{54}H_{84}O_{22}$	$[M-H]^-$	1083.533	1083.5381	-5.1	326.8	317.7	9.1	2
Araliasaponin V	$C_{54}H_{88}O_{23}$	$[M-H_2O-H]^-$	1085.5474	1085.5532	-5.8	327.7			3
Soyasaponin A2	$C_{53}H_{84}O_{23}$	$[M-H_2O-H]^-$	1087.5234	1087.5325	-9.1	330.5			3
Tragopogonsaponin G	$C_{56}H_{82}O_{21}$	$[M-H]^-$	1089.5281	1089.5276	0.5	328.8	321.8	7.0	3
Tragopogonsaponin J	$C_{57}H_{84}O_{22}$	$[M-H_2O-H]^-$	1101.527	1101.527	0.0	329.0			3
Araliasaponin V	$C_{54}H_{88}O_{23}$	$[M-H]^-$	1103.558	1103.5644	-6.4	333.5	324.7	8.8	3
Soyasaponin A2	$C_{53}H_{86}O_{24}$	$[M-H]^-$	1105.5474	1105.5436	3.8	334.9			3
Maltoheptaose	$C_{42}H_{72}O_{36}$	$[M-H]^-$	1151.3667	1151.3731	-6.4	306.2	300.0	6.2	3

References

(1) Velickovic, D.; Agtuca, B. J.; Stopka, S. A.; Vertes, A.; Koppenaal, D. W.; Paga-Tolic, L.; Stacey, G.; Anderton, C. R. *Isme Journal* **2018**, *12*, 2335-2338.
(2) Samarah, L. Z.; Khattar, R.; Tran, T. H.; Stopka, S. A.; Brantner, C. A.; Parlanti, P.; Velickovic, D.; Shaw, J. B.; Agtuca, B. J.; Stacey, G.; Pasa-Tolic, L.; Tolic, N.; Anderton, C. R.; Vertes, A. *Anal. Chem.* **2020**, *92*, 7289-7298.
(3) Stopka, S. A.; Wood, E. A.; Khattar, R.; Agtuca, B. J.; Abdelmoula, W. M.; Agar, N. Y. R.; Stacey, G.; Vertes, A. *Anal. Chem.* **2021**, *93*, 9677-9687.
(4) Zheng, X. Y.; Aly, N. A.; Zhou, Y. X.; Dupuis, K. T.; Bilbao, A.; Paurus, V. L.; Orton, D. J.; Wilson, R.; Payne, S. H.; Smith, R. D.; Baker, E. S. *Chemical Science* **2017**, *8*, 7724-7736.
(5) Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W. M.; Fiehn, O.; Goodacre, R.; Griffin, J. L.; Hankemeier, T.; Hardy, N.; Harnly, J.; Higashi, R.; Kopka, J.; Lane, A. N.; Lindon, J. C.; Marriott, P.; Nicholls, A. W.; Reily, M. D.; Thaden, J. J.; Viant, M. R. *Metabolomics* **2007**, *3*, 211-221.