MULTIDIMENSIONAL ANALYSES FOR HIGH-CONFIDENCE UNTARGETED MOLECULAR ANNOTATION USING STRUCTURALLY SPECIFIC ION MOBILITY-MASS SPECTROMETRY

By

Bailey S. Rose

Dissertation Submitted to the

Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry

May 13, 2022

Nashville, Tennessee

Approved:

John A. McLean, Ph.D.

Renã A. S. Robinson, Ph.D.

Michael P. Stone, Ph.D.

Kevin L. Schey, Ph.D.

This work is dedicated to educators of all levels.

Thank you for your passion and inspiration.

ACKNOWLEDGMENTS

Finally, I must also acknowledge my cat, Eleanor, for her steady companionship as I worked from home for the past few years.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1

THE POTENTIAL OF COLLISION CROSS SECTION AS A MOLECULAR DESCRIPTOR

IN ION MOBIITY-MASS SPECTROMETRY[*]

1.1   Introduction

From the central dogma of molecular biology, studies of genomics, transcriptomics, and

proteomics provide higher order information about gene and protein expression to better

understand implicated phenotypes.[1,2] However, these approaches provide limited information

about real-time production of chemical species related to cellular metabolism as a function of

external stimuli or phenotype of interest. To address the need for rapid characterization of cellular

metabolism, metabolomics seeks to uncover information on a molecular basis by examining

expressed cellular products that can be correlated with a specific phenotype, stimuli, or other

experimental conditions.[3]

While several analytical approaches have been utilized to study metabolism and related cellular

processes (*e.g.* NMR, electrochemistry, etc.),[4,5] mass spectrometry (MS) has gained wide-spread

adoption as a result of its high throughput, low limits of detection, and molecular specificity. Mass

spectrometers can collect chemical information on the microsecond (μs) time scale,[6] and with the

rise of high-resolution, accurate mass techniques such as time of flight (TOF), Orbitrap, and ion

cyclotron instruments, a unique chemical formula can often be generated based solely on mass measurement for a specific analyte signal.[7,8] While identifying a specific chemical formula is advantageous, many metabolic pathways include isomeric molecules covering a range of biological classes, such as carbohydrates (*e.g.* glucose/galactose),[9] nucleosides (*e.g.* adenosine/deoxyguanosine), and lipids (7-dehydrocholesterol/desmosterol).[10] As biological function follows molecular structure, characterization of isomeric species is imperative for complete molecular identification and accurate pathway analysis. In many MS experiments, fragmentation techniques such as collision induced dissociation (CID) or electron transfer dissociation (ETD) are utilized to provide structural information about a specific analyte measured in the study.[11,12] However, as many metabolite isomers are less than 300 Dalton, these compounds often possess identical fragmentation spectra at similar energy thresholds and hence molecular fingerprinting by MS/MS and high resolution precursor mass is often not specific enough to identify a unique molecular structure.[13] Furthermore, as quadrupoles isolate on nominal mass, molecules with different molecular formulas but similar exact mass (*i.e.* nominal mass isobars) cannot be isolated, thereby complicating MS/MS analysis.[14] To address these challenges, pre-separation techniques such as gas and liquid chromatography,[15,16] and more recently ion mobility (IM),[17] have been interfaced prior to mass analysis to provide enhanced structural recognition and increased analyte coverage.

With the advent of commercially available ion mobility-mass spectrometers in 2006,[18] collision cross section (CCS) has become an additional molecular descriptor for untargeted experiments. CCS measurements are being standardized across instrumental platforms using rigid experimental protocols, and as such provide a molecular descriptor independent of system settings which are transferable between laboratories.[17,19–21] These collected CCS measurements provide the

2

capability to distinguish isomeric species in complex mixtures, provided enough resolution is accessible in the IM dimension.[22] In a study of a library of 554 primary metabolite standards (Mass Spectrometry Metabolite Library of Standards, MSMLS), a database of 1246 CCS values measured using uniform field drift tube (DT) IM-MS was established to facilitate chemical identification in untargeted metabolomic workflows. Detailed descriptions of standard preparation, analysis, and database construction from this study can be found in Appendix B, Section B1. The database was then further analyzed to provide insight into the potential benefit of ion mobility as an added separation dimension and molecular descriptor in metabolomics.

1.2   IM-derived CCS for molecular fingerprinting

1.2.1   Beyond mass resolving power: enhancing selectivity

Isomeric compounds are ubiquitous in metabolomic processes across a wide range of biological classes. Of the more than 500 compounds in the MSMLS library, almost one-third (31%) have a chemical formula in common with another compound, forming an isomeric pair. While recent advances in ion mobility have provided increased separation coverage of isomeric species, current untargeted metabolomic workflows identification is based first on primary mass measurement and subsequently supported with additional descriptors such as retention time, isotope ratios, and fragmentation matching. From this viewpoint, it is imperative to describe how much resolving power in the mass dimension is necessary for metabolomic studies.

In the analysis of the MSMLS library, we found that most analytes (64%) are resolvable based only on the mass dimension utilizing 40,000 mass resolving power (*e.g.,* high resolution TOF, see Figure 1.1). Increasing levels of mass resolving power (300,000 for Orbitrap and up to 40,000,000 for FT-ICR, respectively)[23,24] provides minimal increases in resolution of these metabolites (*ca.* 3% more). As *ca.* 30% of the compounds in the MSMLS library are isomers, essentially no level

3
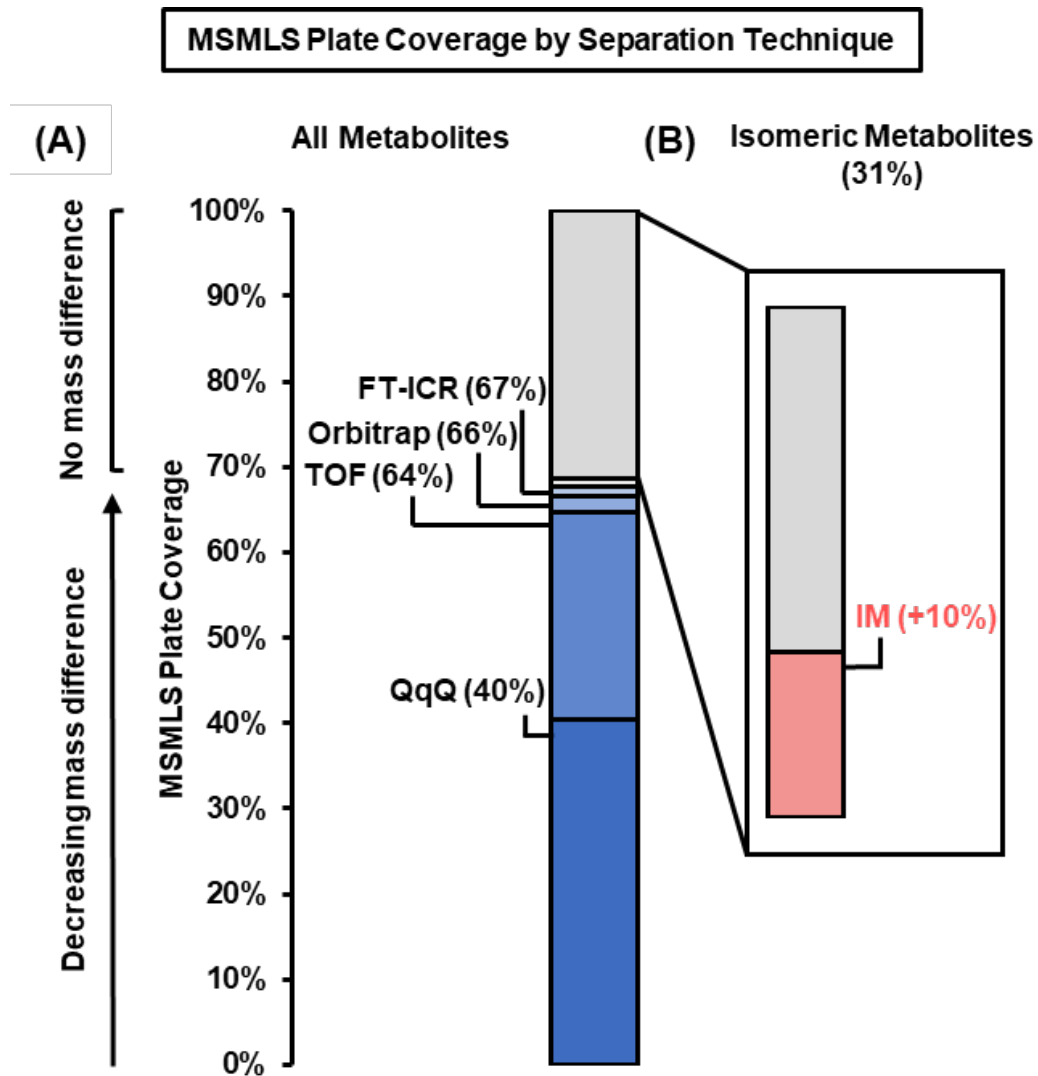
Figure 1.1 MSMLS plate coverage using different separation strategies. (A) Many analytes contained in the library can be resolved in the mass dimension at modest resolving power (TOF Rp = 40,000), with only incremental increases in coverage resulting from the use of an instrument with significantly higher resolving power (FT-ICR Rp = 40,000,000). (B) The addition of IM prior to mass analysis allows for isomeric separation and thus increases plate coverage by 10%.

of increased mass spectrometry efficiency (short of exited state isomer resolution with MS resolving power of *ca.* 10 billion as theorized by Marshall and coworkers.[25]) will be able to resolve these compounds, and hence orthogonal separation techniques are still required. Modest resolving power for commercially available IM instrumentation (*ca*. 70 CCS/ΔCCS) resolves an additional 10% of compounds in the library, which outweighs the benefits of additional mass resolving power beyond 40,000 (*e.g.,* TOF MS).

1.2.2   CCS standardization for database construction

Metabolomic databases commonly used in untargeted analyses (*e.g.* METLIN, HMDB, etc)[26] include multiple descriptors of analyte information (*e.g.* accurate mass, ion adduct form, fragmentation pattern, and retention time) to increase confidence in molecular identification.[27] Similarly, significant efforts have been made in the IM community to establish reliable CCS databases for analyzing unknown features across a range of biochemical classes, including lipids, metabolites, and xenobitics.[19,20,28,29] In Chapter 2 of this work, we detail the curation of the largest standardized repository of CCS values to date, of which a significant portion came from the MSMLS study. The repository contains CCS values acquired from DTIM instrumentation and calibrated using standard methods established for interlaboratory comparison. It includes values derived from relevant molecules of 95 chemical classes (as of this writing), both organic and inorganic. This open-source database, the Unified CCS Compendium, encourages the continuous submission and addition of more data and provides a resource to the community for the further adoption of CCS in untargeted identification workflows.

1.2.3   Mass-mobility correlations

In addition to their applicability as a molecular descriptor for direct database matching, CCS values provide additional information in the form of their conformational relationship to the mass

dimension, which manifests as characteristic class-specific correlations, where molecular classes separate in conformational space due to differences in their structural motifs.[30] In the MSMLS library, we observed several such relationships for individual structural super classes represented in the library, similar to previous IM-MS literature.[30–34] These correlations have shown utility as an additional rapid identifier of biomolecular class for uncharacterized biological samples,[35] making their mathematical description particularly useful. Unlike the canonical biochemical classes (nucleotides, proteins, carbohydrates, and lipids), metabolites exhibit less distinguished structural differences between chemical classes, and so several mathematical fits were investigated to find mass/mobility correlations which exhibit high class specificity. Fits and confidence intervals for representative super classes are shown in Figure 1.2, and detailed analysis parameters and expressions can be found in Appendix B, Section B2 (Eq. B1-B6).

In the work described in Chapter 2 with the CCS Compendium, we explore the utility of generating regression models from the large, curated dataset to mathematically characterize the mass-mobility relationship of all classes and subclasses with sufficient representative data. We show that using these models allows for the filtering of false positive isobaric identifications from an untargeted study. In Chapter 3 of this work, this approach is taken a step further, allowing automated filtering of large lipidomic datasets using the Compendium-based regression models. In a proof-of-concept study, we show that the novel informatics workflow reduced tentative identification candidates of almost 900 molecular features by over 50%, effectively increasing the confidence of the remaining identifications. The additional confidence and structural information provided by incorporating IM-derived chemical data in this way promotes an informed interpretation of the analytical measurements in untargeted experiments.

1.3   Advancing CCS resolving power
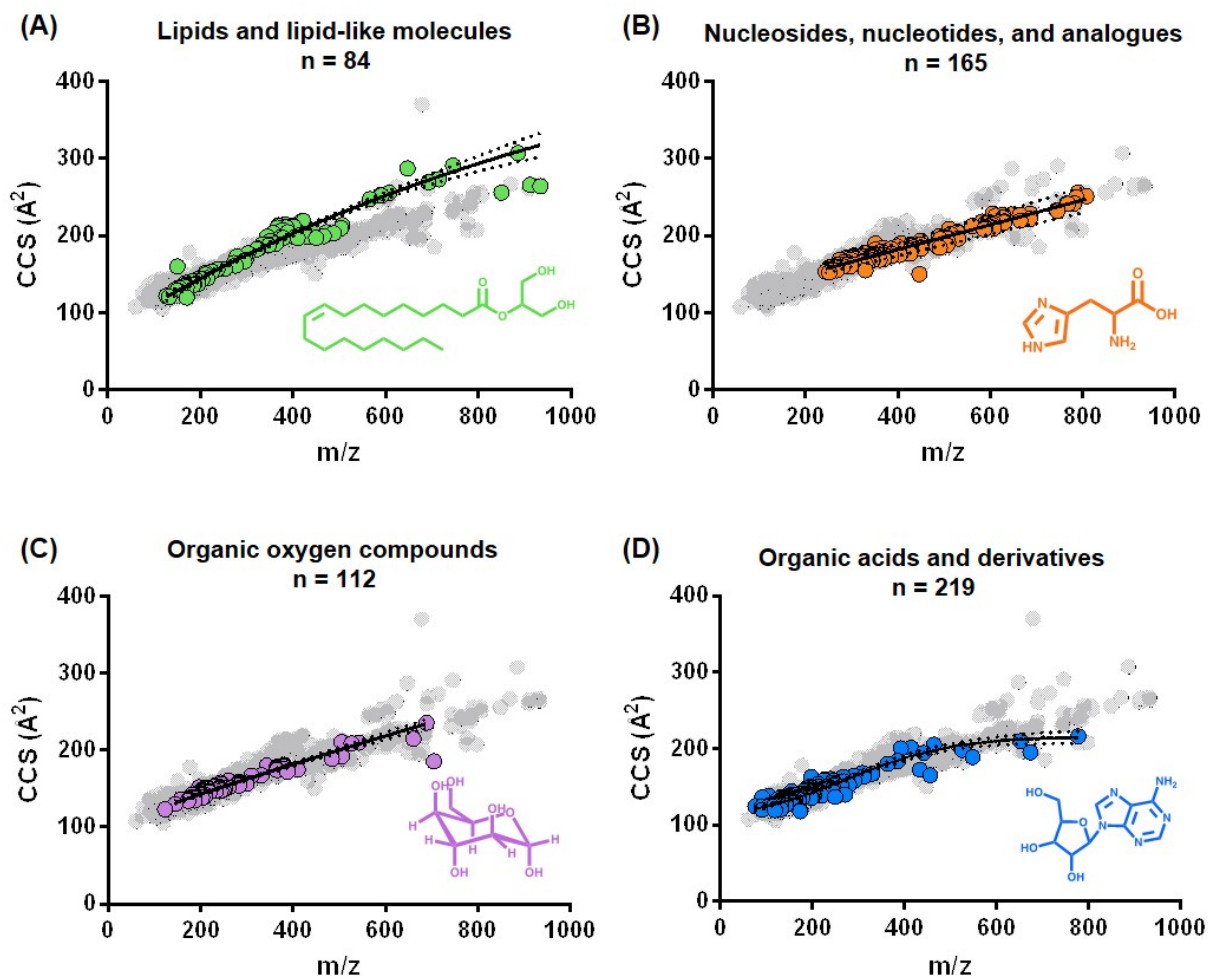
Figure 1.2 Conformational space plots of molecular super classes contained in the MSMLS library. Representative nonlinear regression fits (solid black lines) along with 99% confidence intervals (black dotted lines) are shown for each. Gray dots denote all molecules CCS values obtained in the library. All CCS error bars are smaller than their respective symbols.

1.3.1 Isomeric prevalence in metabolites

Recent advances in ion mobility resolving power have provided increased separation coverage of isomeric species.[36,37] To illustrate the resolving power in the IM dimension needed for untargeted metabolomic experiments, the pairwise matches of all isomers in the MSMLS library with a CCS were analyzed and the resulting pairs were binned by percent difference in cross section (%ΔCCS). In brief, analytes with identical chemical formulas were grouped into isomeric sets and were subsequently matched in a pairwise comparison. Each pairwise match was generated using an enumeration strategy wherein a percent difference in CCS was calculated for each possible combination of isomers. Most isomeric sets consist of 2-3 compounds, whereas the largest isomeric set was comprised of 9 unique analytes (see Figure B3). In one example, there are 5 sugar compounds which share the same chemical formula ($C_6H_{13}O_9PNa^+$, exact *m/z* 283.0195), which results in a total of 10 pairwise isomer matches in this analysis. The percent difference in CCS for all isomer matches were calculated, and the compiled results for all isomer pairings are displayed in Figure 1.3(B) (positive ion mode) and Figure 1.3(C) (negative ion mode). Approximately half of the isomer pairs generated are $\geq$ 2.0% different in CCS and require *ca.* 70 resolving power (CCS/ΔCCS) to separate at half height, which is within the range of conventional DTIM and traveling wave (TWIM) instruments.[37,38] In order to separate additional isomers, more resolving power is required (*ca.* 140 for ~1.0-2.0% difference in CCS, and *ca.* >300 for ~0.5% CCS difference).

1.3.2 High resolution ion mobility

In Chapter 4 of this work, we assess a production prototype high resolution IM (HRIM) platform based on structures for lossless ion manipulation (SLIM) technology. The resolving power (CCS/ΔCCS) of the IM stage was benchmarked across various ion masses and instrument
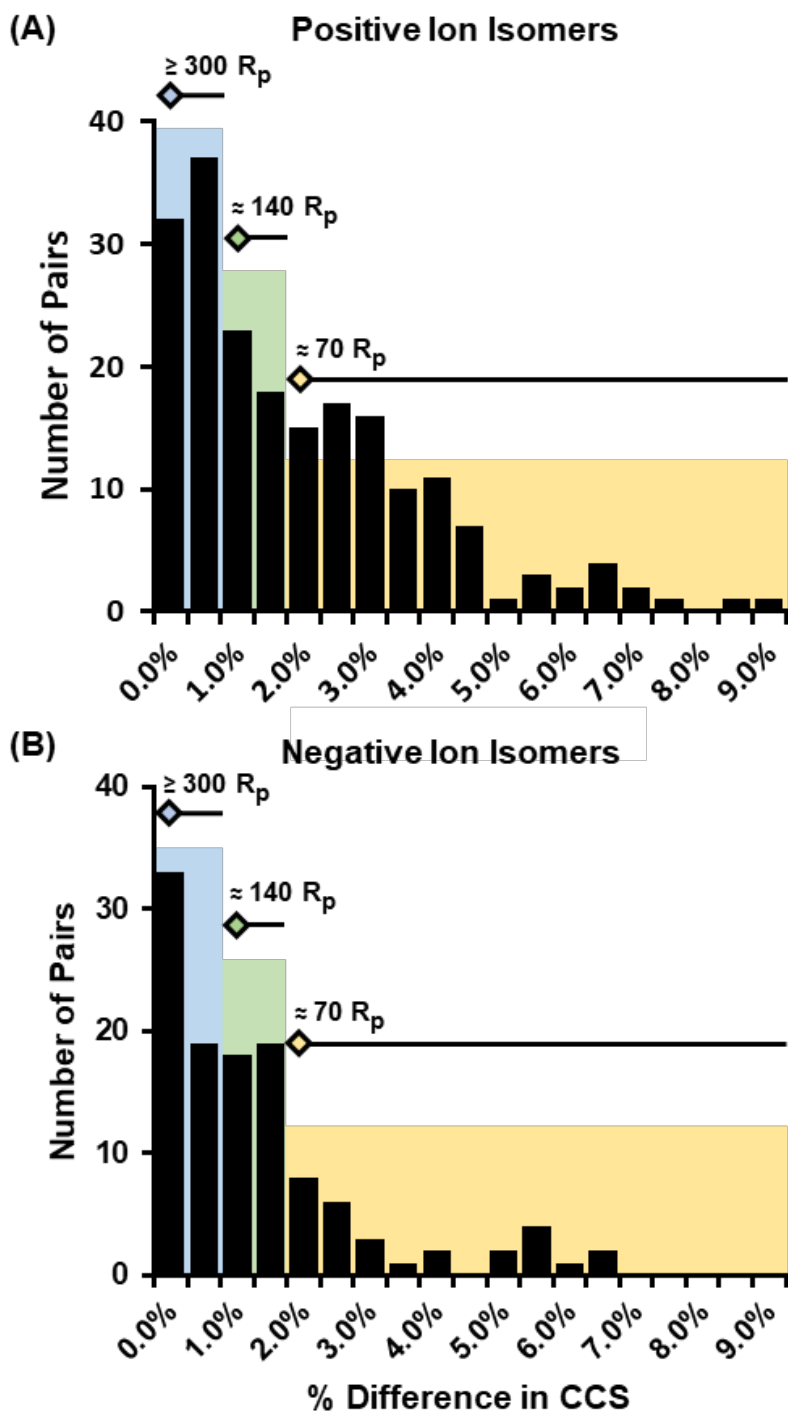
Figure 1.3 Sorted isomer sets in the MSMLS dataset by percent difference CCS. The resulting difficulty in separations is noted in panels (B) for positive and (C) negative ion forms.

parameters (traveling wave speeds and amplitudes), and results indicated that the platform can access resolving powers in excess of 200 across a broad range of masses. This study indicated that the HRIM-MS platform is capable of resolving peaks with CCS values spaced to at least within ~0.6%. With this added resolving power, an estimated 25% increase in isomeric coverage of the MSMLS library could theoretically be accessed in addition to the 10% gain using conventional-resolution DTIM shown in Figure 1.1.

Unlike DTIM measurements, CCS cannot be directly measured by SLIM-based HRIM. Instead, they must be calibrated using calibrant sets with known reference values. In addition to benchmarking resolving power, Chapter 4 also includes one of the first studies of CCS calibration strategies for HRIM measurements across multiple operational ranges. The results show that the calibrated CCS retains its high relative accuracy (< 0.5% bias from reference values) under all surveyed conditions. While this method worked well for a standard calibrant mixture, the CCS bias of lipids calibrated in this way was consistently higher (up to 3%). With the aim of expanding the utility of high-resolution CCS values to broader application spaces, Chapter 5 explores CCS calibration of lipid analytes from five lipid subclasses. In this study, we examined the accuracy, reproducibility, and practical applicability of calibration strategies for a broad range of lipid subclasses and developed a straightforward and generalizable framework for obtaining high-resolution CCS values. By applying subclass-specific corrections to values calibrated with a broadly available general calibrant set, values were obtained with biases < 0.4%. Using this method, a high-resolution CCS database containing over 90 lipid values was generated to support untargeted applications.

1.4   Untargeted multidimensional analysis of complex samples

Separation and characterization of isomeric species in biological extracts often requires multiple steps of chemical separation to gain increased confidence in assigning molecular structure. Using the MSMLS library, we underscore the importance of isomeric characterization in untargeted experiments by the analysis of a commercially available extract of human serum (NIST 1950) which has been characterized previously in traditional GC and LC-MS experiments.[39,40] Detailed methodology and experimental parameters related to this analysis can be found in Appendix B, Section B3.

As an example of the molecular complexity characteristic of untargeted analyses, Figure 1.4A shows the base peak chromatogram in Figure 1.4(A) of the NIST 1950 human serum and the extracted ion chromatogram (lower trace) details a specific molecular feature at *m/z* 203.0528 that elutes into an unresolved broad peak over a *ca.* 2-minute chromatographic window. This broad distribution in the elution profile indicates the potential presence of multiple isomeric forms with similar, yet not identical, retention times. Although TOF MS has high resolving power (*ca.* 40,000), potentially two chemical formulae are within 10 ppm of the measured *m/z* ($C_6H_{12}O_6Na$ and $C_7H_8N_4O_2Na$, at 1.7 ppm and 8.3 ppm respectively; see Figure 1.4(C)). While assignment of this feature to chemical formula $C_6H_{12}O_6Na$ is more probable due to lower observed mass error, isotope ratios were used to confirm this molecular formula assignment, wherein the relative abundance of the M+1 peak in the serum more closely aligns with the isotope model for $C_6H_{12}O_6Na$ as opposed to $C_7H_8N_4O_2Na$ (Figure 1.4(B)). However, even after a specific molecular formula is determined, 9 potential isomers (including both constitutional rearrangements and stereochemistry for this chemical formula) exist within the MSMLS standards, all carbohydrates. These isomers possess almost identical fragmentation profiles (see [M-H]⁻ ion, Figure 1.4(C)), and sophisticated algorithms for identification by MS/MS are needed, an observation which has been

previously noted in other carbohydrate studies.[41] Note that Figure 1.4(C) utilizes the deprotonated ion of $C_6H_{12}O_6$, as the $[M+Na]^+$ species noted in the other panels provides no fragmentation spectra due to ejection of the sodium charge carrier during collisional activation. Although previous studies utilized relative abundance ratios of fragment ions to determine molecular structure, this technique is time intensive and currently is not readily amended to rapid structural determination in untargeted workflows.[41] Similar to the chromatographic profile, ion mobility distributions obtained at three separate time points in the chromatogram (roman numerals) also indicate two separate chemical species present in the serum (Figure 1.4(D)). The collision cross sections measured for these two distributions helps narrow potential chemical structures from 9 potential isomeric forms down to 4 tentative identifications based on a CCS match within 1%. The smaller distribution at 140.7 Å^2 (light red, A) closely aligns with 3 isomers of $C_6H_{12}O_6$ in the standards (fructose, galactose and mannose at *ca.* 141.5 Å^2. The larger distribution at 147.0 Å^2 (light blue, B) closely aligns with α-D-glucose, which is noted at 146.3 Å^2 in the database. Although in this example ion mobility does not provide definitive identification of the compounds observed in the NIST serum, it does significantly reduce the possible candidates from the 9 potential structures noted in the database.

By using CCS as additional metric for tentative identification, additional confidence in identifying molecular signatures can be gained. The NIST serum experiment clearly illustrates the importance of hyphenating sequential separation dimensions based on chemical affinity, gas-phase area, and *m/z* (LC-IM-MS) to obtain the widest scope of molecular coverage in untargeted workflows. The research described in this work, as a whole, aims to support the integration of IM into multidimensional untargeted metabolomics and lipidomics workflows by providing guidelines,
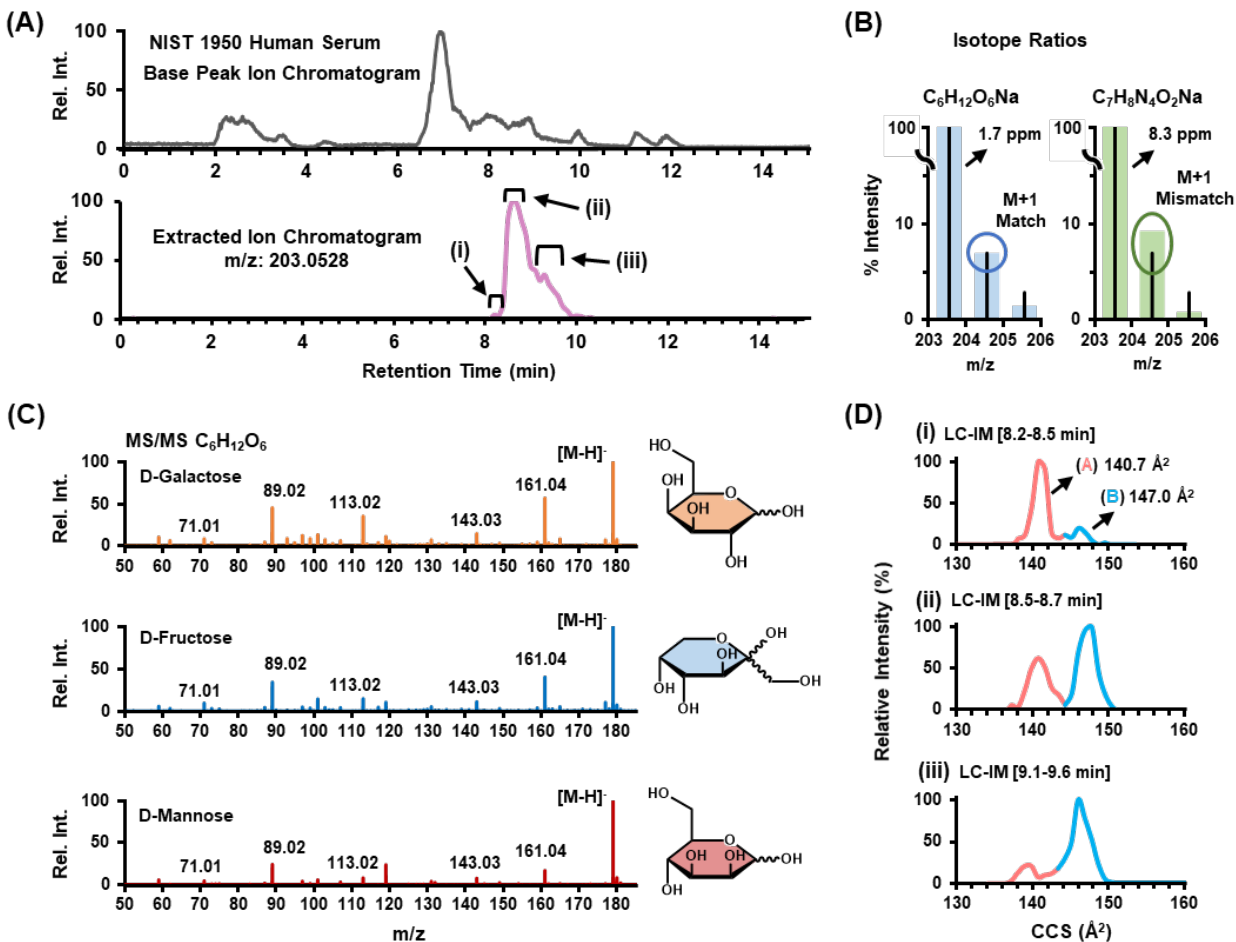
Figure 1.4 Untargeted analysis of NIST 1950 (A) HILIC base peak chromatogram for NIST 1950 human serum sample and (lower trace) the extracted ion chromatogram of *m/z* 203.0528, which consists of two distributions of interest that were further examined by isotope ratio pattern and ion mobility for structural characterization. (B) Expected and measured isotope ratio abundances for two possible chemical formulas corresponding to *m/z* 203 within 10 ppm. The chemical formula C6H12O6 [M+Na]+ more closely aligns with experimental measurements from the NIST serum both on basis of mass accuracy (2 ppm) and isotope ratio pattern (M+1). (C) Fragmentation spectra for isomers of with shared chemical formula C6H12O6 [M-H]-. (D) Selected ion mobility distributions for *m/z* 203 extracted over three time points in the chromatographic dimension.

recommendations, and novel analysis approaches for increasing molecular annotations confidence using CCS values.

1.5   Conclusions

Utilizing orthogonal dimensions of separation in addition to mass analysis is imperative to fully characterize relevant biological pathways in untargeted metabolomics and lipidomics. While additional resolving power in the *m/z* dimension is always advantageous, the diminishing returns of these efforts may not offset the additional analysis time required for ultra-high resolution mass acquisition (*i.e.* FT processes). However, orthogonal separation techniques such as LC and IM can often resolve many isomeric forms, facilitating their identification for a more comprehensive understanding of the biochemical implications of experimental samples. The intrinsic class-specific mass-mobility correlations represent a reproducible method of characterization for biochemical classes which interfaces seamlessly into the timescale of traditional LC/GC-MS workflows. The full potential of adding CCS as a molecular descriptor in untargeted metabolomic and lipidomic analyses will be realized through continued collaborative work toward the standardization of methodology and further innovation in informatic approaches, as seen throughout this work. Following a summary of the work described in the preceding chapters, Chapter 6 discusses its outlook and future directions, including several projects currently underway.

1.6   Acknowledgments

and do not necessarily reflect those of the funding agencies and organizations. EPA does not endorse any products or commercial services mentioned in this publication.

1.7   References

(1)   CRICK, F. Central Dogma of Molecular Biology. *Nature* **1970**, *227* (5258), 561–563. https://doi.org/10.1038/227561a0.

(2)   Joyce, A. R.; Palsson, B. Ø. The Model Organism as a System: Integrating "omics" Data Sets. *Nature Reviews Molecular Cell Biology* **2006**, *7* (3), 198–210. https://doi.org/10.1038/nrm1857.

(3)   Goodacre, R.; Broadhurst, D.; Smilde, A. K.; Kristal, B. S.; Baker, J. D.; Beger, R.; Bessant, C.; Connor, S.; Capuani, G.; Craig, A.; Ebbels, T.; Kell, D. B.; Manetti, C.; Newton, J.; Paternostro, G.; Somorjai, R.; Sjöström, M.; Trygg, J.; Wulfert, F. Proposed Minimum Reporting Standards for Data Analysis in Metabolomics. *Metabolomics* **2007**, *3* (3), 231–241. https://doi.org/10.1007/s11306-007-0081-3.

(4)   Sumner, S.; Snyder, R.; Burgess, J.; Myers, C.; Tyl, R.; Sloan, C.; Fennell, T. Metabolomics in the Assessment of Chemical-Induced Reproductive and Developmental Outcomes Using Non-Invasive Biological Fluids: Application to the Study of Butylbenzyl Phthalate. *Journal of Applied Toxicology* **2009**, *29* (8), 703–714. https://doi.org/10.1002/jat.1462.

(5)   Kimmel, D. W.; Dole, W. P.; Cliffel, D. E. Elucidation of the Role of Lectin-Like OxLDL Receptor-1 in the Metabolic Responses of Macrophages to Human OxLDL. *Journal of Lipids* **2017**, *2017*, 1–9. https://doi.org/10.1155/2017/8479482.

(6)   May, J. C.; McLean, J. A. Ion Mobility-Mass Spectrometry: Time-Dispersive Instrumentation. *Analytical Chemistry*. 2015, pp 1422–1436. https://doi.org/10.1021/ac504720m.

(7)   Kind, T.; Tsugawa, H.; Cajka, T.; Ma, Y.; Lai, Z.; Mehta, S. S.; Wohlgemuth, G.; Kumar, D.; Showalter, M. R.; Arita, M.; Fiehn, O. Identification of Small Molecules Using Accurate Mass MS / MS Search. **2018**, No. March 2017, 513–532. https://doi.org/10.1002/mas.21535.

(8)   Kim, S.; Rodgers, R. P.; Marshall, A. G. Truly "Exact" Mass: Elemental Composition Can Be Determined Uniquely from Molecular Mass Measurement at ~0.1mDa Accuracy for Molecules up to ~500Da. *International Journal of Mass Spectrometry* **2006**, *251* (2–3), 260–265. https://doi.org/10.1016/j.ijms.2006.02.001.

(9)   Hofmann, J.; Hahm, H. S.; Seeberger, P. H.; Pagel, K. Identification of Carbohydrate Anomers Using Ion Mobility-Mass Spectrometry. *Nature* **2015**, *526* (7572), 241–244. https://doi.org/10.1038/nature15388.

(10)   Kyle, J. E.; Zhang, X.; Weitz, K. K.; Monroe, M. E.; Ibrahim, Y. M.; Moore, R. J.; Cha, J.; Sun, X.; Lovelace, E. S.; Wagoner, J.; Polyak, S. J.; Metz, T. O.; Dey, S. K.; Smith, R. D.; Burnum-Johnson, K. E.; Baker, E. S. Uncovering Biologically Significant Lipid Isomers with Liquid Chromatography, Ion Mobility Spectrometry and Mass Spectrometry. *The Analyst* **2016**, *141* (5), 1649–1659. https://doi.org/10.1039/C5AN02062J.

(11)   Yost, R. A.; Enke, C. G.; McGilvery, D. C.; Smith, D.; Morrison, J. D. High Efficiency Collision-Induced Dissociation in an RF-Only Quadrupole. *International Journal of Mass Spectrometry and Ion Physics* **1979**, *30* (2), 127–136. https://doi.org/10.1016/0020-7381(79)80090-X.

(12)   Lareau, N. M.; May, J. C.; McLean, J. A. Non-Derivatized Glycan Analysis by Reverse Phase Liquid Chromatography and Ion Mobility-Mass Spectrometry. *The Analyst* **2015**, *140* (10), 3335–3338. https://doi.org/10.1039/C5AN00152H.

(13)   Dodds, J. N.; May, J. C.; McLean, J. A. Investigation of the Complete Suite of the Leucine and Isoleucine Isomers: Toward Prediction of Ion Mobility Separation Capabilities. *Analytical Chemistry* **2017**, *89* (1), 952–959. https://doi.org/10.1021/acs.analchem.6b04171.

(14)   Ekroos, K.; Chernushevich, I. V.; Simons, K.; Shevchenko, A. Quantitative Profiling of Phospholipids by Multiple Precursor Ion Scanning on a Hybrid Quadrupole Time-of-Flight Mass Spectrometer. *Analytical Chemistry* **2002**, *74* (5), 941–949. https://doi.org/10.1021/ac015655c.

(15)   Kanani, H.; Chrysanthopoulos, P. K.; Klapa, M. I. Standardizing GC-MS Metabolomics. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences* **2008**, *871* (2), 191–201. https://doi.org/10.1016/j.jchromb.2008.04.049.

(16)   Lu, W.; Bennett, B. D.; Rabinowitz, J. D. Analytical Strategies for LC-MS-Based Targeted Metabolomics. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences* **2008**, *871* (2), 236–242. https://doi.org/10.1016/j.jchromb.2008.04.031.

(17)   Mairinger, T.; Causon, T. J.; Hann, S. The Potential of Ion Mobility–Mass Spectrometry for Non-Targeted Metabolomics. *Current Opinion in Chemical Biology* **2018**, *42*, 9–15. https://doi.org/10.1016/j.cbpa.2017.10.015.

(18)   Pringle, S. D.; Giles, K.; Wildgoose, J. L.; Williams, J. P.; Slade, S. E.; Thalassinos, K.; Bateman, R. H.; Bowers, M. T.; Scrivens, J. H. An Investigation of the Mobility Separation of Some Peptide and Protein Ions Using a New Hybrid Quadrupole/Travelling Wave IMS/Oa-ToF Instrument. *International Journal of Mass Spectrometry* **2007**, *261* (1), 1–12. https://doi.org/10.1016/j.ijms.2006.07.021.

(19)   Stow, S. M.; Causon, T. J.; Zheng, X.; Kurulugama, R. T.; Mairinger, T.; May, J. C.; Rennie, E. E.; Baker, E. S.; Smith, R. D.; McLean, J. A.; Hann, S.; Fjeldsted, J. C. An Interlaboratory Evaluation of Drift Tube Ion Mobility-Mass Spectrometry Collision Cross Section Measurements. *Analytical Chemistry* **2017**, *89* (17), 9048–9055. https://doi.org/10.1021/acs.analchem.7b01729.

(20)   Paglia, G.; Williams, J. P.; Menikarachchi, L.; Thompson, J. W.; Tyldesley-Worster, R.; Halldórsson, S.; Rolfsson, O.; Moseley, A.; Grant, D.; Langridge, J.; Palsson, B. O.; Astarita, G. Ion Mobility Derived Collision Cross Sections to Support Metabolomics Applications. *Analytical Chemistry* **2014**, *86* (8), 3985–3993. https://doi.org/10.1021/ac500405x.

(21)   Paglia, G.; Astarita, G. Metabolomics and Lipidomics Using Traveling-Wave Ion Mobility Mass Spectrometry. *Nature Protocols* **2017**, *12* (4), 797–813. https://doi.org/10.1038/nprot.2017.013.

(22)   Pu, Y.; Ridgeway, M. E.; Glaskin, R. S.; Park, M. A.; Costello, C. E.; Lin, C. Separation and Identification of Isomeric Glycans by Selected Accumulation-Trapped Ion Mobility Spectrometry-Electron Activated Dissociation Tandem Mass Spectrometry. *Analytical Chemistry* **2016**, *88* (7), 3440–3443. https://doi.org/10.1021/acs.analchem.6b00041.

(23)   N. Nikolaev, E.; N. Vladimirov, G.; Jertz, R.; Baykut, G. From Supercomputer Modeling to Highest Mass Resolution in FT-ICR. *Mass Spectrometry* **2013**, *2*, S0010–S0010. https://doi.org/10.5702/massspectrometry.S0010.

(24)   Zubarev, R. A.; Makarov, A. Orbitrap Mass Spectrometry. *Analytical Chemistry* **2013**, *85* (11), 5288–5296. https://doi.org/10.1021/ac4001223.

(25)   Marshall, A. G.; Hendrickson, C. L.; Shi, S. D.-H. Peer Reviewed: Scaling MS Plateaus with High-Resolution FT-ICRMS. *Analytical Chemistry* **2002**, *74* (9), 252 A-259 A. https://doi.org/10.1021/ac022010j.

(26)   Smith, C. a; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G. METLIN: A Metabolite Mass Spectral Database. *Therapeutic drug monitoring* **2005**, *27* (6), 747–751.

(27)   Schrimpe-Rutledge, A. C.; Codreanu, S. G.; Sherrod, S. D.; McLean, J. A. Untargeted Metabolomics Strategies—Challenges and Emerging Directions. *Journal of the American Society for Mass Spectrometry* **2016**, *27* (12), 1897–1905. https://doi.org/10.1007/s13361-016-1469-y.

(28)   Paglia, G.; Angel, P.; Williams, J. P.; Richardson, K.; Olivos, H. J.; Thompson, J. W.; Menikarachchi, L.; Lai, S.; Walsh, C.; Moseley, A.; Plumb, R. S.; Grant, D. F.; Palsson, B. O.; Langridge, J.; Geromanos, S.; Astarita, G. Ion Mobility-Derived Collision Cross Section As an Additional Measure for Lipid Fingerprinting and Identification. *Analytical Chemistry* **2015**, *87* (2), 1137–1144. https://doi.org/10.1021/ac503715v.

(29)   Zheng, X.; Aly, N. A.; Zhou, Y.; Dupuis, K. T.; Bilbao, A.; Paurus, V. L.; Orton, D. J.; Wilson, R.; Payne, S. H.; Smith, R. D.; Baker, E. S. A Structural Examination and Collision Cross Section Database for over 500 Metabolites and Xenobiotics Using Drift Tube Ion Mobility Spectrometry. *Chem. Sci.* **2017**, *8*, 7724–7736. https://doi.org/10.1039/C7SC03464D.

(30)   McLean, J. A. The Mass-Mobility Correlation Redux: The Conformational Landscape of Anhydrous Biomolecules. *Journal of the American Society for Mass Spectrometry* **2009**, *20* (10), 1775–1781. https://doi.org/10.1016/j.jasms.2009.06.016.

(31)   May, J. C.; Goodwin, C. R.; Lareau, N. M.; Leaptrot, K. L.; Morris, C. B.; Kurulugama, R. T.; Mordehai, A.; Klein, C.; Barry, W.; Darland, E.; Overney, G.; Imatani, K.; Stafford, G. C.; Fjeldsted, J. C.; McLean, J. A. Conformational Ordering of Biomolecules in the Gas Phase: Nitrogen Collision Cross Sections Measured on a Prototype High Resolution Drift Tube Ion Mobility-Mass Spectrometer. *Analytical Chemistry* **2014**, *86* (4), 2107–2116. https://doi.org/10.1021/ac4038448.

(32)   Hines, K. M.; Ashfaq, S.; Davidson, J. M.; Opalenik, S. R.; Wikswo, J. P.; McLean, J. A. Biomolecular Signatures of Diabetic Wound Healing by Structural Mass Spectrometry. *Analytical Chemistry* **2013**, *85* (7), 3651–3659. https://doi.org/10.1021/ac303594m.

(33)   Hines, K. M.; Ross, D. H.; Davidson, K. L.; Bush, M. F.; Xu, L. Large-Scale Structural Characterization of Drug and Drug-Like Compounds by High-Throughput Ion Mobility-Mass Spectrometry. *Analytical Chemistry* **2017**, *89* (17), 9023–9030. https://doi.org/10.1021/acs.analchem.7b01709.

(34)   Dwivedi, P.; Wu, P.; Klopsch, S. J.; Puzon, G. J.; Xun, L.; Hill, H. H. Metabolic Profiling by Ion Mobility Mass Spectrometry (IMMS). *Metabolomics* **2008**, *4* (1), 63–80. https://doi.org/10.1007/s11306-007-0093-z.

(35)   Goodwin, C. R.; Fenn, L. S.; Derewacz, D. K.; Bachmann, B. O.; McLean, J. A. Structural Mass Spectrometry: Rapid Methods for Separation and Analysis of Peptide Natural Products. *Journal of Natural Products* **2012**, *75* (1), 48–53. https://doi.org/10.1021/np200457r.

(36)   D'Atri, V.; Causon, T.; Hernandez-Alba, O.; Mutabazi, A.; Veuthey, J. L.; Cianferani, S.; Guillarme, D. Adding a New Separation Dimension to MS and LC–MS: What Is the Utility of Ion Mobility Spectrometry? *Journal of Separation Science* **2018**, *41* (1), 20–67. https://doi.org/10.1002/jssc.201700919.

(37)   Dodds, J. N.; May, J. C.; McLean, J. A. Correlating Resolving Power, Resolution, and Collision Cross Section: Unifying Cross-Platform Assessment of Separation Efficiency in Ion Mobility Spectrometry. *Analytical Chemistry* **2017**, *89* (22), 12176–12184. https://doi.org/10.1021/acs.analchem.7b02827.

(38)   Fernandez-Lima, F. A.; Kaplan, D. A.; Park, M. A. Note: Integration of Trapped Ion Mobility Spectrometry with Mass Spectrometry. *Review of Scientific Instruments* **2011**, *82* (12). https://doi.org/10.1063/1.3665933.

(39)   Simón-Manso, Y.; Lowenthal, M. S.; Kilpatrick, L. E.; Sampson, M. L.; Telu, K. H.; Rudnick, P. A.; Mallard, W. G.; Bearden, D. W.; Schock, T. B.; Tchekhovskoi, D. V.; Blonder, N.; Yan, X.; Liang, Y.; Zheng, Y.; Wallace, W. E.; Neta, P.; Phinney, K. W.; Remaley, A. T.; Stein, S. E. Metabolite Profiling of a NIST Standard Reference Material for Human Plasma (SRM 1950): GC-MS, LC-MS, NMR, and Clinical Laboratory

Analyses, Libraries, and Web-Based Resources. *Analytical Chemistry* **2013**, *85* (24), 11725–11731. https://doi.org/10.1021/ac402503m.

(40)   Telu, K. H.; Yan, X.; Wallace, W. E.; Stein, S. E.; Simón-Manso, Y. Analysis of Human Plasma Metabolites across Different Liquid Chromatography/Mass Spectrometry Platforms: Cross-Platform Transferable Chemical Signatures. *Rapid Communications in Mass Spectrometry* **2016**, *30* (5), 581–593. https://doi.org/10.1002/rcm.7475.

(41)   Fang, T. T.; Bendiak, B. The Stereochemical Dependence of Unimolecular Dissociation of Monosaccharide-Glycolaldehyde Anions in the Gas Phase: A Basis for Assignment of the Stereochemistry and Anomeric Configuration of Monosaccharides in Oligosaccharides by Mass Spectrometry via a. *Journal of the American Chemical Society* **2007**, *129* (31), 9721–9736. https://doi.org/10.1021/ja0717313.

CHAPTER 2

STANDARDIZATION OF COLLISION CROSS SECTION REPORTING: BUILDING A

UNIFIED CCS COMPENDIUM TO ANNOTATE AND PREDICT MULTI-OMIC

COMPOUND IDENTITIES[†]

2.1   Introduction

Mass spectrometry (MS) has become a central technique for the investigation of the global profile

of biochemical species in molecular phenomic studies.[1,2] These studies aim to address the grand

challenges of biomedical research including comprehensive descriptions of biological systems,

natural product and drug discovery endeavors, omics sciences to improve health outcomes, and

progress in synthetic biology.[3–6] As the complexity of the systems being studied increases, so must

the ability to increase analyte coverage. Orthogonal separation techniques including gas and liquid

chromatography are often used in conjunction with MS to improve coverage; however, feature

annotation and identification from these types of experiments can be challenging due to analyte

co-elution and retention time variability, among other issues.[7]

These challenges can be addressed with the use of additional analytical separation techniques, such

as ion mobility spectrometry (IM) coupled to MS (IM-MS), which is selective to analyte gas phase

structure.[3,8] One practical benefit of using gas-phase IM is that there are no memory effects or

sample-to-sample carryover due to continuous replacement of the separation gas. Additionally, IM

separations do not require disposable solvents or packed columns and are amenable to all ionizable chemical species. The main advantages of IM-MS are an increase in analytical peak capacity over MS alone as well as the ability to measure gas phase mobility by means of an experimental drift time.[9] This mobility can then be used to calculate an analyte's collision cross section (CCS), a rotationally averaged surface area of the molecule in its gaseous, ionic form. These CCS values are specific and can be compared across different laboratories, making them particularly well-suited for species identification and characterization purposes. Previous studies indicate that the level of reproducibility varies across analyte classes.[10,11] Recent studies using drift tube IM-MS have shown that CCS values can be measured within a 0.30% RSD when data is acquired with a previously established standardized method.[12]

As a result of these advantages, several research groups have used IM-MS to build CCS libraries in which the measured values serve as additional molecular descriptors for assigning identities to unknown analytes. A few of the larger libraries to note include, but are not limited to: Li and colleagues' peptide database which includes > 2300 CCS values,[13] Pagel and colleagues' glycomics database of > 900 CCS values,[14] and Xu and colleagues' small molecule database containing > 1400 CCS values.[15] Additionally, many excellent smaller CCS libraries have been generated for lipids,[16–18] primary metabolites,[18–20] secondary metabolites and other natural products,[18,21,22] as well as illicit substances,[23] among others.

