

Characterization of glycosphingolipids and their diverse lipid forms through two-stage matching of LC-MS/MS spectra

Laura S. Bailey,[†] Fanran Huang,[†] Tianqi Gao,[†] Jinying Zhao,[‡] Kari B. Basso,[†] and Zhongwu Guo^{†,*}

[†]Department of Chemistry and [‡]Department of Epidemiology, University of Florida, Gainesville, Florida 32611, USA

*Corresponding author E-mail address: zguo@chem.ufl.edu

ABSTRACT: Glycosphingolipids (GSLs) play a key role in various biological and pathological events. Thus, determination of the complete GSL compositions in human tissues is essential for comparative and functional studies of GSLs. In this work, a new strategy was developed for GSL characterization and glycolipidomics analysis based upon two-stage matching of experimental and reference MS/MS spectra. In the first stage, carbohydrate fragments, which contain only glycans and thus are conserved within a GSL species, are directly matched to yield a species identification. In the second stage, glycolipid fragments from the matched GSL species, which contain both the lipid and glycans and thus shift due to lipid structural changes, are treated according to lipid rule-based matching to characterize the lipid compositions. This new strategy uses the whole spectrum for GSL characterization. Furthermore, simple databases containing only a single lipid form per GSL species can be utilized to identify multiple GSL lipid forms. It is expected that this method will help accelerate glycolipidomics analysis and disclose new and diverse lipid forms of GSLs.

Glycosphingolipids (GSLs) are a class of diverse and complex glycolipids having a sphingoid, most commonly a ceramide, as the hydrophobic lipid moiety linked to the downstream (also known as the reducing) end of a hydrophilic glycan via a β -glycosidic bond (Figure 1). Both the glycan and the lipid can vary in natural GSLs to generate different GSL species and diverse lipid forms for each species.¹⁻² Based upon the core structures of glycans, GSLs are further classified into different series.³

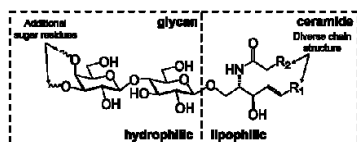


Figure 1. A representative structure of GSL. Structural diversity is resulted from variations in the glycan and ceramide moieties.

GSLs are an essential component of the cell membrane and a key player in various biological and pathological processes.⁴⁻⁶ For example, GSL is the main form of glycoconjugates in vertebrate brains, where >80% of the glycans in the cell glycocalyx are glycolipids.⁷ Therefore, GSLs are directly involved in signal transduction,⁸ cell division, recognition, adhesion and apoptosis,⁹⁻¹⁰ and embryonic and nerve system development.¹¹⁻¹³ In addition, abnormal GSL expression and metabolism are correlated to many diseases, such as cancer¹⁴⁻¹⁵ and Alzheimer's disease.¹⁶⁻¹⁷ Because of their biological significance, GSLs have been a major research focus across different disciplines, among which GSL structural analysis is especially imperative. Studies on GSLs expressed by various cells/tissues (*i.e.*, glycolipidomics analysis) to reveal abnormal GSL expression would not only help gain a deeper understanding of the functions and action mechanisms of GSLs in various processes but also help identify new markers useful for disease diagnosis and therapy.

In the omics era, the high sensitivity of mass spectrometry (MS) makes it one of the most powerful tools for proteomics, lipidomics, and glycomics analyses. Similarly, great effort has been put on MS-based analysis of glycolipidomics as well. Although shotgun MS analysis for GSL profiling is promising,¹⁸⁻²⁰ ion suppression of low abundance species and a high likelihood of isobaric overlaps often necessitate a liquid chromatographic (LC) separation step prior to MS detection. For this purpose, a variety of LC techniques, such as normal and reversed phase²¹⁻²⁸ and hydrophilic interaction LC,²⁹⁻³² in addition to many others (*e.g.*, nano-chip-based capillary electrophoresis³³⁻³⁴ and high-performance thin layer chromatography³⁵), have been explored. All of these techniques can yield general separation of GSL species, whereas reversed phase-LC is the most widely used due to its ease of operation, reproducibility, and compatibility with on-line MS analysis. However, as isobaric differentiation with LC is often limited, ion mobility MS has been employed as a replacement or complement to current separation techniques.³⁶⁻³⁷ Furthermore, tandem MS (MS/MS), which can yield detailed structural information, is commonly included in the analysis. MS/MS can be performed through many different activation techniques, but collision induced dissociation (CID) is the most widely adopted. Therefore, reference libraries are more likely to contain the CID product spectra, or spectra generated by similar higher energy collision dissociation (HCD). Combined, LC-MS/MS and similar technologies have allowed the characterization of many natural GSLs and thereby a better understanding of GSLs in recent years.^{2, 38-39}

Despite these great progresses in the field, GSL analysis is still a significant challenge. Several obstacles have impeded GSL analysis. First, there is not a comprehensive database containing the reference MS/MS spectra of GSLs to enable routine analysis. This is at least partially due to the difficulty in obtaining pure

GSL standards to populate MS/MS databases as GSL synthesis is arduous and GSL isolation from nature is problematic because of their amphiphilicity. Second, the unique and diverse structures of GSLs complicate LC-MS/MS studies. In contrast to glycoprotein analysis, where protein and glycan can be cleaved and analyzed separately,⁴⁰ GSLs are analyzed intact, which is complicated by not only the diversity of GSL species but also assorted lipid forms for each species. Third, a lack of a widely adopted, high throughput analytical method and corresponding interpretative software has further slowed GSL identification and analysis.

To address the above issues, our laboratories are working on several fronts, including developing new synthetic strategies for rapid access to GSLs and their different lipid forms,⁴¹ as well as new MS/MS-based method for high throughput GSL analysis. Here, we report on the development of a new method for GSL and related lipidomics analysis, which is based upon a two-stage comparison of experimental LC-MS/MS data against a reference database. This reference database needs to contain only a limited number of standard GSL MS/MS spectra to determine both GSL species and lipid form.

Materials and Methods

Materials. GSL samples were purchased from Avanti Polar Lipids (GM3 and GM3-d₅), Biosynth Carbosynth (GM2), and ChemCruz Biochemicals (GD2) and used without further purification. GM2 and GD2 were prepared as individual solutions; GM3 and GM3-d₅ as a mixed solution. All were dissolved in methanol to get a final concentration of 50 µg/mL for LC-MS/MS analysis. LC/MS grade formic acid, certified crystalline ammonium formate, and optima LC/MS grade solvents were purchased from Fischer Chemical.

