RYAN, SMITH A. M.S. LC-MS Analysis of Gangliosides: methods development and software solutions. (2023)
Directed by Dr. Qibin Zhang. 118 pp.

Gangliosides are high molecular weight, specialized metabolites found at greatest concentration in the central nervous system. They are glycosphingolipids, possessing a nonpolar ceramide tail and a polar carbohydrate or 'glycan' head. Ganglioside amphiphilicity is essential to their function as membrane lipids and to the formation of lipid rafts on cell surfaces. This amphiphilicity also complicates their separation and analysis. Gangliosides possess several forms of isomerism with isomers coeluting in reverse phase liquid chromatography. Despite these difficulties and biological scarcity, ganglioside analysis is a growing field for their suspected role in various cancers, neurodegenerative disorders, and Type 1 Diabetes (T1D).

Although gangliosides can be analyzed directly using tandem mass spectrometry for their ceramides in the positive ion mode and for their glycans in the negative ion mode, chemical derivatization enhances structural identification through improved fragmentation while also enabling more sophisticated analytical techniques, such as methods based on isotopic and isobaric labeling. With these methods in their infancy for comparative lipidomics, there is little compatible analysis software. We have developed two stable-isotope ganglioside analysis methods and three companion Python programs, one method based on amide formation and isotopic labeling, the second based on oxime formation and isobaric labeling. From a normal mouse brain, the cerebellum, pons/medulla, midbrain, and cortex were dissected, homogenized, and their gangliosides extracted. Both derivatization methodologies were applied to extracts representing 2 mg of tissue, cross-examining the brain regions with 79 gangliosides analyzed across 10 glycan classes. This work greatly expands ganglioside analysis capacity in biological samples and can be applied to study of lipidome remodeling in T1D mice models.

LC-MS ANALYSIS OF GANGLIOSIDES: METHODS DEVELOPMENT AND SOFTWARE SOLUTIONS

by

Ryan A. Smith

A Thesis
Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Greensboro

2023

Approved by

Dr. Qibin Zhang
Committee Chair

DEDICATION

I dedicate this work first and foremost to my loving wife, Jennifer. She has been my steadfast companion and support in North Carolina, far from our Florida home and family network. Already an accomplished writer and business owner, Jennifer has undoubtedly become one of the foremost experts on gangliosides, on the ups and downs of their analysis if not on their chemical nature.

APPROVAL PAGE

This thesis written by Ryan A. Smith has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair

Dr. Qibin Zhang

Committee Members

Dr. Ethan Taylor

Dr. Nadja Cech

April 12, 2023

Date of Acceptance by Committee

April 12, 2023

Date of Final Oral Examination

ACKNOWLEDGEMENTS

This work was conducted at The Center for Translational Biomedical Research at the North Carolina Research Campus in Kannapolis, NC with supervision and mentoring by Dr. Qibin Zhang. Dr. Zhang provided both academic council and professional development throughout the graduate program. He was compassionate toward and accommodating of my needs as a recent car accident survivor, allowed me great latitude in setting and pursuing research goals, and provided essential feedback to guide and assist my growth as an analytical chemist. The high-performance analysis resources of the Center, and my nearly unlimited access to them, enabled the production of a satisfying body of work and the initiation of what promises to be a worthwhile career. For all this, I am very grateful, and I extend my whole-hearted thanks to Dr. Zhang for these opportunities. Invaluable mentoring was also provided by other members of the Center, most notably Dr. Rohit Kamble and Dr. Ashraf Alsagheer, both good men and good researchers, to whom I also extend my whole-hearted thanks.

I would also thank the faculty of the Department of Chemistry and Biochemistry at the University of North Carolina Greensboro, who provided excellent instruction and broadened my expertise as a chemist. Instruction by Dr. Ethan Taylor and Dr. Nadja Cech was of particular value for relevance to this thesis, as was their feedback in its early stages.

Research reported in this publication was supported by the National Institute Of Diabetes And Digestive And Kidney Diseases of the National Institutes of Health under Award Number R01DK123499.

Chapter III has been published in *Bioinformatics Advances* 2023, https://doi.org/10.1093/bioadv/vbad044. What is presented in this thesis is modified from the published version.

TABLE OF CONTENTS

LIST OF TABLES

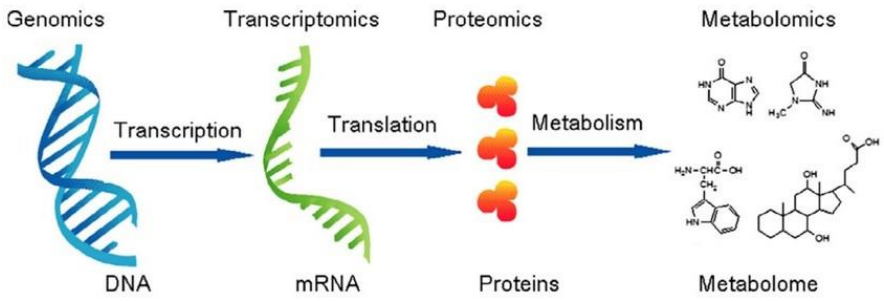
LIST OF FIGURES

CHAPTER I: INTRODUCTION

Metabolomics and Application-Driven Method Development

The 13 year progression and ultimate completion of the Human Genome Project in 2003[1] produced great interest and progress in systems biology as a means to research, detect, and eventually treat human disease.[2-5] Out of systems biology come the 'omics' fields: genomics, transcriptomics, proteomics, and metabolomics. These study DNA, RNA, proteins, and metabolites, respectively, for the identification and tracking of biomarkers.[2] Figure 1 illustrates the ontological progression of the omics fields in relation to their underlying biology.

Figure 1. The Omics Fields of Systems Biology. Ref.[2]

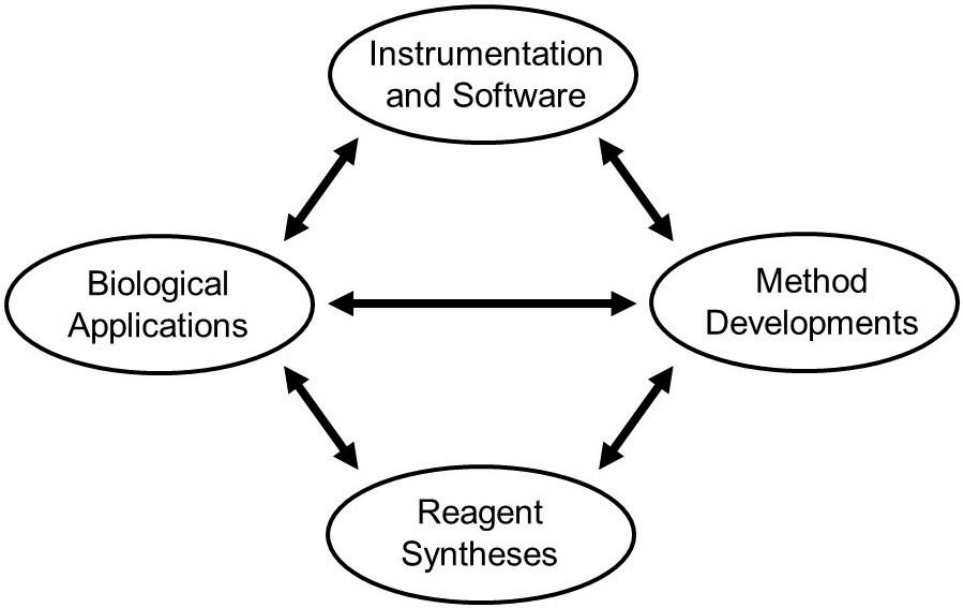


The genome and its transcription contain information for what genotypic risk factors a person has for disease, but even the transcriptome does not necessarily reflect the progression of disease or the impact of exogenous factors. The proteome and to a greater extent the metabolome are the downstream phenotype of these genetic elements expressed with the influence of environmental, age-related, and disease factors.[6, 7] The phenotypic approach to systems biology via metabolomics has long been seen as a plausible end goal for detecting and monitoring cancers[3], cardiovascular diseases[4], and neurodegenerative disorders such as Alzheimer's disease[5]. As the metabolome reflects cellular change in response to stimuli, metabolomics also presents the means for discovery of useful compounds and the elucidation of

cellular mechanisms.[8] Such rationale has been involved in numerous natural product discoveries and refinement for human application.[9] With adequate domain knowledge of the metabolome and sufficiently sophisticated analytical methods, detecting and tracking human disease could become just a question of throughput and affordability.

This thesis work is in the domain of lipidomics. Lipids represent the most hydrophobic subset of the metabolome. With thousands of species in a single cell, lipids perform a variety of biological functions including membrane structure, energy storage, and signal transduction.[10] As with the greater metabolome, variations in the lipidome can indicate dysfunctions of intercellular signaling and of intracellular metabolism, leading to the creation and rapid expansion of the lipidomics field over the last decade.[11] This text describes several method development projects and software innovations for the analysis of gangliosides, a specific class of lipid. Method development such as this may not have a traditional scientific hypothesis but instead expands the capabilities of related inquiry. Figure 2 illustrates the place of method development in the greater scheme of omics science.

Figure 2. Research Feedback Loops in Omics Science.



Method development is not undertaken for its own sake but as part of a research loop. Biological discoveries and desired applications drive method development through the need for more sophisticated or specialized techniques. These two drive the development, manufacture, and sale of reagents. Likewise, the available reagents determine what inquiry is possible and what types of methods may be developed. Instrumentation innovation in hyphenated -MS analyses, liquid chromatography (LC) or gas chromatography (GC) paired with mass spectrometry (MS), have spurred the metabolomics and lipidomics movements, which in turn drive the continued development of high performance analytical supplies and devices.[12]

When faced with large datasets, manual processing and annotation of LC-MS spectra can become prohibitively time-consuming, leading to the development of software solutions for rapid and reproducible processing.[11] The method development herein has the objective of satisfying the greater requirements of planned future biological applications and creatively utilizes preexisting isotopic labeling reagents. Out of necessity, we developed software to fill gaps in the available bioinformatics ecosystem. Future work will apply these methods and software tools to biological study and pursue further tailored reagents through inhouse synthesis.

Gangliosides

Structure and Nomenclature

Gangliosides are glycosphingolipids composed of a hydrophobic ceramide moiety connected by ether bond to a carbohydrate or ‘glycan’ moiety. The inclusion of one or more N-Acetylneuraminic acids (shorthand Neu5Ac or sialic acid) differentiates gangliosides from other glycosphingolipids. Figure 3 illustrates the progressive structure and nomenclature of gangliosides following popular convention initially outlined by Svennerholm in 1994.[13]

Figure 3. Ganglioside Structure and Nomenclature.

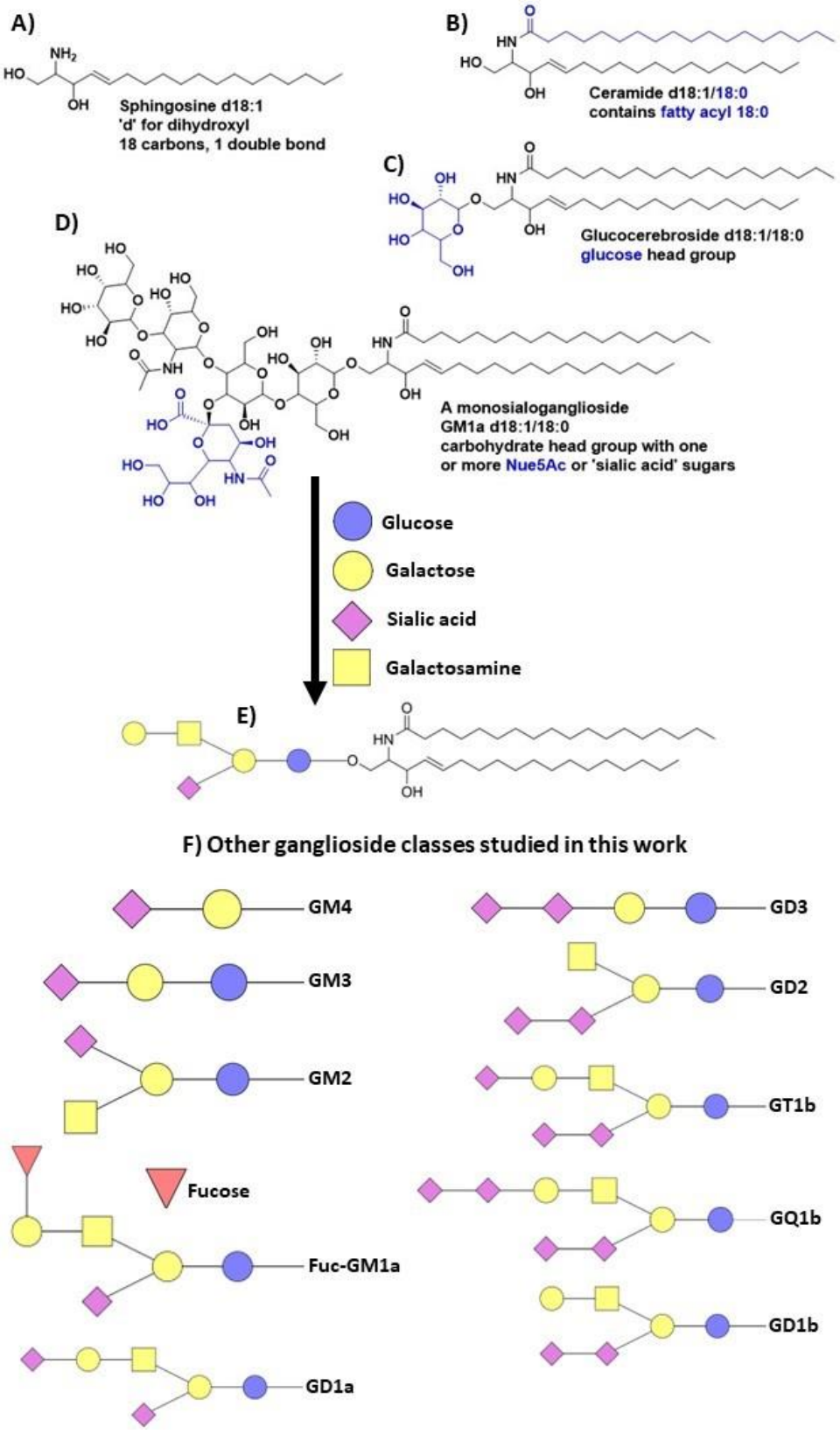


Figure 3A is the sphingosine moiety, often referred to as the long-chain base (LCB) in gangliosides and other ceramide species. The ‘d’ in d18:1 indicates the number of hydroxyls (di-), the ‘18’ indicates that there are 18 carbons, and the ‘1’ indicates that there is 1 double bond. While few methods[14] can precisely determine lipid double bond locations in LC-MS analysis, the double bond between carbons 4 and 5 of the sphingosine backbone is in part responsible for the unique LC-MS/MS fragmentation behavior of ceramides[15]. With this same fragmentation behavior reported for gangliosides in the literature[16-18], this can at least be assumed to be the overwhelmingly dominant isomer. The locations of additional ceramide double bonds and the changes in fragmentation caused by isomerism will be further discussed in Chapter II.

A ceramide (Figure 3B) is the amide joining of a fatty acid and a sphingosine. The fatty acyl moiety follows the same naming conventions as the LCB, its name separated from the LCB’s with a ‘/’ if both tail contents are known. If the total ceramide content was known but not the allocation between the tails, the name would instead be d36:1. A common phenomena in LC-MS analysis of ganglioside is the coelution of species with ceramide isomerism, such as two GM1a d38:1 species, one with a d18:1/20:0 ceramide and the other with a d20:1/18:0 ceramide.[19] With the addition of a glucose residue to the ceramide, Figure 3C is the cerebroside progenitor of most glycosphingolipids, GM4 (Figure 3F) being a notable exception.

GM1a (Figure 3D) is one of the most abundant human gangliosides[20], possessing four neutral sugars and one sialic acid, the sugar negatively charged at physiological pH. The glycan naming convention is a mix of historical precedent and simplification of otherwise burdensome systematic names.[13] The ‘G’ designates ganglioside. In the case of GM1a, it is ‘M’ for monosialoganglioside. There are also di-, tri- , and tetrasialogangliosides with GD1a, GD1b, GT1b, and GQ1b highly expressed in the human central nervous system.[20] The ‘1’

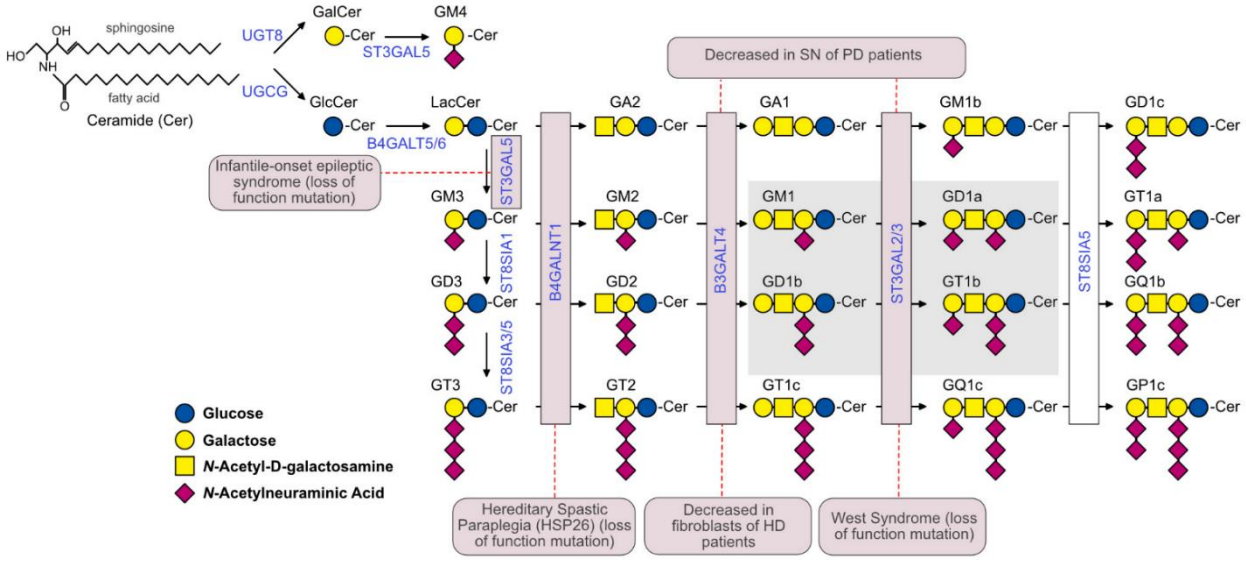
paradoxically accounts for the number of neutral sugars. The original source of the 1-4 ganglioside naming series is their elution order in RPLC. Gangliosides with larger glycans have more hydrophilic interactions and so elute faster. This was later adapted as 5 minus the number of neutral sugars. Hence, a ganglioside with four neutral sugars belongs to the '1' series, and a ganglioside with one neutral sugar belongs to the '4' series. The neutral sugars counted for such calculus are glucose, galactose, and galactosamine, but not fucose, a common glycan modification, making fucosylated GM1a's designation Fuc-GM1a instead of GM0a.[21] The final, lowercase letter indicates the sialic acid positional isomer if relevant. In ganglioside glycans, sialic acids are either linked to galactose resides or other sialic acids.[13] In the '1' ganglioside series with two galactose, there is potential for structural isomerism through sialic acid placement, illustrated by GD1a and GD1b in Figure 3F. There is also linkage isomerism of the glycosidic bonds, such as Neu5Ac-2,6-Neu5Ac vs Neu5Ac-2,3-Neu5Ac.[22] This distinction is not covered in the Figure 3 nomenclature. Diagnosing such linkage isomerism is a difficult and specialized analytical challenge.[22-26] Along with GM1a, the ten ganglioside glycans in Figure 3F represent the eleven ganglioside classes studied in this work.

Biological Abundance, Diversity, and Significance

Gangliosides are present in both plants[27] and animals[28, 29] but with greatest abundance in vertebrates[30] and mammals[10], sometimes apparently absent[31] in plant tissue. Within vertebrates and humans, the concentration of gangliosides in the central nervous system may be 10 to 30 times higher than other tissues.[20] The diversity and abundance of gangliosides in the central nervous system is increasingly associated with higher order cognition.[32] While the enzymatic hardware to produce more than 200 variations of the glycan head exists, the overwhelming majority of mammalian ganglioside mass is relegated to

comparatively few glycan subspecies.[33] Figure 4 illustrates the enzymatic pathways of human brain gangliosides along with several pathologies associated with specific enzymatic disfunctions.[20, 34] Boxed in grey, GM1a, GD1a, GD1b, and GT1b constitute the majority of human brain ganglioside mass with GQ1b, GM2, GM3, GD2, GD3, and GM4 making up almost the entirety of the remaining fraction for both humans and other animals. These 10 classes also represent what is commercially available for method development. With total synthesis of gangliosides being prohibitively difficult or inefficient, the standards that exist are either purified extracts from animals or the result of partial synthesis beginning from similar lipids.

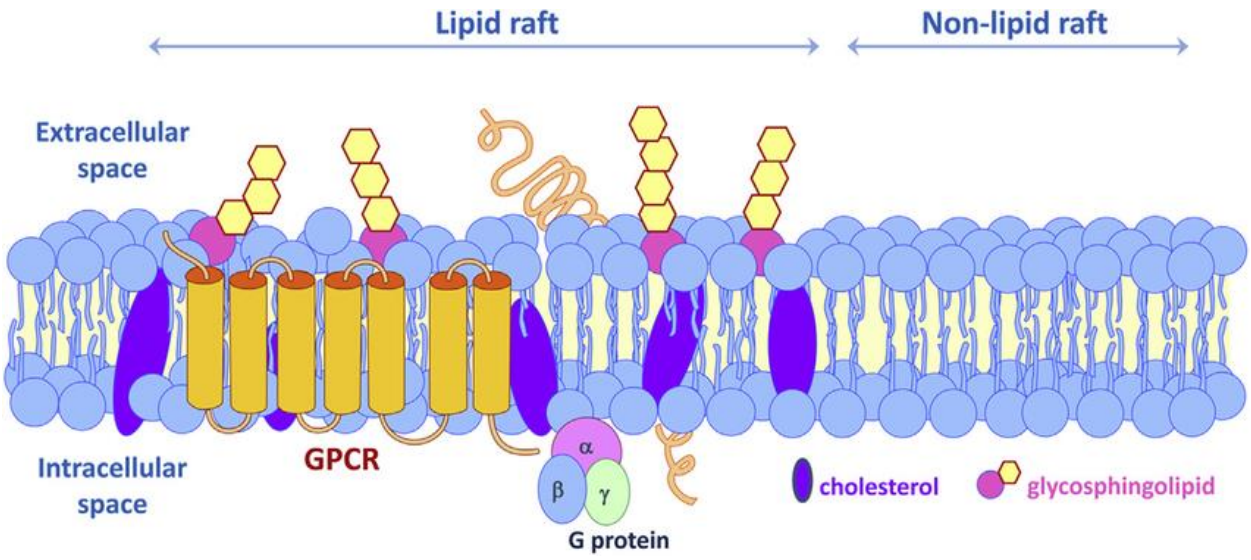
Figure 4. Gangliosides in the Human Brain. Ref.[20]



Given the complexity of these biomolecules and the biochemical capital that must be invested to evolve and produce them, it is clear *a priori* that they are of great cellular significance. Gangliosides are now known to be involved in cell-to-cell signaling[35], defining membrane geometry[36], regulating cancer apoptosis[37], binding of endocrine factors[38], recruitment of transmembrane proteins into lipid rafts[39], recruitment of ions and ligands[40], and RNA virus binding[41]. That gangliosides are predominately found in the cell membrane is

not surprising given their amphiphilic nature, similar to the phospholipids for which the phospholipid bilayer gets its name. In the cell membrane, gangliosides are an essential component of lipid rafts, loss of the rafts and their function observed when gangliosides are experimentally eliminated.[42] Figure 5 illustrates the basic structure of lipid rafts. The numerous hydrogen bond donating and accepting groups of the ganglioside glycan allow for a diversity of ligand interactions outside the membrane while the ceramide moiety leads to favorable hydrophobic interactions with cholesterols and other internal membrane components.[43] The result of these effects is mobile microdomains that recruit transmembrane proteins, cations, and various other factors while playing a major role in signaling.[20]

Figure 5. Lipid Rafts in Cellular Membranes. Ref.[44]



With these diverse and essential functions, it is also not surprising that ganglioside dysfunction has been associated with a variety of diseases. As suggested in Figure 4, a significant body of research concerns the role of gangliosides in neurodegenerative diseases. Ganglioside dysfunction has been associated with Huntington's disease[45], Alzheimer's disease[46], non-pathological age-related neurodegeneration[47], neuron inflammation and

regeneration[35], ischemia[48], and a host of other afflictions of the brain[20, 28, 49]. Often, these diseases progress alongside a measurable decrease[45, 48] in more complex ganglioside glycan structures, such as GM1a, toward less complex glycans, such as GM2 or GM3. This is believed to be the result of glycan-cleaving hydrolases[49] in some cases and the breakdown of Golgi apparatus synthesis[50] of gangliosides in others. This makes gangliosides excellent analytical biomarkers for tracking the progression of neurodegenerative disorders, not just by absolute quantity but through the spectrum of expressed species.

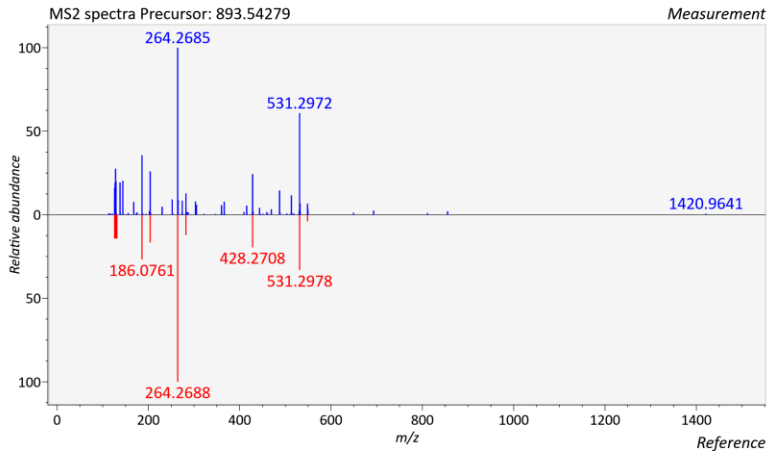
With gangliosides found throughout the human body, they are also known to be involved in diseases beyond the brain. The ganglioside composition of cells changes during cancer pathenogenesis, to signal apoptosis[37], in metastasis[51], and in detection by and evasion of the immune system.[52] GM1a deficiency has further been identified as mechanistically involved in the progression of T1D through autoimmune destruction of pancreatic β-cells and calcium ion channel misregulation.[40] GM1a is a suspected factor in mitigating diabetes-caused ischemia throughout the body in addition to the brain.[48] With many examples of ganglioside depletion associated with disease progression, there have been attempts to apply exogenous gangliosides therapeutically and not without success.[40, 48] Gangliosides have been shown to be involved in epigenetic regulation of neural stem cells, presenting the possibility of therapeutic treatment for congenital disorders.[53] G-coupled proteins, common targets for medicinal active ingredients, are modulated by membrane gangliosides, further building the case for therapeutic application.[54] To rigorously pursue these promising avenues and deeper biological study, more sophisticated analytical techniques are necessary. Current methods are low throughput and lack the more sophisticated capabilities used in analysis of other metabolites and large study cohorts. To increase throughput, most urgently, automated LC-MS processing is necessary.

CHAPTER II: NEEDS AND SOFTWARE DEVELOPED FOR HIGH THROUGHPUT PROCESSING OF GANGLIOSIDE LC-MS/MS DATA SETS

MS-DIAL

MS-DIAL[55] is a relatively new, powerful, and popular[11] metabolomics platform, enabling rapid processing and automated metabolite identification across large cohort sizes. MS-DIAL has a user-friendly interface that poses no intimidation to researchers unfamiliar with R or Python coding. It is free to use, open source, and has a dedicated team behind its ongoing development, now on its eighth year of updates. Like most omics analysis platforms, MS-DIAL performs the essential tasks of aligning LC-MS and GC-MS chromatograms, annotating peaks, discerning adduct species, and assigning isotopic relationships. A unique and powerful feature lies in the ability for users to provide their own annotation databases with tandem mass spectrometry (fragmentation) data, utilizing both fragment mass and relative intensity in identification. This feature made MS-DIAL an early adoption in our analytical method development, both because public ganglioside libraries are incomplete and because fragment mass and fragment relative intensity are essential for confident ganglioside identification.[16-18]

Figure 6. MS-DIAL Ganglioside MS2 Spectra Matching with a Custom Reference Library.



Need of Specialized Tools for Analyzing Ganglioside LC-MS Data in MS-DIAL

While ganglioside libraries exist on such platforms as LIPID MAPS and from other public sources, these libraries are incomplete. They lack critical LC-MS/MS data necessary to make confident annotations, often containing little more than a name, structure, and the molecular formula. Furthermore, numerous species reported in the literature, even at relatively high concentrations in humans, do not have entries in any public database. To enable high throughput automated annotation of gangliosides and derivatized gangliosides by MS-DIAL, it was necessary to make our own libraries. Beginning in Excel, early library prototypes illustrated two problems with this approach: (1) Excel generated libraries lean heavily on human input and so require significant labor to update them with new data and trends, and (2) to generate a library with wide ganglioside coverage beyond what has already been identified in the literature, *in silico* chemistry would be required. A portion of this thesis work was dedicated to the creation of a Python program accomplishing both the goals of ganglioside prediction and rapid library generation for immediate annotation use by MS-DIAL.

After ganglioside identification in MS-DIAL was solved, we found that additional post processing features were needed for MS-DIAL-based quantitation in stable-isotope labeling methodologies. MS-DIAL cannot perform isobaric analysis, and it lacks more sophisticated isotopic labeling features found on other platforms. While there are many other software options for isobaric analysis, we found it beneficial to write a simple script for isobaric post processing of MS-DIAL isobaric data. This was simpler than building a bridge between programs. For isotopic labeling, we found a gap in the greater bioinformatics software ecosystem with far fewer software options. We developed a program and workflow for isotopic labeling quantification through MS-DIAL for any metabolomics study.

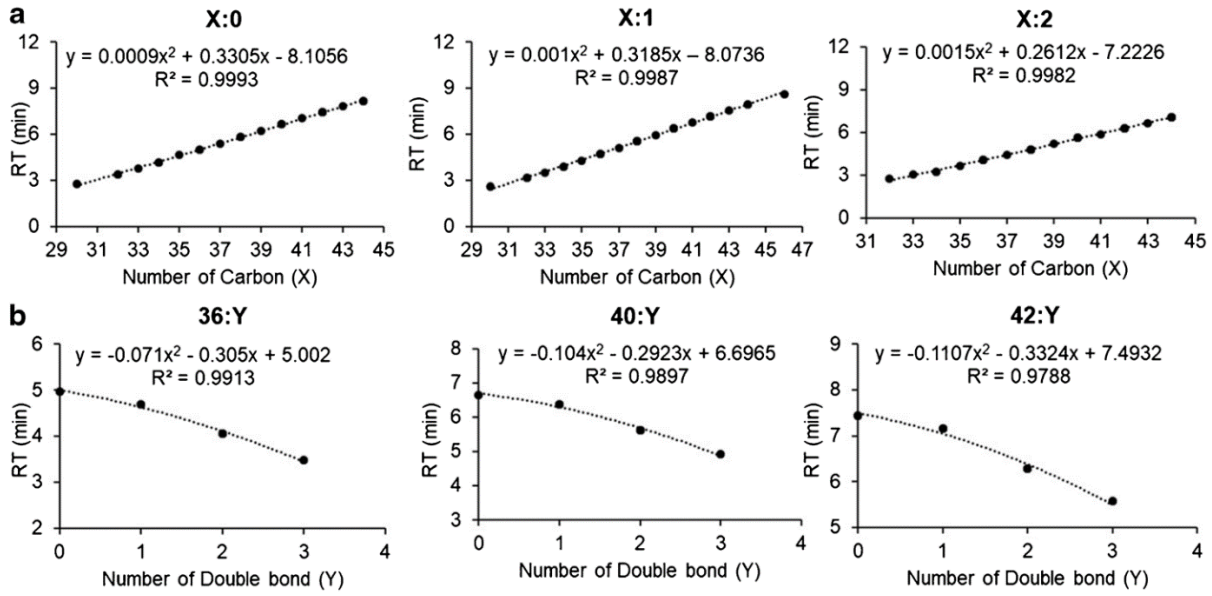
Ganglioside Library Generator

Introduction

Ganglioside Library Generator (GLG) was purpose-built to generate MS-DIAL[55] compatible libraries of derivatized gangliosides for confident MS-DIAL-assisted identification within biological matrixes. At 5812 lines of code, GLG is capable of producing library entries for 1,975,680 unique gangliosides for both derivatization schemes discussed in Chapters IV and V. While the majority of those 2 million gangliosides are theoretical, and some are improbable, the ability to search for uncharacterized gangliosides is an essential advantage of this design.

For LC-MS/MS analysis of gangliosides, an annotation library must contain accurate retention time (RT), precursor mass (MS1), characteristic fragment ion masses (MS2), and the relative intensities of the fragment ions, as different ganglioside species may share fragments at different relative intensities. While programming molecular formulae and the resulting MS1 masses is a straightforward task, predicting RT and MS2 is more involved. Previous work[17] has shown that for a specific glycan class, RPLC RT is determined by total ceramide content. Figure 7 illustrates the changes in RT within the GM3 class based on (a) the number of ceramide carbons and (b) the number of ceramide double bonds. The ceramide interacts with the stationary phase of the C18 column, and so it follows that longer ceramide chains will lead to longer RT and that ‘kinks’ in the ceramide by way of double bonds will lead to shorter RT. The double bonds interfere with favorable hydrophobic ceramide-C18 interactions, much like unsaturated fatty acids of the plasma membrane.

Figure 7. RT Equations for GM3 Gangliosides Based on Ceramide Content. Ref.[17]



While there are programs for fragmentation prediction of small molecules[56-58] and of larger polymers[59, 60], gangliosides are in a middle ground where neither school of fragmentation prediction programs are applicable. Fortunately, there is a growing collection of work[16-19, 51, 61] reporting tandem mass spectrometry data for gangliosides as well as some useful tools[62] for glycan fragmentation if not the entire ganglioside. Taking lessons and observations from these resources and conducting extensive manual data analysis with commercial ganglioside standards has produced a body of data that forms the backbone of GLG's predictive fragmentation formulae. This work is concomitant with Chapters IV and V.

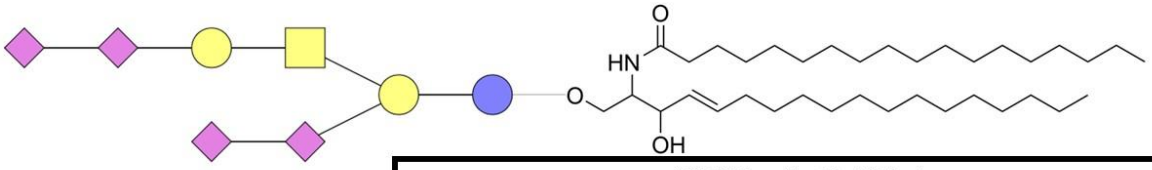
Methods and Materials

The Ganglioside Object

GLG is an example of object-oriented programming with the 'Ganglioside' as the core object. Figure 8 illustrates the input and initialization of a GQ1b d18:1/18:0 ganglioside object in GLG. Comparatively few inputs spawn a comprehensive set of variables defining this molecular species within the GLG architecture. These inputs can be provided manually through

an *à la carte* external program or through a user-friendly GUI for rapid library generation iterating over all user selections, typically in less than 1 minute. Notable here is the specification of Neu5Ac sialic acids. We have found detectable amounts of Neu5Gc sialic acids (one extra oxygen) in standards of porcine origin and so added this feature. Selecting Neu5Gc instead of Neu5Ac changes otherwise static variables to account for this subtle but important difference. There is also the option of specifying a molecule with one or more natural carbon-13, which will update the precursor mass accordingly. For larger gangliosides such as GQ1b d36:1, the [M+1] isotope will be more abundant than the [M+0] isotope, making this distinction useful in such cases where the [M+1] isotope is selected for fragmentation and the [M+0] is not. This collection of variables is used in the subsequent derivatization and fragmentation formulae to account for the diverse reactivities and outcomes of different ganglioside classes and species.

Figure 8. The GLG Ganglioside Object.