While each of these libraries adds to the working knowledge of the IM-MS field, there remain challenges that need to be addressed. The first is reconciling CCS measurements across various IM implementations such as drift tube (DTIMS), traveling wave (TWIMS), ion trapping (TIMS), and structures for lossless ion manipulation (SLIM) techniques. Inherently, these techniques utilize different methodologies for determining the gas-phase CCS, namely DTIMS (and drift tube-based

21

SLIM) utilize the fundamental ion mobility relationship for correlating the measured arrival times directly to CCS, whereas the other IM techniques obtain a CCS value through calibration. In order to reconcile non-DTIMS CCS values with DTIMS values, proper calibrants must be chosen for a given experiment, which can prove challenging.[10,24] An in-depth discussion of considerations for comparing CCS information obtained from different IM techniques can be found in a recent review by Gabelica et al.[25]

Another challenge lies in the difficulty of accurately and efficiently extracting drift time measurements from raw data files in large scale. Currently, most DTIMS drift times for chemical standards are manually extracted, improving accuracy at a cost of throughput. However, several software options exist that aim to automate the extraction of drift times on a large scale and/or predict drift times.[26,27] The recent IM-MS analysis addendum to Skyline is one example that has made considerable strides in these efforts,[28,29] but the IM-MS field is still working towards a streamlined analytic workflow.

Other informatics programs aim to predict CCS values based on experimental data and chemical structure. Some examples of these software include Zhu and colleague's machine learning algorithms for metabolites (MetCCS) and lipids (LipidCCS).[16,19] A major barrier to the success of machine learning CCS prediction is that algorithm training sets are generally not yet large and/or specific enough.[30] An alternate strategy recently described by Colby et al. is the in silico chemical library engine (ISiCLE) workflow which utilizes a combination of molecular dynamics, quantum chemistry, and ion mobility calculations in order to predict CCS values based on theoretical structure information.[31] These CCS prediction efforts are critically important for determining CCS values where empirical measurements on authentic chemical standards are unavailable.

To aid in the mainstream adoption of IM in analyte identification workflows, we explored the potential in curating libraries of empirical CCS values measured via ion mobility into a single, self-consistent compendium. The Unified CCS Compendium presented herein serves as a tool where new data from the community can be vetted using a quality control protocol and subsequently integrated. Included in this curated Compendium are several prevalent calibrant sets (polypeptides, branched phosphazenes, inorganic salt clusters, etc.), as well as molecular standards from a variety of chemical classes measured using DTIMS. These data sets can be used as reference values for other IM-MS techniques. Furthermore, this tool incorporates annotative features (i.e. visualization of chemical locales of molecules) and predictive statistics (chemical structure-based trends) to aid in identifying unknown biochemical species. These predictive trends serve as a powerful filter for increasing confidence in tentative identifications. In order to demonstrate the efficacy of this approach, the structural filtering method was applied to metabolites in a human serum sample. The full interactive visualization of the Compendium, as well as inclusion criteria and guidelines for submitting additional CCS measurements, can be found as an open access tool.[32]

2.2   Experimental Methods

2.2.1   Materials and Instrumentation

Methanol (MeOH), water, acetonitrile (ACN), isopropanol (IPA), and formic acid of Optima grade purity were purchased from Fisher Scientific (Fair Lawn, NJ). Anhydrous methyl-tert-butyl ether (MTBE) was purchased from Sigma Aldrich (St. Louis, MO). Normal human serum was purchased from Utak (Valencia, CA). A mixture of fluoroalkyl phosphazenes, tris(fluoroalkyl)triazines, betaine, and trifluoracetic acid reference standards were purchased from Agilent Technologies (G1969-85000, Santa Clara, CA). In this manuscript, liquid chromatography

MS (LC-MS) and LC-IM-MS data were acquired using a 1290 Infinity LC system and a 6560 IM-QTOF MS (Agilent Technologies).

2.2.2 *Data Sources and Inclusion Parameters*

The 3833 IM-MS measurements included in the Compendium are reported in a series of manuscripts found elsewere.[10,12,18,33–38] In order to provide highly repeatable and reproducible data, the Compendium currently only contains CCS values calculated from the fundamental low-field IM equation (Mason-Schamp relationship) incorporated into a standardized inter-laboratory protocol for single field and stepped field DTIMS acquisition on a commercial uniform-field IM-MS instrument (6560, Agilent).[12,39,40] In-depth information about single and stepped field DTIMS has previously been described.[12] All measurements were acquired in triplicate and aligned with a suite of 13 reference standards (Agilent Technologies) containing symmetrically-branched fluoroalkyl phosphazines, namely hexaxis(fluoroalkoxy)phosphazines, tris(fluoroalkyl)triazines, betaine, and trifluoracetic acid. These reference standards were previously measured with very high precision; and it is currently believed that these CCS values are among the most accurate obtained to date.[12]

In total, there are 1216 single field measurements within the Compendium; and the average relative standard deviation (RSD) for the single field measurements is 0.12%. Compounds were matched to reference standards' values from an inter-laboratory study.[12] The average percent error of Compendium CCS measurements was found to be 0.04% and -0.33% for positive and negative modes, respectively, with all percent error values at < 0.58% for both polarities. The remaining 2617 stepped field values were reconciled by calculating a "true effective length" for each data set (data set defined as a group of measurements collected in a one-day acquisition period) using calibrant measurements within the set. This "true effective length" was then used to align

measurements with reference standards' values. More details and tools to calculate "true effective length" as well as instructions to calibrate acquired stepped field CCS data are accesible in the online interface.[32] Once each data set was individually scaled, the average RSD for stepped field measurements was calculated to be 0.32%. When compared to inter-laboratory reference values, average percent errors were 0.07% and 0.01% for positive and negative modes, respectively. Ninety-one percent of matched values had a percent error < 1% for both polarities. These empirically derived metrics, in conjunction with known errors propagated in this system,[12] were subsequently used as the Compendium's data inclusion criteria. Full descriptions of the inclusion criteria and instructions for submitting data to the Compendium can be accessed from the online user interface.[32]

2.2.3 Data Preparation, Statistical Modeling, and Visualization

Data from all sources were curated into a unified format using the statistical computing programming environment R (R Foundation for Statistical Computing, Vienna, Austria).[41] This unified format includes the following information for each Compendium entry: name, formula, CAS registry number (when available), mass-to-charge (*m/z*), charge state, ion species, size-to-charge (CCS/z), percent RSD, number of observed DTIMS peaks, and annotations for the most intense DTIMS peak(s). Charge-normalization of mobility measurements via CCS/z was utilized to preserve the original drift time scale and analysis consistency.[39] In drift time spectra, ions of similar mass and higher charge states typically have smaller drift times than lower charge state ions; and therefore, appear lower when visualized in drift time vs *m/z* space. Contrastingly, because CCS as calculated from the Mason-Schamp relationship is directly proportional to charge, higher charge state ions have larger CCS values than lower charge state ions when visualized in CCS vs *m/z* space. By charge-normalizing, data can be plotted in CCS/z vs *m/z* space, and measurements

appear as they would in drift time vs *m/z* space. Furthermore, when values were not charge-normalized, statistical modeling could not be standardized and was charge-state dependent. The number of DTIMS peaks observed for each molecule as well as information about which peak was observed to be most intense is included in the Compendium. These data meet the outlined criteria and follow the standardized IM-MS data reporting efforts led by Gabelica et al.[42] Briefly, all observed DTIMS peaks are reported in the online Compendium compound table via a peak number assignment where the smallest CCS/z (earliest drift time) will be assigned number 1 and subsequent peaks will be assigned 2, 3, etc. Compounds with one observed peak will be assigned a "1". Additional information regarding DTIMS peak annotation can be found in Appendix C.

The unified format also includes a hierarchical chemical classification for each compound which includes a kingdom, super class, class, and subclass based on structure. This was performed via the ClassyFire web-based application which operates using a comprehensive chemical ontology (ChemOnt) that classifies each molecule based on its SMILES identifier as an input.[43,44] For example, a phosphatidylcholine would be classified as a member of the organic compound kingdom, the lipids and lipid-like molecules super class, the glycerophospholipid class, and the glycerophosphocholine subclass.

Iterative nonlinear regression modeling was performed using the R program for each chemical class and subclass that contained at least ten data points. Source code for this statistical modeling is provided on the McLean Research Group Github.[45] Each class was tested against three nonlinear regression models: a power fit (PF), a four-parameter sigmoidal fit (4P), and a five-parameter sigmoidal fit. Representative equations for these models are described in Appendix C, eq. 1-3. These models were chosen based on previous work.[34,46,47] The goodness of fit for each model was assessed using the corrected Akaike information criterion (AICc) for each of the three models.

This conservative metric accounts for small sample sizes, bias correction, and varying degrees of freedom in nonlinear candidate models; and has previously been shown to be highly reliable when comparing nonlinear models.[48,49] The model with the lowest AICc value was taken to be the best fit. Ninety-nine percent confidence (CI) and predictive (PI) intervals were calculated as described in Appendix C, eq. C4 and eq. C5, respectively. CI and PI were calculated in the same manner for all nonlinear regressions.

The Unified CCS Compendium was visualized using the following open-source R packages: plotly (v4.7.1), ggplot2 (v2.2.1.900), data.table (v1.10.4-3), plyr (v1.8.4), and shiny.[50–54] Source code for the Compendium GUI can be found on the McLean Research Group Github.[45]

2.2.4   *Evaluation of the Compendium in the Analysis of Human Serum*

Non-endogenous fatty acids 17:0 and 19:0 were used as internal standards and added into 100 µL control human serum. 800 µL of cold MeOH (-20˚C) was subsequently added and the sample was stored at -20˚C overnight to precipitate out proteins. The sample was subsequently centrifuged at 14,000 rpm and 4˚C for five minutes. The supernatant was collected; and 2.4 mL ice cold MTBE and 800 µL ice cold water were added. This MTBE:MeOH:water sample was vortexed then centrifuged at 10,000 rpm and 4˚C for ten minutes. The nonpolar liquid fraction was siphoned, dried under vacuum, and stored at -20˚C until use. Dried fractions were resuspended in 100 µL of 70:18:12 water:IPA:ACN and analyzed via LC-MS and LC-IM-MS. Further details are provided in Appendix C. LC-MS data was analyzed using Progenesis QI (v2.3, Nonlinear Dynamics, Durham, NC). Resulting features were tentatively identified using the Metlin Metabolomics and LipidBlast databases.[55,56] LC-IM-MS raw acquisition files were converted to mzML format using MSConvert (v3.0, ProteoWizard).[57] Drift time values from LC-IM-MS experiments for individual process replicates were extracted using an internally developed python script[45] in which drift times

were matched against the retention time and *m/z* of the tentatively identified compounds. These match functions had a threshold of 30 s (or 1% variation) for retention time and 5 ppm for *m/z*, respectively. Once drift times were extracted from the mzML data files, CCS/z values were calculated from the Mason-Schamp relationship using the averaged drift times. Chemical class probability hierarchies were analyzed using distance of the mean calculations based on where serum CCS/z values fell within the Compendium as compared to the regression models.

2.3   Results and discussion

2.3.1   CCS Compendium Properties

The Unified CCS Compendium compiled in this work consists of a total of 3833 CCS values (see inclusion criteria in the Experimental Section) obtained with uniform drift tube instruments in nitrogen drift gas utilizing a standardized CCS protocol.[12] Measurements consist of 2740 cations and 1093 anions, all of which were acquired in replicates of > 3. Associated measurement RSDs can be found on the web-based Compendium.[32] Thirteen ion species types are represented, as indicated in Figure 2.1(A). The most common species observed were proton coordination (38%), proton loss (27%), and sodium coordination (25%). Ion species were assigned based on the charge source of the molecule. For example, if a compound was observed as [M+2Na-H]+, the ion species was labelled as "+Na." Likewise, if a compound had multiple charge carriers of the same type, such as [M+4H]4+, it was labelled as "+H". Compounds with multiple different equal charge carries, such as [M+H+K]2+ were recorded as both "+H" and "+K". The charge distribution, Figure 2.1(B), in the Compendium ranged from +1 to +31 for cations and -1 to -3 for anions. More than 90% of the compounds were singly or doubly charged. Overall, replicate measurements were highly reproducible as evaluated by RSD. The global average RSD was 0.25%, and 97% of all
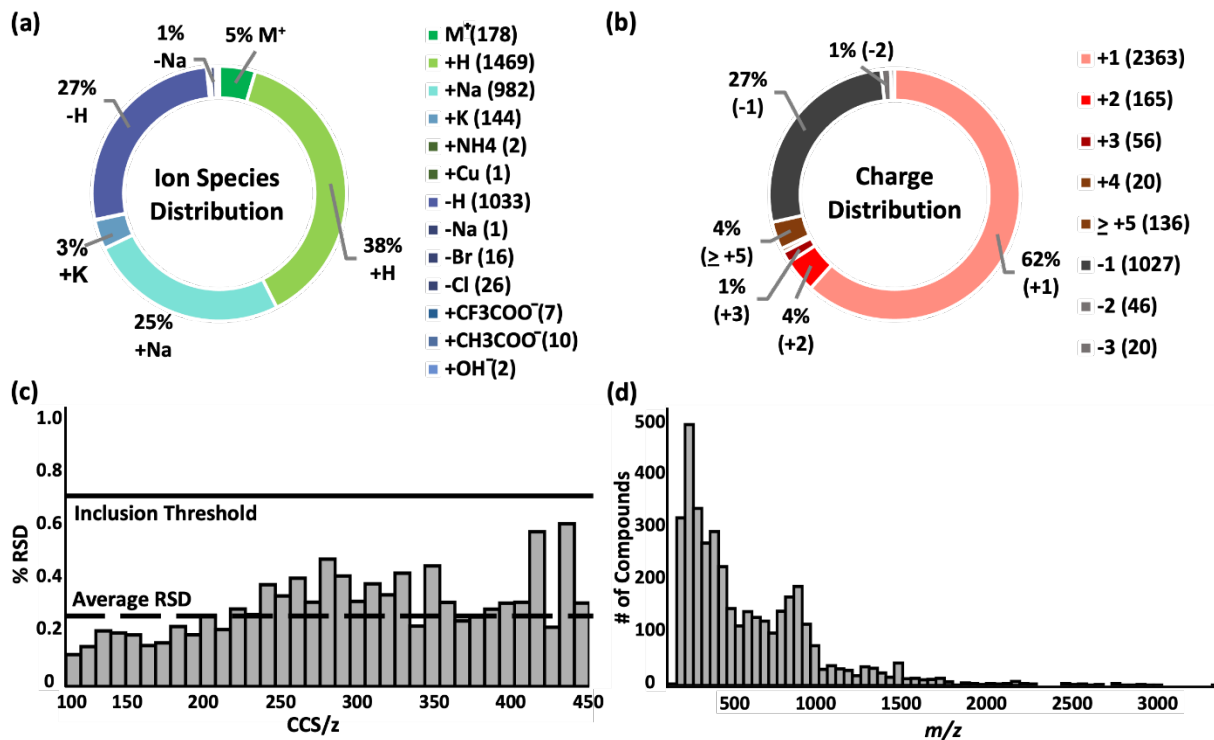
Figure 2.1 Compendium metadata statistics. (A-B) Overall distribution of the 3833 measured ions from (+) and (-) ion polarity modes by ion species and charge state. (C) Relative standard deviation (RSD) of all measurements binned by CCS/z. Global average RSD is 0.25%, and Compendium RSD threshold is 0.7%. (D) Distribution of ions contained in the database as a function of the *m/z*.

compounds had an RSD of < 1.0%. The average RSD per CCS/z bin is shown in Figure 2.1(C). RSD is observed to increase as CCS/z increases due to multiple observed conformers in larger molecules. Under highly controlled interlaboratory experimental conditions, RSD is < 0.3%;[12] and the empirical RSD threshold of 0.7% for the Compendium is a practical limit for data from independent studies. The Compendium data set spans a *m/z* range of ca. 74 to ca. 3300 Da; however, most compounds are less than 1500 Da. The full distribution of compound masses is shown in Figure 2.1(D).

2.3.2   *CCS Compendium Visualization*

The data set was visualized using code written in the R language. The graphical user interface (GUI) of the Unified CCS Compendium is shown in Figure 2.2(A) and is accessible online.[32] The default view for this GUI is to show all data grouped by super class. Users have the ability to zoom and select regions of interest which facilitates maneuvering densely populated areas. By hovering the cursor over any data point, as shown in Figure 2.2(B), users can access specific information regarding the corresponding entry including the compound's name, molecular formula, CAS identity, *m/z*, observed charge species, CCS/z and associated RSD, source citation, and digital object identifier. The interactive GUI can be tailored to the user's needs. Search functionality allows users to find data on any compound within the Compendium's compound table. Users can also isolate a specific data subset based on ion polarity, adduct type, super class, class, and data source. Subsetting data by super class or class reveals its CCS/z vs *m/z* area of occupancy.

The Compendium covers 14 super classes which delineate into 80 classes and 157 subclasses. The distribution of compounds into each super class is summarized in Figure 2.2(C). A list of super classes including *m/z* range and number of compounds per super class is summarized in Table 2.1.
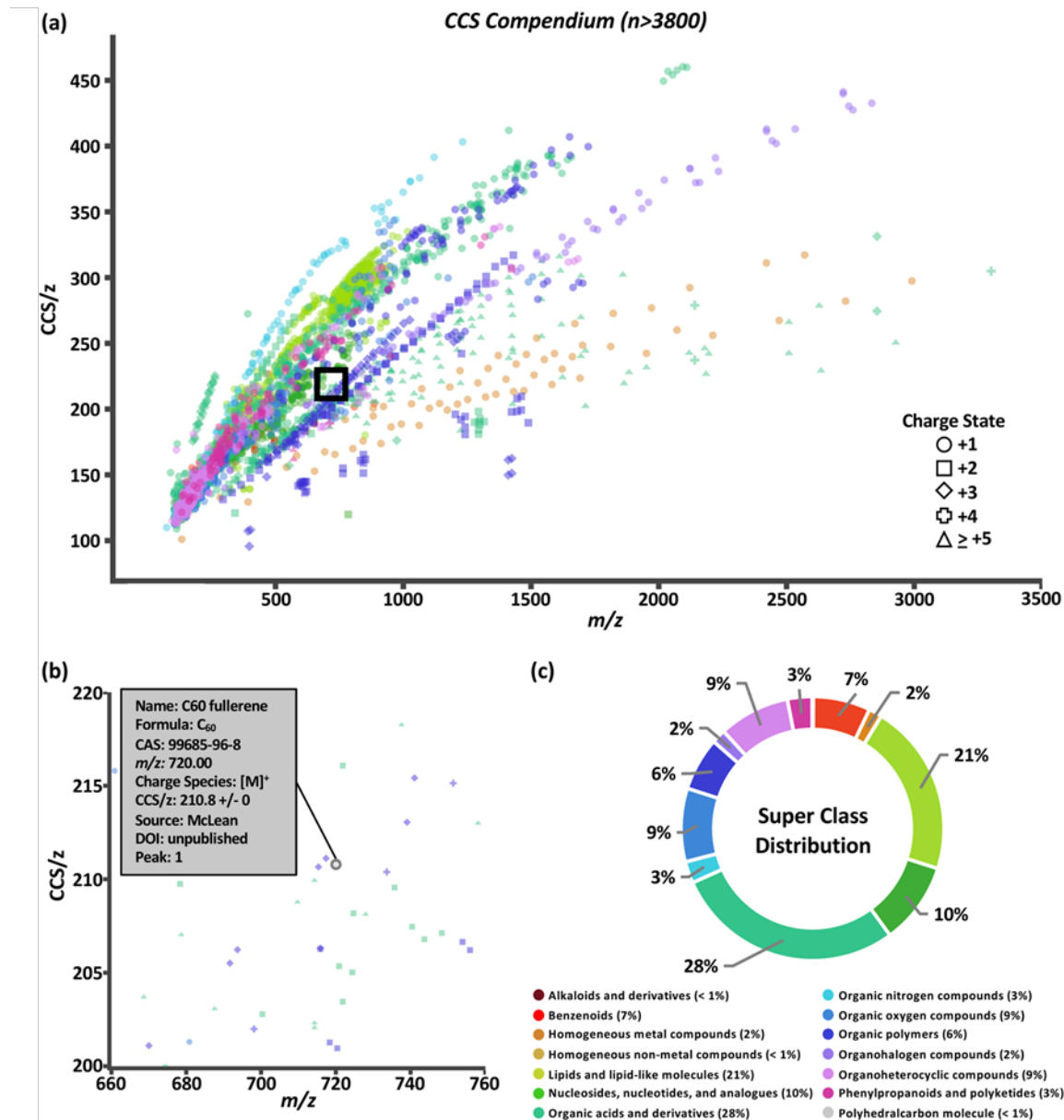
Figure 2.2 Compendium user interface (A) depicting measured data points classified into super classes indicated in the legend above. An enlarged version of the area within the black box is shown in (B) to illustrate how each data point reveals an information box in the online Compendium. (C) Distribution of compounds across the 14 structural super classes.

Super classes and their subsequent classes are further delineated in Table C1. Full classification of individual compounds can be found on the web-based Compendium.[32] Of the 80 classes, 48 had a sufficient (n > 10) number of data points to undergo regression fitting tests. In total, 24 classes and 24 subclasses were modeled. As new data will be added and regression fitting algorithms are iterative, it should be noted that the most up-to-date regression model equations can be found online.[32] A few observations can be made from thce data fitting study. Both four-parameter (Appendix C, eq. C2) and five-parameter (Appendix C, eq. C3) regressions were the best fit more frequently for classes in which m/z range included masses under 200 Da. This suggests a potential minimum observable CCS due to the asymptotic nature of sigmoidal curves. In theory, the IM-derived CCS will converge on the CCS of the neutral drift gas which, for sufficiently low CCS measurements, should manifest as a non-zero y-intercept in these CCS/z vs m/z projections. In the canonical literature, this minimally observable IM measurement is referred to as the gas polarization limit.[39] The smallest CCS/z measurement in the Compendium is 100.81 Å2, for a single cesium cation at m/z 132.90. Presently, more data points are needed to generate functional forms of a global type of fit.

2.3.3   Predictive Structural-Chemical Trends

While the Compendium visualizes the simple, yet fundamental aspects of the relationship between CCS/z and m/z, its highest utility lies in its predictive potential. To support predictive analysis, a 99% CI and 99% PI were generated as described in Appendix C, eq. C4 and eq. C5 for each class fit with a nonlinear regression. Briefly, the CI depicts the value range in which the regression mean is expected to be for normally distributed data.[58] For our data, the mean CCS/z value for a given m/z should be contained within the CI in 99% of cases. The upper and lower CI limits are depicted as the outer solid lines. throughout Figure 2.3. The distance between the two limits is closest where

Table 2.1 Curated CCS Compendium Super Classes

Super Class	*m/z* Range	N
Alkaloids and derivatives	138 – 609	4
Benzenoids	108 – 887	269
Homogeneous metal compounds	132 – 2991	62
Homogeneous non-metal compounds	144	1
Lipids and lipid-like molecules	125 – 1017	810
Nucleosides, nucleotides, and analogues	226 – 809	386
Organic acids and derivatives	89 – 3302	1085
Organic nitrogen compounds	74 – 1233	102
Organic oxygen compounds	105 – 1506	345
Organic polymers	294 – 1724	250
Organohalogen compounds	301 – 2834	66
Organoheterocyclic compounds	96 – 1684	335
Phenylpropanoids and polyketides	133 – 1424	116
Polyhedralcarbon molecules	210 – 227	2

the data point density is highest and prediction error is lowest along the regression model. The 99% PI depicts the 'y' variable value (CCS/z) range expected for 99% of data points at a given 'x' value (*m/z*).[58]

Figure 2.3 is a representative example of this data correlation process. It depicts the super class "Organoheterocyclic compounds" which contain many human metabolites and natural products. Three classes within "Organoheterocyclic compounds" are shown in Figure 2.3(B-D). The "Quinolines and derivatives" (Figure 2.3(B)) and "Imidazopyrimidines" classes (Figure 2.3(C)) were best fit by a 4P regression model. The "Pteridines and derivatives" class (Figure 2.3(D)) was fit best by the PF regression. In these cases, data fit regressions and corresponding CIs and PIs define the CCS/z vs *m/z* space that 99% of data for diazines, imidazopyrimidines, and pteridines and derivatives should occupy. While current AICc values indicate these models are appropriate, the specificity and predictability of these intervals will improve with the inclusion of more data and further delineation of each class into subclasses.

In the Compendium, the 99% interval bands included in the data projections are calculated directly from the Compendium data, therefore most of the empirical measurements within the dataset will fall within these interval bands. As these bands represent a probability, there remains the possibility that CCS values for compound standards will fall outside of these projections, and users should examine these cases on an individual basis to determine if CCS values are repeatably and reproducibly outside of the predicted range. For example, multimers dissociating occurring after the ion mobility measurement but prior to mass analysis (i.e, post-mobility ion activation) would lead to a larger than expected drift time and corresponding CCS. Additionally, CCS values for unknown analytes/isomers obtained from untargeted experiments represent previously
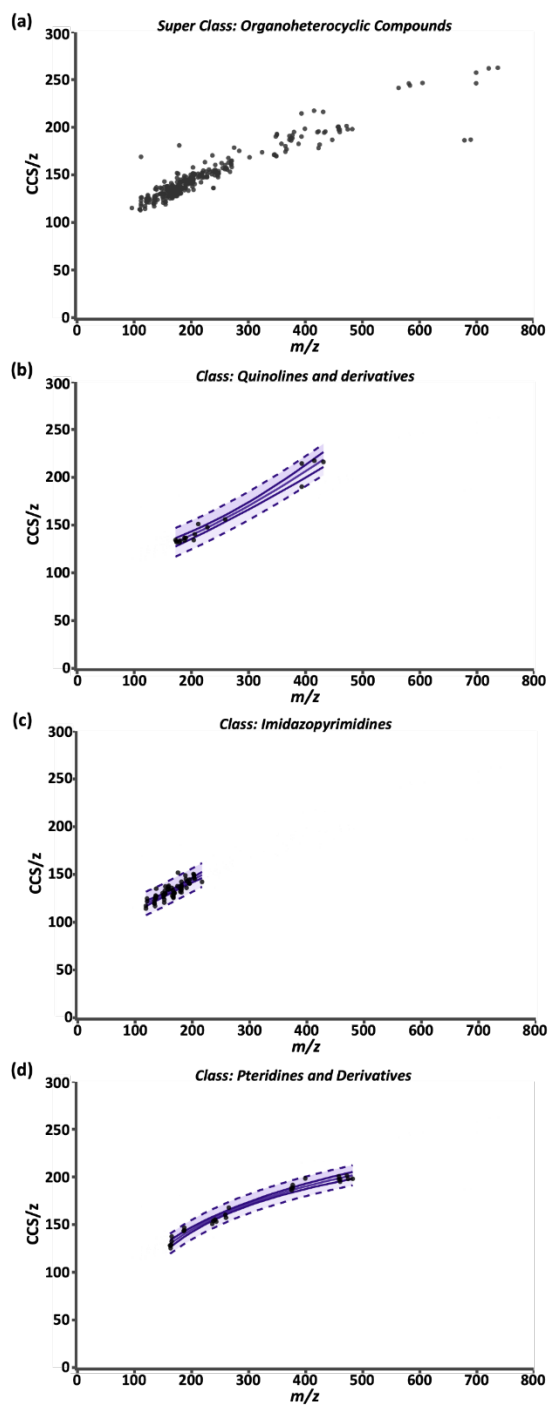
Figure 2.3 Compendium regression models (A) Compendium GUI output of all ion entries within the "Organoheterocyclics" super class. (B) "Quinolines and derivatives" class; and a 4P regression. (C) "Imidazopyrimidines" class; and a 4P regression. (D) "Pteridines and Derivatives"" class; and a PF regression. For (B-D), the center solid line is the regression model, outer solid lines are 99% confidence intervals and the dash lines are 99% prediction intervals.

unmeasured peak features which could fall outside of the interval bands. In these scenarios, the user should exercise caution in determining if the predicted structural class is appropriate.

2.3.4   The Compendium as an Identification Filter

To test the predictability and filtering abilities of the Compendium, metabolites extracted from control human serum were analyzed using classical LC-MS or LC-IM-MS workflows. In the LC-MS data, 4719 deconvoluted features were observed with 955 tentative identifications matched by conservative criteria for exact mass (< 10 ppm) and isotopic distribution (70%) using METLIN and LipidBlast databases.[55,56] In order to append drift time values to these tentative identifications, an in-house python script was developed (available online).[45] Using this script, we can extract drift times at a rate of 4x105 measurements per hour, per sample. Drift times from each of the three technical replicates were aligned to the tentative identifications based on retention time and *m/z*. In these data, most of the aligned drift times were self-consistent with an RSD < 1%. The drift times were averaged and used to calculate CCS/z values using the single-field extension of the Mason-Schamp relationship. Superimposing the tentatively identified serum metabolite data over the Unified CCS Compendium data (Figure 2.4(A)) indicates that the tentatively identified compounds have equivalent mobility-mass correlations as known chemical compounds.

For proof-of-concept purposes, the serum data was subset into compounds tentatively identified as lipids. Compounds in the green highlighted area of Figure 2.4(A) represent the CCS/z vs *m/z* space within the Unified CCS Compendium containing all lipid regressions generated for data in the "Lipids and lipid-like molecules" super class. In total, 550 compounds present in the serum sample were tentatively identified as lipids; and 422 of these compounds overlapped with at least one of the lipid classes and/or lipid subclass regression models. Distance from the mean values were calculated to prioritize the probability that a serum compound belonged to a given lipid class.

36

An example of this process is depicted in Figure 2.4(B) for the compound with *m/z* 744.49 and CCS/z 278.2 Å2 (gold circle). Potential tentative identifications for *m/z* 744.49 included 53 glycerophosphocholine (PC) and glycerophosphoethanolamine (PE) isomers. This unknown compound (gold line, Figure 2.4(B) call-out box) was 9.44 standard deviations from the PE subclass regression model (red line, Figure 2.4(B) call-out box), but at 2.54 standard deviations from the PC subclass regression model (blue line, Figure 2.4(B) call-out box), this compound was within the 99% CI of the PC fit with a difference of about 1.5 Å2 from the mean CCS/z value of PCs (at *m/z* 744). Thus, by comparing an unknown compound to statistical fits of existing data within the Unified CCS Compendium, we made a putative identification and have higher confidence in its assignment.

The molecular identification workflow for *m/z* 744.49 is summarized in Figure 2.4(C). The *m/z* 744.49 was deconvoluted to its neutral mass of 705.53 Da. At unit resolution, there are tens of thousands of potential chemical formulas with a mass of 705 Da. Within 100 ppm mass error of 705.53 Da, there are 7276 possible chemical formulas. Subsequently, there are 653 chemical formulas within 10 ppm mass error and 325 chemical formulas within 5 ppm mass error (the observed mass error). Of these 325 formulas, 173 are known compounds found in the PubChem database. Heuristic filtering based on instrumentation mass accuracy, mass defect, isotope distribution, and information from orthogonal separations enables tentative identification of compounds with a specified level of confidence. In this example, 53 tentative PC and PE identifications were returned after heuristic filtering through Progenesis QI. Using the Compendium, this list can be further narrowed into 21 PC isomers with the neutral mass 705.53 and *m/z* 744.49.
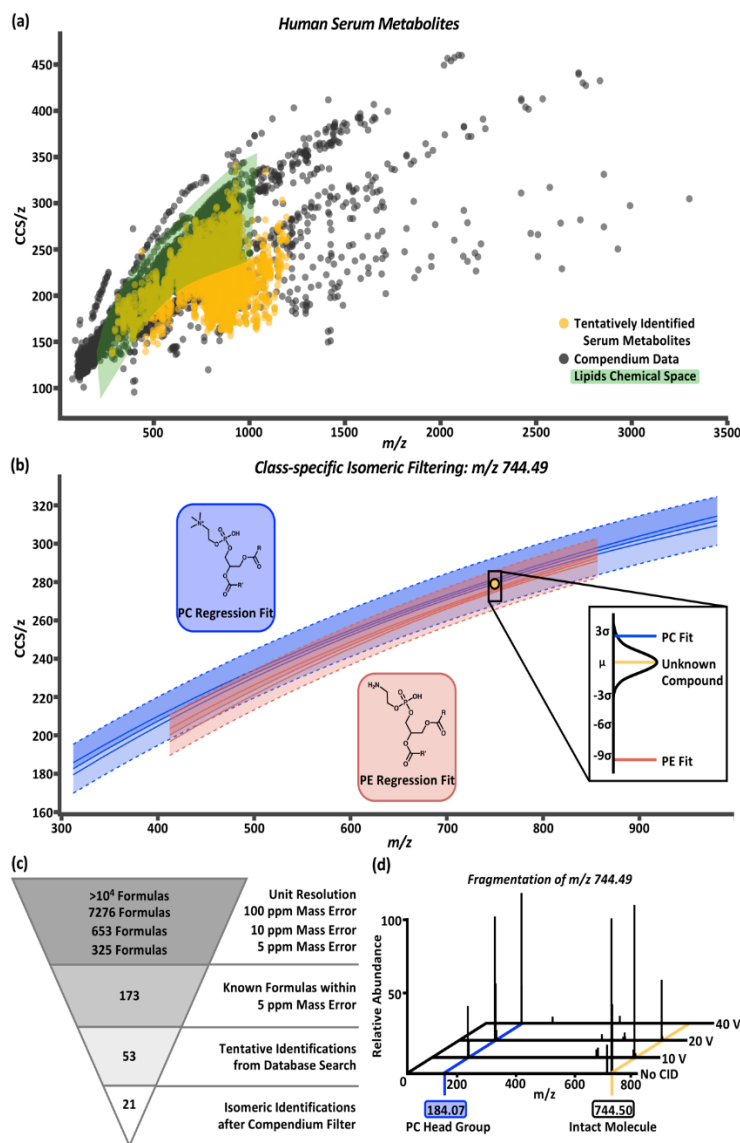
Figure 2.4 Compendium annotation filtering workflow (A) Overlay of human serum metabolites (gold) with the Compendium (black). Green area represents CCS/z vs *m/z* space occupied by any and all lipid subsets within the Compendium. (B) Example plot for class-specific filtering of an unknown serum compound, *m/z* 744.49 and CCS 278.2 Å^2 (gold circle), tentatively identified as a PC (blue regression model) or PE (red regression model). The probability of the unknown compound's class falling within the PC or PE class is shown in the call out box. Based on distance from the mean calculations, the compound falls within 2.54 standard deviations of the PC regression model and 9.44 standard deviations of the PE regression model which indicates the unknown compounds has a higher probability of being a PC than a PE. (C) Molecular identification workflow for the unknown compound depicted in panel b. After Compendium filtering, identifications were reduced to 21 PC isomers with the *m/z* 744.49. (D) Fragmentation of the isolated *m/z* 744.49 at CID 0V, 10V, 20V, and 40V. An increase in the intensity for *m/z* 184.07, corresponding to the phosphocholine head group mass, is observed with increase in collision voltage.

To validate our PC prediction, *m/z* 744.49 underwent mass isolation from the serum matrix and was fragmented using collision induced dissociation at 0 V, 10 V, 20 V, and 40 V. The mass spectra, shown in Figure 2.4(D), demonstrate the increase in the intensity of *m/z* 184.07, the signature *m/z* of a phosphocholine head group, as collision energy increased. While further investigation using chemical standards can lead to high-confidence identifications of unknown compounds, using the CCS filtering workflow presented here allows investigators to achieve high confidence in assigning the chemical class to an unknown molecule using IM-MS datasets. This predictive ability is expected to be particularly important for chemical class and structure annotation of isomers belonging to known compounds from which CCS information has not been previously measured (i.e., an "unknown unknown" isomer), as is the case for most human metabolites which are expected to be isomeric but current undiscovered.[59]

2.4   Conclusions and Future Directions

In this work, we illustrate the utility of IM-MS in quantitatively characterizing biochemical species using a Unified CCS Compendium. Prior to this work, quantitative CCS libraries have been limited in scope to a narrow range of chemical classes, polarities, and adduct types. Therefore, we curated a Unified CCS Compendium obtained from chemical standards representing a wide variety of structures spanning 14 super classes, 80 classes, and 157 subclasses. We anticipate subsequent contributions from the IM-MS community; thus the informatics infrastructure was designed to accommodate future expansion. The current biochemical species contained within the Unified CCS Compendium enabled generation of optimized nonlinear regression models with CI and PI for 45 classes and subclasses. These models enabled prediction and filtering of unknown biochemical species. The capabilities demonstrated in this manuscript establish a foundation for utilizing CCS/z as an additional molecular characterization dimension. The Unified CCS

compendium was used to predict and identify unknown chemical species originating from a serum sample. Future work will focus on expanding the number of entries in the Compendium to improve predictive power.

We anticipate the Unified CCS Compendium to be a collaborative effort of the IM-MS community and invite contributions to this open-access repository for quality-controlled CCS measurements. Specific guidelines for submitting data can be accessed from the interactive online interface.[32] While the Compendium is initially designed to only include DTIMS data, considerations for adding CCS information obtained from other IM techniques will be included in future revisions. The standardized DTIMS CCS measurements contained within the Compendium can serve as calibrant reference values for other IM techniques, which will enable the incorporation of more CCS data into this body of work.

2.5   Acknowledgements

2.6   References

(1)   Houle, D.; Govindaraju, D. R.; Omholt, S. Phenomics: The next Challenge. *Nature Reviews Genetics*. 2010. https://doi.org/10.1038/nrg2897.

(2)   May, J. C.; Gant-Branum, R. L.; McLean, J. A. Targeting the Untargeted in Molecular Phenomics with Structurally-Selective Ion Mobility-Mass Spectrometry. *Current Opinion in Biotechnology* **2016**, *39*, 192–197. https://doi.org/10.1016/j.copbio.2016.04.013.

(3)   May, J. C.; McLean, J. A. Advanced Multidimensional Separations in Mass Spectrometry: Navigating the Big Data Deluge. *Annual Review of Analytical Chemistry* **2016**, *9* (1), 387–409. https://doi.org/10.1146/annurev-anchem-071015-041734.

(4)   Paul, S. M.; Mytelka, D. S.; Dunwiddie, C. T.; Persinger, C. C.; Munos, B. H.; Lindborg, S. R.; Schacht, A. L. How to Improve RD Productivity: The Pharmaceutical Industry's Grand Challenge. *Nature Reviews Drug Discovery* **2010**, *9* (3), 203–214. https://doi.org/10.1038/nrd3078.

(5)   Quinn, R. A.; Navas-Molina, J. A.; Hyde, E. R.; Song, S. J.; Vázquez-Baeza, Y.; Humphrey, G.; Gaffney, J.; Minich, J. J.; Melnik, A. v; Herschend, J.; DeReus, J.; Durant, A.; Dutton, R. J.; Khosroheidari, M.; Green, C.; da Silva, R.; Dorrestein, P. C.; Knight, R. From Sample to Multi-Omics Conclusions in under 48 Hours. *mSystems* **2016**, *1* (2). https://doi.org/10.1128/mSystems.00038-16.

(6)   Chen, R.; Mias, G. I.; Li-Pook-Than, J.; Jiang, L.; Lam, H. Y. K.; Chen, R.; Miriami, E.; Karczewski, K. J.; Hariharan, M.; Dewey, F. E.; Cheng, Y.; Clark, M. J.; Im, H.; Habegger, L.; Balasubramanian, S.; O'Huallachain, M.; Dudley, J. T.; Hillenmeyer, S.; Haraksingh, R.; Sharon, D.; Euskirchen, G.; Lacroute, P.; Bettinger, K.; Boyle, A. P.; Kasowski, M.; Grubert, F.; Seki, S.; Garcia, M.; Whirl-Carrillo, M.; Gallardo, M.; Blasco, M. A.; Greenberg, P. L.; Snyder, P.; Klein, T. E.; Altman, R. B.; Butte, A. J.; Ashley, E. A.; Gerstein, M.; Nadeau, K. C.; Tang, H.; Snyder, M. Personal Omics Profiling Reveals Dynamic Molecular and Medical Phenotypes. *Cell* **2012**, *148* (6), 1293–1307. https://doi.org/10.1016/j.cell.2012.02.009.

(7)   Zimmer, J. S. D.; Monroe, M. E.; Qian, W.-J.; Smith, R. D. Advances in Proteomics Data Analysis and Display Using an Accurate Mass and Time Tag Approach. *Mass Spectrometry Reviews* **2006**, *25* (3), 450–482. https://doi.org/10.1002/mas.20071.

(8)   Zheng, X.; Wojcik, R.; Zhang, X.; Ibrahim, Y. M.; Burnum-Johnson, K. E.; Orton, D. J.; Monroe, M. E.; Moore, R. J.; Smith, R. D.; Baker, E. S. Coupling Front-End Separations, Ion Mobility Spectrometry, and Mass Spectrometry For Enhanced Multidimensional Biological and Environmental Analyses. *Annual Review of Analytical Chemistry* **2017**, *10* (1), 71–92. https://doi.org/10.1146/annurev-anchem-061516-045212.

(9)   McLean, J. A.; Ruotolo, B. T.; Gillig, K. J.; Russell, D. H. Ion Mobility–Mass Spectrometry: A New Paradigm for Proteomics. *International Journal of Mass Spectrometry* **2005**, *240* (3), 301–315. https://doi.org/10.1016/j.ijms.2004.10.003.

(10)   Hines, K. M.; May, J. C.; McLean, J. A.; Xu, L. Evaluation of Collision Cross Section Calibrants for Structural Analysis of Lipids by Traveling Wave Ion Mobility-Mass Spectrometry. *Analytical Chemistry* **2016**, *88* (14), 7329–7336. https://doi.org/10.1021/acs.analchem.6b01728.

(11)   Ridenour, W. B.; Kliman, M.; McLean, J. A.; Caprioli, R. M. Structural Characterization of Phospholipids and Peptides Directly from Tissue Sections by MALDI Traveling-Wave Ion Mobility-Mass Spectrometry. *Analytical Chemistry* **2010**, *82* (5). https://doi.org/10.1021/ac9026115.

(12)   Stow, S. M.; Causon, T. J.; Zheng, X.; Kurulugama, R. T.; Mairinger, T.; May, J. C.; Rennie, E. E.; Baker, E. S.; Smith, R. D.; McLean, J. A.; Hann, S.; Fjeldsted, J. C. An Interlaboratory Evaluation of Drift Tube Ion Mobility–Mass Spectrometry Collision Cross Section Measurements. *Analytical Chemistry* **2017**, *89* (17), 9048–9055. https://doi.org/10.1021/acs.analchem.7b01729.

(13)   Lietz, C. B.; Yu, Q.; Li, L. Large-Scale Collision Cross-Section Profiling on a Traveling Wave Ion Mobility Mass Spectrometer. *Journal of The American Society for Mass Spectrometry* **2014**, *25* (12), 2009–2019. https://doi.org/10.1007/s13361-014-0920-1.

(14)   Struwe, W. B.; Pagel, K.; Benesch, J. L. P.; Harvey, D. J.; Campbell, M. P. GlycoMob: An Ion Mobility-Mass Spectrometry Collision Cross Section Database for Glycomics. *Glycoconjugate Journal* **2016**, *33* (3), 399–404. https://doi.org/10.1007/s10719-015-9613-7.

(15)   Hines, K. M.; Ross, D. H.; Davidson, K. L.; Bush, M. F.; Xu, L. Large-Scale Structural Characterization of Drug and Drug-Like Compounds by High-Throughput Ion Mobility-Mass Spectrometry. *Analytical Chemistry* **2017**, *89* (17), 9023–9030. https://doi.org/10.1021/acs.analchem.7b01709.

(16)   Zhou, Z.; Tu, J.; Xiong, X.; Shen, X.; Zhu, Z. J. LipidCCS: Prediction of Collision Cross-Section Values for Lipids with High Precision to Support Ion Mobility-Mass Spectrometry-Based Lipidomics. *Analytical Chemistry* **2017**, *89* (17), 9559–9566. https://doi.org/10.1021/acs.analchem.7b02625.

(17)   Hernández-Mesa, M.; le Bizec, B.; Monteau, F.; García-Campaña, A. M.; Dervilly-Pinel, G. Collision Cross Section (CCS) Database: An Additional Measure to Characterize Steroids. *Analytical Chemistry* **2018**, *90* (7), 4616–4625. https://doi.org/10.1021/acs.analchem.7b05117.

(18)   Zheng, X.; Aly, N. A.; Zhou, Y.; Dupuis, K. T.; Bilbao, A.; Paurus, V. L.; Orton, D. J.; Wilson, R.; Payne, S. H.; Smith, R. D.; Baker, E. S. A Structural Examination and Collision Cross Section Database for over 500 Metabolites and Xenobiotics Using Drift Tube Ion Mobility Spectrometry. *Chem. Sci.* **2017**, *8*, 7724–7736. https://doi.org/10.1039/C7SC03464D.

(19)   Zhou, Z.; Shen, X.; Tu, J.; Zhu, Z.-J. Large-Scale Prediction of Collision Cross-Section Values for Metabolites in Ion Mobility-Mass Spectrometry. *Analytical Chemistry* **2016**, *88* (22), 11084–11091. https://doi.org/10.1021/acs.analchem.6b03091.

(20)   Paglia, G.; Williams, J. P.; Menikarachchi, L.; Thompson, J. W.; Tyldesley-Worster, R.; Halldórsson, S.; Rolfsson, O.; Moseley, A.; Grant, D.; Langridge, J.; Palsson, B. O.; Astarita, G. Ion Mobility Derived Collision Cross Sections to Support Metabolomics Applications. *Analytical Chemistry* **2014**, *86* (8), 3985–3993. https://doi.org/10.1021/ac500405x.

(21)   Righetti, L.; Bergmann, A.; Galaverna, G.; Rolfsson, O.; Paglia, G.; Dall'Asta, C. Ion Mobility-Derived Collision Cross Section Database: Application to Mycotoxin Analysis. *Analytica Chimica Acta* **2018**, *1014*, 50–57. https://doi.org/10.1016/j.aca.2018.01.047.

(22)   Goodwin, C. R.; Fenn, L. S.; Derewacz, D. K.; Bachmann, B. O.; McLean, J. A. Structural Mass Spectrometry: Rapid Methods for Separation and Analysis of Peptide Natural Products. *Journal of Natural Products* **2012**, *75* (1), 48–53. https://doi.org/10.1021/np200457r.

(23)   Lian, R.; Zhang, F.; Zhang, Y.; Wu, Z.; Ye, H.; Ni, C.; Lv, X.; Guo, Y. Ion Mobility Derived Collision Cross Section as an Additional Measure to Support the Rapid Analysis of Abused Drugs and Toxic Compounds Using Electrospray Ion Mobility Time-of-Flight Mass Spectrometry. *Analytical Methods* **2018**, *10* (7), 749–756. https://doi.org/10.1039/c7ay02808c.

(24)   Chai, M.; Young, M. N.; Liu, F. C.; Bleiholder, C. A Transferable, Sample-Independent Calibration Procedure for Trapped Ion Mobility Spectrometry (TIMS). *Analytical Chemistry* **2018**, *90* (15), 9040–9047. https://doi.org/10.1021/acs.analchem.8b01326.

(25)   Gabelica, V.; Marklund, E. Fundamentals of Ion Mobility Spectrometry. *Current Opinion in Chemical Biology* **2018**, *42*, 51–59. https://doi.org/10.1016/j.cbpa.2017.10.022.

(26)   Blaženović, I.; Kind, T.; Ji, J.; Fiehn, O. Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics. *Metabolites* **2018**, *8* (2), 31. https://doi.org/10.3390/metabo8020031.

(27)   Ma, J.; Casey, C. P.; Zheng, X.; Ibrahim, Y. M.; Wilkins, C. S.; Renslow, R. S.; Thomas, D. G.; Payne, S. H.; Monroe, M. E.; Smith, R. D.; Teeguarden, J. G.; Baker, E. S.; Metz, T. O. PIXiE: An Algorithm for Automated Ion Mobility Arrival Time Extraction and Collision Cross Section Calculation Using Global Data Association. *Bioinformatics* **2017**, *33* (17), 2715–2722. https://doi.org/10.1093/bioinformatics/btx305.

(28)   MacLean, B. X.; Pratt, B. S.; Egertson, J. D.; MacCoss, M. J.; Smith, R. D.; Baker, E. S. Using Skyline to Analyze Data-Containing Liquid Chromatography, Ion Mobility Spectrometry, and Mass Spectrometry Dimensions. *Journal of The American Society for Mass Spectrometry* **2018**, *29* (11), 2182–2188. https://doi.org/10.1007/s13361-018-2028-5.

(29)   Kirkwood, K. I.; Christopher, M. W.; Burgess, J. L.; Littau, S. R.; Pratt, B. S.; Shulman, N.; Tamura, K.; MacCoss, M. J.; MacLean, B. X.; Baker, E. S. Development and Application of Multidimensional Lipid Libraries to Investigate Lipidomic Dysregulation Related to Smoke Inhalation Injury Severity. *bioRxiv* **2021**, 2021.10.13.464246. https://doi.org/10.1021/acs.jproteome.1c00820.

(30) Ripley, B. D. Ripley, B.D. 1996 Pattern Recognition and Neural Networks Cambridge University Press. *Canbridge University Press* **1996**.

(31) Colby, S. M.; Thomas, D. G.; Nunez, J. R.; Baxter, D. J.; Glaesemann, K. R.; Brown, M.; Pirrung, M. A.; Govind, N.; Teeguarden, J. G.; Metz, T. O.; Ryan, S. ISiCLE : A Molecular Collision Cross Section Calculation Pipeline for Establishing Large in Silico Reference Libraries for Compound Identification ISiCLE : A Molecular Collision Cross Section Calculation Pipeline for Establishing Large in Silico Reference. **2018**, No. September.

(32) McLean Research Group. Unified CCS Compendium https://mcleanresearchgroup.shinyapps.io/CCS-Compendium/.

(33) Nichols, C. M.; May, J. C.; Sherrod, S. D.; McLean, J. A. Automated Flow Injection Method for the High Precision Determination of Drift Tube Ion Mobility Collision Cross Sections. *The Analyst* **2018**, *143* (7), 1556–1559. https://doi.org/10.1039/C8AN00056E.

(34) May, J. C.; Goodwin, C. R.; Lareau, N. M.; Leaptrot, K. L.; Morris, C. B.; Kurulugama, R. T.; Mordehai, A.; Klein, C.; Barry, W.; Darland, E.; Overney, G.; Imatani, K.; Stafford, G. C.; Fjeldsted, J. C.; McLean, J. A. Conformational Ordering of Biomolecules in the Gas Phase: Nitrogen Collision Cross Sections Measured on a Prototype High Resolution Drift Tube Ion Mobility-Mass Spectrometer. *Analytical Chemistry* **2014**, *86* (4), 2107–2116. https://doi.org/10.1021/ac4038448.

