analytical chemistry

Bar Coding MS^2 Spectra for Metabolite Identification

Jonathan L. Spalding,[†,§] Kevin Cho,[†,‡] Nathaniel G. Mahieu,[†,‡] Igor Nikolskiy,[†,§] Elizabeth M. Llufrio,[†,‡] Stephen L. Johnson,[§] and Gary J. Patti*[,†,‡]

[†]Department of Chemistry, Washington University in St. Louis, St. Louis, Missouri 63130, United States
[‡]Department of Medicine, Washington University School of Medicine, St. Louis, Missouri 63110, United States
[§]Department of Genetics, Washington University School of Medicine, St. Louis, Missouri 63110, United States

ABSTRACT: Metabolite identifications are most frequently achieved in untargeted metabolomics by matching precursor mass and full, high-resolution MS^2 spectra to metabolite databases and standards. Here we considered an alternative approach for establishing metabolite identifications that does not rely on full, high-resolution MS^2 spectra. First, we select mass-to-charge regions containing the most informative metabolite fragments and designate them as bins. We then translate each metabolite fragmentation pattern into a binary code by assigning 1's to bins containing fragments and 0's to bins without fragments. With 20 bins, this binary-code system is capable of distinguishing 96% of the compounds in the METLIN MS^2 library. A major advantage of the approach is that it extends untargeted metabolomics to low-resolution triple quadrupole (QqQ) instruments, which are typically less expensive and more robust than other types of mass spectrometers. We demonstrate a method of acquiring MS^2 data in which the third quadrupole of a QqQ instrument cycles over 20 wide isolation windows (coinciding with the location and width of our bins) for each precursor mass selected by the first quadrupole. Operating the QqQ instrument in this mode yields diagnostic bar codes for each precursor mass that can be matched to the bar codes of metabolite standards. Furthermore, our data suggest that using low-resolution bar codes enables QqQ instruments to make MS^2-based identifications in untargeted metabolomics with a specificity and sensitivity that is competitive to high-resolution time-of-flight technologies.
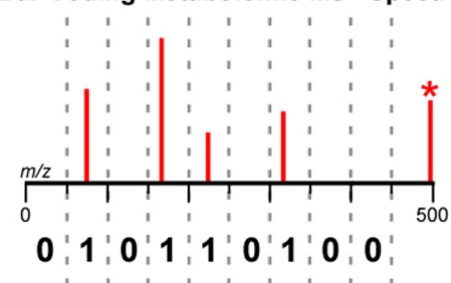
Although profiling features from biological samples by untargeted metabolomics is now routine, establishing the chemical identities of those features remains a major challenge.[1−4] Even when using state-of-the-art liquid chromatography/mass spectrometry (LC/MS) technologies and automated bioinformatic pipelines, analysis of untargeted metabolomic data sets is time-consuming and only a fraction of the thousands of features detected are identified.[5−7] In this sense, untargeted metabolomics as conventionally performed with LC/MS is highly inefficient.

Historically, to support the structural identification of a feature in untargeted metabolomics, high-resolution MS^2 data from in-house standards have been matched to high-resolution MS^2 data from research samples.[8] However, this process is severely limited by the availability of in-house standards in most laboratories. Thus, to help with metabolite identifications, fragmentation data from a variety of instrument platforms have become increasingly available online in recent years.[9−12] METLIN, which is currently the largest public MS^2 database for metabolomics, has experimental high-resolution MS^2 data for over 14 000 metabolites.[13] Yet, despite the availability of these MS^2 data, efficient identification of large numbers of features in untargeted metabolomic experiments continues to be experimentally challenging.[7]

In contrast to untargeted metabolomics, solutions for targeted metabolomics are well established.[14] Most frequently, targeted experiments are performed with a triple quadrupole

(QqQ) mass spectrometer in multiple reaction monitoring (MRM) mode.[15] After empirically identifying precursor-to-product transitions for the metabolites of interest, MRM methods can be optimized to rapidly and efficiently profile those compounds. The resulting data are relatively easy to interpret compared to data from untargeted metabolomic experiments. In addition to their proven sensitivity, quantitative reliability, and robustness, QqQ instruments are also generally less expensive than the high-resolution mass spectrometers that are conventionally used for untargeted metabolomics.[16,17] Indeed, QqQ-based methods have become the gold standard in the pharmaceutical industry.[18] The main limitation of the MRM-based workflow is that it provides narrow (i.e., targeted) coverage of the metabolome.