GQ1b d18:1/18:0
with Neu5Ac sialic acids

Inputs for Ganglioside generation:
- **glycan = GQ1b**
- **LCB = d18:1**
- **FA = 18:0**
- **num_FA_hydroxyls = 0**
- **num_fucose = 0**
- **num_nat_c13 = 0**
- **sialic_acid_type = Neu5Ac**

Inputs submitted to GLG either by external terminal for *à la carte* generation or by user-friendly interface for library generation

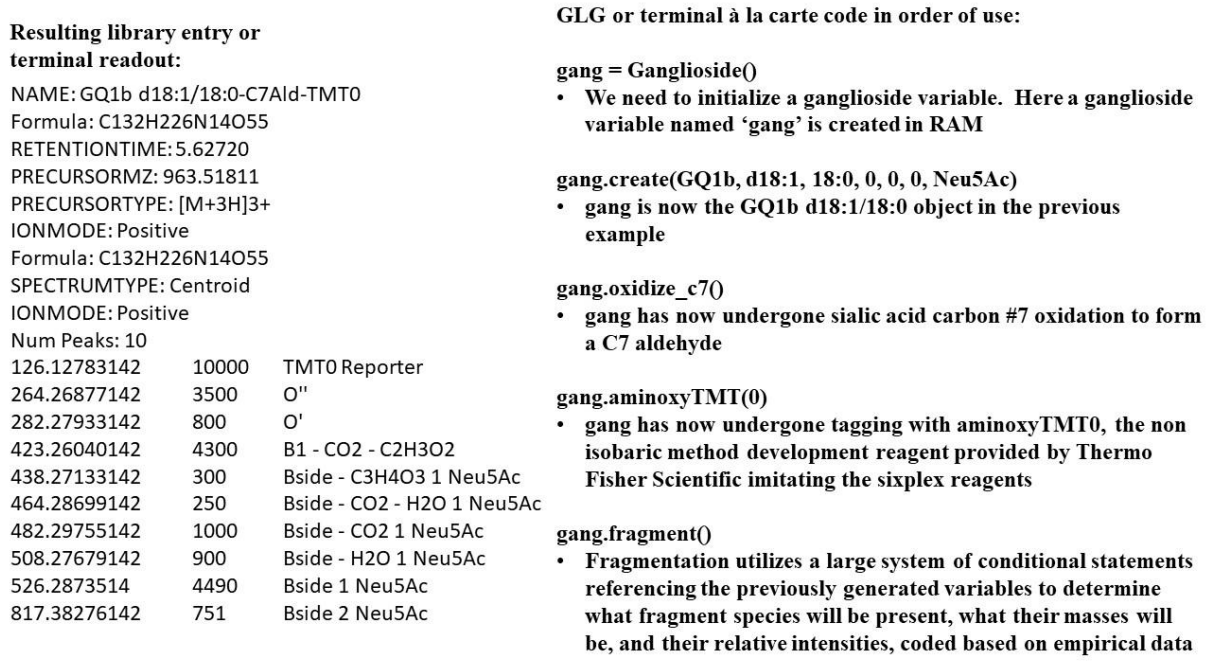
GLG Ganglioside Object

- **sialic_acid_type = Neu5Ac**
- **num_fucose = 0**
- **num_nat_c13 = 0**
- **num_FA_hydroxyls = 0**
- **glycan = GQ1b**
- **glycan_series = 1**
- **glycan_SA_major_character = Q**
- **glycan_SA_minor_character = b**
- **num_sialic_acids = 4**
- **num_terminal_SAs = 2**
- **num_pos1_SAs = 2**
- **num_pos2_SAs = 2**
- **long_chain_base = d18:1**
- **fatty_acid = 18:0**
- **num_LCB_carbons = 18**
- **num_LCB_double_bonds = 1**
- **num_LCB_hydrogens = 36**
- **num_FA_carbons = 18**
- **num_FA_double_bonds = 0**
- **num_ceramide_carbons = 36**
- **num_ceramide_hydrogens = 69**
- **num_ceramide_nitrogens = 1**
- **num_ceramide_oxygens = 3**
- **num_glucose = 1**
- **num_galactose = 2**
- **num_galactosamine = 1**
- **num_alcohols = 25**
- **num_sugar_residues = 8**
- **num_glycan_carbons = 70**
- **num_glycan_hydrogens = 111**
- **num_glycan_nitrogens = 5**
- **num_glycan_oxygens = 52**
- **num_carbons = 106**
- **num_hydrogens = 182**
- **num_nitrogens = 6**
- **num_oxygens = 55**
- **molecular_formula = C106H182N6O55**
- **precursor_plus_one_mz = 2420.17044**
- **name = GQ1b d36:1 0C13 (primary isomer d18:1/18:0)Neu5Ac**

Ganglioside Object Derivatization and Fragmentation

With the ganglioside as a defined object, it can be modified with previously coded functions using single input commands. While defining and testing those functions is challenging, the benefit of object-oriented programming is their easy use later. An inexperienced coder could follow the steps in Figure 9 to create, derivatize, and fragment GQ1b d18:1/18:0 within the GLG architecture, and this is the path followed by GLG for each entry in aminoxyTMT library generation. The text-based MS spectrum at the bottom of the readout contains critical fragment ions specific to this ganglioside species and to this method of derivatization. The libraries used by MS-DIAL are collections of entries like this one.

Figure 9. Derivatization of the Ganglioside Object.



Other supported derivatizations with fragmentation include C8 aldehyde oxidation into aminoxyTMT and amide formation at the sialic acid carboxylate with the isotopic labeling tag cholamine[63]. Permethylation[64] of ganglioside alcohols is also supported but without fragmentation as this derivatization approach was not pursued beyond theory.

Ganglioside Retention Time Prediction

Figure 10 illustrates a complete set of retention time equations for GM3 gangliosides derivatized with cholamine in the Chapter IV protocol, based off a bovine GM3 standard (Matreya Catalog #1503). While most of the Matreya standards have two or three major ganglioside species, the GM3 standard is a mixture of many species, making easy the task of generating RT equations. Where there are two data points on the same vertical line, this represents what we suspect to be isomerism of ceramide double bond placement, often with one of the species being vastly more abundant than the other. When GLG is used to generate a GM3 species, it will use the most appropriate equation here to predict RT.

Figure 10. GLG RT Equations for GM3 Gangliosides After Cholamine Derivatization.

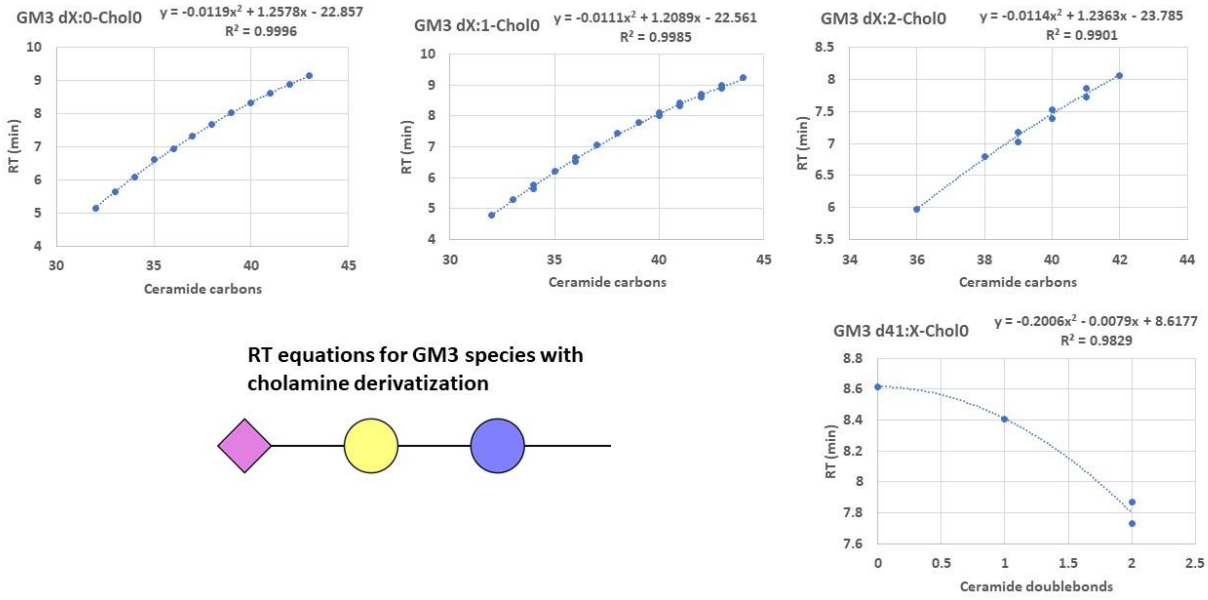


Figure 11. GLG RT Equations for GD2 Gangliosides After Cholamine Derivatization.

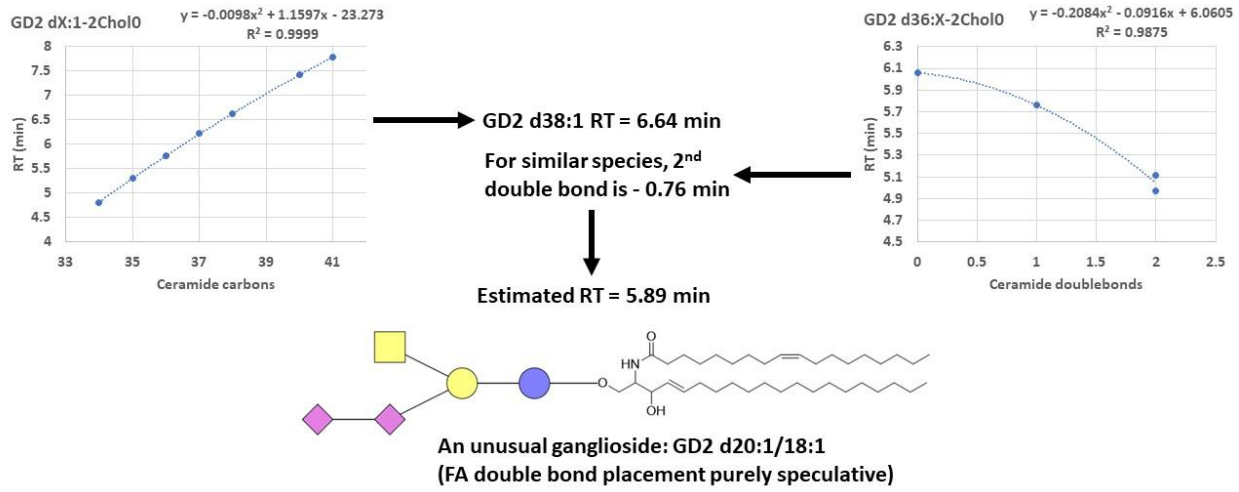


Figure 11 illustrates a more difficult RT prediction case. The rabbit GD2 standard (Matreya Catalog #1527) has far fewer abundant species compared to the previous GM3 standard, and so the RT equations we have made from this ganglioside class are less complete. When calculating RT for the example GD2 d38:2 ganglioside, GLG will first check to see if it has a GD2 dX:2 RT equation. As it does not, GLG next checks if there is a model GD2 d__:X RT equation, which it has for GD2 d36:X. As GD2 d36:2 is very similar to GD2 d38:2, GLG uses an adjustment from the GD2 d36:X equation to supplement the more complete GD2 dX:1 equation and predict the RT for GD2 d38:2. If there were no GD2 d__:X equation, GLG would next check for the existence of the following equation variety for GD2:

$$RT = A(ceramide\ carbons) + B(ceramide\ double\ bonds) + C$$

If such an equation were not available, as might be the case for more exotic or theoretical gangliosides species, GLG would next use an equation of the following variety:

$$RT = A(glycan\ sugars) + B(ceramide\ carbons) + C$$

There is one such equation for gangliosides with 0 double bonds, 1 double bond, and 2 double bonds. If an even more exotic ganglioside with more than 2 double bonds were queried, then there is the final RT prediction equation:

$$RT = A(glycan\ sugars) + B(ceramide\ carbons) + C(ceramide\ double\ bonds) + D$$

In this way, GLG uses the best RT equation available for the specific ganglioside object generated, moving gradually downward to equations with fewer assumptions and greater error. Even this RT equation of last resort is reasonably accurate with average and maximum RT errors of 0.17 min and 0.43 min, respectively, over 156 LC-MS peaks identified in the Matreya standards' chromatograms after cholamine derivatization. The distinction of LC-MS peaks is used here and not species, because each peak can and often does contain multiple species.

The GLG GUI and Operation

Figure 12. The GLG GUI for Rapid Library Generation.

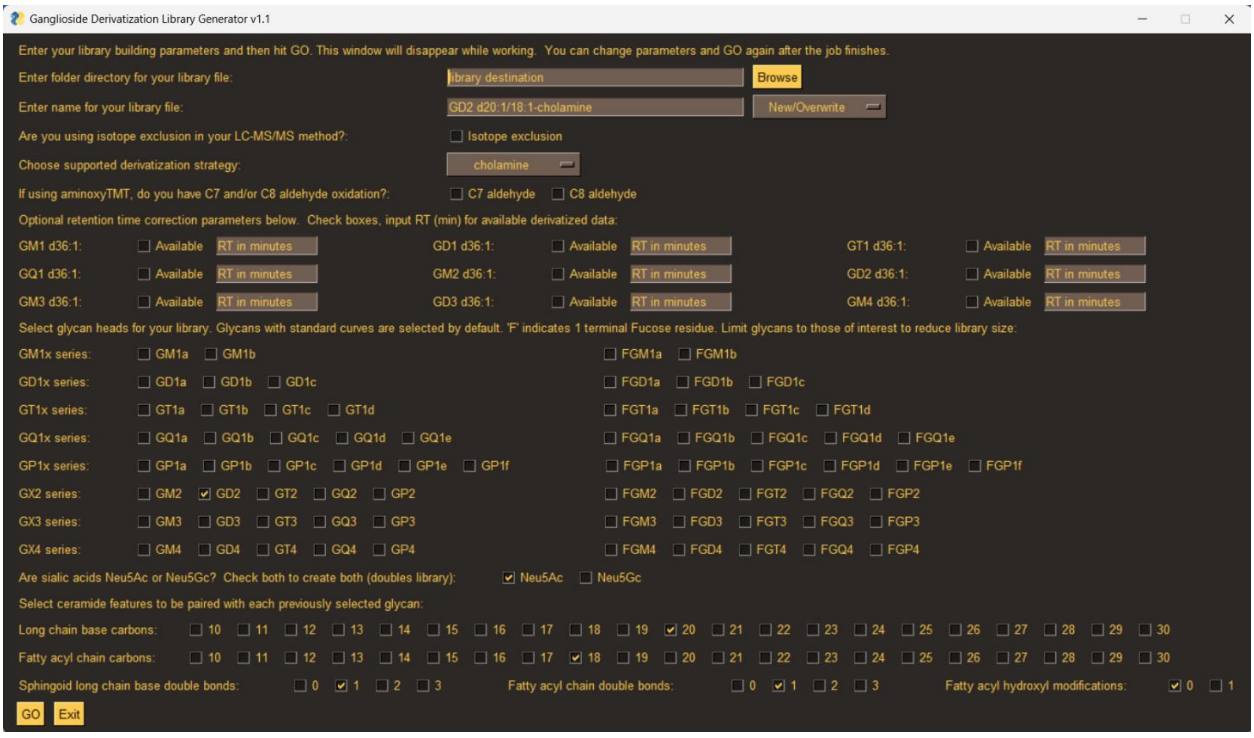


Figure 12 illustrates the current developmental GUI for GLG. This GUI is designed to interface between a non-coding user and the GLG architecture for rapid generation of libraries and immediate use with MS-DIAL. When the user clicks 'GO', every combination of the selected ganglioside features is generated as a ganglioside object and then converted into a library entry. In Figure 12, the selections necessary to make a library with a single entry, GD2 d20:1/18:1 derivatized with cholamine, have been made. From the top down, the user selects a directory where the new library generated will go. They enter a name for the library. By default, GLG will make a new library, but the user can also select from the dropdown menu to amend a previously generated library which can be useful for better tailored ganglioside lists. Next, the user indicates whether or not their LC-MS/MS method utilizes isotope exclusion. If selected, the library will include the necessary [M+1] peaks, that is, gangliosides with a natural carbon-13. Next, the user selects the derivatization strategy: aminoxyTMT0, aminoxyTMTsixplex, or cholamine. If selecting aminoxyTMT0 or aminoxyTMTsixplex, the user must also select which aldehyde intermediate to generate or to generate both.

The next block contains options for RT correction. The RT correction algorithms employed are rudimentary and linear, but they apply to all ganglioside species, not just the nine classes requested. This functionality is useful when switching between instruments with slightly different RT. The RT correction algorithms can be applied to the use of different LC gradients, which we have tested with some success, but large changes in LC gradient will yield poor RT prediction. Next, the user selects ganglioside glycans with 35 non-fucosylated glycans to choose from along with the 35 fucosylated variants. Either Neu5Ac or Neu5Gc sialic acids are selected. Gangliosides generated by GLG are either all Neu5Gc or all Neu5Ac. For example, if both options are selected, GLG will generate GQ1b's with four Neu5Gc and GQ1b's with four

Neu5Ac but not GQ1bs with two of each or other combinations. The architecture would support addition of such function if needed at a later date. Finally, the user picks ceramide elements: LCB carbons, FA carbons, LCB double bonds, FA double bonds, and FA hydroxyls.

Conclusions

While perhaps too experimental to make public, GLG is an invaluable tool for ganglioside analysis. Many of the gangliosides seen in the commercial standards analyzed do not have entries in any public databases, let alone complete MS1, RT, and MS2 data. GLG is a powerful ganglioside investigative tool, of great use even for the comparatively simple task of generating accurate molecular formulae and m/z values for underivatized species. In its original design purpose of identifying derivatized gangliosides in stable-isotope experiments, GLG has saved incalculable manual analysis time and will pay dividends for years to come.

Reporter Ion Analyzer

Introduction

Reporter Ion Analyzer (RIA) is purpose-built for post processing of isobaric analysis data from MS-DIAL[55]. Like isotopic analysis, isobaric analysis relies on the derivatization of analytes with chemical tags enriched with stable isotopes, but unlike isotopic analysis, isobaric analysis is based off the MS2 fragmentation spectrum and reporter ions therein.[16, 65] Figure 13 illustrates a set of isobaric chemical tags, the aminoxyTMTsixplex bundle sold by Thermo Fisher Scientific. The sixplex is isobaric because each of the six tags has the same mass while differing in the placement of ^{13}C and ^{15}N, marked with red stars in Figure 13. HCD fragmentation in tandem mass spectrometry produces reporter ions of appreciably different masses due to the heavy atom positions in the precursor ions. As illustrated in Figure 14 from the popular online isobaric analysis tool Metandem[66], individual samples are tagged, each with

one member of an isobaric set. The samples are then mixed to form a multiplex and analyzed by LC-MS/MS together in one injection. The reporter ion relative intensities in the MS2 spectrum indicate the contribution by each sample to the multiplex.

Figure 13. The Isobaric AminoxyTMTsixplex by Thermo Fisher Scientific

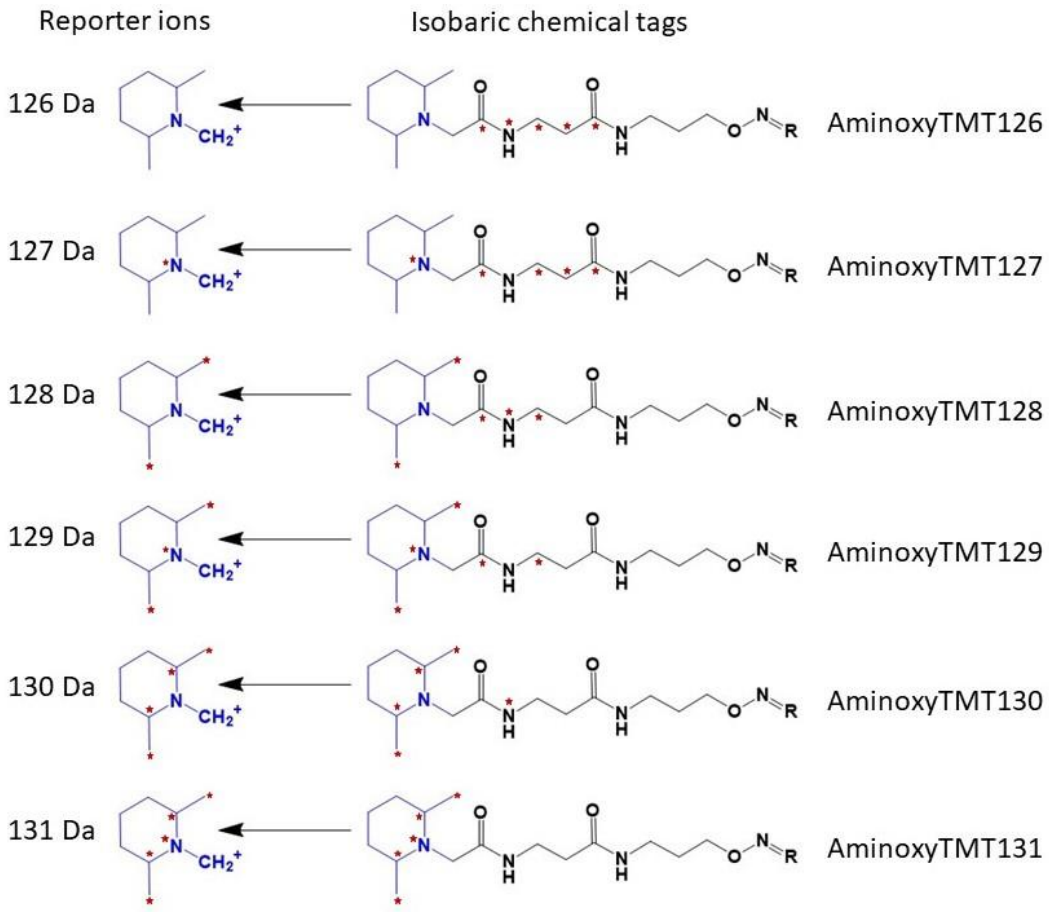
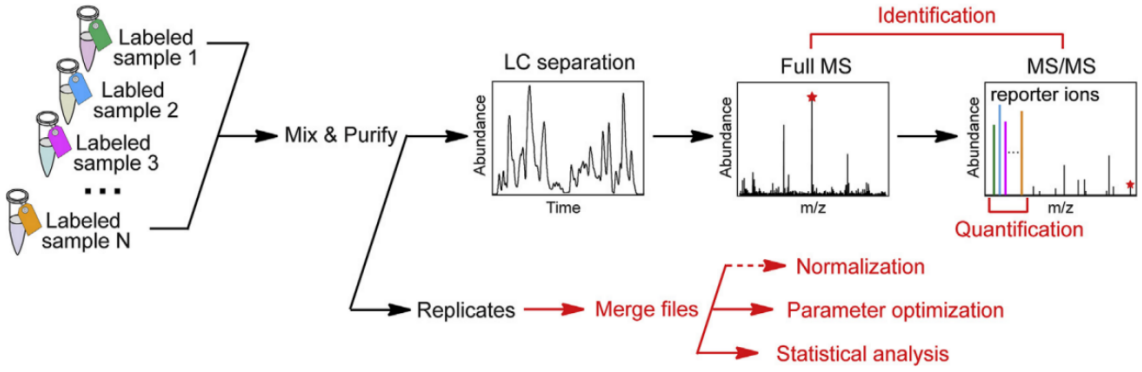
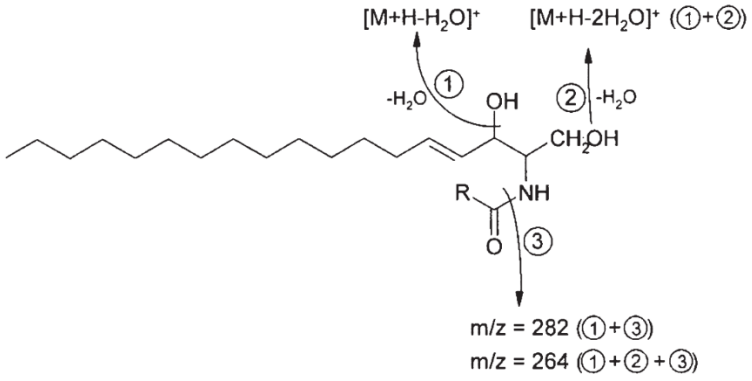


Figure 14. Isobaric Labeling Analysis in Metandem. Ref.[66]



While numerous quality tools for isobaric analysis exist, like Metandem[66], MASIC[67], and other script-based solutions[68-70], none have the capacity for MS-DIAL features deemed essential to the target ganglioside analysis workflow. We explored the possibility of building a software bridge between MS-DIAL and one of these other programs or a script to merge the results, but it became apparent that writing our own script for post processing MS-DIAL and quantifying reporter ions would be easier and of greater long-term value.

Figure 15. Ceramide Fragmentation in Tandem Mass Spectrometry. Ref.[15]



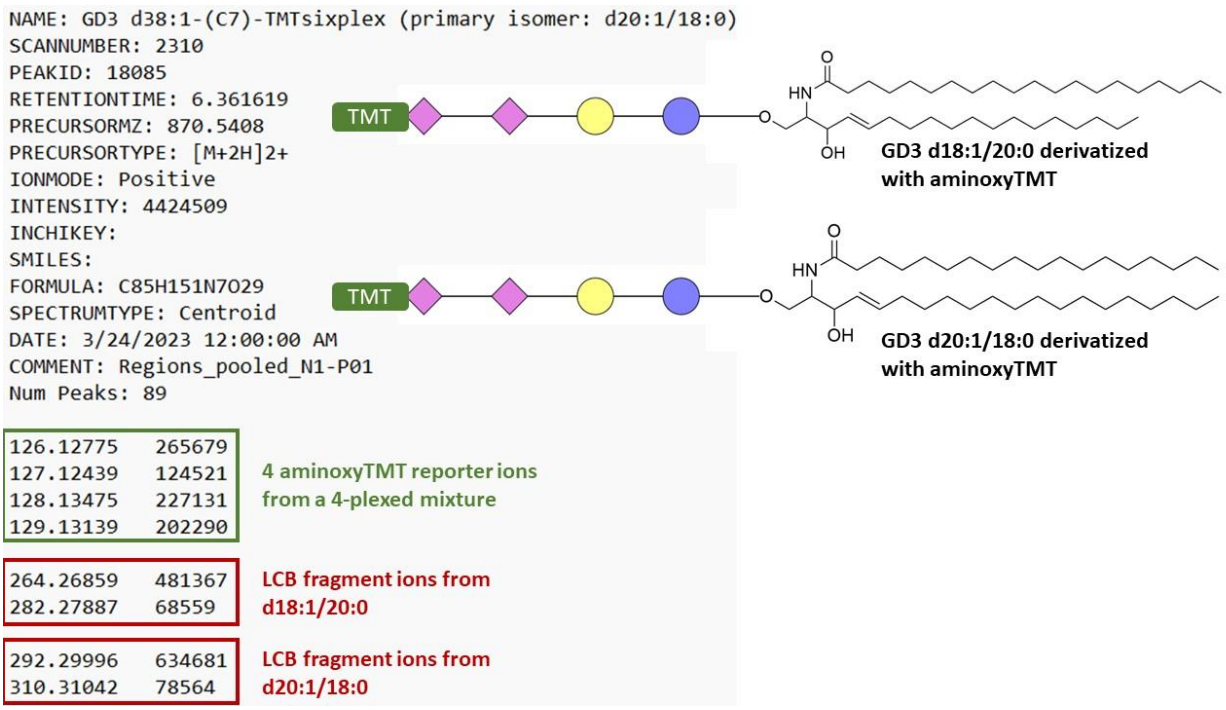
As illustrated in Figure 15, ceramides[15] and the ceramide moiety of gangliosides[16] have a unique fragmentation pattern resulting from the progressive loss of two waters and the fatty acyl. For most gangliosides, the LCB is in the MS2 fragmentation spectrum at high relative intensity as two ions, $[M+H-H_2O]^+$ and $[M+H-2H_2O]^+$.[16] Identifying these ions allows for accurate diagnosis of the total ceramide, both the LCB content explicitly and the FA content implicitly. MS-DIAL and the ganglioside library can determine the primary ceramide isomer along these lines, but it will fail to detect any coeluting lesser ceramide isomers. Treating ceramide fragments computationally as reporter ions, RIA determines the relative contribution of coeluting ceramide isomers to the overall LC-MS peak.

Methods

Peak Collection (.msp) Import

MS-DIAL has multiple export modes. RIA utilizes LC-MS/MS peak .msp files. In this mode, MS-DIAL outputs all identified peaks to individual .msp text files in a designated folder, and RIA scans through the entire folder, looking at each peak file one at a time. Figure 16 illustrates an example peak file outputted by MS-DIAL with the irrelevant fragment ions deleted for clarity. The fragment ions are listed with both their mass-to-charge ratios (left) and their relative intensities (right). This LC-MS peak is GD3 d38:1 from a multiplexed sample of four mouse brain regions, each tagged with a different member of the aminoxyTMT sixplex, details and results of this experiment further discussed in Chapter V. In this particular LC-MS peak, there are two major ganglioside ceramide isomers: GD3 d18:1/20:0 and GD3 d20:1/18:0. RIA reads in each line of this file, creating a Peak object in Python for subsequent processing.

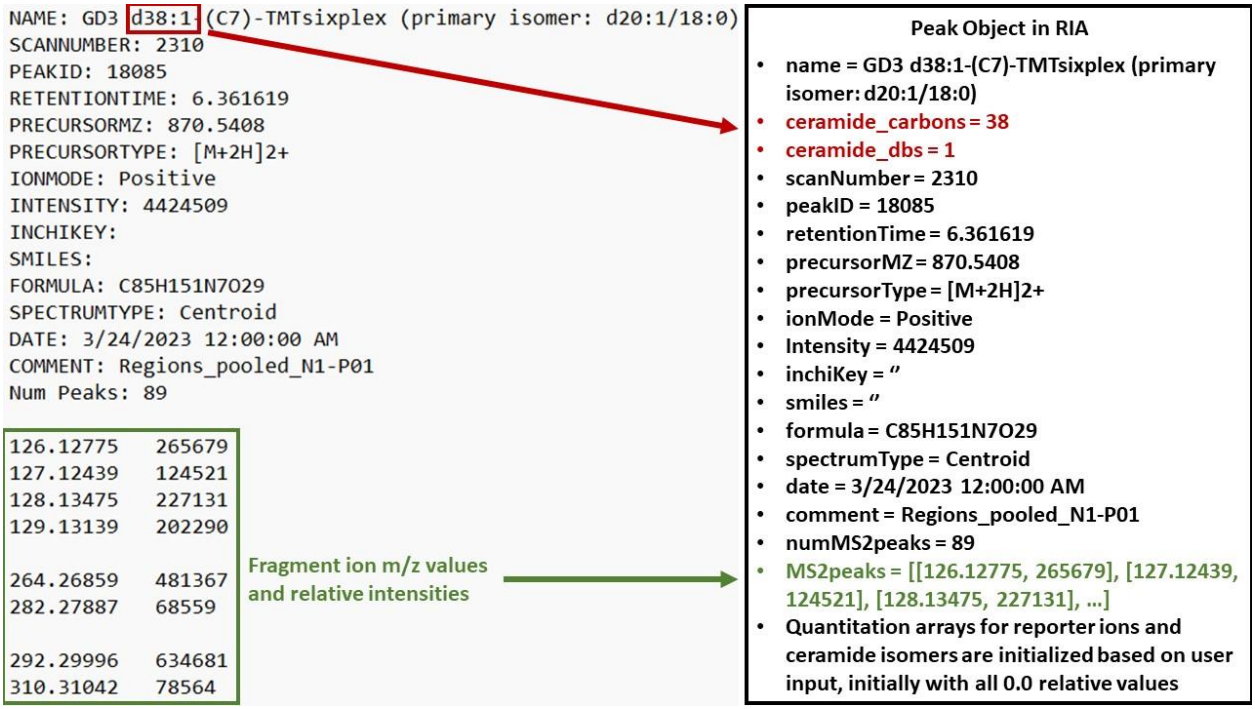
Figure 16. MS-DIAL .msp Peak Output for RIA.



The Peak Object

RIA is another example of object-oriented programming. The essential object here is the 'Peak', though in truth, it is a ganglioside peak and not a general LC-MS peak due to the current narrow focus of the program. With the value of ceramide and isobaric analysis extending beyond gangliosides, expanding RIA capacity to other species is a goal for the near future. Figure 17 illustrates the initial status of the RIA Peak object for the previous MS-DIAL output example with the two GD3 ceramide isomers. The most important variables determined on initialization are the total ceramide content and the MS2 spectrum, coded as paired mass-to-charge ratios and relative intensities as provided by MS-DIAL. The total ceramide content is used to generate a list of possible ceramide isomers and fragment ions.

Figure 17. Creation and Key Variables of the RIA Peak Object.



RIA GUI and Operation

Figure 18 shows the current developmental state of the RIA GUI. The address of the folder containing the MS-DIAL output files is selected. Ceramide isomer analysis is optionally selected. If ceramide analysis is selected, only ceramide species with both $[M+H-H2O]^+$ and $[M+H-2H2O]^+$ characteristic fragments detected will be reported. The aminoxyTMTsixplex reporter ion set is a hardcoded menu option, but RIA can also be run without reporter ion analysis (just ceramide analysis) or with a custom .txt reporter ion list. The user can optionally delete metabolites that are missing one or more defined reporter ions. A MS2 tolerance for matching queried ions vs the measured MS2 spectrum is specified and used in the matching algorithm. When the user clicks 'GO', RIA iteratively checks every MS2 m/z value against the reporter ion list and possible ceramide ion sets. Any matches increment the original signal values of 0.0 with the relative intensity from MS-DIAL, finally weighted out of 100 and reported in an Excel file after processing.

Figure 18. RIA Development GUI.

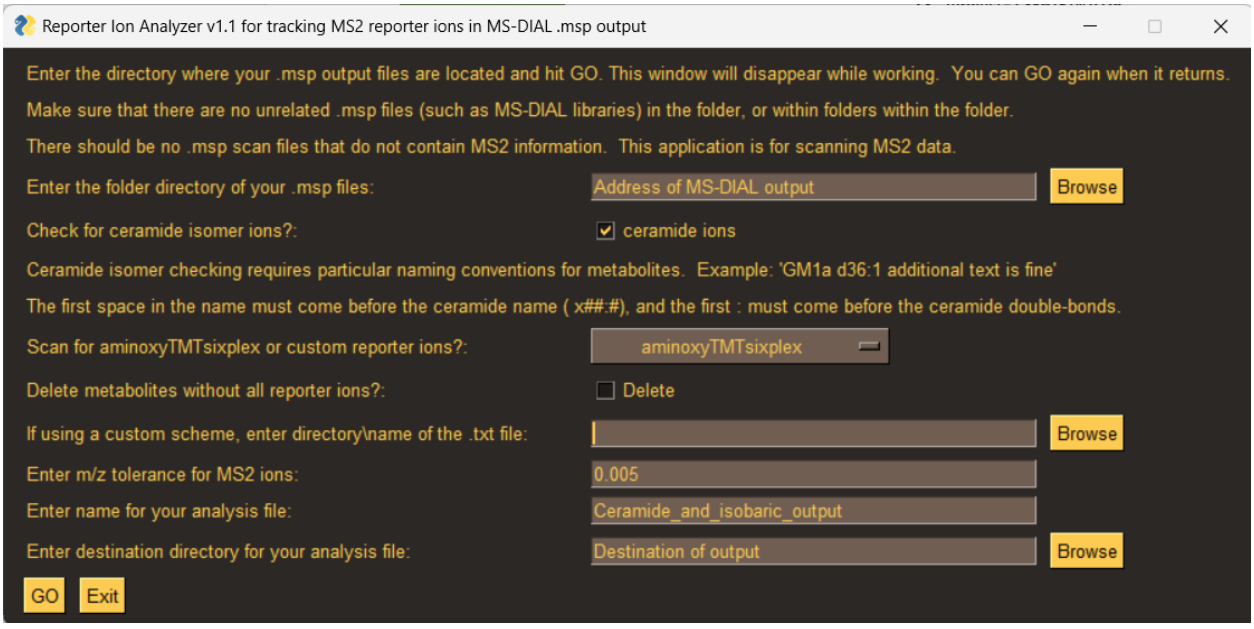


Figure 19. RIA Output and Interpretation.

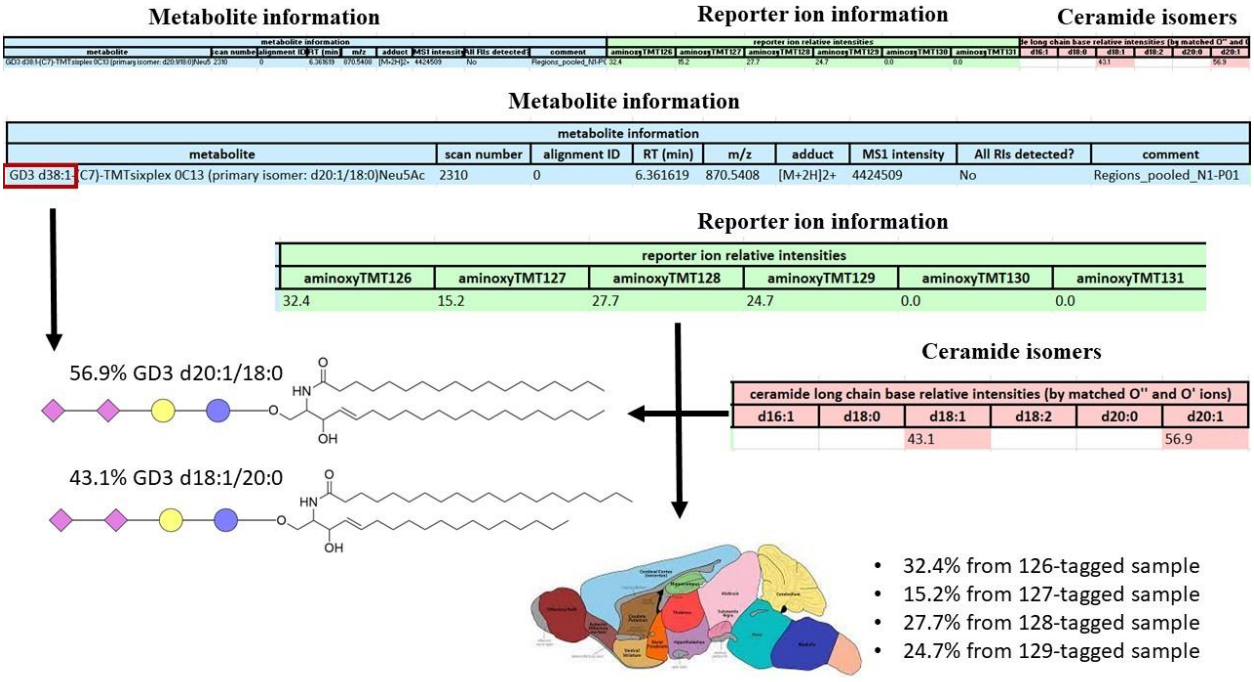


Figure 19 illustrates a small RIA output for the Figure 17 Peak object, ordinarily part of a larger dataset. RIA builds a single Excel file from the collection of peaks analyzed, each peak becoming one row in the output file, carrying forward metabolite information and displaying fragment ion relative quantifications. The reporter ion weighted relative intensities indicate the contribution of the multiplexed samples, which are four different mouse brain regions in this example. The ceramide isomers and original metabolite identification allow us to diagnose the total ceramide isomerism of the peak above a very low abundance threshold. In reporting, RIA refers to the ceramide LCB fragment ions as O' and O'', as is seen elsewhere in the literature.[16] Each apostrophe indicates the loss of a water molecule, while O is the ceramide fragment after losing the glycan and fatty acid.