Liquid Chromatography. The above samples (5 µL) were injected into a Thermo Fischer Scientific Acclaim PepMap RSLC C18 column (300 µm x 15 cm, 2 µm, 100 Å) with a C18 pre-column (3 mm x 2 cm, 75 µm, 100 Å) of the same material on a Thermo Fischer Scientific UltiMate 3000 RSLCnano system equipped with degasser, pump, column compartment, and autosampler. The column was maintained at a constant 40 °C during the analysis, and the autosampler was maintained at 4 °C. Mobile phases (A) 60/40% acetonitrile/water and (B) 90/8/2% isopropanol/acetonitrile/water, both containing 10 mM of ammonium formate and 0.1% formic acid, were used for the gradient pump at a flow rate of 5 µL/min. Samples were loaded to the pre-column and washed for 5 minutes with 98/2% water/acetonitrile containing 0.1% formic acid at a 25 µL/min flow rate, after which the switching valve was activated to initiate column separation. The gradient pump operated at 50%B for 5 minutes before ramping to 98%B at 50 minutes and holding for 20 minutes. Then, the mobile phases were returned to starting conditions at 75 minutes, and the column was re-equilibrated before the next injection.

Mass Spectrometry. MS analysis was performed on an Impact II QqTOF mass spectrometer (Bruker, Billerica, MA) using Apollo electrospray ionization (ESI) in positive mode. The MS was operated at a capillary voltage of 4.0 kV, nebulizer of 0.3 bar, and nitrogen dry gas flow rate of 4.0 L/min and temperature of 200 °C. The instrument was programmed for data dependent acquisition (DDA) for CID using a nitrogen gas collision partner. DDA was selected for singly (mainly) and doubly charged ions in the mass range m/z 500-1500 using collision energies

around 25 and 30 eV, depending on the programmed mass bracket. Active exclusion was used to exclude an ion after a single spectrum for 3 minutes unless the current-to-previous intensity ratio was ≥ 2.0 .

Spectral analysis was performed using Compass (v.5.1; Bruker Daltonics). Carbohydrate fragmentation mass assignments were manually achieved using a theoretical mass list generated by GlycoWorkBench.⁴² Additional details about the theoretical mass list generation and product ion annotation are available in SI.

Results and Discussion

Our investigation was performed on commercial GSL samples, which are derived from biological sources. These samples were subjected to LC-MS/MS analysis using positive mode ESI and CID with DDA to activate mainly singly charged ions in a range of m/z 500-1500. Ammonium formate as a mobile phase additive enhances positive mode ionization efficiency for glycolipids as observed in the literature⁴³ and in this study. The first GSL studied was GM3 (36:1), consisting of a linear trisaccharide (Figure 2A insertion and SI) and a 36-carbon ceramide. Its MS/MS spectrum represents a simple GSL product spectrum (Figure 2A), containing only 12 product ions in total, dominated by 7 highly abundant ions (most of >50% relative abundance). For this and all other GSLs studied herein, due to the presence of ammonium formate in the system, the ammoniated ions were typically equally or more abundant than the protonated ions but they had identical fragmentation patterns, as exemplified by the spectra of protonated and ammoniated GM3-d₅ (SI, Figure S1). Thus, we choose the protonated ions for simplicity sake.

The product ions of GM3 (36:1) (Figure 2A) can be readily interpreted following the general GSL dissociation patterns. Glycolipid fragmentation nomenclature of proposed by Domon and Costello is depicted in Figure 3.⁴⁴ Dissociation of the glycan with charge retention upstream (*i.e.*, on the glycan non-reducing end side) results in A, B, and C-type product ions, which contain carbohydrate only. Here, we call them “carbohydrate fragments” or “carbohydrate product ions”. This type of dissociation is illustrated by the B₁, B₂, and B₃ product ions of GM3. It is important to note that these fragments no longer contain the ceramide thus their masses will not be affected by the lipid structure. Alternatively, fragmentation of the glycan with charge retention downstream (*i.e.*, on the reducing end side) gives X, Y and Z-type ions, which contain both glycans and the ceramide. Here, we call them “glycolipid fragments” or “glycolipid product ions”. This type of dissociation is illustrated by the Y₀, Z₀, Y₁, Z₁, Y₂, and Z₂ product ions of GM3. In addition, according to the nomenclature, B, C, Y, and Z-type dissociations occur along a glycosidic bond, while A and X-type fragmentations involve cleavage of two bonds in the sugar ring. Usually, the former dissociations are predominant for CID activation, as the latter dissociations involve an additional C-C bond cleavage that necessitates higher energies.⁴⁴⁻⁴⁵ It is noteworthy that all these dissociations occur on the glycan and the ceramide remains intact. Thus, this nomenclature should not be confused with fragmentation of the ceramide moiety.

Two GM3 (36:1) product ions are associated with the loss of a water molecule from glycolipid fragments—the precursor ion ($[M - H_2O + H]^+$) and the Z₀ ion ($[Z_0 - H_2O]$). As these ions contain the lipid moiety and follow similar trends to X/Y/Z-type fragments, we will include them in the discussion of glycolipid

fragments. Another product ion with the loss of a water molecule was from B_1 ($[B_1 - H_2O]$). As with the B_1 ion, this ion no longer contains the lipid and, therefore, will be discussed with the carbohydrate fragments. The last GM3 (36:1) product ion to discuss is a small ion at m/z 264, which matches the sphingosine N'' fragment. For all GSLs studied, this ion appeared at a much

lower abundance than other product ions, if it is observed at all. Even for simple sphingolipids, this ion is usually noted to be of low abundance.⁴⁵⁻⁴⁶ A table of all the product ions observed for GM3 (36:1), as well as their assignments, is provided in Tables S1 and S2 of SI.