In this study, we evaluate a novel approach for performing untargeted metabolomics that achieves broad coverage while leveraging the experimental efficiency of a targeted workflow. The basis of our work is a strategy for translating high-resolution MS^2 spectra into low-resolution bar codes without sacrificing the diagnostic specificity of the fragmentation patterns (Figure 1 and Figure 2).[19] The efficiency of the bar codes enables low-resolution, QqQ-based metabolomic work-
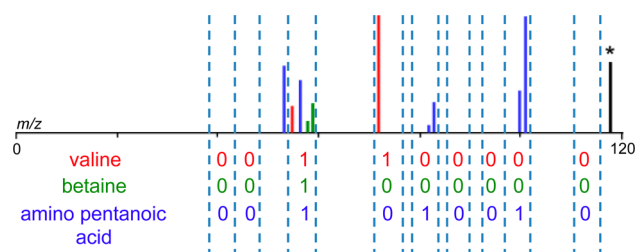
Figure 1. Schematic of MS^2 bar codes. High-resolution MS^2 spectra from three metabolites with the same precursor mass (starred in black) are shown. Each has a characteristic bar code.
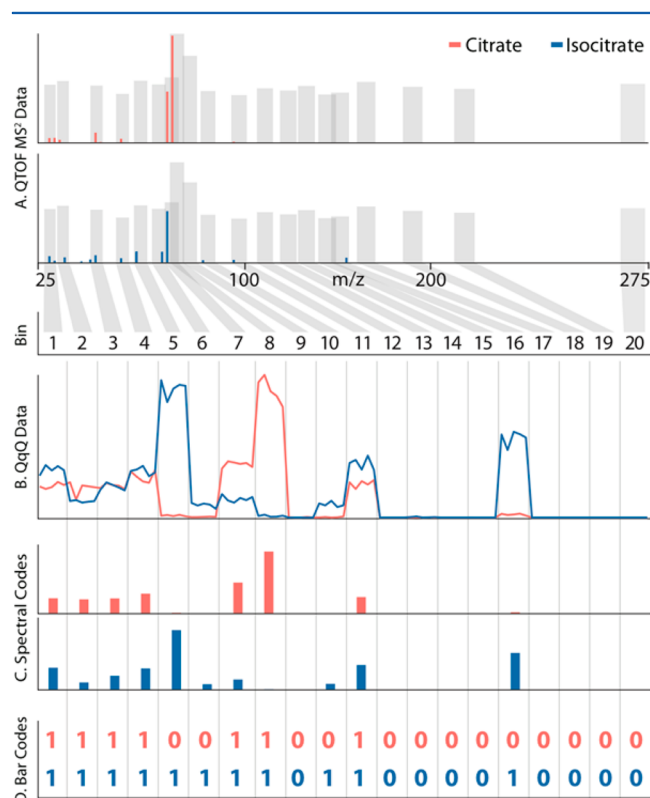

Figure 2. Transforming high-resolution MS^2 spectra into codes. (A) Full MS^2 spectra from a QTOF for citrate and isocitrate. Bins are shown in gray. Bins are staggered in height for clarity. (B) Experimental raw data from a QqQ in bar-coding mode. Bins are demarcated by vertical gray lines. (C) Representation of data as spectral codes. (D) Representation of data as bar codes. Citrate and isocitrate can be distinguished by both spectral codes and bar codes.

flows that have a specificity and sensitivity competitive to that of high-resolution time-of-flight (TOF) instruments.

■ EXPERIMENTAL SECTION

Validating Standard Fragmentation Data. Two solutions containing separate standards were prepared in 1:1 acetonitrile (ACN)/water (H$_2$O) at a concentration of 0.5 mM, enabling independent analysis of isomers without chromatographic separation. The first solution contained sucrose, citrate, tryptophan, and palmitate, while the second contained maltose, isocitrate, and aspartate.

To verify the high-resolution fragmentation data in METLIN for these metabolites, the standard solutions were directly injected at 50 μL/min into an Agilent 6540 UHD QTOF operating in ESI(−) mode. Targeted MS^2 data were collected

at collision energies of 0, 10, 20, and 40 V. MS source conditions were as described previously.[1] The fragmentor was set to 100 V, MS acquisition range was from 25 to 1700 m/z, and acquisition rate was set to 1 spectrum/s.