Conclusions

RIA streamlines what would otherwise be a tedious post MS-DIAL processing task and enables high throughput isobaric quantitation of multiplexed samples following MS-DIAL's automated annotation. Ceramide isomerism reporting is a niche application but one of great use to ganglioside analysis with ceramide variation connected to metabolic function and cellular aging. In its current version, RIA is applicable to any reporter ion of interest through the input for custom ion lists but is only applicable to LC-MS peaks following particular naming conventions inclusive of gangliosides but not necessarily other ceramide-containing lipids.

Peak Pair Pruner

We developed Peak Pair Pruner (PPP) to supplement MS-DIAL's isotopic labeling features. In isotopic labeling, one sample is derivatized with a 'light' tag while another is derivatized with an isotopically enriched 'heavy' tag. Mixing the samples allows for relative quantification between the two peaks. This is of value for ganglioside analysis due to the lack of useful standards for traditional absolute quantification strategies. MS-DIAL does the essential work of pairing isotopic peaks, but it pairs too many, leading to a need for 'pruning' down to the experimentally relevant peaks pairs. MS-DIAL also does not have an appropriate output format for isotopic labeling data, such as is seen on more specialized platforms with tabular reporting of light-to-heavy ratios. Critically, MS-DIAL does not deal with the most significant quantitative hurdle in isotopic labeling, which is isotopic overlap. Gangliosides and other large metabolites have appreciable isotopic peaks in MS1, even as high as [M+6]. If an isotopic tag is used that imitates a high abundance isotope, such as [M+3], there will be overlap of the isotopic envelopes, which must be corrected. The MS-DIAL+PPP workflow addresses MS-DIAL's isotopic labeling gaps and provides quantitative corrections, further discussed in Chapter III.

CHAPTER III: PEAK PAIR PRUNER

Introduction

An essential approach for comparative metabolomics studies is stable isotopic labeling, in which analytes in one sample are derivatized with a ‘light’ tag and those in another sample with a ‘heavy’ or isotopically enriched tag. These samples undergo pooling and mixing strategies to allow for relative quantification in large cohorts. MS-DIAL[55] is one of the most popular general metabolomics analysis platforms[11] and is seeing increasing usage, referenced in 386 articles in 2020, 713 in 2021, and 920 in 2022, respectively on Google Scholar (searched on Jan 6, 2023). It is free and versatile, even annotating metabolites with user-generated tandem mass spectrometry libraries, but its features in handling isotopic labeling data fall short of specialized software. IsoMS[71] is one such specialized software, designed for the dansylation chemistry-based derivatization[72]. IsoMS performs pairing of isotopic peaks and identification based on accurate mass and retention time, providing a tabular report of metabolites with their L/H ratios in the samples. While IsoMS is robust and well suited for its target chemistry, it is a commercial product and there is a desire for more generalized software analyzing isotopic labeling data. MS-IDF[73] presents a non-commercial alternative for isotopic labeling analysis but compares just two chromatograms at a time and so is not suited for large cohort studies with its current release. Leveraging the powerful raw data processing capabilities of MS-DIAL in metabolomics, we built Peak Pair Pruner (PPP) for MS-DIAL post processing, providing the missing, specialized isotopic labeling features with innovation not represented in other software options.

Methods

Alignment Matrix Import

PPP requires an MS-DIAL alignment matrix. Utilizing keywords and isotopic labeling naming convention, PPP collates metabolite, blank, sample, light QC, heavy QC, mix QC, and replicate data into an internal data array.

Isotopic Screening

Matching isotopic relationships from MS-DIAL and experimental parameters provided by the user, PPP searches the internal data array for peak pairs that are potentially due to the user's isotopic labeling experiment, accounting for different charge states and adduct species.

Pea Pair Mass Validation

Mass defect filtering is optionally applied based on user-defined upper and lower mass defect limits. Accurate mass difference between paired peaks is validated against user-defined heavy tag shift and mass ppm tolerance.

Peak Pair Quantitative Corrections

Background peak values are subtracted utilizing a blank. Isotopic overlap between light and heavy tagged analytes is subtracted utilizing the light pool QC.

Peak Pair QC Ratio Validation

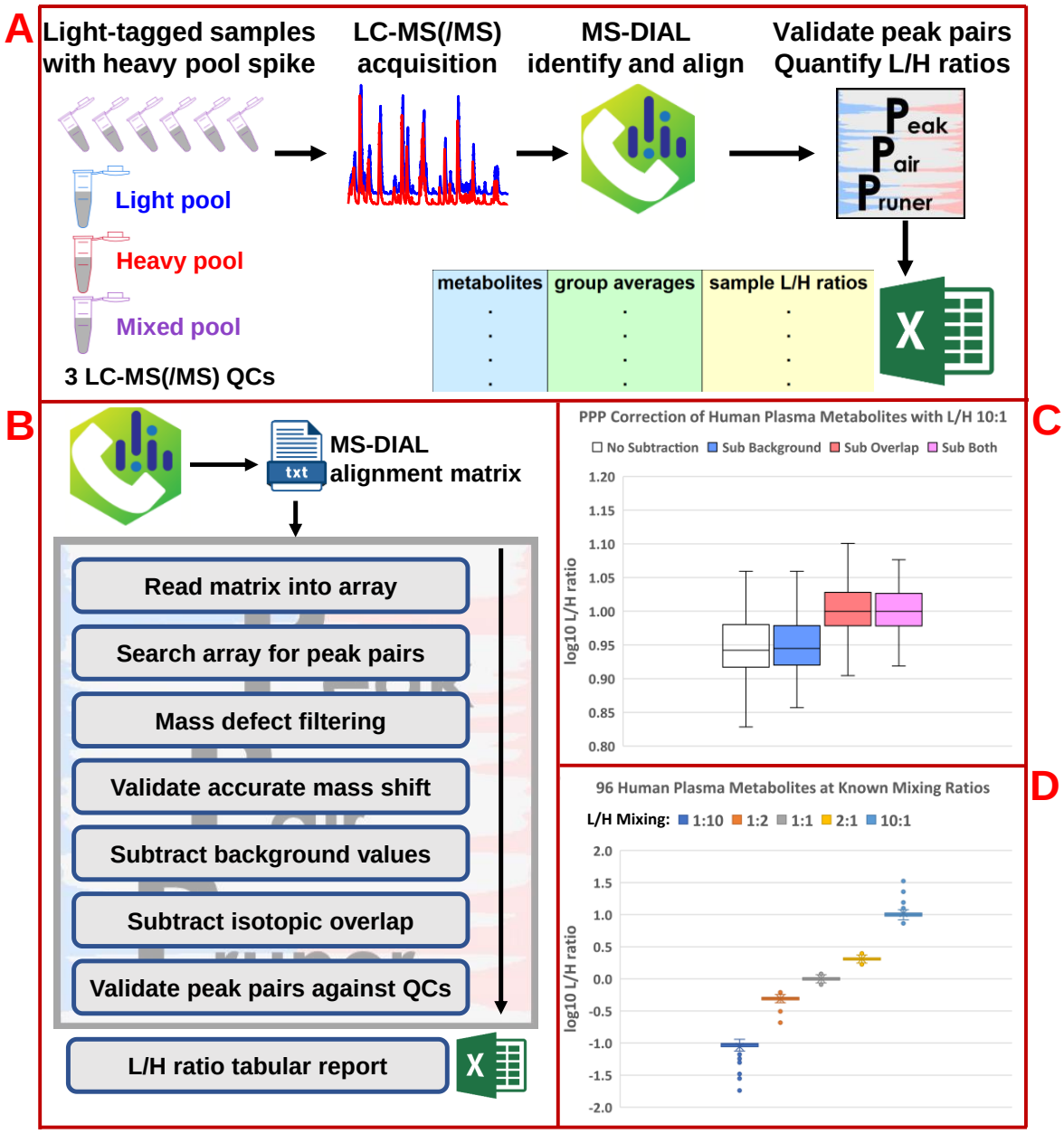
Peak pairs are validated against minimum light QC L/H ratio, minimum heavy QC H/L ratio, theoretical mix QC L/H ratio, and mix QC L/H ratio tolerance.

PPP was implemented in Python utilizing the PySimpleGUI and XlsxWriter packages. We have exported PPP to a single executable program that is independent of its original Python IDE. Technical details are described in the Supplementary Material and the Github ReadMe file.

Results

To assess quantitation and demonstrate the MS-DIAL+PPP workflow's capabilities, we conducted two analyses: (i) dansylation of a 17 amino acid standard mixture and (ii) dansylation of pooled human plasma, both with known L/H ratios of 1:10, 1:2, 1:1, 2:1, and 10:1 with high mass resolution LC-MS data acquisition. In analysis (i), all 17 amino acids were identified by MS-DIAL and validated, quantified by PPP. Further details are in the Supplementary Material. In analysis (ii), MS-DIAL produced an alignment matrix with 3501 peaks, among which PPP found 98 identified potential peak pairs and 701 unknown potential peak pairs. PPP validated 96 identified peak pairs and 378 unknown peak pairs. The drop in unknown peak pairs shows the capability of PPP in validation and removing false positive peak pair identifications, a common issue in metabolomics-related database search. Figure 20 illustrates (20A) the total isotopic labeling workflow with MS-DIAL+PPP, (20B) PPP coding architecture, and (20C, 20D) key results of analysis (ii). Panel C illustrates the usefulness of isotopic overlap subtraction where in the L/H 10:1 sample, the peak pair ratio log10 values are much closer to the expected value of 1.00 following correction. Panel D illustrates the accuracy of MS-DIAL+PPP quantitation over a range of metabolite L/H ratios. Noteworthy is the tailing of values at the extreme ratios of 1:10 and 10:1, indicating that these ratios are close to the limit of quantitation for low abundance metabolites. Further details are in the Supplementary Material.

Figure 20. MS-DIAL+PPP Workflow, PPP Processing, and Validation Results.



(A) MS-DIAL+PPP isotopic labeling workflow in LC-MS metabolomics. Samples are split, mixed to form two pools and aliquots for analysis. One pool is light tagged (light pool QC), while the other is heavy tagged (heavy pool QC). Pools are combined at a known ratio (optimally 1:1) to form the mix pool QC. Analysis aliquots are light tagged and then spiked with

the heavy pool at the known ratio to make analysis samples. Analysis samples undergo LC-MS(/MS) acquisition, then MS-DIAL peak identification and alignment. PPP performs peak pair validation and quantitative correction based on ratios in the QCs. (B) Python coding architecture and processing by PPP. (C) L/H 10:1 dansylated human plasma analysis outcomes with and without PPP quantitative corrections. Metabolite peaks may overlap with background peaks, and chemical tags used in light/heavy analysis may overlap in their isotopic envelopes. PPP can optionally correct for background peaks and for isotopic overlap. In the 10:1 L/H sample, the theoretical log10(L/H) value is 1.00, most closely attained when using both corrections. (D) Workflow quantitative testing of dansylated human plasma across 100-fold range of theoretical L/H values.

Supplementary Material

Materials and Methods

Chemicals and Materials

Pooled K2EDTA human plasma (LOT#:HMN654203) was purchased from BioIVT. Amino acid standard mixtures (product AAS18-10ML) and sodium carbonate were purchased from Sigma-Aldrich. MS pure $^{13}C_2$-dansyl chloride and $^{12}C_2$-dansyl chloride were purchased from Nova Medical Testing (Edmonton, Alberta, Canada). Other LC/MS grade solvents and reagents were purchased from Thermo Fisher Scientific.

Metabolite Sample Preparation

The dansyl chloride isotopic labeling protocol[72] was adapted as follows to label human plasma and a 17 amino acid standard mixtures. Briefly, metabolites were extracted from a pooled BioIVT human plasma sample using methanol precipitation. The extract was dried and reconstituted in 50% acetonitrile for isotopic labeling. Amino acid standard mixture was also

dried first before reconstitution in 50% acetonitrile. Aliquots were then labeled with either the light $^{12}C_2$-dansyl chloride or the heavy $^{13}C_2$-dansyl chloride. After quenching the reaction mixture with 240 mM NaOH, light and heavy labeled samples were recombined to create pools with 1:10, 1:2, 1:1, 2:1, and 10:1 mixing ratios to assess quantitation accuracy.

Mass Spectrometry Analysis

A Thermo Scientific Vanquish Horizon UHPLC System coupled with a Thermo Scientific Q Exactive HF Orbitrap was used to conduct LC-(+)FTMS analysis. A Waters ACQUITY UPLC BEH C18 column (1.7 μm, 2.1 x 100 mm) was used along with the following mobile phases: (A) 0.1% formic acid in water and (B) 0.1% formic acid in acetonitrile. The gradient was 25% B (0 min), 25% B (1.2 min), 99% B (11.2 min), 99% B (16.2 min), 25% B (16.3 min), and 25% B (19 min) to achieve column equilibration. The column was maintained at 40 °C with flowrate of 400 μL/min. The injection volume was 5.00 μL. Each mixture and QC was sampled in triplicate. Ionization for FTMS analysis utilized a HESI source in the positive mode with spray voltage 3 kV, capillary temperature 300 °C, sheath gas flowrate 35 au, aux gas flowrate 10 au, probe heater temperature 300 °C, and S-lens RF level 50 V. FTMS scanning was full MS1 in the range 250 to 1000 m/z with maximum IT 200 ms, AGC target $5x10^5$, and mass resolution 60,000.

MS-DIAL Peak Pair Identification

MS-DIAL 4.92 for Windows x64 was used alongside an in-house metabolite dansylation library. For the amino acid mixture, MS-DIAL data collection proceeded with MS1 tolerance 0.005 Da and a maximum charge number 2. The minimum height was set to 10^6 with mass slice width 0.05 Da. The peak smoothing method was linear weighted moving average with a smoothing level of 3 scans and a minimum peak width of 5 scans. The Sigma window value was

set to 0.1 without exclusion after precursor ions and with isotopic ions w/o MS2Dec preserved. Automated identification was performed with a text file library with retention time tolerance 0.1 min, accurate mass tolerance 0.901 Da, identification score cut off 85%, and only reporting the top hit. Peak alignment was performed in reference to the light pool QC with retention time tolerance 0.04 min, MS1 tolerance 0.004 Da, removing features based on blank information, sample max / blank average minimum fold change 5, keeping all metabolite features, and gap filling by compulsion. Isotope tracking was performed with ^{13}C as the labeled element, the light pool QC as the non-labeled reference file, and the heavy pool QC as the fully-labeled reference file. After MS-DIAL processing, the alignment results were exported in mgf format, choosing export option 'raw data matrixes (Height)', with filtering by blank ion abundances, and filtering by the result of isotope labeled tracking in reference to the heavy pool QC. For the human plasma samples, the parameters were nearly the same with identification retention time tolerance 0.5 min and alignment MS1 tolerance 0.003 Da. Alignment tolerances needed tightening for the human plasma experiment due to greater matrix complexity.

PPP Peak Pair Validation and Ratio Quantification

We have compiled PPP into an individual executable program that runs independently of MS-DIAL. The researcher first provides a matrix alignment file from MS-DIAL, which can have quantitative values in the form of peak areas or of average peak heights. PPP reads the alignment file into an internal array. PPP then searches the internal array for peak pairs that could be originated from the researcher's isotopic labeling experiment. This is where the greatest 'pruning' occurs, as the MS-DIAL alignment file may have tens of thousands of aligned peaks, most of which are unrelated to isotopic labeling. Next, PPP conducts mass defect filtering, keeping peak pairs with mass defects within a window set by a minimum and maximum

mass defect. PPP then validates the accurate mass shift between the peaks of each peak pair, checking that the mass difference conforms to the light and heavy labels within a given ppm tolerance. Next comes background subtraction based on the blank, followed by subtraction of heavy isotope overlap based on the light pool QC. Subtraction of the natural heavy abundance is done through the following equations:

$$\text{Equation 1: } R_{natural\ heavy} = \frac{S_{heavy,light\ QC}}{S_{light,light\ QC}}$$

$$\text{Equation 2: } S_{heavy,corrected} = S_{heavy,measured} - S_{light} R_{natural\ heavy}$$

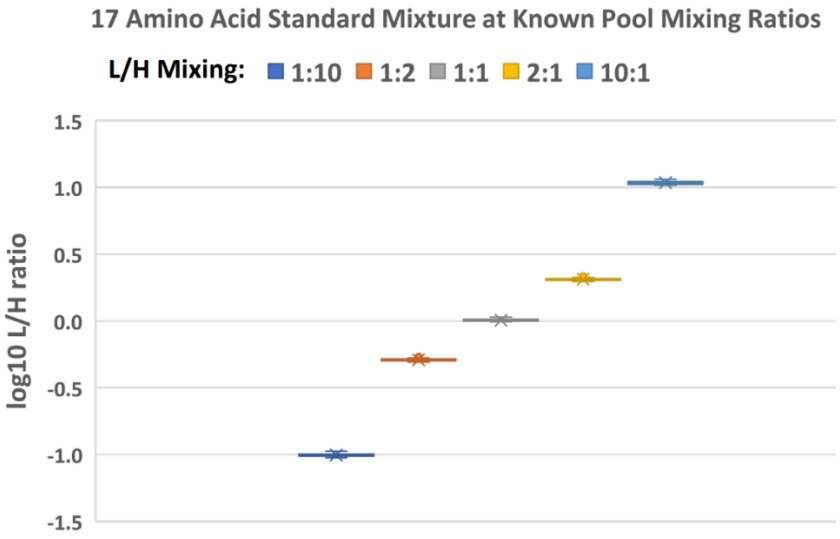
In Equation 1, the natural heavy abundance per light abundance is established from the light pool QC. In Equation 2, the natural heavy abundance for each sample, mix QC, and replicate is subtracted based off the light peak measured in that particular injection. Both subtractions reduce peak pair values to a minimum of zero and are optionally selected by the user such that four choices are possible: no subtraction, background subtraction, overlap subtraction, and both subtractions. If both are selected, background subtraction is applied before overlap subtraction. After quantitative corrections, peak pairs face a final validation against the L/H ratios of the QCs: a minimum L/H ratio in the light pool QC, a minimum H/L ratio in the heavy pool QC, and the L/H ratio in the mix pool QC related to experimental design with a specified tolerance window. Validated peak pairs are outputted as a Microsoft Excel spreadsheet file with metabolite information, group average L/H values, sample L/H values, and mixed pool QC L/H values.

For the amino acid mixture, PPP validated peak pairs in the MS-DIAL alignment result raw data matrix (Height) with the following parameters: mass defect window floor -450 mDa, mass defect window ceiling 300 mDa, minimum L/H ratio in the light pool QC 10.0, minimum

[H/L] ratio in the heavy pool QC 100.0, theoretical mix pool QC L/H ratio 1.0, mix pool QC L/H tolerance 0.2, 1 and 2 tags per molecule, exact peak pair mass shift 2.00671 Da, mass shift tolerance 10.0 ppm, background subtraction enabled, and isotopic overlap subtraction enabled. For the human plasma, the PPP parameters were nearly identical with a mass defect window floor -500 mDa and a mass defect window ceiling 499 mDa. The mass defect window [-500 mDa, 499 mDa] effectively disables mass defect filtering, which is the appropriate choice in the absence of a target metabolite class or *a priori* knowledge of the sample.

Results and Discussion

Figure 21. MS-DIAL+PPP Validation with Dansylated Amino Acids.



Boxplot representation of MS-DIAL+PPP analysis of dansylated AA standard mixture. Light (L) and heavy (H) pools were mixed at 1:10, 1:2, 1:1, 2:1, and 10:1 to assess workflow quantification with expected $\log_{10}$(L/H) values of -1.0, -0.3, 0.0, 0.3, and 1.0, respectively.

Figure 21 illustrates the results of the amino acid dansylation experiment. The amino acid standard mixture from Sigma Aldrich contains 17 amino acids. All 17 light-tagged amino acids were automatically identified in MS-DIAL with an inhouse library. They were paired with

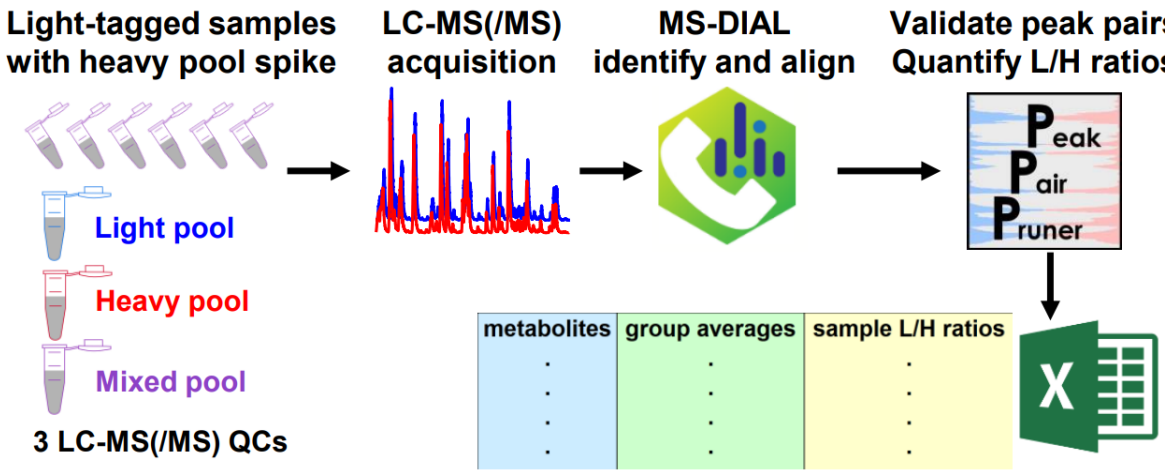
M+2 isotopic peaks in the case of singly tagged amino acids and M+4 isotopic peaks in the case of doubly tagged amino acids along with other experiment-unrelated isotopic relationships. PPP then validated all 17 experimental peak pairs and reported the amino acid L/H ratios in each mixture sample. The PPP-reported L/H values conform very nicely to the experimental pool mixing ratios. As the expected L/H values cover a 100-fold range in this experiment, these values were transformed to $\log_{10}$ values in Excel for presentation.

PPP Isotopic Labeling Metabolomics Workflow Instructions with Examples

Available on Github: https://github.com/QibinZhangLab/Peak_Pair_Pruner

This PPP document is the 'ReadMe' file. It contains instructions and use context.

Figure 22. MS-DIAL+PPP Workflow for Isotopic Labeling LC-MS(/MS) Metabolomics.



Experimental Design Considerations

Past isotopic labeling data may or may not be suitable for this workflow based on the experimental design used. If evaluating data compatibility or planning a new isotopic labeling experiment with PPP in mind, the following samples with LC-MS(/MS) chromatograms are required:

A blank. A reagent blank is best, such as a 1:1 mixture of a light reagent blank and a heavy reagent blank, but even a mobile phase blank will do. Triplicate or higher sampling is recommended. Blanks should run in series with your samples and QCs.

3 pool QCs: light-tagged, heavy-tagged, and mixed. You should pool equal aliquots of all samples prior to derivatization and from this common pool create a light-tagged pool and heavy-tagged pool. These two pools serve as essential PPP QCs. The third pool QC is a mixture of the first two pools at a known ratio. 1:1 L/H mixing is recommended for most reliable quantification, but this can be increased to 2:1 or even 10:1 at high analyte concentrations if the scarcity of the heavy tag demands. Triplicate or higher sampling is recommended. It is also recommended that the heavy-tagged pool QC be analyzed in series with but before other QCs and samples, such as: Blanks, heavy pool, and then the rest.

Light-tagged samples spiked with heavy pool. Any number of samples and replicates is permitted with triplicate or higher sampling recommended if instrument time allocation allows. First derivatize individual samples with your light tag and then spike in heavy-tagged pool at the same ratio as your mix pool QC.

This document discussed the MS-DIAL+PPP workflow while referencing an example dansylated[72] human plasma experiment. The .raw files and .ibf files analyzed are on Metabolomics Workbench Study ID ST002427. The metabolite library, MS-DIAL output, and PPP output are on Github.

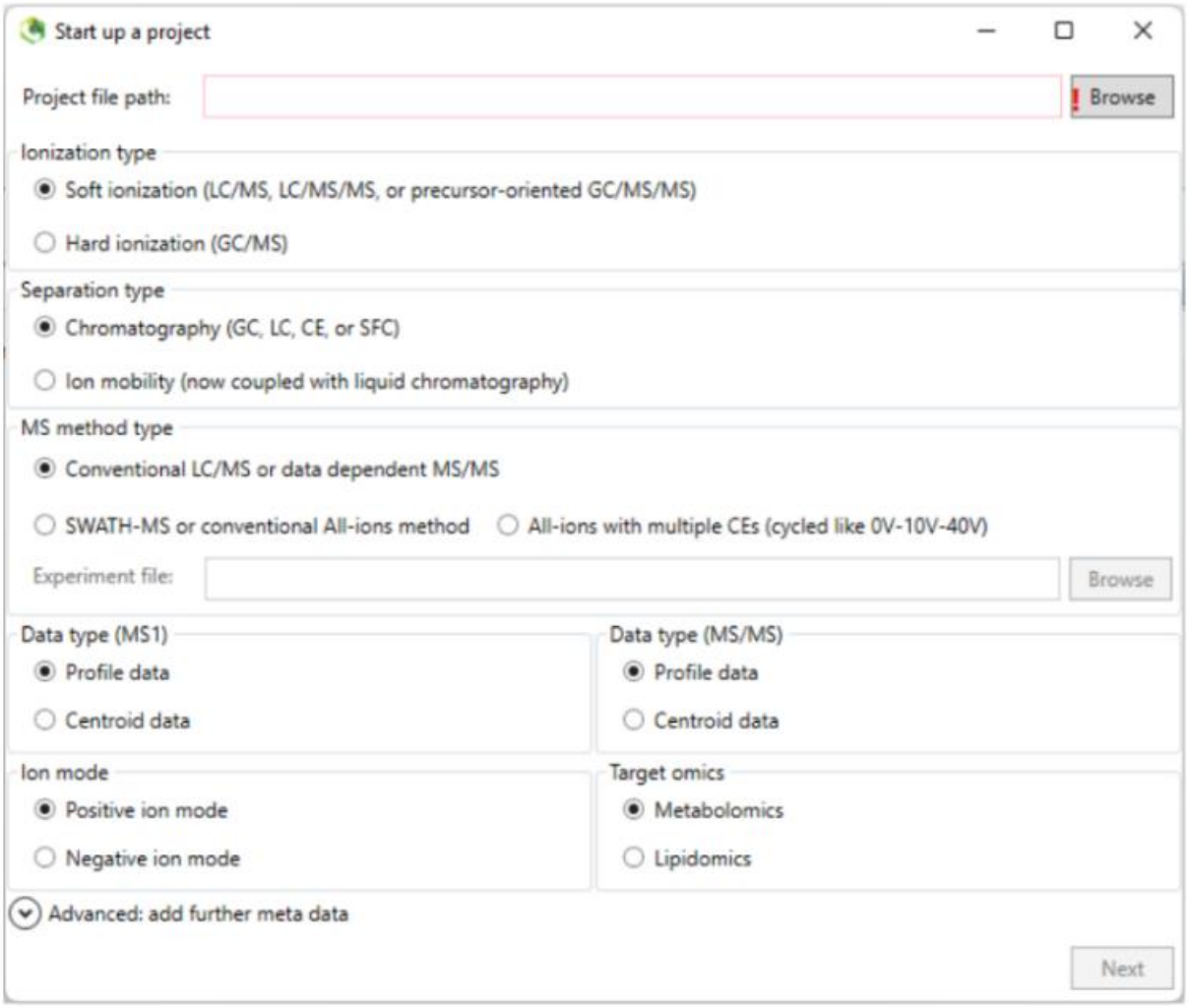
- Light and heavy reagent blanks sampled in duplicate
- Heavy, light, and 1:1 mix pool QCs sampled in triplicate
- Samples imitating a 5 individual cohort with 100-fold metabolite concentration spread:

- 0010P: L/H 1:10 sampled in triplicate
- 0050P: L/H 1:2 sampled in triplicate
- 0100P: L/H 1:1 sampled in triplicate
- 0150P: L/H 2:1 sampled in triplicate
- 1000P: L/H 10:1 sampled in triplicate

MS-DIAL Peak Alignment, Identification, and Isotopic Pairing

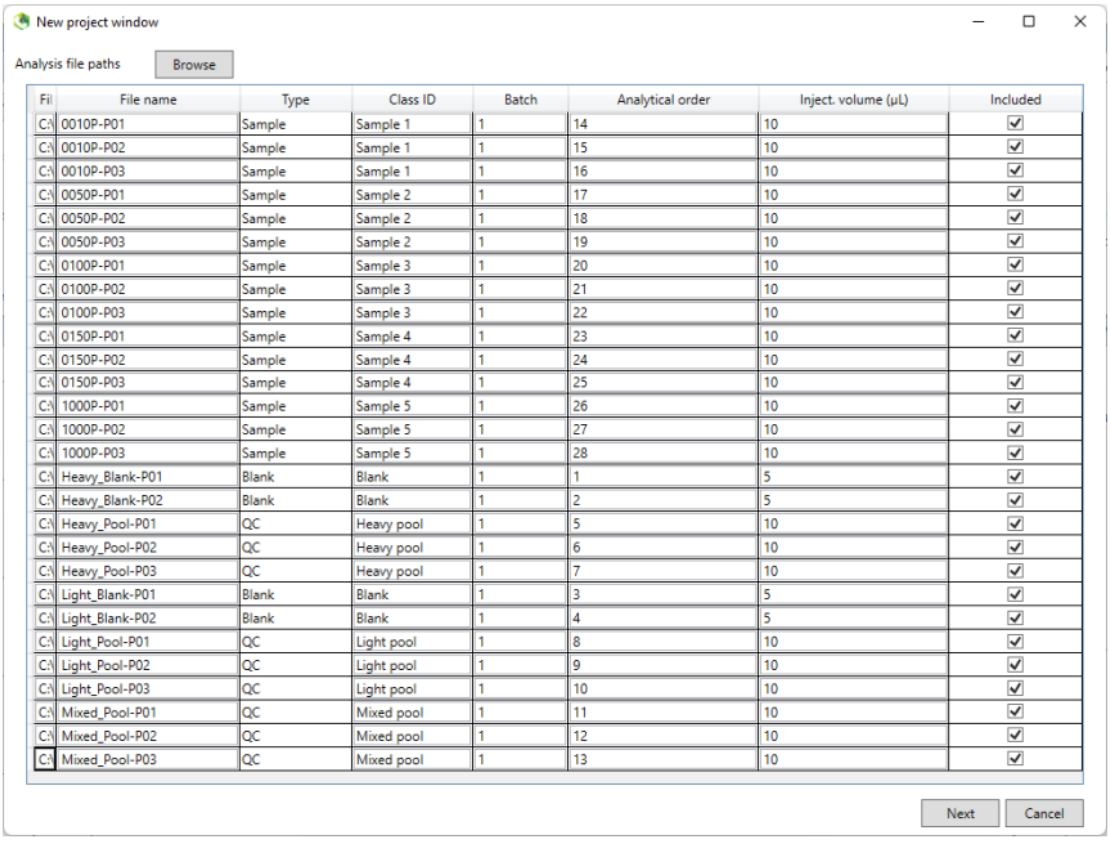
MS-DIAL[55] is a generalized metabolomics platform for LC-MS(/MS) and GC-MS analysis. A MS-DIAL project must be set up in a particular way if the goal is PPP peak pair validation and quantification. This document will walk through these particulars and briefly summarize other parameters for the example human plasma dataset. To learn more about MS-DIAL's capabilities and the function of its parameters, refer to the detailed online tutorial. Our final data was reported utilizing MS-DIAL 4.9.2, but we have also tested it successfully using MS-DIAL 4.9.0 and MS-DIAL 5.1. For this dataset, the main difference between MS-DIAL 4.9.0 and MS-DIAL 4.9.2 or 5.1 is that conversion from .raw to .ibf is no longer required. In this experiment, we found that continuing to use .ibf files was superior. It seems that there is some data loss by MS-DIAL when using .raw instead of first converting to .ibf. As such, you may consider using .ibf on later versions as well. MS-DIAL versions 4.9.0 and earlier come with a simple .ibf converter program that can be used to convert files for 4.9.0 but also for later versions.

Figure 23. MS-DIAL New Project Window.



Make a new folder for your MS-DIAL project to contain all of its related files and set this folder as the project path. When using an inhouse identification library as in this example experiment, you must select Metabolomics beneath Target omics.

Figure 24. MS-DIAL Analysis File Paths and Organization.

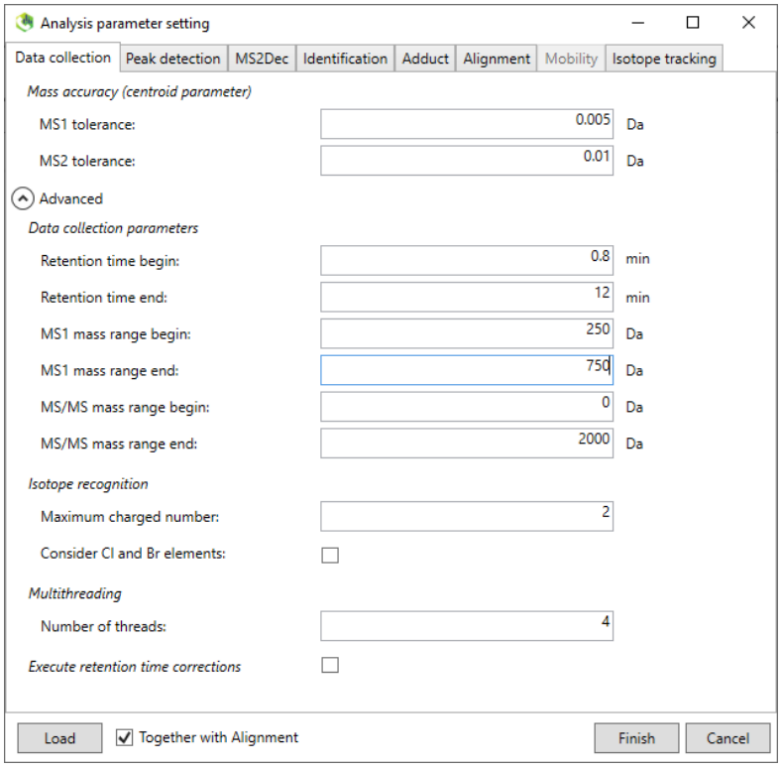


Fil	File name	Type	Class ID	Batch	Analytical order	Inject. volume (µL)	Included
C:\	0010P-P01	Sample	Sample 1	1	14	10	☑
C:\	0010P-P02	Sample	Sample 1	1	15	10	☑
C:\	0010P-P03	Sample	Sample 1	1	16	10	☑
C:\	0050P-P01	Sample	Sample 2	1	17	10	☑
C:\	0050P-P02	Sample	Sample 2	1	18	10	☑
C:\	0050P-P03	Sample	Sample 2	1	19	10	☑
C:\	0100P-P01	Sample	Sample 3	1	20	10	☑
C:\	0100P-P02	Sample	Sample 3	1	21	10	☑
C:\	0100P-P03	Sample	Sample 3	1	22	10	☑
C:\	0150P-P01	Sample	Sample 4	1	23	10	☑
C:\	0150P-P02	Sample	Sample 4	1	24	10	☑
C:\	0150P-P03	Sample	Sample 4	1	25	10	☑
C:\	1000P-P01	Sample	Sample 5	1	26	10	☑
C:\	1000P-P02	Sample	Sample 5	1	27	10	☑
C:\	1000P-P03	Sample	Sample 5	1	28	10	☑
C:\	Heavy_Blank-P01	Blank	Blank	1	1	5	☑
C:\	Heavy_Blank-P02	Blank	Blank	1	2	5	☑
C:\	Heavy_Pool-P01	QC	Heavy pool	1	5	10	☑
C:\	Heavy_Pool-P02	QC	Heavy pool	1	6	10	☑
C:\	Heavy_Pool-P03	QC	Heavy pool	1	7	10	☑
C:\	Light_Blank-P01	Blank	Blank	1	3	5	☑
C:\	Light_Blank-P02	Blank	Blank	1	4	5	☑
C:\	Light_Pool-P01	QC	Light pool	1	8	10	☑
C:\	Light_Pool-P02	QC	Light pool	1	9	10	☑
C:\	Light_Pool-P03	QC	Light pool	1	10	10	☑
C:\	Mixed_Pool-P01	QC	Mixed pool	1	11	10	☑
C:\	Mixed_Pool-P02	QC	Mixed pool	1	12	10	☑
C:\	Mixed_Pool-P03	QC	Mixed pool	1	13	10	☑

This is the most important window to double check. In the new project window, assign the appropriate Type for your samples, pooled QCs, and blanks. Providing the correct Class IDs is essential for downstream quantitation. Your light pool replicates must have the same Class ID (recommend copy-pasting one name), and this name must begin with either ‘Light’ or ‘light’. Your heavy pool replicates must also have the same class ID, and this name must begin with either ‘Heavy’ or ‘heavy’. Your mixed pool replicates must also have the same class ID, and this name must begin with either ‘Mix’ or ‘mix’. Regardless of blank strategy, your blanks must have the same class ID, and this name must begin with either ‘Blank’ or ‘blank’. In this example dataset, two different blanks were used, a light reagent blank and a heavy reagent blank. Due to reagent scarcity, these two blanks were both diluted by a factor of 2 to reach sample volume. To

account for this dilution, the injection volumes were entered such that the reagent blanks have half the injection volumes of the QCs and samples. The function of the Inject. volume column is to account for dilution variations across samples with the actual injection volume being potentially irrelevant. For your cohort samples, the Class ID field represents how you can group LC-MS(/MS) data. In this example, triplicate samplings of the five samples are grouped through the use of five sample Class IDs. Any sample Class ID is allowed so long as it does not satisfy any of the blank, light QC, heavy QC, or mix QC Class ID requirements discussed previously.