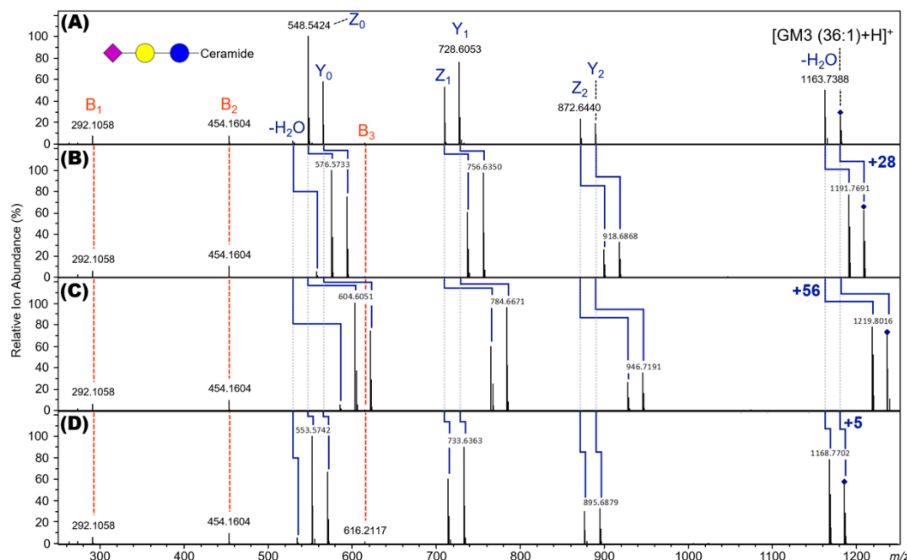


Figure 2: Product spectra of GM3 (A) (36:1), (B) (38:1), (C) (40:1), and (D) (36:1)- d_5 by CID of the protonated precursor ion. (A) also contains a depiction of the GM3 structure and some fragment labeling. The dashed orange lines highlight fragments appearing at the same m/z in all spectra. The solid blue lines highlight fragments that increase in mass from (A) to other spectra. The dotted light gray lines indicate the mass observed in (A). The blue diamond indicates the precursor ion (all protonated) in each product spectrum.

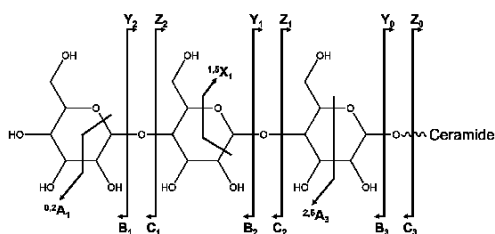


Figure 3: Fragmentation scheme for glycan cleavages.⁴⁴ The line indicates cleaved bond(s) and the arrow direction indicates which side the charge retains. Subscript represents the number of sugar residues retained in the product ion. Superscript represents the two cleaved bonds in a sugar ring (A and X fragments only).

In addition to GM3 (36:1), the sample contained at least two other GSLs. Their product spectra are shown in Figures 2B and 2C. The protonated precursor ions ($[M + H]^+$) of these GSLs were m/z 1209.7794 and 1237.8114, m/z 28 and 56 greater than that of GM3 (36:1) ($[M + H]^+$: m/z 1181.7505). These mass differences corresponded to 2 and 4 methylene groups ($-CH_2-$), respectively. However, the overall fragmentation patterns of all three spectra were very similar, including the relative intensities of their product ions. This follows a trend expected for different lipid forms (*i.e.*, the fragment masses increase or decrease with lipid mass increases or decreases but retaining ion separation), thus, the general appearance of fragmentation patterns remains. Accordingly, we suspected that the product spectra in Figures 2B and 2C were from two heavier lipid forms of GM3 (36:1), namely GM3 (38:1) and GM3 (40:1), respectively. It should be

noted that different lipid forms of a GSL species mean that these GSLs contain the same glycan but different ceramides.

Detailed analysis of the product spectra in Figures 2A, 2B, and 2C provided proofs for the above hypothesis. For example, all three spectra showed the same carbohydrate fragments (B-type ions) at the similar relative intensities, suggesting that they contained the same glycan moiety. These ions are connected in Figures 2A, 2B, and 2C with orange dashed lines. The glycolipid fragments (Y and Z-type ions as well as the dehydrated ions) in Figures 2B and 2C followed a similar trend as their precursor ions. The masses of these ions were found to increase by 28 and 56 mass units, respectively, as compared to those from GM3 (36:1). These shifts obeyed pre-established behavior for lipid fragmentation patterns. The 8 glycolipid product ions that increase in mass as compared to GM3 (36:1) are highlighted with connecting blue lines in Figures 2A, 2B and 2C. All the above evidence suggest that these three glycolipids are different lipid forms of GM3.

An interesting and important observation for the MS/MS spectra of different GM3 lipid forms was the similarity of their carbohydrate product ions. Despite differences in the GSL lipid structure, their carbohydrate product ion masses and relative intensities were the same. In addition, all the glycolipid product ions had the similar relative intensities, although the ion masses changed due to different lipid structures. This may suggest that differences in the ceramide do not have a significant impact on the dissociation pattern for each GSL species.

This observation was studied through comparison of GM3 (36:1) to its isotopically labeled lipid form GM3- d_5 (36:1). The

labeled lipid has five hydrogens on the ceramide acyl chain (R_2 , Figure 1) exchanged for deuterium, making the lipid 5.031 Da heavier than its natural counterpart. It was expected that the H-D exchange within the ceramide would have little to no impact on the fragmentation pattern of GM3, aside from the anticipated mass shift for the glycolipid fragments. As with the other two lipid forms of GM3 (Figures 2B and 2C), the product spectrum of GM3- d_5 (36:1) did show three carbohydrate fragments at the same ion masses (Figure 2D) as those of GM3 (36:1). The Y and Z-type fragments and the dehydrated ions, all of which contained ceramide, increased by m/z 5.031 as compared to that of natural GM3 (36:1). Again, the relative intensities of all product ions of GM3- d_5 (36:1) were the same as those of GM3 (36:1). These results confirmed that deuteration of the lipid in GM3 did not affect either the masses or the relative intensities of carbohydrate fragments nor the relative intensities of glycolipid fragments, as compared to the natural GM3 (36:1).

To further verify the discovery and probe its scope, next, we studied GM2, a more complex GSL containing a branched tetrasaccharide (Figure 4A insertion). The MS/MS spectrum of GM2 (36:1) (Figure 4A) showed 27 product ions (precursor ion and its dehydrated derivative not shown). The observed product ions and their assignments are provided in SI (Tables S3 and S4). The majority of these ions (16 out of 27) were assigned to products of single glycosidic bond cleavages (full B, C, Y, and

Z coverage). Nine were internal carbohydrate fragments formed from the cleavage of two glycosidic bonds. (Cleavage of two or more glycosidic bonds within a glycan with charge remaining on a carbohydrate fragment that is not the non-reducing end piece generates the so-called internal carbohydrate fragment, which is common for complex and/or branched glycans. If the charge remains on non-reducing end or glycolipid fragments, the product ions should be the same as those generated from the conventional scenario.) The possibility of multiple glycosidic bond cleavages in complex glycans would significantly increase the spectral complexity. It should be noted, however, that although internal fragments do increase the spectral congestion, they still follow the carbohydrate/glycolipid fragment behavior we have already established.