QqQ Analysis in Bar-Coding Mode. Fragmentation data for the standards were then reacquired, this time by aggregating fragment intensities for 20 selected m/z bins (Table 1). The

Table 1. Bin Windows Used for Bar Coding[a]

bin ID	lower m/z	upper m/z	bin ID	lower m/z	upper m/z
1	37.0	41.5	11	110.0	116.2
2	42.0	46.6	12	120.0	126.4
3	55.0	59.9	13	129.0	135.6
4	65.0	70.1	14	136.0	142.8
5	72.0	77.3	15	144.0	150.9
6	79.0	84.4	16*	149.0	156.1
7	84.0	89.6	17	159.0	166.3
8*	86.0	91.6	18	177.0	184.7
9	91.0	96.7	19	197.0	205.2
10	98.0	103.9	20*	262.0	271.7

[a]All bins are for a collision energy of 40 V unless indicated by an *, which denotes a collision energy of 20 V.

standard solutions were directly injected at 50 μL/min into an Agilent 6460 QqQ operating in precursor ion scanning mode, with the first quadrupole scanning from 100 to 400 m/z range in 0.1 Da steps and the third quadrupole cycling between the selected m/z bins at each precursor mass. For analysis of zebrafish extracts, an aminopropyl HILIC separation was used as previously described.[20] Bin collision energies were converted from the QTOF collision energies shown in Table 1 to equivalent collision energies for our QqQ 6460, which we estimated based on empirical evidence to be ∼7/10 of QTOF voltages. QqQ source settings were identical to those of the QTOF. However, because the third quadrupole's isolation window was set to match the width of our selected m/z bins, it had to be manually tuned. Tuning was accomplished in MS^2 scan mode by adjusting gain and offset settings for the third quadrupole so that the fwhm of the 113.0 m/z tuning calibrant peak matched our desired bin size for that mass (which was a width of 6.2). The following optimized settings for MS^2 unit-width isolation were then used in precursor ion scanning mode for data collection: width gain, 48; width offset, 2; axis gain, 41; axis offset, 1.5.

Sensitivity Measurements. To determine the sensitivity performance of bar coding, we compared the limits of detection (LODs) for every fragment of each standard compound as measured by three different MS^2 data acquisition methods. First, each standard solution was prepared at a range of concentrations (0.016, 0.033, 0.065, 0.13, 0.36, 0.52, 1.6, 3.1, 6.3, 13, 25, 50, 100, 200, and 400 μM). Next, each solution was injected directly into the instrument for analysis by TOF, bar-coding, and quadrupole MS^2-scan analysis, with a dwell time of 2 s/compound in each analysis mode. For the TOF measurements, the Agilent 6540 UHD QTOF was used in targeted MS^2 mode, with a 1 Da MS^1 isolation window, at collision energies of 20 and 40 V. Other QTOF settings were identical to those used above. For bar coding, the Agilent 6460 was used as described above, but with Q1 isolating each standard compound's precursor mass for 2 s instead of scanning. For the MS^2-scan measurement, the QqQ was returned to normal MS^2 unit-width isolation settings (width

gain, −19.6; width offset, dynamic; axis gain, 3.4; axis offset, dynamic) and then operated in MS^2-scan mode at each standard's precursor mass, scanning from 25 to 425 m/z values with 0.1 Da steps in 2 s, at collision energies of 14 and 28 V. Finally, each fragment's LOD under the different acquisition methods was normalized to its parent compound's LOD as measured by two MRMs targeting only its two most abundant fragments, once again with a dwell time of 2 s/compound. All LODs were defined as the concentration at which the signal-to-noise ratio for a given fragment becomes greater than 2.5, with noise level determined by the intensity signal from a 1:1 ACN/H_2O blank run in parallel.

■ RESULTS AND DISCUSSION

The current standard for supporting metabolite identifications in untargeted metabolomics is matching full, high-resolution MS^2 spectra from research samples to full, high-resolution MS^2 spectra of authentic model compounds.[8] Even with the newest instrumentation platforms and advanced informatic technologies, the process is tedious and inefficient. Here we considered an alternative strategy in which we converted the full, high-resolution MS^2 spectra of authentic model compounds into codes. As we discuss below, these MS^2 codes have high diagnostic specificity for supporting metabolite identifications. Yet, importantly, experimentally measuring a metabolite's code does not require full, high-resolution MS^2 spectra. Rather, codes can be effectively measured with low-resolution QqQ instruments.