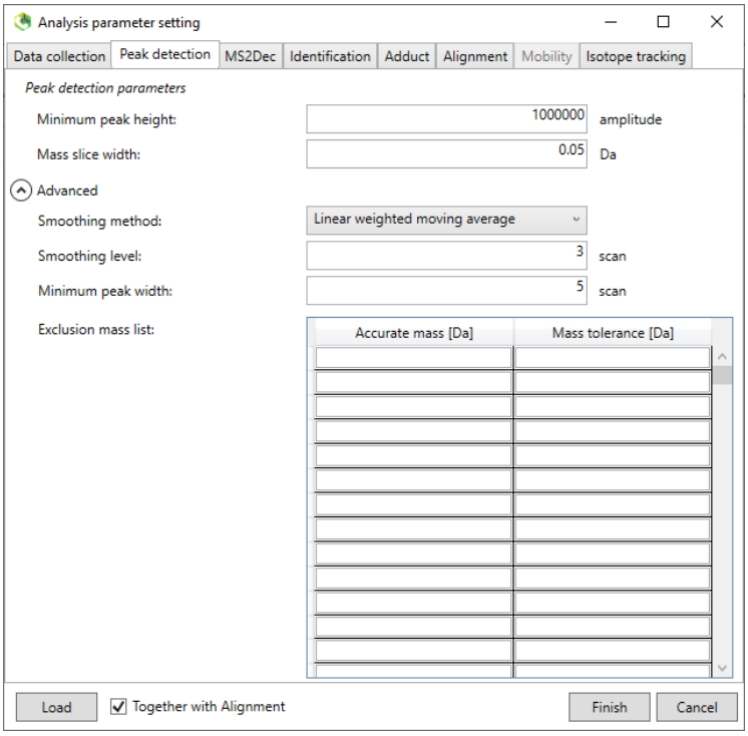
Figure 25. MS-DIAL Data Collection Parameters.



Mass tolerances, retention time windows, and mass windows must be carefully considered based on your LC-MS(/MS) system and the samples analyzed. For this workflow, the maximum charge number should be the maximum charge number in your library. The number of threads refers to the number of processors your computer will dedicate to processing

the data. More threads means faster processing but greater allocation of computer resources to MS-DIAL. Assigning too many threads can lead to freezing. The Load button can be used to load previously saved parameters. The parameter file for this example dataset is on Github. The parameter file will not automatically set your library file, alignment reference file, and isotope tracking parameters. To prevent lost time due to incorrect processing, all parameters should be double-checked even when loading a parameter file.

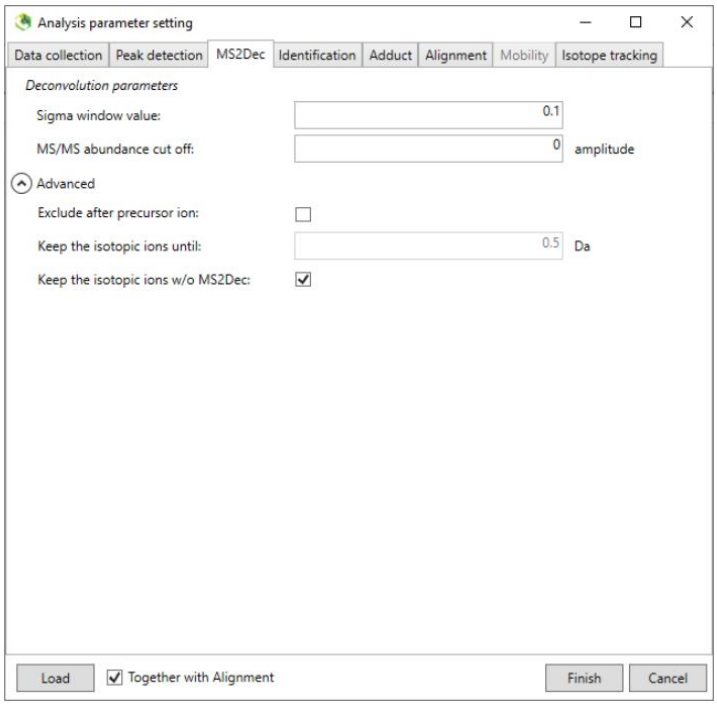
Figure 26. MS-DIAL Peak Detection Parameters.



The minimum peak height is an essential consideration in the PPP workflow. Peaks with intensities below this threshold will be eliminated from MS-DIAL's processing. Raising this value allows for faster data processing and reduces the probability that MS-DIAL will assign incorrect isotopic relationships between analytes and interfering peaks. Lowering this value increases the sensitivity of the workflow and widens the metabolome coverage. Regardless of

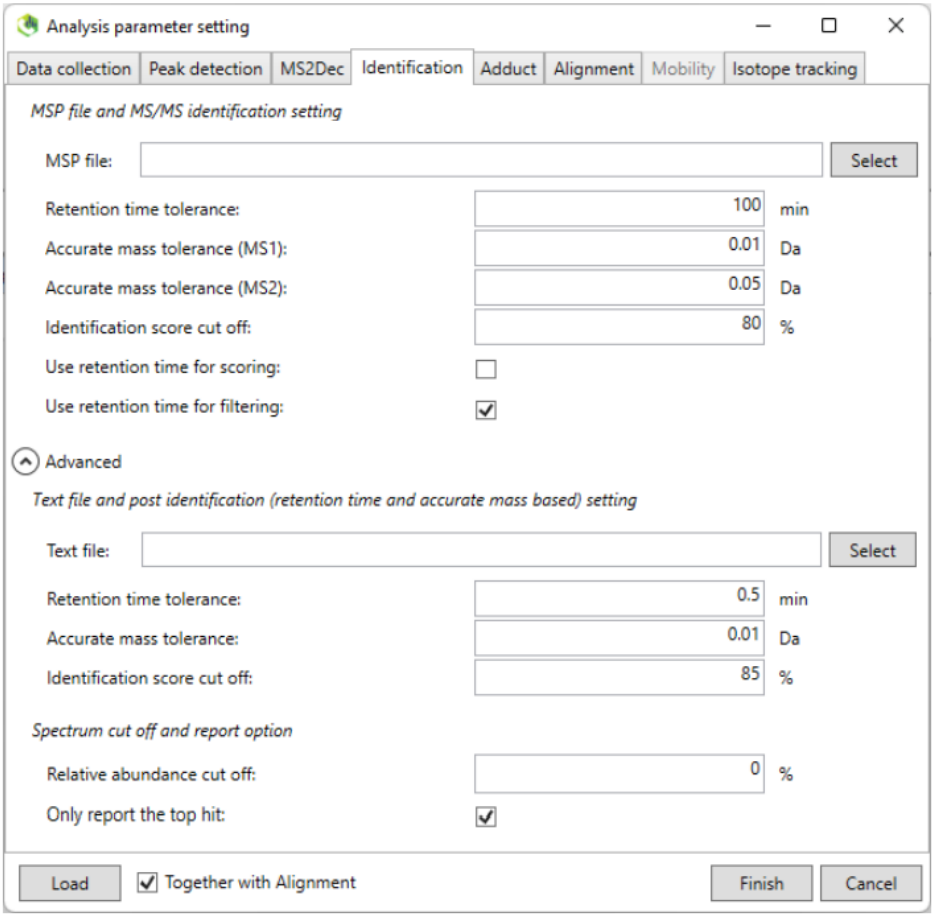
tagging scheme, peak pairing will only be correct for a given analyte if its light-tagged M+1 peak is above the minimum peak intensity. In the example dataset, the highest analyte intensities are in the low-to-mid 10^9 range, and the background peaks are in the mid-10^5 range, making 10^6 a reasonable minimum. An exclusion mass list can be included to potentially improve isotopic peak pairing with a lower minimum intensity by excluding major background peaks. The other parameters refer to MS-DIAL's peak detection and smoothing algorithms. Reducing the smoothing level to 2 may yield better results if you are struggling with partially overlapping isomeric peaks.

Figure 27. MS-DIAL MS2 Deconvolution Parameters.



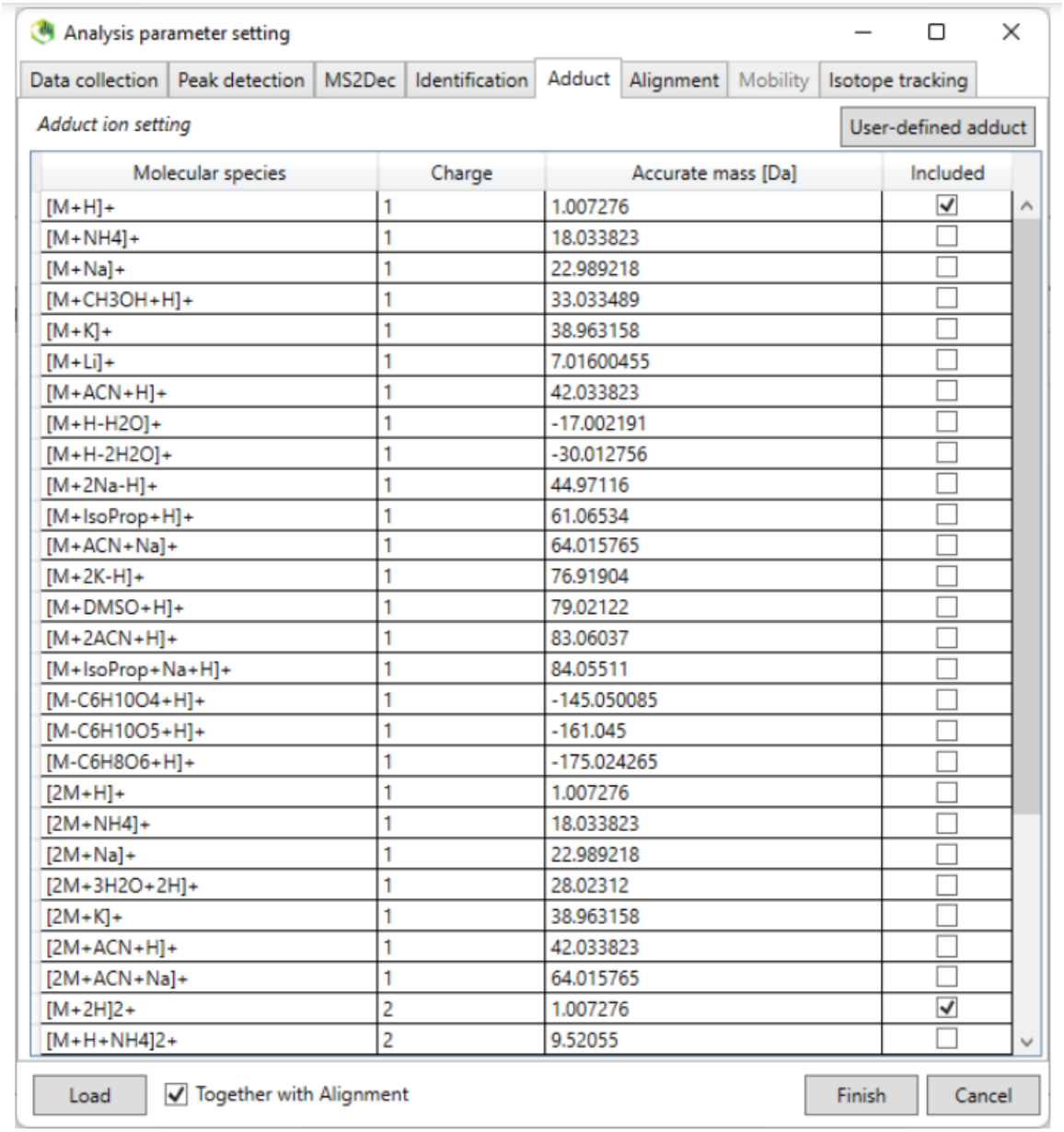
Regardless of whether or not your LC-MS methods include[s] tandem mass spectrometry, the Advanced MS2Deconvolution parameters must be assigned as above for downstream PPP analysis.

Figure 28. MS-DIAL Identification Parameters.



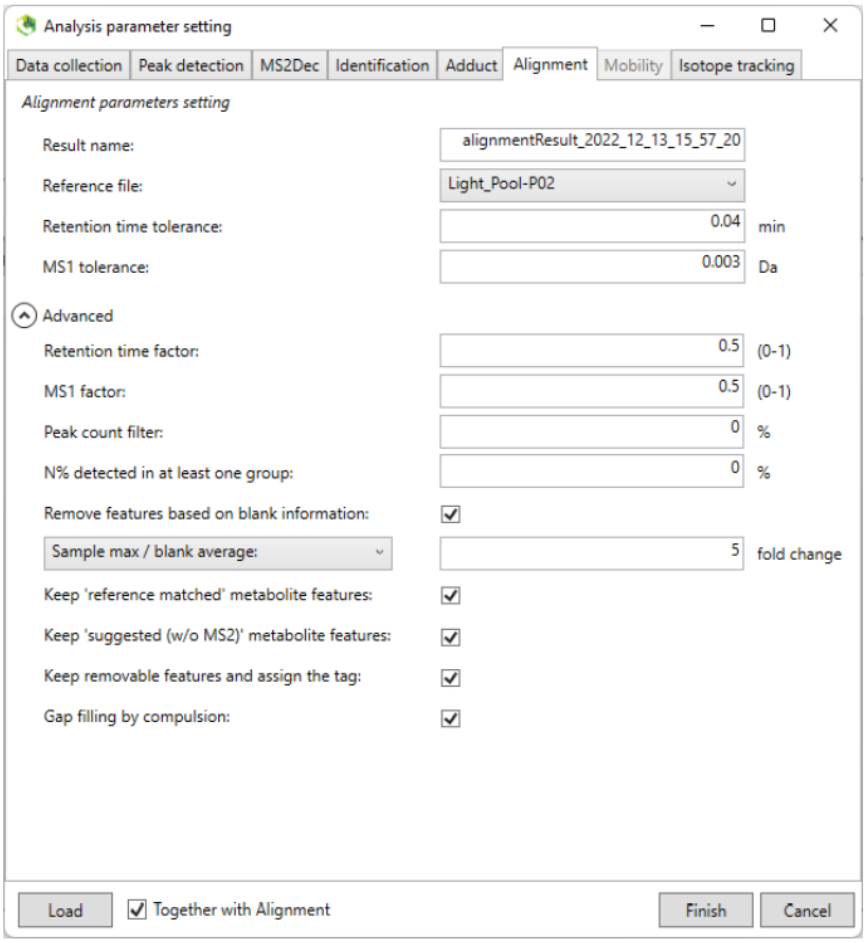
In this example, the library used is a text file library with MS1, RT, and adduct type information. MSP libraries can also be used containing fragmentation data for more confident identification.

Figure 29. MS-DIAL Adduct Parameters.



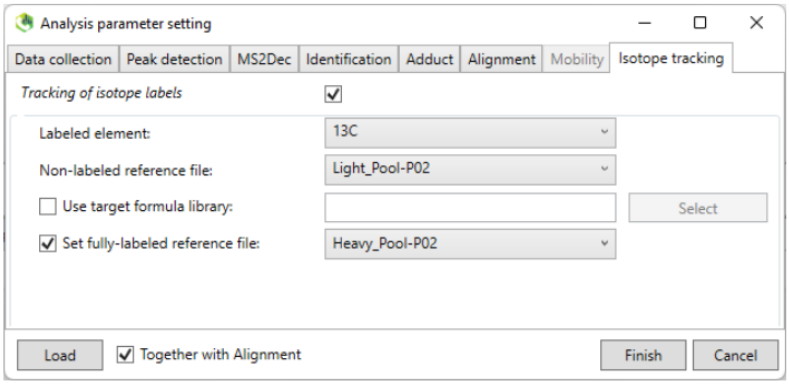
Select adducts relevant to your library file. In this example, the library contains [M+H]+ and [M+2H]2+ adduct modes.

Figure 30. MS-DIAL Alignment Parameters.



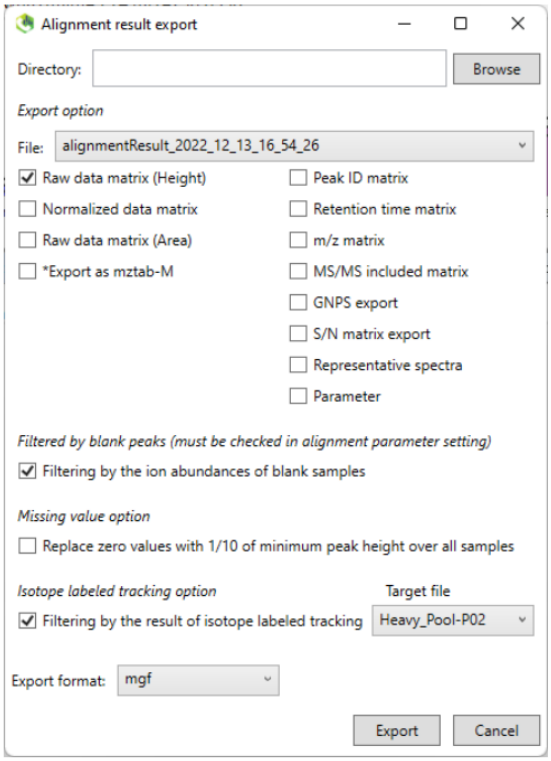
Several alignment parameters are essential for the PPP workflow. Your alignment reference file should be your light pool QC. The retention time tolerance and MS1 tolerance determine the windows within which peaks are aligned between your chromatograms. It is important to make sure your LC system is well equilibrated before and after each injection so that you can make the retention time tolerance small. High mass accuracy allows for a small MS1 tolerance. Reducing these tolerances, like raising the minimum ion intensity, decreases the probability that MS-DIAL will assign false isotopic relationships. Raising the minimum ion intensity and less complicated sample matrixes allows these two tolerances to be relaxed.

Figure 31. MS-DIAL Isotope Tracking Parameters.



Enable tracking of isotope labels and set the labeled element for your tagging scheme. Dansylation relies on carbon-13.[72] Your non-labeled reference file should be the light pool QC. Enable the fully-labeled reference file option and set the file as your heavy pool QC. Double-check all of your parameters and click Finish to begin processing.

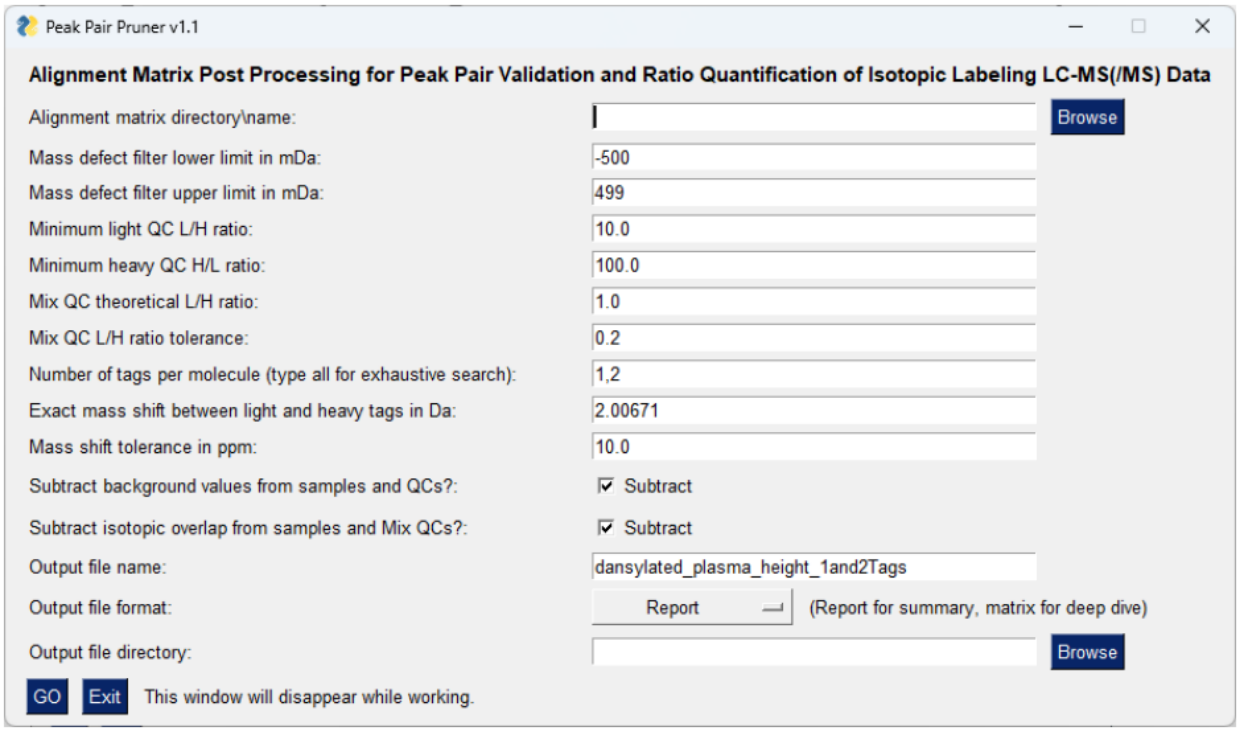
Figure 32. MS-DIAL Alignment Result Export.



Export your alignment results as a raw data matrix. You can export an average peak height raw data matrix or a peak area raw data matrix, but do not select both options for a single export file. Enable filtering by the ion abundances of the blank samples if you want MS-DIAL to eliminate background peaks, but note that it only eliminates or does not eliminate, does not quantitatively correct kept peaks that overlap with the background. Filter by the result of isotope labeled tracking in your heavy pool QC. In the human plasma experiment with FTMS data acquisition and virtually no RT shift between light and heavy peaks, we have had better results with average height than with peak area. Simultaneous acquisition of both light and heavy peaks by the orbitrap allows for peak height to supersede peak area for most accurate quantification. In some later work not included in the PPP paper, we have found that peak area is more useful outside this exceptional case.

PPP Validation and Correction of Peak Pairs

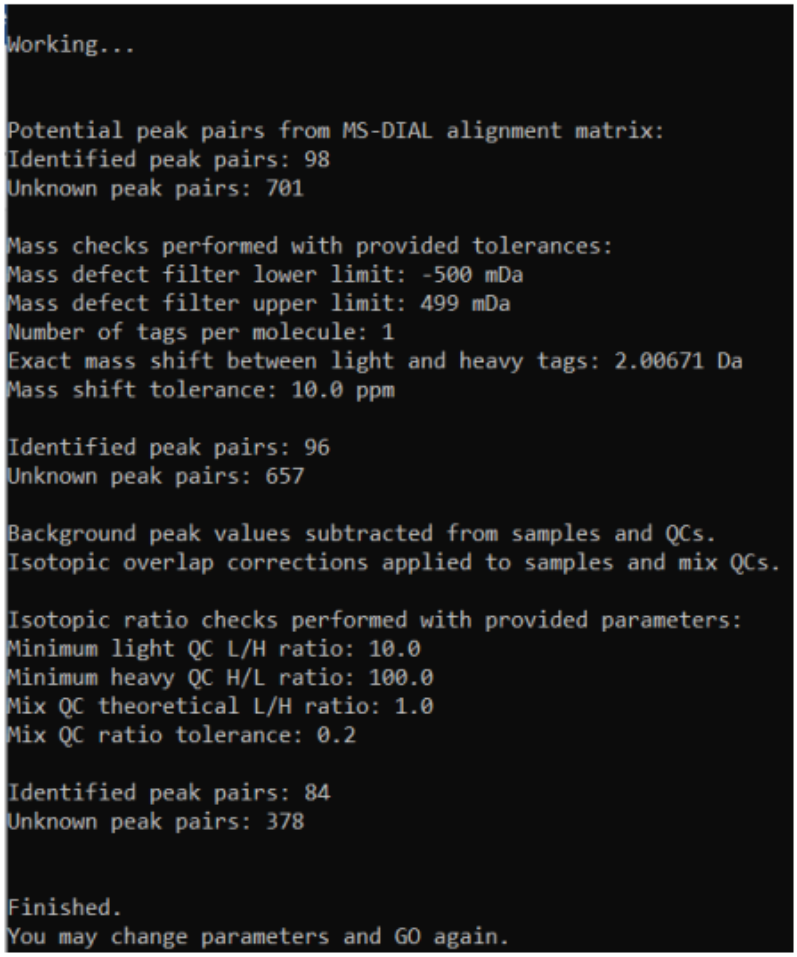
Figure 33. Peak Pair Pruner Primary GUI.



You can now use the example matrix on Github if skipping the MS-DIAL steps. Run the PPP .exe and two windows will appear. Keep both open. The parameter window is shown above with the example data values filled in. Select your MS-DIAL alignment matrix and select an output directory. Optional mass defect filtering can be applied based on your experiment. The current mass defect window of [-500 mDa, 499 mDa] is all inclusive and so effectively disables this filter. If you are analyzing a particular class of analyte, the mass defect window may be useful for focusing your results. Set a minimum acceptable L/H ratio in your light pool QC. This value should be large in your light pool QC as there are no heavy-tagged molecules. However, your light pool analytes will still have natural heavy isotope abundance which may overlap with your heavy tag's m/z values. Enter a minimum acceptable H/L ratio in your heavy

pool QC. Likewise, this value should be large but should not suffer from overlap like the light pool QC. Enter your theoretical mix ratio from your experimental design and acceptable tolerance for peak pairs. 1.0 and 0.2 for these values indicates that a peak pair with a L/H in the range [0.8,1.2] will pass validation by the mix pool QC. Enter the number of tags per analyte molecule. In PPP v1.1, multiple tagging levels can be queued at once. Entering '1,2' will scan for singly and doubly tagged molecules. Entering '2,4,6' would scan for molecules with 2, 4, or 6 tags. Alternatively, enter 'all' for a comprehensive scan of all potential tagging levels at the cost of longer processing time. For the example dataset or alignment matrix, you should run with 1 and 2 tags. Enter the exact mass shift between your light and heavy tags, 2.00671 Da for 2 carbon-13. Enter the mass shift tolerance for checking peak pair mass shifts. Optionally subtract background peak pair values based on the blank. Optionally subtract isotopic overlap (natural heavy abundance) based on the light pool QC. Enter a name for your file. Most applications will benefit from the Report output format, but there is a modified matrix format as well for troubleshooting.

Figure 34. Peak Pair Pruner Data Processing Window.



The second PPP window will show you the progress of data processing (usually very fast) and gives a breakdown of where peak pairs are being eliminated if they fail validations. In this example, 98 peak pairs from MS-DIAL had isotopic relationships conforming to 1 dansyl chloride tag (M+0 and M+2). Two were eliminated by the ppm mass shift check. Twelve were eliminated in the QC checks. These twelve turn out to be doubly tagged molecules (M+0 and M+4), which pass the validation checks when rerunning with 2 as the number of tags per molecule.

Troubleshooting

If you have dramatically fewer validated peak pairs after PPP processing than you expected or than you had identified in MS-DIAL, you should first check the PPP companion window (above). This will tell you where the peaks were eliminated. If the initial isotopic matches from MS-DIAL is too low, then there is a pairing issue in MS-DIAL, which could be related to your alignment tolerances being too strict or your minimum peak height being too high. It could also indicate a chromatographic problem. If there are numerous losses in the mass check steps, make sure you have correctly calculated the mass shift between your light and heavy tag with as many significant figures as are available. Check that the mass defect filtering employed is not too restrictive for your target analytes. If mass accuracy is low, you may need to relax the ppm tolerance. If losses are great in the QC ratio checks, there may be an experimental issue such as differing reactivities and yields between the light and heavy tags or evaporation leading to biased mixing/spiking. Relaxing all tolerances to extremes and looking at the Matrix output format can help diagnose the exact problem. Losses can also be caused by incorrect peak pairing in MS-DIAL, which is not easily diagnosable in PPP. Check the problematic peak pairs in MS-DIAL and assess whether MS-DIAL is paring them correctly or potentially not at all.

Conclusions

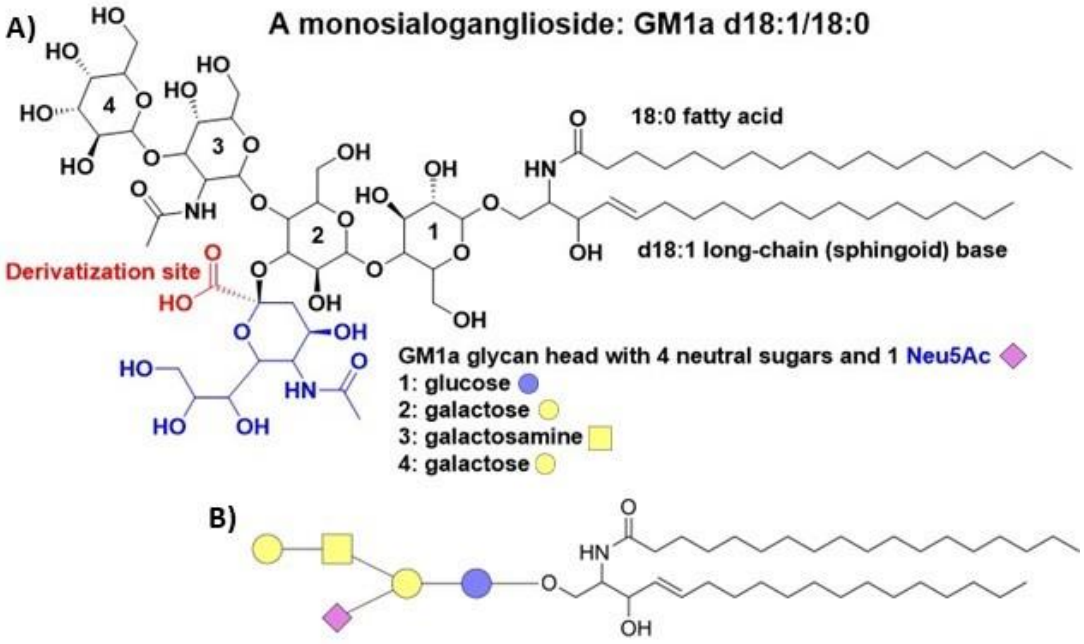
PPP may have a short lifecycle if the active team behind MS-DIAL incorporates its functionality, but for now, the MS-DIAL+PPP workflow represents one of the only free and user-friendly means of conducting isotopic labeling data processing. The manual work of conducting an isotopic labeling experiment is laborious and error prone, but robotic process automation has the potential to make this analytical approach a standard in comparative omics.

CHAPTER IV: HIGH THROUGHPUT LC-MS/MS BASED ANALYSIS OF GANGLIOSIDES USING ISOTOPIC LABELING

Introduction

Gangliosides are glycosphingolipids composed of a hydrophobic ceramide moiety connected by ether bond to a carbohydrate or 'glycan' moiety. The inclusion of one or more N-Acetylneuraminic acids (shorthand Neu5Ac or sialic acid) differentiates gangliosides from other glycosphingolipids. Figure 35 illustrates (A) the structure of a typical monosialoganglioside with naming conventions initially outlined by Svennerholm[13] and (B) glycan structural representation as popularized by glycated peptide research[62] and subsequently adapted for convenient discussion[20, 61] of gangliosides.

Figure 35. Monosialoganglioside GM1a Structure and Cartoon Glycan Representation.



In Figure 35, the amphiphilic nature of the ganglioside is plain to see, responsible for their insertion into the hydrophobic environment of the plasma membrane[47, 51, 74] as well as

their instrumental role in forming and maintaining lipid rafts[27, 32, 43] through hydrogen-bonding and other hydrophilic interactions. Lipid rafts and by extension gangliosides have been mechanistically linked to the progression of numerous pathologies including cancers[52], neurodegenerative disorders[20], and Type 1 Diabetes (T1D)[40]. This makes gangliosides promising biomarkers for tracking disease progression and candidates for exogenous therapies.

Although gangliosides can be analyzed directly using tandem mass spectrometry for their ceramides in the positive ion mode and for their glycans in the negative ion mode[17], chemical derivatization enhances[16] structural identification through improved fragmentation while also enabling more sophisticated analytical techniques such as stable-isotope labeling. Highlighted in red in Figure 35, the carboxylic acid of the ganglioside Neu5Ac residue is the 'thorn' preventing high quality glycan analysis in positive mode tandem mass spectrometry. This preferentially negative functional group leads to glycan fragmentation being most descriptive in the negative ion mode. With cholamine[63] derivatization of carboxylates in other lipids[75] successful at increasing positive mode ionization and with similar amide derivatization of gangliosides previously demonstrated[76], we sought and developed a methodology for cholamine derivatization of gangliosides with the MS-DIAL+PPP isotopic labeling workflow in mind.

Materials and Methods

Chemicals and Materials

11 ganglioside standards were purchased from Matreya for method development, summarized in Table 1. The ganglioside standards are mixtures purified from biological sources and often have a variety of lower abundance species[17]. The major species reported by Matreya for each standard is in the table.

Table 1. Matreya Ganglioside Standard Mixtures with Reported Major Species.

Ganglioside Standards	Matreya Catalog #	Major Mixture Species
Porcine Fuc-GM1a	1526	Fuc-GM1a d18:1/22:0
Bovine GM1a	1061	GM1a d18:1/18:0
Bovine GD1a	1062	GD1a d18:1/18:0
Bovine GD1b	1501	GD1b d18:1/18:0
Bovine GT1b	1063	GT1b d18:1/18:0
Bovine GQ1b	1516	GQ1b d18:1/18:0
Bovine GM2	1542	GM2 d18:1/18:0
Rabbit GD2	1527	GD2 d18:1/18:0
Bovine buttermilk GM3	1503	GM3 d18:1/23:0
Bovine buttermilk GD3	1504	GD3 d18:1/23:0
Chicken egg GM4	1535	GM4 d18:1/m22:0

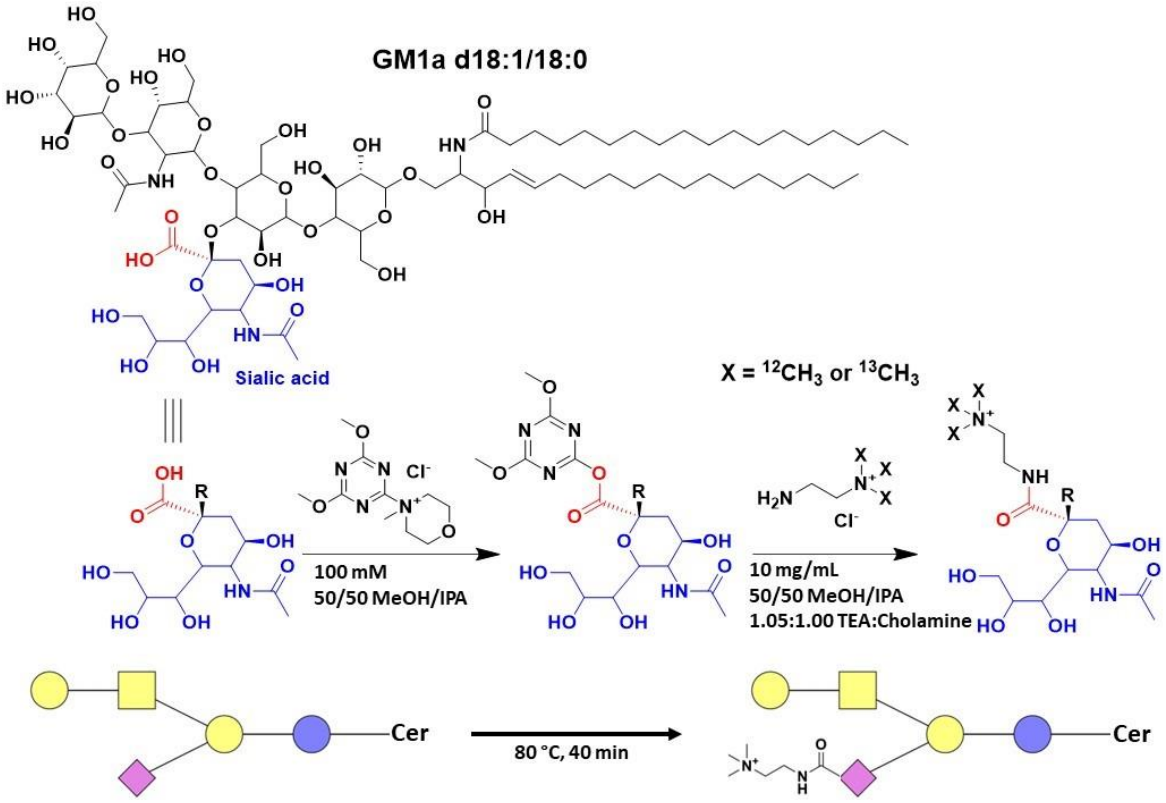
(2-Aminoethyl)trimethylammonium chloride hydrochloride (cholamine) was purchased from Sigma-Aldrich (SKU 284556-5G). Cholamine-$^{13}C_3$ was prepared as previously described in the literature.[75] 4-(4,6-Dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMTMM) was purchased from TCI (product number D2919). All other reagents and solvents were purchased from Thermo Fisher Scientific.