(35) Dodds, J. N.; May, J. C.; McLean, J. A. Investigation of the Complete Suite of the Leucine and Isoleucine Isomers: Toward Prediction of Ion Mobility Separation Capabilities. *Analytical Chemistry* **2017**, *89* (1), 952–959. https://doi.org/10.1021/acs.analchem.6b04171.

(36) May, J. C.; Jurneczko, E.; Stow, S. M.; Kratochvil, I.; Kalkhof, S.; McLean, J. A. Conformational Landscapes of Ubiquitin, Cytochrome c, and Myoglobin: Uniform Field Ion Mobility Measurements in Helium and Nitrogen Drift Gas. *International Journal of Mass Spectrometry* **2017**, *427*, 79–90. https://doi.org/10.1016/j.ijms.2017.09.014.

(37) Leaptrot, K. L.; May, J. C.; Dodds, J. N.; McLean, J. A. Ion Mobility Conformational Lipid Atlas for High Confidence Lipidomics. *Nature Communications* **2019**, *10* (1), 985. https://doi.org/10.1038/s41467-019-08897-5.

(38) Nichols, C. M.; Dodds, J. N.; Rose, B. S.; Picache, J. A.; Morris, C. B.; Codreanu, S. G.; May, J. C.; Sherrod, S. D.; McLean, J. A. Untargeted Molecular Discovery in Primary Metabolism: Collision Cross Section as a Molecular Descriptor in Ion Mobility-Mass Spectrometry. *Analytical Chemistry* **2018**, *90* (24), 14484–14492. https://doi.org/10.1021/acs.analchem.8b04322.

(39) Mason, E. A.; McDaniel, E. W. *Transport Properties of Ions in Gases*; John Wiley & Sons, Ltd.: New York City, NY, 1988.

(40) Siems, W. F.; Viehland, L. A.; Hill, H. H. Improved Momentum-Transfer Theory for Ion Mobility. 1. Derivation of the Fundamental Equation. *Analytical Chemistry* **2012**, *84* (22), 9782–9791. https://doi.org/10.1021/ac301779s.

(41)   The R Core. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria 2014.

(42)   Gabelica, V.; Shvartsburg, A. A.; Afonso, C.; Barran, P.; Benesch, J. L. P.; Bleiholder, C.; Bowers, M. T.; Bilbao, A.; Bush, M. F.; Campbell, J. L.; Campuzano, I. D. G.; Causon, T.; Clowers, B. H.; Creaser, C. S.; de Pauw, E.; Far, J.; Fernandez-Lima, F.; Fjeldsted, J. C.; Giles, K.; Groessl, M.; Hogan, C. J.; Hann, S.; Kim, H. I.; Kurulugama, R. T.; May, J. C.; McLean, J. A.; Pagel, K.; Richardson, K.; Ridgeway, M. E.; Rosu, F.; Sobott, F.; Thalassinos, K.; Valentine, S. J.; Wyttenbach, T. Recommendations for Reporting Ion Mobility Mass Spectrometry Measurements. *Mass Spectrometry Reviews* **2019**, *38* (3), 291–320. https://doi.org/10.1002/mas.21585.

(43)   Djoumbou Feunang, Y.; Eisner, R.; Knox, C.; Chepelev, L.; Hastings, J.; Owen, G.; Fahy, E.; Steinbeck, C.; Subramanian, S.; Bolton, E.; Greiner, R.; Wishart, D. S. ClassyFire: Automated Chemical Classification with a Comprehensive, Computable Taxonomy. *Journal of Cheminformatics* **2016**, *8* (1), 1–20. https://doi.org/10.1186/s13321-016-0174-y.

(44)   Feldman, H. J.; Dumontier, M.; Ling, S.; Haider, N.; Hogue, C. W. V. CO: A Chemical Ontology for Identification of Functional Groups and Semantic Comparison of Small Molecules. *FEBS Letters* **2005**, *579* (21), 4685–4691. https://doi.org/10.1016/j.febslet.2005.07.039.

(45)   McLean Research Group GitHub https://github.com/McLeanResearchGroup.

(46)   Morris, C. B.; May, J. C.; McLean, J. A. Predictive Mathematical Descriptors of Biological Class Trends in Ion Mobility-Mass Spectrometry Analysis. In *62th Annual Conference for the American Society of Mass Spectrometry*; Baltimore, MD, 2014.

(47)   May, J. C.; Morris, C. B.; McLean, J. A. Ion Mobility Collision Cross Section Compendium. *Analytical Chemistry* **2017**, *89* (2), 1032–1044. https://doi.org/10.1021/acs.analchem.6b04905.

(48)   Burnham, K. P.; Anderson, D. R. *Model Selection and Multimodel Interference*, 2nd ed.; Springer-Verlag: New York City, 1998.

(49)   Spiess, A.-N.; Neumeyer, N. An Evaluation of R2 as an Inadequate Measure for Nonlinear Models in Pharmacological and Biochemical Research: A Monte Carlo Approach. *BMC Pharmacology* **2010**, *10* (1), 6. https://doi.org/10.1186/1471-2210-10-6.

(50)   Wickham, H. Ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag: New York City, NY 2016.

(51)   Wickham, H. The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software* **2011**, *40* (1), 1–29. https://doi.org/10.18637/jss.v040.i01.

(52)   Sievert, C.; Parmer, C.; Hocking, T.; Chamberlain, S.; Ram, K.; Corvellec, M.; Despouy, P. Plotly: Create Interactive Web Graphics via "Plotly.Js."

(53)   Dowle, M.; Srinivasan, A. Data.Table: Extension of `data.Frame`. 2017.

(54)   Chang, W.; Cheng, J.; Allaire, J.; Xie, Y.; McPherson, J. Shiny: Web Application Framework for R. 2020.

(55)   Smith, C. a; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G. METLIN: A Metabolite Mass Spectral Database. *Therapeutic drug monitoring* **2005**, *27* (6), 747–751.

(56)   Kind, T.; Liu, K.-H.; Lee, D. Y.; DeFelice, B.; Meissen, J. K.; Fiehn, O. LipidBlast in Silico Tandem Mass Spectrometry Database for Lipid Identification. *Nature Methods* **2013**, *10* (8), 755–758. https://doi.org/10.1038/nmeth.2551.

(57)   Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M.-Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P. A Cross-Platform Toolkit for Mass Spectrometry and Proteomics. *Nature Biotechnology* **2012**, *30* (10), 918–920. https://doi.org/10.1038/nbt.2377.

(58)   Faraway, J. J. Practical Regression and Anova Using R. *Reproduction* **2002**, *21* (July). https://doi.org/10.1016/0360-1315(91)90006-D.

(59)   Wishart, D. S.; Feunang, Y. D.; Marcu, A.; Guo, A. C.; Liang, K.; Vázquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C.; Karu, N.; Sayeeda, Z.; Lo, E.; Assempour, N.; Berjanskii, M.; Singhal, S.; Arndt, D.; Liang, Y.; Badran, H.; Grant, J.; Serra-Cayuela, A.; Liu, Y.; Mandal, R.; Neveu, V.; Pon, A.; Knox, C.; Wilson, M.; Manach, C.; Scalbert, A. HMDB 4.0: The Human Metabolome Database for 2018. *Nucleic Acids Research* **2018**, *46* (D1), D608–D617. https://doi.org/10.1093/nar/gkx1089.

CHAPTER 3

IMPROVING CONFIDENCE IN LIPIDOMIC ANNOTATIONS BY INCORPORATING

EMPIRICAL ION MOBILITY REGESSION ANALYSIS AND CHEMICAL CLASS

PREDICTION[‡]

3.1   Introduction

Lipids and lipid-like molecules play critical roles in a diverse array of biological processes, including membrane structure, signaling, and energy storage.[1] Although the implications of their dysregulation in a number of diseases makes lipids of prime interest for study,[2] global analysis is often hindered by the wide range of chemical and physical properties arising from the structural diversity of this single superclass of molecules.[3–5] Mass spectrometry-based lipidomics has risen to meet this challenge, enabling high sensitivity and high throughput measurements which facilitates not only a broader understanding of lipid metabolism, but also the discovery of significant molecular signatures for further study.[6–8]

Although great strides have been made toward comprehensive annotation of the lipidome, confident compound identification remains a bottleneck in untargeted analyses.[9–11] Comprehensive lipid identification by accurate mass alone is unattainable due to the prevalence of isomeric species as well as the limited number of commercially available analytical standards.[8] Tandem MS (MS/MS) ion fragmentation is often used to aid in the identification of lipid species with support from both experimental and *in silico* libraries.[12–16] However, MS/MS approaches are

challenged by structurally-similar lipids and chimeric fragmentation data resulting from isobaric signals, such as those arising from complex biological matrices. Liquid chromatography (LC) has also been extensively used in lipidomic analyses to improve peak capacity, resolve isomers, and help mitigate ion suppression effects at the ionization source.[17,18] The use of retention time as a chemical descriptor for species identification has been supported by the expanding landscape of retention time libraries and prediction tools,[19,20] but these approaches are not easily applicable across different laboratories and platforms due to the influence of various experimental parameters and matrix effects on retention time variability.[21,22] Additionally, direct sample analysis techniques such as those utilized for MS imaging are incompatible with chromatography and thus cannot take advantage of retention time correlation. Despite these limitations, the integration of complementary analytical techniques into multidimensional MS strategies is necessary to expand coverage and confidence in lipidomic annotations.[23,24]

An analytical separation technique that has had increasing success within the lipidomics community is ion mobility.[19,24,25] This gas phase separation is rapid (milliseconds) and structurally selective, which can resolve isomers/isobars while providing an additional metric for compound annotation, namely the collision cross section (CCS).[26] CCS values provide direct, albeit coarse-grained, structural information for lipids and have been demonstrated to be highly reproducible across laboratories.[27] Further, the millisecond timescale of ion mobility measurements falls within the timescales of the LC and time-of-flight MS dimensions (minutes and microseconds, respectively)[28] and thus these analytical separations (LC-IM-MS) can be performed concurrently on a single sample injection. Many recent efforts have focused on the generation of CCS databases to aid in compound identification efforts.[29–34] Figure 3.1 compares lipid coverage of a large repository of standardized experimental CCS values, the Unified CCS Compendium,[31] to the
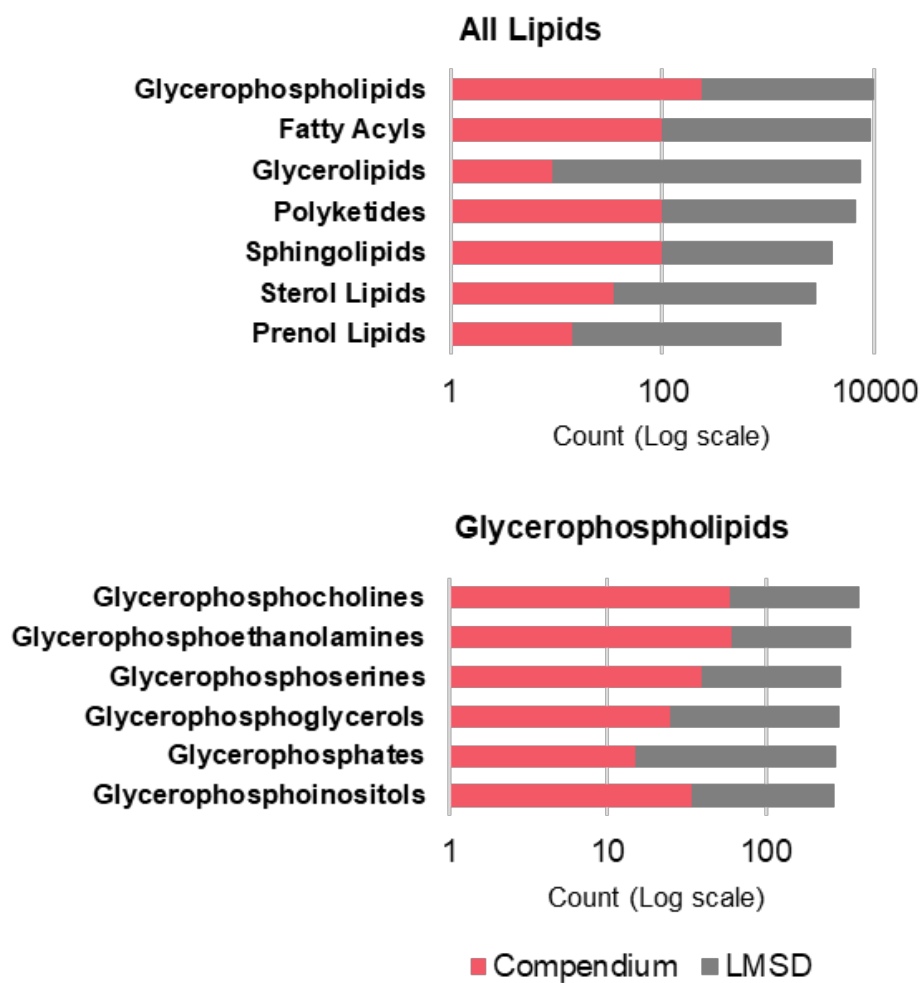
Figure 3.1 Lipid database coverage (log scale) of the LIPID MAPS Structural Data-base (LMSD, grey) and the Unified CCS Compendium (pink).

LIPID MAPS Structural Database.[12] Similar to MS/MS libraries, CCS database initiatives require substantial resources and expertise to curate and are limited by available chemical standards. However, the curation of such empirical databases has been widely successful in compiling CCS values for thousands of lipid species across a variety of classes and subclasses. Additionally, these databases have had significant success as training sets for developing large-scale theoretical CCS libraries using various predictive approaches.[34–38]

Here, we demonstrate an integrated LC-IM-MS lipidomics workflow supported by a sequential chemical class prediction informatics strategy to increase the confidence in lipidomic annotations. In both informatic approaches, the confidence is increased in accordance with the Metabolomics Standards Initiative (MSI) confidence level system, where additional molecular descriptors can be leveraged to effectively narrow the search space of candidate identifications.[10,39,40] First, using characteristic class-specific mobility-mass correlations (available in the Unified CCS Compendium), candidate identifications of lipidomic features are automatically filtered on the basis of their measured CCS values, increasing the confidence in resulting annotations. Second, features which do not yield candidate identifications from conventional database matching can be assigned tentative classes based on their CCS values using a previously developed machine learning framework (i.e., SIFTER).[37] Using this dual approach, increased lipidomic coverage can be achieved with high confidence (putative) in the annotations and resulting biological interpretations.

3.2   Experimental Methods

3.2.1   Materials

Optima grade water, acetonitrile (ACN), isopropanol (IPA), methanol (MeOH), and ammonium acetate; as well as bicinchoninic acid (BCA) assay reagents and albumin standards were purchased from Fisher Scientific (Fair Lawn, NJ). Anhydrous methyl-tert-butyl ether (MTBE), ammonium bicarbonate (NH_4HCO_3), heptadecanoic acid, and nonadecanoic acid were purchased from Sigma Aldrich (St. Louis, MO). Lipid standards including phosphatidylglycerol (PG, chicken egg) extract, phosphatidylinositol (PI, bovine liver) extract, glucosyl(β) sphingosine (d18:1), C17 ceramide (d18:1/17:0), and a mixture of heavy-labeled lipids SPLASH LIPIDOMIX were purchased from Avanti Polar Lipids (Alabaster, AL).

3.2.2 Instrumentation

All data were acquired using a commercial drift tube ion mobility-mass spectrometer (6560A IM-QTOF, Agilent Technologies, Santa Clara, CA) equipped with an electrospray source (Agilent Jet Stream)[41] operated under the following source conditions: gas temperature, 280 °C; drying gas, 12 L/min; nebulizer, 30 psi; sheath gas temperature, 300 °C; sheath gas flow, 11 L/min; capillary voltage (VCap), 3500 V; nozzle voltage, 1800 V; fragmentor, 320 V; and octopole 1 RF V_{pp}, 750 V. The drift tube was operated with 3.95 Torr of nitrogen gas with the pressure regulated in real time using closed-loop pressure controlling system (Alternative Drift Gas Kit, Agilent).[42] Additional drift tube parameters were as follows: ion trap fill time, 20 ms; ion trap release time, 300 μs; drift tube entrance, 1474 V; drift tube exit, 224 V; rear funnel entrance, 217.5 V; rear funnel RF, 150 V_{pp}; rear funnel exit, 45 V; and IM hexapole delta, -8 V. The QTOF stage was operated in Low Mass Range (*m/z* 50-1700), ion slicer operated at High Sensitivity and the digitizer operated at 2 GHz Extended Dynamic Range.

3.2.3 Preparation and analysis of standard lipid extracts

Purified TLC fractions of phosphatidylglycerol (PG) and phosphatidylinositol (PI, Avanti) were prepared to a final concentration of 10 µg/mL in IPA. The standard extracts were analyzed in both positive and negative ion modes via direct infusion IM-MS with a sample flow rate of 10 µL/min. Ion mobility arrival times for calibrant ions (ESI Low Concentration Tuning Mixture, Agilent) were used to calibrate single-field CCS values for lipid features. Where possible, these features were identified by exact mass measurements and the LIPID MAPS Structure Database.[12] This data was then collated and submitted to the Unified CCS Compendium to support downstream identification of PG and PI lipid subclasses.

3.2.4 Brain tissue lipid extraction

Murine brain tissue samples[43] were lysed using ice cold lysis buffer (1:1:2 MeOH:ACN:NH_4HCO_3 0.1M, pH 8.0), followed by sonication using a probe tip sonicator, 10 pulses at 30% power, cooled on ice between samples. Protein concentration was determined using a BCA assay, and samples were normalized to 200 µg total protein in 200 µL lysis buffer. 800 µL of cold MeOH was added to each sample, and the samples were incubated at -80 ˚C overnight to facilitate protein precipitation. These samples were centrifuged for 10 minutes at 10,000 rpm (4 ˚C), supernatant transferred and dried *in vacuo*. Dried samples were resuspended in 150 µL water, vortex thoroughly, followed by 150 µL MeOH, and incubated on ice for 10 minutes. Isotopically labeled lipid mixture (SPLASH LIPIDOMIX, Avanti) was added as an internal standard. A liquid-liquid extraction was performed by adding 1 mL MTBE, vortexing, and subsequently centrifuging for 10 minutes at 10,000 rpm and 4 ˚C. The nonpolar, MTBE layer containing the lipophilic components was removed and dried under vacuum. These fractions were resuspended for LC-IM-MS analysis in 100 µL IPA containing 40 µg/mL heptadecanoic acid and nonadecanoic acid, as well as 10 µg/mL glucosyl(β) sphingosine and N-heptadecanoyl-D-erythro-sphingosine.

3.2.5   Liquid chromatography

The lipophilic tissue extracts were analyzed using a 1290 Infinity LC system (Agilent). Reversed phase LC was performed at a flow rate of 300 µL/min by injecting 4 µL of the sample onto a C18 column (HypersilGold 1.9 µm, 2.1 mm x 100 mm column, Thermo Fisher) held at 40 ˚C with mobile phases consisting of 10 mM ammonium acetate in 9:1 H_2O:MeOH (A) and 5:3:2 IPA:MeOH:ACN (B). The gradient used was 70% B for 1 min, 70-86% B in 2.5 min, 86% B for 6.5 min, 86-100% B in 1 min, 100% B for 6 min, 100-70% B in 0.1 min, and 70% B for 1.9 min. The total duration of the method was 19 minutes, with a 5-minute post time to allow the column to equilibrate between injections.

3.2.6   Data acquisition

Samples were analyzed in both positive and negative ion modes. For each polarity, full scan MS data was acquired for all samples, while auto MS/MS data was acquired only for the pooled samples between batches. Top 2 and top 4 auto MS/MS data were acquired using an absolute threshold of 5000 counts and stepped collision energies at 20 and 40 V. Additionally for CCS determination, LC-IM-MS data from the pooled samples was acquired in triplicate. A previously established single field relationship derived from the fundamental ion mobility equation was used to determine the CCS values from the IM arrival time measurements of all detected features.[27]

3.2.7   Data processing

Initial processing of LC-IM-MS data involved saturation correction and smoothing in both the retention time and drift time dimensions using the PNNL PreProcessor (v. 2.0).[44] Ion mobility arrival times for calibrant ions were acquired at the beginning of experiments and applied offline to the individual data files to determine the single-field CCS values using IM-MS Browser (B.10, Agilent). IM-MS Browser was then used to apply a preset inclusion region of mass-mobility space

to each file, generating extracted files containing data limited to the IM-MS region in which lipidomic data is known to occupy.[45] This IM-MS prefiltering step helps to minimize artifactual and higher order charge state signals which are uncharacteristic of lipids. Finally, 4D feature finding was performed in Mass Profiler (B.10, Agilent). Processing of LC-MS and LC-MS/MS data including retention time alignment, charge carrier deconvolution, and molecular feature finding was performed in Progenesis QI (v2.3, Nonlinear Dynamics, Durham, NC). The resulting deisotoped/deconvoluted features represent discrete molecules and thus are referred to as molecular features. Tentative identifications were assigned to features via accurate mass measurement at a 10 ppm mass tolerance matched against entries from a combination of data repositories including METLIN Metabolite and Chemical Entity Database, Human Metabolome Database, LIPID MAPS Structure Database, and LipidBlast.[12,13,46,47] Here, annotation confidence levels (i.e., tentative identifications) are notated in accordance with the Metabolomics Standards Initiative (MSI) scheme, where higher confidence assignments can be derived from additional pieces of analytical information.[10,39,40]

3.2.8   Classification

All subsequent data processing was performed in the R statistical programming environment (v. 3.6.0) unless otherwise noted. Following tentative identification, all identified molecular features were assigned a hierarchical classification, including a kingdom, superclass, class, and subclass, in accordance with the structure-based comprehensive ontology, ChemOnt.[48] IUPAC International Chemical Identifier strings (InChIKeys) were assigned for each compound annotation using the Chemical Translation Service (CTS) *via* the R package, webchem (v0.4.0).[49,50] These InChIKeys were then used as input to the web application ClassyFire in order to assign each taxonomical classification.[51]

3.2.9   Collision Cross Section Filtering Pipeline

CCS values obtained from LC-IM-MS processing were appended to the tentatively identified molecular features (output from Progenesis QI) using a mass tolerance of 7 ppm and retention time tolerance of 0.5 minutes. Nonlinear least squares regression models were generated for known classes and subclasses within the Unified CCS Compendium using previously developed R scripts.[31] These regression models were used to filter the candidate identifications on two levels (Figure D2): the first filter was applied to all features whose candidate classes or subclasses had a representative regression model, assigning higher confidence to those annotations whose CCS fell within the 99% predictive interval associated with the class and subclass of that identification. The second filter was applied to all features whose annotations were from multiple subclasses and had passed the first filter. This filter calculated the distance of the CCS of the feature from the mean of the regression model of each of its candidate classes or subclasses. The annotations from the class or subclass whose model fell closest to the feature CCS are assigned a higher confidence (Figure D3). All analysis code and R scripts can be found on the McLean Research Group GitHub.[52]

3.3.10 SIFTER Chemical Class Prediction

Chemical class prediction of unknown molecules for which the previous filtering pipeline yielded no results were subjected to predictive analysis using the previously developed Supervised Inference of Feature Taxonomy from Ensemble Randomization (SIFTER) algorithm.[37] SIFTER utilizes a random forest machine learning approach to assign chemical class predictions and has been described elsewhere.[37]

3.3   Results and discussion

3.3.1   Expansion of Compendium Lipid Coverage

To support the broader applicability of the proposed analysis pipeline, efforts were made to expand the lipid subclass coverage of the CCS Compendium to include PG and PI lipids, which are major glycerophospholipid subclasses. Total purified extracts of PG and PI were analyzed via IM-MS, and CCS values of the identified features were calibrated and added to the Compendium. This analysis resulted in 35 new PG CCS values and 34 new PI CCS values, expanding the overall glycerophospholipid coverage (Figure 3.1). Addition of this data to the Compendium enabled the generation of two new subclass regression models for downstream predictive analysis (Figure D3).

3.3.2   Workflow Design and Assessment

An overview of the parallel workflow incorporating the ion mobility CCS data is shown in Figure 3.2. After alignment, deconvolution, and feature finding, the molecular features are searched against accurate mass databases to assign tentative identifications. The features that have accurate mass candidates in a database were assigned tentative identifications and are subjected to CCS filtering to produce more high-confidence IDs (in this case, putative annotations - level 2).[39] Unidentified features which did not match any accurate mass database entries are submitted to SIFTER to predict molecular classifications (Figure 3.2(B)).

Initial assessment of the CCS filtering approach was performed using LC-IM-MS data acquired from a standard mix of isotopically-labeled lipids (SPLASH LIPIDOMIX, Avanti). The masses of the molecular features corresponding to the known components of the mixture were corrected to their unlabeled counterparts to facilitate database matching. These corrected masses were assigned tentative identifications from the LIPID MAPS Structure Database to verify that the filtering pipeline would determine the correct option for each feature as expected based on its CCS value. In this preliminary test with molecular features corresponding to the standard mix components, the correct class was assigned the highest confidence for 24 of 26 features (92%), and the correct
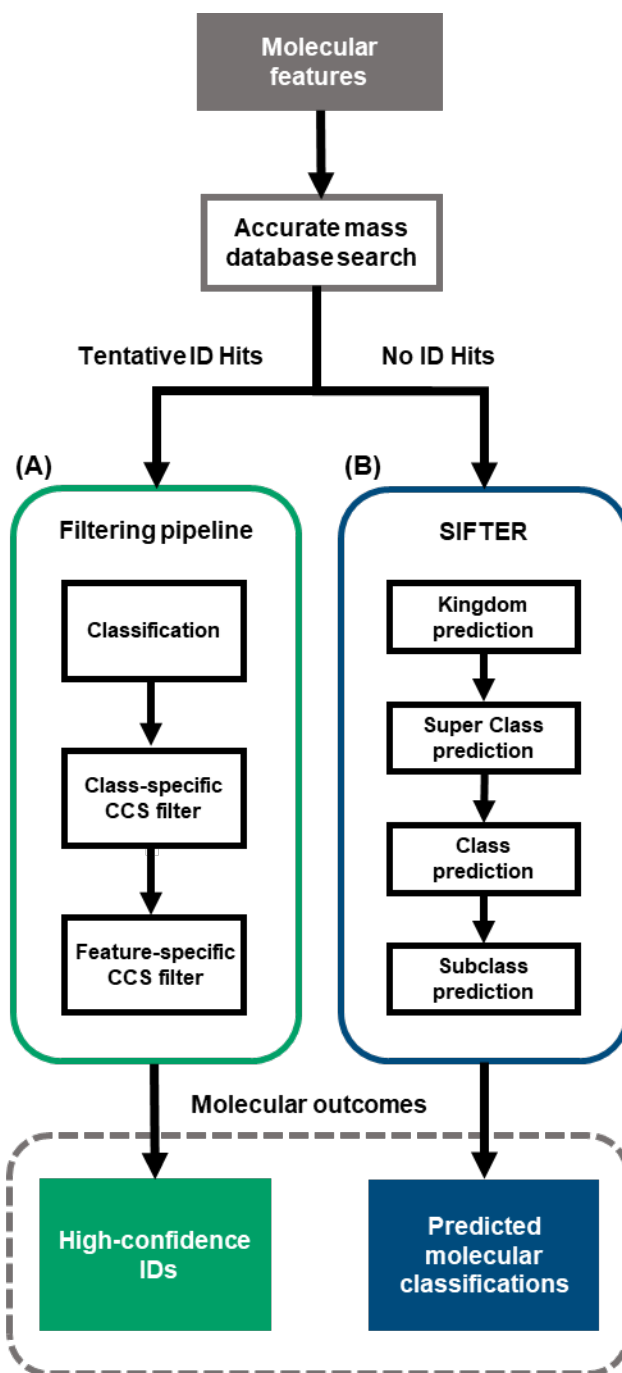
Figure 3.2 Complementary parallel analysis workflows utilizing IM-MS derived measurements, (A) a CCS filtering pipeline to increase the confidence of features with tentatively assigned identifications described in this work, and (B) the SIFTER algorithm to predict the molecular classification of lipidomic features not identified by accurate mass database searching.

subclass was assigned the highest confidence level in 19 of 26 features (73%). Upon more detailed analysis, two incorrect assignments were due to erroneous ion mobility feature selection in the automated peak picking, and one additional incorrect assignment resulted from an absence of the correct candidate subclass in the tentative annotation list. These preliminary tests provided insight into the reliability of the workflow with which to proceed to its application to biological samples. Though these results are promising, the performance is expected to improve as data additions to the Compendium improve the accuracy of the regression models.

3.3.3   CCS Filtering of Lipidomic Data

An untargeted lipidomics experiment was performed on murine brain tissue lipophilic extracts using the CCS filtering pipeline. An illustration of the number of features reduced in each step of the pipeline (Figure 3.2(A)) is contained in Figure 3.3, with features that pass through each level of the filter being assigned higher confidence (Figure 3.3(B)).

Of the 1657 molecular features extracted from the raw data, 1083 (65%) were tentatively identified using accurate mass database searching. During mass database matching, the average number of candidate identifications per compound was 61 and, on average, 75% of these candidates were database entries with the same chemical formula, i.e., mass isomers. The candidate identifications were next assigned a hierarchical structural classification using the ClassyFire algorithm.[51] Of these classified compounds, 39 (~4%) did not have a representative regression model in the training data, and thus could not be assigned any higher confidence other than the initial tentative identification (Figure 3.3(C)).

Following classification, the features were subjected to two stages of CCS-based filtering using both the CCS values and the regression models generated from entries in the Unified CCS
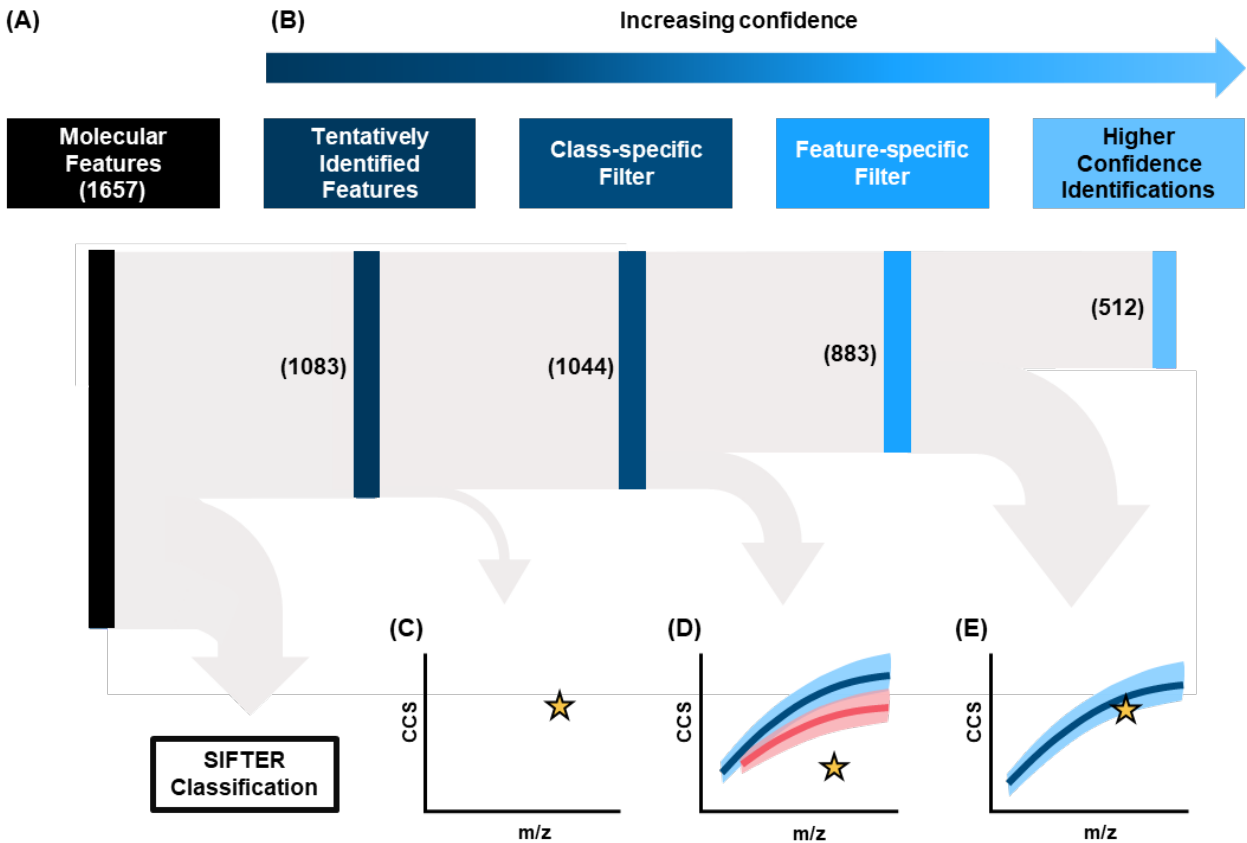
Figure 3.3 Feature reduction workflow for assigning molecular identifications with increasing levels of confidence using the ion mobility-derived CCS. (A) Of 1657 features, 1083 were tentatively identified and thus subjected to CCS filtering. The remaining 574 with no identifications were submitted to the SIFTER algorithm for classification prediction. (B) Results of the CCS filtering pipeline in which increased confidence is assigned from each level of the filter. (C) Out of 1083 tentatively identified features, 39 did not have an ID with a representative regression model. (D) 883 features passed the class-specific filter, while 161 features fell outside the predictive interval of their class model(s). Finally, (E) 512 features were passed through the feature-specific filter. The 371 remaining features had IDs from only one class and did not need the additional filtering level.

Compendium. The first filter level is a class-specific heuristic filter that assigns a higher confidence to an annotation if the CCS of that feature falls within the 99% predictive interval of the class to which the annotation belongs. In addition to assigning increased confidence, this step also serves as a quality assurance step, ensuring that the annotations that pass to the next stage are plausible candidates for the molecular feature. Here, 883 molecular features had annotations that passed this filter, whereas 161 features had CCS values that fell outside of the predictive intervals of all their candidate identifications, meaning these tentative identifications could not be further validated with their CCS information (Figure 3.3(D)).

In the second stage of the filtering pipeline, the proximity of the feature CCS to the mean of the regression models is evaluated. Higher confidence is assigned to an annotation if the mean of the corresponding class regression model falls within a shorter distance to the CCS of the feature than those of other potential classes. Over 500 features had annotations that could be distinguished in this way, while the remaining 371 had only one candidate class or subclass, and therefore did not require further filtering (Figure 3.3(E))).

Using this CCS filtering approach, 82% of the features tentatively identified via accurate mass information (883 out of 1083) were able to be assigned some increased confidence using their CCS values. The average number of potential identification candidates for these high-confidence IDs was decreased from 65 to 31, which represents a reduction of over 50% of the possible compound identities that can be assigned to these features.

In addition to decreasing the overall average candidates per feature, the CCS filtering method also shifted the distribution of candidates significantly (Figure 3.4(A)). Prior to filtering the annotations on the basis of their CCS values, only 34% of the features had 10 or fewer candidates. Of the 883
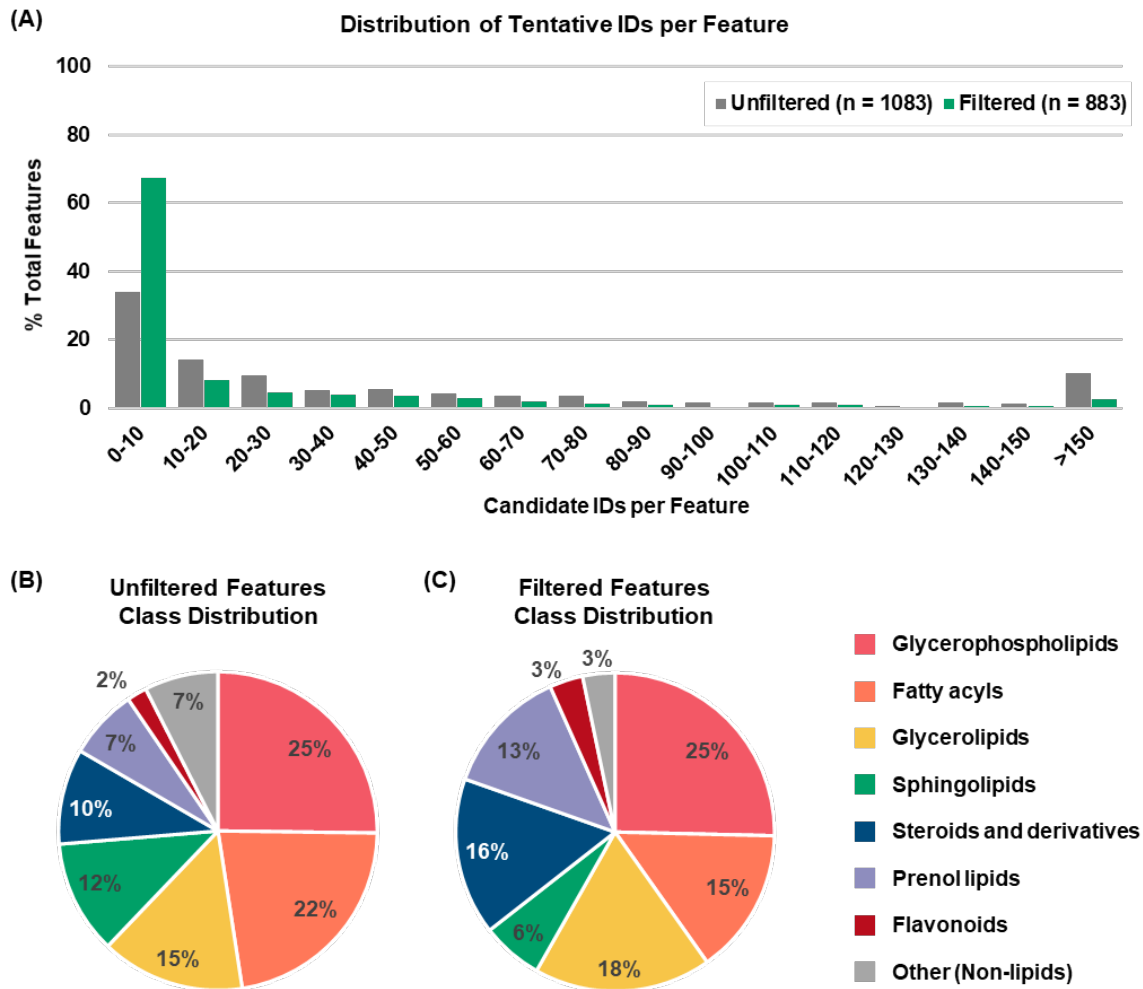
Figure 3.4 Comparison of feature annotations before and after applying the filtering pipeline (A) The list of possible candidate identifications were decreased by filtering on the basis of their CCS values. (B) Lipid class distribution of the unfiltered features with identifications assigned by accurate mass only, as compared to (C) class distribution of features with identifications assigned by the combination of accurate mass and CCS filtering.

features whose candidate identifications had been subjected to filtering, 67% had 10 or fewer remaining candidates. This effective narrowing down of feature identification candidate lists increases the confidence in the remaining options.

For these filtered lists, all the remaining candidate identifications are compounds within the same chemical class, providing high confidence in the class assignments, even in cases in which persisting isomeric/isobaric ambiguity increases the possible number of candidates. This, in turn, lends higher confidence to the resulting lipid class profile in these analyses. Figure 3.4(B) illustrates differences in lipid class distribution among "unfiltered" features (where the "top hit" of the accurate mass database search was used as the identification) and "filtered" features (where the CCS filtering pipeline was used to assign the feature identity). In addition to the shift in distribution of lipid classes between the two groups of features, the share of features with assigned identifications belonging to non-lipid classes is reduced in the filtered group. Out of the 883 initial identifications made using the top candidate, only 441 of these identifications were consistent once the CCS filtering steps were applied (50%), reinforcing the notion that identifications based on mass measurement alone are insufficient for accurate lipid structure assignments.

3.3.4   SIFTER Classification of Unknown Features

The molecular features which were not assigned any tentative identifications or "hits" based on mass database searching were submitted to the SIFTER algorithm.[37] Figure 3.5(B) details the hierarchical class distribution of the 192 features whose classes were successfully predicted by SIFTER. As can be expected from a lipophilic extract, almost half of the predictions are lipids and lipid-like molecules (48%, 93/192). The fact that these compounds were not identified from the initial mass database searching suggests that these databases are incomplete in terms of lipidomic coverage, which might be accepted given that the majority of unidentified compounds derived

Figure 3.5 Molecular classifications successfully predicted with SIFTER machine learning algorithm for experimental features with no tentative identifications assigned by database search.

from biological sources are predicted to be lipids.[53] Notably, a little over half of the predictions from SIFTER were not classified as lipids and lipid-like molecules, but were organic acids (e.g., amino acids and peptides), organic oxygen compounds (e.g., carbohydrates), and organoheterocyclic compounds (e.g., nucleic acids). Since only lipid-specific databases were used for the accurate mass database searching, it is not surprising that most of the unidentified features are predicted to be molecules other than lipids. Further, amino acids and carbohydrates, which make up a large majority of the predicted non-lipids, are common lipid headgroup components whose presence may be partially explained by degradation at the sample level or in-source fragmentation during analysis.

The results from the CCS filtering and the SIFTER classification taken together address two limitations of identification assignment related to database coverage. Searching a large database or multiple databases increases search space and likelihood of a "hit", but also increases the likelihood of false positives and can lower the confidence when the candidate lists are long. Conversely, searching a small, focused database space will result in fewer overall candidate identifications, lowering class coverage. Using additional analytical information, such as CCS provided by IM, mitigates both limitations, and the dual approach demonstrated in this manuscript illustrates a strategy to maximize the chemical information which can be derived from ion mobility measurements

Despite providing increased confidence in molecular annotations, the parallel informatics workflow described in this manuscript comes with challenges. In both the CCS filtering pipeline and the SIFTER prediction, performance is limited by the training set and their resulting regression models. Additionally, the confidence in the assignment decreases with the classification hierarchical structure, *i.e.* there is more confidence in assigning a molecular class than a subclass

because of the limited CCS space covered by the empirically-trained regression models. In other words, the more specific the classification, the more overlap in its model with other similar models. Similarly, classes that are not well-populated in the training set will result in less confidence in annotations of that class. These limitations illustrate the importance of a reliable and representative training set and will diminish as the coverage of empirical databases, such as the Unified CCS Compendium, increases.

3.4   Conclusions

An integrated parallel workflow using IM-derived CCS for high-confidence untargeted lipidomics was demonstrated. This dual workflow combines a previously developed classification prediction algorithm with a novel CCS filtering pipeline. The utility of the workflow was shown using untargeted LC-IM-MS data from a murine brain lipid extract. Filtering candidate compound identifications on the basis of their measured CCS values increased the confidence of the annotations of 883 compounds by narrowing their candidate lists by over 50% on average. Using the SIFTER machine learning algorithm to predict the classification of unidentified features provided insight into the likely classes of 192 compounds which otherwise were not assigned a tentative identification based on accurate mass searching. Both pipelines will continue to yield higher confidence results as the collaborative Unified CCS Compendium training set is expanded. Further integration of these approaches with other analytical techniques, for example tandem MS/MS ion fragmentation or LC retention time searching or prediction, will promote high confidence and increased coverage to improve interpretation and understanding of the complexity of the lipidome.

3.5   Acknowledgements

3.6   References

(1)   Cullis, P. R.; Fenske, D. B.; Hope, M. J. Chapter 1 Physical Properties and Functional Roles of Lipids in Membranes. *New Comprehensive Biochemistry* **1996**, *31* (C), 1–33. https://doi.org/10.1016/S0167-7306(08)60508-6.

(2)   Wymann, M. P.; Schneiter, R. Lipid Signalling in Disease. *Nature Reviews Molecular Cell Biology* **2008**, *9* (2), 162–176. https://doi.org/10.1038/nrm2335.

(3)   Wenk, M. R. The Emerging Field of Lipidomics. *Nature Reviews Drug Discovery* **2005**, *4* (7), 594–610. https://doi.org/10.1038/nrd1776.

(4)   Xu, T.; Hu, C.; Xuan, Q.; Xu, G. Recent Advances in Analytical Strategies for Mass Spectrometry-Based Lipidomics. *Analytica Chimica Acta* **2020**, *1137*, 156–169. https://doi.org/10.1016/j.aca.2020.09.060.

(5)   Chatgilialoglu, C.; Ferreri, C.; Melchiorre, M.; Sansone, A.; Torreggiani, A. Lipid Geometrical Isomerism: From Chemistry to Biology and Diagnostics. *Chemical Reviews* **2014**, *114* (1), 255–284. https://doi.org/10.1021/cr4002287.

(6)   Emília, A.; Rolim, H.; Henrique-araújo, R.; Gomes, E.; Karoline, F.; Alves, D. A.; Gonzaga, L. Lipidomics in the Study of Lipid Metabolism: Current Perspectives in the Omic Sciences. *Gene* **2015**, *554* (2), 131–139. https://doi.org/10.1016/j.gene.2014.10.039.

(7)   Navas-Iglesias, N.; Carrasco-Pancorbo, A.; Cuadros-Rodríguez, L. From Lipids Analysis towards Lipidomics, a New Challenge for the Analytical Chemistry of the 21st Century. Part II: Analytical Lipidomics. *TrAC Trends in Analytical Chemistry* **2009**, *28* (4), 393–403. https://doi.org/10.1016/j.trac.2008.12.004.

(8)   Rustam, Y. H.; Reid, G. E. Analytical Challenges and Recent Advances in Mass Spectrometry Based Lipidomics. *Analytical Chemistry* **2018**, *90* (1), 374–397. https://doi.org/10.1021/acs.analchem.7b04836.

(9)   Blaženović, I.; Kind, T.; Ji, J.; Fiehn, O. Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics. *Metabolites* **2018**, *8* (2), 31. https://doi.org/10.3390/metabo8020031.

(10)   Schrimpe-Rutledge, A. C.; Codreanu, S. G.; Sherrod, S. D.; McLean, J. A. Untargeted Metabolomics Strategies—Challenges and Emerging Directions. *Journal of the American Society for Mass Spectrometry* **2016**, *27* (12), 1897–1905. https://doi.org/10.1007/s13361-016-1469-y.

(11)   Sindelar, M.; Patti, G. J. Chemical Discovery in the Era of Metabolomics. *Journal of the American Chemical Society* **2020**, jacs.9b13198. https://doi.org/10.1021/jacs.9b13198.

(12)   Sud, M.; Fahy, E.; Cotter, D.; Brown, A.; Dennis, E. A.; Glass, C. K.; Merrill, A. H.; Murphy, R. C.; Raetz, C. R. H.; Russell, D. W.; Subramaniam, S. LMSD: LIPID MAPS Structure Database. *Nucleic Acids Research* **2007**, *35* (Database), D527–D532. https://doi.org/10.1093/nar/gkl838.

(13)   Kind, T.; Liu, K.-H.; Lee, D. Y.; DeFelice, B.; Meissen, J. K.; Fiehn, O. LipidBlast in Silico Tandem Mass Spectrometry Database for Lipid Identification. *Nature Methods* **2013**, *10* (8), 755–758. https://doi.org/10.1038/nmeth.2551.

(14)   Koelmel, J. P.; Kroeger, N. M.; Ulmer, C. Z.; Bowden, J. A.; Patterson, R. E.; Cochran, J. A.; Beecher, C. W. W.; Garrett, T. J.; Yost, R. A. LipidMatch: An Automated Workflow for Rule-Based Lipid Identification Using Untargeted High-Resolution Tandem Mass Spectrometry Data. *BMC Bioinformatics* **2017**, *18* (1), 331. https://doi.org/10.1186/s12859-017-1744-3.

(15)   Koelmel, J. P.; Li, X.; Stow, S. M.; Sartain, M. J.; Murali, A.; Kemperman, R.; Tsugawa, H.; Takahashi, M.; Vasiliou, V.; Bowden, J. A.; Yost, R. A.; Garrett, T. J.; Kitagawa, N. Lipid Annotator: Towards Accurate Annotation in Non-Targeted Liquid Chromatography High-Resolution Tandem Mass Spectrometry (LC-HRMS/MS) Lipidomics Using a Rapid and User-Friendly Software. *Metabolites* **2020**, *10* (3), 101. https://doi.org/10.3390/metabo10030101.

(16)   Kochen, M. A.; Chambers, M. C.; Holman, J. D.; Nesvizhskii, A. I.; Weintraub, S. T.; Belisle, J. T.; Islam, M. N.; Griss, J.; Tabb, D. L. Greazy: Open-Source Software for Automated Phospholipid Tandem Mass Spectrometry Identification. *Analytical Chemistry* **2016**, *88* (11), 5733–5741. https://doi.org/10.1021/acs.analchem.6b00021.

(17)   Peterson, B. L.; Cummings, B. S. A Review of Chromatographic Methods for the Assessment of Phospholipids in Biological Samples. *Biomedical Chromatography* **2006**, *20* (3), 227–243. https://doi.org/10.1002/bmc.563.

(18)   Cajka, T.; Fiehn, O. Comprehensive Analysis of Lipids in Biological Systems by Liquid Chromatography-Mass Spectrometry. *TrAC Trends in Analytical Chemistry* **2014**, *61*, 192–206. https://doi.org/10.1016/j.trac.2014.04.017.

(19)   Blaženović, I.; Shen, T.; Mehta, S. S.; Kind, T.; Ji, J.; Piparo, M.; Cacciola, F.; Mondello, L.; Fiehn, O. Increasing Compound Identification Rates in Untargeted Lipidomics Research

with Liquid Chromatography Drift Time-Ion Mobility Mass Spectrometry. *Analytical Chemistry* **2018**, acs.analchem.8b01527. https://doi.org/10.1021/acs.analchem.8b01527.

(20)   Ross, D. H.; Cho, J. H.; Zhang, R.; Hines, K. M.; Xu, L. LiPydomics: A Python Package for Comprehensive Prediction of Lipid Collision Cross Sections and Retention Times and Analysis of Ion Mobility-Mass Spectrometry-Based Lipidomics Data. *Analytical Chemistry* **2020**, No. Ml. https://doi.org/10.1021/acs.analchem.0c02560.

(21)   Beyaz, A.; Fan, W.; Carr, P. W.; Schellinger, A. P. Instrument Parameters Controlling Retention Precision in Gradient Elution Reversed-Phase Liquid Chromatography. *Journal of Chromatography A* **2014**, *1371*, 90–105. https://doi.org/10.1016/j.chroma.2014.09.085.