Figure 4 shows the product spectra for three lipid forms of GM2: (36:1) in 4A; (38:1)—a 28 Da heavier lipid form—in 4B; (40:1)—a 56 Da heavier lipid form—in 4C. As with different lipid forms of GM3, the product spectra in Figures 4B and 4C showed a series of product ions (12 ions) that uniformly shifted to higher masses (corresponding to heavier lipids) as compared to the product ions in Figure 4A. Additionally, an almost equal number of product ions (11 ions) retained the same masses in all three spectra. In Figure 4, the most conserved B, C, and BY-type ions are identified by dashed orange lines, and the shifted Y, Z, and YZ-type ions are identified by solid blue lines.

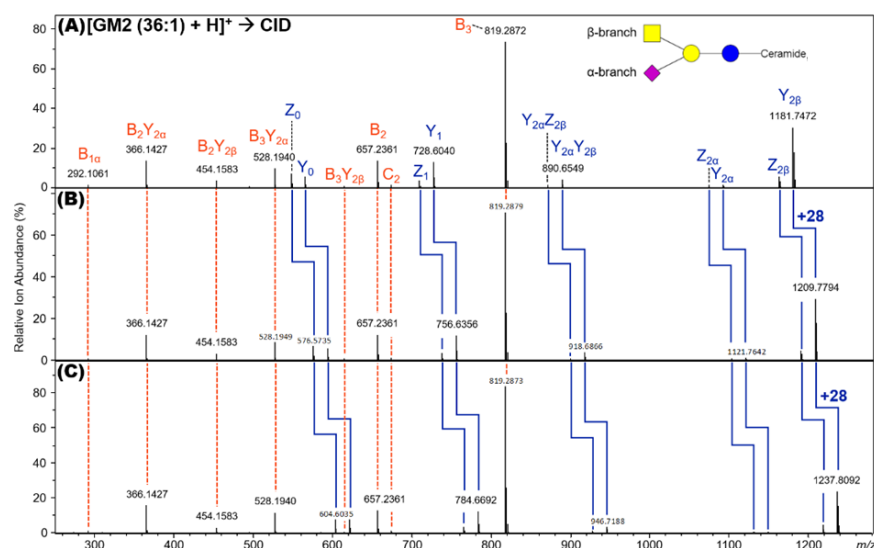


Figure 4: Product spectra for GM2 (A) (36:1), (B) (38:1), and (C) (40:1) by CID of the protonated precursor ion (not shown). (A) also contains a representation of the GM2 structure and some fragment labeling. The orange dashed lines highlight fragments appearing at the same m/z in all three spectra. The blue solid lines highlight fragments that increase by m/z 28 between spectra.

The spectral differences between GM2 and GM3 are evident. For example, the GM2 spectra contained more (nearly twice) product ions than that of GM3. The fragment intensity distributions of GM2 and GM3 were also significantly different. The most abundant product ion for GM2 was a carbohydrate fragment B_3 (m/z 819.2872), but it was a glycolipid fragment Z_0 for GM3. In addition, the abundances of GM2 glycan and glycolipid product ions were comparable, but for GM3, the glycolipid product ions were predominant. The stark differences in the fragmentation patterns of GM2 and GM3 that differ by only one sugar residue illustrate how complicated the product spectra can become as glycan complexity increases. Conversely, the stark

spectral differences among various GSLs form a solid foundation for their MS/MS-based characterization.

As demonstrated above, each GSL species yields a unique product spectrum. It is easily perceivable that the unique spectra of various GSLs can be utilized to differentiate and characterize them.⁴⁷⁻⁵⁰⁹ A more interesting property revealed through above analyses is that each GSL species always contains a unique set of carbohydrate fragments that are not affected by the ceramide at all. This set of conserved product ions is determined only by the unique species structure (*i.e.*, glycan structure). The above data further revealed that the glycolipid fragments from different lipid forms within a GSL species differ only in their ion

mass, due to changes in the ceramide mass; yet, the general spectral pattern (*i.e.*, mass separation between glycolipid fragments and ion intensities) are unaffected by ceramide changes. These properties might be useful in exploring the diverse lipid forms of each GSL species, a significant challenge in the current state of glycolipid analysis.

To address the lipid diversity issue, a rule-based matching strategy was developed for lipidomics analysis,⁵¹⁻⁵⁴ where lipids are identified based on the product spectrum similarity within a lipid class. To identify an experimental lipid whose product spectrum is not physically in the database, the fragments in the reference spectrum are shifted to a higher or lower mass to achieve spectral match with the experimental spectrum. As a result, rule-based matching programs do not require every lipid variation to be in the database but utilize a few example product spectra for each lipid class to make identifications. This is possible due to the fact that most lipid classes are highly predictable in their fragmentation patterns, namely that the vast majority of their fragments are produced by headgroup cleavages, whereas the lipid chains remain intact.

Although the rule-based matching method holds great potentials for lipidomics analysis,^{52,55} it is not directly applicable to GSLs, although GSL cleavages also occur mainly at the headgroup—a glycan, in this case. There are several reasons for that. First, although GSLs and other lipids are alike in that both can produce lipid-free and lipid-containing fragments, the lipid-free fragment for other lipids is usually a single product ion, if it is observed at all, whilst the lipid-containing fragments are more numerous. In contrast, the GSL product spectrum usually contains similar numbers of lipid-containing and lipid-free fragments, making the GSL product spectrum much more complex than that of other lipids. This notably increases the challenge of GSL spectral analysis—a challenge that grows with increased complexity of glycan. Second, the large majority of GSLs share a disaccharide (*i.e.*, lactose) core linked to the ceramide moiety (see the GM3 and GM2 structures in Figures 2A and 4A). As a result, the Z_0/Y_0 , Z_1/Y_1 , and Z_2/Y_2 fragments of different GSL species can have the same product ions (although their intensities may vary among GSL species), which further complicates the product spectral analysis. Third, and most important, as the carbohydrate product masses are conserved and would not shift with different lipid masses, rule-based mass shifts, as described above, would result in mismatch of the carbohydrate fragments and incomplete pattern matching for GSLs.