Sample Preparation

Ganglioside standards were initially dissolved in MeOH to form stock solutions. For cholamine derivatization, 5 nmol ganglioside was transferred to a screwcap plastic tube and dried. In addition to the following derivatization procedure, duplicate standard samples were reconstituted in LC-MS mobile phase and analyzed at equivalent concentration as a benchmark for assessment of ionization improvement through derivatization. Gangliosides were reconstituted with 50 uL 50/50 MeOH/IPA with 10 mg/mL cholamine for light isotopic labeling or cholamine-$^{13}C_3$ for heavy isotopic labeling and 1.05:1.00 molar ratio TEA:cholamine. 50 uL 50/50 MeOH with 100 mM DMTMM was added to catalyze the reaction as illustrated in Figure 36. The mixture was incubated at 80 °C with 600 rpm mixing for 40 min, then placed on ice to cool. Once below room temperature, 100 uL water was added to quench the reaction. Simple

derivatization samples were prepared for LC-MS/MS analysis without further manipulation. Isotopic labeling quantitative testing samples were made by mixing light-tagged samples and heavy-tagged samples at ratios of 1:10, 1:2, 1:1, 2:1, and 10:1, drying these under nitrogen and then reconstituting them in the initial gradient mobile phase. Drying under nitrogen was found to be necessary to avoid degradation and exchange of isotopic labeling tags.

Figure 36. Ganglioside Amide Derivatization with Cholamine and DMTMM.



Tandem Mass Spectrometry Analysis

A Thermo Fisher Scientific Vanquish Horizon UHPLC System coupled with a Thermo Fisher Scientific Orbitrap Exploris 240 was used to conduct LC-(+)FTMS/MS analysis. A Phenomenex Kinetex HPLC 1.7 µm C18 100Å (100 x 2.1 mm) column was used along with the following mobile phases: (A) 0.1% formic acid and 10 mM ammonium formate in 60/40 ACN/H_2O and (B) 0.1% formic acid and 10mM ammonium formate in 88/10/2 IPA/ACN/H_2O.

The gradient was 30% B (- 1 min), 30% B (0 min), 99% B (13 min), 99% B (14 min), 30% B (14.1 min), 30% B (15 min). The column was maintained at 40 °C with flowrate 350 µL/min. The injection volume was 5.00 µL. Ionization for FTMS analysis utilized a HESI source in the positive mode with spray voltage 3 kV, capillary temperature 350 °C, sheath gas flowrate 20 au, aux gas flowrate 5 au, and probe heater temperature 400 °C. FTMS/MS scanning was full MS1 in the range 500 to 2000 m/z with maximum IT 100 ms and mass resolution 120,000, followed by 10 data dependent MS2 scans selecting the top 10 precursor ions with a 2+ charge state with isolation window 1 m/z, stepped normalized HCD collision energies 30 followed by 35 followed by 40, mass resolution 15,000, maximum IT 50 ms, and MS2 spectrum first mass 120 m/z.

MS-DIAL Identification and Isotopic Peak Pairing

MS-DIAL 4.92 for Windows x64 was used alongside an in-house ganglioside derivatization library. For the derivatized standards, MS-DIAL data collection proceeded with MS1 tolerance 0.005 Da and maximum charge number 2. The minimum peak height was set to 10^4 with mass slice width 0.05 Da. The peak smoothing method was linear weighted moving average with smoothing level of 3 scans and minimum peak width of 5 scans. The Sigma window value was set to 0.1 without exclusion after precursor ions and with isotopic ions w/o MS2Dec preserved. Automated identification was performed with a MSP library with retention time tolerance 0.3 min, accurate mass tolerance (MS1) 0.005 Da, accurate mass tolerance (MS2) 0.01 Da, identification score cut off 70%, and retention time used for filtering but not scoring.

For the quantitation validation cohort, the same parameters were used along with peak alignment in reference to a light-tagged sample, retention time tolerance 0.1 min, MS1 tolerance 0.004 Da, and linear extrapolation RT correction by MS-DIAL. Isotope tracking was enabled

with ^{13}C as the labeled element, the light-tagged sample as the non-labeled reference, and the heavy-tagged sample as the fully-labeled reference.

PPP Peak Pair Validation and Quantitation

Relative quantitation between light and heavy peaks in the isotopic labeling validation cohort was carried out through Peak Pair Pruner v1.1 post processing of the MS-DIAL alignment matrix output. PPP proceeded with lower mass defect limit -500 mDa, upper mass defect limit 499 mDa, minimum light QC L/H ratio 1.5, minimum heavy QC L/H ratio 10.0, mix QC theoretical L/H ratio 1.0, mix QC L/H ratio tolerance 0.2, 1 tag per molecule, mass shift between heavy and light tags 3.010065 Da, mass shift tolerance 10.0 ppm, subtraction of background values enabled, and subtraction of isotopic overlap enabled.

Results and Discussion

Upcoming Figures 37 through 45, 47 and 48, and Tables 2 through 12 illustrate the results of cholamine derivatization of the 11 Matreya standards (in order: Fuc-GM1a, GM1a, GM2, GM3, GM4, GD3, GD2, GD1b, GD1a, GT1b, GQ1b). Each Figure has (A) the structure and LC-MS peak for the major species reported by Matreya prior to cholamine derivatization, (B) the structure and LC-MS peak for that same species after cholamine derivatization, and (C) the annotated MS2 spectrum of the derivatized structure as used for library building and subsequent wider annotation by MS-DIAL. Each figure is followed by a table describing the species identified in MS-DIAL using an inhouse MSP library with measured RT (min), relevant monoisotopic m/z (2+ ions for all except GM4 with 1+), species name, peak area in the MS1 chromatogram, and coeluting ceramide isomers. Coeluting ceramide isomers were detected with a small post processing script checking for two characteristic, matching ceramide LCB fragment ions as previously described in the literature[15, 16] and illustrated in the Figures as O'' and O'.

Fucosylated GM1a

Figure 37. Fuc-GM1a d18:1/22:0 Cholamine Derivatization and Fragmentation.

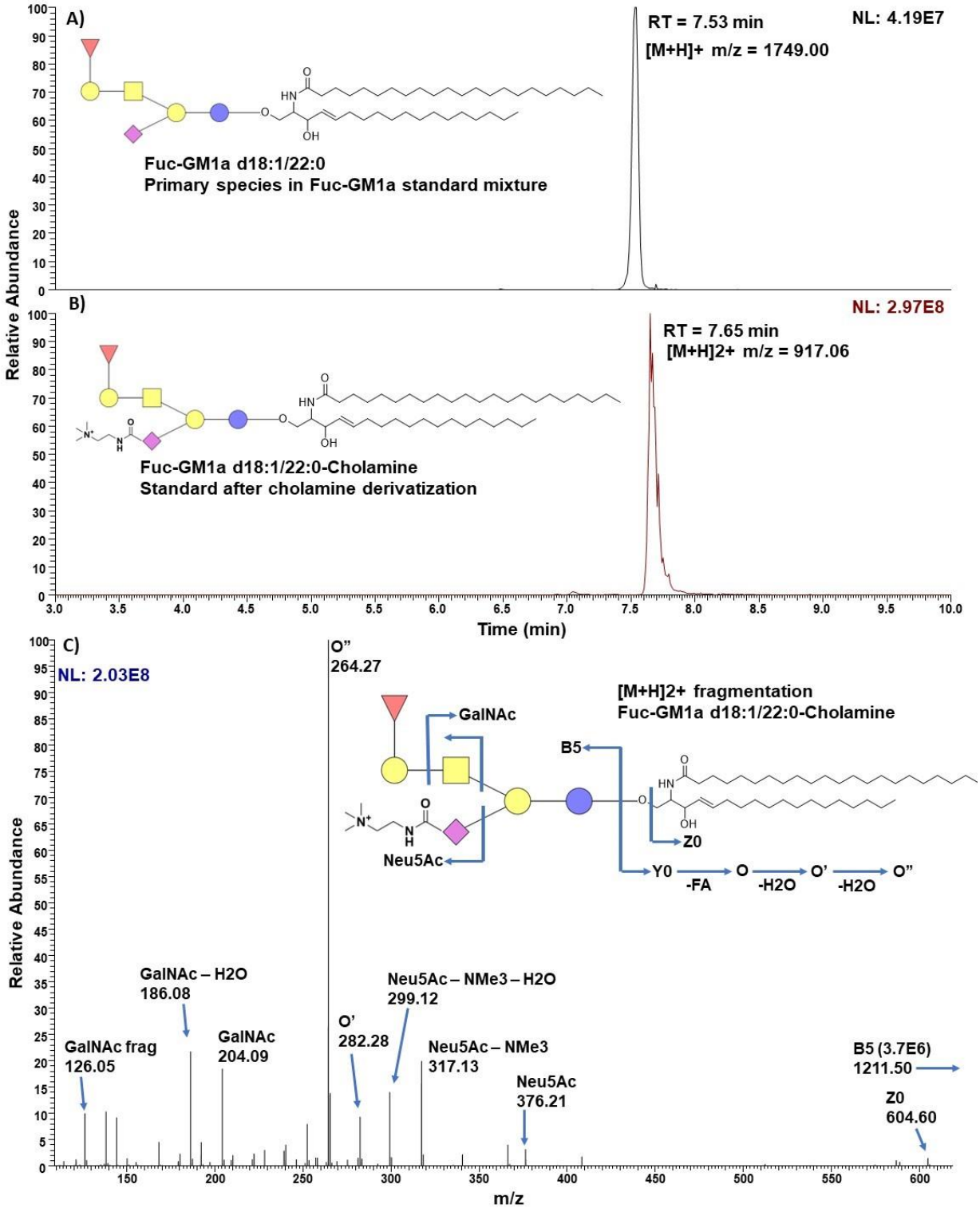


Table 2. Matreya Standard Mixture Fuc-GM1a Gangliosides Derivatized and Identified.

RT	m/z	Gangliosides	Peak Area	Coeluting Isomers
7.66	917.057	Fuc-GM1a d18:1/22:0	3.28E+09	
8.29	931.072	Fuc-GM1a d18:1/24:0	1.93E+09	
6.94	903.042	Fuc-GM1a d18:1/20:0	1.44E+09	
5.19	875.013	Fuc-GM1a d18:1/16:0	5.40E+08	
7.61	930.066	Fuc-GM1a d18:1/24:1	3.64E+08	
7.89	918.067	Fuc-GM1a d18:0/22:0	3.80E+08	
7.99	924.068	Fuc-GM1a d18:1/23:0	2.92E+08	d19:1/22:0
7.20	904.051	Fuc-GM1a d18:0/20:0	2.39E+08	
6.12	889.029	Fuc-GM1a d18:1/18:0	1.25E+08	d16:1/20:0
7.31	910.051	Fuc-GM1a d18:1/21:0	1.26E+08	d17:1/22:0
8.49	932.080	Fuc-GM1a d18:0/24:0	1.23E+08	d20:0/22:0
5.54	876.021	Fuc-GM1a d18:0/16:0	7.43E+07	
8.85	945.092	Fuc-GM1a d20:1/24:0	7.17E+07	d22:1/22:0
6.55	896.037	Fuc-GM1a d17:1/20:0	3.22E+07	d18:1/19:0
8.59	938.083	Fuc-GM1a d18:1/25:0	3.40E+07	d19:1/24:0, d20:1/23:0
6.42	890.036	Fuc-GM1a d18:0/18:0	1.64E+07	d16:0/20:0
5.67	882.021	Fuc-GM1a d18:1/17:0	9.06E+06	
6.82	897.043	Fuc-GM1a d17:0/20:0	3.57E+06	d18:0/19:0

The ability to analyze fucosylated gangliosides in this methodology is a great asset as previous stable-isotope labeling strategies based on periodate oxidation[16] cannot easily be applied to fucose-containing glycans due to fucose destruction by periodate[77]. For Fuc-GM1a d18:1/22:0, there is a large increase in ionization intensity (4E7 to 3E8) in derivatization along with a shift toward a doubly charged ion. This shift is useful for LC-MS/MS analysis, as it allows selection of just 2+ ions for MS2, ignoring the 1+ ions which make up the majority of background peaks. The resulting fragmentation spectrum is very descriptive, containing the intact ceramide, the intact glycan, fragments for diagnosis of ceramide LCB content explicitly and FA content implicitly, and diagnostic glycan fragments. The greater analysis by MS-DIAL in Table 2 agrees nicely with Matreya's product manifest while providing greater detail of low abundance species.

GM1a

Figure 38. GM1a d18:1/18:0 Cholamine Derivatization and Fragmentation.

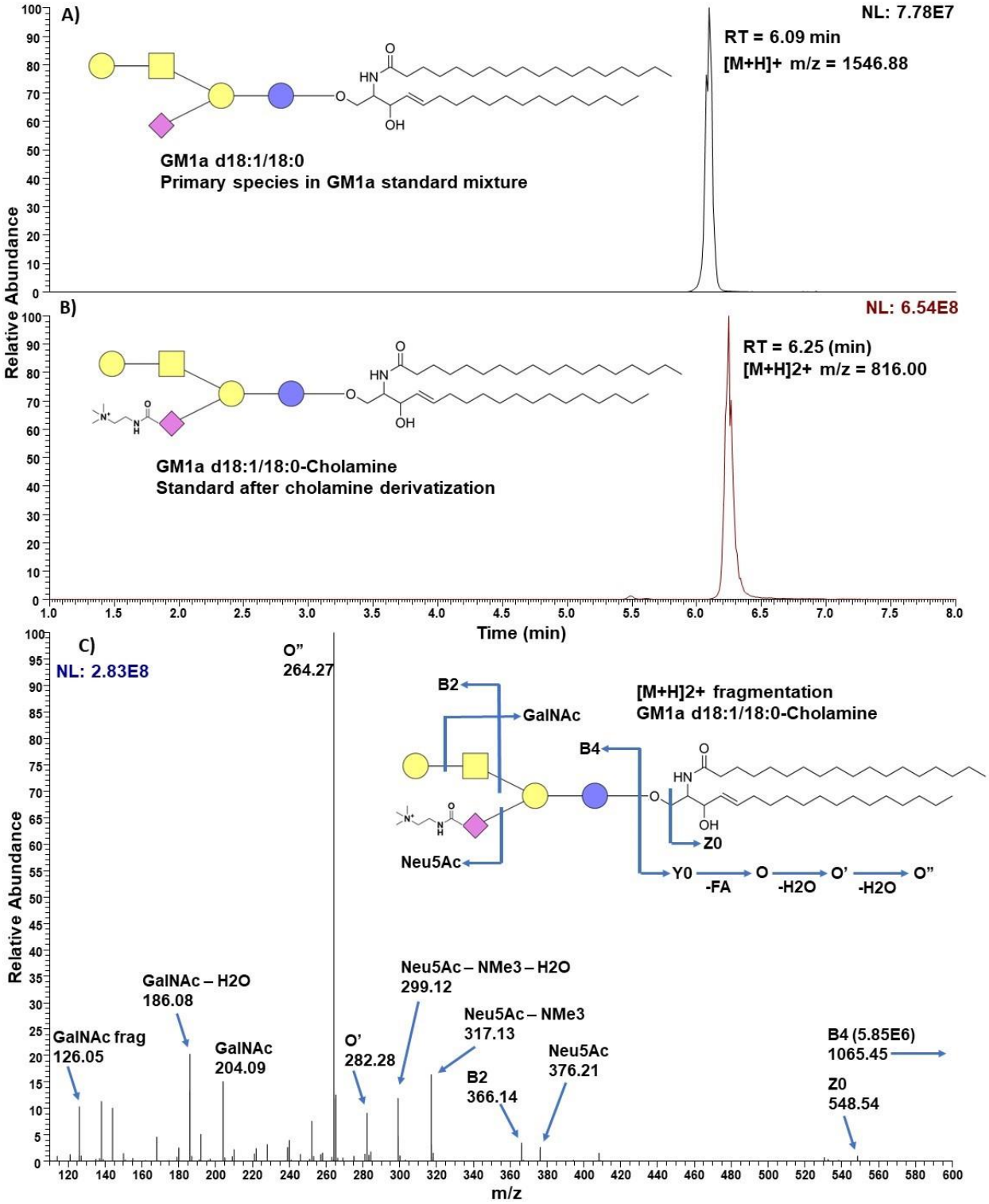


Table 3. Matreya Standard Mixture GM1a Gangliosides Derivatized and Identified.

RT	m/z	Gangliosides	Peak Area	Coeluting Isomers
7.05	830.0141	GM1a d20:1/18:0	7.22E+09	
6.25	815.9990	GM1a d18:1/18:0	6.95E+09	
7.31	831.0215	GM1a d20:0/18:0	6.82E+08	
6.55	817.0061	GM1a d18:0/18:0	4.90E+08	
6.66	823.0056	GM1a d19:1/18:0	3.06E+08	d18:0/19:1, d18:1/19:0
7.76	844.0300	GM1a d20:1/20:0	1.92E+08	d18:1/22:0
5.49	814.9921	GM1a d18:2/18:0	1.60E+08	
5.35	801.9820	GM1a d16:1/18:0	1.32E+08	d18:1/16:0
7.41	837.0201	GM1a d20:1/19:0	5.19E+07	d18:1/21:0, d21:1/18:0
5.60	814.9918	GM1a d18:2/18:0	5.26E+07	
8.40	858.0468	GM1a d18:1/24:0	4.41E+07	
5.81	808.9907	GM1a d18:1/17:0	2.83E+07	d17:0/18:1, d17:1/18:0
6.95	824.0135	GM1a d19:0/18:0	1.37E+07	d18:0/19:0, d20:0/17:0
5.68	802.9917	GM1a d16:0/18:0	9.03E+06	d18:0/16:0
7.99	845.0378	GM1a d20:0/20:0	8.22E+06	d18:0/22:0
5.92	821.9979	GM1a d19:2/18:0	4.80E+06	

As with Fuc-GM1a, there is a large derivatization improvement in ionization for GM1a and a shift toward the 2+ ion alongside a descriptive fragmentation spectrum. In the MS-DIAL output, these results also mostly agree with Matreya's analysis. However, GM1a d20:1/18:0 is the highest abundance species and not GM1a d18:1/18:0 as suggested in the product manifest. Matreya describes the mixture as predominately LCB d18:1 but with variation and with predominately FA 18:0, without analysis of molecular species, and so this highest abundance species is consistent with the limitations of their method.

GM2

Figure 39. GM2 d18:1/18:0 Cholamine Derivatization and Fragmentation.

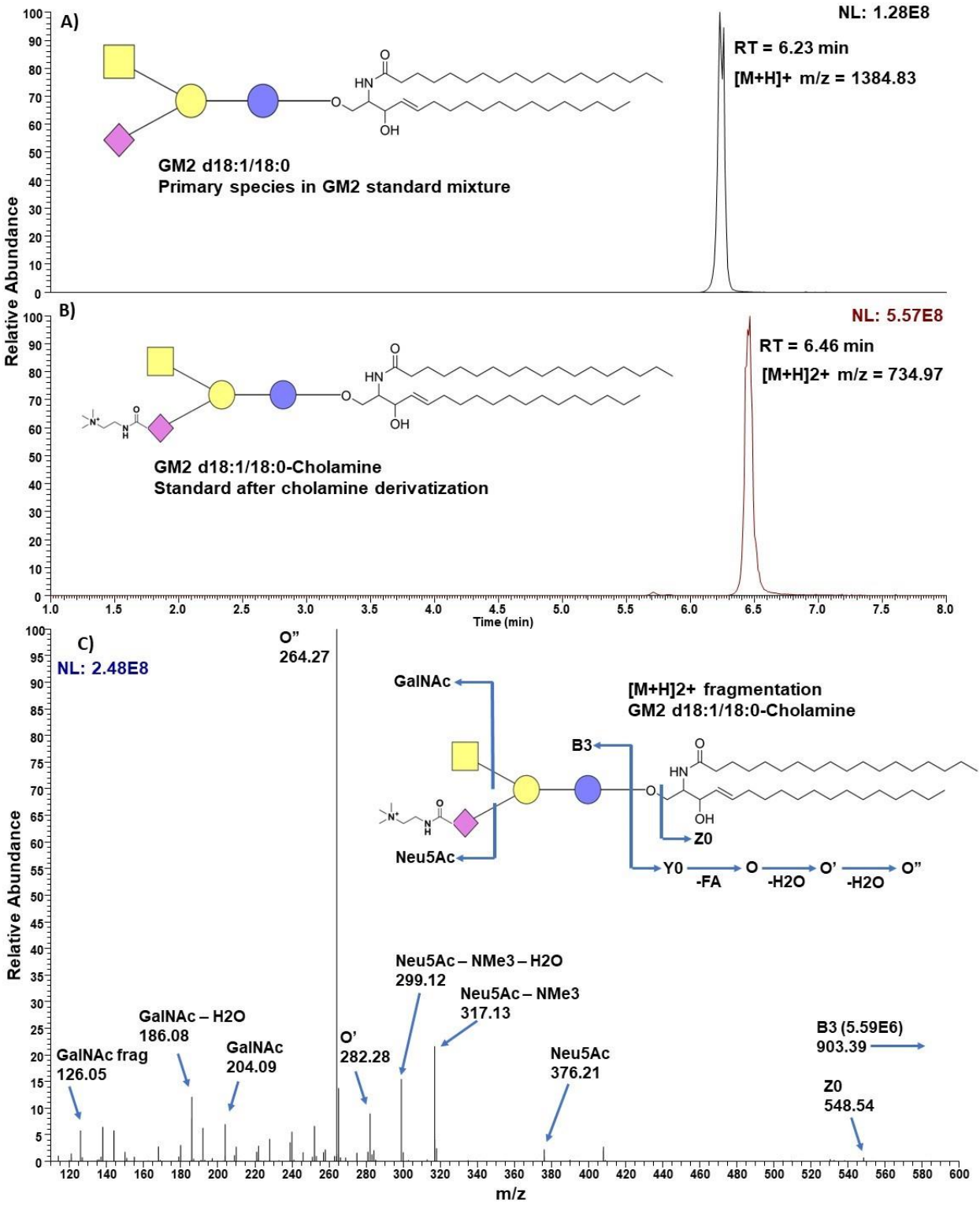


Table 4. Matreya Standard Mixture GM2 Gangliosides Derivatized and Identified.

RT	m/z	Gangliosides	Peak Area	Coeluting Isomers
6.46	734.9717	GM2 d18:1/18:0	6.77E+09	d18:0/18:1
7.23	748.9871	GM2 d20:1/18:0	5.81E+09	
7.50	749.9946	GM2 d20:0/18:0	5.30E+08	
6.75	735.9805	GM2 d18:0/18:0	3.99E+08	
6.86	741.9781	GM2 d19:1/18:0	2.43E+08	d18:0/19:1, d18:1/19:0
7.94	763.0021	GM2 d20:1/20:0	1.61E+08	d18:1/22:0
5.71	733.9644	GM2 d18:2/18:0	1.76E+08	
5.58	720.9561	GM2 d16:1/18:0	1.22E+08	d18:1/16:0
5.82	733.9644	GM2 d18:2/18:0	5.79E+07	
7.60	755.9940	GM2 d20:1/19:0	3.98E+07	d18:1/21:0, d19:0/20:1, d19:1/20:0, d20:0/19:1
6.02	727.9629	GM2 d18:1/17:0	3.62E+07	d17:0/18:1, d17:1/18:0
8.56	777.0185	GM2 d18:1/24:0	3.24E+07	d20:1/22:0
7.58	769.0039	GM2 d18:1/23:1	1.57E+07	
7.14	742.9882	GM2 d19:0/18:0	1.09E+07	d20:0/17:0
5.89	721.9650	GM2 d16:0/18:0	8.62E+06	d18:0/16:0
6.13	740.9727	GM2 d19:2/18:0	5.13E+06	

Like the first two monosialogangliosides, GM2 experiences large ionization improvement and a shift toward the 2+ ion. However, this shift is less pronounced, as species with smaller glycans tend toward single charge states. The fragmentation spectrum is still very descriptive, enabling confident identifications. The results of MS-DIAL analysis agree nicely with the Matreya manifest while as before providing more comprehensive description of lower abundance species.

GM3

Figure 40. GM3 d18:1/23:0 Cholamine Derivatization and Fragmentation.

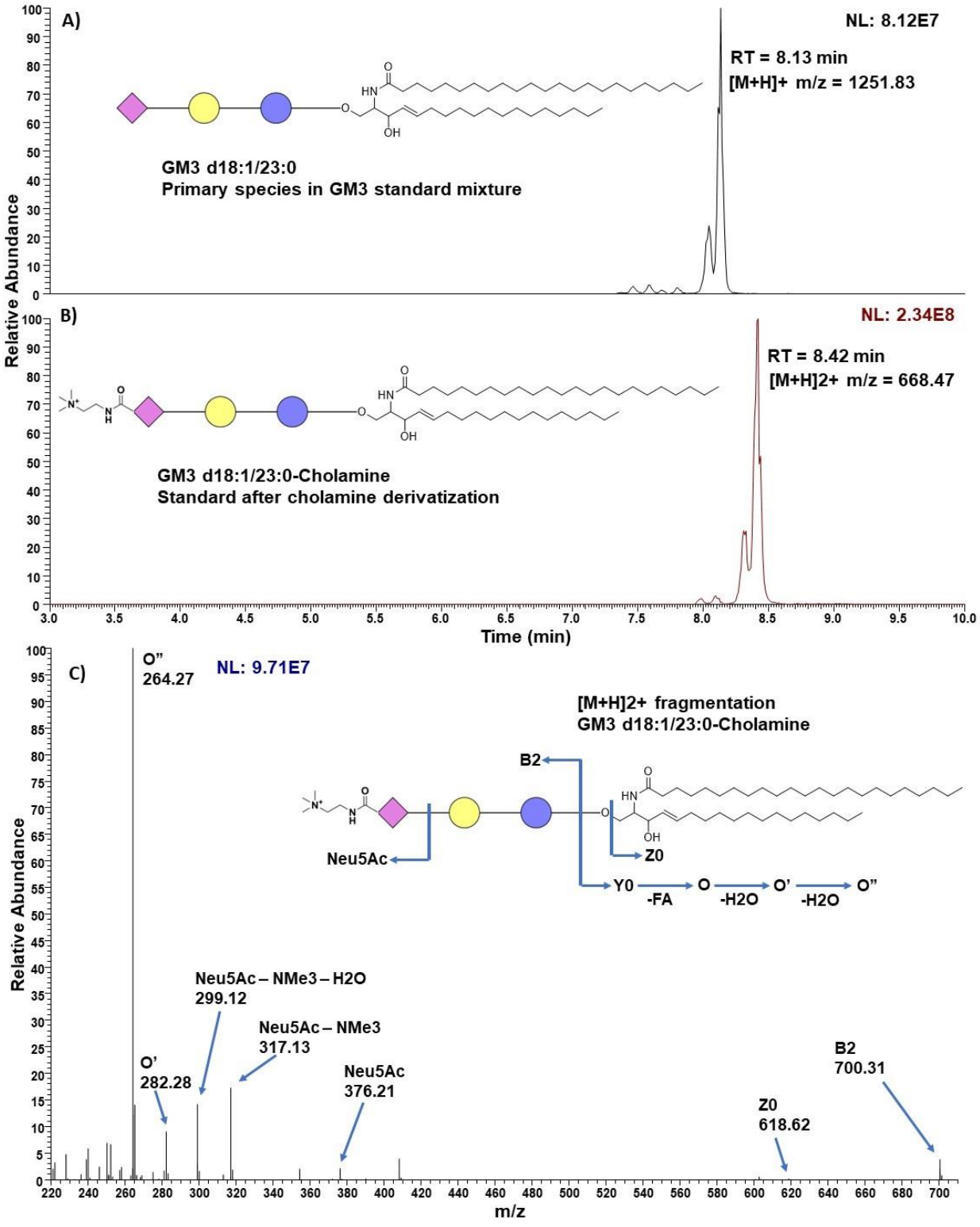


Table 5. Matreya Standard Mixture GM3 Gangliosides Derivatized and Identified.

RT	m/z	Gangliosides	Peak Area	Coeluting Isomers
8.41	668.4714	GM3 d18:1/23:0	3.47E+09	d19:1/22:0
8.10	661.4646	GM3 d18:1/22:0	2.58E+09	d16:1/24:0, d17:1/23:0
8.71	675.4797	GM3 d18:1/24:0	1.87E+09	d19:1/23:0
7.79	654.4549	GM3 d16:1/23:0	1.47E+09	d17:1/22:0
8.02	655.4628	GM3 d16:0/23:0	9.49E+08	d17:0/22:0
7.68	648.4572	GM3 d16:0/22:0	9.06E+08	
8.33	662.4703	GM3 d18:0/22:0	8.12E+08	d16:0/24:0, d17:0/23:0
7.43	647.4476	GM3 d16:1/22:0	7.63E+08	
7.73	667.4629	GM3 d18:1/23:1	6.77E+08	d16:1/25:1, d17:1/24:1, d18:2/23:0
5.76	619.4166	GM3 d18:1/16:0	6.54E+08	
8.62	669.4796	GM3 d18:0/23:0	3.58E+08	d17:0/24:0
8.06	674.4717	GM3 d18:1/24:1	3.17E+08	
8.88	682.4877	GM3 d19:1/24:0	2.90E+08	d18:1/25:0, d20:1/23:0
7.40	660.4556	GM3 d16:1/24:1	2.41E+08	d17:1/23:1, d18:1/22:1, d18:2/22:0
6.09	620.4257	GM3 d18:0/16:0	2.26E+08	d16:0/18:0
7.17	653.4487	GM3 d16:1/23:1	1.51E+08	
8.90	676.4866	GM3 d18:0/24:0	1.38E+08	d20:0/22:0
8.98	682.4877	GM3 d18:1/25:0	1.21E+08	
6.64	633.4324	GM3 d18:1/18:0	1.08E+08	d16:1/20:0
7.32	641.4478	GM3 d16:0/21:0	9.11E+07	d14:0/23:0, d17:0/20:0
6.94	634.4397	GM3 d18:0/18:0	7.05E+07	d14:0/22:0, d16:0/20:0
5.14	606.4095	GM3 d16:0/16:0	5.50E+07	d18:0/14:0
7.06	640.4413	GM3 d16:1/21:0	4.46E+07	d17:1/20:0
9.23	689.4943	GM3 d18:1/26:0	4.45E+07	d20:1/24:0
6.06	626.4246	GM3 d19:1/16:0	3.78E+07	d17:1/18:0
5.28	612.4093	GM3 d17:1/16:0	2.78E+07	d18:1/15:0, d16:0/17:1
4.80	605.4022	GM3 d16:1/16:0	2.64E+07	d18:1/14:0
6.61	627.4318	GM3 d12:0/23:0	2.43E+07	
5.64	613.4178	GM3 d17:0/16:0	1.69E+07	d16:0/17:0
9.16	683.4944	GM3 d20:0/23:0	1.60E+07	d18:0/25:0
6.80	646.4401	GM3 d16:1/22:1	1.20E+07	d18:1/20:1
5.98	632.4257	GM3 d18:1/18:1	1.09E+07	

The GM3 standard mixture is very diverse. Still, there is good agreement with Matreya's manifest. The shoulder in the chromatogram is the M+2 isotope of a species with 1 more ceramide double bond (- 2 Da) with sufficiently different RT to resolve the peaks in MS-DIAL.

GM4

Figure 41. GM4 d18:1/m22:0 Cholamine Derivatization and Fragmentation.

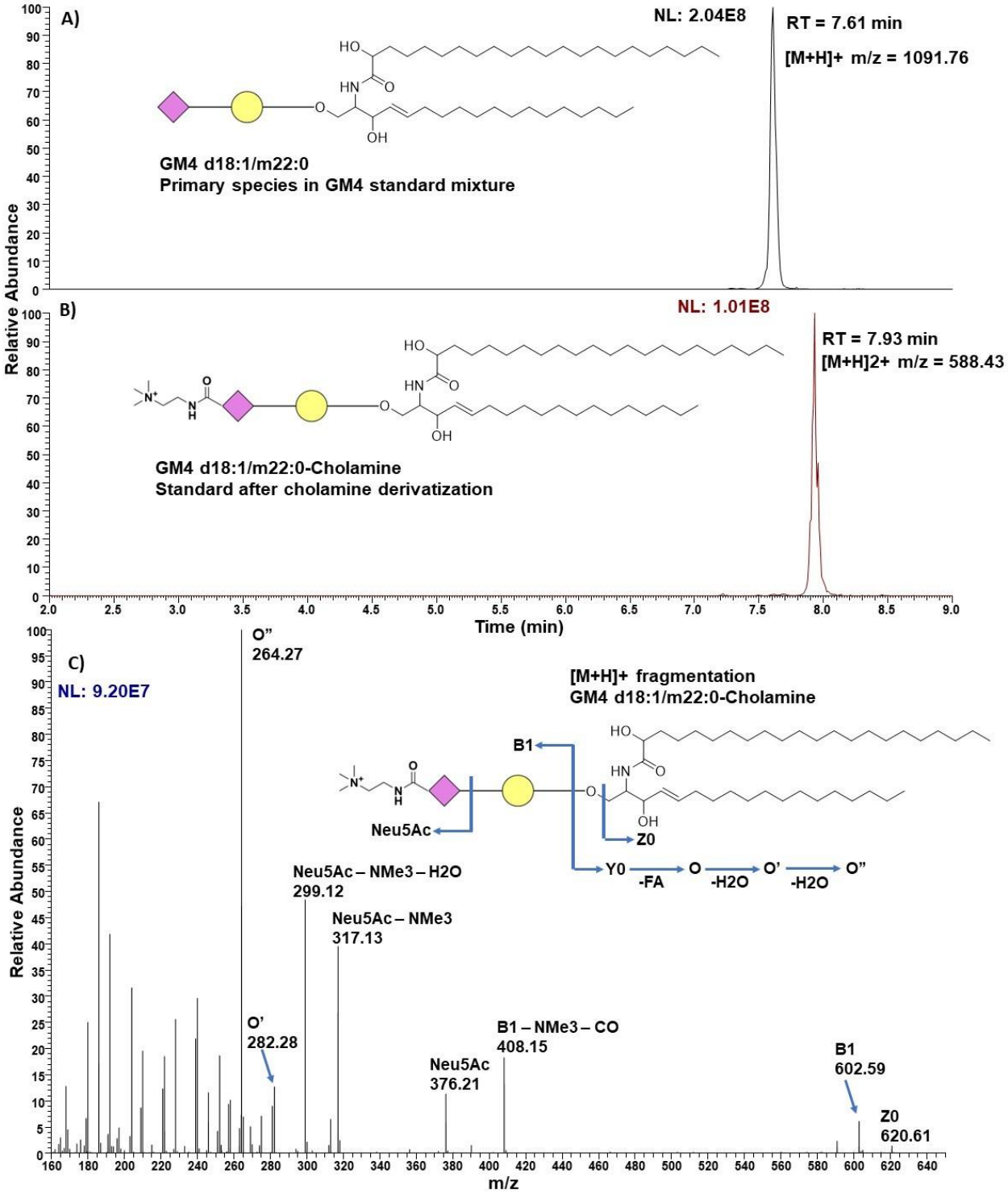


Table 6. Matreya Standard Mixture GM4 Gangliosides Derivatized and Identified.

RT	m/z	Gangliosides	Peak Area
7.93	1175.862	GM4 d18:1/m22:0	3.20E+09
8.56	1203.893	GM4 d18:1/m24:0	2.28E+09
8.25	1189.881	GM4 d18:1/m23:0	1.73E+09
5.68	1075.774	GM4 d18:1/16:0	5.42E+08
5.51	1091.771	GM4 d18:1/m16:0	2.96E+08
8.05	1159.868	GM4 d18:1/22:0	2.71E+08
7.23	1147.829	GM4 d18:1/m20:0	2.63E+08
7.59	1161.848	GM4 d18:1/m21:0	1.37E+08
6.03	1077.788	GM4 d18:0/16:0	1.12E+08
8.85	1217.908	GM4 d18:1/m25:0	1.03E+08
8.36	1173.881	GM4 d18:1/23:0	9.33E+07
8.49	1191.893	GM4 d18:0/m23:0	8.18E+07
8.68	1187.897	GM4 d18:1/24:0	7.32E+07
5.86	1093.786	GM4 d18:0/m16:0	6.40E+07
7.36	1131.833	GM4 d18:1/20:0	3.54E+07
8.78	1205.907	GM4 d18:0/m24:0	2.97E+07
6.58	1103.804	GM4 d18:1/18:0	2.20E+07
7.50	1149.847	GM4 d18:0/m20:0	1.14E+07
6.13	1089.790	GM4 d18:1/17:0	3.07E+06
6.83	1133.817	GM4 d18:1/m19:0	2.26E+06

This last monosialoganglioside, GM4, is the smallest and so suffers most from the smaller glycan to smaller charge state trend. The 2+ ion is appreciable, but fragmentation was repeated with 1+ ions to get better coverage. The GM4 standard mixture is unique for having fatty acids with hydroxyl moieties. The presence of fatty acid hydroxyls can be discerned by mass difference between the Z0 and O'/O'' fragments. The exact carbon address of the fatty acid hydroxyls is not diagnosable by the fragmentation spectrum. The figure depicts FA 22:0 2-OH because this is what is reported by Matreya.

GD3

Figure 42. GD3 d18:1/23:0 Cholamine Derivatization and Fragmentation.