(22)   Taylor, P. J. Matrix Effects: The Achilles Heel of Quantitative High-Performance Liquid Chromatography–Electrospray–Tandem Mass Spectrometry. *Clinical Biochemistry* **2005**, *38* (4), 328–334. https://doi.org/10.1016/j.clinbiochem.2004.11.007.

(23)   Harris, R. A.; Leaptrot, K. L.; May, J. C.; McLean, J. A. New Frontiers in Lipidomics Analyses Using Structurally Selective Ion Mobility-Mass Spectrometry. *TrAC Trends in Analytical Chemistry* **2019**, *116*, 316–323. https://doi.org/10.1016/j.trac.2019.03.031.

(24)   Kyle, J. E.; Zhang, X.; Weitz, K. K.; Monroe, M. E.; Ibrahim, Y. M.; Moore, R. J.; Cha, J.; Sun, X.; Lovelace, E. S.; Wagoner, J.; Polyak, S. J.; Metz, T. O.; Dey, S. K.; Smith, R. D.; Burnum-Johnson, K. E.; Baker, E. S. Uncovering Biologically Significant Lipid Isomers with Liquid Chromatography, Ion Mobility Spectrometry and Mass Spectrometry. *The Analyst* **2016**, *141* (5), 1649–1659. https://doi.org/10.1039/C5AN02062J.

(25)   Leaptrot, K. L.; May, J. C.; Dodds, J. N.; McLean, J. A. Ion Mobility Conformational Lipid Atlas for High Confidence Lipidomics. *Nature Communications* **2019**, *10* (1), 985. https://doi.org/10.1038/s41467-019-08897-5.

(26)   Kliman, M.; May, J. C.; McLean, J. A. Lipid Analysis and Lipidomics by Structurally Selective Ion Mobility-Mass Spectrometry. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **2011**, *1811* (11), 935–945. https://doi.org/10.1016/j.bbalip.2011.05.016.

(27)   Stow, S. M.; Causon, T. J.; Zheng, X.; Kurulugama, R. T.; Mairinger, T.; May, J. C.; Rennie, E. E.; Baker, E. S.; Smith, R. D.; McLean, J. A.; Hann, S.; Fjeldsted, J. C. An Interlaboratory Evaluation of Drift Tube Ion Mobility–Mass Spectrometry Collision Cross Section Measurements. *Analytical Chemistry* **2017**, *89* (17), 9048–9055. https://doi.org/10.1021/acs.analchem.7b01729.

(28)   May, J. C.; McLean, J. A. Ion Mobility-Mass Spectrometry: Time-Dispersive Instrumentation. *Analytical Chemistry*. 2015, pp 1422–1436. https://doi.org/10.1021/ac504720m.

(29)   Paglia, G.; Angel, P.; Williams, J. P.; Richardson, K.; Olivos, H. J.; Thompson, J. W.; Menikarachchi, L.; Lai, S.; Walsh, C.; Moseley, A.; Plumb, R. S.; Grant, D. F.; Palsson, B. O.; Langridge, J.; Geromanos, S.; Astarita, G. Ion Mobility-Derived Collision Cross Section

as an Additional Measure for Lipid Fingerprinting and Identification. *Analytical Chemistry* **2015**, *87* (2), 1137–1144. https://doi.org/10.1021/ac503715v.

(30)   Nichols, C. M.; Dodds, J. N.; Rose, B. S.; Picache, J. A.; Morris, C. B.; Codreanu, S. G.; May, J. C.; Sherrod, S. D.; McLean, J. A. Untargeted Molecular Discovery in Primary Metabolism: Collision Cross Section as a Molecular Descriptor in Ion Mobility-Mass Spectrometry. *Analytical Chemistry* **2018**, *90* (24), 14484–14492. https://doi.org/10.1021/acs.analchem.8b04322.

(31)   Picache, J. A.; Rose, B. S.; Balinski, A.; Leaptrot, K. L.; Sherrod, S. D.; May, J. C.; McLean, J. A. Collision Cross Section Compendium to Annotate and Predict Multi-Omic Compound Identities. *Chemical Science* **2019**, *10* (4), 983–993. https://doi.org/10.1039/C8SC04396E.

(32)   Zheng, X.; Aly, N. A.; Zhou, Y.; Dupuis, K. T.; Bilbao, A.; Paurus, V. L.; Orton, D. J.; Wilson, R.; Payne, S. H.; Smith, R. D.; Baker, E. S. A Structural Examination and Collision Cross Section Database for over 500 Metabolites and Xenobiotics Using Drift Tube Ion Mobility Spectrometry. *Chem. Sci.* **2017**, *8*, 7724–7736. https://doi.org/10.1039/C7SC03464D.

(33)   Poland, J. C.; Leaptrot, K. L.; Sherrod, S. D.; Flynn, C. R.; McLean, J. A. Collision Cross Section Conformational Analyses of Bile Acids via Ion Mobility–Mass Spectrometry. *Journal of the American Society for Mass Spectrometry* **2020**, *31* (8), 1625–1631. https://doi.org/10.1021/jasms.0c00015.

(34)   Zhou, Z.; Luo, M.; Chen, X.; Yin, Y.; Xiong, X.; Wang, R.; Zhu, Z. Ion Mobility Collision Cross-Section Atlas for Known and Unknown Metabolite Annotation in Untargeted Metabolomics. *Nature Communications* **2020**, *11* (1), 4334. https://doi.org/10.1038/s41467-020-18171-8.

(35)   Zhou, Z.; Tu, J.; Xiong, X.; Shen, X.; Zhu, Z. J. LipidCCS: Prediction of Collision Cross-Section Values for Lipids with High Precision to Support Ion Mobility-Mass Spectrometry-Based Lipidomics. *Analytical Chemistry* **2017**, *89* (17), 9559–9566. https://doi.org/10.1021/acs.analchem.7b02625.

(36)   Plante, P.-L.; Francovic-Fontaine, É.; May, J. C.; McLean, J. A.; Baker, E. S.; Laviolette, F.; Marchand, M.; Corbeil, J. Predicting Ion Mobility Collision Cross-Sections Using a Deep Neural Network: DeepCCS. *Analytical Chemistry* **2019**, *91* (8), 5191–5199. https://doi.org/10.1021/acs.analchem.8b05821.

(37)   Picache, J. A.; May, J. C.; McLean, J. A. Chemical Class Prediction of Unknown Biomolecules Using Ion Mobility-Mass Spectrometry and Machine Learning: Supervised Inference of Feature Taxonomy from Ensemble Randomization. *Analytical Chemistry* **2020**, *92* (15), 10759–10767. https://doi.org/10.1021/acs.analchem.0c02137.

(38)   Soper-Hopper, M. T.; Vandegrift, J.; Baker, E. S.; Fernández, F. M. Metabolite Collision Cross Section Prediction without Energy-Minimized Structures. *The Analyst* **2020**, *145* (16), 5414–5418. https://doi.org/10.1039/D0AN00198H.

(39)   Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environmental Science & Technology* **2014**, *48* (4), 2097–2098. https://doi.org/10.1021/es5002105.

(40)   Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W.-M.; Fiehn, O.; Goodacre, R.; Griffin, J. L.; Hankemeier, T.; Hardy, N.; Harnly, J. Proposed Minimum Reporting Standards for Chemical Analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Inititative (MSI). *Metabolomics* **2007**, *3* (3), 211–221. https://doi.org/10.1007/s11306-007-0082-2.Proposed.

(41)   May, J. C.; Goodwin, C. R.; Lareau, N. M.; Leaptrot, K. L.; Morris, C. B.; Kurulugama, R. T.; Mordehai, A.; Klein, C.; Barry, W.; Darland, E.; Overney, G.; Imatani, K.; Stafford, G. C.; Fjeldsted, J. C.; McLean, J. A. Conformational Ordering of Biomolecules in the Gas Phase: Nitrogen Collision Cross Sections Measured on a Prototype High Resolution Drift Tube Ion Mobility-Mass Spectrometer. *Analytical Chemistry* **2014**, *86* (4), 2107–2116. https://doi.org/10.1021/ac4038448.

(42)   Kurulugama, R. T.; Darland, E.; Kuhlmann, F.; Stafford, G.; Fjeldsted, J. Evaluation of Drift Gas Selection in Complex Sample Analyses Using a High Performance Drift Tube Ion Mobility-QTOF Mass Spectrometer. *Analyst* **2015**, *140* (20), 6834–6844. https://doi.org/10.1039/c5an00991j.

(43)   Pfalzer, A. C.; Wilcox, J. M.; Codreanu, S. G.; Totten, M.; Bichell, T. J. V.; Halbesma, T.; Umashanker, P.; Yang, K. L.; Parmalee, N. L.; Sherrod, S. D.; Erikson, K. M.; Harrison, F. E.; McLean, J. A.; Aschner, M.; Bowman, A. B. Huntington's Disease Genotype Suppresses Global Manganese-Responsive Processes in Pre-Manifest and Manifest YAC128 Mice. *Metallomics* **2020**, *12* (7), 1118–1130. https://doi.org/10.1039/d0mt00081g.

(44)   Bilbao, A.; Gibbons, B. C.; Stow, S. M.; Kyle, J. E.; Bloodsworth, K. J.; Payne, S. H.; Smith, R. D.; Ibrahim, Y. M.; Baker, E. S.; Fjeldsted, J. C. A Preprocessing Tool for Enhanced Ion Mobility–Mass Spectrometry-Based Omics Workflows. *Journal of Proteome Research* **2021**. https://doi.org/10.1021/acs.jproteome.1c00425.

(45)   Rose, B. S.; Leaptrot, K. L.; Harris, R. A.; Sherrod, S. D.; May, J. C.; McLean, J. A. High Confidence Shotgun Lipidomics Using Structurally Selective Ion Mobility-Mass Spectrometry. In *Mass Spectrometry-Based Lipidomics: Methods and Protocols*; Hsu, F.-F., Ed.; Springer US: New York, NY, 2021; pp 11–37. https://doi.org/10.1007/978-1-0716-1410-5_2.

(46)   Smith, C. a; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G. METLIN: A Metabolite Mass Spectral Database. *Therapeutic drug monitoring* **2005**, *27* (6), 747–751.

(47)   Wishart, D. S.; Feunang, Y. D.; Marcu, A.; Guo, A. C.; Liang, K.; Vázquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C.; Karu, N.; Sayeeda, Z.; Lo, E.; Assempour, N.; Berjanskii, M.; Singhal, S.; Arndt, D.; Liang, Y.; Badran, H.; Grant, J.; Serra-Cayuela, A.; Liu, Y.; Mandal, R.; Neveu, V.; Pon, A.; Knox, C.; Wilson, M.; Manach, C.; Scalbert, A. HMDB

4.0: The Human Metabolome Database for 2018. *Nucleic Acids Research* **2018**, *46* (D1), D608–D617. https://doi.org/10.1093/nar/gkx1089.

(48) Feldman, H. J.; Dumontier, M.; Ling, S.; Haider, N.; Hogue, C. W. V. CO: A Chemical Ontology for Identification of Functional Groups and Semantic Comparison of Small Molecules. *FEBS Letters* **2005**, *579* (21), 4685–4691. https://doi.org/10.1016/j.febslet.2005.07.039.

(49) Wohlgemuth, G.; Haldiya, P. K.; Willighagen, E.; Kind, T.; Fiehn, O. The Chemical Translation Service-a Web-Based Tool to Improve Standardization of Metabolomic Reports. *Bioinformatics* **2010**, *26* (20), 2647–2648. https://doi.org/10.1093/bioinformatics/btq476.

(50) Szöcs, E.; Muench, D.; Ranke, J. Webchem: Retrieve Chemical Information from the Web. *Zenodo* **2015**. https://doi.org/10.5281/zenodo.33823.

(51) Djoumbou Feunang, Y.; Eisner, R.; Knox, C.; Chepelev, L.; Hastings, J.; Owen, G.; Fahy, E.; Steinbeck, C.; Subramanian, S.; Bolton, E.; Greiner, R.; Wishart, D. S. ClassyFire: Automated Chemical Classification with a Comprehensive, Computable Taxonomy. *Journal of Cheminformatics* **2016**, *8* (1), 1–20. https://doi.org/10.1186/s13321-016-0174-y.

(52) McLean Research Group GitHub https://github.com/McLeanResearchGroup.

(53) May, J. C.; McLean, J. A. Advanced Multidimensional Separations in Mass Spectrometry: Navigating the Big Data Deluge. *Annual Review of Analytical Chemistry* **2016**, *9* (1), 387–409. https://doi.org/10.1146/annurev-anchem-071015-041734.

CHAPTER 4

CCS CALIBRATION STRATEGIES AND SEPARATION PERFORMANCE OF A

PROTOTYPE HIGH-RESOLUTION ION MOBILITY PLATFORM [§]

4.1 Introduction

Ion mobility (IM) has emerged as a robust separation strategy for complex chemical analyses

largely due to its ability to be interfaced with mass spectrometry (MS) to improve peak capacity

and aid in the separation of isobaric signals. Whereas conventional ion mobility does not perform

at the same level of selectivity and resolution as liquid and gas chromatography (LC and GC),[1] IM

separations are several orders of magnitude faster than LC and GC (100s of ms vs. minutes), and

can be integrated online with chromatography techniques or MS imaging to further increase peak

capacity.[2,3] Additionally, IM provides an additional molecular measurement, namely the gas-phase

collision cross section (CCS), which can be used to gain further structural insight and support

compound identifications.[4–6] Whereas many of the technical hurdles associated with integrating

IM with MS have largely been addressed, one of the contemporary challenges of IM has been the

limited resolution offered by the technique. For example, MS routinely operates with resolving

power (R_p) values on the order of tens of thousands (time-of-flight MS) to hundreds of thousands

(Fourier Transform MS, FTMS), yet the resolving power of current commercially-available time-

dispersive IM techniques generally benchmark below 100.[2,7–9] Recent and notable exceptions do

exist, namely trapped ion mobility spectrometry (TIMS) which demonstrates R_p values as high as 400,[10,11] and a recently-introduced, cyclic, multi-pass instrument based on traveling wave ion mobility spectrometry (TWIMS) which is capable of very high R_p values, around 750 for 100 transits (~1 meter/transit) around the drift ring.[12] These IM techniques typically require long scan speeds (e.g., ~1.5 seconds for 100 passes in cyclic TWIMS),[12] and/or target a narrow range of mobilities in order to access high resolutions, which is analogous to the resolution/throughput tradeoffs for FTMS instruments, where resolving power scales with acquisition duration.

In 2014, Smith and coworkers developed a generalized ion optical architecture they termed structures for lossless ion manipulation (SLIM),[13–15] which utilizes 2-dimensional arrays of electrodes patterned upon printed circuit boards (PCBs) that are driven by combinations of dynamic (RF) and static (DC) electrical potentials to allow ion populations to be trapped and accumulated,[16,17] turned at right angles,[18,19] selected through tee-junctions,[11,20] and lifted (elevators and escalators) to different SLIM levels.[21,22] Ion mobility separations in SLIM devices have been demonstrated both with traditional uniform electric fields (drift tube ion mobility spectrometry, DTIMS),[14,18,20,23] and with dynamically-switched DC potentials (TWIMS),[24,25] the latter allowing for ion transfer across long distances without utilizing high electrical potentials.[22,26] Using several right-angle turns, a serpentine path SLIM device operated with traveling waves was demonstrated,[19] and an extended, ~13 meter path length IM spectrometer was subsequently developed,[27,28] which was capable of accessing resolving powers in excess of 300.[21,29] A multi-pass design was later developed, enabling variable path-lengths through multiple transits through the device, e.g., ~1094 meters via 81 passes, yielding averaged resolving powers in CCS-space of ~1860.[30–32] At the time of writing, this represents the highest resolving power published to date. Other SLIM technology developments in ion mobility have included filtering ions based on their

mobilities,[14,32] accumulating/compressing spatially dispersed ion packets (so-called CRIMP),[33] and dual polarity ion confinement and ion-ion reactions.[34–36]

Here, we evaluate the separation capabilities of a pre-production prototype SLIM-based ion mobility spectrometer (SLIM-IM) that utilizes the extended (~13 meter) separation path length design and operates with traveling waves to enable high resolution ion mobility (HRIM) separations prior to mass analysis. We evaluate this platform to provide performance metrics and guidance to the broader scientific community. In contrast to conventional square wave operation of traveling wave IM, we evaluate the effects of both square and sine waveforms on separation performance of this platform. The resolving power of this instrument is benchmarked in CCS-space. Finally, we develop a calibration method for converting SLIM-IM arrival times to CCS and assess the CCS measurement bias of the instrument using a broadly available MS tuning mixture.

4.2   Experimental Methods

4.2.1   *Materials and sample preparation.*

Agilent ESI-L Low Concentration Tune Mix containing fluoroalkyl phosphazene calibrant ions was used as received from the vendor. All solvents were obtained from Fisher Scientific (Optima Grade, Hampton, NH, USA).

4.2.2   *Instrumentation.*

Data were acquired using a prototype serpentine path SLIM ion mobility device (MOBILion Systems, Chadds Ford, PA, USA) integrated with a commercial quadrupole time-of-flight mass spectrometer (6545, Agilent Technologies, Santa Clara, CA, USA). A schematic of this IM-MS platform is shown in Figure 4.1. A liquid chromatography system (1290 Infinity II, Agilent) was used to introduce samples to the IM-MS via flow injection analysis (20 uL injection volume; 0.100
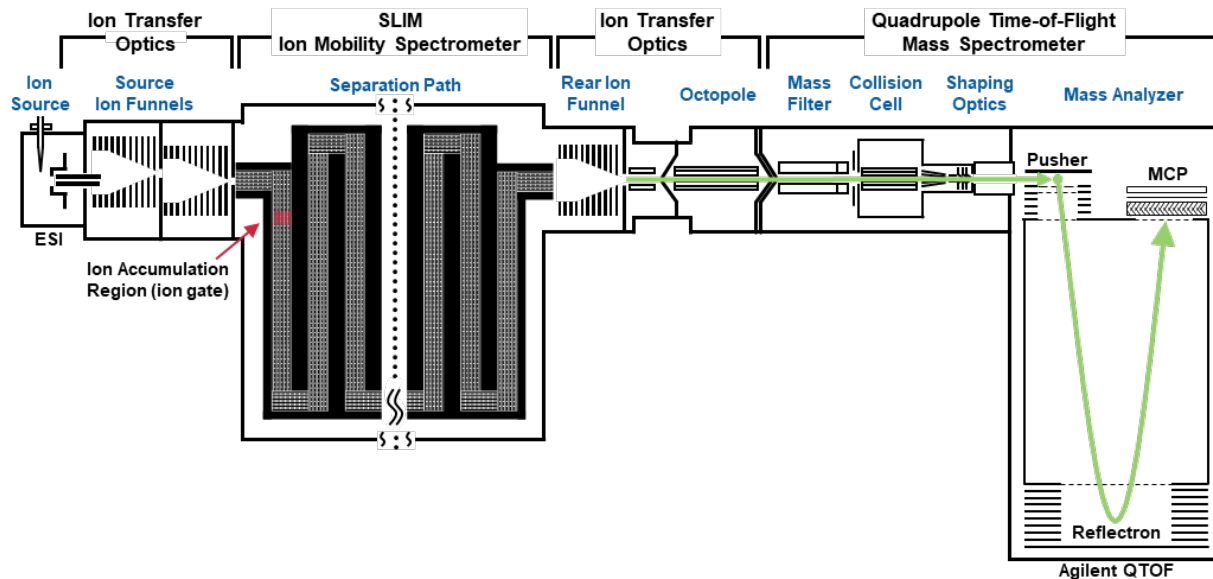
Figure 4.1 Schematic of the prototype SLIM-IM-MS instrument used in this study with significant components annotated.

mL/min flow rate).[37] Samples were ionized via electrospray (Jet Stream, Agilent) operated at 4.0 kV on the entrance capillary and 2.0 kV on the focusing nozzle lens. Ions were transferred to vacuum via a resistive glass capillary and collected by a source ion funnel (200 V_{pp}, 1.1 MHz, 1-10 Torr), where they were focused radially and introduced to the SLIM PCB stack, where a set of SLIM electrodes were operated as an ion accumulation trap. The SLIM-IM separation region utilized ~2.5 Torr of high purity nitrogen gas, which was metered by a gas flow controller (Alicat Scientific, Tucson, AZ) monitored with a capacitance gauge (627F Baratron, MKS Instruments, Andover, MA), which provided a regulation of better than ±0.002 Torr.

4.2.3   SLIM Ion Accumulation and Gating.

Ion gating was achieved in the "Ion Accumulation Region" within the SLIM device (Figure 4.1 callout) which operated a segment of the SLIM as a store-and-release ion trap by applying a repulsive DC potential (±80 V relative to the SLIM DC bias in positive/negative ion mode) for 10 ms (ion accumulation time), and then restoring this potential to the dynamic traveling wave for ion release.[24]   This trapping potential was applied to the last set of dynamic DC electrodes in the segment (corresponding to the last column of dark blue pads in Figure 4.2).

4.2.4   SLIM Separation Path.

The SLIM electrode geometry comprising the IM region is based upon a previous design,[21] which utilizes a 13-meter serpentine ion path length with 44 U-shaped turns (Figure 4.1, SLIM Path). In SLIM, two planar boards with mirrored electrode geometries are stacked (~3 mm apart) to establish the fields necessary for ion manipulation (Figure 4.2(A)). As previously described,[28] the SLIM surface electrode design on each board consists of three electrode types: a pair of outer guard tracks (3 mm width, 15 V) with DC-only potentials; six rows of inner RF-only tracks (~0.5 mm
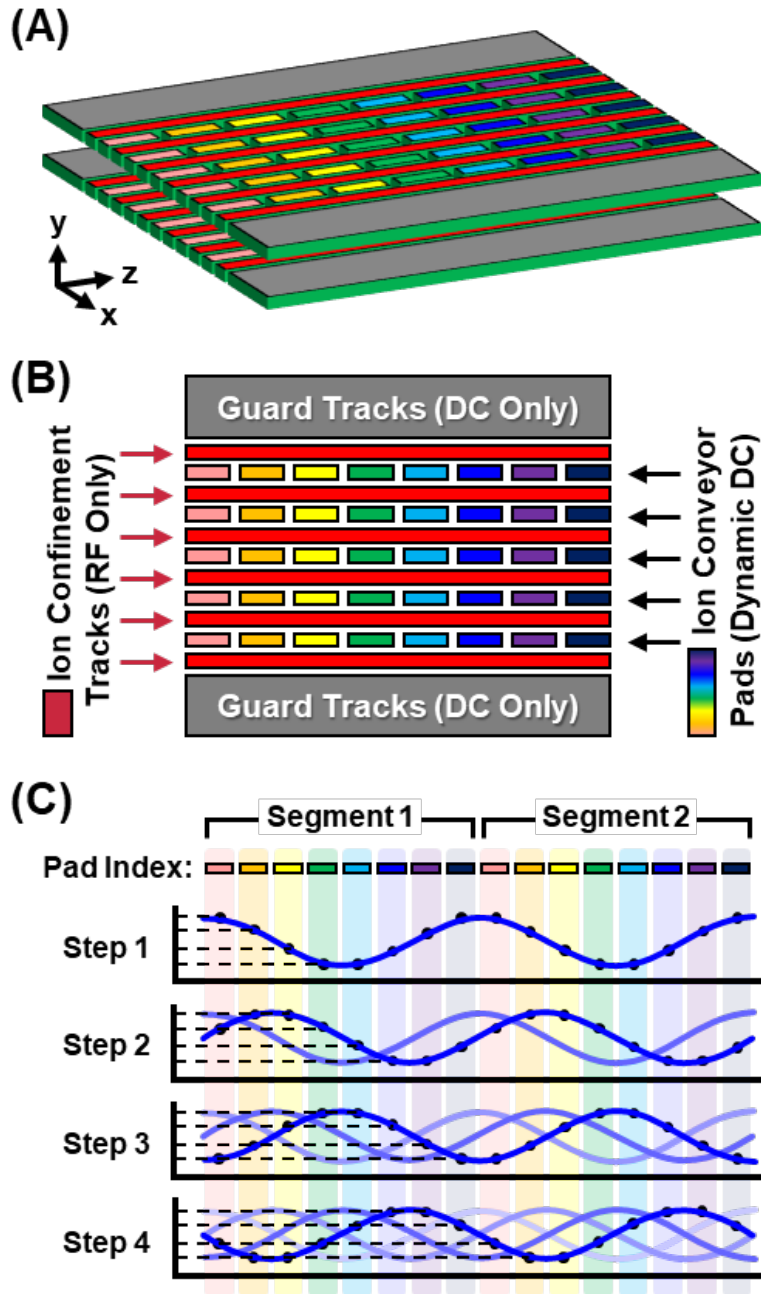
Figure 4.2 SLIM geometry and operational principles. (A) Three-dimensional representation of printed circuit board stacking in the SLIM device for a short section incorporating one complete segment of the TW waveform (8 switching electrodes). (B) SLIM electrode layout. (C) Longitudinal wave propagation, square wave operation, and (D) sine wave operation. Ion conveyor pad electrodes are grouped into sets of eight longitudinal electrodes. At each sequential time point, the wave progresses along the track, propelling ions along the traveling wave at average velocities related to their gas-phase mobilities.

width, 300 V_{pp}, 735 kHz) for ion confinement; and five rows of ion conveyor pads (~0.5 mm width by ~1.0 mm length) which establish the traveling wave potentials used for ion manipulation. In this work, eight conveyor pads in each row (a set) are used to establish one phase of either a square or sinusoidal waveform by applying different potentials to each pad. A digitally generated waveform is applied to each individual segment of electrodes, with adjacent electrodes receiving the same waveform, but shifted by 45 degrees. A total of >1,400 sets (nearly 60,000 pad electrodes) are used across the 13 m serpentine path. With a pad-to-pad distance of 1.125 mm (1.0 mm pad length; 0.125 gap between pads) and the traveling wave stepped across 8 pads (9.0 mm total length) to complete a single phase of the waveform, the wave switching frequencies surveyed in this work (5-25 kHz) correspond to wave speeds between 45 and 225 m/s (Table E1). Both the wave speed and the peak-to-peak wave amplitude (30-40 Vpp) directly influence the ion mobility dispersion observed for traveling wave operation. These parameters are the same as the wave speed and wave height, respectively, commonly used to tune TWIMS instruments.

Following IM separation, ions exit the SLIM boards, are collected by a rear ion funnel (200 V_{pp}, 1.1 MHz, 2.5 Torr), and transferred through an exit quadrupole to the conventional front optics of the Q-TOF for mass analysis. SLIM IM-MS data is acquired using an 8-bit ADC digitizer (U1084A, Keysight Technologies).

4.2.4 Software.

QTOF instrument control and MS data acquisition were accomplished via MassHunter Data Acquisition software (version.09.00). The SLIM-IM module control and data acquisition utilized a custom user interface (GAA Custom Engineering). IM-MS data was viewed and processed using MassHunter IM-MS Browser (version 10). IM traces were integrated across narrow *m/z* ranges and imported into Excel (Microsoft) for further analysis, including peak fitting and peak metric

calculations (arrival time centroid, resolution, resolving power, percent valley, etc.). IM profile

data acquired for the assessment of resolving power across many SLIM-IM conditions were

extracted using IM-MS Browser and further processed and visualized in the R statistical computing

programming environment (R Core Team, Vienna, Austria) using the *tidyverse* suite of tools.[38]

4.2.5 Resolution Calculations.

IM resolution is commonly assessed using "resolving power" which is derived from measurements

of a single peak in the IM dimension.[39] To obtain resolving power values, IM arrival time data

were converted to CCS space.[25] Briefly, the peak centroids (t_p) and peak widths calculated at the

full-width at half the maximum height (Δt_{FWHM}) were obtained by applying normal distribution

fits to the peak of interest within the extracted SLIM-IM data traces. CCS values corresponding to

these peaks were obtained in a separate experiment using a drift tube instrument (6560 IM-QTOF,

Agilent), as previously described.[40,41] The known CCS values of two peaks (CCS_{p1}, CCS_{p2}) within

the same drift spectrum were then used to obtain a CCS difference between two peaks (p1 and p2)

as follows:

$$\Delta CCS_{pp} = |CCS_{p1} - CCS_{p2}| \tag{4.1}$$

A time difference between the two peaks, Δt_{pp}, was similarly calculated from the time centroids of

the peaks, t_{p1} and t_{p2}:

$$\Delta t_{pp} = |t_{p1} - t_{p2}| \tag{4.2}$$

This relationship was then used to convert the FWHM values to CCS-space, which is used to

calculate the CCS-based resolving power, $R_p(CCS/\Delta CCS)$, as follows:

$$R_p(CCS/\Delta CCS) = \frac{CCS}{\Delta t_{FWHM}} \cdot \frac{\Delta t_{pp}}{\Delta CCS_{pp}} \tag{4.3}$$

This CCS-based resolving power is more representative of the separation capabilities for SLIM-IM and can be used to make direct comparisons between the separation capabilities of different ion mobility techniques.[25,27] It is noted here that this arrival time to CCS conversion assumes a linear correspondence, whereas the relationship between TWIMS arrival times and CCS are known to be nonlinear,[42] however the uncertainty is expected to be low due to the fact that the conversion is only applied across the width of a single peak. The average CCS is used when reporting the resolving power from a spectrum containing multiple peaks. Two-peak resolution (R_{pp}) and percent valley (V) calculations were conducted as previously described.[43]

4.2.6. CCS Calibration.

High precision CCS values for the 8 tune mix ions were previously measured from a reference DTIMS instrument operated in nitrogen drift gas ($^{DT}CCS_{N2}$).[40] These reference CCS values (CCS_{ref}) were converted to "reduced CCS" (CCS') by including the ion-neutral reduced mass (μ) and ion charge-state z) dependencies (eq. 4.4) as previously described.[44]

$$CCS' = CCS_{ref} \cdot \sqrt{\frac{\mu}{z}} \tag{4.4}$$

Note here that CCS' is different from the reduced CCS used in the ion transport community to rescale CCS without hard sphere contributions.[45] The tune mix ion raw arrival times were measured at various wave speeds and amplitudes and plots of the reduced CCS versus SLIM-IM arrival time were fitted with nonlinear regression models using the R statistical programming environment. These models were then used as calibration equations to calculate the CCS of the tune mix ions from the SLIM-IM measurements ($^{TW(SLIM)}CCS_{N2}$). As the accuracy of any given CCS measurement is unknown, the percent CCS bias between the calculated and reference DTIMS CCS values was used to assess the performance of the various models:

$$\% \text{ CCS Bias} = \frac{(\text{CCS} - \text{CCS}_{\text{ref}})}{\text{CCS}_{\text{ref}}} \times 100 \tag{4.5}$$

The consensus CCS reference values used here have a reported interlaboratory repeatability of 0.22%,[41] however, expanded uncertainty analysis has estimated that the uncertainty in these drift tube CCS measurements fall within the range of 2.7 to 4.6%,[46] thus limiting the CCS accuracy that can be obtained from this assessment.

4.3   Results and discussion

4.3.1   Accessible Resolving Powers.

Benchmarking experiments were conducted to determine the SLIM-IM conditions in which the highest resolving powers are accessed. For these experiments, two waveforms (square and sine wave), five wave speeds (45, 90, 135, 180, and 225 m/s), and three wave amplitudes (30, 35, and 40 V_{pp}) were evaluated, spanning the optimal transmission and separation range of the instrument. IM profile data was extracted for each of eight tune mix ions that appear prominently in positive ion mode (*m/z* 622, 922, 1222, 1522, 1822, 2122, 2422, 2722). All data was acquired in triplicate, resulting in a total of N=720 R_p(CCS) values for the entire data set. Note that the high RF potentials (300 V_{pp}) applied to the RF electrodes optimized total ion signal, but at a cost of reduced low *m/z* transmission, such that *m/z* 322 did not appear in high abundance and was thus not included in the analysis.

An initial assessment of the two waveforms was conducted which indicated that both waveforms access similar resolving powers, however, the square wave dataset exhibited a narrower optimal range of arrival times which correspond to the highest R_p(CCS) values (Figure E3). Additionally, fewer ions were observed to separate within the square wave data (N=188, or 52% of the total data set) as compared to the sine wave data (N=247, 69%), and recent work from Smith and coworkers

have indicated that square wave operation of SLIM-IM can contribute to more ion heating than sine wave.[47] As such only the sine wave data was evaluated in subsequent experiments. Considering that nearly all traveling wave IM operation to date has utilized a square waveform to approximate a sinusoidal axial potential,[23,48] the broader scale IM separation range observed for operating with a true sine wave in this work implies that the traveling wave technique can be improved using tailored waveforms. The full workup for square wave data can be found in Figure E4. Hereafter only sine wave operation of the SLIM-IM will be discussed.

Results are summarized in Figure 4.3 for optimization of resolving power as a function of the sine wave speed, sine wave amplitude, and analyte mass-to-charge. Only ions which exhibit mobility selective behavior (i.e., not surfing) are considered. Here and elsewhere, the raw arrival times are discussed, and it is important to note that the majority of this measured arrival time represents time spent within the SLIM path (i.e., the true IM drift time), although there are contributions from time spent in other portions of the instrument, including the ion transfer optics to the TOF stage. The scatter plot in Figure 4.3(A) reveals three distinct ranges of arrival times where different resolving power values are accessed.[49] (I) For fast arrival times (<200 ms), so-called "ion surfing" conditions, ion mobilities are too fast to allow IM-selective "roll-over" events to occur. Here, little to no IM separation occurs under most SLIM-IM conditions (see Figure E2). (II) For intermediate arrival times between ca. 200-700 ms, the highest resolving powers are observed. In this region of ion motion, ions are fully subject to mobility-selective ion drift throughout the SLIM separation path. The corresponding box-and-whisker plot in this range of arrival times show that the majority of R_p(CCS) values are between ca. 230-260 (242 mean value for the 400-600 ms bin), with a few data points exhibiting resolving powers in excess of 300. (III) For the slower arrival times beyond ca. 700 ms, the R_p(CCS) magnitude gradually declines, which is interpreted as resulting from peak
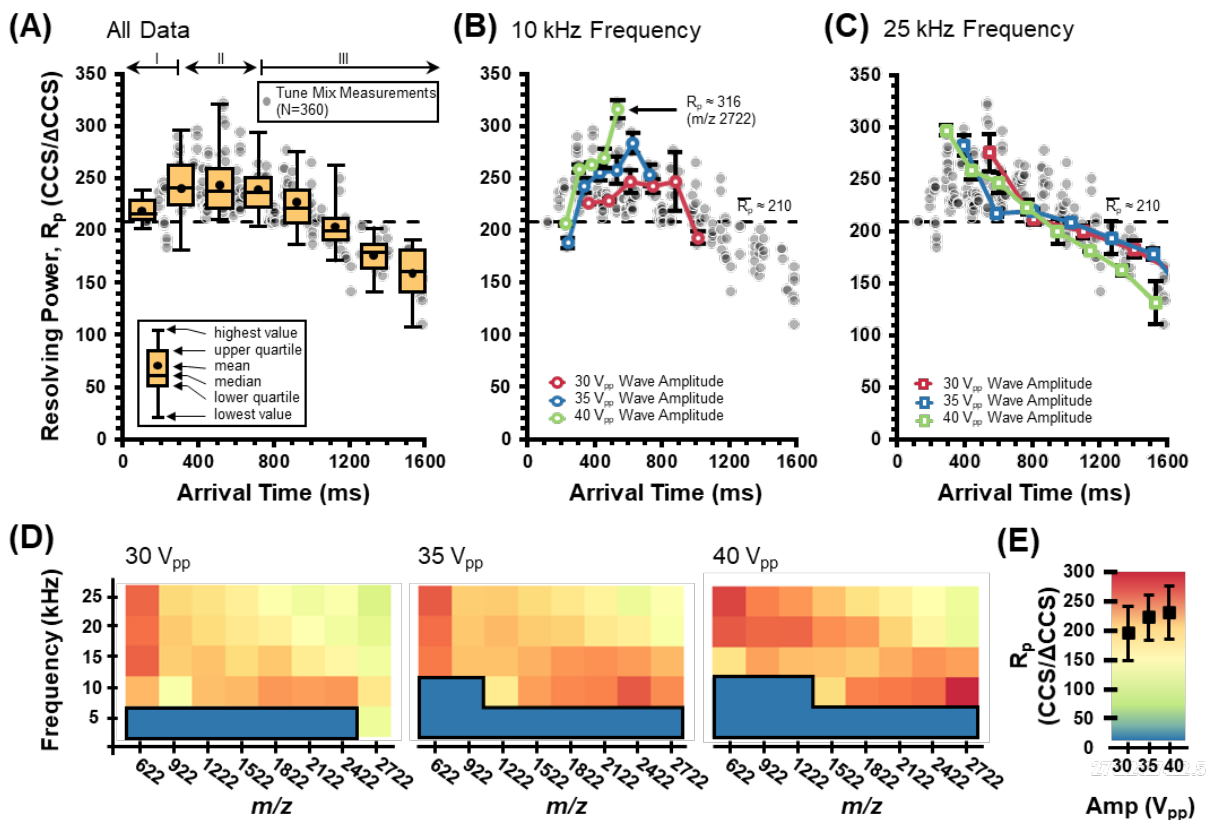
Table 4.3 Visualization of resolving power performance for a sinusoidal traveling wave across the various operational parameters explored in this study. (A) A summary of all CCS-based Rp values (N=247) as a function of arrival time for tune mix components (cy-clophosphazenes) measured across various SLIM-IM parameters, including wave speeds (45, 90, 135, 180, and 225 m/s) and wave amplitudes (30, 35, and 40 Vpp). Box and whisker overlays summarize the data within 200 ms bins. (B) Average Rp(CCS) values (3 replicates per data point) calculated for 90 m/s wave speed at 30, 35, and 40 Vpp wave amplitudes. (C) Average Rp(CCS) values calculated for 225 m/s data. The horizontal dotted line in these scatter plots represents the average Rp across the entire dataset (ca. 210). (D) Heat maps visualizing the resolving powers observed for each tune mix component (x-axis) at each wave speed (y-axis) for wave amplitudes of 30 (left panel), 35 (middle panel), and 40 Vpp (right panel). Each square represents an average of three replicate measurements. Here, the dark blue regions represent conditions in which ions are transmitted, but no IM separation occurs (i.e., ion "surfing" conditions), which is observed at low *m/z*, high wave amplitude, and low wave speeds. (E) Color scale for panel D with average and standard deviation Rp(CCS) values measured for tune mix components at each wave amplitude (x-axis).

broadening due to extended ion-gas diffusion that corresponds to long residence times within the SLIM-IM separation path. The box-and-whisker plots at these long arrival times indicate that the majority of resolving powers continue to decline, falling below 200 for times greater than ~1.2 seconds. Figures 4.3(B) and 4.3(C) contain overlays of the average R_p(CCS) observed at each of the three wave amplitudes for low and high wave speeds (90 and 225 m/s, respectively). Here, the two wave speeds investigated represent conditions where ions either are undergoing transitions from ion surfing to IM-selective drift (Figure 4.3(B), 90 m/s), or where all ions experience continuous ion drift (Figure 4.3(C), 225 m/s). The results are somewhat complicated, but in general, the highest resolving power values are observed for ions near the transition from ion surfing to IM-selective drift (e.g., 90 m/s). At 90 m/s, the highest wave amplitude (40 V_{pp}) generally accesses the highest resolving powers, consistent with previous TWIMS findings,[4] though R_p(CCS) differences no longer appear significant under conditions where all ions have fully transitioned to an IM-selective drifting behavior, e.g., region III (Figure 4.3(C)). Collectively, Figures 4.3(B) and 4.3(C) suggest that operating near the boundary of IM-selective traveling wave behavior can offer a slight increase in R_p(CCS), specifically for operating at the highest wave amplitudes (40 V_{pp}) under the lowest wave speeds which still yield IM separations (90 m/s). In other words, the highest resolving powers are obtained near the onset of ion surfing behavior, and this information can be used to target high resolution for specific analyte systems. Under these conditions, ions are mobility-separated while spending the minimal amount of time within the elevated pressure SLIM separation path which otherwise leads to diffusional broadening of the peaks. Figure 4.3(D) recasts the resolving power data as a function of different *m/z* cyclophosphazene analytes. This projection is useful for illustrating mass-dependent effects on the measured resolving powers. Each square within the heat map represents the R_p(CCS) value (N=3,

averaged) for each tune mix ion measured at each wave speed (y-axis) and wave amplitude (panels) surveyed. The corresponding values for the heat map color scale and overlays of the average $R_p(CCS)$ values observed across each wave amplitude are contained in Figure 4.3(E). Here, the ion-specific data confirms that the highest resolving powers are achieved for all ions at 35 and 40 V_{pp} wave amplitudes, although these projections also reveal that IM separation across the full range of masses surveyed is only achieved under the higher wave speeds, above 135 m/s.

Of note is that the optimal conditions for resolving power are slightly different for higher mobility (lower m/z) species than is observed for lower mobility (higher m/z) ions. When optimizing overall resolving power of the SLIM-IM, we recommend using the higher range of values for both wave speed (e.g., 180-225 m/s) and wave amplitude (e.g., 35-40 V_{pp}) to ensure all ions are undergoing mobility separation. However, to achieve the highest resolving power for a single-species rather than a range of species, the highest wave amplitude (40 V_{pp}) is recommended, and the wave speed can be adjusted to further optimize the separation (e.g., higher speeds for low m/z ions; lower speeds for high m/z). These results suggest that simultaneously scanning wave speed and wave amplitude should allow the highest R_p values to be achieved across a range of m/z, similar to the practice of operating TWIMS instruments using ramped wave heights.

Table 4.1 summarizes the highest R_p values observed and the corresponding SLIM-IM conditions in which they were accessed for each tune mix ion. Overall, 40 V_{pp} wave amplitudes yielded the highest resolving powers. While in some cases the highest R_p value was observed at 35 V_{pp}, for most of those occurrences, the highest and second highest R_p values (Table E2) were within the reproducibility error of one another. Finally, the highest mass tune mix component (m/z 2722) yielded the highest resolving powers, at $R_p(CCS)=316$ for 40 V_{pp}, 90 m/s SLIM-IM conditions. These observations hold for the phosphazenes which are structurally stable compounds, however,

Table 4.1 Highest resolving power values observed

| Tune Mix Ion | Highest R_p Measured (CCS/ΔCCS) [a.] | Corresponding Parameter | | Corresponding Arrival Time (ms) |
		Wave Amplitude (V_{pp})	Wave Speed (m/s)	
m/z 622	297.4±5.0 (3)	40	225	290.5
m/z 922	274.1±16.6 (3)	40	180	437.9
m/z 1222	276.0±22.5 (3)	40	180	462.9
m/z 1522	251.8±13.2 (3)	40	180	596.4
m/z 1822	258.8±3.7 (3)	40	90	304.7
m/z 2122	263.2±1.5 (3)	40	90	381.6
m/z 2422	283.6±9.5 (3)	35	90	622.1
m/z 2722	315.7±8.6 (3)	40	90	535.4
a. Highest R_p is averaged over replicate measurements, denoted in the parenthesis. The time-to-CCS conversion is determined from Equation 3 using the differences between the tune mix ion and the next highest *m/z* ion in the spectrum.				

more fragile ions may not be able to access the same level of resolving power due to different levels of ion heating experienced under the different wave amplitudes and speeds. It is also noted that these results are for singly-charged ions—higher resolving powers are expected when investigating higher charge states.

4.3.2   Collision Cross Section Calibration.

Recent work from Smith and coworkers evaluated the use of different negative mode calibrants for use with a 13-m SLIM-IM (SLIM "SUPER")[47] which has the same geometry as the platform used in this current study, with the exception that the SLIM SUPER incorporates an additional return path to perform multi-pass experiments.[30] In addition to evaluating negative mode, their work assessed both sine and square wave operation under three wave amplitudes (40, 50, and 60 V_{pp}) at a fixed wave speed (200 m/s).

Here, we evaluate positive ion mode, sine wave operation of the HRIM SLIM-IM platform. Because the choice of calibrant strongly affects CCS calibration in TWIMS, pragmatically we chose to only assess tune mix to determine the parameters which yield the lowest CCS errors when other factors are not considered. Figure 4.4(A) contains plots of reduced CCS of the reference values versus SLIM-IM arrival times, fitted with three calibration equations: a power fit, a second order polynomial fit, and a third order polynomial fit. Here, 40 V_{pp} and 180 m/s was chosen for this assessment, as this yielded the overall highest resolving powers where none of the tune mix ions were observed to be surfing (Figure E5). Our results indicate that the CCS errors are lowest when using a 3rd order polynomial fit (R^2=0.9999), with an average CCS bias across all ions of 0.12%. While lower error for a 3rd order fit was also noted in the work from Smith and coworkers,[47] here we observe significant improvement in using a 3rd order polynomial as compared to a conventional power fit (R^2=0.9995; 0.45% bias), summarized in Figure 4.4(B). Using the 3rd

order fit, we also evaluated three different wave speeds (135, 185, and 225 m/s) under which ions are not undergoing surfing behavior, and all three yielded similar CCS error, with the lower wave speed, 135 m/s, exhibiting the lowest absolute CCS bias of 0.07%. Errors associated with the other SLIM-IM wave speeds and amplitudes for the 3rd order polynomial fit are summarized in Figure E6.

For CCS determination, we find that the SLIM-IM conditions which yield the highest resolving powers are also best-suited for CCS calibration, namely 40 V_{pp} and 135 m/s. Regarding the calibration equation itself, we note that all three fits yielded low errors of less than 0.5% bias and emphasize that polynomial fits cannot be extrapolated without high error. Thus, a power fit is recommended when a generalizable calibration is desired, whereas the 3rd order polynomial fit can be used when the lowest CCS errors are needed, and the CCS values fall within the range of calibration. Additional errors are expected when applying these calibrations to other ions, however, this evaluation serves to assess the errors associated with the calibration method itself, as well as the lowest fit errors that can be expected with this approach, namely less than 0.2%. Additional refinements to this calibration method, such as incorporating a time correction to the raw arrival time data or calibrating across a narrower range of ion mobilities, should further improve the error of this approach.

4.4   Conclusions

The accessible resolution and CCS measurement capabilities of a SLIM-based HRIM-MS system was critically evaluated. This pre-production prototype instrument is based directly on a previous IM-MS design and similarly utilizes structures for lossless ion manipulation to enable the transfer and mobility separation of ions across a large distance (~13 meters) for HRIM analyses. The resolving power (CCS/ΔCCS) of the SLIM-IM device was benchmarked to between 230 and 315
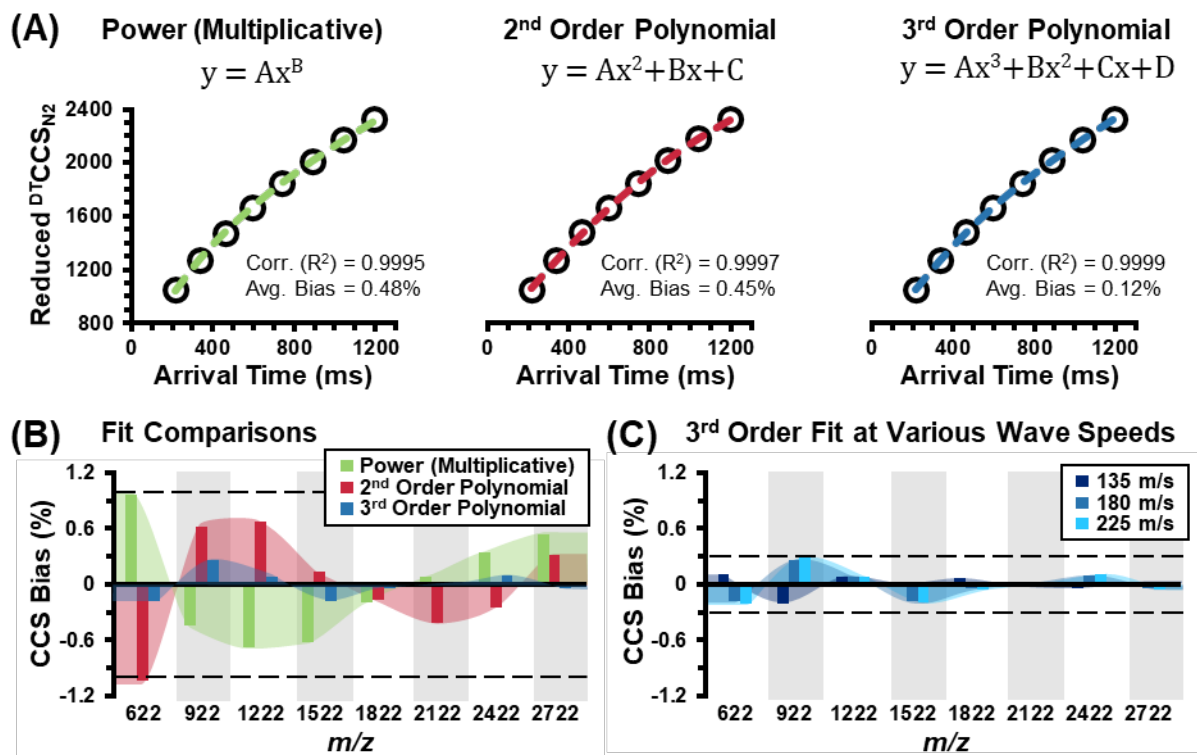
Figure 4.4 Influence of different calibration equations and instrument parameters on calibrating SLIM-IM arrival times of tune mix ions to CCS . (A) The coefficients of determination (R2) and average biases across all ions for three fit equations: a power fit (left panel), a 2nd order polynomial (center panel), and a 3rd order polynomial (right panel). Data was obtained under mobility-selective conditions at 40 Vpp and 180 m/s. (B) Summary of the CCS biases associated with each fit equation. (C) CCS biases observed at 40 Vpp and different wave speeds. Each data point was measured in triplicate.

for a commonly used MS tuning mixture, corresponding to the highest wave amplitudes surveyed in this study (35 and 40 V_{pp}). The optimal resolving powers were observed under conditions where ion arrival times are between 1.5 to 3-times the arrival times associated with surfing-only behavior, which corresponds to ion speeds which are 30-70% greater than the speed of traveling wave itself. Notably, all of the ions from the mixture were transmitted within a short dispersion timeframe (<700 ms) and able to access CCS-based resolving powers in excess of 230, suggesting that this IM-MS platform is well-suited for broadband, untargeted studies.