An illustrative example of the lipid rule-based mismatch for GSL is presented in Figure 5. Let us imagine that a database contains the reference spectrum (Figure 5B) of a GSL species that the experimental GSL (Figure 5A) belongs to but exists in a lighter lipid form. As in the previous cases, the carbohydrate fragments in the reference spectrum would match those in the experimental product spectrum, but the glycolipid fragments in the database spectrum would appear at higher masses than those in the experimental spectrum. The light gray lines highlight all of the ions that need to be matched for a positive identification. Clearly, due to the incomplete match between Figures 5A and 5B, a positive identification cannot be made. Rule-based match endeavors to make identifications by shifting the reference fragments to lower (or higher) masses as needed to match experimental glycolipid fragments, but this results in a uniform shift of all peaks. Now, the resulting spectrum (Figure 5C) has the

glycolipid fragments (green), but no longer carbohydrate fragments (red), matching those in the experimental spectrum. As a result, the application of rule-based matching fails and results in no GSL identification.

A potential solution for the above mismatch problem would be to store only the glycolipid product ions for each GSL reference spectrum in the database. In this case, only a part of the fragments are used for the rule-based matching, as illustrated in Figure 5D. In other words, many of the product ions would need to be ignored for the identification. Using GM2 as an example (Figure 4), *ca.* 50% of the fragments (*i.e.*, glycolipid fragments) would be used for matching purposes but the remaining species-unique glycan-only fragments (*i.e.*, the carbohydrate fragments) would be discarded. As the glycan size increases, the number of carbohydrate fragments also increases, resulting in an even greater percentage of the data being ignored. In most cases, the discarded fragments are also the most abundant product ions. Moreover, some fragments (Z_0/Y_0 , Z_1/Y_1 , and Z_2/Y_2 fragments) retained for matching are shared among different GSL species, further reducing the number of characteristic fragments for GSL identification. Undoubtedly, this strategy may increase the risk of misidentification.

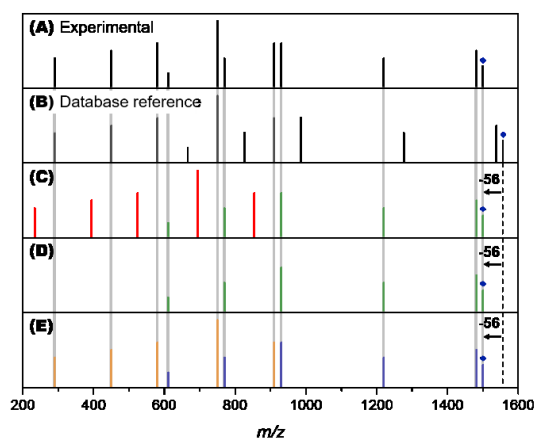


Figure 5: Illustration of databased matching for the identification of an experimental GSL, based on its MS/MS spectrum (A). (B) is a reference spectrum in the database, representing a 4-carbon heavier (+ m/z 56) lipid form of the same species as experimental GSL. Rule-based matching strategies involve storing/shifting all the fragments (C) or only glycolipid fragments (D) in a reference spectrum (red peaks indicate mismatched fragments and green peaks indicate a matched fragments). (E) Proposed 2-stage matching method with the stages color-coded for the carbohydrate fragment matching (orange) and the glycolipid fragment shifting (blue). For all, the blue diamond indicates the precursor ion and the gray lines provided reference to the experimental spectrum.

To overcome these problems, we propose a new two-stage matching method, as depicted in Figure 5E, for more reliable MS/MS-based characterization of GSLs. This method relies on separately matching the carbohydrate and glycolipid fragments of the experimental and reference product spectra in two steps. To illustrate the process, database fragments in Figure 5E are color-coded by stages. The first stage matches only the carbohydrate fragments (orange). As these fragments always retain the same ion masses, the reference and experimental fragments are compared directly. This will identify the glycan structure,

and thus the GSL species. After the glycan is identified, the second matching stage is to treat the glycolipid fragments (blue) from the reference spectrum in accordance to the traditional lipid rules; in this case, shifting the reference fragments to lower masses to match the experimental fragments. As with routine rule-based matching, the mass difference between the reference and experimental spectra indicates the general lipid structure.

As models for this new two-stage method, its application to the previous two GSLs, GM3 (Figure 2) and GM2 (Figure 4), results in complete matches of all the fragments in experimental product spectra of other lipid forms (middle and bottom panels in each figure) with those in the reference product spectrum (top panel of each figure). Furthermore, this would not only result in mass matches but also enable the matching of ion abundances, something that cannot be easily done with glycan sequencing methods, such as those used in most glycomic analyses.⁵⁶

Clearly, this new two-stage matching method can have a number of advantages. Most significantly, it can use all available information in the experimental and reference spectra for GSL characterization, which would reduce the risk of false identification. Furthermore, it requires only a single reference spectrum of one lipid form from each GSL species for the identification and characterization of other lipid forms.

To verify the two-stage matching method for GSL characterization and demonstrate its practicability, we applied it to the analysis of GD2 and its lipid forms. To this end, we used the product spectrum of GD2 (36:1) (see SI) as the reference. Accordingly, a commercial GD2 sample was subjected to LC-

MS/MS analysis, and the resulting LC-MS/MS data, denoted by the base peak chromatogram (BPC), is depicted in Figure 6A. Following the traditional searching methods, where a known precursor mass of GD2 (36:1) (m/z 1675.93 $\pm$ 0.05) was extracted, we identified a distribution containing the desired mass (Figure 6B). After studying the product ions, a product spectrum at a retention time of 25.3 minutes was confirmed to match the GD2 (36:1) reference spectrum (Figure 7A).

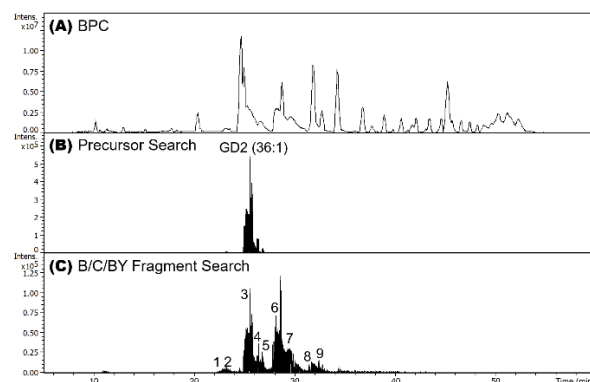


Figure 6: LC-MS/MS chromatographic results for a commercial GD2 sample extracted from biological sources. (A) Base peak chromatogram (BPC). (B) MS/MS extraction for precursor masses associated with GD2 (36:1). (C) MS/MS extraction for fragment masses associated with the B, C, and BY-type carbohydrate fragments of GD2. The counts in (C) indicate the general regions where GD2 lipid forms were detected.