In this study we consider two representations of binned high-resolution MS^2 spectra, which we refer to as spectral codes and bar codes (Figure 1 and Figure 2). To convert a high-resolution MS^2 spectrum into a spectral code, we sum the abundance of fragments in discrete m/z windows. We call each m/z window a bin. The width and location of our selected bins is defined in Table 1. The result is a spectrum where m/z-intensity pairs are replaced by bin-intensity pairs. A simplified representation of a spectral code is a bar code, in which only the presence or absence of fragments in a bin is considered.

When a metabolite produces fragments within a m/z window, we assign a "1" to that bin. When a metabolite does not produce a fragment in a m/z window, we assign a "0" to that bin. To decide if a fragment is present or absent, we first sum the intensities of all fragments in a metabolite's high-resolution MS^2 spectrum. Bins containing intensities less than 1% of the total signal are then designated as 0. The result of the approach is a string of 1's and 0's, or a bar code, for each metabolite. A schematic of the approach is shown in Figure 1. In this study, we exclusively focus on bar-coded MS^2 data, but spectral codes could be employed in an analogous matter.

An important step in bar coding MS^2 spectra is identifying informative bins with high diagnostic specificity. Previously, we determined the most unique precursor-to-product transitions that could be used to construct the fewest number of MRMs for profiling all metabolites in METLIN.[21] We used these informative fragments as the basis for selecting bins. In brief, we optimized the combination of bins that enabled us to resolve as many compounds with MS^2 spectra in METLIN as possible within 0.1 Da. We also considered experimental restrictions related to the tuning and capability of our instrument's quadrupoles. We found that bar codes from 20 bins were theoretically adequate to distinguish 96% of the compounds in the METLIN MS^2 library (Table 1). We note that future optimization of our approach may consider precursor-dependent bin selection or context-dependent selection of precursor masses. Sensitivity could also be improved by gaining programmatic access to the tuning of our instrument to increase the flexibility and transmission efficiency of the third quadrupole's isolation window.

Next, we considered how we could efficiently profile metabolites by using bar codes. We used a QqQ instrument to perform a new mode of data acquisition in which we cycled through 20 wide isolation windows with Q3 (corresponding to our bins) for every precursor mass selected by Q1. We refer to this method of operating a QqQ instrument as "bar-coding mode". When we scan all precursor masses with Q1, we obtain a comprehensive list of bar codes for each sample (Figure 2). We point out that using a QqQ instrument in bar-coding mode is different than using a QqQ instrument in the traditional precursor ion scanning mode or MRM mode where the goal is to use Q3 to isolate a specific fragment with a unique mass value. By using multiple large windows that coincide with our bins, we substantially improve experimental efficiency and coverage with only a minor compromise in specificity. Additionally, the data output consists of a binned spectrum that can be transformed into a string of 1's and 0's for each precursor mass and matched to the bar codes of standard compounds.

To demonstrate the approach, we analyzed a set of standard compounds with a QqQ operating in bar-coding mode. The standards we analyzed included two sets of structural isomers, citrate/isocitrate and maltose/sucrose, which were selected to highlight that bar codes can distinguish between noncoeluting metabolites with the same precursor mass. The raw data for citrate and isocitrate are shown in Figure 2. As above, the intensity of the signal in each bin was integrated and compared to the total. Bins with intensities less than 1% of the total were assigned 0's, while all other bins were assigned 1's. The experimental bar codes we generated matched the theoretical bar codes we calculated by using high-resolution METLIN data (Figure 3).

Figure 3. Predicted and experimental bar codes for several metabolites. Theoretical bar codes were determined on the basis of MS^2 data in METLIN. Experimental data were acquired with a QqQ instrument and then translated into 1's and 0's.

It is important to emphasize that the 20 bins we have used here are optimized for the analysis of biochemically complex samples. Indeed, we were able to identify the same 7 metabolites in a zebrafish extract by using a QqQ in bar-coding mode. This result demonstrates the applicability of bar coding to complex biological samples.