4.5   Acknowledgements

4.6   References

(1)   May, J. C.; McLean, J. A. Advanced Multidimensional Separations in Mass Spectrometry: Navigating the Big Data Deluge. *Annual Review of Analytical Chemistry* **2016**, *9* (1), 387–409. https://doi.org/10.1146/annurev-anchem-071015-041734.

(2)   May, J. C.; McLean, J. A. Ion Mobility-Mass Spectrometry: Time-Dispersive Instrumentation. *Analytical Chemistry*. 2015, pp 1422–1436. https://doi.org/10.1021/ac504720m.

(3)   Kiss, A.; Heeren, R. M. A. Size, Weight and Position: Ion Mobility Spectrometry and Imaging MS Combined. *Analytical and Bioanalytical Chemistry*. 2011. https://doi.org/10.1007/s00216-010-4644-1.

(4)   Paglia, G.; Williams, J. P.; Menikarachchi, L.; Thompson, J. W.; Tyldesley-Worster, R.; Halldórsson, S.; Rolfsson, O.; Moseley, A.; Grant, D.; Langridge, J.; Palsson, B. O.; Astarita, G. Ion Mobility Derived Collision Cross Sections to Support Metabolomics Applications. *Analytical Chemistry* **2014**, *86* (8), 3985–3993. https://doi.org/10.1021/ac500405x.

(5)   Nichols, C. M.; Dodds, J. N.; Rose, B. S.; Picache, J. A.; Morris, C. B.; Codreanu, S. G.; May, J. C.; Sherrod, S. D.; McLean, J. A. Untargeted Molecular Discovery in Primary Metabolism: Collision Cross Section as a Molecular Descriptor in Ion Mobility-Mass Spectrometry. *Analytical Chemistry* **2018**, *90* (24), 14484–14492. https://doi.org/10.1021/acs.analchem.8b04322.

(6)   Wongtrakul-Kish, K.; Walsh, I.; Sim, L. C.; Mak, A.; Liau, B.; Ding, V.; Hayati, N.; Wang, H.; Choo, A.; Rudd, P. M.; Nguyen-Khuong, T. Combining Glucose Units, m/ z, and Collision Cross Section Values: Multiattribute Data for Increased Accuracy in Automated Glycosphingolipid Glycan Identifications and Its Application in Triple Negative Breast Cancer. *Analytical Chemistry* **2019**, *91* (14). https://doi.org/10.1021/acs.analchem.9b01476.

(7)   Zhong, Y.; Hyung, S. J.; Ruotolo, B. T. Characterizing the Resolution and Accuracy of a Second-Generation Traveling-Wave Ion Mobility Separator for Biomolecular Ions. *Analyst* **2011**, *136* (17), 3534–3541. https://doi.org/10.1039/c0an00987c.

(8)   May, J. C.; Dodds, J. N.; Kurulugama, R. T.; Stafford, G. C.; Fjeldsted, J. C.; McLean, J. A. Broadscale Resolving Power Performance of a High Precision Uniform Field Ion Mobility-Mass Spectrometer. *The Analyst* **2015**, *140* (20), 6824–6833. https://doi.org/10.1039/C5AN00923E.

(9)   Kaufmann, A.; Butcher, P.; Maden, K.; Walker, S.; Widmer, M. Does the Ion Mobility Resolving Power as Provided by Commercially Available Ion Mobility Quadrupole Time-of-Flight Mass Spectrometry Instruments Permit the Unambiguous Identification of Small Molecules in Complex Matrices? *Analytica Chimica Acta* **2020**, *1107*, 113–126. https://doi.org/10.1016/j.aca.2020.02.032.

(10)   Adams, K. J.; Montero, D.; Aga, D.; Fernandez-Lima, F. Isomer Separation of Polybrominated Diphenyl Ether Metabolites Using NanoESI-TIMS-MS. *International Journal for Ion Mobility Spectrometry* **2016**, *19* (2–3). https://doi.org/10.1007/s12127-016-0198-z.

(11)   Ridgeway, M. E.; Lubeck, M.; Jordens, J.; Mann, M.; Park, M. A. Trapped Ion Mobility Spectrometry: A Short Review. *International Journal of Mass Spectrometry*. 2018. https://doi.org/10.1016/j.ijms.2018.01.006.

(12)   Giles, K.; Ujma, J.; Wildgoose, J.; Pringle, S.; Richardson, K.; Langridge, D.; Green, M. A Cyclic Ion Mobility-Mass Spectrometry System. *Analytical Chemistry* **2019**, *91* (13), 8564–8573. https://doi.org/10.1021/acs.analchem.9b01838.

(13)   Tolmachev, A. v; Webb, I. K.; Ibrahim, Y. M.; Garimella, S. V. B.; Zhang, X.; Anderson, G. A.; Smith, R. D. Characterization of Ion Dynamics in Structures for Lossless Ion Manipulations. *Analytical Chemistry* **2014**, *86* (18), 9162–9168. https://doi.org/10.1021/ac502054p.

(14)   Ibrahim, Y. M.; Hamid, A. M.; Deng, L.; Garimella, S. V. B.; Webb, I. K.; Baker, E. S.; Smith, R. D. New Frontiers for Mass Spectrometry Based upon Structures for Lossless Ion Manipulations. *The Analyst* **2017**, *142* (7), 1010–1021. https://doi.org/10.1039/C7AN00031F.

(15)   Glen F. Rall, M. J. S. B. M. D. Opening New Paths for Biological Applications of Ion Mobility - Mass Spectrometry Using Structures for Lossless Ion Manipulations. *TrAC Trends in Analytical Chemistry* **2019**, *116* (1), 300–307. https://doi.org/10.1016/j.trac.2019.04.021.

(16)   Zhang, X.; Garimella, S. V. B.; Prost, S. A.; Webb, I. K.; Chen, T. C.; Tang, K.; Tolmachev, A. v.; Norheim, R. v.; Baker, E. S.; Anderson, G. A.; Ibrahim, Y. M.; Smith, R. D. Ion Trapping, Storage, and Ejection in Structures for Lossless Ion Manipulations. *Analytical Chemistry* **2015**, *87* (12), 6010–6016. https://doi.org/10.1021/acs.analchem.5b00214.

(17)   Chen, T. C.; Ibrahim, Y. M.; Webb, I. K.; Garimella, S. V. B.; Zhang, X.; Hamid, A. M.; Deng, L.; Karnesky, W. E.; Prost, S. A.; Sandoval, J. A.; Norheim, R. v.; Anderson, G. A.; Tolmachev, A. v.; Baker, E. S.; Smith, R. D. Mobility-Selected Ion Trapping and Enrichment Using Structures for Lossless Ion Manipulations. *Analytical Chemistry* **2016**, *88* (3), 1728–1733. https://doi.org/10.1021/acs.analchem.5b03910.

(18)   Webb, I. K.; Garimella, S. V. B.; Tolmachev, A. v; Chen, T.; Zhang, X.; Norheim, R. v; Prost, S. A.; LaMarche, B.; Anderson, G. A.; Ibrahim, Y. M.; Smith, R. D. Experimental Evaluation and Optimization of Structures for Lossless Ion Manipulations for Ion Mobility Spectrometry with Time-of-Flight Mass Spectrometry. *Analytical chemistry* **2014**, *86* (18), 9169–9176. https://doi.org/10.1021/ac502055e.

(19)   Garimella, S. V. B.; Ibrahim, Y. M.; Webb, I. K.; Ipsen, A. B.; Chen, T. C.; Tolmachev, A. v.; Baker, E. S.; Anderson, G. A.; Smith, R. D. Ion Manipulations in Structures for Lossless Ion Manipulations (SLIM): Computational Evaluation of a 90° Turn and a Switch. *Analyst* **2015**, *140* (20). https://doi.org/10.1039/c5an00844a.

(20)   Webb, I. K.; Garimella, S. V. B.; Tolmachev, A. v; Chen, T.; Zhang, X.; Cox, J. T.; Norheim, R. v; Prost, S. A.; LaMarche, B.; Anderson, G. A.; Ibrahim, Y. M.; Smith, R. D. Mobility-Resolved Ion Selection in Uniform Drift Field Ion Mobility Spectrometry/Mass Spectrometry: Dynamic Switching in Structures for Lossless Ion Manipulations. *Analytical Chemistry* **2014**, *86* (14), 9632–9637. https://doi.org/10.1021/ac502139e.

(21)   Ibrahim, Y. M.; Hamid, A. M.; Cox, J. T.; Garimella, S. V. B.; Smith, R. D. Ion Elevators and Escalators in Multilevel Structures for Lossless Ion Manipulations. *Analytical Chemistry* **2017**, *89* (3), 1972–1977. https://doi.org/10.1021/acs.analchem.6b04500.

(22)   Hollerbach, A. L.; Li, A.; Prabhakaran, A.; Nagy, G.; Harrilal, C. P.; Conant, C. R.; Norheim, R. v.; Schimelfenig, C. E.; Anderson, G. A.; Garimella, S. V. B.; Smith, R. D.; Ibrahim, Y. M. Ultra-High-Resolution Ion Mobility Separations over Extended Path Lengths and Mobility Ranges Achieved Using a Multilevel Structures for Lossless Ion Manipulations Module. *Analytical Chemistry* **2020**, *92* (11). https://doi.org/10.1021/acs.analchem.0c01397.

(23)   Garimella, S. V. B.; Ibrahim, Y. M.; Webb, I. K.; Tolmachev, A. v.; Zhang, X.; Prost, S. A.; Anderson, G. A.; Smith, R. D. Simulation of Electric Potentials and Ion Motion in Planar Electrode Structures for Lossless Ion Manipulations (SLIM). *Journal of the American Society for Mass Spectrometry* **2014**, *25* (11). https://doi.org/10.1007/s13361-014-0976-y.

(24)   Hamid, A. M.; Garimella, S. V. B.; Ibrahim, Y. M.; Deng, L.; Zheng, X.; Webb, I. K.; Anderson, G. A.; Prost, S. A.; Norheim, R. v.; Tolmachev, A. v.; Baker, E. S.; Smith, R. D. Achieving High Resolution Ion Mobility Separations Using Traveling Waves in Compact Multiturn Structures for Lossless Ion Manipulations. *Analytical Chemistry* **2016**, *88* (18), 8949–8956. https://doi.org/10.1021/acs.analchem.6b01914.

(25)   Li, A.; Garimella, S. V. B.; Ibrahim, Y. M. A Simulation Study of the Influence of the Traveling Wave Patterns on Ion Mobility Separations in Structures for Lossless Ion Manipulations. *Analyst* **2020**, *145* (1). https://doi.org/10.1039/c9an01509d.

(26)   Giles, K.; Pringle, S. D.; Worthington, K. R.; Little, D.; Wildgoose, J. L.; Bateman, R. H. Applications of a Travelling Wave-Based Radio-Frequency-Only Stacked Ring Ion Guide. *Rapid Communications in Mass Spectrometry* **2004**, *18* (20). https://doi.org/10.1002/rcm.1641.

(27)   Deng, L.; Ibrahim, Y. M.; Baker, E. S.; Aly, N. A.; Hamid, A. M.; Zhang, X.; Zheng, X.; Garimella, S. V. B.; Webb, I. K.; Prost, S. A.; Sandoval, J. A.; Norheim, R. v.; Anderson, G. A.; Tolmachev, A. v.; Smith, R. D. Ion Mobility Separations of Isomers Based upon Long Path Length Structures for Lossless Ion Manipulations Combined with Mass Spectrometry. *ChemistrySelect* **2016**, *1* (10), 2396–2399. https://doi.org/10.1002/slct.201600460.

(28)   Deng, L.; Ibrahim, Y. M.; Hamid, A. M.; Garimella, S. V. B.; Webb, I. K.; Zheng, X.; Prost, S. A.; Sandoval, J. A.; Norheim, R. v.; Anderson, G. A.; Tolmachev, A. v.; Baker, E. S.; Smith, R. D. Ultra-High Resolution Ion Mobility Separations Utilizing Traveling Waves in a 13 m Serpentine Path Length Structures for Lossless Ion Manipulations Module. *Analytical Chemistry* **2016**, *88* (18), 8957–8964. https://doi.org/10.1021/acs.analchem.6b01915.

(29)   Dodds, J. N.; May, J. C.; McLean, J. A. Correlating Resolving Power, Resolution, and Collision Cross Section: Unifying Cross-Platform Assessment of Separation Efficiency in Ion Mobility Spectrometry. *Analytical Chemistry* **2017**, *89* (22), 12176–12184. https://doi.org/10.1021/acs.analchem.7b02827.

(30) Deng, L.; Webb, I. K.; Garimella, S. V. B.; Hamid, A. M.; Zheng, X.; Norheim, R. v; Prost, S. A.; Anderson, G. A.; Sandoval, J. A.; Baker, E. S.; Ibrahim, Y. M.; Smith, R. D. Serpentine Ultralong Path with Extended Routing (SUPER) High Resolution Traveling Wave Ion Mobility-MS Using Structures for Lossless Ion Manipulations. *Analytical Chemistry* **2017**, *89* (8), 4628–4634. https://doi.org/10.1021/acs.analchem.7b00185.

(31) Wojcik, R.; Webb, I. K.; Deng, L.; Garimella, S. V. B.; Prost, S. A.; Ibrahim, Y. M.; Baker, E. S.; Smith, R. D. Lipid and Glycolipid Isomer Analyses Using Ultra-High Resolution Ion Mobility Spectrometry Separations. *International Journal of Molecular Sciences* **2017**, *18* (1), 1–12. https://doi.org/10.3390/ijms18010183.

(32) Prabhakaran, A.; Hamid, A. M.; Garimella, S. V. B.; Valenzuela, B. R.; Ewing, R. G.; Ibrahim, Y. M.; Smith, R. D. A Hybrid Constant and Oscillatory Field Ion Mobility Analyzer Using Structures for Lossless Ion Manipulations. *Journal of the American Society for Mass Spectrometry* **2018**, *29* (2). https://doi.org/10.1007/s13361-017-1841-6.

(33) Deng, L.; Garimella, S. V. B.; Hamid, A. M.; Webb, I. K.; Attah, I. K.; Norheim, R. v.; Prost, S. A.; Zheng, X.; Sandoval, J. A.; Baker, E. S.; Ibrahim, Y. M.; Smith, R. D. Compression Ratio Ion Mobility Programming (CRIMP) Accumulation and Compression of Billions of Ions for Ion Mobility-Mass Spectrometry Using Traveling Waves in Structures for Lossless Ion Manipulations (SLIM). *Analytical Chemistry* **2017**, *89* (12), 6432–6439. https://doi.org/10.1021/acs.analchem.7b00189.

(34) Garimella, S. V. B.; Webb, I. K.; Prabhakaran, A.; Attah, I. K.; Ibrahim, Y. M.; Smith, R. D. Design of a TW-SLIM Module for Dual Polarity Confinement, Transport, and Reactions. *Journal of the American Society for Mass Spectrometry* **2017**, *28* (7), 1442–1449. https://doi.org/10.1007/s13361-017-1680-5.

(35) Attah, I. K.; Garimella, S. V. B.; Webb, I. K.; Nagy, G.; Norheim, R. v.; Schimelfenig, C. E.; Ibrahim, Y. M.; Smith, R. D. Dual Polarity Ion Confinement and Mobility Separations. *Journal of The American Society for Mass Spectrometry* **2019**. https://doi.org/10.1007/s13361-019-02138-1.

(36) Attah, I. K.; Nagy, G.; Garimella, S. V. B.; Norheim, R. v.; Anderson, G. A.; Ibrahim, Y. M.; Smith, R. D. Traveling-Wave-Based Electrodynamic Switch for Concurrent Dual-Polarity Ion Manipulations in Structures for Lossless Ion Manipulations. *Analytical Chemistry* **2019**, *91*, 14712–14718. https://doi.org/10.1021/acs.analchem.9b03987.

(37) Nichols, C. M.; May, J. C.; Sherrod, S. D.; McLean, J. A. Automated Flow Injection Method for the High Precision Determination of Drift Tube Ion Mobility Collision Cross Sections. *The Analyst* **2018**, *143* (7), 1556–1559. https://doi.org/10.1039/C8AN00056E.

(38) Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L.; François, R.; Grolemund, G.; Hayes, A.; Henry, L.; Hester, J.; Kuhn, M.; Pedersen, T.; Miller, E.; Bache, S.; Müller, K.; Ooms, J.; Robinson, D.; Seidel, D.; Spinu, V.; Takahashi, K.; Vaughan, D.; Wilke, C.; Woo, K.; Yutani, H. Welcome to the Tidyverse. *Journal of Open Source Software* **2019**, *4* (43), 1686. https://doi.org/10.21105/joss.01686.

(39) Siems, W. F.; Wu, C.; Tarver, E. E.; Hill, H. H.; Larsen, P. R.; McMinn, D. G. Measuring the Resolving Power of Ion Mobility Spectrometers. *Analytical Chemistry* **1994**, *66* (23). https://doi.org/10.1021/ac00095a014.

(40) May, J. C.; Goodwin, C. R.; Lareau, N. M.; Leaptrot, K. L.; Morris, C. B.; Kurulugama, R. T.; Mordehai, A.; Klein, C.; Barry, W.; Darland, E.; Overney, G.; Imatani, K.; Stafford, G. C.; Fjeldsted, J. C.; McLean, J. A. Conformational Ordering of Biomolecules in the Gas Phase: Nitrogen Collision Cross Sections Measured on a Prototype High Resolution Drift Tube Ion Mobility-Mass Spectrometer. *Analytical Chemistry* **2014**, *86* (4), 2107–2116. https://doi.org/10.1021/ac4038448.

(41) Stow, S. M.; Causon, T. J.; Zheng, X.; Kurulugama, R. T.; Mairinger, T.; May, J. C.; Rennie, E. E.; Baker, E. S.; Smith, R. D.; McLean, J. A.; Hann, S.; Fjeldsted, J. C. An Interlaboratory Evaluation of Drift Tube Ion Mobility–Mass Spectrometry Collision Cross Section Measurements. *Analytical Chemistry* **2017**, *89* (17), 9048–9055. https://doi.org/10.1021/acs.analchem.7b01729.

(42) Shvartsburg, A. A.; Smith, R. D. Fundamentals of Traveling Wave Ion Mobility Spectrometry. *Analytical Chemistry* **2008**, *80* (24). https://doi.org/10.1021/ac8016295.

(43) May, J. C.; Knochenmuss, R.; Fjeldsted, J. C.; McLean, J. A. Resolution of Isomeric Mixtures in Ion Mobility Using a Combined Demultiplexing and Peak Deconvolution Technique. *Analytical Chemistry* **2020**. https://doi.org/10.1021/acs.analchem.9b05718.

(44) Smith, D. P.; Knapman, T. W.; Campuzano, I.; Malham, R. W.; Berryman, J. T.; Radford, S. E.; Ashcroft, A. E. Deciphering Drift Time Measurements from Travelling Wave Ion Mobility Spectrometry-Mass Spectrometry Studies. *European Journal of Mass Spectrometry* **2009**, *15* (2), 113–130. https://doi.org/10.1255/ejms.947.

(45) Spangler, G. E. Classical Aspects of the Reduced Collision Cross Section for Elastic Scattering in Ion Mobility Spectrometry. *International Journal for Ion Mobility Spectrometry* **2013**, *16* (3). https://doi.org/10.1007/s12127-012-0117-x.

(46) Causon, T. J.; Hann, S. Uncertainty Estimations for Collision Cross Section Determination via Uniform Field Drift Tube-Ion Mobility-Mass Spectrometry. *Journal of the American Society for Mass Spectrometry* **2020**, *31* (10), 2102–2110. https://doi.org/10.1021/jasms.0c00233.

(47) Li, A.; Conant, C. R.; Zheng, X.; Bloodsworth, K. J.; Orton, D. J.; Garimella, S. V. B.; Attah, I. K.; Nagy, G.; Smith, R. D.; Ibrahim, Y. M. Assessing Collision Cross Section Calibration Strategies for Traveling Wave-Based Ion Mobility Separations in Structures for Lossless Ion Manipulations. *Analytical Chemistry* **2020**, *92* (22), 14976–14982. https://doi.org/10.1021/acs.analchem.0c02829.

(48) Richardson, K.; Langridge, D.; Giles, K. Fundamentals of Travelling Wave Ion Mobility Revisited: I. Smoothly Moving Waves. *International Journal of Mass Spectrometry* **2018**, *428*, 71–80. https://doi.org/10.1016/j.ijms.2018.03.007.

(49) May, J. C.; McLean, J. A. The Influence of Drift Gas Composition on the Separation Mechanism in Traveling Wave Ion Mobility Spectrometry: Insight from Electrodynamic Simulations. *International Journal for Ion Mobility Spectrometry* **2013**, *16* (2), 85–94. https://doi.org/10.1007/s12127-013-0123-7.

CHAPTER 5

A COLLISION CROSS SECTION CALIBRATION STRATEGY FOR LIPID

MEASUREMENTS IN SLIM-BASED HIGH RESOLUTION ION MOBILITY[**]

5.1   Introduction

Ion mobility interfaced with mass spectrometry (IM-MS) has become an important technique for

the analysis of complex biological samples.[1–6] Similar to chromatography, the added structurally-

selective separation capabilities of IM improve analyte coverage and aid in the discrimination of

isomeric and isobaric species that otherwise hinder comprehensive analysis when using MS

alone.[7,8] Although chromatographic methods can be extensively optimized to enhance chemical

selectivity and chromatographic resolution, achieving reproducible retention times across different

laboratories remains a challenge.[9–11] Conversely, there are fewer options to tailor the selectivity in

IM as separations are based directly on an intrinsic physical property of the analyte, namely its

gas-phase structure or structures. Because of this, IM-derived collision cross section (CCS) values

add a highly reproducible metric for filtering and annotating features derived from mass spectra,

while simultaneously providing a specific structural descriptor for molecular species.[12–14]

Collision cross section values can be directly derived from classical electrodynamics using

uniform-field drift tube ion mobility (DTIM) measurements via the fundamental low field Mason-

Schamp equation (^{DT}CCS).[15,16] On the other hand, traveling wave ion mobility (TWIM)-based

techniques use dynamic electric fields for separation, so CCS determination is less straightforward. Though progress has been made toward derivation of a functional first-principles equation from traveling wave fundamentals,[17] in practice traveling wave collision cross section values ([TW]CCS) are typically obtained through calibration from experimentally measured arrival times of calibrants with known or agreed upon [DT]CCS values.[18–22] Calibrating [TW]CCS introduces error in the form of a bias from fundamentally measured reference values (typically DTIM) that is dependent on both the functional equation form used in the calibration as well as the structural similarity of the calibrants and analytes.[20,23–27] However, with appropriate considerations, these calibration methods can regularly achieve biases of < 2% and have shown high interlaboratory reproducibility (< 1%).[13,28–30] This has led to widespread adoption of CCS as a compound annotation parameter and the construction of many CCS libraries to support such studies.[13,28,31–35]

While the conventional resolution range of commercially available DTIM and TWIM systems (40 to 60)[7] has been successful in discriminating many molecular species in complex spectra, there remain many more structurally similar isomeric and isobaric species that require higher resolution.[36,37] Despite the importance of DTIM in providing direct CCS measurements from first-principles theory, DTIM instruments rarely achieve resolving powers above ca. 100. Recent advances in structures for lossless ion manipulation (SLIM) technology have enabled the development of a TW-based high resolution IM (HRIM) system with resolving powers in excess of ~200.[38–40] The increased ion mobility resolution of SLIM-based HRIM has enabled the separation of many biologically relevant compounds of various chemical classes, as well as the elucidation of previously unseen molecular features.[41–44] To support the proper annotation of these additional features, robust calibration methods need to be developed and validated such that SLIM-

based CCS values ($^{TW\ SLIM}$CCS) can be reliably and reproducibly derived from HRIM measurements.

Calibration of CCS for TW-SLIM has been explored previously in a limited capacity using variously modified TWIM calibration protocols. Whereas fundamental differences between the TWIM platforms used in these studies raised concerns regarding the influence of ion heating, the results provided evidence that calibrated CCS values are not significantly influenced by TW separation parameters such as wave height and amplitude, and CCS biases under 2% were achievable in most cases.[40,45,46] Prior investigations into the utility of conventional resolution CCS databases have demonstrated the importance of low bias and high precision (reproducibility and repeatability) for untargeted applications, as lowering database search tolerances below 1% can drastically reduce the number of candidate identifications and improve annotation confidence.[47] With the inherently higher precision of HRIM this is even more salient, as many biologically-relevant isomers exhibit CCS differences of less than 2%.[37] Therefore, $^{TW\text{-}SLIM}$CCS calibration with expected biases under 1% and high reproducibility (< 0.5% RSD) could provide improved confidence in the expansion of database matching using HRIM.

Here, we describe the development and evaluation of a simple, reproducible $^{TW\text{-}SLIM}$CCS calibration strategy focused on lipids. Lipids are a functionally and structurally diverse class of biomolecules with a high prevalence of isomers and distinct CCS trends that can aid in their annotation and characterization.[48,49] Additionally, the various subclasses of lipids occupy a well-defined range of *m/z* and CCS values (Figure 5.1), which allows the parameters of the calibration approach to be confined while enabling the generation of class-specific recommendations for calibrant selection and calibration methodology. To assess the choice of calibrants, we compare the calibrated CCS obtained from structurally similar calibrant sets, as well as a more broadly
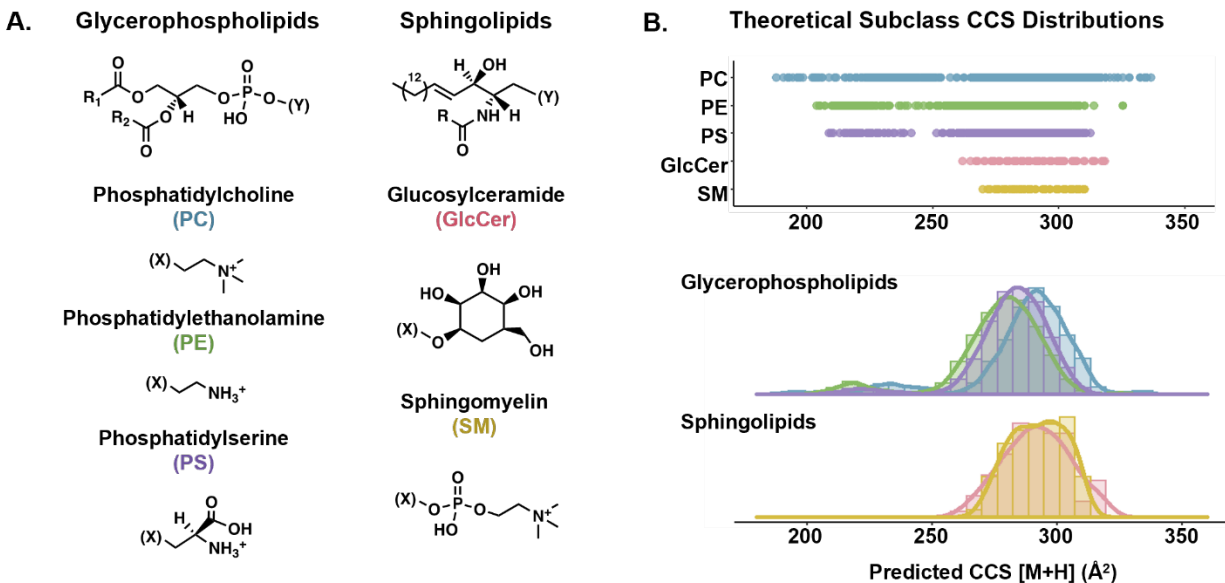
Figure 5.1 Overview of the two lipid classes and five subclasses surveyed in this study. (A) General chemical structure of each class and headgroup of each subclass. (B) Range and distribution of predicted [M+H] CCS values for all lipids from each subclass as rep-resented in the LIPID MAPS Structural Database and predicted using LipidCCS from the Zhu laboratory.

applicable calibrant mixture of current and widespread use. We also examine the utility of a generalizable calibration method combined with subclass-specific CCS correction factors to provide the most precise and accurate CCS determinations from high resolution TW-SLIM measurements.

5.2   Experimental methods

5.2.1   Materials and solvents

High purity (Optima grade) solvents including water, methanol, acetonitrile, chloroform, and formic acid were obtained from Fisher Scientific (Hampton, NH, USA). A tuning mixture containing betaine and a series of symmetrically branched hexakis(fluoroalkoxy)phosphazines (HFAPs, ESI-L low concentration tuning mixture, Agilent) was used for instrument tuning and CCS calibration. Purified TLC fractions of total lipid extracts including phosphatidylcholine (PC, chicken egg), phosphatidylethanolamine (PE, chicken egg), phosphatidylserine (PS, porcine brain), cerebroside (GlcCer, porcine brain), and sphingomyelin (SM, porcine brain), were purchased as lyophilized solids from Avanti Polar Lipids (Birmingham, AL, USA) and were reconstituted in chloroform, then prepared to a final concentration of 10 µg/mL in 1:2 chloroform:methanol for analysis. A deuterated lipid standard mix of varying lipid concentrations (SPLASH, Avanti) was diluted 1:10 in 1:2 chloroform:methanol for analysis.

5.2.2   Instrumentation

Data were acquired using a 13-meter serpentine path SLIM-based HRIM platform (beta prototype, MOBILion Systems) integrated with a time-of-flight mass spectrometer (6546, Agilent Technologies). A schematic of this instrument is shown in Figure F1. Samples were introduced using a liquid chromatography system (1290, Agilent) and ionized by electrospray ionization (Jet Stream, Agilent). TW-SLIM ion mobility experiments were conducted in pure nitrogen drift gas

at ambient temperature, resulting in nitrogen-specific cross section measurements (CCS_{N2}). Measurements for the standard lipid mix were also made using a commercial DTIM-MS system (6560, Agilent) for reference CCS comparison for these lipids.[50,51]

5.2.3   Data acquisition

Lipid extract samples were injected using a 3-minute automated flow injection acquisition with a constant flow solvent of 0.1% formic acid in 1:1 methanol:water at a carrier flow rate of 70 µL/min and an injection volume of 10 µL. Reversed phase liquid chromatography (RPLC) was used for the standard lipid mix, and detailed parameters, solvents, and gradient can be found in Figure F2. In all cases, the ESI source was operated in positive ion mode using the following conditions: nebulizer pressure, 20 psi; sheath gas flow rate, 12 L/min; sheath gas temperature, 275 °C; drying gas flow rate, 5 L/min; drying gas temperature, 325 °C; capillary voltage, 4000 V; entrance nozzle voltage, 2000 V. The SLIM boards were operated at ca. 2.5 Torr. TW-based separation was performed using a wave speed of 180 m/s and a peak-to-peak wave amplitude of 40 V_{pp}. These separation parameters were chosen based on optimal conditions for high resolving power in the *m/z* range of the lipid analytes as described previously.[40] Data was acquired via MassHunter Acquisition (v. 9.0, Agilent) and MOBILion software. For CCS calibration with HFAPs, data for tune mix was acquired in a separate experiment using identical instrument parameters (i.e., external CCS calibration). For calibration using the lipids, lipids observed within the spectra from each extract was selected as calibrants (i.e., internal CCS calibration). For data acquired on the DTIMS instrument, the LC and ESI source conditions were identical to those used on the TW-SLIM system. The drift tube was operated under 3.95 Torr nitrogen gas, and additional drift tube parameters were as follows: ion trap fill time, 20 ms; ion trap release time, 300 µs; drift tube entrance, 1474 V; drift tube exit, 224 V; rear funnel entrance, 217.5 V; rear funnel RF, 150 Vpp;

rear funnel exit, 45 V(matched to the QTOF autotune setting); and IM hexapole delta, -8 V. The QTOF stage was operated in Low Mass Range (*m/z* 50-1700), ion slicer operated at High Sensitivity and the digitizer operated at 2 GHz Extended Dynamic Range. Single-field CCS values were obtained using HFAP drift times as described previously.

5.2.4 Data processing and software

PNNL Preprocessor (version 3.0) was used for IM-MS file conversion and drift bin compression (2-to-1) for the TW-SLIM data.[52] Lipid feature arrival times (peak centroids) were extracted manually within MassHunter IM-MS Browser (developer version). To increase confidence in peak selection, only features falling within the expected lipid IM-MS correlation region were considered for extraction.[53] CCS calibration and bias calculations were performed in Excel (Microsoft).

5.2.5 Guiding calibration theory

Calibration of CCS from commercial TW systems is commonly performed using a set of calibrants with known ^{DT}CCS values.[21] Their experimental arrival time (t_A) is plotted against the corresponding "reduced" ^{DT}CCS values (CCS'), as calculated using eq. 5.1 where *z* represents the ion charge and μ is the reduced mass of the ion-neutral pair.

$$\text{CCS}' = \frac{^{DT}\text{CCS}}{z\sqrt{\frac{1}{\mu}}} \tag{5.1}$$

As described previously, conformational space occupancy arises from average density as related to the cubic volume and squared area of a given set of biomolecules, giving rise to a length squared versus length cubed relationship.[54] Similarly, the arrival time-CCS' relationship has been modeled using many nonlinear equation forms but generally power functions or polynomials are used for TW data.[18,20,21,23] Studies using both power functions and polynomials have shown high

103

reproducibility and relatively low biases under varying TW amplitudes and speeds.[40,45] Trinomial equation forms resulted in the lowest biases from ^{DT}CCS values in these prior studies and therefore is used as the basis for CCS calibration in this work (eq. 5.2).

$$CCS' = A(t_A)^3 + B(t_A)^2 + C(t_A) + D \tag{5.2}$$

Calibrated CCS bias from reference ^{DT}CCS values (CCS_{ref}) is used as a comparison metric to estimate accuracy for the different calibration methods (eq. 5.3). Values from a large database of standardized ^{DT}CCS values, the Unified CCS Compendium, were used as reference values for lipids with a database m/z match. [34]

$$\% \text{ CCS Bias} = \frac{CCS - CCS_{ref}}{CCS_{ref}} \times 100 \tag{5.3}$$

5.3   Results and discussion

Here we evaluate three strategies for CCS calibration of lipid measurements from a SLIM-based HRIM platform: (1) The broadly available HFAP tuning mixture was used for calibration following the protocols set forth in previous TW-SLIM studies which are in turn modeled from conventional TW calibration practices; (2) calibration using subclass-specific lipid calibrant sets was evaluated to determine if chemically similar calibrants significantly impacted the $^{TW\text{-}SLIM}$CCS measurement bias from reference CCS values obtained from DTIM measurements; (3) a correction factor was applied to the calibration obtained using the HFAPs to determine if a more generalizable calibration protocol could be developed with similar or better biases observed from the first two approaches.

5.3.2   HFAPs as TW-SLIM Calibrants

Here, we first evaluate a widely used set of calibrants, HFAPs, for calibration of the lipid features, which is desirable from a standpoint that this tuning mixture is widely accessible and currently utilized for tuning and benchmarking the instrumentation used in this study. Additionally, the ions from this mixture cover the entire experimental arrival time range and reference ^{DT}CCS values of these compounds are available to assess CCS measurement accuracy.[12] Using HFAPs with the calibration method described in eq. 5.1 and 5.2 yielded calibrated $^{TW\text{-}SLIM}$CCS values with high reproducibility (<0.35% RSD for all lipids), however systematic subclass-dependent biases of +2-3% from drift tube values were observed across all five lipid sub-classes (Figure F3). A systematic bias of ~2% is consistent with the results from Hines, et. al. where HFAP was used to calibrate lipids and was attributed to a structural mismatch between the calibrants (phosphazines) and analytes (lipids).[24]

5.3.3 *Lipid-specific calibrants*

It is well-documented that the choice of calibrant plays a significant role in the resulting CCS calibration accuracy.[23–25] As structurally-similar calibrants have been shown to improve CCS bias in many cases, we next curated subclass-specific sets of lipid calibrants which are observed in each lipid extract using the reference ^{DT}CCS values in the Unified CCS Compendium and criteria outlined in Figure 5.2(A). Briefly, features were only considered as calibrants if they had a single symmetrical HRIM profile with no indication of multiple contributing structures (i.e., peak asymmetry or splitting), and a ^{DT}CCS value matching the feature *m/z* was also present in the Compendium. Additionally, quality thresholds of high abundance and multi-replicate observations were used to further screen the candidate calibrants. This selection process resulted in 5-7 calibrant lipids from each subclass extract, each of which spanned the majority of the arrival time range of the analytes (Figure 5.2(B)).
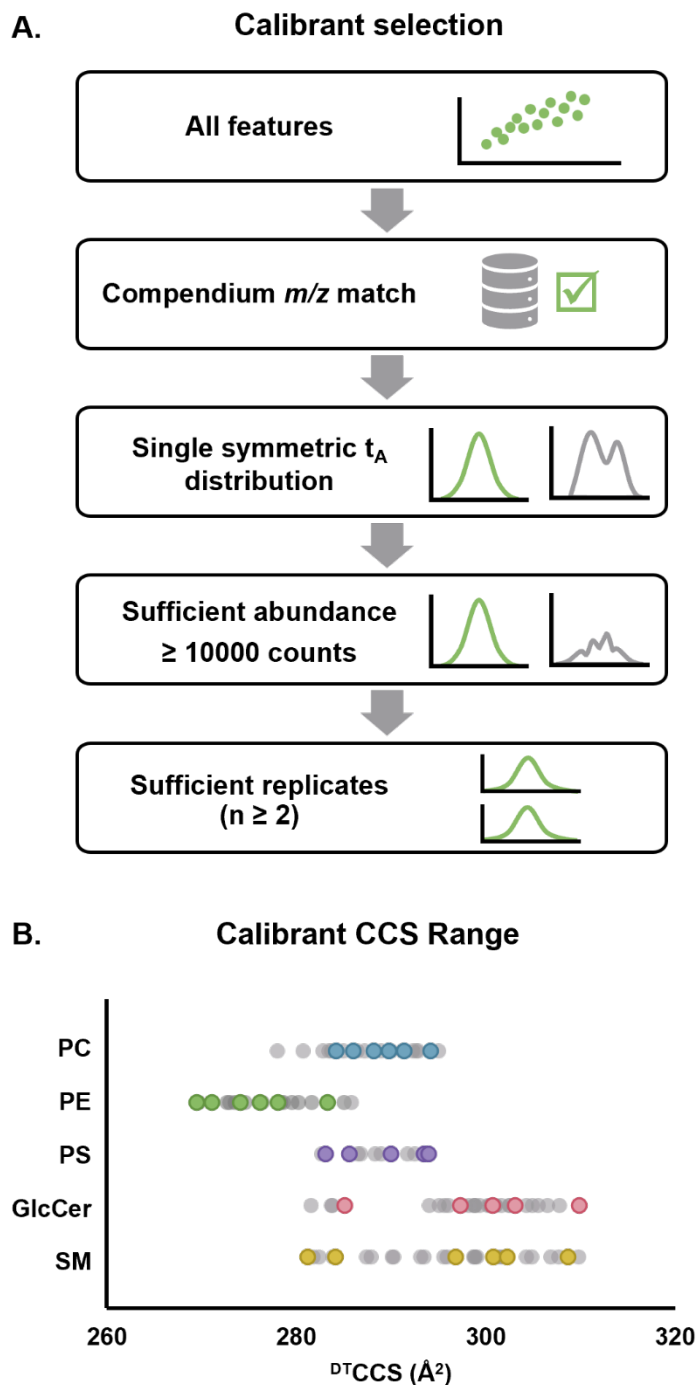
Figure 5.2 Subclass-specific calibrant selection. (A) Criteria used to select calibrant lipids from each subclass extract. (B) CCS range of each set of calibrants chosen as compared to the range of all observed lipid features (grey) in each class.

Using lipids of the same subclass to calibrate the analytes was found to lower the CCS bias substantially to an absolute average bias of 0.48%. All lipid-calibrated $^{\text{TW-SLIM}}$CCS values were found to be within 2% of the reference $^{\text{DT}}$CCS values, and 80% were within 1% bias (Figure 5.3(A)). While promising, this strategy requires a relatively large number of calibrants with known CCS values, which may not be achievable in most situations. Additionally, for this approach, it is important to choose calibrants which span the full range of analyte arrival times due to the high errors associated with extrapolating trinomial fits. Whereas the more commonly used power fits perform better when the calibration range must be extrapolated, these equations can result in higher CCS biases than the trinomial fit used here.[40] Thus a calibration procedure that incorporates calibrants spanning a broad range of arrival times with a trinomial calibration equation is desirable.

5.3.4 Generalizable calibration using HFAPs

Aside from using new calibrant sets, there is precedent for adjusting calibrated CCS bias using corrections of varying degrees of complexity. Including a correction factor to the $^{\text{TW}}$CCS calibration has been used to address various systematic contributions to bias, including instrumental time delays, ion motion disparities, and calibrant structural differences, while providing a more generalizable calibration protocol accessible to a broader community of researchers.[12,17,26,55] Here, we apply a subclass-specific semi-empirical correction to the trinomial fit based on the HFAP calibrants. To achieve this, each $^{\text{TW-SLIM}}$CCS determined from HFAP calibration was rescaled to the average bias of its respective subclass using a simple linear correction factor. This strategy lowered the bias to an absolute average of 0.38%, which is a lower bias that what was observed with using lipid-specific calibrants (Figure 5.3(B)). Using a subclass-specific correction factor also resulted in less bias variability than the lipid calibrants, with 98% of biases under 1% and all values under 1.5% (Figure 5.4). In addition to providing lower variability,
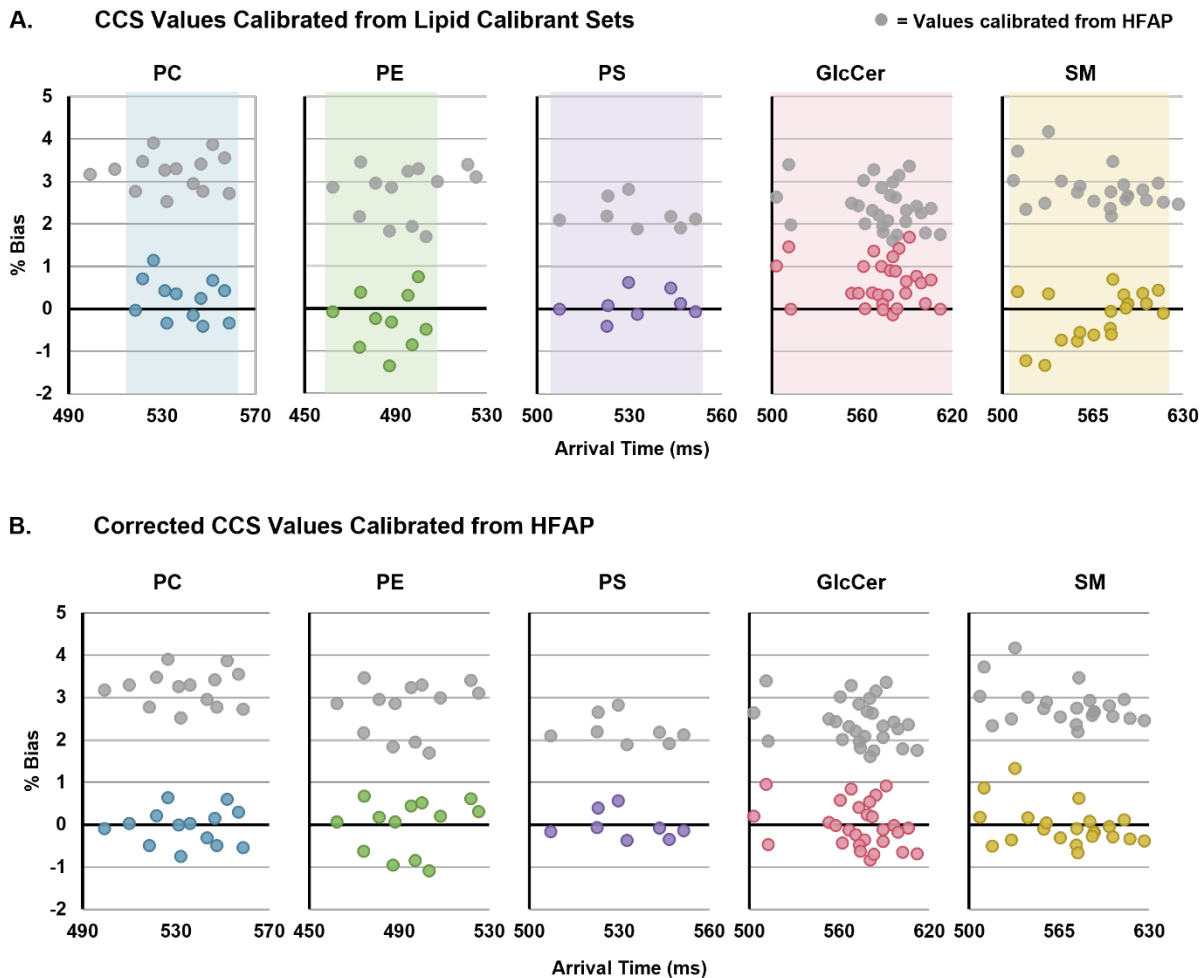
Figure 5.3 CCS biases from triplicate ^{DT}CCS values two calibration strategies. In both panels, grey values represent those calibrated using HFAP calibrants with no added correction factor. (A) Lipids calibrated from lipid calibrants within the same subclass. Shaded regions represent the arrival time ranges of the calibrant sets. (B) Lipids calibrated from HFAP calibrants with an added subclass-specific correction factor.

using HFAPs with a correction factor is a more straightforward and broadly applicable calibration approach. HFAPs are widely available, exogenous compounds that cover a broad experimental arrival time range, and their ^{DT}CCS values have been thoroughly vetted by the community. This mixture is also stable and structurally defined, whereas lipids are more prone to solution-phase degradation and may also have unresolved structural contributions to their mobility profiles. For these reasons, the correction factor applied to HFAP-based calibration is a more practical strategy for obtaining the CCS of lipids from HRIM measurements.

Using the corrected HFAP calibration, an empirical HRIM-derived database of over 90 $^{TW\text{-}SLIM}$CCS values was compiled (Table F2), including over 20 lipids that were previously unresolved by conventional resolution DTIM. The cerebrosides, in particular, produced many new spectral features, likely as a result of isomeric variations in the sugar headgroup. In many subclasses, new conformational trendlines were observed in the high-resolution dataset that were obscured by higher abundance isobars in conventional resolution. In all cases, mapping the correlations of the high-precision calibrated CCS values will be essential for the characterization of the newly resolved features. The publication of this database provides a significant resource to the community and may be applied to future studies in HRIM lipid annotation and characterization. Similar to databases for molecular annotation at varying MS resolving power, we find this is also, and potentially more, necessary for ion mobility-derived CCS values.

5.3.5 *Application to other lipids*

To test the generalizability of the correction approach to broader lipidomic applications, a standard mix of heavy-labelled lipids from various classes was analyzed. For comparison, ^{DT}CCS values for this standard lipid mix were measured and have been published to the Unified CCS Compendium.[56] The observed features were extracted, identified, and then subjected to calibration
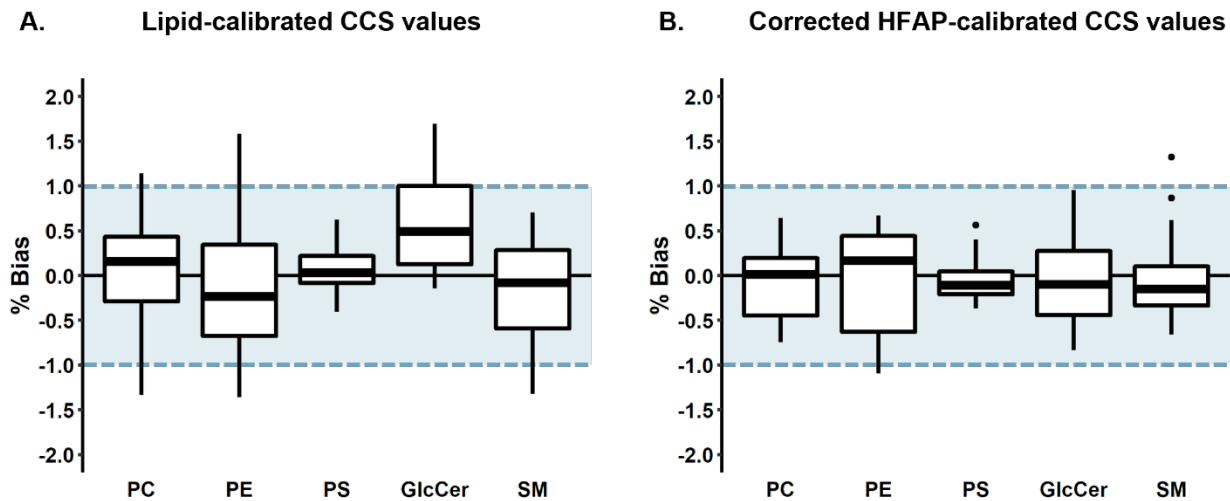
Figure 5.4 CCS bias distributions for two calibration strategies (A) CCS values calibrated from calibrants within the same subclass and (B) values calibrated from HFAPs with an applied correction factor. Center lines represent the median of each subclass. Markers outside whiskers represent statistical outliers. Blue shaded regions represent the target bias of ±1%.

using HFAP and an averaged correction factor from the extract lipids that were originally evaluated in this study. This approach overcorrected the CCS and resulted in negative biases ranging from 0.5 to 2.4, as shown in Figure 5.5(A). The standard lipids of subclasses from which the correction factor was derived (i.e. PE, PS, and SM) had biases within the range of their corresponding subclasses (-1.1 to 1.3%, Figure 5.3(B)), but the glycerolipids (DG and TG) exhibited higher biases in general (up to -1.6%). In addition to lipid class, arrival time range also contributed to these biases (Figure 5.5(B)). Although lysophosphatidylcholine (LPC) is within the classes examined in this study, its arrival time falls below the range of those lipids, and accordingly, the observed bias was higher (2.4%). These results suggest a smaller magnitude correction may be useful for general use in lipidomics experiments where subclass information is unknown.

While the average subclass specific correction was 2.7%, a smaller correction of 1.5% was applied as a more "general" correction to the SPLASH mix standards. Applying this more conservative correction resulted in lower biases, with the highest absolute value at 1.1% (Figure 5.5(C), Table F1). As expected, application of the smaller correction factor to the other lipid extract species resulted in higher absolute biases than their subclass-specific corrections (Figure 5.5(D)). However, most values (94%) still fell within 2% bias, which is comparable to the expected CCS bias in conventional TWIM experiments and is sufficient for many applications. While not explicitly explored here, for cases where the specific lipid subclass is known, applying a second subclass-specific correction would result in even lower biases as observed here.