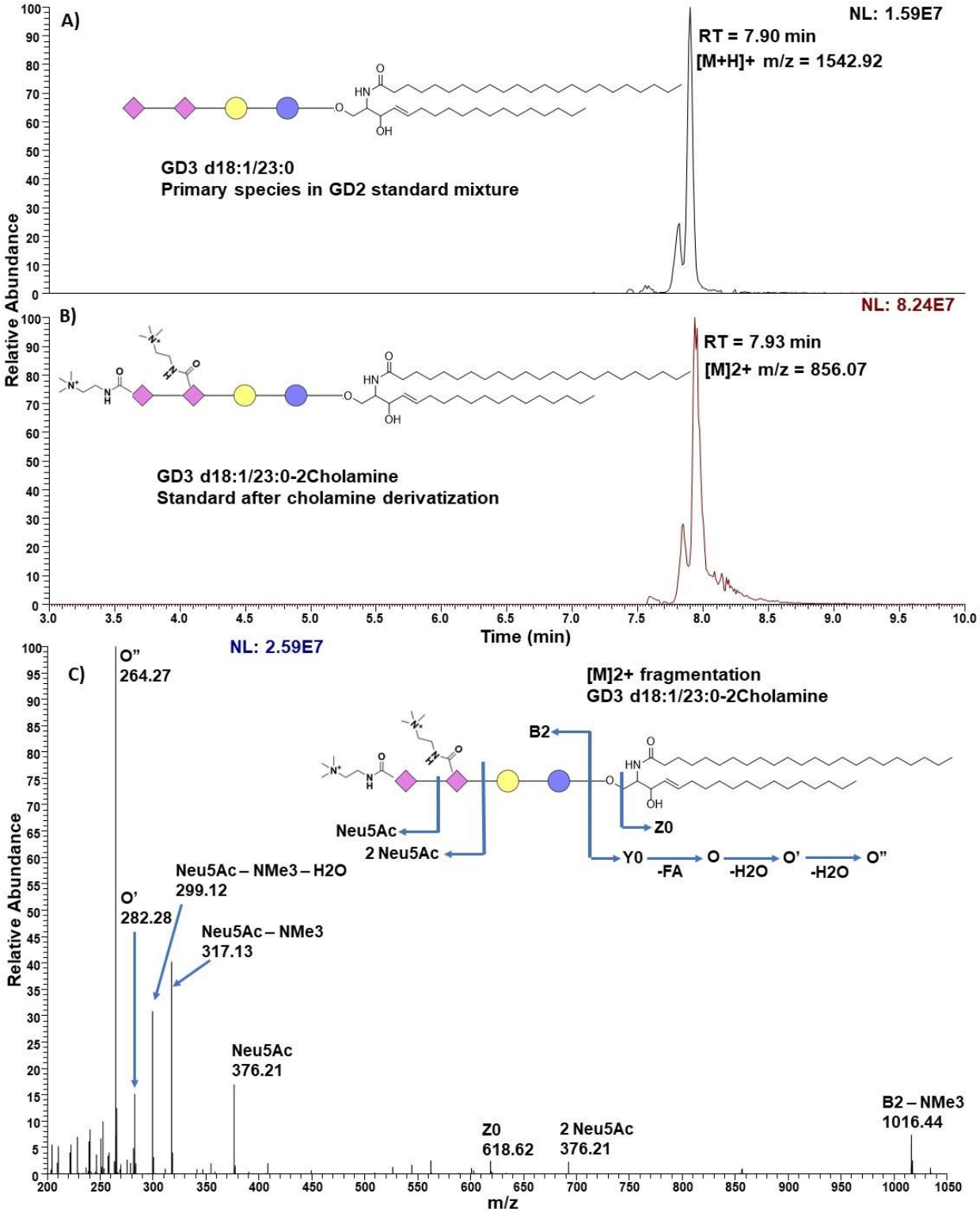


Table 7. Matreya Standard Mixture GD3 Gangliosides Derivatized and Identified.

RT	m/z	Gangliosides	Peak Area	Coeluting Isomers
7.94	856.0731	GD3 d18:1/23:0	1.17E+09	d19:1/22:0
7.60	849.0652	GD3 d18:1/22:0	9.59E+08	d16:1/24:0, d17:1/23:0
7.26	842.0547	GD3 d16:1/23:0	6.24E+08	d17:1/22:0
8.27	863.0795	GD3 d18:1/24:0	5.13E+08	
6.88	835.0473	GD3 d16:1/22:0	2.79E+08	
5.06	807.0177	GD3 d18:1/16:0	3.09E+08	
7.85	850.0693	GD3 d16:0/24:0	2.84E+08	d17:0/23:0, d18:0/22:0
7.50	843.0626	GD3 d16:0/23:0	2.72E+08	
7.14	836.0558	GD3 d16:0/22:0	2.05E+08	
7.22	855.0648	GD3 d18:1/23:1	2.55E+08	d16:1/25:1, d17:1/24:1, d18:2/23:0
7.57	862.0695	GD3 d18:1/24:1	1.52E+08	d18:2/4:0, d19:1/23:1, d20:2/22:0
5.41	808.0245	GD3 d18:0/16:0	5.56E+07	d16:0/18:0
8.47	870.0859	GD3 d19:1/24:0	9.74E+07	d18:1/25:0, d20:1/23:0
6.85	848.0568	GD3 d16:1/24:1	7.90E+07	d17:1/23:1, d18:1/22:1, d18:2/22:0
6.01	821.0341	GD3 d18:1/18:0	3.77E+07	d16:1/20:0
6.60	841.0487	GD3 d16:1/23:1	4.98E+07	
4.05	793.0027	GD3 d16:1/16:0	2.17E+07	d18:1/14:0
4.57	800.0092	GD3 d17:1/16:0	2.15E+07	
4.40	794.0079	GD3 d16:0/16:0	1.99E+07	
6.46	828.0420	GD3 d16:1/21:0	1.42E+07	
6.75	829.0474	GD3 d16:0/21:0	1.80E+07	
6.33	822.0421	GD3 d18:0/18:0	1.70E+07	d16:0/20:0
5.39	814.0236	GD3 d19:1/16:0	1.74E+07	d17:1/18:0
5.97	815.0322	GD3 d35:0	8.31E+06	-

GD3 is the first disialoganglioside in this progression and the smallest in the standard suite. The disialogangliosides have stronger 2+ ionization because they have two permanent formal charges via two cholamine tags. In the derivatized chromatogram, there is a shoulder from the isotopic peak of another species, but also some tailing of the peak. Peak tailing is observed for all disialogangliosides in this method and can be much improved by column choice. Superficially porous columns were found to be superior to fully porous columns.

GD2

Figure 43. GD2 d18:1/18:0 Cholamine Derivatization and Fragmentation.

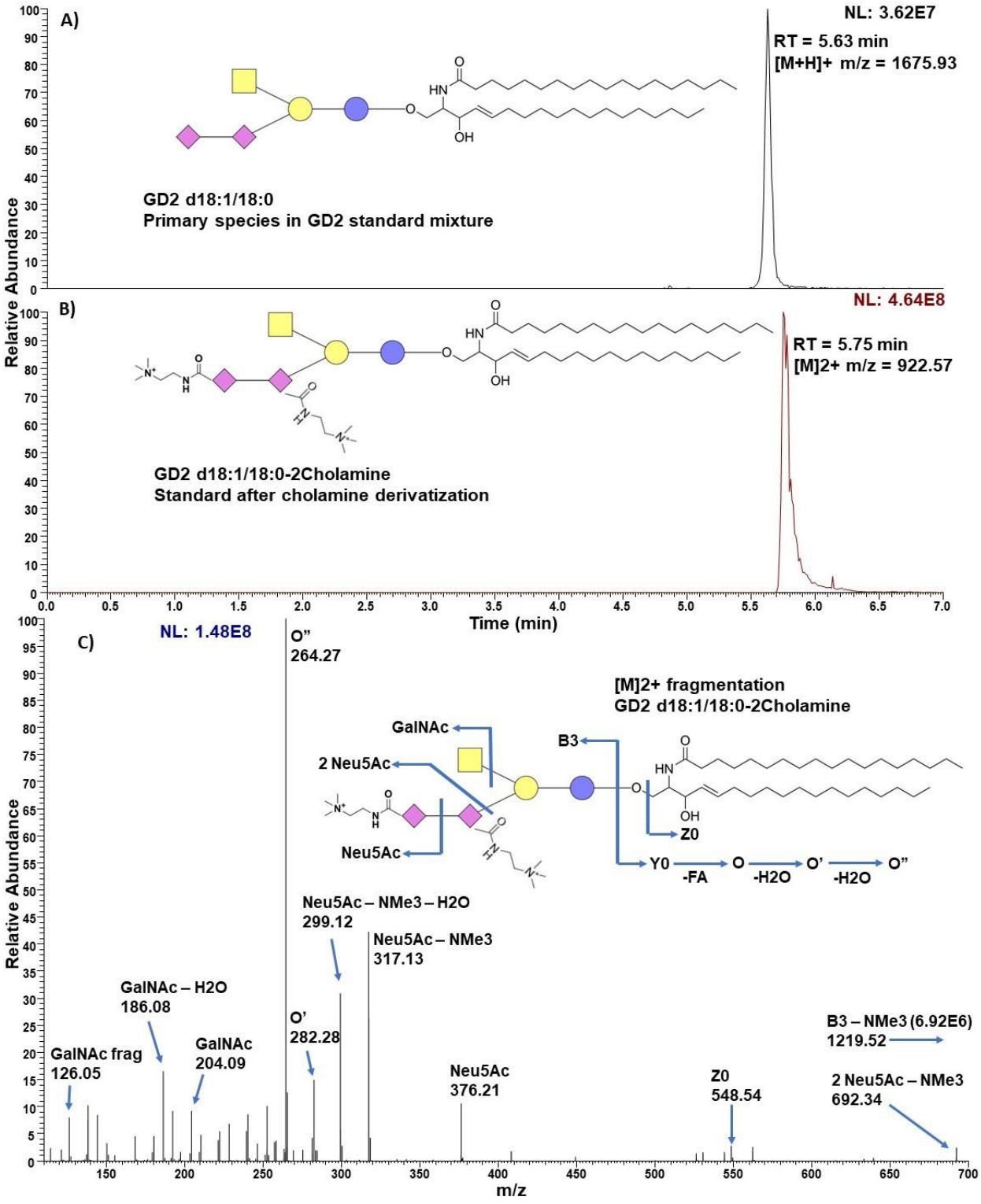


Table 8. Matreya Standard Mixture GD2 Gangliosides Derivatized and Identified.

RT	m/z	Gangliosides	Peak Area	Coeluting Isomers
5.76	922.5729	GD2 d18:1/18:0	5.83E+09	
6.62	936.5884	GD2 d20:1/18:0	4.97E+09	d18:1/20:0
6.22	929.5801	GD2 d19:1/18:0	2.24E+08	d18:1/19:0, d20:1/17:0
7.42	950.6036	GD2 d20:1/20:0	1.12E+08	d18:1/22:0
4.97	921.5658	GD2 d18:2/18:0	5.77E+07	
5.11	921.5630	GD2 d18:2/18:0	3.88E+07	
4.81	908.5569	GD2 d18:1/16:0	2.37E+07	d16:1/18:0
7.79	957.6104	GD2 d18:1/23:0	1.74E+07	d20:1/21:0
7.68	950.5999	GD2 d20:1/20:0	1.77E+07	
5.31	915.5674	GD2 d18:1/17:0	1.08E+07	d17:1/18:0

GD2 is a slightly larger disialoganglioside and so has better 2+ ionization. Interestingly, the GD2 ceramide diversity is very similar to the GM2 ceramide diversity, as was the GD3's to the GM3's. The GM3 and GD3 mixtures were both sourced from bovine buttermilk, and so their ceramide similarity is not surprising. However, GM2 and GD2 were from bovine and rabbit sources, respectively, and so it is curious that these two are so similar.

GD1b

Figure 44. GD1b d18:1/18:0 Cholamine Derivatization and Fragmentation.

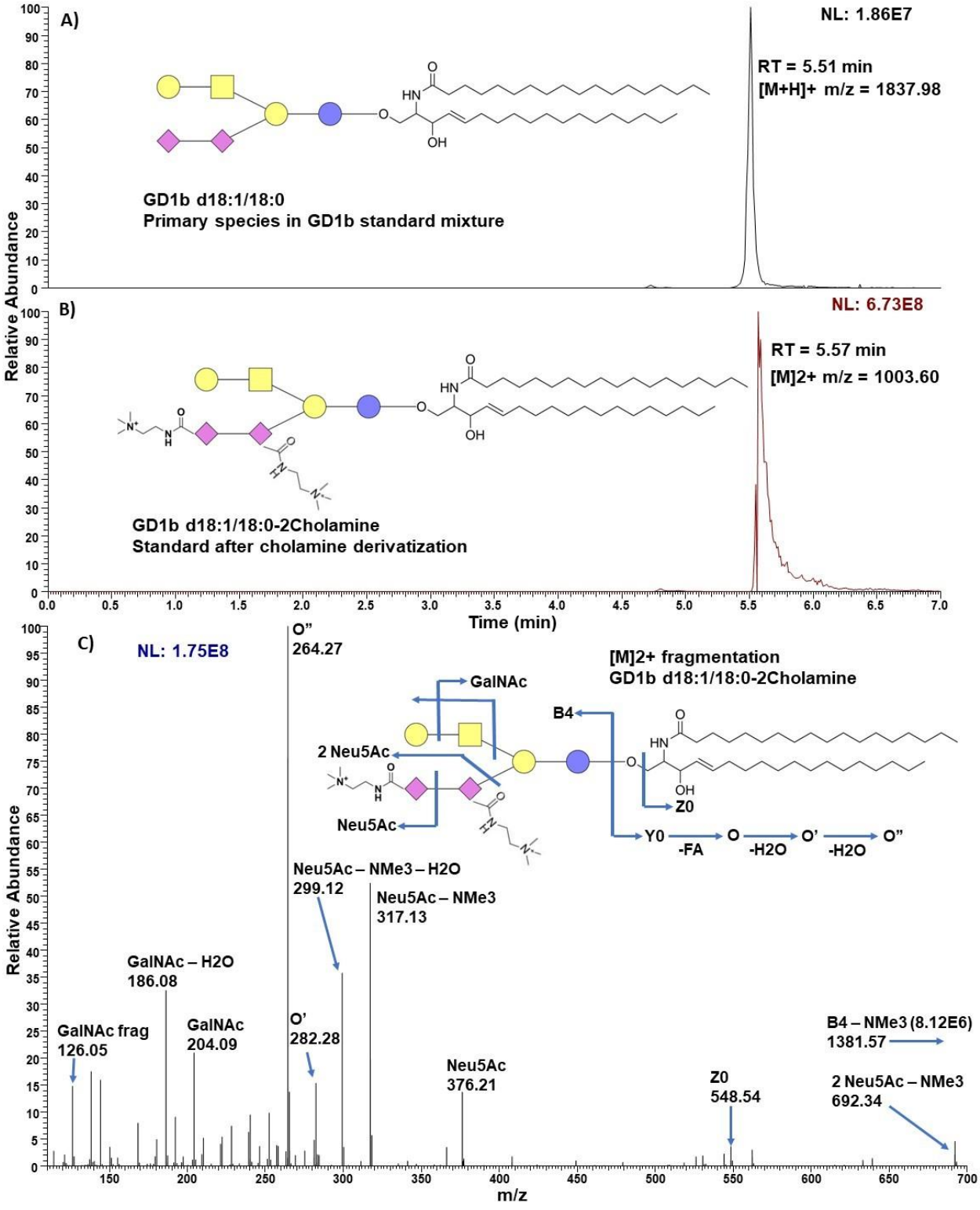


Table 9. Matreya Standard Mixture GD1b Gangliosides Derivatized and Identified.

RT	m/z	Gangliosides	Peak Area	Coeluting Isomers
6.45	1017.6150	GD1b d20:1/18:0	1.09E+10	
5.59	1003.5980	GD1b d18:1/18:0	7.15E+09	
5.90	1004.6020	GD1b d18:0/18:0	4.59E+08	
4.80	1002.5910	GD1b d18:2/18:0	1.69E+08	
4.65	989.5844	GD1b d16:1/18:0	1.30E+08	d18:1/16:0
7.27	1031.6280	GD1b d20:1/20:0	1.39E+08	d18:1/22:0
5.14	996.5920	GD1b d18:1/17:0	2.77E+07	d17:1/18:0

GD1b is the first of two glycan structural isomers in the standard suite (GD1b and GD1a), and so the fragmentation pattern here is critical for accurate glycan identification. The Neu5Ac-Neu5Ac fragment is characteristic of GD1b and not of GD1a. In the MS-DIAL analysis, there is another example of where the top Matreya species is the #2, inferior to another d20:1 LCB variant.

GD1a

Figure 45. GD1a d18:1/18:0 Cholamine Derivatization and Fragmentation.

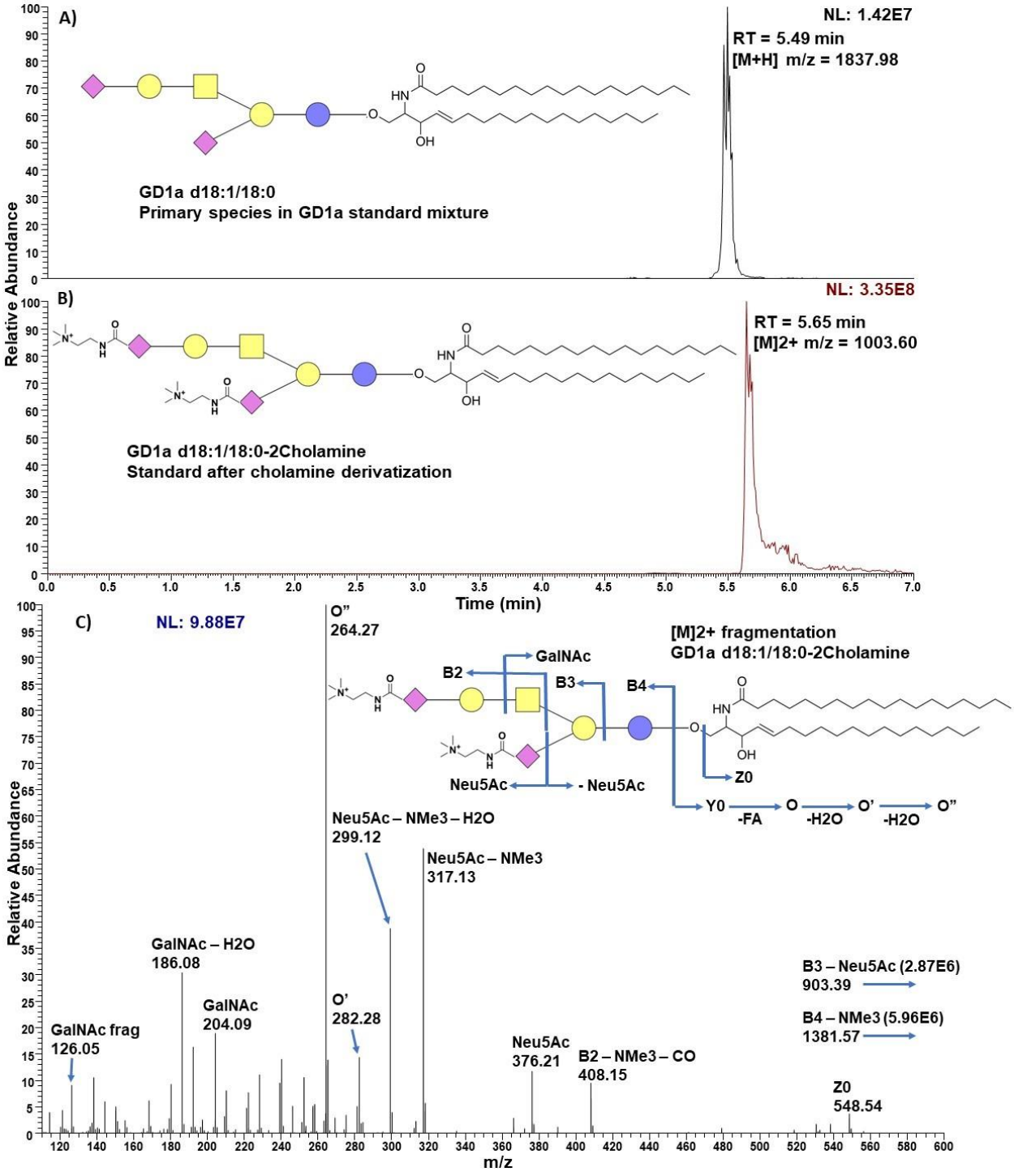


Table 10. Matreya Standard Mixture GD1a Gangliosides Derivatized and Identified.

RT	m/z	Gangliosides	Peak Area	Coeluting Isomers
6.52	1017.6150	GD1a d20:1/18:0	7.70E+09	
5.67	1003.5980	GD1a d18:1/18:0	7.60E+09	
6.16	1010.6080	GD1a d19:1/18:0	1.07E+08	d18:0/19:1, d18:1/19:0
4.80	989.5842	GD1a d16:1/18:0	5.39E+07	d18:1/16:0

GD1a is the second GD1 glycan isomer in the set. It does not have the Neu5Ac-Neu5Ac fragment that was diagnostic for GD1b. For its own diagnostic fragment, it has GalNAc-Gal-Neu5Ac, which is characteristic of the GD1a structure and not of GD1b. As per the MS-DIAL analysis, this is a relatively simple mixture in good agreement with Matreya's manifest.

Figure 46. Lactone Formation Observed in Cholamine Derivatization of GT1b and GQ1b.

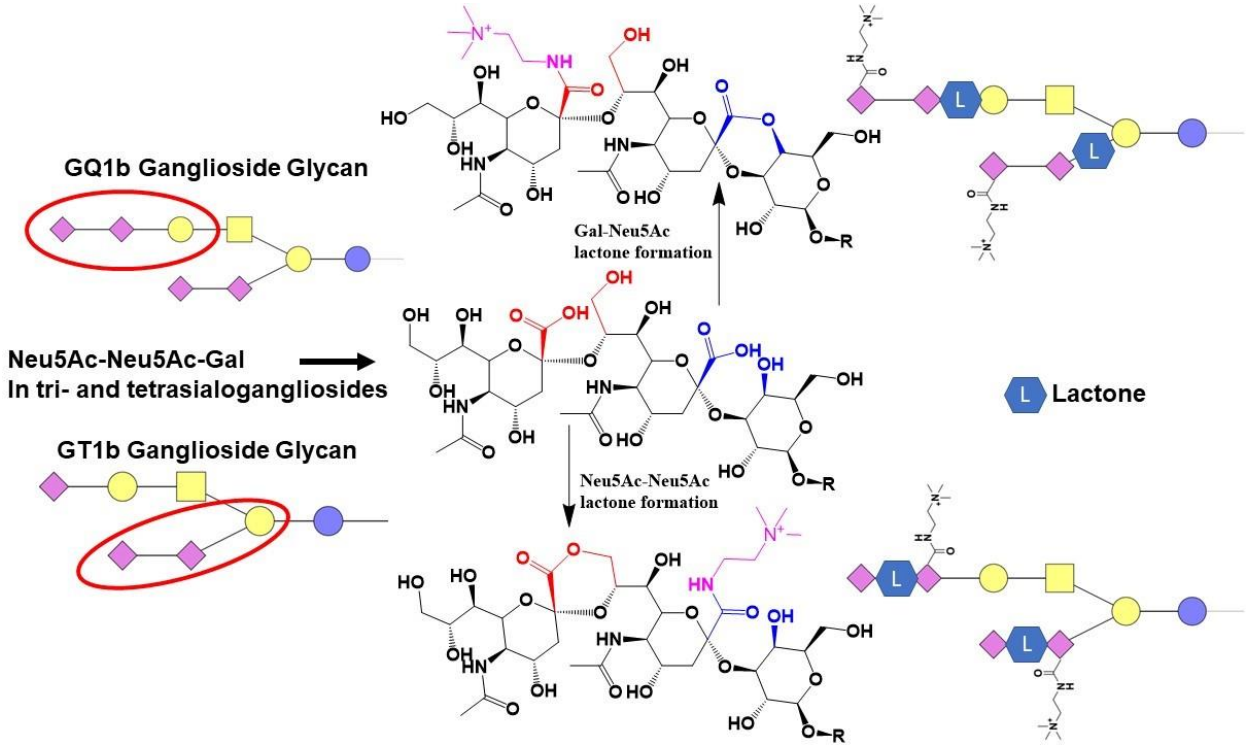


Figure 46 illustrates unique lactone generation seen with GT1b and GQ1b and with none of the mono- or disialogangliosides. It seems that the additional coulombic stress of adding a third or fourth permanent formal charge is too much for the cholamine derivatization.

Fragmentation observed is consistent with multiple outcomes, namely Neu5Ac-Neu5Ac and Neu5Ac-Gal lactone formation, these products eluting with slightly different retention times.

It is noteworthy that there are not as many products as would be expected if operating on the assumption that the lactone outcome is random. For both GT1b and GQ1b, the major product is only Neu5Ac-Neu5Ac lactones. Products with only Neu5Ac-Gal lactones are also observed at lower yield. What is not appreciably observed are most or all of the permutations mixing these outcomes that one might expect, leading to the possibility that there are different reactants rather than different products, such as the analytically challenging glycan disaccharide linkage isomers. By accident, we have stumbled upon a chemistry very similar to previously demonstrated lactone generation for the diagnosis of ganglioside glycan linkage isomers.[22, 23, 25] As such, it seems possible that this derivatization can be used to discern glycan linkage isomers for tri- and tetrasialogangliosides. However, we currently lack the standards necessary to confidently make this claim, making this utility a question for future research.

GT1b

Figure 47. GT1b d18:1/18:0 Cholamine Derivatization and Fragmentation.

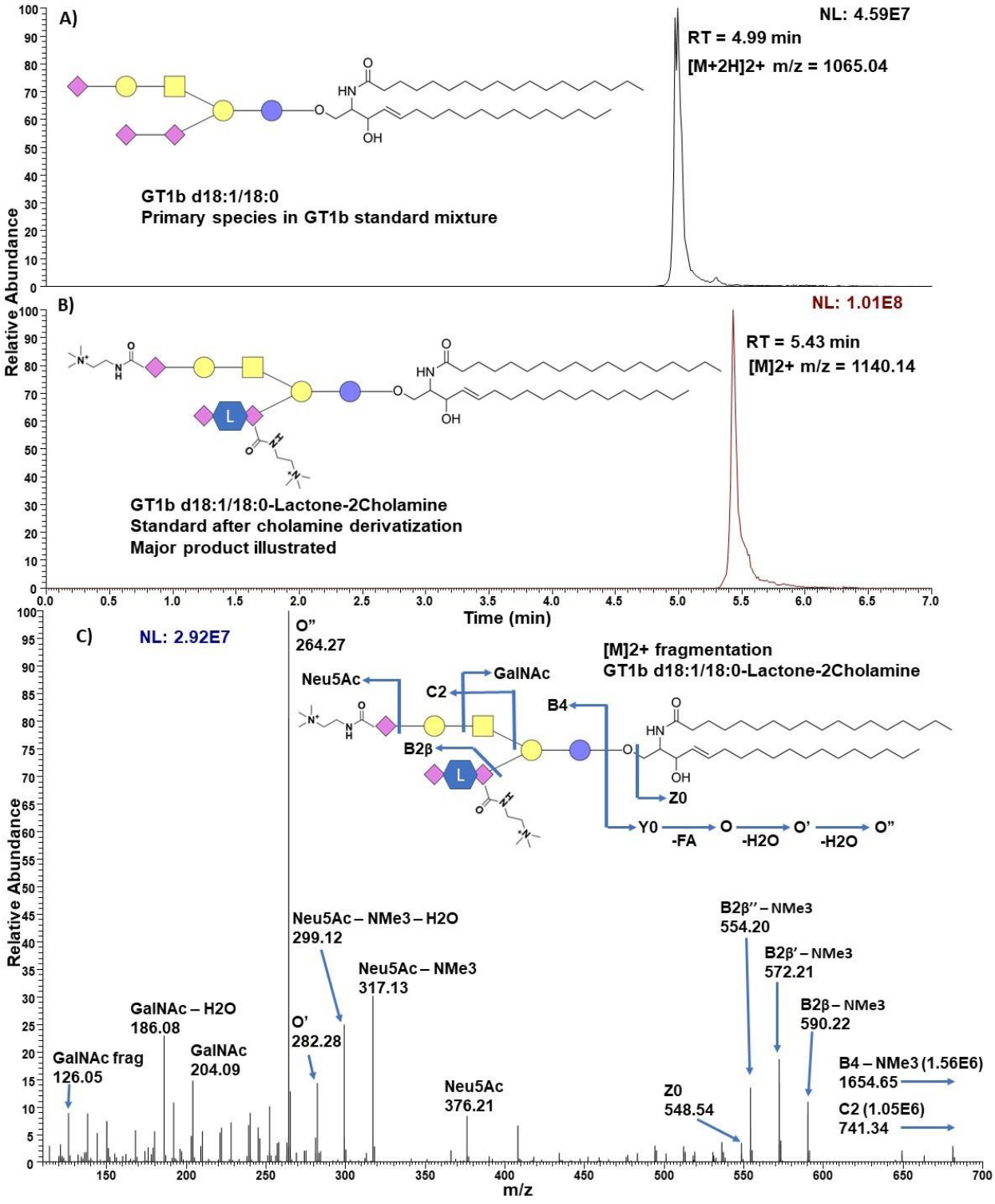


Table 11. Matreya Standard Mixture GT1b Gangliosides Derivatized and Identified.

RT	m/z	Gangliosides	Peak Area	Coeluting Isomers
6.28	1154.157	GT1b d20:1/18:0	1.01E+09	
5.44	1140.142	GT1b d18:1/18:0	6.99E+08	
6.60	1155.164	GT1b d20:0/18:0	8.99E+07	
7.08	1168.176	GT1b d20:1/20:0	8.91E+06	d18:1/22:0

As described in Figure 46, GT1b undergoes lactone formation in the cholamine derivatization reaction. Figure 47B illustrates the major product, the Neu5Ac-Neu5Ac lactone. There is a shoulder on the right side of the LC-MS peak, which is the lesser Neu5Ac-Gal lactone product. The lactone and its position produce unique fragment ions, which were instrumental in diagnosing the nature of these products and the difference between them.

GQ1b

Figure 48. GQ1b d18:1/18:0 Cholamine Derivatization and Fragmentation.

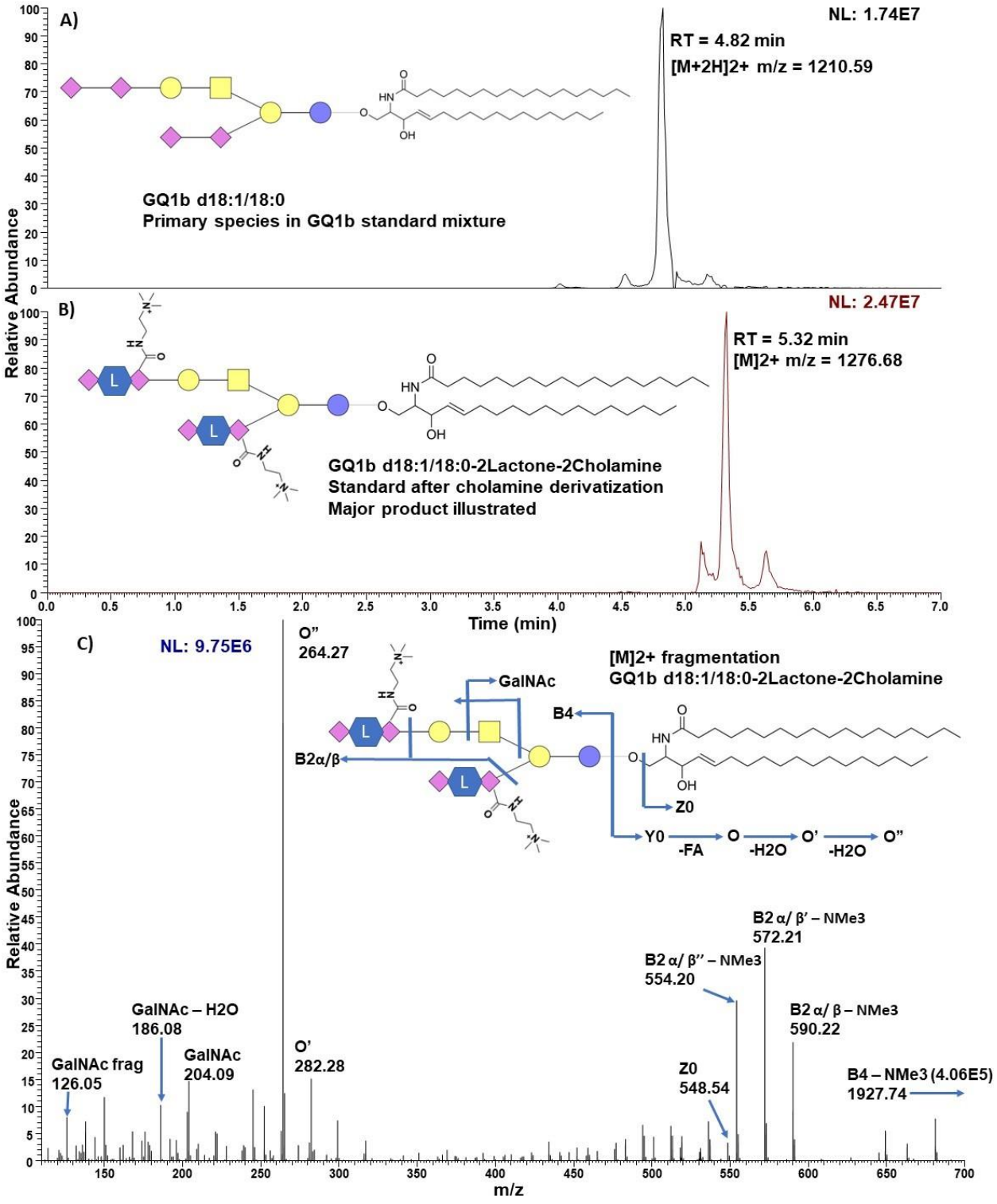


Table 12. Matreya Standard Mixture GQ1b Gangliosides Derivatized and Identified.

RT	m/z	Gangliosides	Peak Area	Coeluting Isomers
6.17	1290.701	GQ1b d20:1/18:0	2.66E+08	d18:1/20:0
5.32	1276.681	GQ1b d18:1/18:0	1.36E+08	

While the GQ1b standard mixture appears to be a simple one, there is evidence of unknown isomerism, even in the underivatized chromatogram. There are three product peaks in the derivatized chromatogram that appear to be the result of three initial ganglioside peaks in the underivatized chromatogram. Answering whether these are ultimately the result of linkage isomers or another form of isomerism, such as the undocumented presence of GQ1a or GQ1c, will require an orthogonal analysis strategy or more specific standards. The other two products are however certainly GQ1 species with two cholamine tags and two lactones.

Quantitation Validation Cohort

Figure 49. GM3 Cohort Results in Cholamine Isotopic Labeling Quantitation

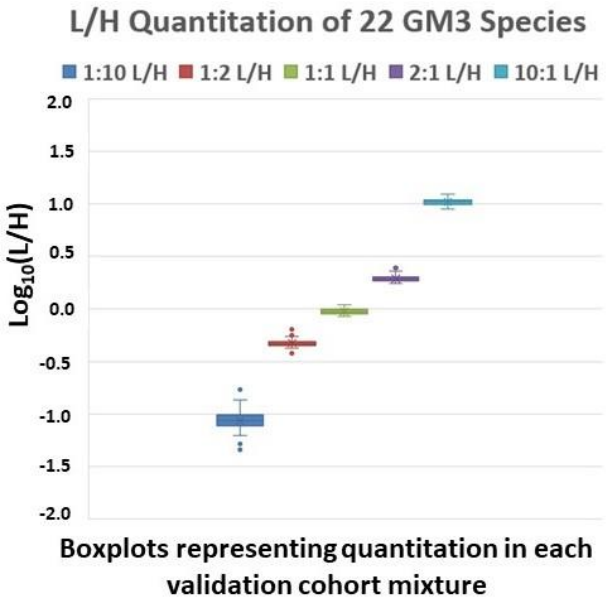


Figure 49 illustrates the key results of the isotopic labeling relative quantification cohort with GM3. The five mixtures composing the cohort have known light-to-heavy (L/H) mixing

ratios of 1:10, 1:2, 1:1, 2:1, and 10:1 with expected measurement log(L/H) values of -1.0, -0.30, 0.0, 0.30, and 1.0, respectively. These results show that the concept of conducting ganglioside relative quantification through isotopic labeling with cholamine is sound.

Conclusions

While the results with the standards are very promising, and 260 ganglioside identifications is higher than most method publications report, this workflow has several considerations for its use in biological application which may cause difficulty. First, the labeling reaction is by necessity completely dry; even condensation inside an extract vial from the freezer may cause issues. This can be solved through careful practice. Next, the 80 °C reaction condition leads to refluxing of the MeOH/IPA solvent mixture, requiring the use of screwcap vials or otherwise making the samples airtight, but such solutions are readily available. Third and most significantly, as the reaction is fundamentally amide bond formation between a carboxylate and a primary amine, there are numerous interfering agents in a biological sample.

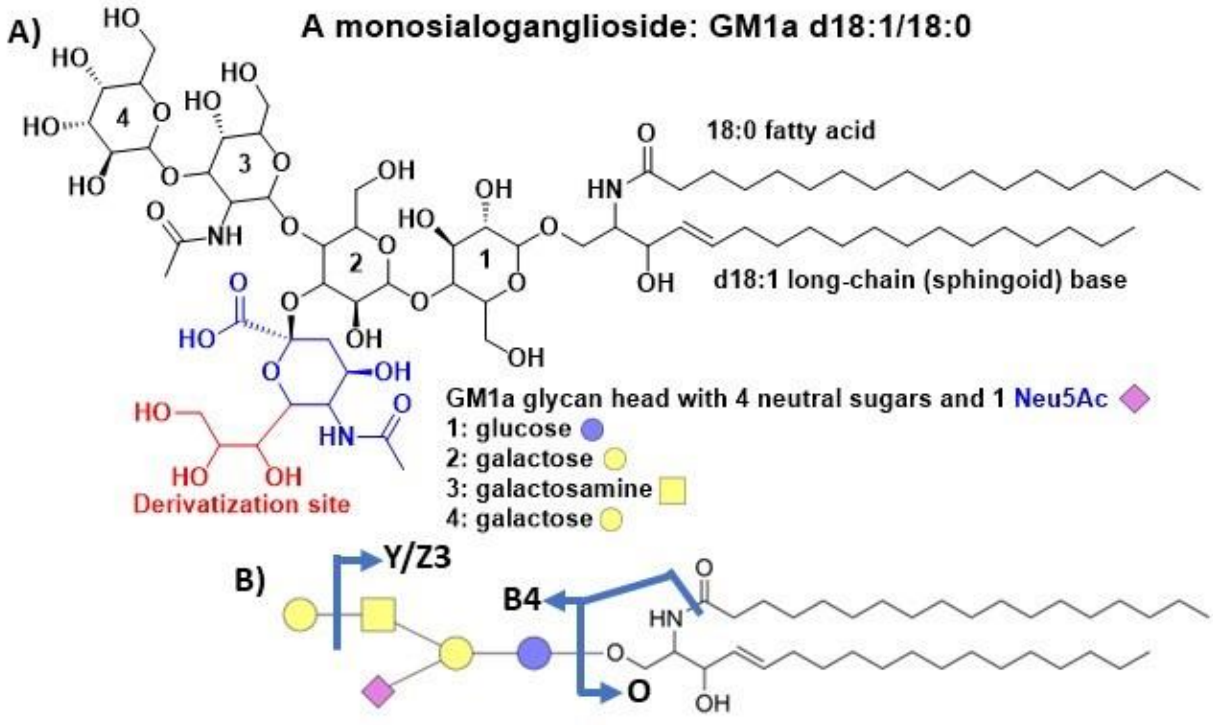
For example, amino acids possess both a carboxylate and a primary amine, allowing for 'ultrapolymerization' of metabolites by the reaction procedure, observed as fluffy white precipitates and resulting in incomplete ganglioside derivatization through reagent consumption. We have experimented with several biological samples and extraction methodologies but have not yet found the right balance between eliminating interfering agents and making an efficient, reproducible workflow. That the method works well with the standards extracted and purified from biological sources provides assurance that this can be done. Once this problem is solved, cholamine derivatization will be a powerful tool for the whole ion analysis of gangliosides and provides a means for isotopic labeling relative quantification in their comparative analyses.

CHAPTER V: HIGH THROUGHPUT LC-MS/MS BASED ANALYSIS OF GANGLIOSIDES USING ISOBARIC LABELING

Introduction

Gangliosides are glycosphingolipids composed of a hydrophobic ceramide moiety connected by ether bond to a carbohydrate or 'glycan' moiety. The inclusion of one or more N-Acetylneuraminic acids (shorthand Neu5Ac or sialic acid) differentiates gangliosides from other glycosphingolipids. Figure 50 illustrates (A) the structure of a typical monosialoganglioside with naming conventions initially outlined by Svennerholm[13] and (B) glycan structural representation and fragmentation as popularized by glycated peptide research[62] and subsequently adapted for convenient discussion[20, 61] of gangliosides.

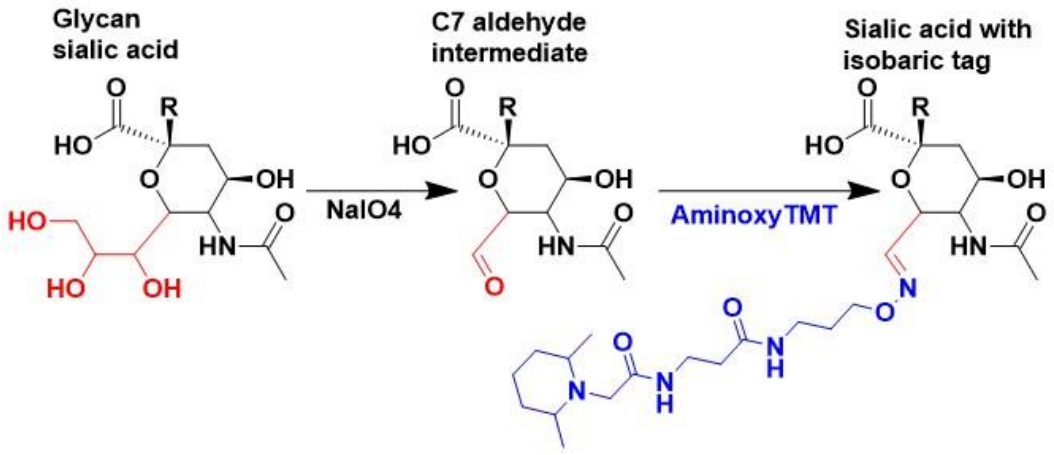
Figure 50. Ganglioside Nomenclature and Structure with Highlighted Derivatization Site.



In Figure 50, the amphiphilic nature of the ganglioside is plain to see, responsible for their insertion into the hydrophobic environment of the plasma membrane[47, 51, 74] as well as their instrumental role in forming and maintaining lipid rafts[27, 32, 43] through hydrogen-bonding and other hydrophilic interactions. Lipid rafts and by extension gangliosides have been mechanistically linked to the progression of numerous pathologies including cancers[52], neurodegenerative disorders[20], and Type 1 Diabetes (T1D)[40]. This makes gangliosides promising biomarkers for tracking disease progression and candidates for exogenous therapies. Although gangliosides can be analyzed directly using tandem mass spectrometry for their ceramides in the positive ion mode and for their glycans in the negative ion mode[17], chemical derivatization enhances[16] structural identification through improved fragmentation while also enabling more sophisticated analytical techniques such as stable-isotope labeling.

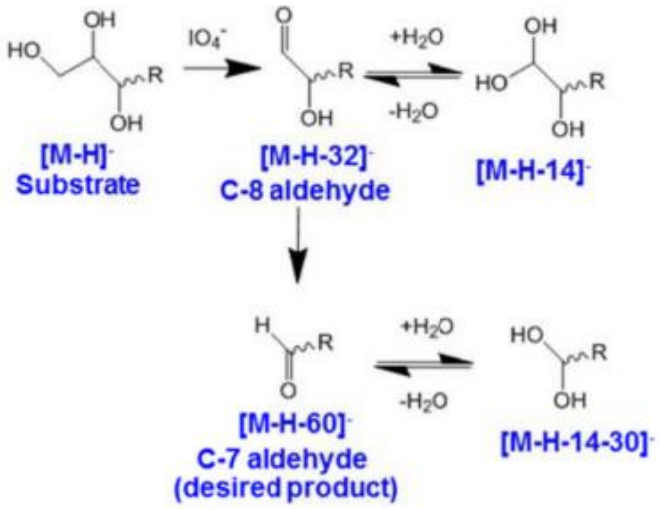
Highlighted in red in Figure 51, the glycerol arm moiety of sialic acid has long been the target of derivatization studies of both glycated peptides and gangliosides.[16, 74, 77, 78] Figure 51 illustrates a simplified view of a previously reported[16] chemical derivatization of gangliosides intended for multiplexed analysis with the isobaric tag aminoxyTMT from Thermo Fisher Scientific.

Figure 51. Fundamental Progression of Glycan Isobaric Tagging with AminoxyTMT.



This method depends on formation of an aldehyde intermediate through oxidation by periodate. The hydroxylamine functional group terminus of the aminoxyTMT reagent is selective for ketones and aldehydes. With this aldehyde being the only target on the ganglioside, the second reaction can proceed to completion using remarkably few equivalents of the tagging reagent. However, diol cleavage by periodate produces small carbonyl byproducts, such as formaldehyde and formic acid. These additional carbonyls would react with aminoxyTMT and were eliminated in the original method by SPE desalting, which also removed unquenched periodate. While effective, SPE introduces variability and reduces total method sensitivity. Additionally, the aldehyde intermediate was part of a more complex equilibrium, see Figure 52.

Figure 52. Aldehyde Intermediate Instability. Ref.[16]



The glycerol arm can be oxidized to either the C7 or C8 aldehyde, which would lead to two different tagging products in the second reaction. Under continued oxidation conditions, the C8 aldehyde can be further oxidized to the C7 aldehyde, as was reported in optimized conditions. These aldehydes also exist in equilibrium with geminal diols, which are not hydroxylamine reactive. On repetition with a diverse set of ganglioside standards, we found that different glycans have different ratios of these two products under the same reaction conditions, and we

feared that the same glycan might even have different product ratios depending on biological matrix factors. With the aldehyde intermediate concerns in mind and a desire to eliminate the SPE desalting step, we critically examined each step of the original procedure and found solutions to both problems, then applying the updated methodology to biological samples.

Materials and Methods

Materials

Sodium (meta)periodate (SKU S1818), glycerol (SKU G5516), and sodium acetate (SKU S2889) was purchased from Sigma-Aldrich. The aminoxyTMTsixplex isobaric tagging reagents were purchased from Thermo Fisher Scientific (catalog number 90401) along with all other reagents and solvents.

Mouse Brain Dissection and Homogenization

One normal mouse brain, previously stored at -80 °C was dissected on ice following guidance available online from Neurogenetics at UT Health Science Center (https://www.mbl.org/anatomy_images/fixed/mba.html) into regional samples and massed by difference in bead beating homogenization tubes. The regions used in this study and their masses are in Table 13.

Table 13. Mouse Brain Fraction Masses.

Brain Fraction	Mass (mg)
Cerebellum	62.53
Pons, medulla	53.31
Right midbrain	27.33
Right posterior cortex	38.05

5 mM ammonium formate was added to each sample to reach 0.05 mg tissue per μL. Homogenization was conducted by a Bertin Precellys bead beating homogenizer at 8000 rpm

with four 30 sec cycles with 30 sec breaks between. Homogenates were immediately subjected to ganglioside extraction.

Ganglioside Extraction

To extract gangliosides from the homogenate, the Folch method was used with minor modifications as follows. For each sample, 40 μL aqueous homogenate representing 2 mg tissue was aliquoted into 160 μL 1:2 TCM:MeOH. The resulting mixture was briefly vortexed and then sonicated 5 min followed by 600 g centrifugation for 15 min. The single-phase supernatant was transferred to a new vial. 33 μL water was added to achieve phase separation. The mixture was vortexed briefly, sonicated 5 min, and centrifuged at 16,000 g for 10 min at 4 °C. The upper aqueous phase was collected and dried. The dry ganglioside extracts were stored at -80 °C.

Ganglioside Derivatization

Ganglioside extracts were reconstituted in 200 uL cold 200 mM pH = 5.5 HOAc/NaOAc buffer with 1 mM $NaIO_4$. While working with standards, we found that increasing the buffer concentration from the original 100 mM to 200 mM improved aldehyde intermediate ratios. Experiments with changing the buffer salts, periodate concentration, and pH did not produce useful results. Samples were then incubated for 5.5 hr at 1 °C in the dark at 600 rpm. While longer incubation times eventually lead to oxidation of glycan sugar rings, we found that extending the original incubation time from 1 hr to 5.5 hr dealt with most C8 aldehyde intermediates without appreciable losses to ring oxidation. Among the standards, there were still some that produced C8 aldehydes, mostly smaller glycan species such as GM3 and GM4. The reaction was quenched with 60 μL 10.0 mM glycerol. This is 3 eq glycerol per $NaIO_4$, which we found sufficient to fully quench the excess periodate and eliminate that particular need for SPE desalting. The mixtures were further incubated a minimum of 2 hr to fully quench.

Next, the samples were dried thoroughly, which we found eliminates the small molecule carbonyl byproducts completely, such that the later tagging reaction is not impeded. This otherwise achieves the second goal of SPE desalting. Samples were then reconstituted with 200 uL 50/50/0.1 MeOH/H_2O/HOAc with 0.1 mg aminoxyTMT. The original solvent composition was 95/5/0.1 MeOH/H_2O/HOAc for two reasons: (1) to help prevent hydrolysis of the oxime bond formed and (2) to make easier subsequent drying steps. While subsequent drying steps were certainly slower with higher water content, oximes hydrolyze very slowly and appreciable losses were not observed between methodologies. This higher water content was necessary to carry forward the salt from previous steps without SPE desalting. After reconstitution, samples were allowed 10 min to react at room temperature with periodic vortexing. To encourage full tagging with low tagging reagent equivalence, samples were then subject to two cycles of complete drying and reconstitution. After the second reconstitution, 100 uL 10% acetone in the same 50/50/0.1 MeOH/H_2O/HOAc buffer was added to quench excess tagging reagent. Samples were then dried and reconstituted in mobile phase solvent for multiplexing.

Multiplexing Strategy

Figure 53. Members of the AminoxyTMTsixplex with Reporter Ions.

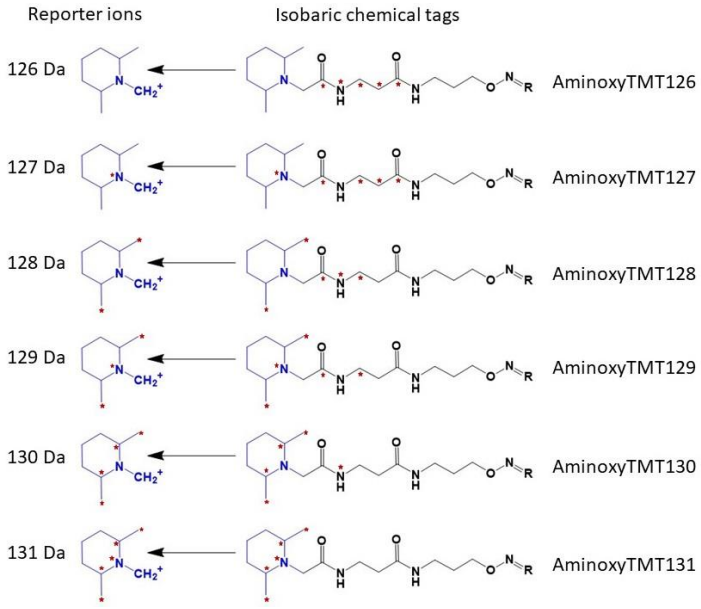


Figure 53 illustrates the six isobaric reagents of the aminoxyTMTsixplex. The red stars indicate heavy atom (^{15}N and ^{13}C) positions. In isobaric analysis, each sample is tagged with one of the reagents and then multiplexed for simultaneous LC-MS/MS injection. The reporter ions in the MS2 spectrum indicate each sample's contribution to a given LC-MS peak.

Isobaric tagging reagents are precious, so creative multiplexing is essential. We ran three samples of each region with Vials 1, 2, and 3 in Table 14 representing the primary, multiplexed samples. We were also able to run one QC vial for each region, Vials 4 through 7, wherein the reporter ions should be of approximately equal intensity. There was an error in preparing Cortex N=2 such that its derivatized ganglioside content is much lower than can reasonably be expected. As a consequence, Vial 2 (N=2 multiplex) has 36 species that pass the Iglewicz and Hoaglin outlier test with a Z-score less than - 3.5, and this sample is omitted in subsequent data. There are such examples from this multiplex as GD1a d36:1, a top 4 cortex high abundance species, having a cortex contribution of 4.2% instead of mean 58% in the other multiplexes. The error

most likely occurred in extraction, in which it is necessary to achieve equilibrium between the aqueous and chloroform layers and then take the aqueous layer without any contamination from the precipitate layer that forms between them.

Table 14. Multiplexing of Tagged Samples.

Vial	TMT126	TMT127	TMT128	TMT129	TMT130	TMT131
1	Cereb. N=1	Pons. N=1	Midb. N=1	Cort. N=1		
2	Midb. N=2	Cort. N=2			Cereb. N=2	Pons. N=2
3			Cereb. N=3	Pons. N=3	Midb. N=3	Cort. N=3
4	Cereb. N=1		Cereb. N=3		Cereb. N=2	
5		Pons. N=1		Pons. N=3		Pons. N=2
6	Midb. N=2		Midb. N=1		Midb. N=3	
7		Cort. N=2		Cort. N=1		Cort. N=3

Tandem Mass Spectrometry Analysis

A Thermo Fisher Scientific Vanquish Horizon UHPLC System coupled with a Thermo Fisher Scientific Orbitrap Exploris 240 was used to conduct LC-(+)FTMS/MS analysis. A Phenomenex Kinetex HPLC 1.7 μm C18 100Å (100 x 2.1 mm) column was used along with the following mobile phases: (A) 0.1% formic acid and 10 mM ammonium formate in 60/40 ACN/H_2O and (B) 0.1% formic acid and 10mM ammonium formate in 88/10/2 IPA/ACN/H_2O. The gradient was 30% B (- 1 min), 30% B (0 min), 99% B (13 min), 99% B (14 min), 30% B (14.1 min), 30% B (15 min). The column was maintained at 40 °C with flowrate 350 μL/min. The injection volume was 5.00 μL. Ionization for FTMS analysis utilized a HESI source in the positive mode with spray voltage 3 kV, capillary temperature 350 °C, sheath gas flowrate 20 au, aux gas flowrate 5 au, and probe heater temperature 400 °C. FTMS/MS scanning had three time-dependent loops to maximum ganglioside coverage: full MS1 over 680 m/z to 1520 m/z from RT 3 min to 10 min with MS2 scans each cycle of the top 10 2+ ions, full MS1 over 1200 m/z to 1350 m/z from RT 5 min to 7 min with MS2 scans each cycle of the top 5 1+ ions, and full MS1

over 1300 m/z to 1600 m/z from RT 7 min to RT 10 min with MS2 scans each cycle of the top 5 1+ ions. All MS1 scans had maximum IT 100 ms and mass resolution 120,000. All MS2 scans had isolation window 1 m/z, stepped normalized HCD collision energies 30 then 35 then 40, mass resolution 15,000, maximum IT 50 ms, and MS2 spectrum first mass 120 m/z.

MS-DIAL Identification and Isotopic Peak Pairing

MS-DIAL 4.92 for Windows x64 was used in metabolomics mode alongside an in-house ganglioside derivatization library. MS-DIAL data collection proceeded with MS1 tolerance 0.005 Da and maximum charge number 2. The minimum peak height was set to 10^4 with mass slice width 0.05 Da. The peak smoothing method was linear weighted moving average with smoothing level of 3 scans and minimum peak width of 5 scans. The Sigma window value was set to 0.1 without exclusion after precursor ions and with isotopic ions w/o MS2Dec preserved. Automated identification was performed with a MSP library with retention time tolerance 0.3 min, accurate mass tolerance (MS1) 0.005 Da, accurate mass tolerance (MS2) 0.01 Da, identification score cut off 60%, and retention time used for filtering but not scoring. Peak alignment in reference to a multiplex sample, retention time tolerance 0.1 min, MS1 tolerance 0.003 Da, and linear extrapolation RT correction by MS-DIAL. Post-processing of MS-DIAL data used a simple script to organize reporter ion and ceramide data, matching measured MS2 fragment m/z to reporter ion m/z within a 0.005 m/z tolerance. Ceramide analysis matching utilized the O’ and O’’ LCB fragments reported in the literature.[15, 16]

Essential MS2 Ions

Table 15. Ganglioside AminoxyTMT Diagnostic and Quantitative MS2 Ions.

TMT Reporter Ions

Tag	m/z
TMT0	126.12783
TMT126	126.12783
TMT127	127.12483
TMT128	128.13454
TMT129	129.13154
TMT130	130.14125
TMT131	131.13825

Spingoid LCB Ions

	m/z	
LCB	O'	O"
d16:0	256.26367	238.25311
d16:1	254.24801	236.23745
d16:2	252.23235	234.22179
d18:0	284.29499	266.28443
d18:1	282.27933	264.26877
d18:2	280.26367	262.25311
d20:0	312.32631	294.31575
d20:1	310.31065	292.30009
d20:2	308.29499	290.28443

Sialic Acid Species and Intermediate Diagnostic Ions

	m/z			
	C7 Aldehyde		C8 Aldehyde	
Fragment	Neu5Ac	Neu5Gc	Neu5Ac	Neu5Gc
SA+TMT	531.29777	547.29267	561.30833	577.30323

Neu5Ac-C7-Derivatized Glycan Diagnostic Fragments

	m/z		
Class	Glycan (B)	2nd Fragment	2nd Mass
GM1a	1218.87839	GalNAc	204.08667
GM2	1056.90389	Y2α/B2	693.35057
GM3	853.92629	B1	693.35057
GM4	691.95179	-	-
GD1a	1748.86351	Y2α/B3	693.35057
GD1b	1509.84069	2 Neu5Ac	822.39318
GD2	1347.86619	2 Neu5Ac	822.39318
GD3	1144.88859	2 Neu5Ac	822.39318
GT1b	2039.82581	2 Neu5Ac	822.39318
GQ1b	2330.78811	2 Neu5Ac	822.39318

Table 15 includes an incomplete but key list of MS2 fragment ions used for quantitation, reaction monitoring, and species identification. The TMT reporter ions are the essential fragments used for comparative quantitation. The LCB ions are useful for determining ceramide

content.[15] The derivatized sialic acid has high intensity characteristic fragments based on the aldehyde intermediate and based on sialic acid subtype, Neu5Ac and Neu5Gc. While many more glycan fragments exist in this methodology[16], some essential ones are provided. The intact glycan is sometimes not detectable, especially with the larger species, but it can be deduced by difference using the typically detectable intact ceramide ion Z0, which can be further used for specific ceramide species determination with the LCB ions.

In MS-DIAL, identifying ions were matched within a 0.01 Da tolerance while in post processing for reporter ion and ceramide isomer quantitation, ions were matched within a 0.005 Da tolerance.

Results and Discussion

LC-MS Peaks and Annotations

Table 16 illustrates the major ganglioside LC-MS peaks from the six chromatographic injections with annotated species name, RT, m/z, adduct type, and MS1 intensity. All are C7 aldehyde oxidized and completely TMTsixplex tagged. All are Neu5Ac variants. We expected to find no or nearly no Neu5Gc species given the animal studied, but the total absence of C8 aldehyde products is surprising. We optimized the method to reduce C8 product but never eliminated it with standards. It seems the biological matrix took care of this problem for us. We had hypothesized that the oxidation issue was in part due to the formation of ganglioside micelles and other hydrophobic structures in the aqueous oxidation buffer. We modeled this in YASARA[79] and found that gangliosides form a variety of hydrophobic structures in very little simulation time, some of them completely protecting the glycerol arm moiety, others partly shielding it such that periodate could only access carbons 8 and 9, leading to C8 aldehyde formation. These structures form due to favorable ganglioside-ganglioside interactions. In the

biological matrix, there are far more other molecules to interact with, sufficiently disrupting the ganglioside-ganglioside interactions.

Table 16. Major LC-MS Peaks with MS2 Acquisition and Confident Identification.

Metabolite	RT (min)	m/z	Adduct	MS1 Intensity
		Multiplex 1		
GD1a d34:1	4.68	1144.671	[M+2H]2+	1.35E+06
GD1a d36:1	5.39	1158.688	[M+2H]2+	5.37E+07
GD1a d38:1	6.01	1172.704	[M+2H]2+	1.38E+07
GD1b d34:1	4.69	1025.074	[M+2H]2+	1.65E+06
GD1b d36:1	5.41	1039.090	[M+2H]2+	5.33E+07
GD1b d36:2	4.80	1038.080	[M+2H]2+	3.07E+06
GD1b d38:1	6.04	1053.108	[M+2H]2+	1.55E+07
GD1b t38:2	5.29	1060.096	[M+2H]2+	2.55E+06
GD3 d36:1	5.78	856.525	[M+2H]2+	2.26E+07
GD3 d38:1	6.36	870.541	[M+2H]2+	4.42E+06
GD3 d40:1	6.90	884.557	[M+2H]2+	1.17E+06
GD3 d41:1	7.14	891.565	[M+2H]2+	5.28E+05
GM1a d34:1	5.03	879.529	[M+2H]2+	6.76E+06
GM1a d36:1	5.71	893.543	[M+2H]2+	2.08E+08
GM1a d38:1	6.31	907.557	[M+2H]2+	7.34E+07
GM1a d42:2	6.82	934.583	[M+2H]2+	3.59E+06
GM1a t41:1	7.32	936.584	[M+2H]2+	1.99E+05
GM2 d36:1	5.87	812.516	[M+2H]2+	6.47E+07
GM2 d36:2	5.29	811.509	[M+2H]2+	3.14E+06
GM2 d38:1	6.44	826.533	[M+2H]2+	2.78E+07
GM3 d36:1	6.04	710.977	[M+2H]2+	1.86E+07
GM3 d38:1	6.63	724.992	[M+2H]2+	4.95E+06
GM4 d36:1	5.94	1258.896	[M+H]+	7.85E+05
GM4 d38:1	6.55	1286.928	[M+H]+	2.72E+05
GM4 t36:1	5.82	1274.892	[M+H]+	5.73E+05
GQ1b d36:1	4.96	1449.781	[M+2H]2+	5.11E+06
GT1b d34:0	4.61	1291.230	[M+2H]2+	2.18E+05
GT1b d34:1	4.39	1290.220	[M+2H]2+	1.43E+06
GT1b d36:1	5.11	1304.236	[M+2H]2+	1.53E+08
GT1b d36:2	4.52	1303.225	[M+2H]2+	2.60E+06
GT1b d38:1	5.77	1318.249	[M+2H]2+	5.42E+07
		Multiplex 3		
GD1a d34:1	4.66	1144.675	[M+2H]2+	1.17E+06

GD1a d36:1	5.43	1158.687	[M+2H]2+	2.38E+07
GD1a d38:1	6.03	1172.703	[M+2H]2+	1.55E+07
GD1b d34:1	4.67	1025.073	[M+2H]2+	8.43E+05
GD1b d36:1	5.43	1039.089	[M+2H]2+	2.42E+07
GD1b d36:2	4.80	1038.085	[M+2H]2+	2.00E+06
GD1b d38:1	6.01	1053.108	[M+2H]2+	1.17E+07
GD3 d36:1	5.78	856.526	[M+2H]2+	1.11E+07
GD3 d38:1	6.38	870.545	[M+2H]2+	9.98E+05
GD3 d40:1	6.87	884.556	[M+2H]2+	7.93E+05
GD3 d41:1	7.11	891.566	[M+2H]2+	5.12E+05
GM1a d34:1	5.02	879.528	[M+2H]2+	6.60E+06
GM1a d36:1	5.68	893.541	[M+2H]2+	1.29E+08
GM1a d38:1	6.30	907.557	[M+2H]2+	7.64E+07
GM1a d42:2	6.79	934.582	[M+2H]2+	1.98E+06
GM2 d36:1	5.83	812.515	[M+2H]2+	2.09E+07
GM2 d36:2	5.27	811.510	[M+2H]2+	2.23E+06
GM2 d38:1	6.43	826.532	[M+2H]2+	1.09E+07
GM3 d36:1	6.03	710.979	[M+2H]2+	1.09E+07
GM3 d38:1	6.62	724.994	[M+2H]2+	3.09E+06
GM3 d38:1	6.62	725.000	[M+2H]2+	1.43E+06
GM4 d38:1	6.50	1286.928	[M+H]+	9.82E+04
GM4 t36:1	5.82	1274.892	[M+H]+	2.51E+05
GQ1b d34:1	4.23	1435.766	[M+2H]2+	4.77E+04
GQ1b d36:1	4.94	1449.785	[M+2H]2+	9.44E+06
GQ1b d38:1	5.66	1463.798	[M+2H]2+	9.47E+05
GT1b d34:0	4.59	1291.229	[M+2H]2+	2.10E+05
GT1b d34:1	4.38	1290.218	[M+2H]2+	1.68E+06
GT1b d36:1	5.12	1304.235	[M+2H]2+	1.10E+08
GT1b d36:2	4.54	1303.230	[M+2H]2+	2.40E+06
GT1b d38:1	5.78	1318.251	[M+2H]2+	2.53E+07
		Cerebellum QC		
GD1a d34:1	4.67	1144.670	[M+2H]2+	1.63E+06
GD1a d36:1	5.38	1158.687	[M+2H]2+	2.22E+07
GD1a d38:1	5.99	1172.702	[M+2H]2+	1.25E+07
GD1b d34:1	4.68	1025.073	[M+2H]2+	1.76E+06
GD1b d36:1	5.40	1039.088	[M+2H]2+	4.24E+07
GD1b d36:2	4.81	1038.084	[M+2H]2+	1.38E+06
GD1b d38:1	6.04	1053.106	[M+2H]2+	6.88E+06
GD1b t38:2	5.31	1060.095	[M+2H]2+	1.07E+06
GD2 d36:1	5.50	958.065	[M+2H]2+	6.44E+06
GD3 d36:1	5.77	856.524	[M+2H]2+	2.96E+07

GD3 d38:1	6.36	870.540	[M+2H]2+	3.10E+06
GD3 d42:2	6.84	897.563	[M+2H]2+	2.88E+06
GD3 t39:2	5.70	884.539	[M+2H]2+	2.03E+07
GM1a d34:1	5.02	879.528	[M+2H]2+	7.22E+06
GM1a d36:1	5.70	893.541	[M+2H]2+	1.43E+08
GM1a d36:2	5.14	892.536	[M+2H]2+	4.60E+06
GM1a d38:1	6.31	907.560	[M+2H]2+	3.91E+07
GM1a d42:2	6.79	934.583	[M+2H]2+	1.98E+06
GM2 d36:1	5.86	812.515	[M+2H]2+	5.09E+07
GM2 d38:1	6.44	826.532	[M+2H]2+	9.56E+06
GM3 d36:1	6.04	710.975	[M+2H]2+	3.32E+07
GM3 d38:1	6.62	724.993	[M+2H]2+	3.77E+06
GM3 d42:2	7.07	752.017	[M+2H]2+	3.85E+06
GM4 d36:1	5.94	1258.895	[M+H]+	8.07E+05
GM4 d38:1	6.52	1286.923	[M+H]+	3.51E+05
GM4 t36:1	5.82	1274.891	[M+H]+	4.59E+05
GQ1b d34:1	4.23	1435.767	[M+2H]2+	2.56E+05
GQ1b d36:1	4.96	1449.779	[M+2H]2+	5.23E+06
GQ1b t38:2	4.75	1470.788	[M+2H]2+	5.16E+05
GT1b d34:0	4.64	1291.228	[M+2H]2+	5.96E+05
GT1b d34:1	4.38	1290.219	[M+2H]2+	3.77E+06
GT1b d36:1	5.10	1304.234	[M+2H]2+	1.00E+08
GT1b d36:2	4.51	1303.230	[M+2H]2+	1.23E+06
GT1b d38:1	5.75	1318.248	[M+2H]2+	4.01E+07
GT1b d40:1	6.33	1332.265	[M+2H]2+	2.01E+06
GT1b t35:0	5.33	1306.244	[M+2H]2+	5.25E+06
Pons, Medulla QC				
GD1a d36:1	5.33	1158.687	[M+2H]2+	2.47E+06
GD1a d38:1	5.98	1172.703	[M+2H]2+	1.33E+06
GD1b d34:1	4.67	1025.075	[M+2H]2+	2.65E+05
GD1b d36:1	5.39	1039.088	[M+2H]2+	3.10E+07
GD1b d36:2	4.79	1038.081	[M+2H]2+	4.36E+05
GD1b d38:1	5.98	1053.105	[M+2H]2+	1.71E+07
GD1b d40:1	6.55	1067.126	[M+2H]2+	3.43E+05
GD1b d42:2	6.53	1080.133	[M+2H]2+	1.31E+05
GD1b t38:2	5.28	1060.094	[M+2H]2+	1.99E+05
GD2 d36:1	5.47	958.064	[M+2H]2+	2.57E+06
GD2 d38:1	6.09	972.082	[M+2H]2+	1.56E+06
GD3 d36:1	5.77	856.524	[M+2H]2+	4.59E+06
GD3 d38:1	6.33	870.540	[M+2H]2+	1.56E+06
GD3 d40:1	6.86	884.556	[M+2H]2+	1.51E+05

GD3 d41:1	7.10	891.564	[M+2H]2+	1.14E+05
GD3 d42:2	6.83	897.566	[M+2H]2+	3.38E+05
GD3 t39:2	5.68	884.539	[M+2H]2+	1.21E+07
GM1a d36:1	5.68	893.541	[M+2H]2+	9.79E+07
GM1a d36:2	5.13	892.536	[M+2H]2+	3.13E+06
GM1a d38:1	6.28	907.560	[M+2H]2+	2.50E+07
GM1a d40:1	6.80	921.575	[M+2H]2+	4.81E+05
GM1a d42:2	6.78	934.582	[M+2H]2+	5.35E+05
GM2 d36:1	5.82	812.518	[M+2H]2+	1.27E+07
GM2 d38:1	6.42	826.532	[M+2H]2+	4.58E+06
GM2 d40:1	6.91	840.550	[M+2H]2+	1.38E+05
GM2 d42:2	6.89	853.555	[M+2H]2+	7.43E+04
GM3 d38:1	6.58	724.993	[M+2H]2+	2.49E+05
GM4 d36:1	5.92	1258.894	[M+H]+	8.40E+04
GM4 d40:1	7.02	1314.958	[M+H]+	5.88E+04
GM4 t36:1	5.81	1274.889	[M+H]+	6.28E+04
GQ1b d36:1	4.95	1449.786	[M+2H]2+	3.64E+06
GQ1b d38:1	5.60	1463.796	[M+2H]2+	1.68E+06
GQ1b t38:2	4.74	1470.792	[M+2H]2+	7.49E+04
GT1b d34:1	4.36	1290.224	[M+2H]2+	1.10E+05
GT1b d36:0	5.30	1305.237	[M+2H]2+	3.02E+06
GT1b d36:1	5.09	1304.234	[M+2H]2+	3.31E+07
GT1b d36:2	4.48	1303.229	[M+2H]2+	1.92E+05
GT1b d38:1	5.74	1318.248	[M+2H]2+	2.04E+07
GT1b d38:2	5.17	1317.237	[M+2H]2+	2.31E+05
GT1b d40:1	6.31	1332.270	[M+2H]2+	8.22E+05
GT1b t37:0	5.93	1320.261	[M+2H]2+	1.20E+06
GT1b t38:0	5.76	1327.257	[M+2H]2+	2.02E+05
		Midbrain QC		
GD1a d34:1	4.67	1144.675	[M+2H]2+	1.28E+06
GD1a d36:1	5.38	1158.687	[M+2H]2+	9.86E+07
GD1a d38:1	5.98	1172.703	[M+2H]2+	1.53E+07
GD1a d40:1	6.52	1186.717	[M+2H]2+	1.13E+06
GD1b d34:1	4.67	1025.077	[M+2H]2+	2.12E+06
GD1b d36:1	5.38	1039.093	[M+2H]2+	1.08E+08
GD1b d36:2	4.79	1038.083	[M+2H]2+	5.29E+06
GD1b d38:1	6.02	1053.105	[M+2H]2+	2.87E+07
GD1b d40:1	6.58	1067.121	[M+2H]2+	1.66E+06
GD1b d42:2	6.52	1080.128	[M+2H]2+	6.67E+05
GD1b t38:2	5.31	1060.094	[M+2H]2+	2.02E+06
GD2 d36:1	5.48	958.063	[M+2H]2+	1.11E+07

GD2 d38:1	6.07	972.082	[M+2H]2+	3.80E+06
GD3 d36:1	5.77	856.524	[M+2H]2+	1.60E+07
GD3 d38:1	6.34	870.540	[M+2H]2+	4.42E+06
GD3 d40:1	6.85	884.555	[M+2H]2+	7.64E+05
GD3 d41:1	7.08	891.564	[M+2H]2+	8.47E+05
GD3 d42:2	6.82	897.566	[M+2H]2+	1.71E+06
GD3 t39:2	5.70	884.539	[M+2H]2+	3.37E+07
GM1a d34:1	5.00	879.528	[M+2H]2+	6.40E+06
GM1a d36:1	5.70	893.545	[M+2H]2+	2.44E+08
GM1a d36:2	5.12	892.535	[M+2H]2+	1.83E+07
GM1a d38:1	6.27	907.559	[M+2H]2+	8.95E+07
GM1a d39:1	6.55	914.566	[M+2H]2+	1.22E+06
GM1a d40:1	6.79	921.575	[M+2H]2+	2.91E+06
GM1a d42:2	6.77	934.582	[M+2H]2+	3.56E+06
GM1a d43:2	7.00	941.593	[M+2H]2+	2.41E+05
GM2 d36:1	5.83	812.517	[M+2H]2+	7.82E+07
GM2 d36:2	5.28	811.510	[M+2H]2+	3.64E+06
GM2 d38:1	6.42	826.532	[M+2H]2+	2.23E+07
GM3 d36:1	6.02	710.978	[M+2H]2+	2.43E+07
GM3 d38:1	6.61	724.993	[M+2H]2+	3.36E+06
GM3 d40:1	7.08	739.001	[M+2H]2+	6.73E+05
GM3 d42:2	7.06	752.016	[M+2H]2+	2.76E+06
GM4 d36:1	5.93	1258.894	[M+H]+	5.57E+05
GM4 d38:1	6.50	1286.926	[M+H]+	1.65E+05
GM4 d40:1	7.02	1314.957	[M+H]+	4.98E+05
GM4 t36:1	5.82	1274.890	[M+H]+	5.29E+05
GM4 t47:5	7.93	1420.999	[M+H]+	6.74E+04
GQ1b d34:1	4.20	1435.773	[M+2H]2+	7.44E+04
GQ1b d36:1	4.96	1449.786	[M+2H]2+	1.69E+07
GQ1b d36:2	4.34	1448.775	[M+2H]2+	2.61E+05
GQ1b d38:1	5.60	1463.795	[M+2H]2+	3.77E+06
GQ1b t38:2	4.73	1470.787	[M+2H]2+	4.78E+05
GT1b d34:1	4.35	1290.224	[M+2H]2+	1.28E+06
GT1b d36:1	5.09	1304.233	[M+2H]2+	1.78E+08
GT1b d36:2	4.48	1303.230	[M+2H]2+	4.36E+06
GT1b d38:1	5.73	1318.248	[M+2H]2+	6.64E+07
GT1b d40:1	6.32	1332.270	[M+2H]2+	4.91E+06
GT1b d41:1	6.58	1339.275	[M+2H]2+	1.81E+06
		Cortex QC		
GD1a d34:1	4.64	1144.674	[M+2H]2+	2.67E+06
GD1a d36:1	5.35	1158.686	[M+2H]2+	1.28E+08

GD1a d38:1	5.96	1172.702	[M+2H]2+	3.44E+07
GD1a d40:1	6.51	1186.722	[M+2H]2+	1.86E+06
GD1a t37:1	5.38	1173.691	[M+2H]2+	2.91E+07
GD1b d34:1	4.67	1025.077	[M+2H]2+	2.91E+06
GD1b d36:1	5.38	1039.088	[M+2H]2+	9.55E+07
GD1b d36:2	4.78	1038.083	[M+2H]2+	8.91E+06
GD1b d38:1	5.99	1053.105	[M+2H]2+	3.40E+07
GD1b d40:1	6.54	1067.121	[M+2H]2+	1.71E+06
GD1b t38:2	5.30	1060.094	[M+2H]2+	2.69E+06
GD3 d36:1	5.78	856.523	[M+2H]2+	1.97E+07
GD3 d38:1	6.35	870.539	[M+2H]2+	6.92E+06
GD3 d40:1	6.86	884.555	[M+2H]2+	1.70E+06
GD3 d41:1	7.08	891.564	[M+2H]2+	8.70E+05
GD3 d42:2	6.83	897.562	[M+2H]2+	1.90E+06
GD3 t39:2	5.69	884.539	[M+2H]2+	3.11E+07
GM1a d34:1	5.00	879.527	[M+2H]2+	8.50E+06
GM1a d36:1	5.69	893.544	[M+2H]2+	2.25E+08
GM1a d36:2	5.12	892.536	[M+2H]2+	2.14E+07
GM1a d38:1	6.28	907.559	[M+2H]2+	6.91E+07
GM1a d40:1	6.81	921.575	[M+2H]2+	2.36E+06
GM1a d42:2	6.78	934.582	[M+2H]2+	2.77E+06
GM2 d36:1	5.84	812.518	[M+2H]2+	5.80E+07
GM2 d36:2	5.28	811.510	[M+2H]2+	3.66E+06
GM2 d38:1	6.40	826.531	[M+2H]2+	1.52E+07
GM3 d36:1	6.04	710.978	[M+2H]2+	1.45E+07
GM3 d38:1	6.59	724.993	[M+2H]2+	5.70E+06
GM3 d40:1	7.08	739.010	[M+2H]2+	2.53E+06
GM3 d42:2	7.06	752.016	[M+2H]2+	3.16E+06
GQ1b d34:1	4.20	1435.772	[M+2H]2+	7.21E+04
GQ1b d36:1	4.94	1449.786	[M+2H]2+	8.99E+06
GT1b d34:1	4.34	1290.224	[M+2H]2+	2.63E+06
GT1b d36:1	5.08	1304.233	[M+2H]2+	2.41E+08
GT1b d36:2	4.48	1303.229	[M+2H]2+	8.77E+06
GT1b d38:1	5.73	1318.248	[M+2H]2+	5.97E+07
GT1b d40:1	6.30	1332.270	[M+2H]2+	4.60E+06
GT1b d41:1	6.59	1339.276	[M+2H]2+	1.41E+06
GT1b t35:0	5.30	1306.243	[M+2H]2+	6.04E+06
GT1b t35:2	3.59	1304.219	[M+2H]2+	4.72E+04

QC Reporter Ions

Table 17. Weighted Reporter Ion Relative Intensities in the Brain Region QCs.

QC	TMT126	TMT128	TMT130	TMT127	TMT129	TMT131
Cerebellum	50.8%	49.2%				
Midbrain		50.1%	49.9%			
Pons, medulla				48.2%	51.8%	
Cortex					48.3%	51.7%

Table 17 summarizes the most critical result of QC tandem mass spectrometry. The reporter ion relative intensity for each sample from the same region should be approximately equal, 50% for two samples, and these results are reasonably close. We expect that tighter error bars can be achieved through further physical optimization of the workflow.

GD1a Species Across and Ceramide Distribution Within Brain Regions

Table 18. GD1a Relative Quantification Across Brain Regions and Ceramides in Regions.

Major LCMS Peak	Weighted Relative Quantifications			
Ceramide Isomers	Detected, Identified		Not Detected	
Gangliosides	Cerebellum	Pons, Medulla	Midbrain	Cortex
GD1a d34:1	23 ± 2	2.7 ± 1.2	23.9 ± 0.9	50 ± 3
GD1a d18:1/16:0				
GD1a d36:1	8.1 ± 0.3	3.4 ± 1.4	31 ± 3	58 ± 4
GD1a d18:1/18:0				
GD1a d38:1	20 ± 3	9 ± 3	24.7 ± 1.1	46 ± 7
GD1a d18:1/20:0				18.0
GD1a d20:1/18:0				82.0
GD1a d40:1				
GD1a d18:1/22:0				59.8
GD1a d20:1/20:0				40.2
GD1a t37:1				
GD1a d18:1/m19:0				

Table 18 summarizes the results of GD1a analysis, one of ten ganglioside classes characterized. In the left-most column, in darker blue, are the major LC-MS peaks. In lighter

blue are the specific ceramide configurations detected within the peaks. On the right side are four columns, one for each brain region, where green indicates that this species was detected and confirmed through confident MS2 identification and red that it was not detected with confident MS2 identification. Boxed values indicate the local basis for relative quantification; each set of boxed values adds up to 100. Vertical boxes are tentative ceramide isomer quantifications within a QC. Horizontal boxes are the primary result of the analysis, relative quantification across brain regions with 95% confidence intervals. While ceramide isomer quantifications are not listed with confidence intervals, these values are inherently more confident due to the absence of mixing different sample types. The 9 tables to follow have the same information scheme.

The GD1a data is typical for analysis of gangliosides in the brain. There is both greater quantity and diversity of GD1a in the cortex, followed by the midbrain, and then the typically lower concentration and diversity regions, the cerebellum, pons, and medulla. There are perplexing trends in the ceramide content that would need a deeper metabolism study to explain, such as that within the cortex, LCB d20:1 is apparently more abundant than LCB d18:1 when paired with FA 18:0 but not when paired with FA 20:0. There is one exotic species in the set, GD1a t37:1, only identified in the cortex region, with a FA 19:0 X-OH. Care must be taken in designing automated protocols for making such identifications, as the mass difference caused by a fatty acid hydroxyl is identical to the mass difference caused by 1 Neu5Gc residue, both 1 oxygen. The absence of characteristic Neu5Gc MS2 ions allows for this identification.

GD1b Species Across and Ceramide Distribution Within Brain Regions

Table 19. GD1b Relative Quantification Across Brain Regions and Ceramides in Regions

Major LCMS Peak	Weighted Relative Quantifications			
Ceramide Isomers	Detected, Identified		Not Detected	
Gangliosides	Cerebellum	Pons, Medulla	Midbrain	Cortex
GD1b d34:1	25 ± 9	4 ± 13	37 ± 10	34 ± 11
GD1b d18:1/16:0				
GD1b d36:1	13.9 ± 0.6	10.71 ± 0.09	39 ± 11	37 ± 12
GD1b d18:1/18:0				
GD1b d36:2	8.48 ± 0.10	6 ± 4	40.4 ± 0.9	45 ± 4
GD1b d18:2/18:0				
GD1b d38:1	13.1 ± 0.7	17 ± 10	41 ± 12	29 ± 2
GD1b d18:1/20:0	20.1	14.3	19.7	14.6
GD1b d20:1/18:0	79.9	85.7	80.3	85.4
GD1b d40:1				
GD1b d18:1/22:0		35.2	46.6	43.9
GD1b d20:1/20:0		64.8	53.4	56.1
GD1b d42:2				
GD1b d18:1/24:1				
GD1b t38:2	16*	10*	25*	49*
GD1b d18:1/m20:0				

Next is GD1b. The ability to resolve the GD1a and GD1b glycan isomers is a strength of the underlying method in which GD1a and GD1b lead to different products.[16] The asterisked values on the GD1b t38:2 row indicate that this ganglioside was only MS2 scanned in one multiplex, being of very low abundance. Different ceramide distribution between regions is an interesting outcome. Some studies have suggested that the d18:1-d20:1 LCB paradigm between tissues indicates cell age and degree of differentiation.[33]

GD2 Species Across and Ceramide Distribution Within Brain Regions

Table 20. GD2 Relative Quantification Across Brain Regions and Ceramides in Regions.

Major LCMS Peak	Weighted Relative Quantifications			
Ceramide Isomers	Detected, Identified		Not Detected	
Gangliosides	Cerebellum	Pons, Medulla	Midbrain	Cortex
GD2 d36:1				
GD2 d18:1/18:0				
GD2 d38:1				
GD2 d18:1/20:0		17.4	22.4	
GD2 d20:1/18:0		82.6	77.6	

While still one of the top 10 glycans, GD2 is very low abundance in mammalian brains, its increase often associated with cancer pathenogenesis.[80, 81] As this was a normal mouse brain, relatively low abundance of GD2 is expected.

GD3 Species Across and Ceramide Distribution Within Brain Regions

Table 21. GD3 Relative Quantification Across Brain Regions and Ceramides in Regions.

Major LCMS Peak	Weighted Relative Quantifications			
Ceramide Isomers	Detected, Identified		Not Detected	
Gangliosides	Cerebellum	Pons, Medulla	Midbrain	Cortex
GD3 d36:1	37 ± 2	14 ± 2	24 ± 3	25 ± 7
GD3 d18:0/18:1	0.6			
GD3 d18:1/18:0	99.4			
GD3 d38:1	28 ± 12	15.3 ± 0.2	24 ± 11	32 ± 22
GD3 d18:1/20:0	45.3	34.2	37.8	43.5
GD3 d20:1/18:0	54.7	65.8	62.2	56.5
GD3 d40:1	37 ± 4	14 ± 8	23 ± 5	26 ± 9
GD3 d18:1/22:0				92.4
GD3 d20:1/20:0				7.6
GD3 d41:1	31 ± 4	15 ± 9	29 ± 12	25 ± 17
GD3 d18:1/23:0				
GD3 d42:2				
GD3 d18:1/24:1				
GD3 t39:2				
GD3 d18:0/m21:2	0.7	0.5	0.8	0.4
GD3 d18:1/m21:1	99.3	99.5	99.2	99.6

While GD2 is so scarce in the mouse brain that it nearly avoided detection, there are far more GD3 hits. In GD3 are the first examples where the cortex is not the main contributor. The cortex and midbrain historically have 1 – 3 times more total ganglioside mass, but they of course will not have more of every individual ganglioside. GD3 is believed to be involved in early vertebrate tissue development, being more highly expressed in developing embryonic brains. It would be interesting if the trend seen here related to developmental stages of the vertebrate brain.

GM1a Species Across and Ceramide Distribution Within Brain Regions

Table 22. GM1a Relative Quantification Across Brain Regions and Ceramides in Regions.

Major LCMS Peak	Weighted Relative Quantifications			
Ceramide Isomers	Detected, Identified		Not Detected	
Gangliosides	Cerebellum	Pons, Medulla	Midbrain	Cortex
GM1a d34:1	15.8 ± 1.0	16 ± 7	31.2 ± 0.2	37 ± 6
GM1a d16:1/18:0	6.7		14.2	8.1
GM1a d18:1/16:0	93.3		85.8	91.9
GM1a d36:1	12 ± 2	22 ± 7	35 ± 3	30 ± 11
GM1a d18:1/18:0				
GM1a d36:2				
GM1a d18:1/18:1	43.8		8.1	10.4
GM1a d18:2/18:0	56.2		91.9	89.6
GM1a d38:1	15.2 ± 0.5	31.5 ± 0.3	34 ± 5	20 ± 4
GM1a d18:1/20:0		9.8	9.8	15.2
GM1a d20:1/18:0		90.2	90.2	84.8
GM1a d39:1				
GM1a d18:1/21:0			88.1	
GM1a d20:1/19:0			11.9	
GM1a d40:1				
GM1a d18:1/22:0		42.8	54.0	62.7
GM1a d20:1/20:0		57.2	46.0	37.3
GM1a d42:2	17 ± 8	32 ± 18	30 ± 3	21 ± 6
GM1a d18:1/24:1				
GM1a d43:2				
GM1a d18:1/25:1				
GM1a t41:1	16*	31*	24*	30*

For GM1a, there are more ceramide comparisons, in which there is wide variation even between ganglioside species in the same region, again pointing to different metabolic pathways.

GM2 Species Across and Ceramide Distribution Within Brain Regions

Table 23. GM2 Relative Quantification Across Brain Regions and Ceramides in Regions.

Major LCMS Peak	Weighted Relative Quantifications			
Ceramide Isomers	Detected, Identified		Not Detected	
Gangliosides	Cerebellum	Pons, Medulla	Midbrain	Cortex
GM2 d36:1	20 ± 6	27 ± 9	31.4 ± 1.2	21 ± 2
GM2 d18:0/18:1			0.6	
GM2 d18:1/18:0			99.4	
GM2 d36:2	9 ± 3	12.5 ± 1.4	27.5 ± 0.6	51 ± 4
GM2 d18:1/18:1			16.2	24.5
GM2 d18:2/18:0			83.8	75.5
GM2 d38:1	20 ± 2	37 ± 7	30.3 ± 1.2	12 ± 4
GM2 d18:1/20:0	9.7	8.9	9.0	13.4
GM2 d20:1/18:0	90.3	91.1	91.0	86.6
GM2 d40:1				
GM2 d20:1/20:0				
GM2 d42:2				
GM2 d18:1/24:1				

GM2 d38:1 presents one of very few species found at higher concentration in the pons, medulla fraction. Pons, medulla also appears to have the greater diversity of GM2. However, this could just be due to other regions not getting affirmative MS2 scans due to overlapping elution with high abundance species.

GM3 Species Across and Ceramide Distribution Within Brain Regions

Table 24. GM3 Relative Quantification Across Brain Regions and Ceramides in Regions.

Major LCMS Peak	Weighted Relative Quantifications			
Ceramide Isomers	Detected, Identified		Not Detected	
Gangliosides	Cerebellum	Pons, Medulla	Midbrain	Cortex
GM3 d36:1	33 ± 3	7.5 ± 0.7	31 ± 3	29 ± 7
GM3 d18:0/18:1	0.5		0.6	0.7
GM3 d18:1/18:0	99.5		99.4	99.3
GM3 d38:1	37.78 ± 0.14	15 ± 5	24.1 ± 0.9	23 ± 6
GM3 d18:1/20:0	72.6		46.1	58.5
GM3 d20:1/18:0	27.4		53.9	41.5
GM3 d40:1				
GM3 d18:1/22:0				
GM3 d42:2				
GM3 d18:1/24:1				

GM3 presents more cases of higher concentration outside the cortex and midbrain. Once again, the ratio between d18:1 and d20:1 appears to be tissue specific.

GM4 Species Across and Ceramide Distribution Within Brain Regions

Table 25. GM4 Relative Quantification Across Brain Regions and Ceramides in Regions.

Major LCMS Peak	Weighted Relative Quantifications			
Ceramide Isomers	Detected, Identified		Not Detected	
Gangliosides	Cerebellum	Pons, Medulla	Midbrain	Cortex
GM4 d36:1	23*	41*	23*	12*
GM4 d38:1	35 ± 28	20 ± 57	38 ± 52	8 ± 23
GM4 d40:1				
GM4 t36:1	16 ± 4	12 ± 6	27 ± 4	45 ± 6
GM4 t47:5				

GM4 ceramide analysis is a casualty of transitioning from standards to biological samples. While GM4 2+ ions were appreciable in standards, the biological matrix has promoted their 1+ ions, which are less descriptive in fragmentation. GM4 is the main reason why the LC-MS method has multiple scanning loops. They were added to catch these GM4 1+ species. Though there is capture of isobaric data, the ceramide fragments become unreliable in the 1+ charge state. As in the GM4 standard, GM4 more commonly has a fatty acid hydroxyl, making GM4 t36:1 higher abundance and more reliably quantified than GM4 d36:1, which has very large confidence intervals.

GQ1b Species Across and Ceramide Distribution Within Brain Regions

Table 26. GQ1b Relative Quantification Across Brain Regions and Ceramides in Regions.

Major LCMS Peak	Weighted Relative Quantifications			
Ceramide Isomers	Detected, Identified		Not Detected	
Gangliosides	Cerebellum	Pons, Medulla	Midbrain	Cortex
GQ1b d34:1	48*	0	24*	28*
GQ1b d36:1	22.6 ± 0.4	12 ± 3	43.4 ± 1.3	22 ± 4
GQ1b d18:1/18:0				
GQ1b d36:2				
GQ1b d38:1	21*	21*	42*	16*
GQ1b d20:1/18:0				
GQ1b t38:2				
GQ1b d18:1/m20:0				

GQ1b along with GT1b representation is another strength of the method. While some methods report only monosialogangliosides[76], and others track disialogangliosides, few can cover mono-, di, tri-, and tetrasialogangliosides, and these are typically without[17] derivatization.

GT1b Species Across and Ceramide Distribution Within Brain Regions

Table 27. GT1b Relative Quantification Across Brain Regions and Ceramides in Regions.

Major LCMS Peak	Weighted Relative Quantifications			
Ceramide Isomers	Detected, Identified		Not Detected	
Gangliosides	Cerebellum	Pons, Medulla	Midbrain	Cortex
GT1b d34:0	65 ± 3	0	16 ± 8	19 ± 11
GT1b d34:1	50.1 ± 0.7	3 ± 3	19 ± 2	28 ± 6
GT1b d18:1/16:0				
GT1b d36:0				
GT1b d18:0/18:0				
GT1b d36:1	19.1 ± 0.4	8 ± 3	34 ± 3	39 ± 7
GT1b d18:1/18:0				
GT1b d36:2	9.08 ± 0.14	3 ± 2	34 ± 2	54 ± 4
GT1b d18:2/18:0				
GT1b d38:1	22.9 ± 0.3	15.77 ± 0.10	33 ± 5	28 ± 5
GT1b d18:1/20:0			20.5	13.8
GT1b d20:1/18:0			79.5	86.2
GT1b d40:1				
GT1b d41:1				
GT1b t35:0				
GT1b d18:0/m17:0				
GT1b t35:2				
GT1b t37:0				
GT1b t38:0				

GT1b is surprisingly diverse in the mouse brain. Better ceramide description can be achieved by performing MS2 of 3+ ions, but we found that there were too many 2+ ions to reliably get 3+ scans unless yet another scanning loop was added. When rescanning was tried with a dedicated 3+ fragmentation loop, overall results worsened such that this method version was not pursued further.

Conclusions

Isobaric analysis represents an innovative way of dealing with common sources of variation in the analysis of biological samples while increasing throughput. This derivatization increases the ionization and fragmentation quality of gangliosides such that detailed annotation is possible. After refining the original isobaric analysis method, we applied it to biological samples and found 60 major LC-MS peaks containing a total of 79 ganglioside molecular species while performing relative quantification between brain regions and of ceramide content distribution. These modes of quantification are of potential use for biological study and comparative metabolomics. Considering the relatively short gradient time and small sample mass, 60 peaks is a testament to the capabilities of high-performance LC-MS instrumentation. Greater ganglioside coverage could be achieved through larger sample mass or through longer gradient time. While there is still room for improvement, this work will inform future study of gangliosides in mice models.

CHAPTER VI: PERSPECTIVES

The first challenge encountered in high sample count ganglioside analysis is the labor cost of identification. Confident and descriptive identification requires scrutiny of the MS2 spectrum with care given to specific ions as well as their intensities. In the absence of curated libraries and automated annotation, this challenge leads to limited ganglioside coverage or low throughput. We decided to allocate significant research and development time to the task of writing ganglioside library generating software to later increase throughput and coverage. We were not sure if that decision would pay off, and it was a significant risk with the reduction in time for method development. However, even the early prototypes of the software expedited method development through assistance with reaction monitoring.

With the library, we were able to fully explore stable-isotope strategies. Isotopic labeling has a quantitative obstacle in isotopic overlap. Heavy tags with greater shifts, such as cholamine-d9 (M+9) can address this problem. We attempted to synthesize and use this heavier tag but struggled to get high purity for reliable 1:1 reactivity vs the light tag. MS-DIAL was able to handle this isotopic shift and correctly paired the peaks, suggesting this is a viable solution with a high purity reagent. Simultaneously exploring a more generalized solution, we developed PPP for all metabolomics isotopic labeling with its algorithmic correction of isotopic overlap.

Following the difficulties in applying cholamine to biological samples, we switched to the isobaric approach. These methodologies were both pursued throughout the thesis work, each with its own strengths and weaknesses. The isobaric method worked much better with the biological samples prepared, even better than it had performed with the standards. Future work should further explore this derivatization, possibly generating comprehensive ganglioside standards for absolute quantification through oxidative clipping of the glycerol arm.

REFERENCES

1. Abramowicz, M., *The Human Genome Project in retrospect.* Adv Genet, 2003. **50**: p. 231-61; discussion 507-10.
2. Zhao, Y.Y. and R.C. Lin, *UPLC-MS(E) application in disease biomarker discovery: the discoveries in proteomics to metabolomics.* Chem Biol Interact, 2014. **215**: p. 7-16.
3. Claudino, W.M., et al., *Metabolomics in cancer: a bench-to-bedside intersection.* Crit Rev Oncol Hematol, 2012. **84**(1): p. 1-7.
4. Kordalewska, M. and M.J. Markuszewski, *Metabolomics in cardiovascular diseases.* J Pharm Biomed Anal, 2015. **113**: p. 121-36.
5. Hao, L., et al., *Comparative Evaluation of MS-based Metabolomics Software and Its Application to Preclinical Alzheimer's Disease.* Sci Rep, 2018. **8**(1): p. 9291.
6. Agin, A.H., D.; Ruhland, E.; Chao de la Barca, J. M.; Zumsteg, J.; Moal, V.; Gauchez, A. S.; Namer, I. J.; , *MEtabolomics - an overview. From basic principles to potential biomarkers (part 1).* Med. Nucleaire. , 2016. **40**: p. 4-10.
7. Dufour-Rainfray, D.L., M.; Boulard, P.; Guidotti, M.; Delaye, J.-B.; Ribeiro, M.-J.; Gauchez, A.-S.; Balageas, A.-C.; Emond, P.; Agin, A., *Metabolomics - an overview. From basic principles to potential biomarkers (part 2).* Med. Nucleaire., 2020. **44**: p. 158-163.
8. Fuhrer, T. and N. Zamboni, *High-throughput discovery metabolomics.* Curr Opin Biotechnol, 2015. **31**: p. 73-8.
9. Duhrkop, K., et al., *SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information.* Nat Methods, 2019. **16**(4): p. 299-302.
10. van Meer, G. and A.I. de Kroon, *Lipid map of the mammalian cell.* J Cell Sci, 2011. **124**(Pt 1): p. 5-8.
11. Rakusanova, S.F., O.; Cajka, T.; , *Toward building mass spectrometry-based metabolomics and lipidomics atlases for biological and clinical research.* Trends Anal. Chem. , 2023. **158**: p. 116825.
12. Patel, K.N., et al., *Introduction to hyphenated techniques and their applications in pharmacy.* Pharm Methods, 2010. **1**(1): p. 2-13.
13. Svennerholm, L., *Designation and schematic structure of gangliosides and allied glycosphingolipids.* Prog Brain Res, 1994. **101**: p. XI-XIV.
14. Harris, R.A., et al., *Determining Double Bond Position in Lipids Using Online Ozonolysis Coupled to Liquid Chromatography and Ion Mobility-Mass Spectrometry.* Anal Chem, 2018. **90**(3): p. 1915-1924.
15. Liebisch, G., et al., *Quantitative measurement of different ceramide species from crude cellular extracts by electrospray ionization tandem mass spectrometry (ESI-MS/MS).* J Lipid Res, 1999. **40**(8): p. 1539-46.
16. Barrientos, R.C. and Q. Zhang, *Isobaric Labeling of Intact Gangliosides toward Multiplexed LC-MS/MS-Based Quantitative Analysis.* Anal Chem, 2018. **90**(4): p. 2578-2586.
17. Li, Z. and Q. Zhang, *Ganglioside isomer analysis using ion polarity switching liquid chromatography-tandem mass spectrometry.* Anal Bioanal Chem, 2021. **413**(12): p. 3269-3279.

18. Li, H., et al., *Combinatory Data-Independent Acquisition and Parallel Reaction Monitoring Method for Deep Profiling of Gangliosides.* Anal Chem, 2020. **92**(15): p. 10830-10838.
19. Barrientos, R.C. and Q. Zhang, *Recent advances in the mass spectrometric analysis of glycosphingolipidome - A review.* Anal Chim Acta, 2020. **1132**: p. 134-155.
20. Sipione, S., et al., *Gangliosides in the Brain: Physiology, Pathophysiology and Therapeutic Applications.* Front Neurosci, 2020. **14**: p. 572965.
21. Muthing, J., et al., *Isolation and structural characterization of fucosylated gangliosides with linear poly-N-acetyllactosaminyl chains from human granulocytes.* Glycobiology, 1996. **6**(2): p. 147-56.
22. Reiding, K.R., et al., *High-throughput profiling of protein N-glycosylation by MALDI-TOF-MS employing linkage-specific sialic acid esterification.* Anal Chem, 2014. **86**(12): p. 5784-93.
23. de Haan, N., et al., *Linkage-specific sialic acid derivatization for MALDI-TOF-MS profiling of IgG glycopeptides.* Anal Chem, 2015. **87**(16): p. 8284-91.
24. Holst, S., et al., *Linkage-Specific in Situ Sialic Acid Derivatization for N-Glycan Mass Spectrometry Imaging of Formalin-Fixed Paraffin-Embedded Tissues.* Anal Chem, 2016. **88**(11): p. 5904-13.
25. Nishikaze, T., et al., *Differentiation of Sialyl Linkage Isomers by One-Pot Sialic Acid Derivatization for Mass Spectrometry-Based Glycan Profiling.* Anal Chem, 2017. **89**(4): p. 2353-2360.
26. Hanamatsu, H., et al., *Sialic Acid Linkage Specific Derivatization of Glycosphingolipid Glycans by Ring-Opening Aminolysis of Lactones.* Anal Chem, 2018. **90**(22): p. 13193-13199.
27. Martin, S.W., B.J. Glover, and J.M. Davies, *Lipid microdomains--plant membranes get organized.* Trends Plant Sci, 2005. **10**(6): p. 263-5.
28. Halstead, S.K., et al., *Serum anti-GM2 and anti-GalNAc-GD1a IgG antibodies are biomarkers for acute canine polyradiculoneuritis.* J Small Anim Pract, 2022. **63**(2): p. 104-112.
29. Li, Y.T., et al., *Association of GM4 ganglioside with the membrane surrounding lipid droplets in shark liver.* J Lipid Res, 2002. **43**(7): p. 1019-25.
30. Irvine, R.A. and T.N. Seyfried, *Phylogenetic conservation of ganglioside GD3 expression during early vertebrate ontogeny.* Comp Biochem Physiol B Biochem Mol Biol, 1994. **109**(4): p. 603-12.
31. Cherry, J.M.B., T. H.; Morre, D. J.; , *The absence of gangliosides in a higher plant.* Experientia, 1978. **34**: p. 1433-1434.
32. Chiricozzi, E., et al., *GM1 Ganglioside Is A Key Factor in Maintaining the Mammalian Neuronal Functions Avoiding Neurodegeneration.* Int J Mol Sci, 2020. **21**(3).
33. Sonnino, S. and V. Chigorno, *Ganglioside molecular species containing C18- and C20-sphingosine in mammalian nervous tissues and neuronal cell cultures.* Biochim Biophys Acta, 2000. **1469**(2): p. 63-77.
34. Yu, R.K., et al., *Pathological roles of ganglioside mimicry in Guillain-Barre syndrome and related neuropathies.* Adv Exp Med Biol, 2011. **705**: p. 349-65.
35. Schnaar, R.L., *Glycolipid-mediated cell-cell recognition in inflammation and nerve regeneration.* Arch Biochem Biophys, 2004. **426**(2): p. 163-72.

36. Da Silva, J.S., et al., *Asymmetric membrane ganglioside sialidase activity specifies axonal fate.* Nat Neurosci, 2005. **8**(5): p. 606-15.
37. Ha, S.H., et al., *Exogenous and Endogeneous Disialosyl Ganglioside GD1b Induces Apoptosis of MCF-7 Human Breast Cancer Cells.* Int J Mol Sci, 2016. **17**(5).
38. Dalton, G., et al., *Soluble klotho binds monosialoganglioside to regulate membrane microdomains and growth factor signaling.* Proc Natl Acad Sci U S A, 2017. **114**(4): p. 752-757.
39. Posse de Chaves, E. and S. Sipione, *Sphingolipids and gangliosides of the nervous system in membrane function and dysfunction.* FEBS Lett, 2010. **584**(9): p. 1748-59.
40. Wu, G., et al., *Ganglioside GM1 deficiency in effector T cells from NOD mice induces resistance to regulatory T-cell suppression.* Diabetes, 2011. **60**(9): p. 2341-9.
41. Das, A., et al., *Gangliosides are essential endosomal receptors for quasi-enveloped and naked hepatitis A virus.* Nat Microbiol, 2020. **5**(9): p. 1069-1078.
42. Sjogren, B. and P. Svenningsson, *Depletion of the lipid raft constituents, sphingomyelin and ganglioside, decreases serotonin binding at human 5-HT7(a) receptors in HeLa cells.* Acta Physiol (Oxf), 2007. **190**(1): p. 47-53.
43. Sezgin, E., et al., *The mystery of membrane organization: composition, regulation and roles of lipid rafts.* Nat Rev Mol Cell Biol, 2017. **18**(6): p. 361-374.
44. Villar, V.C., S.; Zheng, X.; Jose, P.;, *Localization and signaling of GPCRs in lipid rafts.* Methods Cell Biol. , 2016. **132**.
45. Alpaugh, M., et al., *Disease-modifying effects of ganglioside GM1 in Huntington's disease models.* EMBO Mol Med, 2017. **9**(11): p. 1537-1557.
46. Svennerholm, L., et al., *Alzheimer disease - effect of continuous intracerebroventricular treatment with GM1 ganglioside and a systematic activation programme.* Dement Geriatr Cogn Disord, 2002. **14**(3): p. 128-36.
47. Svennerholm, L., et al., *Membrane lipids in the aging human brain.* J Neurochem, 1991. **56**(6): p. 2051-9.
48. Zhang, J.Z., et al., *Monosialotetrahexosy-1 ganglioside attenuates diabetes-enhanced brain damage after transient forebrain ischemia and suppresses phosphorylation of ERK1/2 in the rat brain.* Brain Res, 2010. **1344**: p. 200-8.
49. Annunziata, I., R. Sano, and A. d'Azzo, *Mitochondria-associated ER membranes (MAMs) and lysosomal storage diseases.* Cell Death Dis, 2018. **9**(3): p. 328.
50. van Echten-Deckert, G. and M. Gurgui, *Golgi localization of glycosyltransferases involved in ganglioside biosynthesis.* Curr Drug Targets, 2008. **9**(4): p. 282-91.
51. Lee, J., et al., *Comprehensive Profiling of Surface Gangliosides Extracted from Various Cell Lines by LC-MS/MS.* Cells, 2019. **8**(11).
52. Cavdarli, S., et al., *Profiling of O-acetylated Gangliosides Expressed in Neuroectoderm Derived Cells.* Int J Mol Sci, 2020. **21**(1).
53. Itokazu, Y., Y.T. Tsai, and R.K. Yu, *Epigenetic regulation of ganglioside expression in neural stem cells and neuronal cells.* Glycoconj J, 2017. **34**(6): p. 749-756.
54. Fantini, J. and F.J. Barrantes, *Sphingolipid/cholesterol regulation of neurotransmitter receptor conformation and function.* Biochim Biophys Acta, 2009. **1788**(11): p. 2345-61.
55. Tsugawa, H., et al., *MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis.* Nat Methods, 2015. **12**(6): p. 523-6.

56. Koopman, J. and S. Grimme, *Calculation of Mass Spectra with the QCxMS Method for Negatively and Multiply Charged Molecules.* J Am Soc Mass Spectrom, 2022. **33**(12): p. 2226-2242.
57. Ruttkies, C., et al., *MetFrag relaunched: incorporating strategies beyond in silico fragmentation.* J Cheminform, 2016. **8**: p. 3.
58. Allen, F., et al., *CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra.* Nucleic Acids Res, 2014. **42**(Web Server issue): p. W94-9.
59. Lin, Y.M., C.T. Chen, and J.M. Chang, *MS2CNN: predicting MS/MS spectrum based on protein sequence using deep convolutional neural networks.* BMC Genomics, 2019. **20**(Suppl 9): p. 906.
60. Kirik, U., J.C. Refsgaard, and L.J. Jensen, *Improving Peptide-Spectrum Matching by Fragmentation Prediction Using Hidden Markov Models.* J Proteome Res, 2019. **18**(6): p. 2385-2396.
61. Bailey, L.S., et al., *Characterization of Glycosphingolipids and Their Diverse Lipid Forms through Two-Stage Matching of LC-MS/MS Spectra.* Anal Chem, 2021. **93**(6): p. 3154-3162.
62. Ceroni, A., et al., *GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans.* J Proteome Res, 2008. **7**(4): p. 1650-9.
63. Tayyari, F., et al., *15N-cholamine--a smart isotope tag for combining NMR- and MS-based metabolite profiling.* Anal Chem, 2013. **85**(18): p. 8715-21.
64. Haverkamp, J., J.P. Kamerling, and J.F. Vliegenthart, *Methylation analysis determination of acylneuraminic acid residue type 2 leads to 8 glycosidic linkage. Application to GT 1b ganglioside and colominic acid.* FEBS Lett, 1977. **73**(2): p. 215-9.
65. Bouchal, P., et al., *Biomarker discovery in low-grade breast cancer using isobaric stable isotope tags and two-dimensional liquid chromatography-tandem mass spectrometry (iTRAQ-2DLC-MS/MS) based quantitative proteomic analysis.* J Proteome Res, 2009. **8**(1): p. 362-73.
66. Hao, L., et al., *Metandem: An online software tool for mass spectrometry-based isobaric labeling metabolomics.* Anal Chim Acta, 2019. **1088**: p. 99-106.
67. Monroe, M.E., et al., *MASIC: a software program for fast quantitation and flexible visualization of chromatographic profiles from detected LC-MS(/MS) features.* Comput Biol Chem, 2008. **32**(3): p. 215-7.
68. Hao, L., et al., *Relative quantification of amine-containing metabolites using isobaric N,N-dimethyl leucine (DiLeu) reagents via LC-ESI-MS/MS and CE-ESI-MS/MS.* Analyst, 2015. **140**(2): p. 467-75.
69. Capellades, J., et al., *geoRge: A Computational Tool To Detect the Presence of Stable Isotope Labeling in LC/MS-Based Untargeted Metabolomics.* Anal Chem, 2016. **88**(1): p. 621-8.
70. Chokkathukalam, A., et al., *mzMatch-ISO: an R tool for the annotation and relative quantification of isotope-labelled mass spectrometry data.* Bioinformatics, 2013. **29**(2): p. 281-3.
71. Zhou, R., et al., *IsoMS: automated processing of LC-MS data generated by a chemical isotope labeling metabolomics platform.* Anal Chem, 2014. **86**(10): p. 4675-9.

72. Guo, K. and L. Li, *Differential 12C-/13C-isotope dansylation labeling and fast liquid chromatography/mass spectrometry for absolute and relative quantification of the metabolome.* Anal Chem, 2009. **81**(10): p. 3919-32.
73. Wang, S., et al., *MS-IDF: A Software Tool for Nontargeted Identification of Endogenous Metabolites after Chemical Isotope Labeling Based on a Narrow Mass Defect Filter.* Anal Chem, 2022. **94**(7): p. 3194-3202.
74. Gahmberg, C.G. and L.C. Andersson, *Selective radioactive labeling of cell surface sialoglycoproteins by periodate-tritiated borohydride.* J Biol Chem, 1977. **252**(16): p. 5888-94.
75. Yang, R., et al., *Quantitative analysis of n-3 polyunsaturated fatty acids and their metabolites by chemical isotope labeling coupled with liquid chromatography - mass spectrometry.* J Chromatogr B Analyt Technol Biomed Life Sci, 2021. **1172**: p. 122666.
76. Huang, Q.L., D.; Xin, B.; Cechner, K.; Zhou, X.; Wang, H.; Zhou, A. , *Quantification of monosialogangliosides in human plasma through chemical derivatization for signal enhancement in LC-ESI-MS.* Anal. Chim. Acta., 2016. **929**: p. 31-38.
77. Andrews, A.T.H., G. M.; Kent, P. W., *The periodate oxidation of bovine bone sialoprotein, and some observations on its structure.* Biochem. J. , 1969. **111**(5): p. 621-627.
78. Reuter, G., et al., *A detailed study of the periodate oxidation of sialic acids in glycoproteins.* Glycoconj J, 1989. **6**(1): p. 35-44.
79. Land, H. and M.S. Humble, *YASARA: A Tool to Obtain Structural Guidance in Biocatalytic Investigations.* Methods Mol Biol, 2018. **1685**: p. 43-67.
80. Yesmin, F., et al., *Extracellular vesicles released from ganglioside GD2-expressing melanoma cells enhance the malignant properties of GD2-negative melanomas.* Sci Rep, 2023. **13**(1): p. 4987.
81. Inagaki, F.F., et al., *Disialoganglioside GD2-Targeted Near-Infrared Photoimmunotherapy (NIR-PIT) in Tumors of Neuroectodermal Origin.* Pharmaceutics, 2022. **14**(10).