Having established the specificity of our bar-coding approach, we next considered its sensitivity relative to other methods of obtaining MS[2] data. For the same set of standards used above, we measured the limit of detection for each fragment with bar-coding, QqQ-based MS[2] scanning, and TOF experiments. We used a QTOF in targeted MS[2] mode to generate precursor isolation and fragmentation for the TOF experiment comparable to that of the QqQ experiments. We then normalized these data to the limit of detection measured for each compound by monitoring its two most abundant fragments with MRM experiments. All experiments performed were allotted the same dwell time per compound. Strikingly, we found that measuring fragmentation patterns with 20 bar codes has a sensitivity comparable to that of measuring fragmentation patterns with a TOF (Figure 4). As expected, the sensitivity of
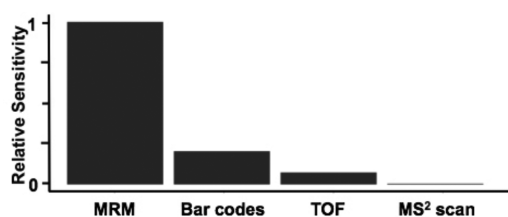

Figure 4. Relative sensitivity of different MS[2] acquisition methods. Our analysis is based on a comparison of the limits of detection for each fragment of our standards (see text for details). We normalized to the limit of detection measured by MRM experiments in which the two most abundant fragment intensities were monitored for each compound. The data shown here represent the average of each experiment.

QqQ-based MS[2] scanning was 2 orders of magnitude worse. Interestingly, bar-code experiments were more sensitive than predicted based on the duty-cycle considerations of MRM and TOF type experiments. We considered the duty cycle of a MRM experiment as one over the number of transitions (1/20 or 5%) and the duty cycle of a TOF as 10−20% due to ion loss during pulsing.[22] We rationalize these results based on the enhanced transmission efficiency of larger quadrupole isolation windows and the increased signal that results from multiple ion fragments contributing to the same bin.[1]

We have shown here that bar coding provides a combination of specificity, sensitivity, and compound coverage for metabolomics that is competitive to other methods of MS[2] data acquisition. We imagine that bar coding could be integrated into metabolomic workflows in several ways. We outline two possibilities below.

One possible approach for generating a large number of metabolite identifications is a QqQ-based workflow in which the instrument scans the full range of potential metabolite precursor masses in bar-coding mode. When using the 20 bins calculated in Table 1, this approach will theoretically enable the identification of 96% of the compounds in the METLIN MS[2] database from complex biological samples (assuming, as with any MS[2]-based identification method, that chromatography is used to prevent compounds with the same mass from coeluting). As we have demonstrated, LC coupled bar coding enabled us to identify a select group of metabolites in zebrafish extract. However, a practical consideration is the trade-off between sensitivity and the number of precursor masses selected. Scanning over the entire range of possible metabolite precursor masses with Q1 results in a low duty cycle and will be

particularly problematic for low-abundance metabolites. Therefore, we imagine a workflow in which the precursor mass space is divided into smaller ranges and analyzed over multiple LC/MS sample runs. As an alternative solution, precursor mass space could be constrained on the basis of a prescan in which Q1 operates in rf-only mode and a summed code is detected. This prescan provides information about the possible precursors that generated the summed code and thus limits the number of potential precursor masses present.

A second possible application of bar coding could be for targeted studies in place of MRM experiments. In this application, Q1 only selects precursor masses of interest while Q3 operates in bar-coding mode. An advantage of this approach is that it enables analytical specificity and sensitivity to be accomplished on a QqQ instrument without the labor-intensive development of MRM methods.

CONCLUSIONS

By considering the presence or absence of fragments in 20 discrete m/z bins, we have transformed high-resolution MS[2] spectra into bar codes of 1's and 0's that can be measured with low-resolution QqQ instruments. Notably, this informatic compression does not substantially reduce specificity with respect to metabolite identification when we select bins that contain the most informative MS[2] fragments (Table 1). Additionally, we have shown that QqQ instruments operating in bar-coding mode can acquire MS[2] spectra with a sensitivity comparable to that of TOF instruments. Therefore, we believe that the application of bar codes is a promising and unexplored avenue for untargeted metabolomics.

AUTHOR INFORMATION

Corresponding Author
*Phone: 314-935-3512. E-mail: gjpattij@wustl.edu.
Notes
The authors declare no competing financial interest.

REFERENCES

(1) Nikolskiy, I.; Mahieu, N. G.; Chen, Y. J.; Tautenhahn, R.; Patti, G. J. *Anal. Chem.* **2013**, *85*, 7713−9.
(2) Sumner, L. W.; Lei, Z.; Nikolau, B. J.; Saito, K. *Nat. Prod. Rep.* **2015**, *32*, 212−29.
(3) Salek, R. M.; Steinbeck, C.; Viant, M. R.; Goodacre, R.; Dunn, W. B. *GigaScience* **2013**, *2*, 13.
(4) Dunn, W. B.; Erban, A.; Weber, R. J. M.; Creek, D. J.; Brown, M.; Breitling, R.; Hankemeier, T.; Goodacre, R.; Neumann, S.; Kopka, J.; Viant, M. R. *Metabolomics* **2013**, *9*, S44−S66.
(5) Benton, H. P.; Ivanisevic, J.; Mahieu, N. G.; Kurczy, M. E.; Johnson, C. H.; Franco, L.; Rinehart, D.; Valentine, E.; Gowda, H.; Ubhi, B. K.; Tautenhahn, R.; Gieschen, A.; Fields, M. W.; Patti, G. J.; Siuzdak, G. *Anal. Chem.* **2015**, *87*, 884−91.
(6) Gerlich, M.; Neumann, S. *J. Mass Spectrom.* **2013**, *48*, 291−8.
(7) Schymanski, E. L.; Neumann, S. *Metabolites* **2013**, *3*, 412−39.
(8) Patti, G. J.; Yanes, O.; Siuzdak, G. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 263−9.

(9) Kind, T.; Liu, K. H.; Lee, D. Y.; DeFelice, B.; Meissen, J. K.; Fiehn, O. *Nat. Methods* **2013**, *10*, 755−758.

(10) Tautenhahn, R.; Cho, K.; Uritboonthai, W.; Zhu, Z.; Patti, G. J.; Siuzdak, G. *Nat. Biotechnol.* **2012**, *30*, 826−8.

(11) Stein, S. *Anal. Chem.* **2012**, *84*, 7274−82.

(12) Sud, M.; Fahy, E.; Cotter, D.; Azam, K.; Vadivelu, I.; Burant, C.; Edison, A.; Fiehn, O.; Higashi, R.; Nair, K. S.; Sumner, S.; Subramaniam, S. *Nucleic Acids Res.* **2016**, *44*, D463.

(13) Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G. *Ther. Drug Monit.* **2005**, *27*, 747−51.

(14) Roberts, L. D.; Souza, A. L.; Gerszten, R. E.; Clish, C. B. Targeted Metabolomics. In *Current Protocols in Molecular Biology*; John Wiley & Sons, Inc.: Hoboken, NJ, 2012; Chapter 30, Unit 30.2, pp 1−24, DOI: 10.1002/0471142727.mb3002s98.

(15) Bajad, S. U.; Lu, W.; Kimball, E. H.; Yuan, J.; Peterson, C.; Rabinowitz, J. D. *J. Chromatogr A* **2006**, *1125*, 76−88.

(16) Henry, H.; Sobhi, H. R.; Scheibner, O.; Bromirski, M.; Nimkar, S. B.; Rochat, B. *Rapid Commun. Mass Spectrom.* **2012**, *26*, 499−509.

(17) Perkel, J. Quantification of Complex Samples Using Multiple Reaction Monitoring. http://www.biocompare.com/Editorial-Articles/117311-Quantification-of-Complex-Samples-Using-Multiple-Reaction-Monitoring (accessed Dec. 23, 2015).

(18) Shushan, B. *Mass Spectrom. Rev.* **2010**, *29*, 930−44.

(19) Patti, G. J. When Clinical Lab Tests Fail: New Paradigms of Metabolic Regulation. *Mass Spectrometry: Applications to the Clinical Lab*, Plenary Lecture, San Diego, CA, March 1−5, 2014.

(20) Mahieu, N. G.; Spalding, J.; Patti, G. J. *Bioinformatics* **2015**, btv564.

(21) Nikolskiy, I.; Siuzdak, G.; Patti, G. J. *Bioinformatics* **2015**, *31*, 2017−23.

(22) Bateman, R. H.; Carruthers, R.; Hoyes, J. B.; Jones, C.; Langridge, J. I.; Millar, A.; Vissers, J. P. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 792−803.