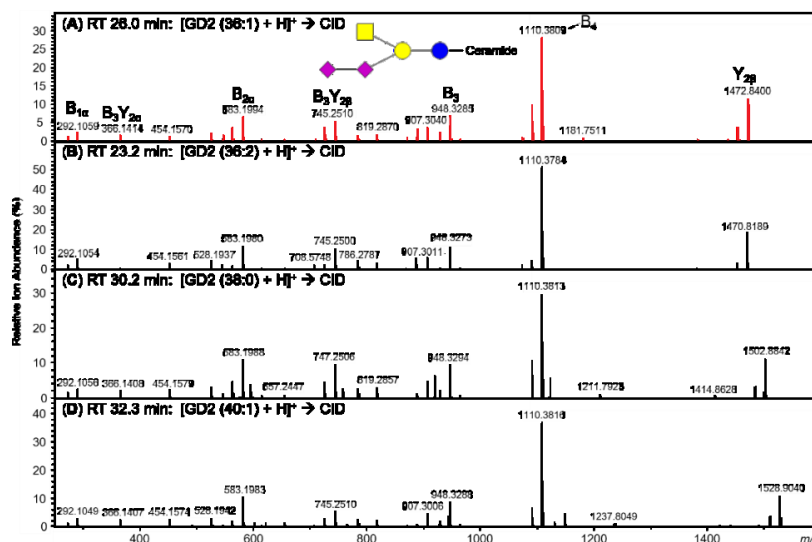


Figure 7: Product spectra of reference GD2 (36:1) (A, red) and other abundant GD2 lipid forms detected via two-stage searching of the LC-MS/MS data shown in Figure 6C (top to bottom): (B) GD2 (36:2), (C) (38:0), and (D) GD2 (40:1). Some of the more abundant GD2 product ions are labeled in (A). Product spectra for all 9 detected GD2 lipid forms are provided in SI (Figure S2).

Afterwards, a second search was conducted, in which the two-stage method was employed. For this, the first searching stage was implemented by extracting all product spectra for the B, C, and BY carbohydrate fragments associated with GD2. This generated four general distributions between 25 and 32 minutes (Figure 6C), as opposed to the one distribution detected via precursor searching. These spectra were then filtered based on not just adherence to the carbohydrate spectral fingerprint but also match with anticipated glycolipid fragments (second

matching stage). This resulted in the detection of 8 additional GD2 lipid forms, a significant increase in the number of GD2 annotation. The more abundant lipid forms identified by the two-stage matching method are presented in Figures 7B-7D. Compared to the reference GD2 (36:1), these lipid forms were carbon light, carbon heavy, reduced and oxidized, such as GD2 (36:2), (38:0), and (40:1), all of which were observed in brain tissues.⁵⁷ The product spectra for all discovered GD2 lipid

forms and their fragment ions and assignments, are provided in the SI (Figure S2 and Tables S5-S10).

Conclusions

Like other lipids, where headgroup cleavages dominate the CID product spectra, the GSL product spectra are dominated by cleavages in the glycan. Unlike other lipids, however, where the characteristic product ions are mainly the lipid fragments containing partial headgroup structures, GSLs produce two sets of equally intense and species-specific product ions. Cleavages of the glycan with the charge remaining on its non-reducing end side give product ions without the ceramide moiety, which we call "carbohydrate fragments". The masses and relative intensities of these ions are only determined by the glycan structure and conserved within each GSL species. Conversely, cleavages of the glycan with the charge remaining on its reducing end side generate product ions that retain the ceramide moiety, which we call "glycolipid fragments". These ions behave like that of other lipid classes, namely that the product ion masses change with the changes in ceramide, a behavior that has been described for lipids and is expectable for GSL. Furthermore, we have also observed that the fragmentation pattern of a GSL species is essentially unaffected by lipid forms, thus both the masses and the relative intensities for the carbohydrate fragments are independent of the lipid structure; however, the glycolipid fragments show increasing (or decreasing) masses as dictated by lipid forms, while the ion relative intensities remain unchanged.

Based on these observations, we proposed a new strategy to characterize GSLs by a two-stage matching process. In the first stage, experimental spectra are directly matched against the reference carbohydrate fragments to yield a GSL species identification. In the second stage, reference glycolipid fragments are treated by the lipid rule-based matching method to identify the ceramide composition (*i.e.*, the carbon and double bond numbers). As a result, the new method makes use of the complete GSL product spectra for the identification. It is our hope that this method may be used to disclose not only the glycan species but also the lipid forms of GSLs. We also expect that a searching software based on this algorithm combined with an extensive database will facilitate high throughput GSL analysis. In addition, we anticipate that its principle may be applied to other glycolipids to close the characterization gap currently hindering glycolipidomics analysis.

Although this two-stage matching method has only been demonstrated in controlled experiments (*i.e.*, using "pure" commercial samples), investigations have also been performed on complex biological extracts (data not shown). These investigations have resulted in more GSL identifications than targeted search alone, leading us to believe that this method may be beneficial in finding previously unidentified GSLs. A more thorough discussion of these results as well as the application of our two-stage matching method to complex biological GSL samples is currently in preparation.

Finally, although negative mode ESI is more broadly used to analyze acidic GSLs (*i.e.*, ganglioside), positive mode is also common⁵⁸⁻⁵⁹ and even better for neutral GSLs.⁶⁰ We selected the latter as the reported GSL dissociation patterns in negative and positive mode are similar⁶¹ and the two-stage matching method should be equally applicable to both results. Moreover, we aim

to build a database containing negative and positive mode spectra for both acidic and neutral GSLs but an in-depth comparison of positive and negative mode spectra is outside the scope of the current study. In addition, this study and above discussions have been focused on low energy dissociation using CID, one of the prominent methods for lipidomics analysis currently, but provides limited details about lipid structures. As the two-stage matching method is based on dissociation of the glycans, it should be easily applicable to other activation schemes. To determine the lipid structure, however, including locations of double bonds and any other functional groups, higher energy methods, such as electron-capture dissociation (ECD) and ultraviolet photodissociation (UVPD), should be used.⁶²⁻⁶⁶ Alternatively, MS³, either using CID for each activation step or by combining different activation schemes, has also been proved useful in further elucidation of GSL structure.⁶⁷ In brief, the combination of various MS techniques with the new searching methodology will make full characterization of the glycan and lipid composition of GSLs easily achievable.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information as a PDF file containing the methods to generate theoretical fragments and annotate product ions, tables of all matched product ion masses and their fragment assignments for various lipid forms of GM3, GM2, and GD2, and product spectra of different GD2 lipid forms is available free of charge on the ACS Publications website.