5.4   Conclusions

In this work we evaluated the accuracy and practical applicability of various CCS calibration strategies for five lipid subclasses on a SLIM-based TWIM platform. Using a simple trinomial calibration based on a HFAP tuning mixture resulted in subclass-dependent systematic biases of
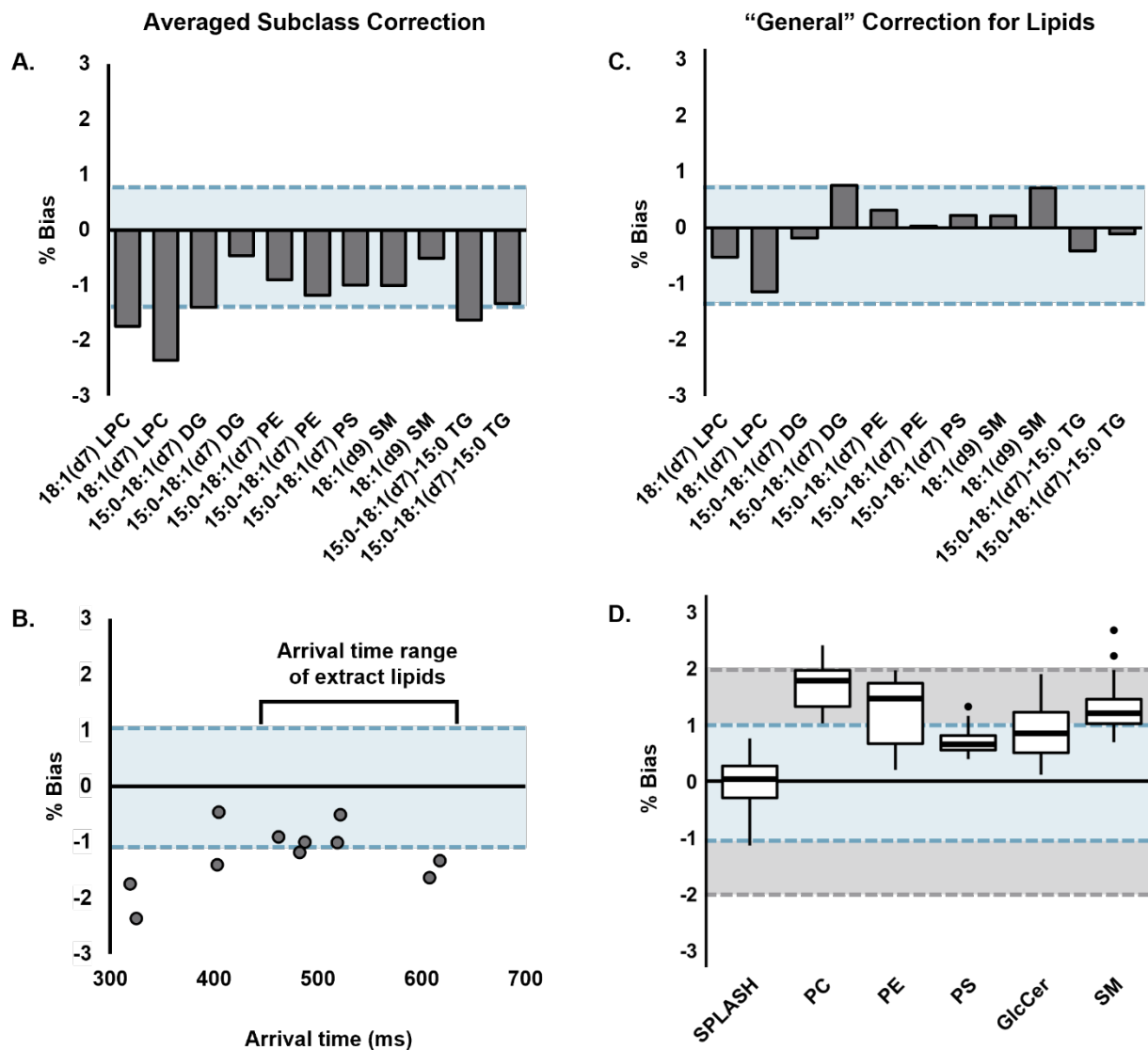
Figure 5.5 CCS biases from ^{DT}CCS values of heavy-labelled lipid standard mix components. In all panels, blue shaded regions represent the target bias of ±1%. (A) Standard lipids calibrated with HFAP and an averaged correction factor derived from the five standard extracts examined in this study. Analytes are arranged in order of increasing measured arrival times. (B) The arrival time range of the standards as compared to those of the lipid extract species. (C) Standard biases when calibrated with HFAP and a "general" correction factor of 1.5%. (D) Biases of the standard mix (SPLASH) as well as lipid extract species when calibrated using the general correction factor. The grey shaded region represents the expected bias of conventional TWIMS, ±2%.

2-3% from reference ^{DT}CCS values. While curation of custom calibrant sets of lipids within each subclass lowered the bias to within 0.5% on average, using subclass specific semi-empirical correction factors with the more generalizable HFAP calibration provided the lowest biases (< 0.4%) and variability (98% of values under 1% bias). This HFAP-based correction strategy provides a straightforward and accessible method for obtaining highly reproducible lipid CCS values and, with empirically determined correction factors, it can be generalized to other compounds. Using this calibration method, we curated a HRIM CCS database of over 90 calibrated lipid values from all five subclasses, including those of many newly resolved lipid features. The routine acquisition of high precision CCS values with bias under 1% will enable the construction of new HRIM libraries, as well as standardization and inter-laboratory comparison, paving the way for further characterization of newly elucidated spectral features in support of untargeted applications.

5.5 Acknowledgments

5.6 References

(1)   Burnum-Johnson, K. E.; Zheng, X.; Dodds, J. N.; Ash, J.; Fourches, D.; Nicora, C. D.; Wendler, J. P.; Metz, T. O.; Waters, K. M.; Jansson, J. K.; Smith, R. D.; Baker, E. S. Ion Mobility Spectrometry and the Omics: Distinguishing Isomers, Molecular Classes and Contaminant Ions in Complex Samples. *TrAC Trends in Analytical Chemistry* **2019**, *116*, 292–299. https://doi.org/10.1016/j.trac.2019.04.022.

(2)   Mairinger, T.; Causon, T. J.; Hann, S. The Potential of Ion Mobility–Mass Spectrometry for Non-Targeted Metabolomics. *Current Opinion in Chemical Biology* **2018**, *42*, 9–15. https://doi.org/10.1016/j.cbpa.2017.10.015.

(3)   Zheng, X.; Wojcik, R.; Zhang, X.; Ibrahim, Y. M.; Burnum-Johnson, K. E.; Orton, D. J.; Monroe, M. E.; Moore, R. J.; Smith, R. D.; Baker, E. S. Coupling Front-End Separations, Ion Mobility Spectrometry, and Mass Spectrometry for Enhanced Multidimensional Biological and Environmental Analyses. *Annual Review of Analytical Chemistry* **2017**, *10* (1), 71–92. https://doi.org/10.1146/annurev-anchem-061516-045212.

(4)   McLean, J. A.; Ruotolo, B. T.; Gillig, K. J.; Russell, D. H. Ion Mobility–Mass Spectrometry: A New Paradigm for Proteomics. *International Journal of Mass Spectrometry* **2005**, *240* (3), 301–315. https://doi.org/10.1016/j.ijms.2004.10.003.

(5)   Sherrod, S. D.; McLean, J. A. Systems-Wide High-Dimensional Data Acquisition and Informatics Using Structural Mass Spectrometry Strategies. *Clinical Chemistry* **2016**, *62* (1), 77–83. https://doi.org/10.1373/clinchem.2015.238261.

(6)   May, J. C.; Gant-Branum, R. L.; McLean, J. A. Targeting the Untargeted in Molecular Phenomics with Structurally-Selective Ion Mobility-Mass Spectrometry. *Current Opinion in Biotechnology* **2016**, *39*, 192–197. https://doi.org/10.1016/j.copbio.2016.04.013.

(7)   Dodds, J. N.; May, J. C.; McLean, J. A. Correlating Resolving Power, Resolution, and Collision Cross Section: Unifying Cross-Platform Assessment of Separation Efficiency in Ion Mobility Spectrometry. *Analytical Chemistry* **2017**, *89* (22), 12176–12184. https://doi.org/10.1021/acs.analchem.7b02827.

(8)   Dodds, J. N.; May, J. C.; McLean, J. A. Investigation of the Complete Suite of the Leucine and Isoleucine Isomers: Toward Prediction of Ion Mobility Separation Capabilities. *Analytical Chemistry* **2017**, *89* (1), 952–959. https://doi.org/10.1021/acs.analchem.6b04171.

(9)   Žuvela, P.; Skoczylas, M.; Jay Liu, J.; Bączek, T.; Kaliszan, R.; Wong, M. W.; Buszewski, B. Column Characterization and Selection Systems in Reversed-Phase High-Performance Liquid Chromatography. *Chemical Reviews* **2019**, *119* (6), 3674–3729. https://doi.org/10.1021/acs.chemrev.8b00246.

(10)   Witting, M.; Böcker, S. Current Status of Retention Time Prediction in Metabolite Identification. *Journal of Separation Science* **2020**, *43* (9–10), 1746–1754. https://doi.org/10.1002/jssc.202000060.

(11)   Mouchahoir, T.; Schiel, J. E.; Rogers, R.; Heckert, A.; Place, B. J.; Ammerman, A.; Li, X.; Robinson, T.; Schmidt, B.; Chumsae, C. M.; Li, X.; Manuilov, A. v.; Yan, B.; Staples, G.

O.; Ren, D.; Veach, A. J.; Wang, D.; Yared, W.; Sosic, Z.; Wang, Y.; Zang, L.; Leone, A. M.; Liu, P.; Ludwig, R.; Tao, L.; Wu, W.; Cansizoglu, A.; Hanneman, A.; Adams, G. W.; Perdivara, I.; Walker, H.; Wilson, M.; Brandenburg, A.; DeGraan-Weber, N.; Gotta, S.; Shambaugh, J.; Alvarez, M.; Yu, X. C.; Cao, L.; Shao, C.; Mahan, A.; Nanda, H.; Nields, K.; Nightlinger, N.; Barysz, H. M.; Jahn, M.; Niu, B.; Wang, J.; Leo, G.; Sepe, N.; Liu, Y.-H.; Patel, B. A.; Richardson, D.; Wang, Y.; Tizabi, D.; Borisov, O. v.; Lu, Y.; Maynard, E. L.; Gruhler, A.; Haselmann, K. F.; Krogh, T. N.; Sönksen, C. P.; Letarte, S.; Shen, S.; Boggio, K.; Johnson, K.; Ni, W.; Patel, H.; Ripley, D.; Rouse, J. C.; Zhang, Y.; Daniels, C.; Dawdy, A.; Friese, O.; Powers, T. W.; Sperry, J. B.; Woods, J.; Carlson, E.; Sen, K. I.; Skilton, S. J.; Busch, M.; Lund, A.; Stapels, M.; Guo, X.; Heidelberger, S.; Kaluarachchi, H.; McCarthy, S.; Kim, J.; Zhen, J.; Zhou, Y.; Rogstad, S.; Wang, X.; Fang, J.; Chen, W.; Yu, Y. Q.; Hoogerheide, J. G.; Scott, R.; Yuan, H. New Peak Detection Performance Metrics from the MAM Consortium Interlaboratory Study. *Journal of the American Society for Mass Spectrometry* **2021**, *32* (4), 913–928. https://doi.org/10.1021/jasms.0c00415.

(12) Stow, S. M.; Causon, T. J.; Zheng, X.; Kurulugama, R. T.; Mairinger, T.; May, J. C.; Rennie, E. E.; Baker, E. S.; Smith, R. D.; McLean, J. A.; Hann, S.; Fjeldsted, J. C. An Interlaboratory Evaluation of Drift Tube Ion Mobility–Mass Spectrometry Collision Cross Section Measurements. *Analytical Chemistry* **2017**, *89* (17), 9048–9055. https://doi.org/10.1021/acs.analchem.7b01729.

(13) Paglia, G.; Williams, J. P.; Menikarachchi, L.; Thompson, J. W.; Tyldesley-Worster, R.; Halldórsson, S.; Rolfsson, O.; Moseley, A.; Grant, D.; Langridge, J.; Palsson, B. O.; Astarita, G. Ion Mobility Derived Collision Cross Sections to Support Metabolomics Applications. *Analytical Chemistry* **2014**, *86* (8), 3985–3993. https://doi.org/10.1021/ac500405x.

(14) Luo, M. du; Zhou, Z. W.; Zhu, Z. J. The Application of Ion Mobility-Mass Spectrometry in Untargeted Metabolomics: From Separation to Identification. *Journal of Analysis and Testing*. 2020. https://doi.org/10.1007/s41664-020-00133-0.

(15) Mason, E. A.; McDaniel, E. W. *Transport Properties of Ions in Gases*; John Wiley & Sons, Ltd.: New York City, NY, 1988.

(16) Siems, W. F.; Viehland, L. A.; Hill, H. H. Improved Momentum-Transfer Theory for Ion Mobility. 1. Derivation of the Fundamental Equation. *Analytical Chemistry* **2012**, *84* (22), 9782–9791. https://doi.org/10.1021/ac301779s.

(17) Richardson, K.; Langridge, D.; Giles, K. Fundamentals of Travelling Wave Ion Mobility Revisited: I. Smoothly Moving Waves. *International Journal of Mass Spectrometry* **2018**, *428*, 71–80. https://doi.org/10.1016/j.ijms.2018.03.007.

(18) Ruotolo, B. T.; Benesch, J. L. P.; Sandercock, A. M.; Hyung, S. J.; Robinson, C. V. Ion Mobility-Mass Spectrometry Analysis of Large Protein Complexes. *Nature Protocols* **2008**, *3* (7), 1139–1152. https://doi.org/10.1038/nprot.2008.78.

(19) Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T. Collision Cross Sections of Proteins and Their Complexes: A Calibration Framework and Database for Gas-

Phase Structural Biology. *Analytical Chemistry* **2010**, *82* (22), 9557–9565. https://doi.org/10.1021/ac1022953.

(20) Bush, M. F.; Campuzano, I. D. G.; Robinson, C. V. Ion Mobility Mass Spectrometry of Peptide Ions: Effects of Drift Gas and Calibration Strategies. *Analytical Chemistry* **2012**, *84* (16), 7124–7130. https://doi.org/10.1021/ac3014498.

(21) Gabelica, V.; Shvartsburg, A. A.; Afonso, C.; Barran, P.; Benesch, J. L. P.; Bleiholder, C.; Bowers, M. T.; Bilbao, A.; Bush, M. F.; Campbell, J. L.; Campuzano, I. D. G.; Causon, T.; Clowers, B. H.; Creaser, C. S.; de Pauw, E.; Far, J.; Fernandez-Lima, F.; Fjeldsted, J. C.; Giles, K.; Groessl, M.; Hogan, C. J.; Hann, S.; Kim, H. I.; Kurulugama, R. T.; May, J. C.; McLean, J. A.; Pagel, K.; Richardson, K.; Ridgeway, M. E.; Rosu, F.; Sobott, F.; Thalassinos, K.; Valentine, S. J.; Wyttenbach, T. Recommendations for Reporting Ion Mobility Mass Spectrometry Measurements. *Mass Spectrometry Reviews* **2019**, *38* (3), 291–320. https://doi.org/10.1002/mas.21585.

(22) Richardson, K.; Langridge, D.; Dixit, S. M.; Ruotolo, B. T. An Improved Calibration Approach for Traveling Wave Ion Mobility Spectrometry: Robust, High-Precision Collision Cross Sections. *Analytical Chemistry* **2021**, *93* (7), 3542–3550. https://doi.org/10.1021/acs.analchem.0c04948.

(23) Forsythe, J. G.; Petrov, A. S.; Walker, C. A.; Allen, S. J.; Pellissier, J. S.; Bush, M. F.; Hud, N. v.; Fernández, F. M. Collision Cross Section Calibrants for Negative Ion Mode Traveling Wave Ion Mobility-Mass Spectrometry. *Analyst* **2015**, *140* (20), 6853–6861. https://doi.org/10.1039/c5an00946d.

(24) Hines, K. M.; May, J. C.; McLean, J. A.; Xu, L. Evaluation of Collision Cross Section Calibrants for Structural Analysis of Lipids by Traveling Wave Ion Mobility-Mass Spectrometry. *Analytical Chemistry* **2016**, *88* (14), 7329–7336. https://doi.org/10.1021/acs.analchem.6b01728.

(25) Gelb, A. S.; Jarratt, R. E.; Huang, Y.; Dodds, E. D. A Study of Calibrant Selection in Measurement of Carbohydrate and Peptide Ion-Neutral Collision Cross Sections by Traveling Wave Ion Mobility Spectrometry. *Analytical Chemistry* **2014**, *86* (22), 11396–11402. https://doi.org/10.1021/ac503379e.

(26) Deschamps, E.; Schmitz-Afonso, I.; Schaumann, A.; Dé, E.; Loutelier-Bourhis, C.; Alexandre, S.; Afonso, C. Determination of the Collision Cross Sections of Cardiolipins and Phospholipids from Pseudomonas Aeruginosa by Traveling Wave Ion Mobility Spectrometry-Mass Spectrometry Using a Novel Correction Strategy. *Analytical and Bioanalytical Chemistry* **2019**. https://doi.org/10.1007/s00216-019-02194-2.

(27) Ridenour, W. B.; Kliman, M.; McLean, J. A.; Caprioli, R. M. Structural Characterization of Phospholipids and Peptides Directly from Tissue Sections by MALDI Traveling-Wave Ion Mobility-Mass Spectrometry. *Analytical Chemistry* **2010**, *82* (5). https://doi.org/10.1021/ac9026115.

(28) Paglia, G.; Angel, P.; Williams, J. P.; Richardson, K.; Olivos, H. J.; Thompson, J. W.; Menikarachchi, L.; Lai, S.; Walsh, C.; Moseley, A.; Plumb, R. S.; Grant, D. F.; Palsson, B.

O.; Langridge, J.; Geromanos, S.; Astarita, G. Ion Mobility-Derived Collision Cross Section as an Additional Measure for Lipid Fingerprinting and Identification. *Analytical Chemistry* **2015**, *87* (2), 1137–1144. https://doi.org/10.1021/ac503715v.

(29)   Righetti, L.; Dreolin, N.; Celma, A.; McCullagh, M.; Barknowitz, G.; Sancho, J. v.; Dall'Asta, C. Travelling Wave Ion Mobility-Derived Collision Cross Section for Mycotoxins: Investigating Interlaboratory and Interplatform Reproducibility. *Journal of Agricultural and Food Chemistry* **2020**, *68* (39), 10937–10943. https://doi.org/10.1021/acs.jafc.0c04498.

(30)   Hernández-Mesa, M.; D'Atri, V.; Barknowitz, G.; Fanuel, M.; Pezzatti, J.; Dreolin, N.; Ropartz, D.; Monteau, F.; Vigneau, E.; Rudaz, S.; Stead, S.; Rogniaux, H.; Guillarme, D.; Dervilly, G.; le Bizec, B. Interlaboratory and Interplatform Study of Steroids Collision Cross Section by Traveling Wave Ion Mobility Spectrometry. *Analytical Chemistry* **2020**, *92* (7), 5013–5022. https://doi.org/10.1021/acs.analchem.9b05247.

(31)   Righetti, L.; Bergmann, A.; Galaverna, G.; Rolfsson, O.; Paglia, G.; Dall'Asta, C. Ion Mobility-Derived Collision Cross Section Database: Application to Mycotoxin Analysis. *Analytica Chimica Acta* **2018**, *1014*, 50–57. https://doi.org/10.1016/j.aca.2018.01.047.

(32)   Ross, D. H.; Cho, J. H.; Xu, L. Breaking Down Structural Diversity for Comprehensive Prediction of Ion-Neutral Collision Cross Sections. *Analytical Chemistry* **2020**, *92* (6), 4548–4557. https://doi.org/10.1021/acs.analchem.9b05772.

(33)   Picache, J. A.; May, J. C.; McLean, J. A. Crowd-Sourced Chemistry: Considerations for Building a Standardized Database to Improve Omic Analyses. *ACS Omega* **2020**, acsomega.9b03708. https://doi.org/10.1021/acsomega.9b03708.

(34)   Picache, J. A.; Rose, B. S.; Balinski, A.; Leaptrot, K. L.; Sherrod, S. D.; May, J. C.; McLean, J. A. Collision Cross Section Compendium to Annotate and Predict Multi-Omic Compound Identities. *Chemical Science* **2019**, *10* (4), 983–993. https://doi.org/10.1039/C8SC04396E.

(35)   May, J. C.; Morris, C. B.; McLean, J. A. Ion Mobility Collision Cross Section Compendium. *Analytical Chemistry* **2017**, *89* (2), 1032–1044. https://doi.org/10.1021/acs.analchem.6b04905.

(36)   May, J. C.; McLean, J. A. Advanced Multidimensional Separations in Mass Spectrometry: Navigating the Big Data Deluge. *Annual Review of Analytical Chemistry* **2016**, *9* (1), 387–409. https://doi.org/10.1146/annurev-anchem-071015-041734.

(37)   Nichols, C. M.; Dodds, J. N.; Rose, B. S.; Picache, J. A.; Morris, C. B.; Codreanu, S. G.; May, J. C.; Sherrod, S. D.; McLean, J. A. Untargeted Molecular Discovery in Primary Metabolism: Collision Cross Section as a Molecular Descriptor in Ion Mobility-Mass Spectrometry. *Analytical Chemistry* **2018**, *90* (24), 14484–14492. https://doi.org/10.1021/acs.analchem.8b04322.

(38)   Hamid, A. M.; Garimella, S. V. B.; Ibrahim, Y. M.; Deng, L.; Zheng, X.; Webb, I. K.; Anderson, G. A.; Prost, S. A.; Norheim, R. v.; Tolmachev, A. v.; Baker, E. S.; Smith, R. D. Achieving High Resolution Ion Mobility Separations Using Traveling Waves in Compact

Multiturn Structures for Lossless Ion Manipulations. *Analytical Chemistry* **2016**, *88* (18), 8949–8956. https://doi.org/10.1021/acs.analchem.6b01914.

(39) Deng, L.; Ibrahim, Y. M.; Hamid, A. M.; Garimella, S. V. B.; Webb, I. K.; Zheng, X.; Prost, S. A.; Sandoval, J. A.; Norheim, R. v.; Anderson, G. A.; Tolmachev, A. v.; Baker, E. S.; Smith, R. D. Ultra-High Resolution Ion Mobility Separations Utilizing Traveling Waves in a 13 m Serpentine Path Length Structures for Lossless Ion Manipulations Module. *Analytical Chemistry* **2016**, *88* (18), 8957–8964. https://doi.org/10.1021/acs.analchem.6b01915.

(40) May, J. C.; Leaptrot, K. L.; Rose, B. S.; Moser, K. L. W.; Deng, L.; Maxon, L.; DeBord, D.; McLean, J. A. Resolving Power and Collision Cross Section Measurement Accuracy of a Prototype High-Resolution Ion Mobility Platform Incorporating Structures for Lossless Ion Manipulation. *Journal of the American Society for Mass Spectrometry* **2021**, *32* (4), 1126–1137. https://doi.org/10.1021/jasms.1c00056.

(41) Chouinard, C. D.; Nagy, G.; Webb, I. K.; Garimella, S. V. B. B.; Baker, E. S.; Ibrahim, Y. M.; Smith, R. D. Rapid Ion Mobility Separations of Bile Acid Isomers Using Cyclodextrin Adducts and Structures for Lossless Ion Manipulations. *Analytical Chemistry* **2018**, *90* (18), 11086–11091. https://doi.org/10.1021/acs.analchem.8b02990.

(42) Nagy, G.; Attah, I. K.; Garimella, S. V. B.; Tang, K.; Ibrahim, Y. M.; Baker, E. S.; Smith, R. D. Unraveling the Isomeric Heterogeneity of Glycans: Ion Mobility Separations in Structures for Lossless Ion Manipulations. *Chemical Communications* **2018**, 1–4. https://doi.org/10.1039/C8CC06966B.

(43) Wojcik, R.; Webb, I. K.; Deng, L.; Garimella, S. V. B.; Prost, S. A.; Ibrahim, Y. M.; Baker, E. S.; Smith, R. D. Lipid and Glycolipid Isomer Analyses Using Ultra-High Resolution Ion Mobility Spectrometry Separations. *International Journal of Molecular Sciences* **2017**, *18* (1), 1–12. https://doi.org/10.3390/ijms18010183.

(44) Wormwood Moser, K. L.; van Aken, G.; DeBord, D.; Hatcher, N. G.; Maxon, L.; Sherman, M.; Yao, L.; Ekroos, K. High-Defined Quantitative Snapshots of the Ganglioside Lipidome Using High Resolution Ion Mobility SLIM Assisted Shotgun Lipidomics. *Analytica Chimica Acta* **2021**, *1146*, 77–87. https://doi.org/10.1016/j.aca.2020.12.022.

(45) Li, A.; Conant, C. R.; Zheng, X.; Bloodsworth, K. J.; Orton, D. J.; Garimella, S. V. B.; Attah, I. K.; Nagy, G.; Smith, R. D.; Ibrahim, Y. M. Assessing Collision Cross Section Calibration Strategies for Traveling Wave-Based Ion Mobility Separations in Structures for Lossless Ion Manipulations. *Analytical Chemistry* **2020**, *92* (22), 14976–14982. https://doi.org/10.1021/acs.analchem.0c02829.

(46) Lee, J.; Bilbao, A.; Conant, C. R.; Bloodsworth, K. J.; Orton, D. J.; Zhou, M.; Wilson, J. W.; Zheng, X.; Webb, I. K.; Li, A.; Hixson, K. K.; Fjeldsted, J. C.; Ibrahim, Y. M.; Payne, S. H.; Jansson, C.; Smith, R. D.; Metz, T. O. AutoCCS: Automated Collision Cross-Section Calculation Software for Ion Mobility Spectrometry–Mass Spectrometry. *Bioinformatics* **2021**, 2–10. https://doi.org/10.1093/bioinformatics/btab429.

(47) Zheng, X.; Aly, N. A.; Zhou, Y.; Dupuis, K. T.; Bilbao, A.; Paurus, V. L.; Orton, D. J.; Wilson, R.; Payne, S. H.; Smith, R. D.; Baker, E. S. A Structural Examination and Collision Cross Section Database for over 500 Metabolites and Xenobiotics Using Drift Tube Ion Mobility Spectrometry. *Chem. Sci.* **2017**, *8*, 7724–7736. https://doi.org/10.1039/C7SC03464D.

(48) Leaptrot, K. L.; May, J. C.; Dodds, J. N.; McLean, J. A. Ion Mobility Conformational Lipid Atlas for High Confidence Lipidomics. *Nature Communications* **2019**, *10* (1), 985. https://doi.org/10.1038/s41467-019-08897-5.

(49) Hancock, S. E.; Poad, B. L. J.; Batarseh, A.; Abbott, S. K.; Mitchell, T. W. Advances and Unresolved Challenges in the Structural Characterization of Isomeric Lipids. *Analytical Biochemistry* **2017**, *524*, 45–55. https://doi.org/10.1016/j.ab.2016.09.014.

(50) May, J. C.; Goodwin, C. R.; Lareau, N. M.; Leaptrot, K. L.; Morris, C. B.; Kurulugama, R. T.; Mordehai, A.; Klein, C.; Barry, W.; Darland, E.; Overney, G.; Imatani, K.; Stafford, G. C.; Fjeldsted, J. C.; McLean, J. A. Conformational Ordering of Biomolecules in the Gas Phase: Nitrogen Collision Cross Sections Measured on a Prototype High Resolution Drift Tube Ion Mobility-Mass Spectrometer. *Analytical Chemistry* **2014**, *86* (4), 2107–2116. https://doi.org/10.1021/ac4038448.

(51) May, J. C.; Dodds, J. N.; Kurulugama, R. T.; Stafford, G. C.; Fjeldsted, J. C.; McLean, J. A. Broadscale Resolving Power Performance of a High Precision Uniform Field Ion Mobility-Mass Spectrometer. *The Analyst* **2015**, *140* (20), 6824–6833. https://doi.org/10.1039/C5AN00923E.

(52) Bilbao, A.; Gibbons, B. C.; Stow, S. M.; Kyle, J. E.; Bloodsworth, K. J.; Payne, S. H.; Smith, R. D.; Ibrahim, Y. M.; Baker, E. S.; Fjeldsted, J. C. A Preprocessing Tool for Enhanced Ion Mobility–Mass Spectrometry-Based Omics Workflows. *Journal of Proteome Research* **2021**. https://doi.org/10.1021/acs.jproteome.1c00425.

(53) Rose, B. S.; Leaptrot, K. L.; Harris, R. A.; Sherrod, S. D.; May, J. C.; McLean, J. A. High Confidence Shotgun Lipidomics Using Structurally Selective Ion Mobility-Mass Spectrometry. In *Mass Spectrometry-Based Lipidomics: Methods and Protocols*; Hsu, F.-F., Ed.; Springer US: New York, NY, 2021; pp 11–37. https://doi.org/10.1007/978-1-0716-1410-5_2.

(54) McLean, J. A. The Mass-Mobility Correlation Redux: The Conformational Landscape of Anhydrous Biomolecules. *Journal of the American Society for Mass Spectrometry* **2009**, *20* (10), 1775–1781. https://doi.org/10.1016/j.jasms.2009.06.016.

(55) Chai, M.; Young, M. N.; Liu, F. C.; Bleiholder, C. A Transferable, Sample-Independent Calibration Procedure for Trapped Ion Mobility Spectrometry (TIMS). *Analytical Chemistry* **2018**, *90* (15), 9040–9047. https://doi.org/10.1021/acs.analchem.8b01326.

(56) McLean Research Group. Unified CCS Compendium https://mcleanresearchgroup.shinyapps.io/CCS-Compendium/.

CHAPTER 6

CONCLUSIONS AND OUTLOOK

6.1   Summary

Untargeted metabolomics and lipidomics seek to comprehensively profile small molecules and lipids found in complex biological samples to gain a better understanding of the pathways and networks that govern large scale biological processes.[1,2] Largely due to its sensitivity and specificity, mass spectrometry (MS) has become an increasingly prevalent analytical strategy for untargeted omics studies.[3–5] A truly global analysis of the highly complex biochemical activity of these systems remains elusive though, largely as a result of challenges in confident annotation of untargeted data.[6,7] Even with high-resolution mass spectrometers, comprehensive identification by accurate mass alone is unattainable due to the prevalence of isomeric species as well as the limited number of commercially available analytical standards, especially in lipidomics.[8] In a global MS study of a complex biological sample, one feature can be assigned hundreds of tentative identifications which are indistinguishable without additional information. Due to the predictable fragmentation patterns of many biochemical classes, MS/MS is often used to aid in the identification of lipid species, increasing its confidence to a "probable structure". However, even MS/MS approaches are incapable of resolving all species in biological matrices.[9,10] Thus the combination and integration of several analytical techniques into a single workflow is often necessary to expand coverage and increase confidence of lipidomic annotations in order to facilitate meaningful biological conclusions.[11,12]

120

Ion mobility (IM) has gained widespread use in recent years for structural analysis.[13–15] This gas phase separation technique separates ions based on their shape and size, and its millisecond timescale nests well between the timescales of liquid chromatography (LC) and MS. In addition to a complementary dimension of separation, IM can provide a structurally selective metric for compound annotation, the collision cross section (CCS).[11,16] CCS values are highly reproducible and can differentiate many isomeric species that have variation in their conformation despite their identical mass. They also manifest class-specific conformational trends that can be utilized in identification or filtering workflows.[17,18] Because of these advantages, many recent efforts have focused on the generation of class-specific CCS databases to aid in compound identification.[19–21] Such databases have seen success in annotating many datasets, but have been relatively narrow in scope.[16,22] This work aimed to develop tools and strategies to enable the broader adoption of CCS as an additional molecular descriptor in support of high-confidence untargeted molecular annotation.

In Chapter 2, we explored the potential of pooling existing CCS databases into one self-consistent repository to facilitate the broader adoption of IM to current molecular identification workflows. A collaborative library of multi-omic CCS values was curated and standardized to provide resources and guidelines for the IM community. This dataset, the Unified CCS Compendium, not only allows empirical matching to a large set of experimental CCS values, but also unlocks the predictive potential of statistical modeling, both of which provide tools with which to address the challenge of identification confidence. Some of the predictive applications of the Compendium are explored in Chapter 3, where we developed an informatic pipeline to increase identification confidence in lipidomic identifications without the need for exact matching from an experimental or in silica MS library. This integrated high-confidence workflow provides support for the

application of IM to untargeted studies, and when combined with LC and tandem MS data, it can further increase coverage in untargeted studies, thus increasing confidence in the resulting biological interpretations

While these database matching and predictive informatics strategies provide support for the application of IM to identification workflows, there exist numerous cases for which the resolution of commercially available drift tube IM instrumentation is insufficient for separation of isomers. Because biological function is dictated by structure and isomers are prolific in nature, it is of great importance to pursue techniques capable of distinguishing isomeric species. SLIM technology, based on scalable printed circuit board architectures, allow for extended pathlength high resolution IM separations (HRIM).[27–29] In Chapter 4, we surveyed the resolving power capabilities and CCS calibration strategies for a prototype HRIM system. The fundamental calibration study in Chapter 4 was used as a foundation for Chapter 5, where it was applied to the analysis of a broad range of lipid species. The consensus values from the Compendium allowed for the evaluation of calibration accuracy of these species and allowed for subclass-specific optimization of the calibration protocol. The efforts described in Chapters 4 and 5 provide a comprehensive evaluation of both HRIM instrumental performance and CCS calibration to establish protocols for highly selective and accurate IM measurements. These more structurally specific CCS measurements not only enable the analysis of previously unresolved lipid species but will also ultimately improve the performance of IM-based informatics for untargeted studies.

Overall, the guidelines, recommendations, and strategies developed throughout this work enable the adoption and integration of IM and CCS into untargeted workflows for molecular identification and characterization. The added structural resolution from the integration of IM into -omics studies will facilitate a broader annotation of relevant pathway components, including isomeric species.

This global profile will enable a more comprehensive understanding of molecular dysregulation in large-scale biological processes.

6.2   Future directions

6.2.1   Compendium expansion and updates

The Unified CCS Compendium, described in Chapter 2, was developed as a collaborative resource to which data can be added from the community. Since the initial publication of this work in 2019, over 500 new CCS values have been added to the Compendium thanks to the efforts of multiple research groups around the world. The expansion has included measurements from previously unrepresented compound classes of interest such as steroid conjugates[23] and perfluoroalkyl substances (PFAS)[24,25]. The Compendium has been widely used in numerous works for compound annotation,[26,27] prediction,[28,29] and method validation.[30,31] Continued work toward the expansion and maintenance of this shared resource will enable further refinement of the predictive tools built with the Compendium as a training set, as well broader adoption of CCS for IM workflows as the field continues to grow into new application areas.

Currently, guidelines are in place to ensure only highly vetted data is added to the database to maintain the quality of the consensus values. However, working toward data volume in addition to data quality will be advantageous in the further advancement of this tool. The implementation of a confidence system would allow the addition of published values that pre-date the Compendium and its standardized guidelines. Data quality could be ranked as a measure of metrics such as relative standard deviation, replicate number, or bias from gold-standard values. The quality value could then be used to indicate confidence in subsets of CCS values, allowing users to filter by quality to apply appropriate considerations in their downstream use of the data. Importantly, the largescale addition of data enabled by this system would enable more refined and

advanced predictive models, leading to more confident annotation and prediction of CCS values in untargeted workflows.

6.2.2   CCS-based filtering for untargeted lipidomics

The CCS filtering informatics workflow described in Chapter 3 was applied in a proof-of-concept lipidomics study to filter tentative identifications, which resulted in candidate lists of 50% fewer tentative identifications on average. Since the initial study, the workflow has also been applied to lipidomics studies from more complex biological systems. In one pilot study, a global lipidomics experiment was performed on murine plasma from a knockout model of a transmembrane lipid transfer protein.[32] The resulting molecular feature lists were subjected to the IM filtering workflow, resulting in 289 features with increased identification confidence, of which 105 were significantly altered in the knockout model as compared to the wild type. Notably, eicosanoid species were significantly altered, suggesting that eicosanoid production may be an important link to the expressed phenotype of the knockout animals. The subclass-specific information provided by the CCS filtering approach was integral to findings that are currently being interrogated in follow-up assays.

Continued improvements and additions to the IM filtering workflow will be integral to its adoption and success in the broader IM community. One example of innovation already underway is the addition of a complementary filter based on Kendrick mass defect (KMD), which has been used previously in lipid class identification.[33] In this approach, exact mass values from the Compendium were used to calculate KMD and create class-specific linear regression models similar to the CCS-based models in the Compendium (Figure 6.1(A)). Features that do not fall within the predictive interval can then be filtered from further analysis (Figure 6.1(B)). In a discovery study, this
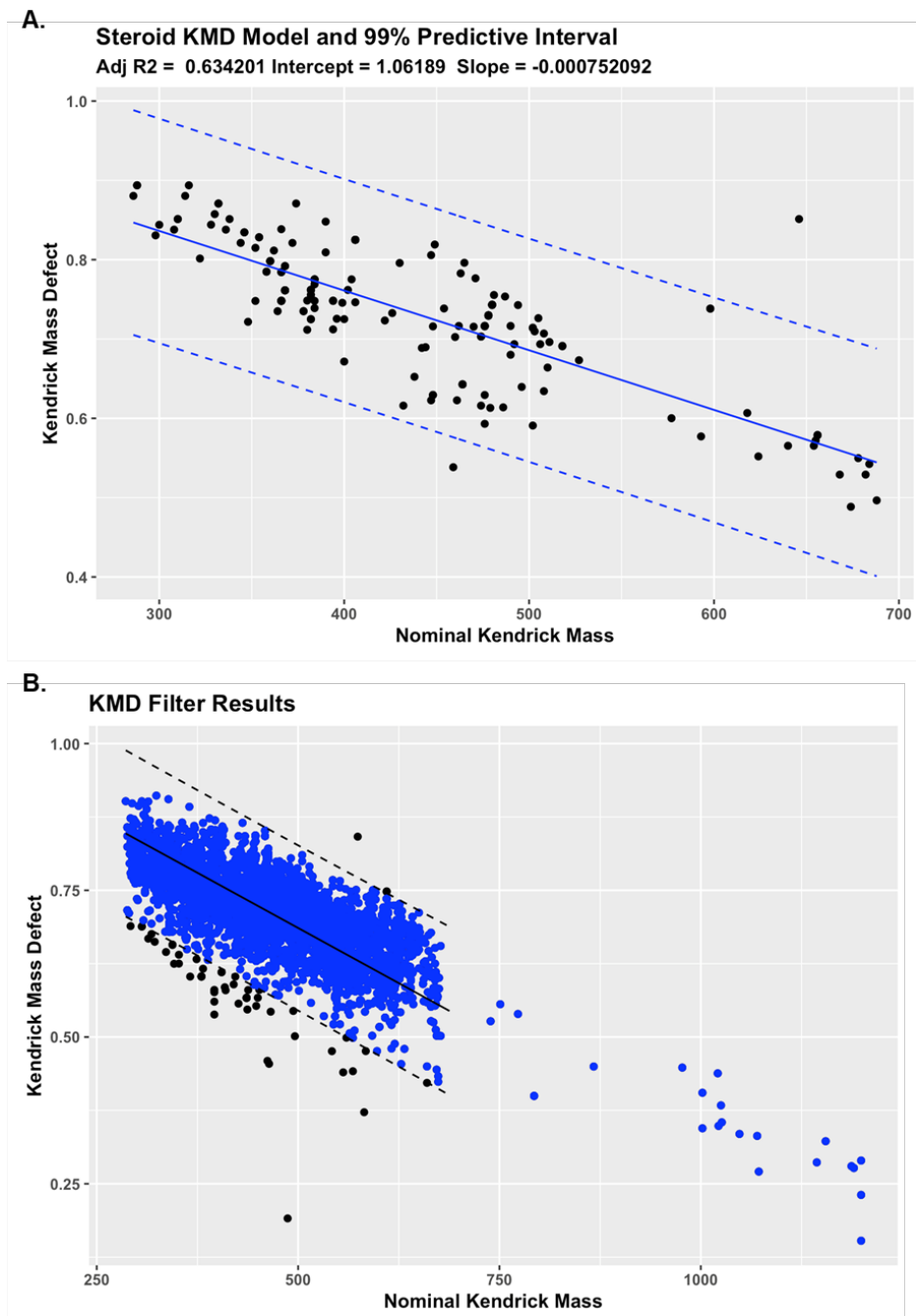
Figure 6.1 Kendrick mass defect filtering. (A) KMD model mean (solid blue line) and 99% predictive interval (dashed line) generated based on the exact mass of steroids and derivatives in the Unified CCS Compendium (represented as black markers). (B) KMD filter results of features in an untargeted study, where 57 features (black) were rejected as they fell outside the predictive interval (dashed line).

combined workflow was used to elucidate phase II anabolic androgenic steroid metabolites in urine and resulted in the rejection of 834 tentatively identified metabolites from IM filtering, with an additional 57 rejected from KMD filtering.

This combined filtering approach highlights the value of integrating complementary analytical approaches for higher confidence annotations in untargeted studies. Further automated informatic integration of other dimensions such as LC and tandem MS, in addition to IM, will be a key focus in the future directions of this project.

6.2.3   HRIM derived CCS values for lipid characterization

In Chapters 5 and 6, we benchmarked a HRIM platform and developed recommendations and calibration strategies specific to analyte mass range and chemical class. Current work is focused on the application of these protocols to lipid extract characterization. The high resolving power afforded by the HRIM platform allows for the separation and elucidation of previously unseen isomeric and isobaric lipid species. Figure 6.2 shows an example of a HRIM 2D spectrum of a phosphatidylcholine (PC) total extract. In this spectrum, there are clearly separated features in the arrival time dimension that do not separate in mass (e.g. *m/z* 782.5672). By using the characteristic trends in acyl chain length and degree of unsaturation described by Leaptrot, et. al.,[14] these species can be assigned tentative identifications.

Mapping of these lipid species across various biological extracts, as well as calibration of their CCS values will facilitate the construction of HRIM databases, supporting future work in untargeted lipidomic applications of HRIM. These more structurally specific CCS measurements will not only enable the analysis of previously unresolved lipid species but will also ultimately improve the performance of IM-based informatics for lipidomics.
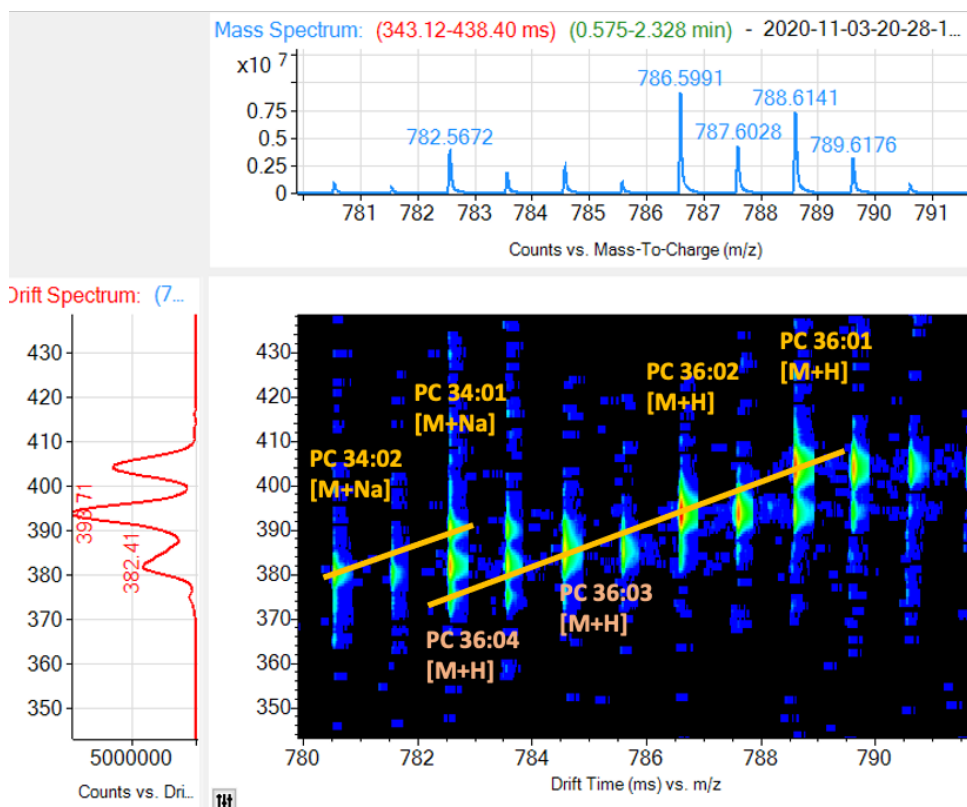
Figure 6.2 HRIM 2D spectrum of a phosphatidylcholine extract showing tentative identifications of spectral features based on accurate mass. The feature at *m/z* 782.5672 has three arrival time features, which can be annotated in part by their characteristic acyl chain trend lines (orange lines).

6.3   Concluding remarks and outlook

The grand challenge of untargeted, discovery-based metabolomics and lipidomics remains rooted in the molecular annotation bottleneck. Though great strides have been made in recent years, the biochemical complexity of biological matrices continues to preclude a truly global profile of the metabolites and lipids in any given sample. Ion mobility is uniquely suited to address this challenge, as it fits within the analytical timescale of an LC-MS experiment, it provides an additional dimension of separation with which to further deconvolute molecular complexity, and it also can be used to obtain an additional molecular descriptor for identification, the CCS. However, there is still work to be done before the full potential of this powerful approach can be realized. Because CCS values must be experimentally measured, database construction is arduous and limited by the availability of standards. Therefore, interlaboratory standardization and subsequent pooling of databases, as well as standards-free approaches using machine learning and other predictive strategies will be instrumental in advancing the field beyond targeted assays. Further, the addition of the IM dimension adds a layer of data complexity, which, as of this writing, still incurs a significant amount of manual analysis time or development of custom analysis tools. Broader adoption of IM outside of specialized analytical research will require significant work toward the further advancement of informatic tools and software for data handling and analysis. The work described in this dissertation seeks to lay a fundamental foundation upon which new innovative strategies can be built to enable the deployment of IM for high confidence global untargeted metabolomics.

6.4   Acknowledgements

I would like to thank Prof. Katrina Leaptrot, Mr. David Koomen, and Ms. Allison Reardon for their efforts in the continuation and advancement of the projects discussed in this chapter. This

work was supported in part using the resources of the Center for Innovative Technology (CIT) at Vanderbilt University.

6.5   References

(1)   Patti, G. J.; Yanes, O.; Siuzdak, G. Metabolomics: The Apogee of the Omics Trilogy. *Nature Reviews Molecular Cell Biology* **2012**, *13* (4), 263–269. https://doi.org/10.1038/nrm3314.

(2)   May, J. C.; Gant-Branum, R. L.; McLean, J. A. Targeting the Untargeted in Molecular Phenomics with Structurally-Selective Ion Mobility-Mass Spectrometry. *Current Opinion in Biotechnology* **2016**, *39*, 192–197. https://doi.org/10.1016/j.copbio.2016.04.013.

(3)   Wenk, M. R. The Emerging Field of Lipidomics. *Nature Reviews Drug Discovery* **2005**, *4* (7), 594–610. https://doi.org/10.1038/nrd1776.

(4)   Emília, A.; Rolim, H.; Henrique-araújo, R.; Gomes, E.; Karoline, F.; Alves, D. A.; Gonzaga, L. Lipidomics in the Study of Lipid Metabolism: Current Perspectives in the Omic Sciences. *Gene* **2015**, *554* (2), 131–139. https://doi.org/10.1016/j.gene.2014.10.039.

(5)   Navas-Iglesias, N.; Carrasco-Pancorbo, A.; Cuadros-Rodríguez, L. From Lipids Analysis towards Lipidomics, a New Challenge for the Analytical Chemistry of the 21st Century. Part II: Analytical Lipidomics. *TrAC Trends in Analytical Chemistry* **2009**, *28* (4), 393–403. https://doi.org/10.1016/j.trac.2008.12.004.

(6)   Blaženović, I.; Kind, T.; Ji, J.; Fiehn, O. Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics. *Metabolites* **2018**, *8* (2), 31. https://doi.org/10.3390/metabo8020031.

(7)   Schrimpe-Rutledge, A. C.; Codreanu, S. G.; Sherrod, S. D.; McLean, J. A. Untargeted Metabolomics Strategies—Challenges and Emerging Directions. *Journal of the American Society for Mass Spectrometry* **2016**, *27* (12), 1897–1905. https://doi.org/10.1007/s13361-016-1469-y.

(8)   Rustam, Y. H.; Reid, G. E. Analytical Challenges and Recent Advances in Mass Spectrometry Based Lipidomics. *Analytical Chemistry* **2018**, *90* (1), 374–397. https://doi.org/10.1021/acs.analchem.7b04836.

(9)   Ekroos, K.; Chernushevich, I. V.; Simons, K.; Shevchenko, A. Quantitative Profiling of Phospholipids by Multiple Precursor Ion Scanning on a Hybrid Quadrupole Time-of-Flight Mass Spectrometer. *Analytical Chemistry* **2002**, *74* (5), 941–949. https://doi.org/10.1021/ac015655c.

(10)   Koelmel, J. P.; Kroeger, N. M.; Gill, E. L.; Ulmer, C. Z.; Bowden, J. A.; Patterson, R. E.; Yost, R. A.; Garrett, T. J. Expanding Lipidome Coverage Using LC-MS/MS Data-Dependent Acquisition with Automated Exclusion List Generation. *Journal of the American*

Society for Mass Spectrometry **2017**, *28* (5), 908–917. https://doi.org/10.1007/s13361-017-1608-0.

(11)   Kyle, J. E.; Zhang, X.; Weitz, K. K.; Monroe, M. E.; Ibrahim, Y. M.; Moore, R. J.; Cha, J.; Sun, X.; Lovelace, E. S.; Wagoner, J.; Polyak, S. J.; Metz, T. O.; Dey, S. K.; Smith, R. D.; Burnum-Johnson, K. E.; Baker, E. S. Uncovering Biologically Significant Lipid Isomers with Liquid Chromatography, Ion Mobility Spectrometry and Mass Spectrometry. *The Analyst* **2016**, *141* (5), 1649–1659. https://doi.org/10.1039/C5AN02062J.

(12)   Harris, R. A.; Leaptrot, K. L.; May, J. C.; McLean, J. A. New Frontiers in Lipidomics Analyses Using Structurally Selective Ion Mobility-Mass Spectrometry. *TrAC Trends in Analytical Chemistry* **2019**, *116*, 316–323. https://doi.org/10.1016/j.trac.2019.03.031.

(13)   Blaženović, I.; Shen, T.; Mehta, S. S.; Kind, T.; Ji, J.; Piparo, M.; Cacciola, F.; Mondello, L.; Fiehn, O. Increasing Compound Identification Rates in Untargeted Lipidomics Research with Liquid Chromatography Drift Time–Ion Mobility Mass Spectrometry. *Analytical Chemistry* **2018**, *90* (18), 10758–10764. https://doi.org/10.1021/acs.analchem.8b01527.

(14)   Leaptrot, K. L.; May, J. C.; Dodds, J. N.; McLean, J. A. Ion Mobility Conformational Lipid Atlas for High Confidence Lipidomics. *Nature Communications* **2019**, *10* (1), 985. https://doi.org/10.1038/s41467-019-08897-5.

(15)   Tu, J.; Zhou, Z.; Li, T.; Zhu, Z. J. The Emerging Role of Ion Mobility-Mass Spectrometry in Lipidomics to Facilitate Lipid Separation and Identification. *TrAC - Trends in Analytical Chemistry* **2019**, *116*, 332–339. https://doi.org/10.1016/j.trac.2019.03.017.

(16)   Nichols, C. M.; Dodds, J. N.; Rose, B. S.; Picache, J. A.; Morris, C. B.; Codreanu, S. G.; May, J. C.; Sherrod, S. D.; McLean, J. A. Untargeted Molecular Discovery in Primary Metabolism: Collision Cross Section as a Molecular Descriptor in Ion Mobility-Mass Spectrometry. *Analytical Chemistry* **2018**, *90* (24), 14484–14492. https://doi.org/10.1021/acs.analchem.8b04322.

(17)   Fenn, L. S.; McLean, J. A. Biomolecular Structural Separations by Ion Mobility–Mass Spectrometry. *Analytical and Bioanalytical Chemistry* **2008**, *391* (3), 905–909. https://doi.org/10.1007/s00216-008-1951-x.

(18)   May, J. C.; Goodwin, C. R.; Lareau, N. M.; Leaptrot, K. L.; Morris, C. B.; Kurulugama, R. T.; Mordehai, A.; Klein, C.; Barry, W.; Darland, E.; Overney, G.; Imatani, K.; Stafford, G. C.; Fjeldsted, J. C.; McLean, J. A. Conformational Ordering of Biomolecules in the Gas Phase: Nitrogen Collision Cross Sections Measured on a Prototype High Resolution Drift Tube Ion Mobility-Mass Spectrometer. *Analytical Chemistry* **2014**, *86* (4), 2107–2116. https://doi.org/10.1021/ac4038448.

(19)   Struwe, W. B.; Pagel, K.; Benesch, J. L. P.; Harvey, D. J.; Campbell, M. P. GlycoMob: An Ion Mobility-Mass Spectrometry Collision Cross Section Database for Glycomics. *Glycoconjugate Journal* **2016**, *33* (3), 399–404. https://doi.org/10.1007/s10719-015-9613-7.

(20) Chouinard, C. D.; Beekman, C. R.; Kemperman, R. H. J.; King, H. M.; Yost, R. A. Ion Mobility-Mass Spectrometry Separation of Steroid Structural Isomers and Epimers. *International Journal for Ion Mobility Spectrometry* **2017**, *20* (1–2), 31–39. https://doi.org/10.1007/s12127-016-0213-4.

(21) Zheng, X.; Smith, F. B.; Aly, N. A.; Cai, J.; Smith, R. D.; Patterson, A. D.; Baker, E. S. Evaluating the Structural Complexity of Isomeric Bile Acids with Ion Mobility Spectrometry. *Analytical and Bioanalytical Chemistry* **2019**. https://doi.org/10.1007/s00216-019-01869-0.

(22) Zheng, X.; Aly, N. A.; Zhou, Y.; Dupuis, K. T.; Bilbao, A.; Paurus, V. L.; Orton, D. J.; Wilson, R.; Payne, S. H.; Smith, R. D.; Baker, E. S. A Structural Examination and Collision Cross Section Database for over 500 Metabolites and Xenobiotics Using Drift Tube Ion Mobility Spectrometry. *Chem. Sci.* **2017**, *8*, 7724–7736. https://doi.org/10.1039/C7SC03464D.

(23) Davis, D. E.; Leaptrot, K. L.; Koomen, D. C.; May, J. C.; Cavalcanti, G. D. A.; Padilha, M. C.; Pereira, H. M. G.; McLean, J. A. Multidimensional Separations of Intact Phase II Steroid Metabolites Utilizing LC–Ion Mobility–HRMS. *Analytical Chemistry* **2021**, acs.analchem.1c02163. https://doi.org/10.1021/acs.analchem.1c02163.

(24) Dodds, J. N.; Hopkins, Z. R.; Knappe, D. R. U.; Baker, E. S. Rapid Characterization of Per- and Polyfluoroalkyl Substances (PFAS) by Ion Mobility Spectrometry–Mass Spectrometry (IMS-MS). *Analytical Chemistry* **2020**, acs.analchem.9b05364. https://doi.org/10.1021/acs.analchem.9b05364.

(25) Belova, L.; Caballero-Casero, N.; van Nuijs, A. L. N.; Covaci, A. Ion Mobility-High-Resolution Mass Spectrometry (IM-HRMS) for the Analysis of Contaminants of Emerging Concern (CECs): Database Compilation and Application to Urine Samples. *Analytical Chemistry* **2021**, *93* (16), 6428–6436. https://doi.org/10.1021/acs.analchem.1c00142.

(26) Dodds, J. N.; Wang, L.; Patti, G. J.; Baker, E. S. Combining Isotopologue Work Fl Ows and Simultaneous Multidimensional Separations to Detect, Identify, and Validate Metabolites in Untargeted Analyses. *Analytical Chemistry* **2022**, *94* (5), 2527–2535. https://doi.org/10.1021/acs.analchem.1c04430.

(27) Feuerstein, M. L.; Kurulugama, R. T.; Hann, S.; Causon, T. Novel Acquisition Strategies for Metabolomics Using Drift Tube Ion Mobility-Quadrupole Resolved All Ions Time-of-Flight Mass Spectrometry (IM-QRAI-TOFMS). *Analytica Chimica Acta* **2021**, *1163*, 338508. https://doi.org/10.1016/j.aca.2021.338508.

(28) Plante, P.-L.; Francovic-Fontaine, É.; May, J. C.; McLean, J. A.; Baker, E. S.; Laviolette, F.; Marchand, M.; Corbeil, J. Predicting Ion Mobility Collision Cross-Sections Using a Deep Neural Network: DeepCCS. *Analytical Chemistry* **2019**, *91* (8), 5191–5199. https://doi.org/10.1021/acs.analchem.8b05821.

(29) Picache, J. A.; May, J. C.; McLean, J. A. Chemical Class Prediction of Unknown Biomolecules Using Ion Mobility-Mass Spectrometry and Machine Learning: Supervised

Inference of Feature Taxonomy from Ensemble Randomization. *Analytical Chemistry* **2020**, *92* (15), 10759–10767. https://doi.org/10.1021/acs.analchem.0c02137.

(30) Lee, J.; Bilbao, A.; Conant, C. R.; Bloodsworth, K. J.; Orton, D. J.; Zhou, M.; Wilson, J. W.; Zheng, X.; Webb, I. K.; Li, A.; Hixson, K. K.; Fjeldsted, J. C.; Ibrahim, Y. M.; Payne, S. H.; Jansson, C.; Smith, R. D.; Metz, T. O. AutoCCS: Automated Collision Cross-Section Calculation Software for Ion Mobility Spectrometry–Mass Spectrometry. *Bioinformatics* **2021**, 2–10. https://doi.org/10.1093/bioinformatics/btab429.

(31) Richardson, K.; Langridge, D.; Dixit, S. M.; Ruotolo, B. T. An Improved Calibration Approach for Traveling Wave Ion Mobility Spectrometry: Robust, High-Precision Collision Cross Sections. *Analytical Chemistry* **2021**, *93* (7), 3542–3550. https://doi.org/10.1021/acs.analchem.0c04948.

(32) Norris, A.; Graham, T.; Stafford, J.; Zhu, L. Exploring the Role of ATP10A in Diet-induced Obesity, Insulin Resistance, and Type 2 Diabetes. *The FASEB Journal* **2021**, *35* (S1). https://doi.org/10.1096/fasebj.2021.35.s1.00248.

(33) Lerno, L. A.; German, J. B.; Lebrilla, C. B. Method for the Identification of Lipid Classes Based on Referenced Kendrick Mass Analysis. *Analytical Chemistry* **2010**, *82* (10), 4236–4245. https://doi.org/10.1021/ac100556g.

APPENDIX A

REFERENCE OF ADAPTION FOR CHAPTERS

Chapter 1.   Charles M. Nichols, James N. Dodds, **Bailey S. Rose**, Jaqueline A. Picache, Caleb B. Morris, Simona G. Codreanu, Jody C. May, Stacy D. Sherrod, John A. McLean, "Untargeted Discovery in Primary Metabolism: Collision Cross Section as a Molecular Descriptor in Ion Mobility-Mass Spectrometry," Analytical Chemistry, 2018, 90 (24), 14484-14492.

Chapter 2.   Jaqueline A. Picache, **Bailey S. Rose**, Andrzej Balinski, Katrina L. Leaptrot, Stacy D. Sherrod, Jody C. May, John A. McLean, "Collision Cross Section Compendium to Annotate and Predict Multi-omic Compound Identities," Chemical Science, 2019, 10 (4), 983-993.

Chapter 3.   **Bailey S. Rose**, Jody C. May, Jaqueline A. Picache, Simona G. Codreanu, Stacy D. Sherrod, John A. McLean, "Improving Confidence in Lipidomic Annotations by Incorporating Empirical Ion Mobility Regression Analysis and Chemical Class Prediction," Bioinformatics, *Submitted.*

Chapter 4.   Jody C. May, Katrina L. Leaptrot, **Bailey S. Rose**, Kelly L. W. Moser, Liulin Deng, Daniel DeBord, John A. McLean, "Resolving Power and Collision Cross Section Measurement Accuracy of a Prototype High-Resolution Ion Mobility Platform Incorporating Structures for Lossless Ion Manipulation," Journal of the American Society for Mass Spectrometry, 2021, 32 (4), 1126-1137.

Chapter 5.   **Bailey S. Rose**, Allison R. Reardon, Jody C. May, John A. McLean, "Strategies for Lipid Collision Cross Section Calibration of SLIM-based High Resolution Ion Mobility Measurements," Journal of the American Society for Mass Spectrometry. *Submitted*.

APPENDIX B

SUPPLEMENTARY MATERIALS FOR CHAPTER 1

B1   MSMLS Study Sample Preparation and Acquisition Parameters

B1.1   MSMLS Sample Preparation

The Mass Spectrometry Metabolite Library of Standards (MSMLS, IROA technologies) is supplied as dried standards distributed across seven 96-well plates (Sigma-Aldrich; St. Louis, MO) and each well contains 5 μg of analytical standard. All solvents used to reconstitute the analytes prior to analysis, including water (H_2O), methanol (MeOH), acetonitrile (ACN), isopropanol (IPA), and chloroform (CHCl3) were Optima LC-MS grade purchased from Fisher Scientific (Fair Lawn, NJ). Stock solutions of the hydrophilic standards were prepared by adding 100 μL 1:9 (MeOH: H_2O) to each well prior to mixing on a waving rotator for 5 minutes. The stocks were then distributed in 20 μL aliquots throughout five 96-well plates (Waters part no. 186005837). Stock plates that were not immediately analyzed were capped and transferred to -80 °C for storage. Working solutions of the hydrophilic standards were prepared by adding 80 μL of water with 0.1% formic acid to the 20 μL stock solutions, sealed with plate covers (Waters part no. 186006332), and subsequently mixed on a waving rotator for 5 minutes. The hydrophobic analyte set was prepared similarly, where stock solutions were prepared with 100 μL 2:1:1:0.3 (MeOH: CHCl3: IPA: H_2O), and distributed in 20 μL aliquots throughout five 96-well plates. Working solutions

were prepared by adding 80 µL of 1:1 (MeOH: IPA). The concentration of the working solutions used for IM-MS analysis was 10 µg/mL.

B1.2   Collision Cross Section Measurements

CCS measurements for the MSMLS were obtained on a commercially available drift tube ion mobility-mass spectrometer (DTIMS, Agilent 6560) operated with nitrogen gas (3.95 Torr) at room temperature (~25 °C) and using both single-field and stepped-field approaches. The single-field CCS values reported here were measured in triplicate, while the stepped-field values were collected in a single acquisition. Stepped-field measurements were acquired using an automated flow injection analysis (FIA) stepped-field approach. Briefly, the FIA method was performed with a liquid chromatography system (Agilent 1290) modified with a 100 µL sample loop (Agilent part no. G4226-87303) coupled to an IM-MS (6560, Agilent). 20 µL of the working solution was injected from the 96-well plate with 1:1 (water: isopropanol) as the carrier solvent. For traditional stepped-field CCS determination by FIA, following a 0.5 s delay, an entrance potential was stepped every 0.5 min. in increments of 100 V from 1074 V to 1674 V; the first step from 1074 to 1174 occurred at 1.0 minute rather than 0.5 min. For single-field CCS determination using FIA, 4 µL of sample was injected into the carrier solvent at a flow of 800 µL/min. Data was collected for 0.5 min, followed by a 0.4 min postrun flushing cycle. A drift tube entrance voltage of 1574 V was used. DTIMS exhibits a linear relationship between drift time and CCS, and single-field CCS values are determined by first measuring the drift time of ions (ESI Low Concentration Tuning Mix, Agilent) with a known CCS. The calibrant ions were infused for 0.5 minutes while IM-MS spectra are collected; calibration experiments were preformed intermittently to ensure instrument stability. IM-MS Browser (Agilent, B.08) was used to plot the linear regression of the calibration ions for single field experiments, and the instrumental coefficients β and T_{fix}, were extracted and

were prepared by adding 80 µL of 1:1 (MeOH: IPA). The concentration of the working solutions used for IM-MS analysis was 10 µg/mL.

B1.2   Collision Cross Section Measurements

CCS measurements for the MSMLS were obtained on a commercially available drift tube ion mobility-mass spectrometer (DTIMS, Agilent 6560) operated with nitrogen gas (3.95 Torr) at room temperature (~25 °C) and using both single-field and stepped-field approaches. The single-field CCS values reported here were measured in triplicate, while the stepped-field values were collected in a single acquisition. Stepped-field measurements were acquired using an automated flow injection analysis (FIA) stepped-field approach. Briefly, the FIA method was performed with a liquid chromatography system (Agilent 1290) modified with a 100 µL sample loop (Agilent part no. G4226-87303) coupled to an IM-MS (6560, Agilent). 20 µL of the working solution was injected from the 96-well plate with 1:1 (water: isopropanol) as the carrier solvent. For traditional stepped-field CCS determination by FIA, following a 0.5 s delay, an entrance potential was stepped every 0.5 min. in increments of 100 V from 1074 V to 1674 V; the first step from 1074 to 1174 occurred at 1.0 minute rather than 0.5 min. For single-field CCS determination using FIA, 4 µL of sample was injected into the carrier solvent at a flow of 800 µL/min. Data was collected for 0.5 min, followed by a 0.4 min postrun flushing cycle. A drift tube entrance voltage of 1574 V was used. DTIMS exhibits a linear relationship between drift time and CCS, and single-field CCS values are determined by first measuring the drift time of ions (ESI Low Concentration Tuning Mix, Agilent) with a known CCS. The calibrant ions were infused for 0.5 minutes while IM-MS spectra are collected; calibration experiments were preformed intermittently to ensure instrument stability. IM-MS Browser (Agilent, B.08) was used to plot the linear regression of the calibration ions for single field experiments, and the instrumental coefficients β and T_{fix}, were extracted and

used to convert raw ion drift times to CCS. The resulting single- and stepped-field CCS library can be found in the Supporting Information.

B1.3 IM-MS Source and Drift Cell Conditions

To obtain high coverage of analytes within the MSMLS, both electrospray (Agilent Jet Stream, AJS) and chemical ionization (APCI) sources were used. The majority of the samples collected with the AJS in both ion modes were measured using the following conditions: gas temperature, 250 °C; drying gas, 8 L/min; nebulizer, 60 psig; sheath gas temperature, 300 °C; sheath gas flow, 11 L/min; capillary voltage (VCap), 3500 V; nozzle voltage, 800 V; fragmentor, 340 V; octopole 1 RF Vpp, 750 V. All metabolites were first investigated using the AJS source; those which were not observed in either ion polarity were subsequently investigated using the APCI source under the following conditions: gas temperature, 250 °C; vaporizer, 200 °C; drying gas, 7 L/min; nebulizer, 30 psig; VCap, 3800 V; corona, 5 μA; fragmentor, 350 V; octopole 1 RF Vpp 750 V. Some of the low *m/z* ions (typically ≤ 200 Da) exhibited metastable ion dissociation in the DTIMS which resulted in uncorrelated mobilities. In these cases, we increased the fragmentor potential to > 350 V and decreased the Trap Funnel RF to ≤ 80 V_{pp} to culminate the ion signal into a single IM distribution. The IM-MS settings for the CCS values reported herein are as follows: 0.9 frames/s; 18 IM transients/frame; 60 ms max drift time; 600 TOF transients/IM transient; 20000 μs trap fill time; 180 μs trap release time; drift tube exit voltage, 224 V; rear funnel entrance voltage, 217.5 V; rear funnel exit voltage, 45 V.

B2   Nonlinear regression analysis

Iterative nonlinear regression modeling for the super classes was performed using GraphPad Prism 7, and 99% confidence intervals were generated for each biomolecular super class. Three fits were tested for each super class: power fit (PF), 4-parameter sigmoidal (4P), and 5-parameter sigmoidal (5P). The most parsimonious fit was chosen by a probabilistic comparison of the corrected Akaike information criterion (AICc) values.

Power Association (General)

$$y = y_0 + (\text{plateau} - y_0) * (1 - e^{-Kx}) \qquad \text{Eq. B1}$$

4P sigmoidal (General)

$$y = y_o + \frac{y_{max} - y_0}{1 + 10^{(\log y_{50} - x) \cdot H}} \qquad \text{Eq. B2}$$

Lipids and lipid-like molecules- Power association

$$y = 75.41 + (516.9 - 75.41) * (1 - e^{-0.000851x}) \qquad \text{Eq. B3}$$

Organic oxygen compounds- Power association

$$y = 101.9 + (935.0 - 101.9) * (1 - e^{-0.000251x}) \qquad \text{Eq. B4}$$

Nucleosides, nucleotides, and analogues- Power association

$$y = 117.6 + (1.52 \times 10^7 - 117.6) * (1 - e^{-1.06 \times 10^{-8}x}) \qquad \text{Eq. B5}$$

Organic acids and derivatives – 4P sigmoidal

$$y = 104.8 + \frac{216.3 - 104.8}{1 + 10^{(277.4 - x) \cdot 0.00376}} \qquad \text{Eq. B6}$$
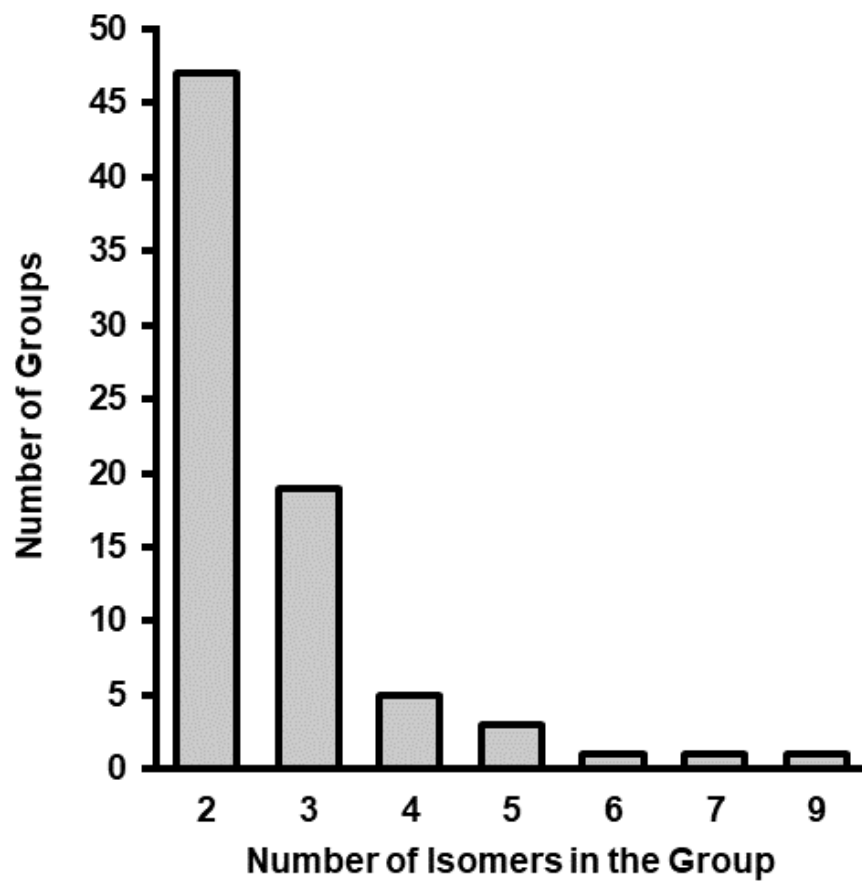
Figure B1. The distribution of isomeric families within the MSMLS. Most isomeric sets contain 2 or 3 isomers per group, and the largest set contained 9 isomers.

## B3	Preparation and analysis of NIST 1950 Serum

### B3.1	Preparation of human serum

Protein precipitation was performed by adding 800 µL of ice cold MeOH to 100µL NIST 1950 serum and stored at -80 ˚C for one hour. The sample was centrifuged at 14,000 rpm for 5 minutes before collecting the supernatant. Next, 2.4 mL ice cold methyl tert-butyl ether and 800 µL ice cold water were added. The sample was vortexed then centrifuged at 10,000 rpm at 4˚C for 10 minutes. The polar and nonpolar fractions were separated and dried separately *in vacuo.* Samples were stored at -20 ˚C until analysis. Dried fractions were resuspended in 200 µL of the initial mobile phase solvent and analyzed via LC-IM-MS.

### B3.2	Liquid chromatography

LC-IM-MS was performed on the prepared NIST 1950 serum using HILIC chromatography for the hydrophilic layer of the liquid-liquid extraction. For this method, 4 µL of sample was injected onto a column heated to 40 °C. The Millipore SeQuant Zic-HILIC (2.1 x 100 mm, 3.5 µm) column was used with mobile phase A and B being 9:1 and 1:9 (water: acetonitrile, buffered with 5 mM ammonium formate), respectively. The mobile phase flow rate was 200 µL/min. The gradient was initially held at 98 %B from 0 to 1 minutes, decreased to 45 %B from 1 to 20 minutes, held at 45 %B from 20 to 22 minutes, increased to 98 %B from 22 to 40 minutes, and subsequently held at 98 %B from 40 to 45 minutes before the next injection.

APPENDIX C

SUPPLEMENTARY MATERIALS FOR CHAPTER 2

Section C1   Ion Mobility Peak Annotation

In some of the acquired measurements, multiple ion mobility peaks were observed. In an effort

to preserve information, all observed peaks were annotated by assigning a peak number to each

peak. Peak Number where "1" refers to the smallest observed CCS and each subsequent peak is

assigned 2, 3, and so on. An example is shown in Figure S1.
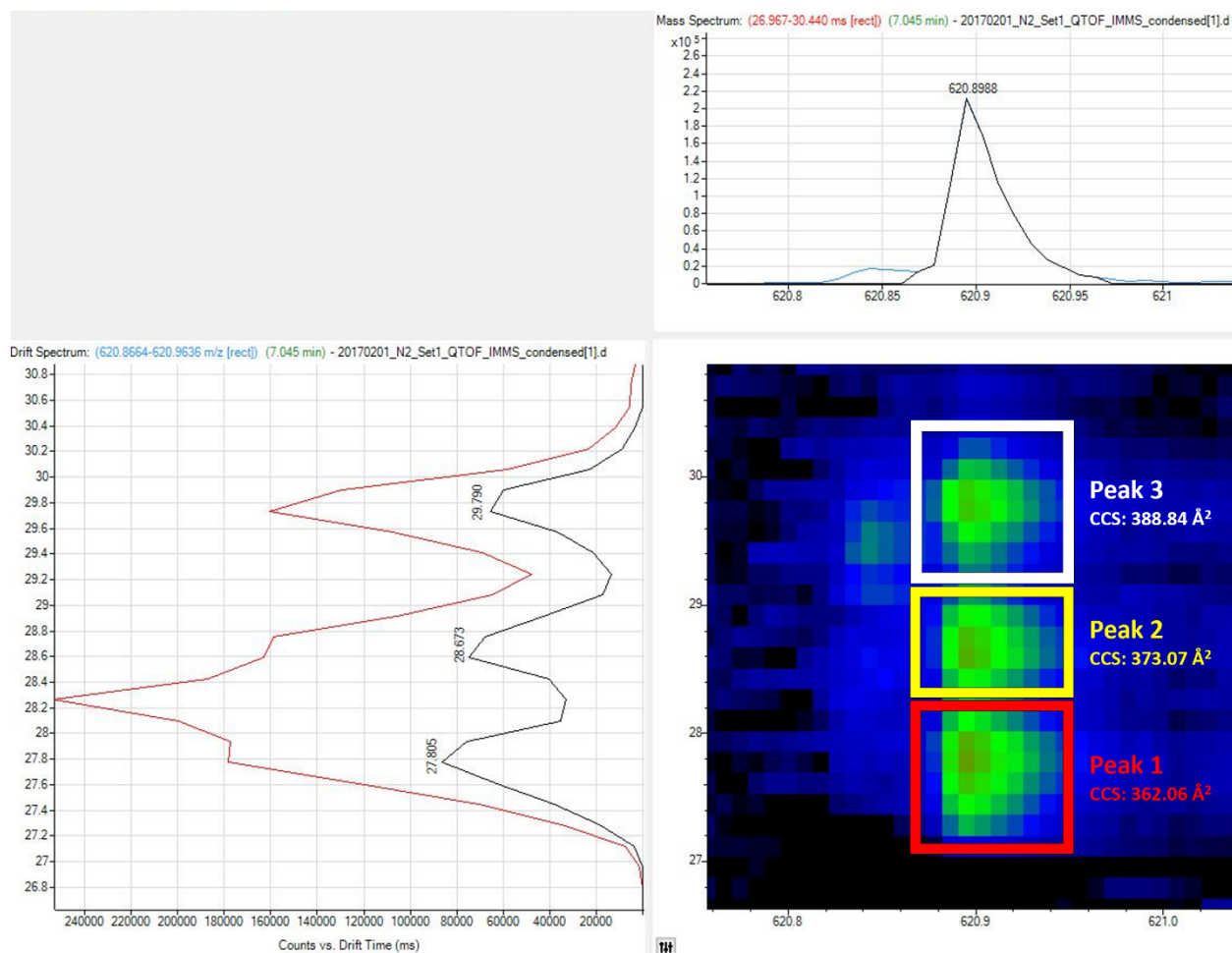
Figure C1. Illustration of annotating CCS values for analytes with multiple ion mobility peaks. If only one peak is observed, the number "1" is assigned.

Section C2   LC-MS and LC-IM-MS acquisition parameters

Serum samples were analyzed via liquid chromatographic separation using an C18 Zorbax RRHD (1.8µm) column on a 1290 Infinity LC system (Agilent Technologies). Solvent A was water with 0.1% formic acid; and Solvent B was 3:2 isopropanol:acetonitrile with 0.1% formic acid. 2 µl of sample were injected via autosampler and separations occurred using a 30 min gradient described in Fig. S5a at 200 µl/min. Post-LC separation, analytes were ionized using an electrospray ionization source (Jet Stream, Agilent Technologies) at 300°C and a VCap voltage of 3500 V. The drying gas flow rate was 8 L/min, while the sheath gas flowed at 11 L/min. When data was acquired using LC-IM-MS mode, ion mobility separations were performed using a uniform field drift tube with high-purity nitrogen drift gas at 3.95 Torr at room temperature (~298 K). A single field analysis at 1574 V was performed on a standardized calibrant mixture (Agilent Tune Mix) to normalize sample drift times. Time-of-flight scan range was 100 *m/z* to 1700 *m/z*.

Figure C2(A)
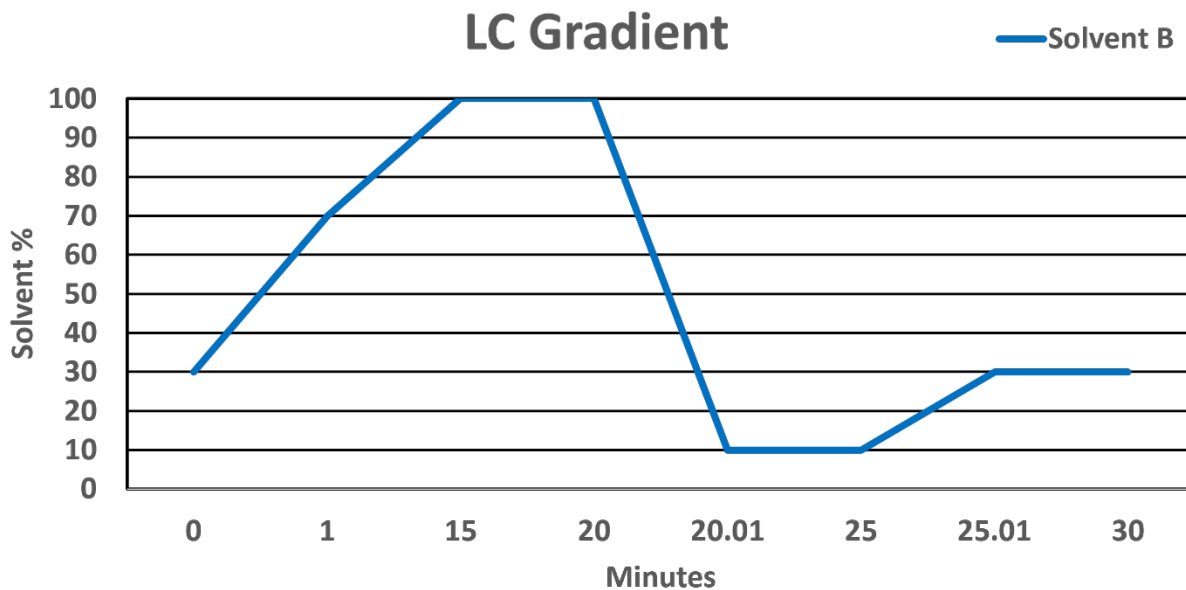
Figure C2(B)
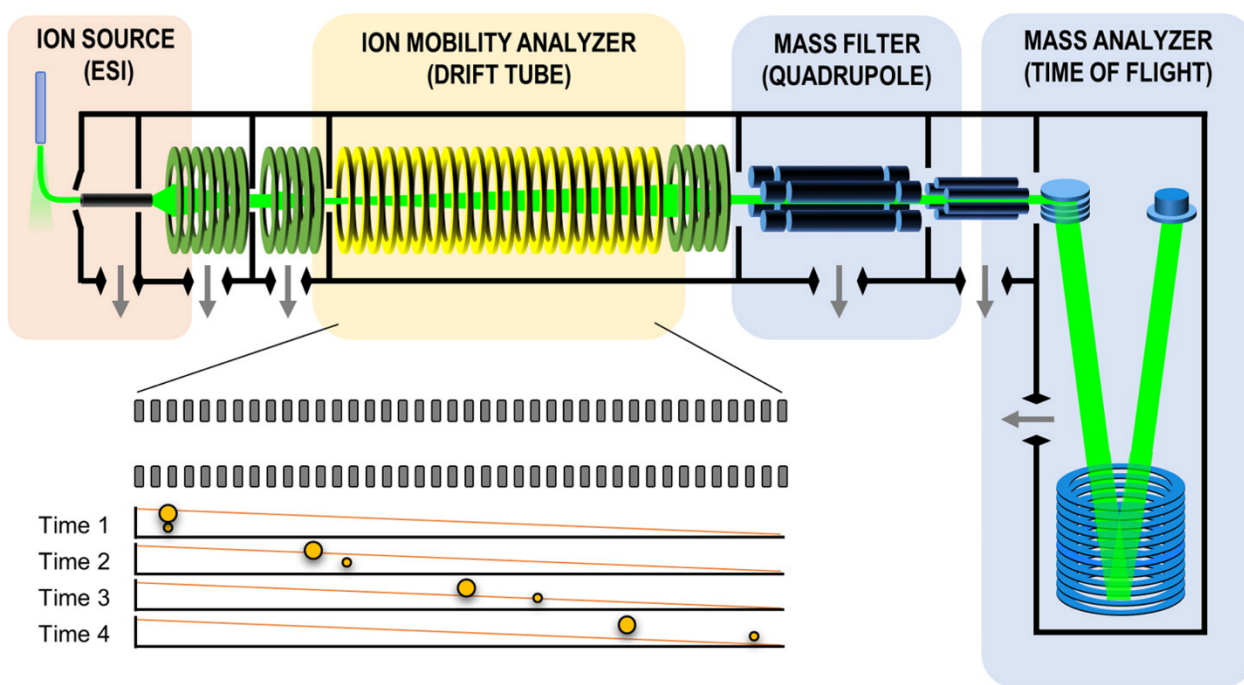
Figure C2. (A) LC Gradient. (B). Agilent 6560 instrument schematic.

Section C3   Nonlinear regression equations

Power Fit

$$y = a \cdot x^{-k} + y_0 \qquad\qquad \text{(Eq. C1)}$$

a is the curve max – curve min; k is the curve rate

Four-Parameter Sigmoidal Fit (4P)

$$y = y_0 + \frac{y_{max}-y_0}{1+10^{(\log y_{50}-x)\cdot H}} \qquad\qquad \text{(Eq. C2)}$$

y_{50} is x at curve half-maximum; H is the Hill Slope

Five-Parameter Sigmoidal Fit (5P)

$$y = y_0 + \frac{y_{max}-y_0}{\left(1+10^{((\log y_{50}-x)\cdot H}\right)^{S}} \qquad\qquad \text{(Eq. C3)}$$

S is the curve symmetry parameter

Confidence Interval

$$z \cdot s_{y,x} \cdot \left(\frac{1}{n} + \frac{(x-\bar{x})^2}{SS_x}\right)^{1/2} \qquad\qquad \text{(Eq. C4)}$$

z is standard deviations z score based on interval percentage

(z-score for 99% is 2.576);

$S_{y,x}$ is the standard error of the x and y data inputs;

SS_x is the sum of the squared deviations from the x input mean

Predictive Intervals

$$z \cdot s_{y,x} \cdot \left(1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{SS_x}\right)^{1/2} \qquad \text{(Eq. C5)}$$

Table C1.   Table of all Super Classes and Classes represented in the unified CCS Compendium at date of submission.

Super Class	Class	*m/z* Range	N
Alkaloids and derivatives	Yohimbine alkaloids	609	1
		138 – 164	3
Benzenoids	Anthracenes	178 – 271	9
	Benzene and substituted derivatives	108 – 886	159
	Fluorenes	166	1
	Indanes	300	1
	Naphthalenes	128 – 254	25
	Pentacenes	278, 280	2
	Phenanthrenes and derivatives	178 – 303	19
	Phenols	109 – 208	31
	Pyrenes	202 – 304	22
Homogeneous metal compounds	Homogeneous transition metal compounds	132 – 2991	62
Homogeneous non-metal compounds	Non-metal oxoanionic compounds	200	1
Lipids and lipid-like molecules	Fatty acyls	125 – 935	223
	Glycerolipids	253 – 746	8
	Glycerophospholipids	171 – 1017	334
	Prenol lipids	137 – 886	21
	Sphingolipids	548 – 8989	146
	Steroids and steroid derivatives	287 – 648	78
Nucleosides, nucleotides, and analogues	(5'->5')-dinucleotides	662 – 783	29
	5'-deoxyribonucleosides	250 – 408	19
	Flavin nucleotides	455 – 809	9
	Imidazole ribonucleosides and ribonucleotides	337 – 362	4
	Nucleoside and nucleotide analogues	24, 268	2
	Purine nucleosides	250 – 613	49
	Purine nucleotides	280 – 790	125
	Pyrimidine nucleosides	226 – 281	23
	Pyrimidine nucleotides	304 – 646	124
Organic acids and derivatives	Carboximidic acids and derivatives	131, 154	2
	Carboxylic acids and derivatives	89 – 2110	623
	Keto acids and derivatives	115 – 184	11
	Hydroxy acids and derivatives	103 – 239	12
	Organic carbonic acids and derivatives	155	1
	Organic phosphonic acids and derivatives	124 – 205	13

Super Class	Class	*m/z* Range	N
	Organic sulfonic acids and derivatives	124 – 213	4
	Peptidomimetics	225 – 1317	23
	Proteins	493 – 3302	139
	Sulfinic acids and derivatives	108, 111	2
Organic acids and derivatives	Tryptic peptides	288 – 1580	254
Organic nitrogen compounds	Organonitrogen compounds	74 – 1233	101
Organic oxygen compounds	Organic oxoanionic compounds	227 – 411	6
	Organooxygen compounds	105 – 1505	340
Organic polymers	Cyclic Peptides	1111 – 1704	20
	Polypeptides	294 – 1724	230
Organohalogen compounds	Organofluorides	301 – 2834	66
Organoheterocyclic compounds	Azoles	127 – 458	12
	Benzimidazoles	145 – 225	4
	Benzodioxoles	191 – 272	4
	Benzopyrans	421, 424	2
	Dihydrofurans	173 – 350	6
	Dithiolanes	228	1
	Furofurans	199, 200	2
	Imidazopyrimidines	119 – 218	60
	Indoles and derivatives	148 – 381	69
	Lactams	348 – 738	10
	Lactones	153 – 350	6
	Naphthofurans	821 – 1684	14
	Pteridines and derivatives	162 – 483	28
	Pyridinecarboxylic acids and derivatives	140	1
	Pyridines and derivatives	96 – 285	52
	Pyrroles	110	1
	Quinolines and derivatives	172 – 431	17
	Tetrahydroisoquinolines	178 – 181	2
	Tetrapyrroles and derivatives	563 – 1378	9
	Triazines	215 – 325	8
Phenylpropanoids and polyketides	Anthracyclines	540 – 1320	6
	Cinnamaldehydes	133, 134	2
	Cinnamic acids and derivatives	149 – 360	5
	Coumarins and derivatives	161, 186	2
	Flavonoids	269 – 1424	44
	Isoflavonoids	140 – 418	16
	Linear 1,3-diarylpropanoids	255 – 280	4
	Macrolactams	786 - 825	3

	Macrolides and analogues	661– 955	15
	Phenylpropanoic acids	165 – 284	11
	Tetracyclines	410	8
Polyhedralcarbon molecules		720, 840	2

SUPPLEMENTARY MATERIALS FOR CHAPTER 3
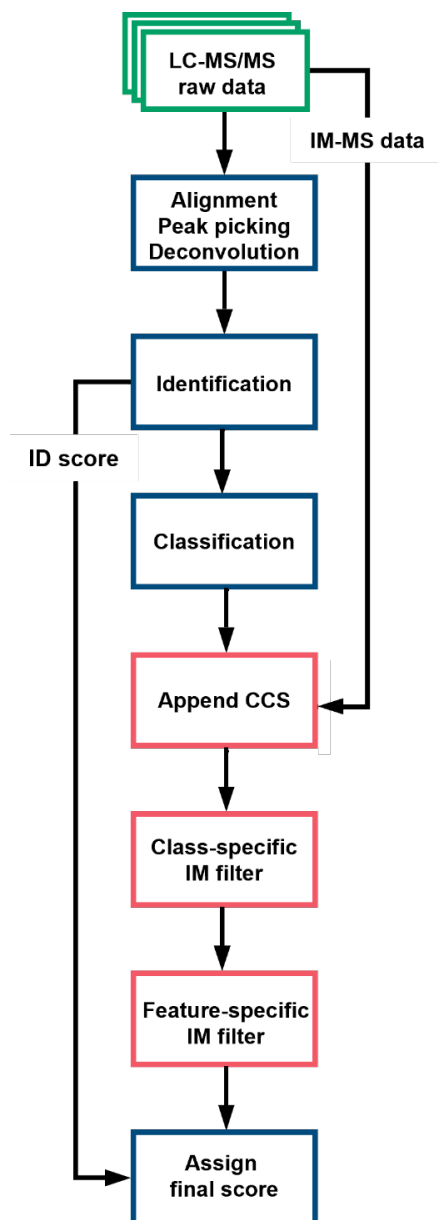
Figure D1. Detailed flow description of the integration of the filtering pipeline (pink) with a typical untargeted lipidomics annotation workflow.
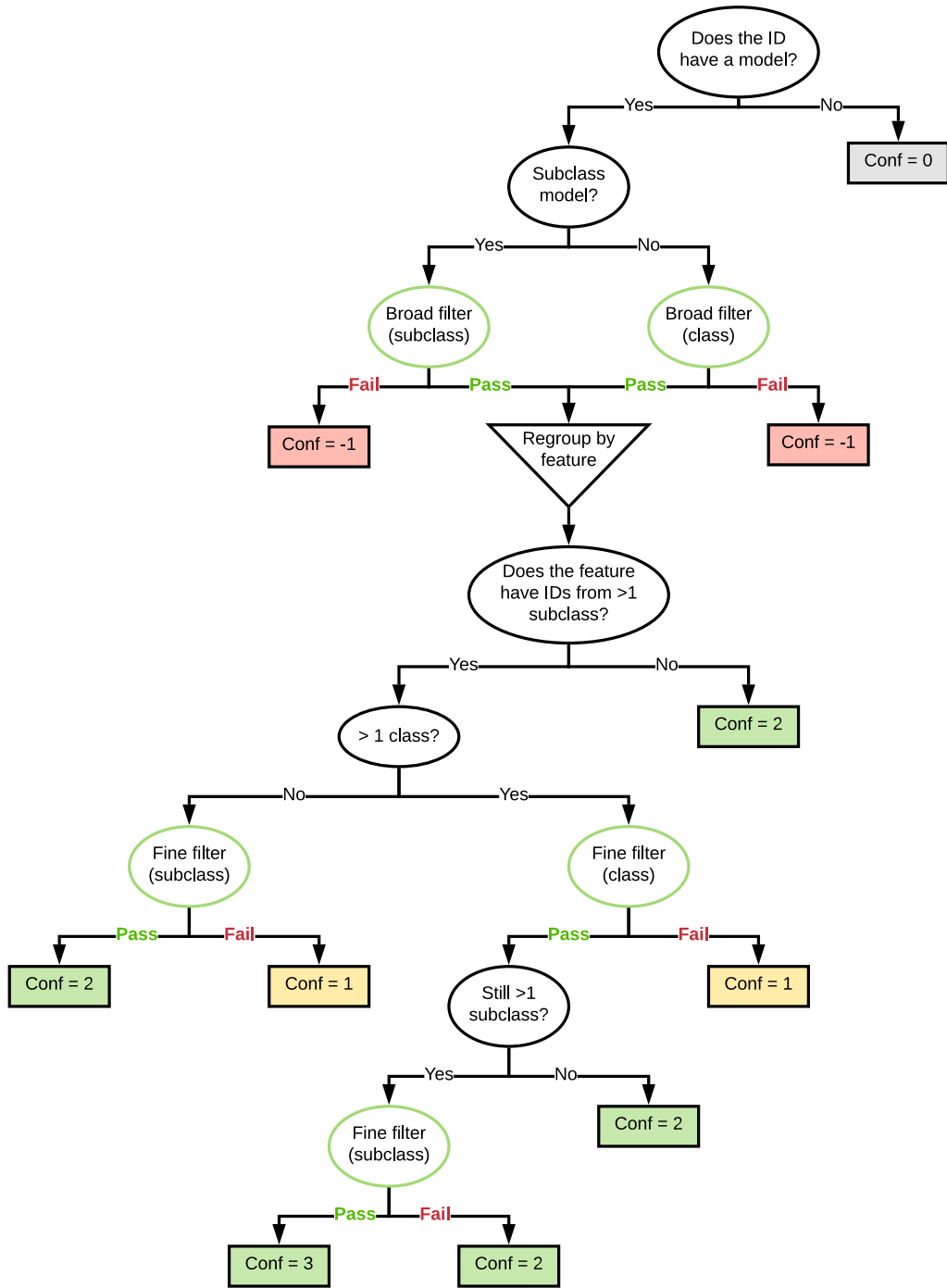
Figure D2. Decision tree describing the IM filtering process.
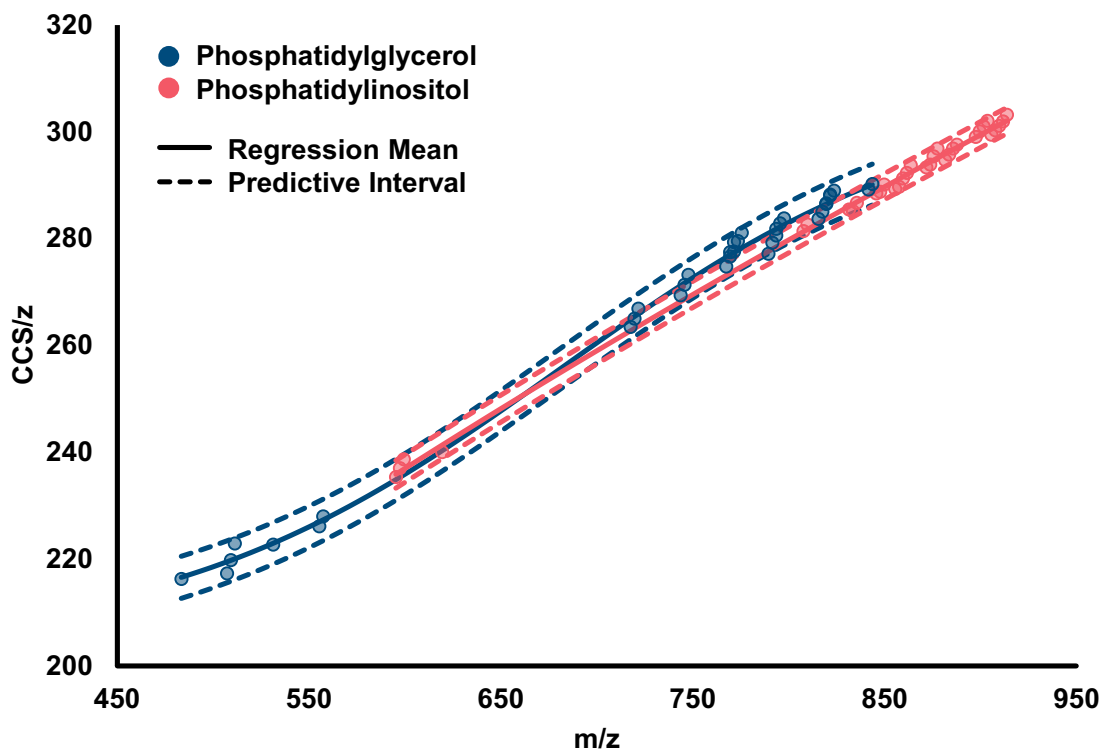
Figure D3. Phosphotidylglycerol (PG) and phosphotidyinositol (PI) conformational plot with regression models, including 99% predictive intervals. The PG regression model uses a 4-parameter sigmoidal equation, and the PI regression model uses a power equation. These models were built with methods outlined in Chapter 2.

Table D1. Collision cross section values of phosphatidylglycerol and phosphatidylinositol lipids measured via drift tube ion mobility with nitrogen (^{DT}CCS$_{N2}$) in positive and negative ionization mode. Standard deviations are reported from three replicate measurements (n=3).

Name	Molecular Formula	MW (Da)	^{DT}CCS$_{N2}$ (Å^2)		
			[M+Na]$^+$	[M+2Na-H]$^+$	[M-H]$^-$
PG 16:00	C22H45O9P	484.28	--	--	216.3±0.636
PG 18:00	C24H49O9P	512.31	--	--	222.9±1.208
PG 18:01	C24H47O9P	510.30	--	--	219.8±0.411
PG 18:02	C24H45O9P	508.28	--	--	217.3±0.130
PG 20:04	C26H45O9P	532.28	--	--	222.6±0.678
PG 22:05	C28H47O9P	558.30	--	--	227.9±1.886
PG 22:06	C28H45O9P	556.28	--	--	226.1±0.452
PG 32:00	C38H75O10P	722.51	--	--	266.8±0.070
PG 32:01	C38H73O10P	720.49	--	--	265.0±0.113
PG 32:02	C38H71O10P	718.48	--	--	263.4±0.429
PG 34:01	C40H77O10P	748.53	279.3±0.552	280.5±0.298	273.1±0.171
PG 34:02	C40H75O10P	746.51	277.4±0.138	279.2±0.231	271.3±0.058
PG 34:03	C40H73O10P	744.49	--	277.2±1.343	269.3±0.059
PG 36:01	C42H81O10P	776.56	--	288.2±0.225	281.1±0.401
PG 36:02	C42H79O10P	774.54	--	286.6±0.723	279.5±0.362
PG 36:03	C42H77O10P	772.53	--	285.0±0.653	277.5±0.174
PG 36:04	C42H75O10P	770.51	--	283.6±0.837	276.6±0.229
PG 36:05	C42H73O10P	768.49	--	--	274.8±0.250
PG 38:04	C44H79O10P	798.54	--	290.2±0.716	283.7±0.193
PG 38:05	C44H77O10P	796.53	--	289.1±0.263	282.8±0.328
PG 38:06	C44H75O10P	794.51	--	--	281.7±0.326
PG 40:05	C46H81O10P	824.56	--	--	288.9±0.415
PG 40:06	C46H79O10P	822.54	--	--	288.0±0.136
PG 40:07	C46H77O10P	820.53	--	--	286.5±0.364
PI 18:00	C27H52O12P	599.32	--	--	238.4±0.240
PI 18:01	C27H50O12P	597.30	--	--	236.7±0.221
PI 18:02	C27H48O12P	595.29	--	--	235.0±1.412
PI 20:04	C29H48O12P	619.29	--	--	239.8±0.365

PI 32:00	C41H79O13P	810.53	--	--	282.3±0.891
PI 32:01	C41H77O13P	808.51	--	--	281.1±0.690
PI 34:01	C43H81O13P	836.54	--	--	286.4±0.239
PI 34:02	C43H79O13P	834.53	--	--	284.8±0.461
PI 34:03	C43H77O13P	832.51	--	--	285.1±1.107
PI 35:01	C44H83O13P	850.56	--	--	289.8±0.310
PI 35:02	C44H81O13P	848.54	--	--	288.4±0.194
PI 35:03	C44H79O13P	846.53	--	--	288.1±0.141
PI 36:01	C45H85O13P	864.57	--	--	293.3±0.346
PI 36:02	C45H83O13P	862.56	--	--	292.0±0.101
PI 36:03	C45H81O13P	860.54	--	--	290.8±0.069
PI 36:04	C45H79O13P	858.53	--	--	289.4±0.186
PI 36:05	C45H77O13P	856.51	--	--	289.0±0.384
PI 37:01	C46H87O13P	878.59	--	--	296.5±1.107
PI 37:02	C46H85O13P	876.57	--	--	295.1±0.179
PI 37:03	C46H83O13P	874.56	--	--	293.5±0.151
PI 37:04	C46H81O13P	872.54	--	--	293.0±0.279
PI 38:03	C47H85O13P	888.57	--	--	297.3±0.665
PI 38:04	C47H83O13P	886.56	--	--	296.4±0.514
PI 38:05	C47H81O13P	884.54	--	--	295.4±0.531
PI 38:06	C47H79O13P	882.53	--	--	294.2±0.281
PI 39:02	C48H89O13P	904.60	--	--	301.7±0.535
PI 39:03	C48H87O13P	902.59	--	--	300.7±1.086
PI 39:04	C48H85O13P	900.57	--	--	299.6±0.941
PI 39:05	C48H83O13P	898.56	--	--	298.6±0.437
PI 40:04	C49H87O13P	914.59	--	--	302.8±0.828
PI 40:05	C49H85O13P	912.57	--	--	301.6±0.829
PI 40:06	C49H83O13P	910.56	--	--	300.8±0.759
PI 40:07	C49H81O13P	908.54	--	--	299.9±0.520
PI 40:08	C49H79O13P	906.53	--	--	299.1±0.721

SUPPLEMENTARY MATERIALS FOR CHAPTER 4
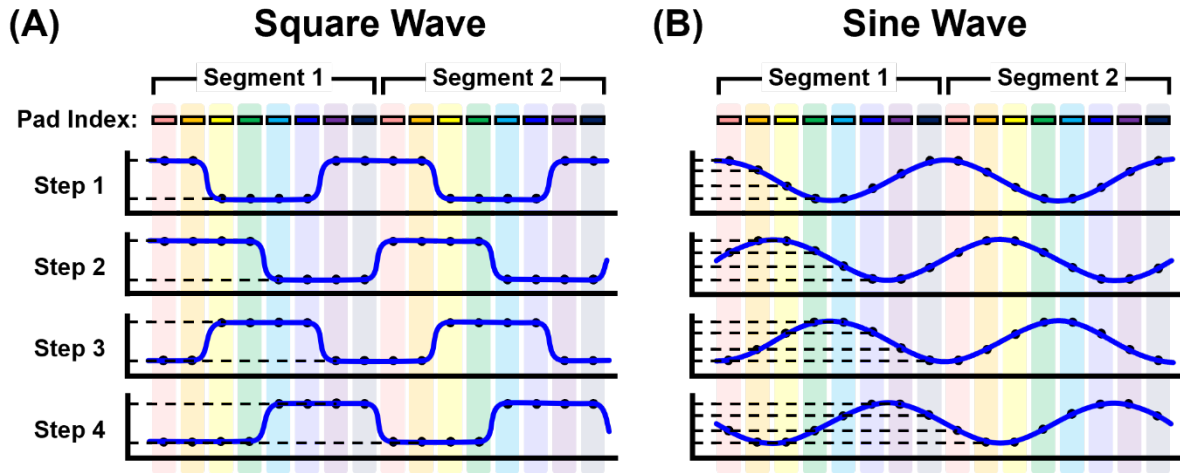
Figure E1. Traveling wave operation of SLIM IM illustrated for two SLIM segments (8 pads each) using (A) a square waveform, and (B) a sine waveform.
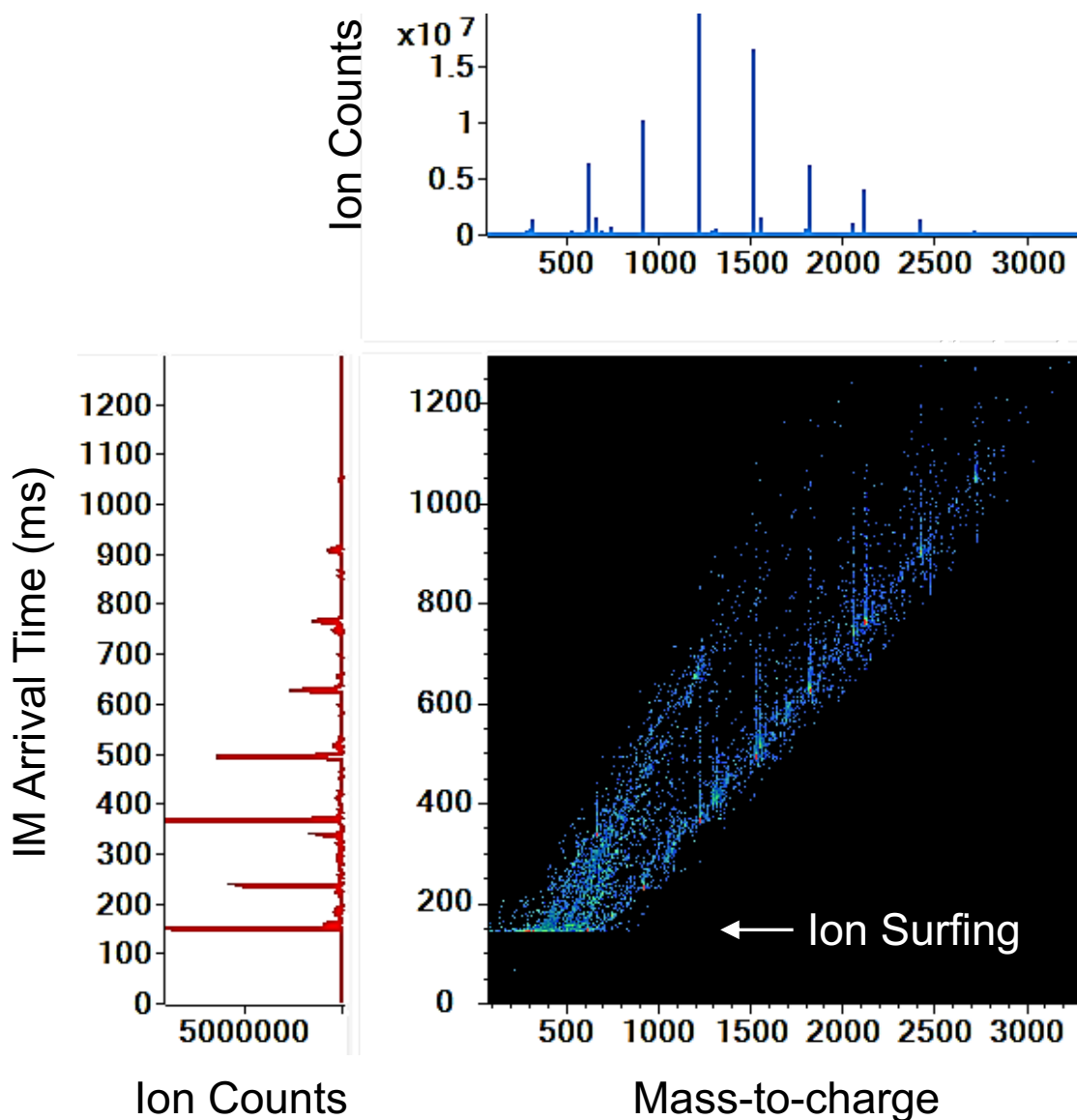
Figure E2. A multidimensional IM-MS spectrum of tune mix ions illustrating the spectrum observed under conditions where the lower *m/z* ions are "surfing" the traveling wave and no IM separation occurs. Here, SLIM conditions were 90 m/s and 30 Vpp. Under surfing conditions, all ions exhibit the same arrival times, which corresponds to the wave propagation time through the SLIM device (90 m/s / 13 m = ~144 ms).
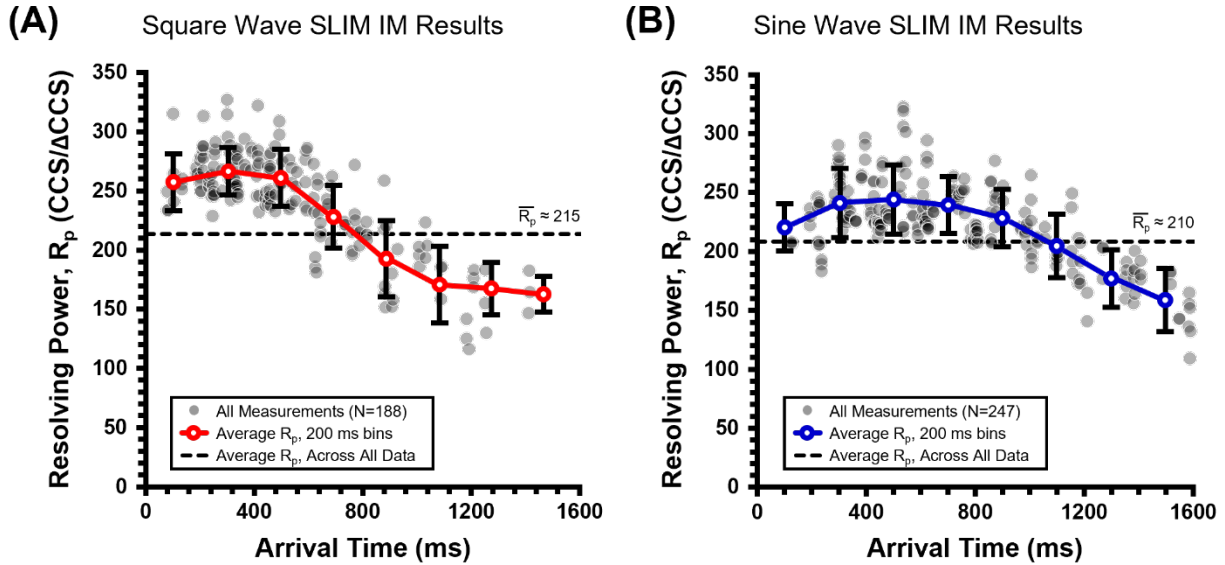
Figure E3. Plots projecting the single-peak resolving power in CCS-space, R_p(CCS), as a function of the measured arrival times for (A) square wave, and (B) sine wave operation of the traveling wave. Plots contain data points for all wave amplitudes (30, 35, and 40 V_{pp}) and wave speeds (45, 90, 135, 180, and 225 m/s) surveyed in these experiments, with data points omitted for ions which do not exhibit IM-selective behavior (that is, ions which "surf" the wave and thus are not resolved). For square wave, 52,2% of the measurements are IM-selective (188 out of 360), whereas for sine wave, 68.6% of the ions (247 out of 360) are observed to separate via the traveling wave. In both projections, the circle markers connected with solid traces represent the average R_p(CCS), determined for 200 ms arrival time bins, and the horizontal dotted line corresponds to the average R_p(CCS) across the entire dataset. Whereas both datasets exhibit similar resolving powers, the square wave data (panel A) transmits fewer ions under IM-selective conditions across the parameters surveyed.

156
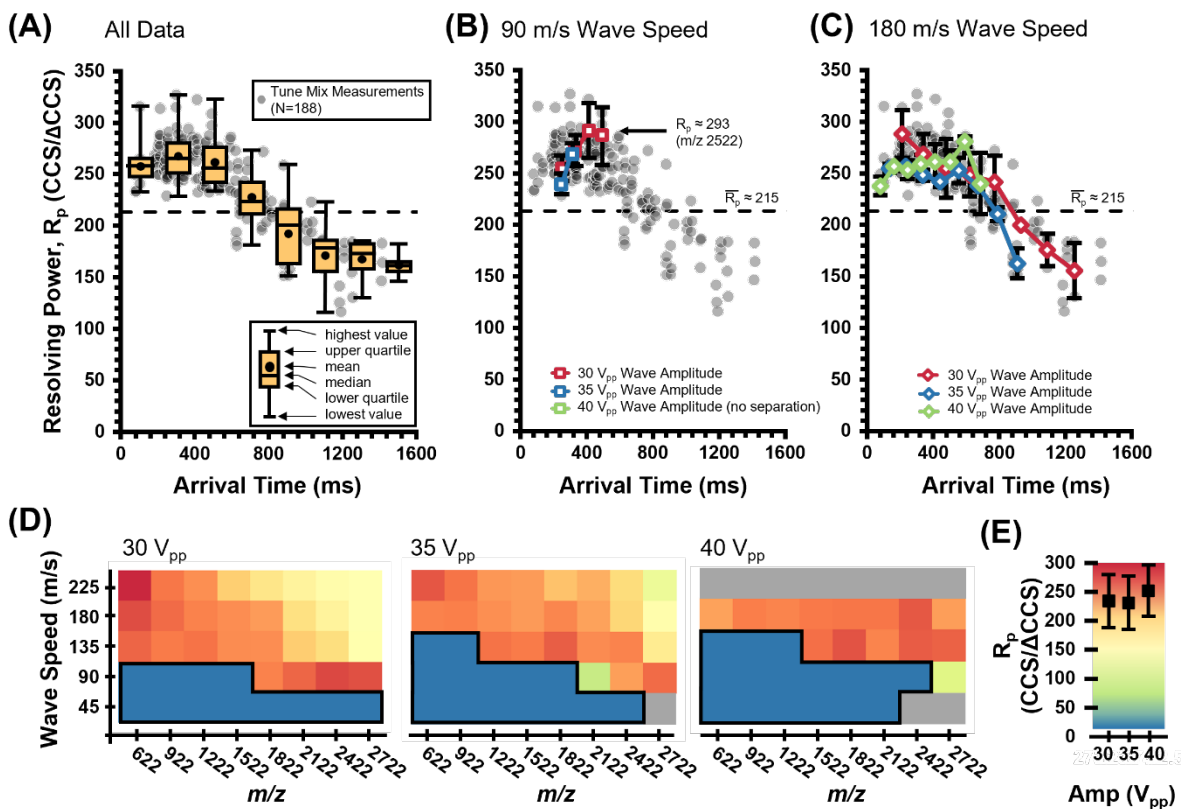
Figure E4. Plots assessing the resolving power performance for a square waveform under the traveling wave parameters surveyed in this work. (A) CCS-based R_p values (N=188) as a function of the measured arrival time for tune mix components (cyclophosphazenes) measured across various SLIM IM parameters, including wave speeds (45, 90, 135, 180, and 225 m/s) and wave amplitudes (30, 35, and 40 Vpp). Box and whisker overlays summarize the data within 200 ms bins. (B) Average R_p(CCS) values (3 replicates per data point) determined for 90 m/s wave speed at 30, 35, and 40 V_{pp} wave amplitudes. Under the highest amplitude evaluated (40 V_{pp}), no ions are transmitted under IM-selective arrival times for square wave. (C) Average R_p(CCS) values calculated for 180 m/s data. The horizontal dotted line in these scatter plots represents the average R_p across the entire dataset (ca. 215). (D) Heat maps visualizing the resolving powers for each tune mix component (x-axis) at each wave speed (y-axis) for wave amplitudes of 30 (left panel), 35 (middle panel), and 40 V_{pp} (right panel). Each square represents an average of three replicate measurements. Here, dark blue squares represent conditions in which ions are transmitted, but no IM separation occurs (i.e., ion "surfing" conditions), whereas grey boxes correspond to conditions where ion transmission is poor. (E) Color scale for panel D with average and standard deviation values overlaid.
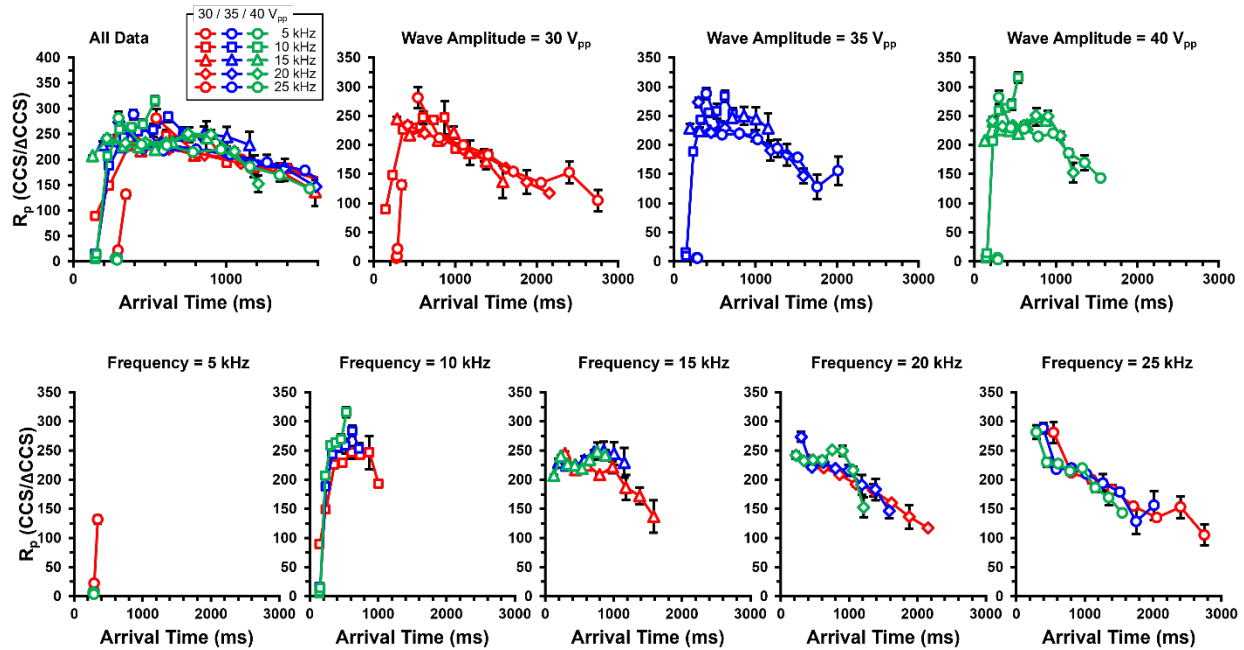
Figure E5. Resolving power (CCS/ΔCCS) results for all tune mix ions, separated into the different wave amplitudes and wave speeds surveyed. Here, ions which are not fully separated by the traveling waves (surfing ions) are also included.
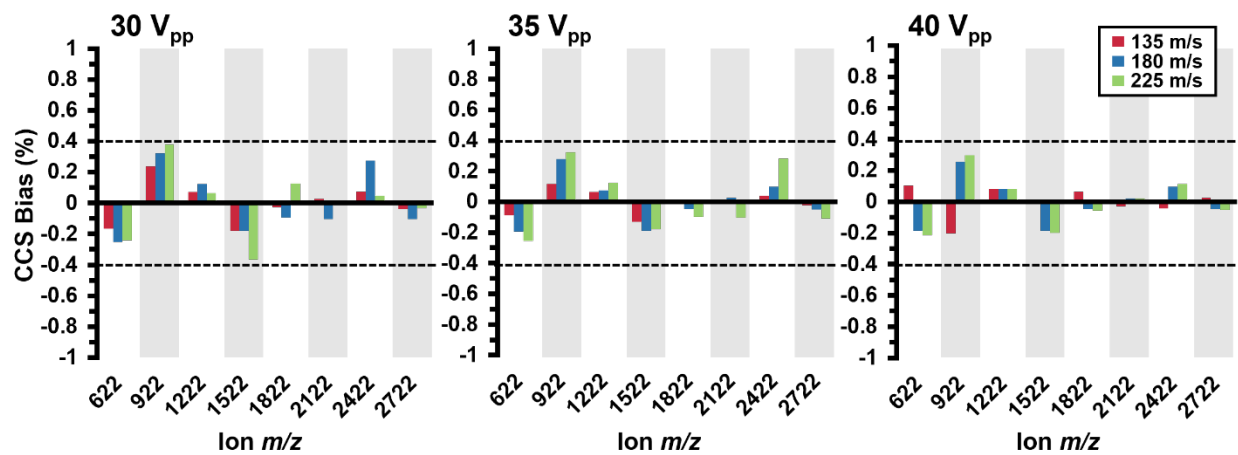
Figure E6. CCS calibration biases observed across all SLIM IM parameters for a 3rd order polynomial fit. Dotted lines represent ±0.4% bias of the calibrated CCS compared with the reference CCS values.

Table E1. Corresponding Wave Speeds for SLIM IM Wave Switching Frequencies.

SLIM IM TW Output Drive Frequency (kHz)	Equivalent Wave Speed (m/s) [a.]
5	45
10	90
15	135
20	180
25	225

a. To calculate the wave speed, the drive frequency (in kHz) is multiplied by the pad-to-pad distance (1.125 mm) and the number of pads traversed to complete one phase of the wave form (8 pads).

Table E2. The Second Highest Resolving Powers Observed

Tune Mix Ion	Highest R_p Measured (CCS/ΔCCS) [a.]	Corresponding Parameter		Corresponding Arrival Time (ms)
		Wave Amplitude (V_{pp})	Wave Speed (m/s)	
m/z 622	283.7±20.7 (3)	40	180	215.0
m/z 922	258.7±8.0 (3)	40	225	437.9
m/z 1222	247.1±9.7 (3)	40	225	601.2
m/z 1522	243.0±7.1 (2)	35	90	330.8
m/z 1822	255.1±7.1 (3)	35	90	426.5
m/z 2122	257.9±12.7 (3)	35	90	523.8
m/z 2422	268.8±8.9 (3)	40	90	458.2
m/z 2722	253.0±10.1 (3)	35	90	721.8

a. Highest R_p is averaged over replicate measurements, denoted in the parenthesis. The time-to-CCS conversion is determined from Equation 3 in the main text using the differences between the tune mix ion and the next highest *m/z* ion in the spectrum.

SUPPLEMENTARY MATERIALS FOR CHAPTER 5
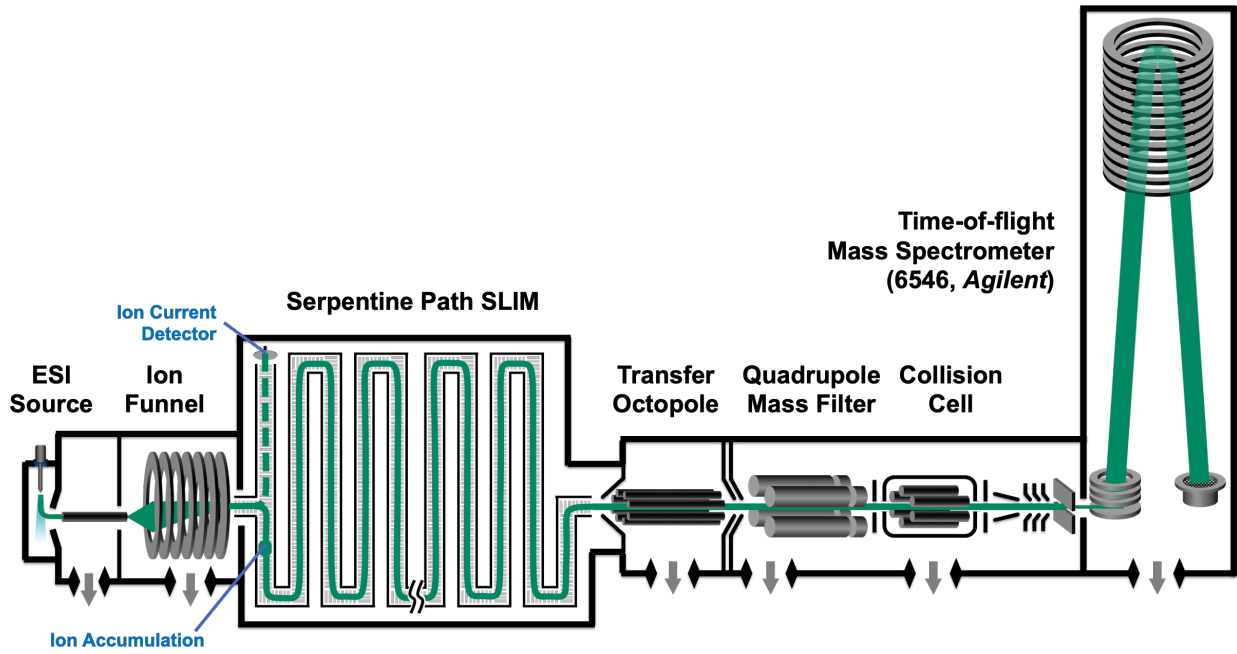
Figure F1. Schematic of the beta prototype HRIM-MS platform used in this study.
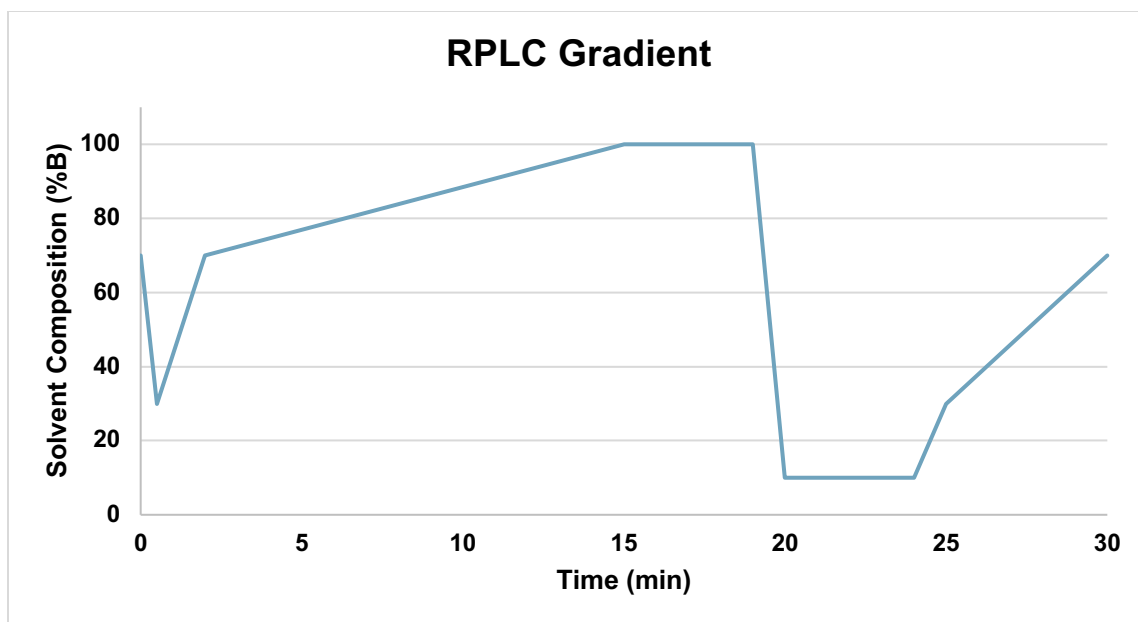
Figure F2. The standard lipid mix was analyzed using a 1290 Infinity LC system (Agilent). Reversed phase LC was performed at a flow rate of 250 μL/min using a C18 column (HypersilGold 1.9 μm, 2.1 mm x 100 mm column, Thermo Fisher) held at 40 ˚C with mobile phases consisting of 10 mM ammonium formate and 0.1% formic acid in H2O (A) and 60:36:4 IPA:ACN:H2O (B). The 30 minute gradient is illustrated above.
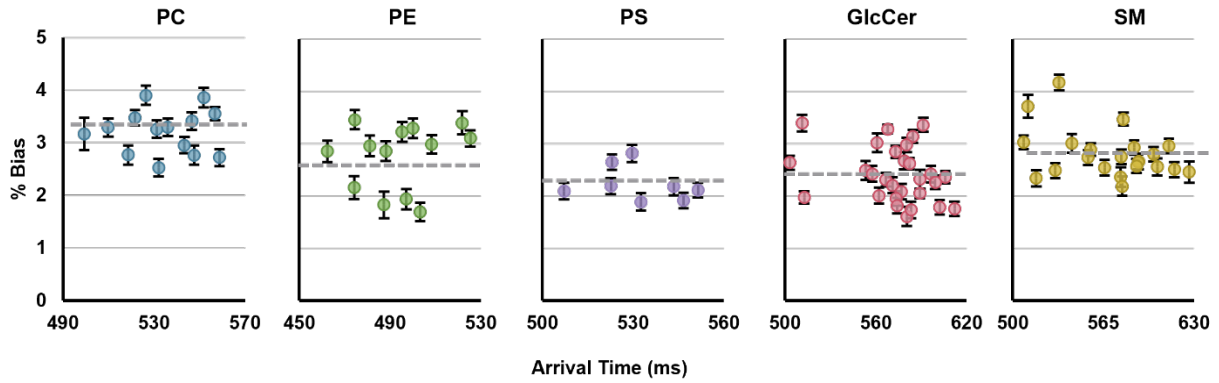
Figure F3. CCS calibration bias for all values calibrated from triplicate measurements using HFAPs with a 3rd order polynomial as compared to established DTCCS values. Dashed lines represent average biases for each subclass, and error bars represent the measurement standard deviation.
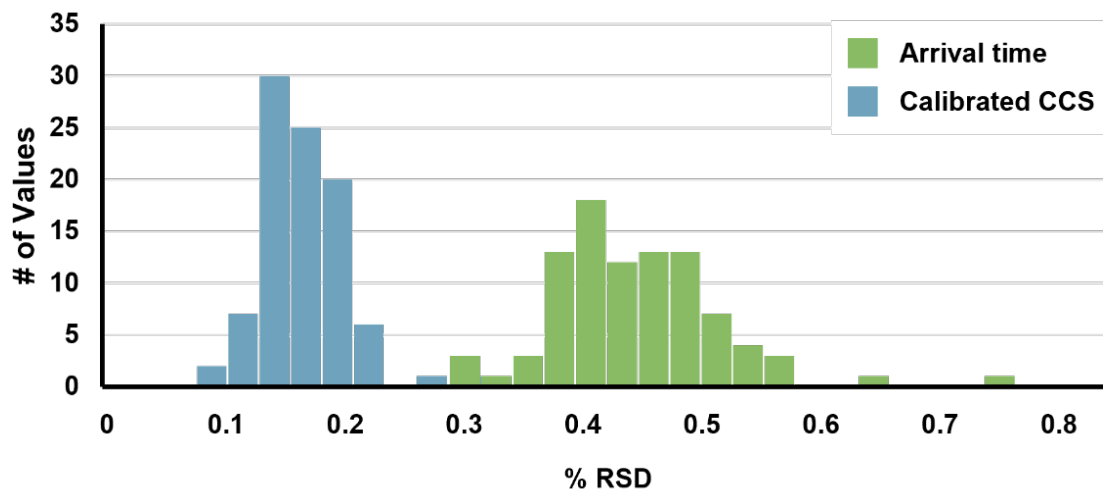
Figure F4. Inter-day reproducibility of experimental arrival times (green) and calibrated CCS values from all lipid features (n = 92).

Table F1. Subclass specific semi-empirically derived correction factors used for correction of HFAP-calibrated CCS values.

Class	Subclass	Correction factor
Glycerophospholipids	PC	0.0321
	PE	0.0275
	PS	0.0223
Sphingolipids	GlcCer	0.0241
	SM	0.0281
SPLASH Mix		0.0150

Table F2. Table of calibrated and corrected CCS values for all lipids using the correction factors in Table S1. ^{DT}CCS values in the left column are sourced from the Unified CCS Compendium. Lipid features from each subclass that were not identified are annotated as their subclass and measured *m/z*. When arrival time features were observed for one *m/z*, peak numbers are annotated in ascending order of arrival time.

Tentative ID				Feature Description			
Name	Adduct	^{DT}CCS (Å²)	Bias (%)	*m/z* (measured)	Peak #	$^{TW-SLIM}$CCS (Å²)	RSD (%)
Phosphatidylcholines							
PC 34:01	[M+H]+	283.1	0.21	760.5899	1	283.7	0.14
PC 34:01	[M+H]+	283.1	0.64	760.5902	2	284.9	0.18
PC 34:01	[M+Na]+	286.0	0.00	782.5720	1	286.0	0.16
PC 34:02	[M+H]+	280.6	0.03	758.5742	1	280.7	0.17
PC 34:02	[M+Na]+	284.2	-0.49	780.5566	1	282.8	0.18
PC 34:03	[M+H]+	278.2	-0.09	756.5605	1	277.9	0.31
PC 36:01	[M+H]+	289.4	0.15	788.6210	1	289.8	0.17
PC 36:01	[M+H]+	289.4	0.60	788.6215	2	291.1	0.19
PC 36:01	[M+Na]+	291.4	-0.50	810.6054	1	290.0	0.18
PC 36:01	[M+Na]+	291.4	0.29	810.6051	2	292.2	0.13
PC 36:02	[M+H]+	287.1	0.03	786.6056	1	287.2	0.17
PC 36:02	[M+Na]+	289.8	-0.31	808.5896	1	288.9	0.15
PC 36:03	[M+Na]+	288.2	-0.74	806.5746	1	286.1	0.17
PC 38:03	[M+Na]+	294.2	-0.55	834.6071	1	292.6	0.17
PC (*m/z* 782.57)	--	--	--	782.5742	1	293.3	0.19
PC (*m/z* 804.56)	--	--	--	804.5588	1	296.4	0.14
PC (*m/z* 832.59)	--	--	--	832.5906	1	302.9	0.17
PC (*m/z* 836.63)	--	--	--	836.6243	1	305.0	0.17
	Average absolute % bias		**0.33**			**Average % RSD**	**0.17**
Phosphatidylethanolamines							
PE 34:01	[M+H]+	271.1	0.67	718.5438	1	272.9	0.18
PE 34:01	[M+Na]+	277.1	0.44	740.5270	1	278.3	0.19
PE 34:02	[M+H]+	269.5	0.06	716.5275	1	269.7	0.21
PE 34:02	[M+Na]+	274.1	0.17	738.5098	1	274.6	0.20
PE 36:01	[M+H]+	278.1	0.51	746.5752	1	279.5	0.19
PE 36:01	[M+Na]+	283.3	0.61	768.5599	1	285.0	0.22
PE 36:01	[M+Na]+	283.3	-1.09	768.5587	2	280.2	0.18
PE 36:01	[M+2Na-H]+	284.9	0.31	790.5421	1	285.8	0.16
PE 36:02	[M+H]+	276.2	0.07	744.5589	1	276.4	0.19
PE 36:02	[M+Na]+	281.0	0.20	766.5410	1	281.6	0.18
PE 36:02	[M+Na]+	281.0	-0.85	766.5448	2	278.6	0.19
PE 36:03	[M+H]+	274.4	-0.63	742.5467	1	272.7	0.22
PE 36:03	[M+Na]+	278.7	-0.96	764.5288	1	276.0	0.25
PE (*m/z* 740.53)	--	--	--	740.5295	1	273.5	0.20
	Average absolute % bias		**0.51**			**Average % RSD**	**0.20**
Phosphatidylserines							
PS 36:01	[M+H]+	285.6	0.40	790.5650	1	286.7	0.15

167

PS 36:02	[M+H]+	283.1	-0.16	788.5505	1	282.6	0.16
PS 38:04	[M+H]+	286.7	0.56	812.5484	1	286.5	0.15
PS 38:04	[M+H]+	286.7	-0.07	812.5490	2	288.3	0.16
PS 40:04	[M+H]+	293.5	-0.34	840.5801	1	292.5	0.15
PS 40:05	[M+H]+	291.9	-0.08	838.5661	1	291.7	0.17
PS 40:06	[M+H]+	290.0	-0.37	836.5494	1	288.9	0.17
PS 40:06	[M+Na]+	294.0	-0.14	858.5327	1	293.6	0.14
	Average absolute % bias		**0.27**			**Average % RSD**	**0.16**

Cerebrosides

GlcCer 36:01	[M+H-H$_2$O]+	281.0	0.20	710.6001	1	281.5	0.14
GlcCer 36:01	[M+H-H$_2$O]+	281.0	0.95	710.5992	2	283.7	0.16
GlcCer 36:01	[M+Na]+	285.1	-0.47	750.5925	1	283.8	0.12
GlcCer 40:00	[M+Na]+	298.9	-0.23	808.6726	1	298.2	0.15
GlcCer 40:00	[M+Na]+	298.9	0.54	808.6729	2	300.5	0.14
GlcCer 40:01	[M+H-H$_2$O]+	293.9	0.05	766.6641	1	294.0	0.18
GlcCer 40:01	[M+H-H$_2$O]+	293.9	0.58	766.6653	2	295.6	0.19
GlcCer 40:01	[M+Na]+	297.3	-0.44	806.6560	1	296.0	0.15
GlcCer 40:02	[M+H]+	295.1	-0.02	782.6565	1	295.1	0.15
GlcCer 40:02	[M+H]+	295.1	0.84	782.6593	2	297.6	0.08
GlcCer 42:00	[M+Na]+	304.3	-0.02	836.7027	1	304.2	0.16
GlcCer 42:00 OH	[M+Na]+	306.7	-0.08	852.6966	1	306.5	0.12
GlcCer 42:01	[M+H-H$_2$O]+	299.5	0.23	794.6925	1	300.2	0.17
GlcCer 42:01	[M+H-H$_2$O]+	299.5	0.70	794.6928	2	301.6	0.13
GlcCer 42:01	[M+Na]+	302.9	-0.83	834.6867	1	300.4	0.19
GlcCer 42:01	[M+Na]+	302.9	-0.12	834.6864	2	302.5	0.18
GlcCer 42:01 OH	[M+Na]+	305.4	-0.18	850.6816	1	304.9	0.13
GlcCer 42:02	[M+H]+	300.4	-0.36	810.6905	1	299.3	0.15
GlcCer 42:02	[M+H]+	300.4	0.18	810.6890	2	300.9	0.11
GlcCer 42:02	[M+H]+	300.4	0.92	810.6884	3	303.2	0.14
GlcCer 42:02	[M+H-H$_2$O]+	297.6	-0.13	792.6776	1	297.2	0.13
GlcCer 42:02	[M+H-H$_2$O]+	297.6	0.41	792.6778	2	298.8	0.12
GlcCer 42:02	[M+Na]+	300.7	-0.63	832.6701	1	298.8	0.15
GlcCer 42:02 OH	[M+Na]+	303.1	-0.70	848.6660	1	301.0	0.16
GlcCer 42:04 OH	[M+H]+	300.2	-0.48	822.6523	1	298.8	0.14
GlcCer 43:02	[M+Na]+	303.6	-0.39	846.6836	1	302.4	0.09
GlcCer 44:02	[M+Na]+	307.5	-0.66	860.7005	1	305.5	0.14
GlcCer 46:05 OH	[M+H]+	309.9	-0.69	876.6957	1	307.8	0.14
	Average absolute % bias		**0.43**			**Average % RSD**	**0.14**

Sphingomyelins

SM 34:01	[M+H]+	281.2	0.18	703.5808	1	281.7	0.13
SM 34:01	[M+Na]+	279.9	0.87	725.5647	1	282.3	0.22
SM 36:01	[M+H]+	288.4	-0.36	731.6116	1	287.4	0.15
SM 36:01	[M+Na]+	284.1	1.32	753.5268	1	287.9	0.15
SM 36:02	[M+H]+	285.3	-0.51	729.5966	1	283.9	0.15
SM 38:00	[M+H]+	289.8	0.16	745.6293	1	290.3	0.18
SM 38:01	[M+H]+	293.4	-0.11	759.6422	1	293.1	0.15
SM 38:01	[M+Na]+	293.3	0.04	781.6248	1	293.4	0.12

SM 40:00	[M+H]+	300.8	0.08	789.6903	1	301.0	0.14
SM 40:01	[M+H]+	299.1	-0.10	787.6743	1	298.8	0.14
SM 40:01	[M+Na]+	297.2	0.62	809.6581	1	299.0	0.13
SM 40:02	[M+H]+	296.8	-0.31	785.6599	1	295.9	0.15
SM 41:01	[M+H]+	302.3	-0.18	801.6911	1	301.7	0.15
SM 41:02	[M+H]+	300.1	-0.49	799.6761	1	298.6	0.19
SM 42:00	[M+H]+	306.5	0.11	817.7225	1	306.8	0.14
SM 42:01	[M+H]+	304.4	-0.05	815.7058	1	304.3	0.14
SM 42:02	[M+H]+	302.2	-0.19	813.6896	1	301.6	0.15
SM 42:02	[M+Na]+	302.0	-0.28	835.6737	1	301.2	0.14
SM 42:03	[M+H]+	300.8	-0.66	811.6745	1	298.8	0.19
SM 44:01	[M+H]+	305.7	-0.29	827.7070	1	304.8	0.17
SM 44:02	[M+H]+	311.0	-0.39	843.7410	1	309.8	0.21
SM 44:03	[M+H]+	308.7	-0.34	841.7226	1	307.6	0.14
SM (*m/z* 733.63)	--	--	--	733.6272	1	290.1	0.13
SM (*m/z* 761.66)	--	--	--	761.6590	1	295.6	0.14
	Average absolute % bias		**0.35**			**Average % RSD**	**0.15**
SPLASH Mix							
18:1(d7) LPC	[M+H]+	232.9	-0.53	529.4026	1	231.7	0.09
18:1(d7) LPC	[M+Na]+	236.0	-1.14	551.3809	1	233.4	0.10
15:0-18:1(d7) DG	[M+H-H$_2$O]+	255.6	0.75	570.5576	1	256.7	0.16
15:0-18:1(d7) DG	[M+Na]+	257.1	-0.18	610.5485	1	257.6	0.12
15:0-18:1(d7) PE	[M+H]+	271.7	0.31	711.5758	1	272.6	0.10
15:0-18:1(d7) PE	[M+Na]+	277.8	0.03	733.5595	1	277.9	0.11
15:0-18:1(d7) PS	[M+H]+	278.6	0.23	755.5674	1	279.2	0.10
18:1(d9) SM	[M+H]+	286.3	0.71	738.6596	1	287.5	0.10
18:1(d9) SM	[M+Na]+	286.9	0.22	760.6386	1	288.4	0.11
15:0-18:1(d7)-15:0 TG	[M+NH$_4$]+	312.1	-0.11	829.8109	1	309.3	0.10
15:0-18:1(d7)-15:0 TG	[M+Na]+	310.6	-0.41	834.7644	1	311.7	0.10
	Average absolute % bias		**0.42**			**Average % RSD**	**0.11**

CURRICULUM VITAE

Bailey S. Rose

Permanent address:

3375 Brown Circle

Cookeville, TN 38506

baileysutton@gmail.com

931-239-6464

Present address:

320 Old Hickory Blvd. Apt 703

Nashville, TN 37221

bailey.s.rose@vanderbilt.edu

931-239-6464

EDUCATION

Vanderbilt University   Nashville, TN

Ph.D. in Chemistry   May 2022

Dissertation: Multidimensional Analyses for High-Confidence Untargeted Molecular Annotation Using Structurally Specific Ion Mobility Mass-Spectrometry

Advisor: Dr. John A. McLean

Belmont University   Nashville, TN

B.S. Chemistry   May 2017

Advisor: Dr. Thom Spence/Dr. Justin Stace

RESEARCH EXPERIENCE

Graduate Research Assistant   2017-2022

Department of Chemistry

Vanderbilt University

Nashville, TN

Principal Investigator: Dr. John McLean

Explored strategies for improving lipid and metabolite identifications in untargeted -omics studies using ion mobility-derived collision cross sections (CCS). Developed standardization and reporting recommendations for CCS values derived from various types of ion mobility measurements and used these recommendations to build databases and implement novel algorithms for obtaining high-confidence molecular identifications. Other projects included LC-MS/MS method optimization for lipidomics and application of lipidomic methodology and data analysis pipelines to various biological systems in collaboration with the Center for Innovative Technology at Vanderbilt.

Undergraduate Research Assistant   2016-2017

Department of Chemistry

Belmont University

Nashville, TN

Principal Investigator: Dr. Thom Spence

Completed optical design, alignment, and construction of a cavity ring-down system using a nitrogen pumped dye laser. After complete optical alignment of the high-reflectivity mirrors, a decay signal was generated proving the proper function of the cavity for further gas-phase analysis. An iron (II) terpyridal complex was used to assess and optimize the sensitivity of the ring-down spectroscopic analysis.

Analytical Chemist Summer Intern   June-August 2016

Carlisle Construction Materials

Carlisle, PA

Supervisor: Dr. Vivian Zhang

Completed analytical projects relating to production quality control and warranty claims. Developed rheological methods and SOPs for new product lines in development, and trained staff on use and maintenance of the rheometer.

Undergraduate Research Assistant   June-November 2015

Department of Chemistry

Belmont University

Nashville, TN

Principal Investigator: Dr. Justin Stace

Synthesized an enzyme-mimic copper(II) Schiff-base complex and explored the kinetics of its reaction with oxalic acid using UV-visual data from a stopped-flow rapid-mixing accessory. Performed multiple characterization tests and experiments on the product to deduce the overall mechanism. Explored new reactions of the copper complex with other reducing agents as well as new complexes with the existing ligand.

TEACHING EXPERIENCE

Teaching Assistant   2017-2021

Department of Chemistry

Vanderbilt University

Courses: General Chemistry Lab I, Analytical Chemistry Lab, Capstone Advanced Integrated Lab

PUBLICATIONS

Reardon, A. R.; **Rose, B. S.;** Leaptrot, K. L.; Robinson, J. R.; Gilliam, S. G.; May, J. C.; McLean, J. A. High Resolution Lipid Atlas. **2022**, *in preparation.*

Rose, B. S.; May, J. C.; Reardon, A. R.; McLean, J. A. A Collision Cross Section Calibration Strategy for Lipid Measurements in SLIM-based High Resolution Ion Mobility. *J. Am. Soc. Mass Spectrom*. **2022**, *submitted*.

Rose, B. S.; May, J. C.; Picache, J. A.; Codreanu, S. G.; Sherrod, S. D.; McLean, J. A. Improving Confidence in Lipidomic Annotations by Incorporating Empirical Ion Mobility Regression Analysis and Chemical Class Prediction. *Bioinf.* **2022,** *submitted.*

Rose, B. S.; Leaptrot, K. L.; Harris, R. A.; Sherrod, S. D.; May, J. C.; McLean, J. A. A Shotgun Lipidomics Workflow for High-Confidence Lipid Analysis using a Drift Tube Ion Mobility-Tandem Mass Spectrometer. *Methods Mol Biol*. **2021**, 2306, 11-37.

May, J. C.; Leaptrot, K. L.; **Rose, B. S.;** Moser, K. L. M.; Deng, L.; Maxon, L.; DeBord, D.; McLean, J. A. Resolving Power and Collision Cross Section Measurement Accuracy of a Prototype High-Resolution Ion Mobility Platform Incorporating Structures for Lossless Ion Manipulation. *J. Am. Soc. Mass Spectrom*. **2021**, *32*(4), 1162-1137.

Picache, J. A.; **Rose, B. S.**; Balinski, A.; Leaptrot, K. L.; Sherrod, S. D.; May, J. C.; McLean, J. A. Collision cross section compendium to filter and predict multi-omic compound identities. *Chem. Sci*. **2019**, 10, 983-993.

Nichols, C. M.; Dodds, J. N.; **Rose, B. S.**; Picache, J. A.; Codreanu, S. G.; May, J. C.; Sherrod, S. D.; McLean, J. A. Untargeted Molecular Discovery in Primary Metabolism: Collision Cross Section as a Molecular Descriptor in Ion Mobility-Mass Spectrometry. *Anal. Chem.* **2018**, 90(24), 14484-14492.

PRESENTATATIONS

Rose, B. S., Reardon, A. R.; May, J. C., McLean, J. A. Structural Characterization of Previously Unresolved Lipids with SLIM-based High Resolution Ion Mobility (HRIM)- Mass Spectrometry. Poster presented at the 70th ASMS Conference on Mass Spectrometry and Allied Topics. Minneapolis, MN, 2022.

Rose, B. S., Leaptrot, K. L., Reardon, A. R.; May, J. C., McLean, J. A. Lipid Collision Cross Section Calibration Strategies for High Resolution SLIM-based Traveling Wave Ion Mobility Separations. Poster presented at the 69th ASMS Conference on Mass Spectrometry and Allied Topics. Philadelphia, PA, 2021.

Rose, B. S., Leaptrot, K. L., Wormwood, K. L. M., May, J. C., McLean, J. A. Collision cross section calibration strategies for a prototype SLIM-based ion mobility instrument. Poster presented at the 68th ASMS Conference on Mass Spectrometry and Allied Topics. Virtual conference, 2020.

Rose, B. S.; Picache, J. A; May, J.C.; McLean, J.A. Evaluation of Surface Induced Achieving a Consistent and Unified Community-wide Molecular Database: The Collision Cross Section Compendium. Oral presented at the 1st Annual Agilent North American Ion Mobility Workshop. Raleigh, NC, 2019.

Rose, B. S.; Codreanu, S. G.; May, J. C.; Sherrod, S. D.; McLean, J.A. Integrated multidimensional liquid chromatography-ion mobility-tandem mass spectrometry (LC-IM-MS/MS) workflow for high-confidence annotations in untargeted lipidomics. Oral presented at the 67th ASMS Conference on Mass Spectrometry and Allied Topics. Atlanta, GA, 2019.

Rose, B. S.; Codreanu, S. G.; May, J. C.; Sherrod, S. D.; McLean, J.A. Development of a LC-IM-MS/MS workflow for high confidence lipidomics. Poster presented at the Southeaster Lipid Research Conference. Nashville, TN, 2018.

Rose, B. S.; Ligon, L. A.; Spence, T. G.; Cavity ring-down spectroscopy of gas-phase ions. Poster presented at the National Meeting of the American Chemical Society. San Francisco, CA, 2017.

Rose, B. S.; Colling, S.; Williams, S.; Krauss, J.; Stace, J.; Ligand exchange of an enzyme-mimic schiff-base copper(II) complex. Poster presented at the National Meeting of the American Chemical Society. San Diego, CA, 2016.

HONORS AND AWARDS

American Society for Mass Spectrometry Student Travel Award	2020
May Institute Student Travel Award	2019
Vanderbilt Institute of Chemical Biology Fellowship	2018
Vanderbilt Chemistry Mitchum E. Warren Fellowship	2017

PROFESSIONAL SOCIETIES/ORGANIZATIONS

American Society of Mass Spectrometry