AUTHOR INFORMATION

Corresponding Author

*Zhongwu Guo, E-mail address: zguo@chem.ufl.edu.

Author Contributions

The manuscript was written through contributions of all authors, and all authors have approved the final version of the manuscript.

ACKNOWLEDGMENT

This work is partly supported by an NSF grant (CHE-1800279) to ZG. The MS instrument was funded by the NIH (S10 OD021758).

REFERENCES

1. Stults, C. L. M.; Sweeley, C. C.; Macher, B. A. *Methods Enzym.* **1989**, *179*, 167-214.
2. Merrill Jr., A. H.. *Chem. Rev.* **2011**, *111*, 6387-6422.
3. Yu, R. K.; Tsai, Y.; Ariga, T.; Yanagisawa, M. *J Oleo Sci.* **2011**, *60*, 537-544.
4. Todeschini, A. R.; Hakomori, S. *Biochim. Biophys. Acta* **2008**, *1780*, 421-433.
5. Wennekes, T.; van den Berg, R. J. B. H. N.; Boot, R. G.; van der Marel, G. A.; Overkleef, H. S.; Aerts, J. M. F. G. *Angew. Chem. Int. Ed.* **2009**, *48*, 8848-8869.
6. D'Angelo, G.; Capasso, S.; Sticco, L.; Russo, D. *FEBS J.* **2013**, *280*, 6338-6353.
7. Schnaar, R. L.; Suzuki, A.; Stanley, P. In *Essentials of Glycobiology*, 2nd ed.; Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Eds. Cold Spring Harbor Laboratory Press: Cold Spring Harbor 2008; pp 129-142.

8. Russo, D.; Parashuraman, S.; D'Angelo, G. *Int. J. Mol. Sci.* **2016**, *17*, 1732-1-23.
9. Garcia-Ruiz, C.; Morales, A.; Fernandez-Checa, J. C. *Apoptosis* **2015**, *20*, 607-620.
10. Lopez, P. H. H.; Schnaar, R. L. *Curr. Opin. Struct. Biol.* **2009**, *19*, 549-557.
11. Schnaar, R. L. *J. Mol. Biol.* **2016**, *428*, 3325-3336.
12. Schengrund, C.-L. *Trend. Biochem. Sci.* **2015**, *40*, 397-406.
13. Palmano, K.; Rowan, A.; Guillermo, R.; Guan, J.; McJarrow, P. *Nutrients* **2015**, *7*, 3891-3913.
14. Groux-Degroote, S.; Guerardel, Y.; Delannoy, P. *ChemBioChem* **2017**, *18*, 1146-1154.
15. Qamsari, E. S.; Nourazarian, A.; Bagheri, S.; Motallebnezhad, M. *Asi. Pac. J. Cancer Prev.* **2016**, *17*, 1643-1647.
16. van Echten-Deckert, G.; Walter, J. *Prog. Lipid Res.* **2012**, *51*, 378-393.
17. Ariga, T.; Wakade, C.; Yu, R. K. *Int. J. Alzh. Dis.* **2011**, *ID 193618*, 1-14.
18. Li, T.; Ohashi, Y.; Nagai, Y. *Carbohydr. Res.* **1995**, *273*, 27-40.
19. Han, X.; Cheng, H. *J. Lipid Res.* **2005**, *46*, 163-175.
20. Sarbu, M.; Dehelean, L.; Munteanu, C. V.; Vukelić, Ž.; Zamfir, A. D. *Anal. Biochem.* **2017**, *521*, 40-54.
21. Shaner, R. L.; Allegood, J. C.; Park, H.; Wang, E.; Kelly, S.; Haynes, C. A.; Sullards, M. C.; Merrill, A. H. *J. Lipid Res.* **2009**, *50*, 1692-1707.
22. Li, M.; Tong, X.; Lv, P.; Feng, B.; Yang, L.; Wu, Z.; Cui, X.; Bai, Y.; Huang, Y.; Liu, H. *J. Chromatogr. A* **2014**, *1372*, 110-119.
23. Boutin, M.; Menkovic, I.; Martineau, T.; Vaillancourt-Lavigne, V.; Toupin, A.; Auray-Blais, C. *Anal. Chem.* **2017**, *89*, 13382-13390.
24. Lee, H.; Lerno Jr, L. A.; Choe, Y.; Chu, C. S.; Gillies, L. A.; Grimm, R.; Lebrilla, C. B.; German, J. B. *Anal. Chem.* **2012**, *84*, 5905-5912.
25. Ikeda, K.; Shimizu, T.; Taguchi, R. *J. Lipid Res.* **2008**, *49*, 2678-2689.
26. Sørensen, L. K. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 3625-3633.
27. Li, Q.; Sun, M.; Yu, M.; Fu, Q.; Jiang, H.; Yu, G.; Li, G. *Glycoconj. J.* **2019**, *36*, 419-428.
28. Polo, G.; Burlina, A. P.; Kolamunnage, T. B.; Zampieri, M.; Dionisi-Vici, C.; Strisciuglio, P.; Zaninotto, M.; Plebani, M.; Burlina, A. B. *Clin. Chem. Lab. Med.* **2017**, *55*, 403-414.
29. Kirsch, S.; Zarei, M.; Cindric, M.; Müthing, J.; Bindila, L.; Peter-Katalinić, J. *Anal. Chem.* **2008**, *80*, 4711-4722.
30. Lavoie, P.; Boutin, M.; Auray-Blais, C. *Anal. Chem.* **2013**, *85*, 1743-1752.
31. Hájek, R.; Jirásko, R.; Lisa, M.; Cífková, E.; Holčápek, M. *Anal. Chem.* **2017**, *89*, 12425-12432.
32. Ikeda, K.; Taguchi, R. *Rapid Commun. Mass Spectrom.* **2010**, *24*, 2957-2965.
33. Almeida, R.; Mosoarca, C.; Chirita, M.; Udrescu, V.; Dinca, N.; Vukelić, Z.; Allen, M.; Zamfir, A. D. *Anal. Biochem.* **2008**, *378*, 43-52.
34. Serb, A. F.; Sisu, E.; Vukelić, Z.; Zamfir, A. D. *J. Mass Spectrom.* **2012**, *47*, 1561-1570.
35. Fabris, M.; Rozman, M.; Sajko, T.; Vukelić, Z. *Biochimie* **2017**, *137*, 56-68.
36. Sarbu, M.; Clemmer, D. E.; Zamfir, A. D. *Biochimie* **2020**, *177*, 226-237.
37. Sarbu, M.; Vukelić, Z.; Clemmer, D. E.; Zamfir, A. D. *Analyst* **2018**, *143*, 5234-5246.
38. Farwanah, H.; Kolter, T. *Metabolites* **2012**, *2*, 134-164.
39. Sarbu, M.; Zamfir, A. D. *Electrophoresis* **2018**, *39*, 1155-1170.
40. Tissot, B.; North, S. J.; Ceroni, A.; Pang, P.-C.; Panico, M.; Rosati, F.; Capone, A.; Haslam, S. M.; Dell, A.; Morris, H. R. *FEBS Lett.* **2009**, *583*, 1728-1735.
41. Li, Q.; Jaiswal, M.; Rohokale, R. S.; Guo, Z. *Org. Lett.* **2020**, *22*, published online, DOI: 10.1021/acs.orglett.0c02847.
42. Ceroni, A.; Maass, K.; Geyer, H.; Geyer, R.; Dell, A.; Haslam, S. M. *J. Proteome Res.* **2008**, *7*, 1650-1659.
43. Cajka, T.; Fiehn, O. *Metabolomics* **2016**, *12*, 34.
44. Domon, B.; Costello, C. E. *Glycoconj. J.* **1988**, *5*, 397-409.
45. Costello, C. E.; Vath, J. E. *Methods Enzymol.* **1990**, *193*, 738-768.
46. Khalil, M. B.; Hou, W.; Zhou, H.; Elisma, F.; Swayne, L. A.; Blanchard, A. P.; Yao, Z.; Bennett, S. A. L.; Figeys, D. *Mass Spectrom. Rev.* **2010**, *29*, 877-929.
47. Lapadula, A. J.; Hatcher, P. J.; Hanneman, A. J.; Ashline, D. J.; Zhang, H.; Reinhold, V. N. *Anal. Chem.* **2005**, *77*, 6271-6279.
48. Maxwell, E.; Tan, Y.; Tan, Y.; Hu, H.; Benson, G.; Aizikov, K.; Conley, S.; Staples, G. O.; Slys, G. W.; Smith, R. D.; Zaia, J. *PLOS ONE* **2012**, *7*, e45474.
49. Yu, C.-Y.; Mayampurath, A.; Hu, Y.; Zhou, S.; Mechref, Y.; Tang, H. *Bioinformatics* **2013**, *29*, 1706-1707.
50. Goldberg, D.; Sutton-Smith, M.; Paulson, J.; Dell, A. *Proteomics* **2005**, *5*, 865-875.
51. Kind, T.; Liu, K.-H.; Lee, D. Y.; DeFelice, B.; Meissen, J. K.; Fiehn, O. *Nat. Methods* **2013**, *10*, 755-758.
52. Koelmel, J. P.; Kroeger, N. M.; Ulmer, C. Z.; Bowden, J. A.; Patterson, R. E.; Cochran, J. A.; Beecher, C. W. W.; Garrett, T. J.; Yost, R. A. *BMC Bioinform.* **2017**, *18*, 331, 1-11.
53. Kind, T.; Okazaki, Y.; Saito, K.; Fiehn, O. *Anal. Chem.* **2014**, *86*, 11024-11027.
54. Ma, Y.; Kind, T.; Vaniya, A.; Gennity, I.; Fahrman, J. F.; Fiehn, O. *J. Cheminformatics* **2015**, *7*, 53, 1-5.
55. Yetukuri, L.; Ekroos, K.; Vidal-Puig, A.; Orešič, M. *Mol. Biosyst.* **2008**, *4*, 121-127.
56. Tang, H.; Mayampurath, A.; Yu, C.-Y.; Mechref, Y. *Curr. Protoc. Protein Sci.* **2014**, *76*, 2.15.1-2.15.7.
57. Zamfir, A. D.; Serb, A.; Vukelić, Z.; Flangea, C.; Schiopu, D.; Fabris, D.; Kalanj-Bognar, S.; Capitan, F.; Sisu, E. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 2145-2159.
58. Saville, J. T.; Smith, N. J. C.; Fletcher, J. M.; Fuller, M. *Anal. Chim. Acta* **2017**, *955*, 79-85.
59. Si, W.; Liu, Y.; Xiao, Y.; Guo, Z.; Jin, G.; Yan, Y.; Shen, A.; Zhou, H.; Yang, F.; Liang, X. *Talanta* **2020**, *208*, 120366.
60. Shioiri, Y.; Kurimoto, T.; Ako, A.; Daikoku, S.; Ohtake, A.; Ishida, H.; Kiso, M.; Suzuki, K.; Kanie, O. *Anal. Chem.* **2009**, *81*, 139-145.
61. Lee, H.; Lerno, L. A.; Choe, Y.; Chu, C. S.; Gillies, L. A.; Grimm, R.; Lebrilla, C. B.; German, J. B. *Anal. Chem.* **2012**, *84*, 5905-5912.
62. Macias, L. A.; Feider, C. L.; Eberlin, L. S.; Brodbelt, J. S. *Anal. Chem.* **2019**, *91*, 12509-12516.
63. Jones, J. W.; Thompson, C. J.; Carter, C. L.; Kane, M. A. *J. Mass Spectrom.* **2015**, *50*, 1327-1339.
64. Zhao, C.; Xie, B.; Chan, S.-Y.; Costello, C. E.; O'Connor, P. B. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 138-150.
65. Schaller-Duke, R. M.; Bogala, M. R.; Cassady, C. J. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 1021-1035.
66. Ruhaak, L. R.; Xu, G.; Li, Q.; Goonatilleke, E.; Lebrilla, C. B. *Chem. Rev.* **2018**, *118*, 7886-7930.
67. Calvano, C. D.; Ventura, G.; Sardanelli, A. M.; Losito, I.; Palmisano, F.; Cataldi, T. R. I., *Anal. Biochem.* **2019**, *581*, 113348.

TOC graphic:

