

**Development and Application of Experimental and Informatic Advances to Characterize  
the Lipidomes of Diverse Microbial Systems**

A Dissertation Presented for the  
Doctor of Philosophy  
Degree  
The University of Tennessee, Knoxville

Matthew Keller  
December 2024

### **Acknowledgements:**

I first want to thank my doctoral advisor, Dr. Robert Hettich. Thank you for sticking your neck out for me and helping me to join the GST program out of the usual cycle. In your lab, I've been able to grow scientifically and personally. In many other circumstances I doubt I would have been able to be successful in my graduate studies, but thanks to your advisement I am a much more competent and confident scientist today than when I began. Thank you also for providing the funding necessary to pursue my research under a research assistantship. Thank you again for all your support.

I also want to thank the members of my Ph.D. committee for their support in this process. Thank you to Drs. Heidi Goodrich-Blair, Timothy Tschaplinski, Bode Olukolu, Hugh O'Neill, and Elizabeth Fozo. Thank you for the time you have invested in my graduate work and for the comments and suggestions to make my research better.

I want to thank many fellow graduate students I have worked with over the years. Thank you to Dr. David Reeves for his help getting started after I joined the Hettich Lab. Thank you to the other fellow Bob-Mobsters: Dr. Iván Villalobos Solis, Dr. Samantha Peters, Dr. Payal Chirania, Dr. Him Shrestha, Manasa Appidi, Katie Ostrouchov, Megan Davin, Sameer Mudbhari, Andrew Stai, and Eda Öksüz. Thank you for the friendship, suggestions, and help you've given me during my time with GST, I'm going to miss you guys. You have made working much more enjoyable than it would otherwise have been. A special thank you to Steven Tavis, I owe you a lot of credit for my growth in statistical analysis and experimental design. I also appreciate the friendship and good times we've shared over the years. I want to thank the other members of the Bioanalytical Mass Spectrometry group at ORNL, especially Drs. Paul Abraham, Richard Giannone, Jack Cahill, and Vilmos Kertész. I also want to thank Diana Ramirez for her friendship and profitable collaborations on multiple research projects. It's been a pleasure to have you as a collaborator. A big shout-out to the other folks in Bob's journal club. It's definitely been fun talking shop with you all.

I also owe a good deal to many other scientists I have collaborated with on various research projects. Thank you to Drs. Hugh O'Neill, Frank Löffler, Tomás Rush, and Allison Werner, to name a few.

I owe even more to my loving family who has supported me during my graduate studies and before. Thank you Rose for being a wonderful wife! Your love and support have, perhaps

single-handedly, kept me sane despite the rigours of grad school life. Thank you for managing the household affairs in a Josephian manner so I that have had no concerns except the food on my plate. A big thank you to my parents Calvin and Connie Keller. Humanly speaking, I owe you both everything. Without your love and care I would probably still be hitting my head into the walls and not speaking. Thank you for the patience and love expressed in homeschooling stubborn ol' me until college and setting me up well for success there (despite the roach motel you had for a "college resort"). I also want to thank the McPherson clan who have been a blessing to me. Through friendship with David, Duncan, and Andrew and Joyce McPherson's Shakespeare camp I learned how to have semi-normal interactions with other people. That may not sound like a significant achievement, but let's be honest, I was a tough case.

Most of all, though, I owe everything to the Lord God and His son, Jesus the Messiah. All these blessings have been from your hand, but more than all these you have given me eternal life through the sacrifice of Jesus. I thank you for the time you've given me to study your amazing creation in graduate school. Whatever is well done and scientifically sound, of my dissertation I dedicate to my Lord.

## **Abstract**

Lipidomics concerns itself with the large-scale measurement of lipids. This burgeoning field, while showing promise, faces several significant technical, informatic, and biological challenges that severely limit its performance. These challenges are partially driven by the limited benchmarking and thorough validation studies associated with most new fields, as well as intrinsic difficulties in lipid analyses. These include uncertainties in extraction and sample preparation, inaccurate automated lipid identification/quantification software, and poor statistical and interpretive practices. In my doctoral research, I have sought to improve upon the groundwork others have laid in lipidomics by developing tools and analyses that will provide a firmer foundation, which can be utilized by research yet to come. Along the way, I have also applied lipidomics to address a variety of important biological questions such as the effect of lipo-chitooglucosaccharides on lipid signaling in fungi, the covalent incorporation of polyfluorinated alkyl substances into bacterial lipids, and the design of *in vivo* deuterated lipids for neutron scattering.

## Table of Contents

### Chapter I - Lipidomics: A Brief Introduction with a Particular Focus on Experimental

|                                                                                              |    |
|----------------------------------------------------------------------------------------------|----|
| Methodology .....                                                                            | 1  |
| The What, Why, and How of Lipidomics: .....                                                  | 1  |
| The Biology of Lipids .....                                                                  | 1  |
| Lipidomics .....                                                                             | 2  |
| Experimental Techniques for Lipid Measurement: .....                                         | 4  |
| Planar Chromatography .....                                                                  | 4  |
| Nuclear Magnetic Resonance .....                                                             | 5  |
| Mass Spectrometry .....                                                                      | 5  |
| Liquid Chromatography-Mass Spectrometry Lipidomics .....                                     | 7  |
| Automated Analysis of Untargeted LC-MS Lipidomics Data .....                                 | 9  |
| Selection and Development of My LC-MS Lipidomics Method: .....                               | 10 |
| Selection of liquid chromatography-mass spectrometry as the analytical platform .....        | 10 |
| Original LC-MS method identified and subsequent improvements .....                           | 12 |
| Selection and Operation of Mass Spectrometer for Untargeted Lipidomics .....                 | 14 |
| Lipid Extraction .....                                                                       | 16 |
| Areas in Lipidomics Where Further Development is Needed: .....                               | 17 |
| Summary of my Doctoral Research: .....                                                       | 18 |
| Machine Learning for Rapid and Reliable Lipid Identification in Untargeted Lipidomics ..     | 19 |
| Empirical Assessment of the Repeatability of Untargeted Lipidomics and Proteomics            |    |
| Experiments .....                                                                            | 19 |
| Non-model Fungal Lipidomics and Lipo-chitooglucosaccharide Signaling Molecules .....         | 20 |
| In vivo Deuteration of Bacterial Phospholipids .....                                         | 20 |
| Covalent Incorporation of PFAS Into Bacterial Phospholipids .....                            | 21 |
| A Multi-Omics Investigation of <i>Clostridium autoethanogenum</i> 's Cultivation Temperature | 22 |
| Summary .....                                                                                | 22 |
| Chapter II - LipoCLEAN: The Development of a Machine Learning-Based Rescoring Method         |    |
| for Confident Assignment of Lipid Identifications .....                                      | 23 |
| Citation: .....                                                                              | 24 |
| Context: .....                                                                               | 24 |

|                                                                                                                          |    |
|--------------------------------------------------------------------------------------------------------------------------|----|
| Abstract: .....                                                                                                          | 25 |
| Introduction:.....                                                                                                       | 26 |
| Materials and Methods: .....                                                                                             | 27 |
| Manual Classification Strategy.....                                                                                      | 28 |
| Lipid Tandem Mass Spectrometry Fragmentation Analysis.....                                                               | 28 |
| Datasets Used .....                                                                                                      | 32 |
| Software Architecture .....                                                                                              | 32 |
| Model Development .....                                                                                                  | 33 |
| Instrument Cross Validation .....                                                                                        | 33 |
| Lipid Class Cross Validation .....                                                                                       | 33 |
| Model Training Stability Test .....                                                                                      | 33 |
| Using LipoCLEAN.....                                                                                                     | 34 |
| Results and Discussion: .....                                                                                            | 34 |
| Conclusions:.....                                                                                                        | 37 |
| Chapter III - A Critical Examination of the Statistical Properties of Lipidomics Data .....                              | 38 |
| Citation: .....                                                                                                          | 39 |
| Abstract:.....                                                                                                           | 39 |
| Introduction:.....                                                                                                       | 39 |
| Materials and Methods: .....                                                                                             | 41 |
| Bacterial Culture and Sampling.....                                                                                      | 41 |
| Lipid Extraction Procedure.....                                                                                          | 42 |
| Lipidomics Instrumental Analysis .....                                                                                   | 42 |
| Lipid Identification and Quantitation .....                                                                              | 43 |
| Data Processing .....                                                                                                    | 43 |
| Results and Discussion: .....                                                                                            | 44 |
| Overall Lipid Composition.....                                                                                           | 44 |
| Lipidome Summary Statistics.....                                                                                         | 46 |
| Assessment of Quantification Method .....                                                                                | 48 |
| Conclusions:.....                                                                                                        | 48 |
| Chapter IV - Non-model fungal lipidomics reveals lipo-chitooglucosaccharides selectively preserve native lipidomes ..... | 50 |

|                                                                                                                                                                             |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Citation: .....                                                                                                                                                             | 51 |
| Abstract: .....                                                                                                                                                             | 51 |
| Introduction:.....                                                                                                                                                          | 51 |
| Aspergillus and Laccaria lipidomes .....                                                                                                                                    | 51 |
| Mycology of lipo-chitooligosaccharides .....                                                                                                                                | 52 |
| Methods: .....                                                                                                                                                              | 53 |
| Fungal Growth.....                                                                                                                                                          | 53 |
| Lipid extraction and measurement by LC-MS .....                                                                                                                             | 54 |
| Data Analysis .....                                                                                                                                                         | 56 |
| Results:.....                                                                                                                                                               | 57 |
| Summary of Lipidomics Results Across Datasets .....                                                                                                                         | 57 |
| Testing (L)CO-Modulation of Lipidomes .....                                                                                                                                 | 58 |
| Discussion: .....                                                                                                                                                           | 58 |
| Native Lipidome of Aspergillus fumigatus and Laccaria bicolor .....                                                                                                         | 58 |
| Effect of Temperature on A. fumigatus Lipids .....                                                                                                                          | 59 |
| Impact and Concluding Thoughts .....                                                                                                                                        | 62 |
| Chapter V - Deuterated Lipids: A Comprehensive Examination of Lipidome-Wide Deuteration<br>and the Principals of Designing Regiospecifically Deuterated Phospholipids ..... | 63 |
| Part 1. Characterization of the <i>in vivo</i> deuteration of native phospholipids by mass spectrometry<br>yields guidelines for their regiospecific customization: .....   | 65 |
| Citation: .....                                                                                                                                                             | 65 |
| Abstract: .....                                                                                                                                                             | 65 |
| Introduction:.....                                                                                                                                                          | 66 |
| Materials and Methods: .....                                                                                                                                                | 67 |
| Growth of E. coli with Labeled Lipids .....                                                                                                                                 | 67 |
| Enfors media preparation .....                                                                                                                                              | 68 |
| Lipid Extraction Procedure.....                                                                                                                                             | 68 |
| Lipidomic LC-MS Analysis .....                                                                                                                                              | 69 |
| LC-MS Data Analysis.....                                                                                                                                                    | 69 |
| Lipid Extraction Procedure for SANS Experiments .....                                                                                                                       | 71 |
| Analysis of Lipid Deuteration .....                                                                                                                                         | 71 |

|                                                                                                                                                     |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Small Angle Neutron Scattering (SANS).....                                                                                                          | 72 |
| Results and Discussion: .....                                                                                                                       | 72 |
| Lipid Identifications .....                                                                                                                         | 72 |
| Deuterium Incorporation and Profiling .....                                                                                                         | 73 |
| Lipidome Scale Deuterium Incorporation .....                                                                                                        | 74 |
| Deuterium Localization .....                                                                                                                        | 75 |
| Hydrogen Sourcing within Lipid Biosynthesis .....                                                                                                   | 76 |
| Prediction of Lipid Deuteration.....                                                                                                                | 77 |
| SANS analysis of Lipid Extracts .....                                                                                                               | 78 |
| Conclusions:.....                                                                                                                                   | 80 |
| Part 2. Bioproduction and small-angle neutron scattering characterization of deuterated<br>phosphatidylcholine from <i>Escherichia coli</i> : ..... | 80 |
| Citation: .....                                                                                                                                     | 80 |
| Abstract: .....                                                                                                                                     | 80 |
| Introduction:.....                                                                                                                                  | 81 |
| Material and Methods: .....                                                                                                                         | 83 |
| Chemicals .....                                                                                                                                     | 83 |
| Strains and culture conditions .....                                                                                                                | 83 |
| Separation of total lipid extract .....                                                                                                             | 83 |
| Fraction of total lipid extract for PC purification.....                                                                                            | 84 |
| Chemical synthesis of D-choline.....                                                                                                                | 84 |
| Mass spectrometric characterization of lipid samples.....                                                                                           | 85 |
| Lipid Identification.....                                                                                                                           | 85 |
| Analysis of Lipid Deuteration .....                                                                                                                 | 86 |
| Lipid Class Abundance Calculation .....                                                                                                             | 86 |
| SANS analysis of lipid samples .....                                                                                                                | 87 |
| Results and Discussion: .....                                                                                                                       | 87 |
| Characterization of PC in total lipid extract.....                                                                                                  | 87 |
| Chain distribution in PC .....                                                                                                                      | 88 |
| Control of deuterium labeling level and deuterium labeling position.....                                                                            | 88 |
| Perdeuterated choline and PC.....                                                                                                                   | 89 |

|                                                                                                                     |     |
|---------------------------------------------------------------------------------------------------------------------|-----|
| Small-Angle Neutron Scattering Analysis.....                                                                        | 90  |
| Conclusions: .....                                                                                                  | 91  |
| Chapter VI - First Report of the Covalent Incorporation of Polyfluorinated Alkyl Substances into Phospholipids..... | 93  |
| Citation: .....                                                                                                     | 94  |
| Context:.....                                                                                                       | 94  |
| Abstract:.....                                                                                                      | 94  |
| Introduction:.....                                                                                                  | 95  |
| Materials and Methods: .....                                                                                        | 96  |
| Cultivation .....                                                                                                   | 96  |
| Lipid Extraction and Lipidomic Analysis.....                                                                        | 96  |
| Data Processing .....                                                                                               | 97  |
| Calculation of a PFAS Membrane Incorporation Factor (MIF) .....                                                     | 97  |
| Results and Discussion: .....                                                                                       | 98  |
| Identification of PFASylated lipids in <i>Pseudomonas</i> sp. strain 273 .....                                      | 98  |
| Overall FTCA incorporation.....                                                                                     | 100 |
| PFAS incorporation observed in other organisms .....                                                                | 101 |
| Investigation of substrate range for covalent incorporation of FTCA in strain 273.....                              | 101 |
| Limit of detectable FTCA incorporation .....                                                                        | 102 |
| Formation of Fluoromembranes Potentially Impacts Environmental PFAS Pools and Bioaccumulation.....                  | 102 |
| Future Directions of Fluorinated Lipid Research: .....                                                              | 104 |
| Concluding Thoughts .....                                                                                           | 104 |
| Chapter VII - A Multi-Omic Investigation of the Effects of Temperature on the Bioreactor                            |     |
| Cultivation of <i>Clostridium autoethanogenum</i> .....                                                             | 105 |
| Citation: .....                                                                                                     | 106 |
| Context:.....                                                                                                       | 106 |
| Abstract:.....                                                                                                      | 107 |
| Introduction:.....                                                                                                  | 107 |
| Materials and Methods: .....                                                                                        | 109 |
| <i>Clostridium autoethanogenum</i> fermentations, sampling, and product determination.....                          | 109 |

|                                                                                                                               |     |
|-------------------------------------------------------------------------------------------------------------------------------|-----|
| GC-MS based metabolomics measurements, data normalization, and analysis .....                                                 | 110 |
| Proteomics sample preparation and measurement by reversed phase chromatography and tandem mass spectrometry .....             | 110 |
| Proteomics data analysis.....                                                                                                 | 112 |
| MTBE-based extraction of <i>C. autoethanogenum</i> lipids .....                                                               | 113 |
| Measurement of lipids by Reversed Phase-based liquid chromatography mass spectrometry .....                                   | 114 |
| Lipidomics data analysis .....                                                                                                | 114 |
| Results and Discussion: .....                                                                                                 | 115 |
| The gas fermenter <i>C. autoethanogenum</i> shifts from ethanol to acetate production at higher growth temperature.....       | 115 |
| Proteomics and metabolomics reveal temperature-induced metabolic changes in <i>C. autoethanogenum</i> .....                   | 116 |
| Examination of enzymes in the central metabolism pathways of <i>C. autoethanogenum</i> .....                                  | 117 |
| Accumulation of fatty acids and their glycerol conjugated forms occurs at 40 °C .....                                         | 119 |
| Integrated analysis of lipidome and proteome uncovers membrane remodeling with potential glycolipid production at 40 °C ..... | 121 |
| Temperature induces expression of bacterial microcompartments in <i>C. autoethanogenum</i>                                    | 123 |
| Conclusions:.....                                                                                                             | 125 |
| Chapter VIII - Conclusions and future perspectives .....                                                                      | 126 |
| Summary of Research Achievements: .....                                                                                       | 126 |
| Lipidomics Research Needs .....                                                                                               | 126 |
| Instrumental Developments: .....                                                                                              | 127 |
| Ion Mobility .....                                                                                                            | 127 |
| Improved Fragmentation Technology .....                                                                                       | 128 |
| Experimental Rigour:.....                                                                                                     | 129 |
| References:.....                                                                                                              | 132 |
| Appendix:.....                                                                                                                | 165 |
| Vita.....                                                                                                                     | 269 |

## List of Tables

|                                                                                                                                                               |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Table 2-1. The datasets that were manually annotated for model training and validation. ....                                                                  | 186 |
| Table 2-2: The number of manually annotated putative lipid IDs per class within each dataset. ....                                                            | 188 |
| Table 3-1. Chromatography conditions used in the lipidomics replicability experiment. ....                                                                    | 195 |
| Table 3-2. Normality test of lipid summary statistics .....                                                                                                   | 201 |
| Table 3-3. Comparison of quantification method precision. ....                                                                                                | 206 |
| Table 4-1. Chromatography conditions used in the lipidomics replicability experiment .....                                                                    | 208 |
| Table 4-2. Number of Lipid Species per Lipid Class .....                                                                                                      | 208 |
| Table 4-3. Average Carbon Number and Unsaturation by Lipid Class .....                                                                                        | 209 |
| Table 4-4. P-values obtained from ANOVA tests on lipidome summary statistics .....                                                                            | 213 |
| Table 4-5. Results from PERMANOVA.....                                                                                                                        | 213 |
| Table 5-1. Recipe for one liter of Enfors media. ....                                                                                                         | 217 |
| Table 5-2. Sample ID and growth conditions.....                                                                                                               | 217 |
| Table 5-3. LC gradient. ....                                                                                                                                  | 218 |
| Table 5-4. Average lipid component volumes used for SLD calculation. ....                                                                                     | 219 |
| Table 5-5. Best fit deuterium probabilities ( $p_D$ ) for PE lipids. ....                                                                                     | 222 |
| Table 5-6. Best fit deuterium probabilities ( $p_D$ ) for PG lipids.....                                                                                      | 223 |
| Table 5-7. Deuterium incorporation (best fitting $p_D$ distribution) specific to fatty acids.....                                                             | 224 |
| Table 5-8. Comparison of fatty acid specific deuterium incorporation calculation. ....                                                                        | 230 |
| Table 5-9. Lipid bilayer thickness at different $D_2O$ concentrations. ....                                                                                   | 231 |
| Table 5-10. Relative abundance of phospholipids used for CMP calculation.....                                                                                 | 234 |
| Table 5-11. Contrast match point from experimental SANS data and MS data.....                                                                                 | 235 |
| Table 5-12. Calculated CMP at different solvents based on headgroup deuteration level. ....                                                                   | 235 |
| Table 6-1. Orbitrap tandem MS of PE 8:3 FTCA_18:1 .....                                                                                                       | 245 |
| Table 6-2. Comparison of PFAS membrane incorporation factors determined for different bacterial taxa<br>cultivated in the presence of 5 $\mu M$ 8:3 FTCA..... | 251 |
| Table 7-1. Proteins with the highest fold changes with temperature. ....                                                                                      | 264 |
| Table 8-1. Replication Rates of Cancer Biology Research. ....                                                                                                 | 268 |

## List of Figures

|                                                                                                         |     |
|---------------------------------------------------------------------------------------------------------|-----|
| Figure 1-1. Representative 2-D Lipid TLC.....                                                           | 165 |
| Figure 1-2. Elution profiles of related phosphatidylglycerol lipids in reversed phase chromatography... | 166 |
| Figure 1-3. Evolution of RPLC Gradients Employed.....                                                   | 167 |
| Figure 1-4. Cross section of an orbitrap mass analyzer. ....                                            | 168 |
| Figure 1-5. LTQ-Orbitrap Block Diagram. ....                                                            | 169 |
| Figure 1-6. Injection Volume Solvent Overload .....                                                     | 170 |
| Figure 2-1. PE 35:4 fragmentation spectrum. ....                                                        | 171 |
| Figure 2-2. Manual validation of MS-DIAL lipid IDs.....                                                 | 172 |
| Figure 2-3. The manual annotation workflow.....                                                         | 174 |
| Figure 2-4. Fragment ions of PG lipids.....                                                             | 175 |
| Figure 2-5. Example fragmentation spectrum of PG 16:0_16:0. ....                                        | 175 |
| Figure 2-6. Fragment ions of PE lipids. ....                                                            | 176 |
| Figure 2-7. Example fragmentation spectrum of PE 16:0_16:0 .....                                        | 176 |
| Figure 2-8. Fragment ions of PC lipids.....                                                             | 177 |
| Figure 2-9. Example fragmentation spectrum of PC 18:1_18:1. ....                                        | 177 |
| Figure 2-10. Fragment ions of PS lipids. ....                                                           | 178 |
| Figure 2-11. Example fragmentation spectrum of PS 18:2_18:2.....                                        | 178 |
| Figure 2-12. Fragment ions of PI lipids. ....                                                           | 179 |
| Figure 2-13. Example fragmentation spectrum of PI 18:2_18:2.....                                        | 180 |
| Figure 2-14. Example fragment ions of a carnitine lipid. ....                                           | 180 |
| Figure 2-15. QE Fragmentation Spectrum of Car 16:0.....                                                 | 181 |
| Figure 2-16. Example Fragmentation Ions of a TAG 16:0_16:0_16:0. ....                                   | 182 |
| Figure 2-17. Fragmentation Spectrum of a TAG Lipid.....                                                 | 183 |
| Figure 2-18. Fragmentation Ion Structures for Cer Lipids. ....                                          | 184 |
| Figure 2-19. The architecture of the tool LipoCLEAN.....                                                | 185 |
| Figure 2-20. Model performance on the validation dataset. ....                                          | 186 |
| Figure 2-21. Assessment of the repeatability of the model training process.....                         | 187 |
| Figure 2-22. Assessing generalization to new instruments. ....                                          | 189 |
| Figure 2-23. Investigating the incompatibility between TOF and Orbitrap data. ....                      | 190 |
| Figure 2-24. Investigating the compatibility between Q Exactive and Orbitrap Velos Pro data.....        | 191 |
| Figure 2-25. Assessing generalization to new lipid classes. ....                                        | 192 |
| Figure 2-26. Investigating poor generalization performance. ....                                        | 193 |
| Figure 3-1. P-values obtained with and without log-transformation of the raw data.....                  | 194 |

|                                                                                                     |     |
|-----------------------------------------------------------------------------------------------------|-----|
| Figure 3-2. Proportion of significant analytes that were only significant in one operation. ....    | 194 |
| Figure 3-3. Example of instrumental drift and LOWESS correction. ....                               | 195 |
| Figure 3-4. The effect of log-transformation on distributional properties of lipidomic data. ....   | 196 |
| Figure 3-5. P-value distribution obtained from lipidomics T-test empirical null. ....               | 197 |
| Figure 3-6. P-value distribution obtained from lipidomics permutational T-test empirical null. .... | 198 |
| Figure 3-7. Power based on LOWESS of FC vs frequency of significant plot at given BH cutoff. ....   | 199 |
| Figure 3-8. Observed fold changes (control/solvent stress) from <i>P. putida</i> . ....             | 200 |
| Figure 3-9. Summary Statistic Q-Q Plots. ....                                                       | 202 |
| Figure 3-10. Empirical null for summary statistics. ....                                            | 203 |
| Figure 3-11. Examples of Misintegration Performed by MS-DIAL. ....                                  | 204 |
| Figure 3-12. Examples of Misintegration Performed by XCMS. ....                                     | 205 |
| Figure 3-13. Distributional properties from different quantification methods. ....                  | 207 |
| Figure 4-1. Lipid Class Fatty Acid Characteristics. ....                                            | 210 |
| Figure 4-2. Ceramide lipids between different temperatures. ....                                    | 211 |
| Figure 4-3. Proportions of different lipid classes in the three datasets. ....                      | 212 |
| Figure 4-4. PCA of the lipidomics data for <i>L. bicolor</i> . ....                                 | 214 |
| Figure 4-5. PCA of the lipidomics data for <i>A. fumigatus</i> at 37 °C. ....                       | 215 |
| Figure 4-6. PCA of the lipidomics data for <i>A. fumigatus</i> at 25 °C. ....                       | 216 |
| Figure 5-1. Structure and fragmentation of example PE lipid. ....                                   | 218 |
| Figure 5-2. Retention time regression of identified lipids. ....                                    | 219 |
| Figure 5-3. Lipid abundances of PE and PG lipids in the unlabeled sample. ....                      | 220 |
| Figure 5-4. Example mass spectra and isotope fitting. ....                                          | 221 |
| Figure 5-5. Best fit deuterium percent for individual lipids. ....                                  | 225 |
| Figure 5-6. Weighted average deuterium incorporations. ....                                         | 226 |
| Figure 5-7. Lipid deuteration linear regressions. ....                                              | 227 |
| Figure 5-8. Deuterium incorporation polynomial regressions. ....                                    | 228 |
| Figure 5-9. PG Isotope Localization for Selected Lipids in H-Gly/100 and D-Gly/0. ....              | 229 |
| Figure 5-10. Polynomial regression of D <sub>2</sub> O and fatty acid deuteration. ....             | 230 |
| Figure 5-11. Fatty acid II synthesis scheme. ....                                                   | 231 |
| Figure 5-12. SANS contrast series of <i>E coli</i> at three different conditions. ....              | 232 |
| Figure 5-13. Contrast matching point calculated over the Q range. ....                              | 233 |
| Figure 5-14. Demonstration of PC production from engineered <i>E coli</i> . ....                    | 236 |
| Figure 5-15. Relative abundance of PC lipids identified in all culture conditions. ....             | 236 |
| Figure 5-16. Proton NMR spectra of isolated PC with different deuteration levels. ....              | 237 |

|                                                                                                                                       |     |
|---------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 5-17. SAXS data for deuterium labeled PC lipids.....                                                                           | 237 |
| Figure 5-18. SANS data and contrast match point analysis for deuterium labeled PC lipids.....                                         | 238 |
| Figure 5-19. <i>Q</i> -dependency of neutron contrast match point for non- and isotopically labeled phospholipids forms.....          | 239 |
| Figure 6-1. Structure and fragmentation of representative PE lipids containing up to two 8:2 FTCA as FA tails.....                    | 240 |
| Figure 6-2. Expected ion structures for the tandem MS of PE 34:1 .....                                                                | 241 |
| Figure 6-3. Expected ion structures for the tandem MS of PE 18:1_8:3 FTCA .....                                                       | 242 |
| Figure 6-4. HRAM tandem MS spectrum of free 8:3 FTCA.....                                                                             | 243 |
| Figure 6-5. Expected ion structures for the tandem MS of PE 18:1_8:3 FTCA .....                                                       | 244 |
| Figure 6-6. HRAM tandem MS of FTCA-containing PE Lipids.....                                                                          | 246 |
| Figure 6-7. HRAM tandem MS of FTCA-containing PG Lipids .....                                                                         | 247 |
| Figure 6-8. Possible structures for the FTCA fragment ions observed in the tandem MS spectra. ....                                    | 248 |
| Figure 6-9. Relative compositions of PE lipids. ....                                                                                  | 249 |
| Figure 6-10. Relative compositions of PG lipids. ....                                                                                 | 250 |
| Figure 6-11. Schematic illustrating the covalent incorporation of aqueous PFAS (i.e., 8:3 FTCA) into the bacterial lipid bilayer..... | 252 |
| Figure 6-12. Tandem MS of PFASylated lipids with 5:3, 7:3, 8:3, and 10:2 FTCA.....                                                    | 253 |
| Figure 6-13. PFASylated lipid signal as a function of FTCA concentration.....                                                         | 254 |
| Figure 7-1. End fermentation-product profile of <i>C. autoethanogenum</i> at 30 and 40 °C. ....                                       | 255 |
| Figure 7-2. Abundance changes in the proteome and metabolome drive growth at the two temperatures. ....                               | 256 |
| Figure 7-3. Proteome correlation heatmap.....                                                                                         | 257 |
| Figure 7-4. Central metabolic pathways in <i>C. autoethanogenum</i> . ....                                                            | 258 |
| Figure 7-5. Proteome KEGG pathway enrichment analysis. ....                                                                           | 259 |
| Figure 7-6. Fatty acids accumulate at high temperature with changes in fatty acid metabolism.....                                     | 260 |
| Figure 7-7. Lipidomics results and lipid proteome fold-changes. ....                                                                  | 261 |
| Figure 7-8. Anaerobic plasmalogen homolog. ....                                                                                       | 262 |
| Figure 7-9. Glycosyl transferases fold-changes. ....                                                                                  | 263 |
| Figure 7-10. Bacterial Microcompartment related proteins. ....                                                                        | 265 |
| Figure 8-1. Mass-to-charge collisional cross section relationship. ....                                                               | 266 |
| Figure 8-2. TIMS cell diagram .....                                                                                                   | 267 |

## **Chapter I - Lipidomics: A Brief Introduction with a Particular Focus on Experimental Methodology**

### **The What, Why, and How of Lipidomics:**

#### *The Biology of Lipids*

Lipids, as a class of biomolecules, can be notoriously difficult to define. Many metabolite classes are defined by shared biosynthetic pathways or chemical functional groups. However, lipids, as commonly understood, not only include lipid synthase-derived fatty acid-based compounds (phospholipids, glycerolipids, *et cetera*) but also include polyketides, sphingosines, and sterol-derivatives with their own distinct biosynthetic routes. Defining lipids by function is similarly challenging. The functions of lipids are myriad, including structural, energy, and signaling roles. Yet despite this, there is a general understanding of what lipids are. The most distinctive characteristic of lipids is their polarity. Lipids are generally non-polar with positive logP values (making them hydrophobic); however, a strict definition based on logP values does not exist, nor would it be helpful. For example, para-coumaric acid, clearly a “metabolite” in contradistinction from lipids, has a logP of 1.5<sup>1</sup> (predicted by XLogP3-AA<sup>2</sup>) while lyso-phosphatidylethanolamine, clearly a lipid, only has a logP of 0.3.<sup>3</sup> Rather, lipids should be considered as low-molecular weight biological compounds with generally low polarities that are often derived from a few chemical families. While clunky, this description underscores the fact that lipids are a class more easily recognized than defined.

The biological roles of lipids are as varied as the class is disparate. One obvious function of lipids is as the building-blocks of cell membranes. Glycerophospholipids form the major constituent of the plasma membrane across almost all kingdoms of life. These lipids have polar head-groups and non-polar fatty acid tails which will spontaneously form into bilayers. These bilayers provide a segregated internal environment for a living cell so that harmful environmental substances can be excluded while retaining high-value compounds internally (while porins can be used for necessary exchanges). Membranes, including subcellular organelles, also provide a substrate in which proteins can be embedded. This can help provide structural integrity for membrane proteins as well as spatially constraining enzymes and/or substrates. For these structural roles to be met, the physical properties of the bilayer must be well managed. If the lipids should completely melt, the membrane can be lost and the cell destroyed. On the other hand, if the membrane solidifies, the permeability of the membrane and

functionality of membrane proteins will be impaired and the cell can die anyway. Temperature and other physiochemical stressors can affect membrane integrity, forcing organisms to employ a variety of mechanisms to preserve an ideal membrane composition. This is known as homeoviscous adaptation and includes adaptation of lipids to changing environments (tail length or unsaturation) and production of lipid mixtures that naturally broaden the phase transitions (mixing sterols with phospholipids). In addition to this “simple” structural role, some lipids have important signaling functions, especially in multicellular organisms. Phosphatidylinositol lipids, for example, are involved in various intracellular regulatory pathways in many organisms such as the modulation of protein kinase B. Similarly, various sterol derivatives (such as testosterone and estrogens) are critical intercellular signaling compounds and need no introduction. Finally, a more prosaic, though still important, function of lipids is as energy storage compounds, especially for acylglycerol lipids. With this array of important functions, it is clear that lipids are a class of biomolecules worth investigating. That said, while some lipids may provide actionable information even when considered in isolation (such as testosterone), others require a more holistic view to discern the biological state, especially for membrane and storage lipids. Similarly, resolution of the lipid species may or may not be necessary to address the biological question. For example, bulk triacylglycerol-class concentration is sufficient to predict disease risks while ceramide-induced endoplasmic reticulum stress is highly dependent on the chain length of ceramides.<sup>4</sup> But how shall these be measured? What analytical platforms are available for lipid analysis, and what will allow for accurate lipid measurement on a large scale?

### *Lipidomics*

There are a great variety of lipid analytical techniques and experimental designs, and lipid measurements themselves are nothing particularly new. However, with technological developments of the 1980s and 1990s, especially in mass spectrometry, a slightly new conception of lipid analysis came about.<sup>5</sup> This branch eventually took the name “lipidomics”, a moniker that has come to denote the large-scale analysis of lipids.<sup>5</sup> Lipidomics has been established as an important tool for biological research and can be found in applications covering diverse fields such as structural biology, microbiology, biotechnology, and epidemiology.<sup>4-17</sup>

An early example of lipidomics can be seen in a 2007 lipidomic investigation of traumatic brain injury (in rats) with specific focus on oxidized lipids.<sup>18</sup> In this work, thin layer chromatography (TLC) was used for lipid separation. TLC spots corresponding to different lipid

classes were scraped off for bulk class quantification as well as analysis by direct infusion tandem mass spectrometry. This allowed high abundance lipid species to be identified by mass spectrometry while also providing an overall class analysis. Through this process, the authors report the identification 73 lipid molecular species, covering 8 lipid classes. This is a robust and detailed analysis, and probably represents one of the better lipidomics studies of its time. The authors characterized lipid oxidation following traumatic brain injury and found that cardiolipin was the first phospholipid class to be perturbed following traumatic brain injury, followed later by phosphatidylserine, phosphatidylethanolamine, and phosphatidylcholine.

A more recent example (published in 2023) of lipidomics involved the study of atopic dermatitis in mouse models.<sup>19</sup> Here an untargeted liquid chromatography-mass spectrometry method was employed that, when combined with automated analysis software, resulted in 745 reported lipid identifications. This represents an order of magnitude increase in lipid coverage since 2007, although the automated analysis number may be somewhat inflated with false positives.<sup>20</sup> Nevertheless, this untargeted analysis indicated certain types of ceramide lipids were negatively associated with atopic dermatitis. Through follow-up experiments, including the targeted LC-MS lipidomics of 54 lipids, the authors were able to demonstrate the existence of a causal relationship between very long-chain (C22 or more) nonhydroxy-fatty acid-containing dihydrosphingosines and atopic dermatitis, thereby opening possible therapeutic options.<sup>19</sup> These findings may not have been achievable without high performance lipidomics experimentation, underscoring the utility of lipidomics and the importance of using accurate experimental techniques.

While a few years old (from 2017), a paper describing an interlaboratory test of quantitative lipidomics workflows gives a good sense of the scope and characteristics of lipidomics as currently practiced.<sup>21</sup> In this work, 31 laboratories participated in a comprehensive, quantitative analysis of the National Institute of Standards and Technology's standard reference material for human plasma. The project sought to quantify as many lipids as possible in the plasma reference material and provide a benchmark for future reference. Across all 31 labs, a total of 1527 unique lipids were measured.<sup>21</sup> However, limiting this to consistently quantified lipids for which a consensus value was achieved drops the number to 339 lipid species.<sup>21</sup> These species covered more than 12 lipid classes (including sterols, ceramides, glycerolipids, and glycerophospholipids) which underscores the analytical challenge in lipidomics. The

measurement must have adequate specificity to provide discrimination between similar lipids while capturing a large range of different lipid types. This interlaboratory comparison demonstrates the nontrivial scope and challenge of modern lipidomics both in terms of the chemical diversity and consistency of replication (consensus lipids were only 22% of the total reported between the different labs). It is worth noting that the number of identifiable lipids is a function of the organism of interest. While 300+ may be reasonable for humans, other organisms have greatly reduced lipid diversity. *E. coli*, for example, is mostly restricted to three lipid classes (phosphatidylglycerol, phosphatidylethanolamine, and cardiolipin) and therefore has many fewer unique lipid species.<sup>22-25</sup> Thus, approaches that are suitable for the low complexity *E. coli* may not translate well to higher complexity systems. Of note, this interlaboratory comparison used exclusively mass spectrometry-based methodologies. MS has become by far the dominant analytical tool for cutting-edge lipidomics measurements. Indeed, established and well-respected leaders in lipidomics research, like Oliver Fiehn, Erin Baker, Facundo Fernandez, and many others, all use LC-MS based techniques.<sup>16, 17, 26-29</sup> In fact, mass spectrometry has become so ubiquitous among untargeted lipidomics profiling that the Lipidomics Standards Initiative's interactive Minimal Reporting Checklist does not have an option for non-MS experiments. The natural question when considering this is what makes LC-MS ideal for lipidomics? Or perhaps the consensus is misguided, and another approach could be equally or more useful than LC-MS? We will consider these topics in the following section.

### **Experimental Techniques for Lipid Measurement:**

#### *Planar Chromatography*

One of the older family of techniques for large-scale lipid measurements is planar chromatography. Thin-layer chromatography (TLC) is possibly the most well-known planar technique and has been used widely for lipid measurement. This is a simple approach to generate a rough measurement of bulk lipid classes.<sup>30</sup> The lipids are adsorbed to a paper or other planar substrate (TLC uses a functionalized matrix bound to a solid support). This is then “developed” by the flow of solvent across the substrate which will “pull” along the lipids at different rates depending on their relative affinities to the substrate vs the solvent. **Figure 1-1** shows the result of a two-dimensional TLC lipid analysis as a representative example of what is possible with this technique (all tables and figures are located in the appendix). The presence of lipids can be determined by treating the TLC plates with a variety of possible reagents such as iodine or

molybdenum.<sup>30</sup> Iodine complexes with double bonds in the lipids, making it fairly non-specific but not applicable to fully saturated lipids.<sup>30</sup> Molybdenum, on the other hand, binds to phosphate esters, making it general to phospholipids and without as significant bias.<sup>31</sup> This can also be used for approximate quantification of lipid classes, based on the size and darkness of the stain, similar to blotting techniques in biological research. This is a versatile technique that is robust and time efficient and is likely to remain in use for the foreseeable future. However, TLC is not truly capable of achieving the goals of lipidomics. Its major limitation is the inability to provide specificity beyond lipid classes. All that can be reasonably assessed is bulk class abundances, and the quantitative precision of this technique is only marginal. Additionally, in many cases, full resolution of different lipid classes cannot be achieved, even when two-dimensional TLC is employed.

### *Nuclear Magnetic Resonance*

While never broadly adopted, nuclear magnetic resonance (NMR) spectroscopy has found some use for lipid analyses. NMR is a measurement technique based on the spin states of nuclei in high magnetic fields. This unique measurement principle means it is the nuclei of lipids that are being measured. Because of this, the NMR response of different lipids can be directly related to the molar ratio of atoms belonging to those lipids. The drawback of this is that there are negligible differences between the nuclei of lipids belonging to the same class, prohibiting lipid species resolution. Practically then, NMR tends to be used as a bulk lipid class measurement technique, perhaps by quantifying different phospholipid classes with <sup>31</sup>P NMR.<sup>32</sup> <sup>33</sup> This works similarly to TLC but with more accurate and precise quantitation. NMR also has relatively high detection limits, in the range of micromolar.<sup>33</sup> Perhaps the biggest limitation to the application of NMR to lipidomics research is its prohibitive cost, both in the initial purchase and the operating costs. Unless the infrastructure is already present, investing in NMR for bulk class analyses is not viable.

### *Mass Spectrometry*

Another lipid assay with a notable history is fatty-acid methyl ester (FAMES) analysis by gas chromatography-mass spectrometry. In some ways, this is the inverse of TLC methods in that FAMES provides quantitative measurements of fatty acid tail lengths while losing all lipid class/headgroup information. This analysis involves the cleavage of fatty acid tails from their constituent lipids and the formation of methyl esters. This can easily be performed via the

treatment of a lipid extract with methanolic sulfuric acid.<sup>34</sup> These resulting fatty acid methyl esters are then separated via gas chromatography before measurement by mass spectrometry. This provides a quantitative readout of the individual fatty acid species and can be useful in elucidating homeoviscous adaptation of the lipidome. However, due to the cleavage of fatty acids from their headgroups, the FAMES measurement is a convolution of the fatty acids of all lipids present in the sample. This significantly limits the interpretability of the results of most FAMES analyses as fatty acid changes cannot be linked to specific lipid classes. Because of this focus, FAMES is often used in oil research where fatty acids are of primary importance (biodiesel, for instance),<sup>35-37</sup> but is found in other applications as well.<sup>38, 39</sup>

Another mass spectrometry-based option is known as shotgun lipidomics. Shotgun lipidomics refers to the direct infusion (no separation) of lipid extracts into a mass spectrometer. Lipid species would then be identified based simply on the observed  $m/z$  values, or with some amount of additional tandem mass spectrometry.<sup>40, 41</sup> Shotgun or direct infusion lipidomics was one of the first techniques to be associated with the term lipidomics. This is likely due to the fact shotgun lipidomics is the first method capable of measuring many specific lipid species at the same time rather than bulk or average analyses. Shotgun lipidomics was made possible by way of technological developments in mass spectrometry. Lipids are naturally non-volatile and will generally degrade before evaporating. This makes lipids unsuitable for electron ionization and chemical ionization which both require molecules to be in the gas phase before ionization. This made mass spectrometry of lipids not practicable before the development of ionization techniques workable with solution phase analytes. The first such major technique was fast atom bombardment which uses a beam of high kinetic energy neutral atoms to desorb and ionize analytes deposited on a matrix. This is a “soft” ionization technique that induced a minimum of fragmentation during the ionization process. This enabled discoveries such as the predominance of plasmalogens among phospholipids in the myocardial sarcolemma.<sup>42</sup> While fast atom bombardment was nearly revolutionary in mass spectrometry, with the development of electrospray ionization (ESI), fast atom bombardment quickly receded in significance, due to ESI’s ease of use, increased sensitivity, and even softer ionization process.<sup>5, 42</sup> Shotgun lipidomics also relied on developments in mass analyzer technology. Single quadrupole mass spectrometers with unit-mass ( $\pm 0.5 m/z$ ) resolution were not sufficient to characterize lipid mixtures. High resolution and/or tandem mass spectrometers were also necessary to increase

selectivity. Searching the mass of PG 33:1 (733.5025  $m/z$ ) in the LIPID MAPS database with a tolerance of 0.5  $m/z$  returns 306 lipids with 156 unique molecular formulae that could match the searched  $m/z$ . Dropping this to 0.005  $m/z$ , which is reasonably achievable mass accuracy in Orbitraps, gives only 25 lipids with 8 unique formulae. Because of this, the development of tandem quadrupoles and ion traps provided the necessary capabilities to enable shotgun lipidomics. This is still a relatively simple process that can resolve lipids to the species level (class and fatty acid tails) in some situations. However, the lack of a separation technique comes with significant limitations. First, due to the direct infusion of all lipids simultaneously, there is a significant potential for ionization suppression and adduction. Best practice then requires the infusion of dilute samples, which in turn limits the detection limits of the method.<sup>40</sup> An additional drawback is the inability to distinguish isobaric species. For example, while PC 32:0 and PE 35:0 share the same formula (and thus  $m/z$ ), they would be chromatographically resolved by most chromatography methods. However, in shotgun lipidomics these would require specific tandem mass spectrometry scans to distinguish the two. This is also the case for isomer-saturation pairs. A double  $^{13}\text{C}$  isotope of PC 32:1 is functionally isobaric with PC 32:0. While these would be separated by reversed-phase chromatography, shotgun lipidomics (even with tandem mass spectrometry) will be unable to distinguish the two. While all these techniques have certain beneficial characteristics, they are unfit for purpose when the comprehensive characterization of lipidomes with species-level resolution is desired.

#### *Liquid Chromatography-Mass Spectrometry Lipidomics*

To achieve greater species-level resolution of lipids, one can add liquid chromatography to separate lipids before analysis by mass spectrometry. Liquid chromatography separates a mixture of chemicals through differential affinities for a stationary phase (chemically modified silica bead matrix) and a mobile phase (solvent). The solvent flows across the stationary phase or column and will elute from the column at different times based on how strongly analytes “stick” to the stationary phase. For best resolution and efficiency, a mobile phase gradient is employed where the solvent composition changes over the course of the chromatography. In the case of lipid separation, the stationary phase of choice is either a C18 (functionalized with an eighteen-carbon alkyl group) or C30 reversed phase column (functionalized with a thirty-carbon alkyl group). These both provide high methylene-selectivity and will fully resolve lipid species differing by only one carbon or one degree of unsaturation. This separation is performed online

with the mass spectrometer such that lipids are immediately analyzed by the mass spectrometer following elution from the column, allowing for both identification and quantification of the resolved lipid species.

LC-MS lipidomics analyses can be done in a targeted or untargeted fashion. Targeted analyses require the analyst to define a given set of lipids to be measured. The chromatography will generally be tailored specifically for given analytes, and the mass spectrometer will acquire data “focused” on those analytes, typically in the form of scheduled multiple-reaction monitoring (MRM) or parallel reaction monitoring (PRM) scans (forms of tandem mass spectrometry). These analytical methods are generally developed with standards and optimized over several iterations. Although this targeted approach tends to have the best quantitative accuracy and simplifies lipid identification, this makes the instrument blind to lipids not preselected. An ideal use-case for targeted lipidomics, then, would be measurements of large numbers of samples from a biological system with a well-characterized lipidome. If relatively few samples are to be measured, the time to specifically adapt an LC-MS method to the given analytes of interest may not be worth it. Similarly, if the lipid species present in an organism or biological system are not fully known, it would be better to start with an untargeted analysis of some kind to identify the species present. Untargeted analyses are designed to have broad specificity. Generally, the MS acquisition will be a data-dependent method where intense ions are isolated and fragmented without the significant use of prior knowledge. This gives the system the potential to measure anything in the sample that will ionize, eliminating the need for defining lipids of interest ahead of time. However, because the system will measure anything that ionizes, the data are much messier than in targeted methods and are more prone to ion suppression and matrix effects. Identification in an untargeted workflow can be a challenge since confirmation with standards is the gold-standard for any analytical technique. This is not feasible in a truly untargeted approach. Rather, lipids are identified (with varying degrees of confidence) based on similarity to library or reference data. The use of such reference data is improved by the predictable fragmentation patterns of lipids as compared to metabolites more broadly. In fact, rules-based theoretical fragmentation spectra match very well the spectra observed in experimental data, at least with respect to the identity of ions observed. Still, the data generated from untargeted LC-MS analyses can be quite large requiring automation of the identification process.

### *Automated Analysis of Untargeted LC-MS Lipidomics Data*

There are many different automated lipid identification software available including Lipostar, Lipid Data Analyzer, LipidXplorer, Lipid Hunter and Lipid Wizard.<sup>43-47</sup> LipidXplorer was designed for use with shotgun lipidomics (no chromatography) and utilizes a rules-based identification workflow that can putatively annotate features based on exact mass alone, but can also make some use of fragmentation data.<sup>45</sup> Lipid Data Analyzer works only on MS1 data and looks at exact mass and isotope patterns.<sup>43</sup> Lipid Wizard makes extensive use of retention time information but does not work with fragmentation spectra, limiting its ability to distinguish related lipids.<sup>47</sup> Lipostar and LipidHunter both make use of exact mass and expected fragmentation patterns and work with LC-MS data.<sup>44, 46</sup> Other programs are available that are capable of processing both metabolomics and lipidomics data such as MS-DIAL, MZmine, and Compound Discoverer.<sup>48-50</sup> Based on citation counts, MS-DIAL and MZmine are far more popular programs than the lipid specific software, although this is, to some degree, to be expected given their greater versatility (citations assessed using the Publish or Perish software with Google Scholar as the database).<sup>51</sup> Compound Discoverer is a proprietary program maintained by Thermo Scientific and does not have publications associated with it. This makes it challenging to rank with respect to popularity, although it likely is less commonly used than MS-DIAL or MZmine since it is not vendor-neutral. For years MS-DIAL has been the only vendor-neutral all-in-one solution for metabolomics and lipidomics data processing since MZmine only recently acquired the ability to natively perform spectral similarity matching. MS-DIAL follows a pattern similar to other programs in its data processing, consisting of feature detection, followed by exact mass filtering, and tandem MS matching. For lipids, MS-DIAL has a built-in library that consists of both *in-silico* rules-based fragmentation spectra and experimental spectra that covers a wide range of lipids. While a rigorous assessment of the accuracy and other performance metrics has not yet been performed for MS-DIAL and related programs, there is a growing consensus that any given untargeted analysis program is not sufficiently reliable to be used without additional validation.<sup>20</sup> Indeed, the “gold standard” validation is an examination of the putative identifications by an expert taking into account chromatographic behaviour as well as mass spectral data. This makes the untargeted analysis programs only marginally useful. They can be helpful in pointing in the right direction but are not capable of giving results sufficiently reliable for use in an automated fashion. This is a severe bottleneck for untargeted lipidomics (as

well as metabolomics) that, while possibly ameliorable by analytical improvements in separation or fragmentation techniques, likely needs to be addressed through informatic improvements.

### **Selection and Development of My LC-MS Lipidomics Method:**

#### *Selection of liquid chromatography-mass spectrometry as the analytical platform*

Considering the available options for lipid analysis, the appropriate method will in large part be determined by the science question of a given experiment. I have had several different research projects, all with different science questions; however, a consensus platform clearly arises from a consideration of the experimental metrics needed. The chronologically first lipidomics project I pursued involved the deuteration of *E. coli* lipids to support neutron scattering (Chapter VI). To this end, the deuteration states of the lipids had to be quantified, ideally to within a few percent accuracy. In this case, mass spectrometry was a necessary component of the analytical platform. Deuteration of biomolecules will only marginally change their chemical properties and so methods like TLC would provide no discriminatory power between deuterated and protiated lipids. NMR could theoretically distinguish between deuterated and protiated lipids since deuterium will be silent in  $^1\text{H}$  (proton) NMR (and vice versa with  $^2\text{H}$  NMR); however, NMR is functionally a bulk analysis method which would only provide information regarding the average deuteration. While certainly useful, there is a desire to know the variability of deuteration states, both the spread within a given lipid species and the deuteration differences between lipid species. Similarly, NMR would not be able to link measured fatty acid tail deuteration to specific headgroups. By the same token, GC-MS FAMES analysis is also unfit for purpose. This would provide a measurement of the fatty acid deuteration (including the distribution thereof) but would not link specific fatty acids to headgroups providing species-level information. Shotgun mass spectrometry then could be a viable option. The mass analysis of intact lipids would provide the species-specific deuteration states and would clarify whether deuteration varies between lipid classes. However, the congestion of the mass spectra would present a real issue. The deuteration states of related lipids would be highly convoluted and individual deuteration states of specific lipid species could not be reliably characterized. For example, a PG 34:2 with four deuterium and a PG 34:1 with 2 deuterium would require  $>240,000$  resolution at an  $m/z$  of 749. Given that the orbitrap's resolution decreases as a function of the square root of the  $m/z$ , this would correspond to a needed resolution of 472,000 at  $m/z$  200 (a common benchmark for comparing resolving powers). This

resolution is only achievable by only the most recent high-end orbitrap mass spectrometers (and ion cyclotron resonance instruments). A better option is to make use of the high methylene selectivity of reversed phase chromatographies which easily resolve lipids different by even a single degree of unsaturation (see **Figure 1-2**). Lipid species that still coelute, PE 30:0 and PE 32:1, for example, are trivial to resolve in the  $m/z$  dimension. The 26 deuterium isotopologue of PE 30:0 has the same nominal mass as the monoisotopic PE 32:1 (688.6393 and 688.4918  $m/z$ ). However, these require a resolution of only 4700 to distinguish. This is the equivalent of just over 9000 at  $m/z$  200, well within the range of high-resolution mass spectrometers.

Other projects have similarly called for the specific use of the liquid chromatography-mass spectrometry platform for lipid measurements. In chapter IV, I describe the lipidomics analysis of two relatively understudied fungi. The goal for that project being to both characterize the native lipidomes of the fungi and quantify the effect of lipo-chitooglucosaccharide treatment on the lipidomes. Since there is little specifically known about the lipid composition of these fungi, mass spectrometry is indicated as the ideal experiment instrument. Thin-layer chromatography could be useful to characterize the general types of lipids present, but again lipid-species analysis requires mass spectrometry of intact lipids. Similarly, to profile the effect of LCO treatment, TLC could be sufficient if large changes in lipid class abundance occurred, but in order to assess multiple effects (class changes, fatty acid length changes, *et cetera*) at the same time, LC-MS is the analytical platform of choice. While isomer-saturation pairs are not as significant a drawback in this experiment, the use of shotgun lipidomics is contraindicated by the high ionization suppression a complex mixture of lipids is likely to cause. During the electrospray ionization process, there is a limited ion count that can reasonably be achieved. This induces “competition” between analytes and those analytes that can more readily ionize can “drown out” those analytes of poorer ionization efficiency. Even if this weren’t the case, there are physical restraints to the number of ions the orbitrap itself can contain before space-charge effects prohibit accurate measurements. This makes shotgun lipidomics intrinsically limited in terms of dynamic range (the highest and lowest concentrations that can be measured at the same time). By spreading out the analytes and measuring them sequentially as in liquid chromatography, the dynamic range burden on the mass spectrometer is relaxed, allowing the measurement of lower abundance lipids.

While LC-MS lipidomics is the foundation of my research and development, other lipid analysis methods have been employed in my research as well. My research in chapter V not only utilizes LC-MS untargeted lipidomics but also TLC and NMR lipid measurements. This underscores the fact these methods are often complementary and not mere replacements of each other. The selection of an LC-MS method for lipidomics does require care, however, and in the following section I describe of my method of choice.

#### *Original LC-MS method identified and subsequent improvements*

While hydrophilic interaction liquid chromatography (HILIC) methods are sometimes used for untargeted lipidomics, they are less popular than reversed phase approaches with about 70% of lipidomics datasets on repositories Metabolomics Workbench<sup>52</sup> and MetaboLights<sup>53, 54</sup> using RP chromatography.<sup>55</sup> The retention mechanisms of RP and HILIC are distinct. RP chromatography works primarily through adsorption of analytes to the non-polar matrix and resulting differential partitioning between adsorption to the stationary phase and dissolution in the mobile phase. HILIC retention mechanisms vary somewhat due to the diversity of stationary phases used. Some are simply repackaged normal phase columns like bare silica, while others have a zwitterionic polymer matrix as a stationary phase. However, the primary mechanism is thought to be partitioning between the mobile phase proper and the aqueous-enriched solvent “layer” immediately surrounding the stationary phase. This is, therefore, a liquid-liquid process rather than the solid-liquid partitioning of RP. That said, adsorption and ion-exchange mechanisms are often a factor in HILIC separation as well. Because of these distinct mechanisms, the expected elution of lipids is very different between RP and HILIC chromatographies. RP is often described as separating primarily by fatty acid tails while HILIC separates by headgroup. This is somewhat true, although headgroup absolutely effects the retention of lipids in RP chromatography. In HILIC this gives rise to high level of coelution of lipids of the same class, which presents a similar isomer-saturation pair problem as shotgun lipidomics suffers from, making double <sup>13</sup>C isotopes and unsaturation difficult to distinguish.<sup>56, 57</sup> Conversely, lipids that coelute in RP chromatography tend to be very different in *m/z* space, making MS resolution of these species trivially easy.<sup>7, 57</sup> Another drawback of HILIC is that because of its complex retention mechanisms, many analytes tend to experience peak broadening, especially for analytes with different possible charge states in solution. RP peaks, on the other hand, tend to be narrower and thereby provide better sensitivity and peak capacity.<sup>57</sup>

I selected a reversed-phase nano-LC-MS method from the literature as a starting point to establish an LC-MS analysis method for untargeted lipidomics.<sup>58</sup> This method used 60:40 acetonitrile:water for solvent A (the aqueous solvent) and 90:10 isopropanol:acetonitrile for solvent B (the lipophilic solvent). Because of the greater hydrophobicity of lipids, these solvents are much stronger in hydrophobicity than many that are used from proteomics or polar metabolomics methods. The original gradient can be seen in **Figure 1-3.A**. However, this has a redundant two-wash procedure that can be removed to increase throughput. Additionally, the gradient slope can be adjusted based on the density of lipids eluting at a given time. Lower slopes can allow for better discrimination between closely eluting lipids while faster slopes shorten the overall length of the gradient. After several iterations, the gradient I used for most of my analyses is shown in **Figure 1-3.B** with the total time reduced from 110 mins to 90 mins. Because this reduction in time was due to optimizations to the gradient and not a reduction in effective separation time, this came at not cost to the overall performance of the separation. To improve elution and washing at the high-organic region of the chromatography, a method was tested that incorporated a third solvent. Solvent C consisted of 6:3:1 isopropanol:butanol:acetonitrile. Butanol (being a four-carbon alcohol) is a “stronger” solvent for hydrophobic analytes than isopropanol. The rationale was that butanol would allow for more selectivity between extremely hydrophobic lipids as well as ensuring the chromatographic wash was successful in removing all lipids from the column. This was shown to be unnecessary and was not pursued further. The only other modification to the mobile phase was the addition of 1% water to solvent B (90:10:1). This small proportion of water had no noticeable effect on the chromatographic performance of the method, but did greatly simplify the dissolution of ammonium acetate into solvent B. Ammonium acetate is present at 0.05 g/L in both solvents A and B. It serves as an ionization additive which can pair and form counter ions to improve the ionizability of the analytes in the sample. It can also undergo redox reactions that help balance the charge generated during electrospray ionization. I recently streamlined this gradient further and more finely tuned its elution parameters. The result can be seen in **Figure 1-3.C**.

The column used in the original paper was an in-house packed column. I similarly prepared the columns used in-house. I utilized 100 µm inner diameter silica capillaries which were packed with Kinetex 1.7 µm C18 stationary phase beads with 100 Å pores from Phenomenex. After forming an electrospray tip by laser pulling the silica capillary, the stationary

phase was forced into the column as a methanol slurry by means of a pressure cell. Columns were typically packed with stationary phase to a length between 15-20 cm. After optimizing the waste column to provide high backpressure, the flow rate was set to 0.05 mL/min. This would typically result in a maximum pressure of ~400 bar. This provided a good flow rate through the column (hundreds of nL/min) while minimizing the solvent/sample proportion directed to waste.

#### *Selection and Operation of Mass Spectrometer for Untargeted Lipidomics*

The most important MS figures of merit for lipidomics are likely mass accuracy, resolution, and scan speed. Mass resolution was briefly discussed in the context of deuterium labeling and resolving powers of 10,000 was determined to be sufficient to distinguish overlapping isotopologue distributions. This is achievable with linear-trapping quadrupole (LTQ) instruments despite not generally being considered “high resolution”. However, the scan speed to perform this is quite slow, and a 2000  $m/z$  range scan will take about a whole second. On the other hand, even previous generation orbitrap instruments can cover a 2000  $m/z$  range at 30,000 resolving power in only 0.6 seconds. Orbitraps also have greatly superior mass accuracy than do LTQ analyzers. A typical scan in an LTQ might have the mass accuracy of  $\pm 0.1$   $m/z$  while an orbitrap can achieve  $<10$  ppm accuracy (typically less than  $\pm 0.005$   $m/z$ ). Searching the  $m/z$  of PG 33:1 on the LIPID MAPS database with a 0.1  $m/z$  tolerance returns 209 possible lipids with 99 unique molecular formulae as opposed to the orbitrap’s 25 and 8. This puts a greatly reduced burden on the tandem mass spectrometry to be able to accurately distinguish between lipid species. A large part of why the orbitrap is able to achieve much better resolution and mass accuracy is because it essentially measures every ion multiple times. The LTQ scans across the mass range, ejecting ions to electron multiplier tubes for detection. The point along the scan where the ion is detected corresponds to its  $m/z$  and, importantly, the measurement only happens once. In the orbitrap, ions are caught in a centripetal orbit between a central (spindle) and outer (barrel) electrodes. Due to the curve of the electrodes, ions also oscillate radially along the electrodes where the frequency of this oscillation is inversely proportional to the  $m/z$  of the ion. Oscillations of all ions present are measured simultaneously over a period of time, which essentially provides a measurement averaging effect allowing for more accurate results. A figure showing the architecture of an orbitrap mass analyzer and the ion flight path is shown in **Figure 1-4**.

However, in order to confidently identify lipid species, especially in an untargeted fashion, it is still necessary to generate tandem mass spectra for the putative lipid species. The simplest approach to do this is to perform a data-dependent acquisition (DDA) experiment. DDA experiments require a balancing of the instrument duty cycle as applied to generating fragmentation scans as well as the acquisition of full-scan data. The full-scans are important because they provide the information from which extracted ion chromatograms (EICs) are generated. If few datum points are acquired across a chromatographic peak, the EICs generated will be highly variable and will not provide a stable quantitative estimate of the actual peak area. However, if too much duty cycle is spent on full-scans, the number of identifiable lipids will eventually start to decrease as proportionally fewer fragmentation scans are acquired. While orbitrap scans are desirable for the MS1 acquisition, these would be prohibitively slow for MS2 acquisition in the LTQ-Orbitrap Pro instrument. This is due, in part, to the ion accumulation time required for the orbitrap fragmentation scans which, between lossy ion optics and an intrinsically less responsive detector, require much more time than LTQ tandem MS scans. Lipid fragmentation patterns tend to be somewhat sparse with maybe 4-8 peaks depending on the lipid. These do not require particularly high resolution or mass accuracy to accurately characterize a lipid's fragmentation pattern. My method, therefore, uses a top-ten DDA with an orbitrap full-scan followed by 10 LTQ tandem MS scans. This seems to strike a favorable duty-cycle balance with a full cycle at about 0.5 Hz. A typical lipid elution peak with my chromatography is roughly one minute, thereby resulting in about 30 datapoints for many EICs. Dynamic exclusion is also employed which helps to avoid redundant MS2 measurements of the same ion by "ignoring" ions for a given period of time after targeting them for a MS2 scan. This allows subsequent top-ten DDA cycles to fragment lower abundance ions without increasing the MS2 duty cycle proportion. In general, while more optimization can always be performed, this method has been more than sufficient to allow for sensitive and robust analysis of lipidomes across a range of experiments. Overall then, combining the high resolution orbitrap capabilities with the rapid scanning linear trapping quadrupole, the LTQ-Orbitrap Velos Pro provides the needed figures of merit to achieve high quality lipidomics data. A diagram of the structure of this instrument is shown in **Figure 1-5**. It is worth noting that the Velos Pro was not the only viable option. A quadrupole-Exactive (orbitrap) or quadrupole-Time of Flight instruments would also work well. However, the Velos Pro was readily available and sufficient for the required task.

## *Lipid Extraction*

Before the LC-MS analysis can ever take place, however, lipidomic samples must first be generated. After sample collection, which generally involves centrifuging or filtering cells from a liquid medium, lipids must be extracted from the sample. Lipid extractions have been utilized long before “lipidomics” as a discipline was conceived of. The most influential bulk lipid extraction is the Folch extraction which was developed in the 1950s and utilizes a chloroform-methanol-water biphasic extraction procedure.<sup>59,60</sup> This results in a partition of nonpolar components dissolving in the chloroform-rich lower phase and the dissolution of polar components in the upper aqueous phase with insoluble material (proteins and nucleic acids) forming a precipitate between the phases. While this is an effective preparation method, the use of chloroform has been considered a drawback given its time-sensitivity and suspected carcinogen status (IARC class 2B). Additionally, the formation of the lipid-rich phase below the aqueous and insoluble debris is inconvenient when performing the extraction. The Bligh and Dyer extraction simplified the Folch method slightly but still used chloroform as the primary organic solvent.<sup>61</sup> More recently, the Matyash protocol has been developed as a substitute for the Folch.<sup>62</sup> This uses methyl *tert*-butyl ether (MTBE) as the primary organic solvent along with methanol and water to form a biphasic extract. In addition to avoiding the possible carcinogenicity of chloroform, MTBE is also less dense than water and will form a lipid-rich upper phase simplifying extract collection. For both protocols, the extraction solvent volumes will be in part dependent on the sample amounts. From the different samples analyzed, I have found that 50 mg of wet cell pellet is sufficient to provide good lipid signal while allowing the extraction to be performed in 2 mL centrifuge tubes. The Matyash protocol has become the default extraction procedure I use. However, there are times when this method is not ideal. When extracting fungal samples with reasonably high biomass, there was a large amount of what appear to be undissolved lipids forming during the extraction. Fungi have higher proportions of highly nonpolar lipids such as triacylglycerols (TAGs) and ergosterols. To better ensure collection of lipids from these fungi, I switched to a modified Folch procedure since chloroform is a stronger nonpolar solvent than MTBE. This illustrates the need to be validating and adapting an analytical workflow as new sample types and experiments emerge.

While chloroform is a stronger solvent than the MTBE mixture, both can be problematic if used directly in the LC. Despite being a split-flow setup, the nano chromatography method

employed still has a low total volume being used. This low volume limits the dilution of the injection volume, which can help sample conservation, but has the side effect of decreasing solvent compatibility. The extraction solvent polarity is such that many lipids will remain dissolved and not partition to the stationary phase. This leads many lipids to elute in the dead volume or else causes severe peak splitting (**Figure 1-6.A**). Because of this, after extracting lipids, a drying and redissolution step is employed. I often use a vacuum centrifuge (speedvac) for the drying step as this operates quickly for high organic solvents and has limited potential for sample loss or contamination. The solvent used for redissolution must be balanced such that it will dissolve lipids without being too strongly non-polar that lipids do not partition to the stationary phase. A solvent consisting of 8:23:69 butanol:isopropanol:water has been recommended previously for lipidomics experiments.<sup>58</sup> However, I have found this to be somewhat too weak a solvent for lipids and am not always able to fully dissolve the lipid extract. Other compositions I have used include 10:30:60 butanol:isopropanol:water and even 1:1:1. Equivolume mixtures are fairly strong non-polar solvents and limit the injection volume that can be used before stationary phase partitioning becomes problematic ( $\leq 5 \mu\text{L}$ , see **Figure 1-6**). My current “default” redissolution solvent is 10:30:60 butanol:isopropanol:water (10Bu) which generally balances solvent strength and compatibility well.

#### **Areas in Lipidomics Where Further Development is Needed:**

While I have modified and optimized the described LC-MS method and sample prep techniques, these were not developed *de novo*. These experimental procedures were based on research and developments already performed by others in the field. This groundwork has enabled the relatively rapid adaptation of established procedures to our instruments and research areas, and I am indebted to those who laid this groundwork. That being the case, there are still many areas that need focused development. This is true both on a fundamental technical level as well as in more niche applications of lipidomics (like isotope labeling). Two especially important limitations in lipidomics are the lack of reliable automated lipid identification and validated statistical procedures for lipidomics data analysis. Lipid identification is considered the primary bottleneck in untargeted lipidomics. Although various programs have developed to automate lipid identification, it is not clear that these programs have developed to the point of significant utility. This can be seen in a *Nature Communications* Matters Arising discussion where a paper was criticized for their low rigor in lipid identifications.<sup>20</sup> While the discussion was broad, the

main focus was on the unreliability of automated lipid spectral matching. The authors, many of whom are part of the Lipidomics Standards Initiative, wrote that “exclusively relying on annotations by a single software without additional means of validation often leads to unacceptably high rates of false positive identification.”<sup>20</sup> Indeed, similar concerns have been raised elsewhere, producing some controversy in reported lipidomics results.<sup>63-67</sup> While the response of the authors of the original *Nature Communications* paper covered other points, they raised a practical objection to the concerns of the Matters Arising article. They wrote, “it is not clear how the approach advocated by Köfeler et al.—up to the detailed manual inspection of all spectra by experts and even the synthesis of reference molecules—should be the future of omics-type investigations.”<sup>68</sup> Clearly the need for rapid accurate lipid identification from untargeted data is a live issue in the community. Even more difficult is the identification of modified or otherwise non-standard lipids from untargeted data, which can be challenging even with manual inspection of the data.

The other major limitation, as I see it, relates to the quantitative and statistical analysis of lipidomics data. LC-MS omics data is generally assumed, explicitly or implicitly, by practitioners to have certain statistical properties. For example, the common use of ANOVA and T-tests assume a normal distribution of the sample mean (and a chi-distributed sample standard deviation), but if logarithmic transformation is applied before testing, the assumption of normality changes to one of log-normality. However, the assumptions practitioners make actually vary from individual to individual and are rarely based on strong empirical evidence. This makes the application of various statistical procedures a matter of “gut feeling” as one publication put it.<sup>69</sup> This has led to inconsistent practices such as calculating arithmetic means vs geometric means, parametric vs non-parametric tests, and many different data transformation, normalization, and imputation approaches. Clearly, research is needed to clarify what assumptions can be made regarding the structure of LC-MS lipidomics data and what statistical approaches are robust and reliable.

### **Summary of my Doctoral Research:**

My research has sought to bring together both rigorous analytical method development and practical biological analysis for the advancement of lipidomics in the biological sciences. Not only have I directly attacked the previously identified issues of lipid identification and

statistical procedures, I have also applied lipidomics to a variety of important biological systems to advance our understand of fundamental and applied biology.

#### *Machine Learning for Rapid and Reliable Lipid Identification in Untargeted Lipidomics*

A major bottleneck of untargeted lipidomics is lipid identification. A growing consensus in the field recognizes that automated analysis software do not have sufficient accuracy to permit their use without additional validation. Another project of mine addressed just this issue. In collaboration with Tomás Rush (ORNL), Steven Tavis, and Andrew Stai (both fellow Genome Science and Technology students), I developed a machine learning model (called LipoCLEAN) to filter automated lipid identifications for likely false positives. To do this, automated lipid identifications were manually curated by an examination of a putative lipid's chromatographic behavior, observed isotope distribution, and tandem mass spectrometry fragmentation pattern. Using this process, 3,408 putative lipid identifications across four datasets were manually curated as true or false positives, of which I personally curated 3,105. Predictors used to help identify false positives include the signal to noise ratio, the tandem MS cosine similarity score, and the root-mean-squared error of the observed isotope abundances. These were used in a gradient-boosted decision trees model to pre-predict expected high confidence identifications. From these, a  $m/z$  recalibration was performed to stabilize mass error as an additional predictor. After another round of pre-prediction, a retention time-LOWESS regression was performed to calibrate the reversed phase chromatography expected lipid elution order to the experimental method. This allows retention time error to be used as a fifth predictor. All predictors are then combined into a random forest of shallow linear trees where the leaves are logistic regressions. This model categorized lipid identifications as “probable true positive,” “probable false positive,” or “uncertain.” While the standard software had a false-discovery rate (FDR) of 45%, the probable true positives as reported by our model had an FDR of only 7% while achieving an 80% recall (true positive rate). If uncertain lipids are included, the recall increases to 95% while the FDR increases to 17%. A paper describing this model is in review at *Analytical Chemistry*.

#### *Empirical Assessment of the Repeatability of Untargeted Lipidomics and Proteomics*

##### *Experiments*

While untargeted lipidomics and proteomics have become widely used, there is still little empirical evidence of the experiment-to-experiment reproducibility of analytical and statistical approaches. Many statistical approaches have important assumptions (such as the distributional

properties of the data) that have never been validated to hold for LC-MS omics data. In collaboration with Steven Tavis, Diana Ramirez, and Frank Löffler, I have undertaken to test the distributional properties of LC-MS omics data and the robustness of common statistical approaches. While this project is still in progress, untargeted lipidomics has been performed on 92 samples of broth cultured bacteria. These were split between two sample types with 46 biological replicates each, allowing for the confident assessment of statistical properties. From these measurements, statistical modeling has shown the biological distribution of lipids do not hold to a normal (Gaussian) distribution. T-tests performed on this distribution result in significantly biased results. T- and log-normal distributions were also tested, but neither fit the data better than a normal distribution. These results indicate that non-parametric statistical approaches are the most robust approaches for LC-MS omics analyses. This project is still in progress.

#### *Non-model Fungal Lipidomics and Lipo-chitoogliosaccharide Signaling Molecules*

In the first application of LipoCLEAN, I investigated the lipidomes of two understudied fungi, *Aspergillus fumigatus* and *Laccaria bicolor*. This was done in collaboration with Tomás Rush at ORNL. Here I performed the first detailed lipidomic investigation of these fungi. LipoCLEAN was used for identification and novel statistical practices were employed for quantitative analysis. Neither of these genera have had their lipidomes characterized making this work useful to clarify the breadth of fungal lipids and help guide future research. At the same time, we investigated the effects of lipo-chitoogliosaccharides (LCOs) on the lipidomes of these fungi. LCOs have been shown to impact fungal morphology as well as many other aspects of physiology (protein and metabolite abundances). Surprisingly, despite the seeming wide physiological effects of LCOs, we found the lipidome to be largely unaffected by the treatment of LCOs. From this we hypothesize that the native lipidome is well suited for fungal health and the signaling effects of LCOs selectively preserve that native composition. A manuscript describing this is in preparation.

#### *In vivo Deuteration of Bacterial Phospholipids*

In collaboration with Qiu Zhang, Wellington Leite, Hong-Hai Zhang, Brian Sanders, and Hugh O'Neill at ORNL, one major avenue of my research has been the *in vivo* deuteration of bacterial lipids. Deuterated lipids are useful for many applications but especially for neutron-based experimental techniques. Neutron scattering and reflectometry both benefit from contrast

matching, where selective deuteration of sample components renders them “transparent” to neutrons, thereby allowing cleaner measurements of unmatched components of interest. Through extensive lipidomics measurements and computational modeling, I thoroughly characterized the *in vivo* deuteration of *E. coli* lipids from a variety of deuterated feedstocks. I introduced the use of a computationally derived one-number summary of deuterium labeling distributions and developed guidelines for the design of specifically deuterated phospholipids. These guidelines allow for the control of both the amount and the location of deuteration. This enables large-scale synthesis of selectively deuterated phospholipids with a native composition of lipid species. This project resulted in the publication of “Characterization of the *in vivo* deuteration of native phospholipids by mass spectrometry yields guidelines for their regiospecific customization” in *Analytical Chemistry* (DOI: 10.1021/acs.analchem.3c03750). Follow-up experiments have extended the application of *in vivo* lipid deuteration to transgenic *E. coli* that was engineered to produce non-native lipids. This allows for a better mimic of mammalian lipid profiles. While useful for basic research, such steps are also critical for research focused on human health and physiology. This work is completed, and we are finishing a report for submission.

#### *Covalent Incorporation of PFAS Into Bacterial Phospholipids*

I have also provided the first evidence of the covalent incorporation of per- and polyfluoroalkyl substances (PFAS) into bacterial lipids. In collaboration with Yongchao Xie, Diana Ramirez, Gao Chen, Cynthia Swift, Amanda May, Shawn Campagna, and Frank Löffler (UTK), and Anita Thapalia, and Trent Key (Exxon Mobil), I have identified bacterial phospholipids that contain a fluorotelomere in place of a fatty acid. Untargeted lipidomics was performed on bacteria grown with 8:3-fluorotelomere carboxylic acid (FTCA) present in the media at environmentally relevant concentrations. Identifications were achieved by interrogation of tandem mass spectrometry. Both lipid-specific fragment ions and multiple fragments explained by the presence of FTCA incorporation were observed. FTCA-specific fragmentation pathways were confirmed by comparison to tandem mass spectrometry of free FTCA. FTCA incorporation was observed in the cardiolipin, phosphatidylethanolamine, phosphatidylglycerol, and phosphatidic acid lipid classes. I have shown incorporation to occur in *Pseudomonas sp.* strain 273, *Pseudomonas putida*, *E. coli*, and *Enterococcus faecalis*, implying this PFAS incorporation may be a broad feature among many bacteria. This novel discovery represents a previously unknown mechanism for PFAS toxicity, bioaccumulation, and environmental

longevity, and is likely to reshape our understanding of PFAS in the environment. A manuscript describing this research is currently under consideration at *Nature*.

#### *A Multi-Omics Investigation of Clostridium autoethanogenum's Cultivation Temperature*

Working in collaboration with Payal Chirania (UTK), Richard Giannone, Nancy Engle, Timothy Tschaplinski (all ORNL), and Steven Brown (LanzaTech), we performed a detailed multi-omics characterization of the physiological differences in *Clostridium autoethanogenum* when grown at different temperatures. *C. autoethanogenum* is a carbon dioxide/monoxide fermenting bacterium used industrially to produce biofuels from waste gasses. Cultivation temperature changes the gas fermentation product profile and is important for the production of biofuel or other bioproducts. This is investigated using proteomics, lipidomics, and metabolomics. Of particular interest is our findings that suggest *C. autoethanogenum* produces less glycerophospholipids at high temperatures in preference for glycolipids. This somewhat unusual lipid remodeling may be a homeoviscous adaption of *C. autoethanogenum*. A manuscript describing this project is in preparation.

#### *Summary*

Overall, this work broadly has two goals. The first is to develop and validate analytical techniques that will provide a robust foundation for future experimentation. This can be seen in my development of a machine learning model for improved lipid identification, my characterization of statistical properties of lipidomics data, and my development of guidelines for *in vivo* lipid identification. The second objective of my dissertation research has been the investigation of specific biological questions through the application of lipidomics. This is most clearly seen in my study of fungal lipidomes and their responses to lipo-chitooglucosaccharide signaling, the investigation of the covalent incorporation of PFAS into phospholipids, and the effects of temperature on the lipidome of *C. autoethanogenum*. In some cases these goals overlap. For example, not only did we report the PFASylation of bacterial phospholipids, I also characterized their fragmentation patterns facilitating their subsequent analysis in research projects yet to come. My hope is that the impact of my research will be realized in lipidomics research performed by many others to come that is more rigorous and robust than currently practiced.

## **Chapter II - LipoCLEAN: The Development of a Machine Learning-Based Rescoring Method for Confident Assignment of Lipid Identifications**

### Citation:

This chapter is primarily derived from the following manuscript (current in review): “LipoCLEAN: A Machine Learning Filter to Improve Untargeted Lipid Identification Confidence.” Steven Tavis, Matthew J. Keller, Andrew J. Stai, Tomás A. Rush, and Robert L. Hettich. 2024, under review at *Analytical Chemistry*. Steven Tavis is a co-first author with me on this manuscript, along with coauthors Andrew Stai, Tomás Rush, and corresponding author Robert Hettich. In this work, I performed LC-MS lipidomics experiments, validated thousands of putative lipid identifications, helped identify predictors for model development, and helped draft and edit the manuscript.

### Context:

Lipidomics experimental methods have been well developed since its inception as discussed in Chapter I. The major bottlenecks remaining primarily involve data analysis, both qualitative lipid identification and quantitative statistical analysis. With sufficient up-front work, lipid data analysis can be made trivial if the scope is limited to a targeted selection of lipids. However, for truly untargeted analyses this is not an option. Instead, automated analysis software are used that seek to identify and quantify lipids from untargeted lipidomics datasets with no external information. These software will “scan” the  $m/z$ -retention time matrix for chromatographic peaks, match the  $m/z$  to a lipid database, and check any associated tandem MS for library matches. While a generally sound principle, this does not always achieve reliable results. Early in my doctoral research I critically examined the performance of such untargeted analysis software. Having implemental untargeted metabolomics/lipidomics LC-MS methods, I utilized software to perform lipid/metabolite identification. On the metabolite side, issues were obvious when anthropogenic chemicals such as remdesvir were detected in *E. coli* metabolite extracts. While implausible lipid identification were less obvious, a manual validation of reported identifications revealed many false positives. This was observed both in the identification of noise as lipids and the misidentification of true lipids. Because of this I implemented manual lipid identification for all the lipidomics research projects I was involved in. For projects with organisms that have relatively simple lipidomes, like *E. coli* or *Pseudomonas putida*, this was a functional, though somewhat time-consuming process. However, when I started studying filamentous fungi with relatively poorly characterized

lipidomes, it became clear that this manual approach would not be tractable to cover the several thousand lipid identifications reported by automated software.

A further examination of the data revealed possible trends between the reported identifications and associated matching metrics. For example, features of a low precursor signal-to-noise seemed more likely to be false positives than high signal-to-noise. Similarly, a high cosine similarity score between an experimental spectrum and the reference was somewhat associated with fewer false positives. After noticing this, I collaborated with a fellow graduate student, Steven Tavis, to build a machine learning model to predict likely true and false positives from automated lipid identifications using such matching metrics. In order to provide a large dataset to build this model I performed the manual curation of 3,105 automated lipid identifications covering three different instrument types. An additional 303 identifications were provided by another graduate student, Andrew Stai, whom I had trained in lipid analysis. This provided a dataset of 3,408 curated identifications with which to build a machine learning model. Though I worked closely with Steven Tavis to identify metrics useful as predictors, he wrote all the Python code to construct the machine learning model and computationally validate its performance.

This project was born from a practical need I identified in my research. Through months of work, we have successfully answered that need and provided a high-throughput way of rapidly identifying lipids from untargeted LC-MS data. Beyond that we have built a novel lipid identification workflow that is available to any practitioner of lipidomics. We have also provided a strong proof-of-principle that this technique ought to be adopted by the developers of untargeted lipidomics software to natively provide high confidence lipid identifications and it is my hope that this project can help initiate a paradigm shift in lipidomics resulting in higher confidence analyses. This project would likely not have been possible had I not already spent hundreds of hours validating lipid identifications in the LC-MS data. The expertise in parsing tandem MS data that I gained in this process has been summarized in the methods section of this chapter. This information itself should be a valuable resource to lipidomics investigators. Tandem MS discussion aside, this chapter is mostly derived from a manuscript currently under review at *Analytical Chemistry*.

#### **Abstract:**

In untargeted lipidomics experiments, putative lipid identifications generated by automated analysis software require substantial manual filtering to arrive at useably high confidence data. However, identification softwares do not make full use of the available data to assess the quality of lipid identifications. Here, we present a machine-learning based model to provide coherent, holistic quality scores based on multiple lines of evidence. Underutilized metrics such as isotope ratios and chromatographic behavior allow for much higher accuracy of identification confidence. We find that approximately 50% of tandem mass spectrometry-based automated lipid identifications are incorrect but that multidimensional rescoring reduces false discoveries to only 7% while retaining 80% of true positives. Our method works with most chromatography methods and generalizes across a family of MS instruments. LipoCLEAN is available at <https://github.com/stavis1/LipoCLEAN>.

### **Introduction:**

Any untargeted omics-scale analysis of lipid samples requires automated tools for the interpretation of mass spectrometry (MS) data. However, there is a growing consensus in the lipidomics community that the accuracy of current automated analysis software is insufficient for most tasks and that these tools must be supplemented with time consuming manual filtering.<sup>20, 63, 64</sup> It is not surprising that this is the case, since the large majority of chromatographic features in a typical lipidomics experiment are likely the result of degeneracy or contamination.<sup>70, 71</sup> Most automated analysis tools only consider a small fraction of the potential degenerate forms of a lipid (such as adduction, in-source fragmentation, etc.), which means that most features are impossible for these tools to correctly annotate. Because the base rate of features being annotatable lipids is so low, even tools with a very low per-feature error rate can end up yielding a large percentage of incorrect calls. Additionally, there are a variety of related lipid classes and species which share many mass spectral characteristics, increasing the chance of non-random false identifications. These performance issues of lipid identification software have contributed to confusion regarding the confidence and specificity of lipid identifications, which has in turn led to controversy over reported lipidomics results.<sup>20, 64-67</sup>

Previous attempts to address this issue have revealed that about half of automated identifications are incorrect. Aicheler *et al.* used retention time predictions to filter the results of OpenMS<sup>72</sup> searches, which removed more than 50% of the initial identifications.<sup>73</sup> This proof of concept work demonstrates the power of including more diverse information in the final

classification of a lipid but it requires a new training dataset for each chromatography method. Collisional cross section (CCS) data were used by Rose *et al.*<sup>74</sup> to filter the results of MS<sup>1</sup> only identifications from SIFTER.<sup>75</sup> The addition of this extra line of evidence removed 47% of initial putative identifications. Unfortunately, this approach both requires CCS measurements, which are not possible on the majority of mass spectrometers, and requires that all putative lipids have enough correctly annotated members of their class for a class-specific regression to be constructed.

We chose to use MS-DIAL<sup>48,49</sup> as the starting point for our filtering tool. The lack of rigorous benching studies for untargeted lipidomics analysis software makes the choice of program difficult; however, MS-DIAL works for a wide variety of data types (chromatography, ion mobility, direct infusion, etc.) and is capable of processing both lipidomics and metabolomics data. The relative popularity of software tools is hard to access. For example, proprietary software like Compound Discoverer frequently do not have publications describing their introduction, while multi-purpose tools like MZmine<sup>50</sup> and MS-DIAL are cited for more applications than lipidomic specific tools like Lipid Data Analyzer.<sup>43</sup> That said, MS-DIAL is certainly a common choice for data analysis, with its first publication garnering 262 citations per year, while the paper introducing MS-DIAL 4 has received 132 citations per year (based on a Google Scholar search using Publish or Perish 8).<sup>48,49,51</sup> Our lab has preferred the use of MS-DIAL for lipidomics experiments due to its user friendliness, free and open source nature, and good performance in internal tests relative to other programs. Despite this, we have observed non-negligible error-rates that led us to develop LipoCLEAN, a machine-learning based post-processing tool for filtering MS-DIAL-derived lipid identifications from untargeted lipidomics experiments.

Confidence in lipid identifications is substantially improved by using LipoCLEAN to combine multiple mechanistically distinct lines of evidence into a single coherent score. The use of machine learning for this task allows subtle information to be extracted from the relationships between multiple lines of evidence. We are particularly concerned with ensuring that our approach is applicable to a broad range of lipidomics experiments. To that end, we perform *m/z* corrections and nonlinear retention time mapping, which makes our approach robust to differences in instrument calibration and chromatography methods.

## **Materials and Methods:**

### *Manual Classification Strategy*

Lipid identifications reported by MS-DIAL (version 4.60 was used throughout) were validated by a manual inspection of the data, as detailed below. While MS-DIAL will often report lipid annotations at the molecular species level (as defined by the LIPID MAPS convention), we considered annotations as specific only to the species level (where fatty acid tails are summed and reported based on the overall carbon count and degrees of unsaturation).<sup>76</sup> This was based on the observation that tail isomers are not always chromatographically resolved, making the molecular species level distinctions somewhat arbitrary. Indeed, many fragmentation spectra show pairs of fatty acids that correspond to different molecular species that have the same summed carbon count/degrees of unsaturation (see example in **Figure 2-1**). This level of specificity is achievable by many different experimental designs and is sufficient for most applications of untargeted lipidomics.

To be manually validated, a feature must have a chromatographic peak of moderate quality at the specified retention time, an isotopic envelope broadly consistent with the proposed lipid, and tandem mass spectrometry scan(s) with the presence of expected fragment ions from the given lipid species and with major fragment ions explainable by the lipid species or observed co-isolated ions. Lipids that were considered high confidence based on manual validation were used as retention time anchors for additional validation of lipid annotations. Lipid retention times are known to have predictable patterns in reversed phase chromatography with a roughly linear mass dependency for a given subclass of lipid.<sup>20, 77</sup> Lipids falling outside a plausible range from this retention relationship were identified as false positives, while annotations with borderline fragmentation patterns at an expected retention time were confirmed as true positives. This validation workflow is summarized in **Figures 2-2** and **2-3**.

### *Lipid Tandem Mass Spectrometry Fragmentation Analysis*

In order to validate lipid identifications using tandem MS data, the fragmentation patterns of the analytes must be characterized. Lipids fragment in predictable ways and lipid classes share fragmentation pathways. The fragmentation patterns of several lipids classes are described below. While many of the details here have been synthesized over time through reading literature and manually characterized observed fragmentation patterns, I do owe a specific debt to the LIPID MAPS fragment ion predictor and the work of Pulfer and Murphy that this was built on.<sup>78</sup>

I have tried to reference other studies that have been useful in specific instances below, but readers should be aware this is not an exhaustive listing of the relevant literature.

Phosphatidylglycerol (PG) lipids predominately ionize in negative mode as  $[M-H]^-$ , and only the negative mode fragmentation will be considered here. An example structure of a PG lipid ion (PG 16:0/16:0 specifically) is shown in **Figure 2-4.A**. Similar to most ions in collisional induced dissociation (CID), PG lipids and all the lipids I cover here, fragment at heteroatoms in the structure. This is because carbon-carbon bonds are quite strong and are unlikely to break without a much more energetic dissociation process. For many lipids classes containing fatty acids moieties, the most intense fragment ions are the fatty acid fragment ions. These dissociate from the parent ion by a heterolytic cleavage between the fatty acid and central glycerol moiety. Carboxylic acids are very good leaving groups and thus are often part of fragmentation spectra. A PG fatty acid fragment ion example can be seen in **Figure 2-4.B**. Neutral loss of the fatty acids moieties can also generate fragment ions. There are two mechanisms that can generate the neutral loss of the fatty acids. One mechanism loses fatty acids as a carboxylic acid (**Figure 2-4.C**) and the other mechanism forms a ketene (**Figure 2-4.D**). Both mechanisms have subtle rearrangements which results in the net charge remaining on the parent ion which is now missing the fatty acid and is appropriately shifted in mass. Finally, PG lipids can also fragment by means of a combined ketene loss and glycerol headgroup loss (**Figure 2-4.E**). These ions can all be seen in a fragmentation spectrum of a PG 16:0/16:0 lipid in **Figure 2-5**.

The fragmentation of PE lipids in negative mode (as a  $[M-H]^-$  ion) is very similar to that of PG. Phosphatidylethanolamine (PE) ionizes approximately equally well in positive and negative mode, which reflects its natural zwitterionic character. The negative mode PE fragmentation is more informative than positive mode and can be used to discern the specific lipid tails present. Fatty acid fragment ions and fatty acids neutral losses (both ketene and carboxylic) can all be observed from these lipids, just as in PG tandem mass spectrometry (see **Figure 2-6.A-D**). Unlike PG, however, there are no combination ketene and headgroup losses from PE lipids. An example spectrum is shown in **Figure 2-7**. In positive mode, the fragmentation is much less rich and typically only the headgroup neutral loss is observed.

Phosphatidylcholine (PC) lipids are also naturally zwitterionic. These will ionize readily in positive mode, however, like PE, their fragmentation will only generate a headgroup ion. PC lipids cannot ionize in negative mode as an  $[M-H]^-$  ion since they have a permanent positive

charge in the choline headgroup and can only dissociate a single proton from its phosphate group. However, PC lipids will adduct with ionization additives such as acetate or formate and form negative ions nonetheless. Given the chromatography solvents I use have ammonium acetate as the ionization modifier, I tend to largely observe  $[M+\text{acetate}]^-$  ions (see **Figure 2-8.A**). Fragmentation of this adduct states predominately results in loss of methyl acetate, however fatty acids ions and combination methyl acetate and fatty acid neutral losses can also be observed (**Figure 2-8.B-D**). However, as can be observed in **Figure 2-9**, these other fragment ions can be of rather low intensity.

Phosphatidylserine (PS) can ionize in positive mode, however I have not undertaken to characterize the expected fragmentation pattern on PS in positive mode. In negative mode PS, as  $[M-H]^-$  ion, will fragment at the fatty acids, forming negatively charged fatty acid fragment ions (**Figure 2-10.A-B**). In addition to the fatty acids, the serine headgroup also shows a strong propensity to dissociate. This gives rise to serine neutral loss ions as well as combination ketene/carboxylic fatty acid plus serine neutral losses (**Figure 2-10.C-E**).

Phosphatidylinositol (PI) lipids lack an obvious charge carrier with which to form positive ions but rather readily ionize in negative mode as  $[M-H]^-$  ions (**Figure 2-12.A**). These ions can give very rich fragmentation patterns similar to those described above with ketene and carboxylic fatty acid losses, combination fatty acid (ketene and carboxylic) and inositol losses, and fatty acid fragment ions (**Figure 2-12.B-G**). There are additional inositolphosphate ions that can be formed as well as glycerophosphate-inositol ions (**Figure 2-12.H**). The studies performed by Hsu and Turk as well as Wenk *et al.* were helpful in confirming these fragmentation spectra.<sup>79</sup>  
<sup>80</sup> An example fragmentation spectrum of PI 18:2\_18:2 is shown in **Figure 2-13**.

Acylcarnitines have a quaternary amine with a permanent positive charge can be detected in positive ion mode (see **Figure 2-14.A**). This has a relatively simple fragmentation spectrum, showing a trimethylamine neutral loss and the carnitine ion (carboxylic acid fatty acid neutral loss), see **Figure 2-14.B-C**. However, should the MS2 scan include the 50-100  $m/z$  range, relatively abundance trimethylammonium and hydroxyfuran ions (fragments of carnitine) will also be observed (see **Figure 2-14.D**) An example spectrum of a 16:0 Car is shown in **Figure 2-15**. This spectrum was from a quadrupole-Exactive instrument. All the other spectra shown in this chapter were generated from and LTQ-Orbitrap Pro. This fragmentation pathway is based on the descriptions elucidated in previous publications.<sup>81, 82</sup>

Acylglycerols including triacylglycerol (TG) and diacylglycerol (DG) both fragment similarly in CID mass spectrometry. TG and DG will form positively charged ions via adduction with ammonium ( $\text{NH}_4^+$ ). The fragmentation of TG is shown in **Figure 2-16**, but DG also follows a similar pattern. The most abundant fragmentation ions are combination fatty acid (carboxylic)-ammonia neutral losses.<sup>83, 84</sup> This directly shows the constituent fatty acid that make up the acylglycerol. Other fragment ions have been described, including fatty acid ions and ammonia and water neutral losses.<sup>84</sup> However, as can be seen in **Figure 2-17**, these are often of low relative intensity.

Ceramides (Cer) are a broad class of lipids that contain a sphingosine and a fatty acid moiety. I have worked with the data derived from ceramides with non-hydroxy (or “normal”) fatty acids combined with a standard sphingosine (Cer\_NS) or dihydroxysphingosine (Cer\_NDS). While I don’t have all the details of the fragmentation of these lipids fully characterized, I do have certain possible fragmentation structures that are generally consistent with prior work with ceramides.<sup>85, 86</sup> These ceramide seem to ionize both as  $[\text{M}+\text{H}]^+$  ions but can also undergo in-source fragmentation to form  $[\text{M}-\text{H}_2\text{O}+\text{H}]^+$  ions. These ion forms generate somewhat different fragmentation spectra. Water loss from the  $[\text{M}+\text{H}]^+$  ion is clearly a favorable fragmentation pathway, and water loss is observed in the tandem MS. Along with that are sphingosine fragment ions with up to two water losses. The  $[\text{M}-\text{H}_2\text{O}+\text{H}]^+$  ion can undergo an additional water loss, or a formaldehyde neutral loss. Other fragment peaks observed include a sphingosine fragment with two water losses, a sphingosine fragment with a water and a formaldehyde neutral losses, and a fatty amide fragment ion. Some example fragment ions can be seen in **Figure 2-18**. The sphingosine and fatty amide fragments appear to be unique to Cer\_NS while the sphingosine fragment minus water and formaldehyde seems to be mostly unique to Cer\_NDS.

These fragmentation descriptions cover most of the classes that were manually annotated. There were also class derivatives, like lyso- or ether versions of the described lipids. These generally fragment similarly, although fatty ether tails do not generally fragment. Tandem MS only one of several metrics used to assess annotation accuracy, and low quality fragmentation spectra did not always lead to the feature being dismissed, rather a wholistic view of all the evidence was assessed.

### Datasets Used

Four datasets were utilized for the training and validation of LipoCLEAN (see **Table 2-1**). Two were generated in-house, and two were taken from the MetaboLights repository.<sup>54</sup> The in-house datasets were derived from two different fungal species, *Laccaria bicolor* strain S238N, an ectomycorrhizal fungus of poplar, and *Aspergillus fumigatus* strain AFU293, an opportunistic human pathogen. Lipids were extracted using a Folch-derivative<sup>60</sup> and analyzed by nano liquid chromatography-tandem mass spectrometry using a Dionex UltiMate 3000 coupled to a ThermoFisher LTQ-Orbitrap Velos Pro. Reversed-phase chromatography was used (C18) along with a top-ten data-dependent acquisition MS method. Fullscan measurements were performed with the Orbitrap analyzer at 30,000 resolution from 150-2000  $m/z$ . Fragmentation scans were performed at 33 normalized collision energy with a 2  $m/z$  isolation window.

### Software Architecture

LipoCLEAN works with text files exported from MS-DIAL with no additional operations by the user required. A fuller description of the internal processing will be provided in the publication, but a simplified explanation is given here. Five parameters regarding the putative identifications are used as predictors. These are all reported directly by MS-DIAL or calculatable from what MS-DIAL reports. These include the MS<sup>2</sup> dot product, the signal-to-noise ratio of the feature, the mass error (after *in-silico* recalibration), the isotope intensity mean-squared-error, and the retention time error. MS-DIAL reports an expected retention time for all its identifications. These are based on measured or predicted retention times defined by their in-house reverse phase chromatography method.<sup>49</sup> Since our chromatography method is also reversed phase, the elution order of lipids should be conserved despite differences in the exact gradients. A LOWESS regression of the MS-DIAL expected retention times vs the observed retention times is performed which related observed retention times to their expected time in MS-DIAL's standardized method. Deviation from this regression provides the retention time error predictor. A random forest of linear trees is used to filter all putative lipid identifications into “predicted positive”, “predicted negative”, and “requires manual validation” categories (see **Figure 2-19** for graphical representation of the model).

### *Model Development*

After substantial preliminary investigations based on datasets 2-4, sixteen different models were tested to identify the best model architecture. For the preliminary models (**Figure 2-19.A and B**), logistic regressions and gradient boosted decision trees were investigated. For the final model, logistic regression, gradient boosted decision trees, Bayesian additive regression trees,<sup>87</sup> and a random forest of linear trees were tested. All testing during this final development was conducted on datasets 2 and 3 (**Table 2-1**). Decisions were made on the basis of 5-fold cross validation experiments on these datasets. The validation dataset, number 1, was not used for decision making until the final validation experiment (**Figure 2-20**).

### *Instrument Cross Validation*

For datasets 2-4 (**Table 2-1**), we split the data 80:20 into train and test subsets. A model was trained on the train subset for each dataset independently. Each of the three models were then tested on all three datasets. When they were tested on the dataset they were trained on, only the test split was used. When the model was tested on a different dataset, the whole dataset was used.

### *Lipid Class Cross Validation*

Lipid class validation was conducted on datasets 2 and 3 (**Table 2-1**). For each lipid class, we trained a new model on all data except for the lipids in that class. We then evaluated the model on the lipids in our test class. We assessed the predictor distributions for the outlier lipids phosphatidylinositol (PI) and phosphatidylglycerol (PG) lipids. First, all training data was subset down to only annotated positive entries. Then, for PI and PG separately, we fit both a PLS-DA (partial least squares-discriminant analysis) model and a logistic regression model that used the predictors from the final classifier (**Figure 2-19.C**) to predict whether a lipid belonged to the query class.

### *Model Training Stability Test*

We constructed 32 random 80:20 train-test splits of datasets 2 and 3 (**Table 2-1**). For each split, we trained a new model on the training subset and assessed its performance on the test subset. The outcome of this test was determined by a visual inspection of a plot of the receiver operating characteristic curves for these test sets (**Figure 2-21**).

### *Using LipoCLEAN*

LipoCLEAN was written in python and can be run as a python module, a docker container, or a standalone command-line executable. Installation instructions can be found at <https://github.com/stavis1/LipoCLEAN>. All parameters and a list of MS-DIAL export files to process are passed to the tool in a single TOML formatted text document. An example parameters file can be generated by the tool with the '--print' command line option. The elution time corresponding to the column void volume, which is used in the initial filtering step, can be set on a per-file basis. The flexibility of the LOWESS retention time regression and the score cutoffs used by the models (**Figure 2-19**) can also be controlled. The tool produces an extensive set of quality control plots which can be used to make decisions about these parameters.

### **Results and Discussion:**

Automated lipid identifications are not of sufficient quality to use for most purposes without manual reanalysis. The observed false discovery rate for MS-DIAL, among the lipids we manually annotated, ranged from 35-62% (**Table 2-1**). One common source of error was the identification of double  $^{13}\text{C}$  isotopologues of a given lipid as a species with one less degree of unsaturation. For example, the double carbon isotope of phosphatidylethanolamine (PE) 32:1 could be mis-annotated as PE 32:0 since the species would have very close masses and fragmentation patterns. This can be recognized by examining the isotope pattern, setting a tight mass accuracy bound, or investigating the fragmentation pattern. Another common source of false positives includes noise regions being identified and integrated as a lipid despite being minutes removed from the actual peak. Examples of true and false positive identifications are shown in **Figure 2-2**.

An important caveat to the observed false discovery rates in **Table 2-1** is that our manual reanalysis strategy was not based on a simple random sampling of putative lipid identifications. Because of this, we cannot make unbiased estimates of the false discovery rate among all putative lipid identifications. However, our analysis covered 25 lipid classes (**Table 2-2**), including more than 100 examples each from triacylglycerol, cardiolipin, phosphatidylethanolamine, phosphatidylcholine, phosphatidylglycerol, phosphatidylserine, phosphatidylinositol, and diacylglycerol lipids. This nonrandom sampling of lipids is a consequence of our annotation strategy, which requires a detailed understanding of fragmentation patterns and benefits from class specific retention time regressions.

Our analysis is strictly concerned with putative identifications that, *a priori*, have a reasonable chance of providing a confident match to a biological lipid. We use a series of rules-based filters to segregate entries that fail this criterion into a ‘not considered’ text file. In the text export which our tool operates on, MS-DIAL will report all observed features and will provide some suggested annotations based on mass and retention time for features without MS<sup>2</sup> spectra. It is reasonable to export this information, as it can be useful in many contexts, but our tool is solely concerned with providing high confidence identifications. Doing so requires MS<sup>2</sup> spectral analysis and any entry without an MS<sup>2</sup> match is removed. Among those that remain, we found that 12% of entries across our datasets either had no chemical formula or the formula was for deuterated standards which are non-biological. A further 0.7% were not lipids, and of the remaining entries 1.4% had retention times in the void volume. All of these entries are also filtered out.

Different instruments can plausibly produce different distributions of the predictor variables which could limit the generalizability of our model. Before finalizing our model, we tested this by training three separate models, one for each of the datasets 2-4 (**Table 2-1**). Each of these models were then tested on each of the training datasets in turn (**Figure 2-22**). Dataset 4, which was acquired on a TripleTOF 6600, was found to not be compatible with either of the other datasets (**Figure 2-22 A, D, H, and I**). This is primarily due to differences in the distribution of dot products and m/z errors (**Figure 2-23**), which indicates that the compatibility is due to the MS instrument type and not differences in the chromatography method. The likely cause of this discrepancy is the lower resolution and mass accuracy of TOF instruments compared to Orbitrap based systems. Although there is an excess of false positives when using a model trained on Orbitrap Velos Pro data to filter Q Exactive data (**Figure 2-22.B**), the compatibility between these two instruments is substantially higher (**Figure 2-22.B, F, and 2-24**). In a series of cross validation experiments on a model trained on datasets 2 and 3 (**Figure 2-21**), we found that the false positives observed above had only an 8% probability of being classified as positive by the combined model in a held-out subset. Taken together, these findings indicate that a model trained on datasets 2-3 should work well for the Orbitrap instrument family. More generally, we expect that a trained model should generalize within an instrument family but large differences between instruments will require new models.

Having identified compatible datasets, we trained an Orbitrap model on datasets 2 and 3 (**Table 2-1**). The performance of this model was assessed on dataset 1 (also Orbitrap). Since this was a hold-out set not used for model development, it will provide unbiased estimates of real-world model behavior (**Figure 2-20**). Our model achieved good separation between true and false positives with an area under the receiver operating characteristic (ROC) curve of 0.97. At the cutoff for predicting a positive ID, our observed false discovery rate was 7% and we identified 80% of the annotated true positives. Including the reanalyze category in this analysis identifies 95% of annotated true positives while worsening the false discovery rate to 17%.

While our Orbitrap model already shows good performance metrics, there still is tension between the desire to use the maximum amount of data for training the deployed model, and the need to validate the quality of the model on a dataset that has never been used for decision making. One potential solution to this problem is to train a preliminary model on all of the data used in development, assess its performance on the pristine, held-out dataset, then train the deployment model on all data together. This only works if the training process very repeatedly produces models with similar performance. We assessed the validity of this requirement by generating 32 random 80:20 train-test splits of datasets 2 and 3 (**Table 2-1**), then training a new model on each train subset, and finally assessing its performance on the test subset. The quality of these models were all highly similar, as assessed by an inspection of the ROC curves (**Figure 2-21**). A consequence of this is that we can safely train the final, distributed model on datasets 1-3 and assume that its performance will be at least as good as the performance of the model trained on datasets 2 and 3 and validated on dataset 1 (**Figure 2-20**). We did not include dataset 4 in this model due to the incompatibility results shown in **Figure 2-22**. However, we also trained a separate model on dataset 4. This model achieved a test set AUC ROC of 0.95, and at the predicted positive cutoff we observed a false discovery rate of 9% and a recall of 79%. Both trained models are available at <https://github.com/stavis1/LipoCLEAN/releases>. To assess the runtime of LipoCLEAN, we ran the model trained with datasets 1-3 on seven previous lipidomics experiments run in our lab (data not shown). When generating no quality control figures, the tool took an average of 9 seconds per file, whereas when all plots were generated it took a mean time of 30 seconds.

Predictor distributions could also plausibly vary by lipid class. To assess this, we ran leave-one-class-out cross validation experiments in which each class in turn was censored from

datasets 1-3 (**Table 2-1**), a new model was trained on this subset, and this model was tested on the held out lipid class data (**Figure 2-25**). For most classes the model translated well to the new domain. However, the models trained without PG or PI lipids had poor performance on their test sets. This poor performance does not appear to be the result of a substantial shift in the underlying predictor distributions (**Figure 2-26**), and the number of positive examples in both groups is relatively small so these results could be simple sampling error artifacts. No decisions about the model architecture were made on the basis of these results because the analysis was conducted after finalizing the model, so it was possible to use datasets 1-3 in this experiment.

### **Conclusions:**

In this work, LipoCLEAN has been developed for filtering putative lipid identifications from RPLC-MS/MS experiments analyzed with MS-DIAL. Our tool uses a machine learning approach to coherently and non-linearly integrate multiple mechanistically distinct lines of evidence into a single quality score. Filtering on this score substantially improves the reliability of automated lipid identifications over the use of MS-DIAL's scoring function, with a reduction in FDR from 45% to 7%. The high degree of improvement observed here shows that, much like in the related field of proteomics, multidimensional rescoring should be included as a standard step in untargeted lipidomics processing pipelines. The inclusion of this step represents a trivial fraction of the total processing time and is therefore scalable to large and complex experiments. We have shown that a trained model can be used on multiple instruments within an instrument family. We have also shown that a trained model will usually generalize well to lipid classes that were not present in its training dataset. We therefore encourage lipidomics researchers to evaluate LipoCLEAN for their analysis pipeline, as well as performing their own manual validation of automated identifications to further validate LipoCLEAN and extend its applicability to more instrument types. It is our eventual hope, however, that authors of lipidomics search tools will incorporate similar rescoring models directly into their software. This will allow for higher confidence identifications with little to no effect on software runtime while requiring the training of, most likely, only a few models to cover the majority of lipidomics workflows.

## **Chapter III - A Critical Examination of the Statistical Properties of Lipidomics Data**

### **Citation:**

The figures and text of this chapter are all my contribution. This work will likely be combined with other research into a manuscript describing statistical properties of lipidomics and proteomics data. Collaborators on this broader project, and likely coauthors, include Steven Tavis, Diana Ramirez, Frank Löffler and Robert L. Hettich. For the work described in this dissertation, Diana Ramirez in Frank Löffler's lab grew the bacterial samples. Steven Tavis, Diana Ramirez, and I harvested the samples. Andrew Stai assisted me in the lipid extractions. I performed the measurements and data analysis with suggestions from Steven Tavis.

### **Abstract:**

Limited validation of statistical analyses and approaches represents a major limitation of current lipidomics experiments. There has been no empirical assessment of the accuracy and sensitivity of commonly used statistical tests in lipidomics data. Additionally, the distributional properties of LC-MS omics data is uncharacterized. This makes the choice of any given statistical analysis procedure unprincipled and unreliable. Here we address this gap by performing a large-scale replication project with 92 samples split between only two conditions. These samples were generated from bacterial isolates cultured with and without solvent stress. Care was taken to ensure the large sample size did not introduce unusual variability in the data. From these data, a variety of statistical properties are assessed, including the normality and precision of replicate measurements. Certain lipidome summary statistic values are introduced and partially validated on the large-scale dataset. These summary values are useful in providing a biologically interpretable quantitative summary of the overall state of the lipidome. Finally, a comparison of different quantification procedures is provided that demonstrates manual quantification is superior to automated quantification. Additionally, among automated quantification procedures, peak height is found to be more reliable than peak area for quantification due to common errors in peak boundary identification. This project provides an important test of the statistical properties of lipidomics data and provides a large dataset that can be mined for further insights into lipidomics and metabolomics data analysis.

### **Introduction:**

Despite the wide adoption of metabolomic and lipidomic analyses, there is little reported in the way of rigorous benchmarking of the statistical characteristics of the data generated. Most work relating to the statistical analysis of metabolomic/lipidomic data has focused on best

practices to limit variability and avoid introducing bias. Broadhurst *et al.* writing in 2018, detailed useful procedures for the use of quality controls to correct for measurement drift and batch effects.<sup>88</sup> They also describe procedures to quantify variability at different stages of the analytical pipeline. While Broadhurst *et al.* mention the possibility the measured data may not have Gaussian noise (that is, be normally distributed), they do not specifically address what distributional properties the data are expected to have nor what statistical procedures would be appropriate. Another paper which gets closer to addressing the statistical properties of lipidomics data was published a couple years earlier by Trutschel *et al.*<sup>69</sup> Here they apply a hierarchical design of experiment to quantify sources of variability in metabolomics. Working with plant samples (*Arabidopsis* leaves), they find that biological variability produces about 54% of total variability, while sample preparation variability and analytical variability account for 29% and 17% of the total variability, respectively. The scale of the variability was such that the commonly used biological triplicate only had sufficient power ( $\geq 80\%$ ) to reliably detect a 1.9 log-fold change (difference of the means of the log data). Here too, though, the focus was predominately on variability estimation and not on the distributional or other statistical properties that have to be satisfied for common tests and analyses. Indeed, outside of epidemiological studies, there have been no empirical measurements of what statistical assumptions hold true for metabolomics or lipidomics data. Some have attempted to examine the normality assumption by plotting the density of observed intensities across an entire sample or dataset. Such operations generally produce results that appear to be roughly log-normal.<sup>89</sup> This is not the correct assessment, however, since the common statistical approaches employed (such as ANOVA or t-tests) are analyte-specific operations. Each t-test in a binary comparison, for example, is calculated for each analyte individually and independently from all other analytes. Thus, the distribution of analytes intensities across a sample is not relevant, but rather the distribution of intensities for a given analyte across biological replicates is what matters. Indeed, the t-test (and most hypothesis tests) base their assumptions on the null hypothesis, and therefore, it is the properties of the null (biological replicates of a given analyte) that matter. Unfortunately, no study exists with sufficient replication to assess the per-analyte distributional properties of lipidomics data. Interestingly, there can be very dramatic differences in the statistically “significant” results between data processed with and without log transformation. **Figure 3-1** shows a scatterplot of the p-values obtained from a dataset of 108 lipid species from *Clostridium autoethanogenum*.

Indeed, surveying a range of publicly available and in-house LC-MS omics experiments shows high dissimilarity between Benjamini-Hochberg “significant” results obtained with and without log-transformation (**Figure 3-2**).

There have been certain other attempts to assess reproducibility in lipidomics. However, these studies focus on specific aspects outside the scope of statistical robustness in untargeted lipidomics. For example, an international “ring trial” was performed where participants used a common approach albeit with different instruments across many different labs to quantify variation in results.<sup>90</sup> However, the approach used was a targeted lipidomics workflow and absolute quantitation was performed. This is a different workflow and analysis type that does not immediately apply to untargeted lipidomics with relative quantitation. In order to elucidate the statistical properties of the lipidomics data and to validate appropriate statistical tests and procedures, a large-scale replication experiment was performed. Here I generated forty-six biological replicates of a broth-culture bacterium in two different growth conditions for a total of ninety-two samples. This allows for a quantitative assessment of the distributional properties of untargeted lipidomics data. The bacterial samples were grown in the Löffler lab with the assistance of Diana Ramirez. I am also indebted to Steven Tavis for providing the initial inspiration for this project and his guidance in statistical matters.

### **Materials and Methods:**

#### *Bacterial Culture and Sampling*

*Pseudomonas putida* strain GB1 was cultured in a mineral salt medium with 25 mM acetate as the carbon source. The medium has been described previously and contained 6.34 g/L of Na<sub>2</sub>HPO<sub>4</sub>, 1.33 g/L of KH<sub>2</sub>PO<sub>4</sub>, 1.0 g/L of (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.2 g/L of MgSO<sub>4</sub>·7H<sub>2</sub>O, the Wolin vitamin mix,<sup>91</sup> and a trace metal solution.<sup>92, 93</sup> In order to test statistical repeatability and population-level estimands, a high replicate count was necessary. To that end, 96 samples were cultured with 48 control samples containing just the minimal media and 48 solvent stress samples where the minimal medium was supplemented with both 2 µM perfluorooctanesulfonic acid (PFOS) and 0.6 mM trichloroethylene (TCE). PFOS and TCE were used because of their environmental relevance, but the focus this research project was the statistical properties of the data, not the immediate biological interpretation of the data. Cultures were performed in 50 mL of media in rubber-stoppered bottles. Cultures were grown for approximately 16 hours while shaking at 120 RPM. Previous tests had found our sample collection took 10 minutes for every

set of six samples (sample collection included a second sampling of each culture for a proteomics experiment not reported here). The starting inoculation was staggered so that sample collection on all 96 samples would occur at approximately the same time post-inoculation. An sample of about 20 mL of the culture was taken and centrifuged to pellet the cells. The supernatant was decanted and samples frozen at  $-80^{\circ}\text{C}$  before extraction.

#### *Lipid Extraction Procedure*

A Matyash lipid extraction protocol derivative was used to isolate bacterial lipids.<sup>62</sup> The procedure was as follows. First, 700  $\mu\text{L}$  of methyl *tert*-butyl ether (MTBE), 210  $\mu\text{L}$  of methanol, and 170  $\mu\text{L}$  of water were added to each sample. Bacterial cells were suspended in the solvent and transferred to 2 mL centrifuge tubes. These were then bath sonicated for approximately 5 minutes to help disrupt cells and dissolve lipids. Following this, the samples were incubated on ice for about 40 minutes to allow for lipid partitioning to the organic phase. Centrifugation was then performed at 10k rpm for 7 min. The upper phase of the resulting supernatant was collected and transferred to a separate vial. 600  $\mu\text{L}$  of an upper phase resulting from a 4:1.25:1 mixture of MTBE, methanol, and water was added to the remaining sample for an additional lipid collection. After resuspension, samples were subjected to mixing on an orbital shaker while sitting on ice. Following this, samples were centrifuged and the resulting upper phase was collected and mixed with the first collection. Lipid extracts were dried by vacuum centrifugation. For an internal standard, 2  $\mu\text{L}$  of the EquiSplash internal standard mixture (Avanti Polar Lipids) was added to each sample. Lipids were then dissolved with the addition of 200  $\mu\text{L}$  of 1:1:1 butanol, isopropanol, and water. Bath sonication was employed to help dissolve lipids. Samples were aliquot into autosamplers and kept in a refrigerated ( $4^{\circ}\text{C}$ ) autosampler prior to analysis.

#### *Lipidomics Instrumental Analysis*

A split-flow nano liquid chromatography method was used with a 60:40 acetonitrile and water mobile phase A and a 90:10:1 isopropanol, acetonitrile, and water mobile phase B. Both mobile phase A and B also contained 0.42 g/L ammonium acetate as an ionization additive. A multistep gradient was employed as described in **Table 3-1**. The chromatographic flow rate was set to 0.05 mL/min, but since splitflow chromatography was employed, the exact flow rate through the column is uncertain (probably on the order of hundreds of nL per minute). The column employed was made in-lab from a 100  $\mu\text{m}$  inner diameter silica capillary. The column was filled with Kinetex 1.7  $\mu\text{m}$  C18 stationary phase beads with 100 Å pores (Phenomenex). The

mass spectrometer was operated with a top 10 DDA method. One full scan was acquired at 30,000 resolution in the orbitrap followed by ten ion trap scans sequentially fragmenting the most abundant ions. This was performed 33 normalized collisional energy (CID) with a 2  $m/z$  isolation window. TCE and control samples were alternated in the acquisition queue to minimize analytical bias from instrumental drift.

#### *Lipid Identification and Quantitation*

Phosphatidylglycerol, phosphatidylethanolamine, and cardiolipin lipid species were identified and quantified in a purely manual manner, the project having commenced before LipoCLEAN was fully validated. This manual analysis involved the examination of observed extracted ion chromatograms (at a mass tolerance of 10 parts per million) and fragmentation spectra (if any) for PG and PE lipids with between 20 and 40 carbons and between 0 and 2 degrees of unsaturation. The CL analysis was performed similarly searching for anything between 60 and 80 carbon with between 0 to 6 degrees of unsaturation. Retention time vs  $m/z$  regressions were also performed to confirm identifications and check for outliers. Quantification was performed by manual integration of the extracted ion chromatograms using the EI-MAVEN program.<sup>94</sup> After LipoCLEAN was validated, MS-DIAL with LipoCLEAN post processing was used to generate a list of lipids to compare MS-DIAL's automated extracted ion chromatogram quantitative with the manual quantitation performed in EI-MAVEN. LipoCLEAN-predicted true positive lipids were manually quantified in EI-MAVEN to compare manual quantification to MS-DIAL's reported peak area and peak height quantifications.

#### *Data Processing*

Lipid species were blank-filtered to remove potential lipid contaminants. This was done by removing lipids where neither of the sample averages exceeded the blank average by at least 3 times. To correct for instrumental drift over the course of the analysis, a LOWESS-based drift correction was employed.<sup>88, 95</sup> To correct the solvent stress samples, lipid observations in the control samples were fit with a LOWESS regression and interpolated to predict the effect of drift on the TCE measurements. Each TCE sample intensity was scaled by dividing by the interpolated value and multiplying by the average LOWESS interpolation value. The same procedure was used to correct the control samples except that the solvent stress samples were used to build the LOWESS. Several statistical tests were employed including Scipy's goodness of fit test in Python that tests the plausibility of a given set of data being derived from a defined

distribution.<sup>96, 97</sup> Various bootstrappings of statistical tests (such as the T-test) were also performed in Python, using the Numpy package.<sup>98</sup> Summary statistics of the overall lipid composition were calculated. These were the average fatty acid length (which is a weighted average of fatty acid lengths by lipid intensities), average fatty acid unsaturation (a weighted average of fatty acid unsaturation), and cationic to anionic glycerophospholipid ratio. The cationic to anionic glycerophospholipid ratio is the ratio of summed intensity of cationic/zwitterionic glycerophospholipids classes (in this case PE and LPE) over the summed intensity of anionic glycerophospholipids classes (in this case PG, LPG, and CL).

## **Results and Discussion:**

### *Overall Lipid Composition*

Manual identification of lipids resulted in 78 unique lipid species. These included the LPG, PG, LPE, PE, and CL classes. After blank filtering the data, 70 lipid species remained. Considering *P. putida* has a relatively simple lipidome, this is a reasonable range of species and covers the expected classes. While 96 samples were grown, two samples in each condition were lost during lipid extraction due to operator error. We still had 46 replicates in each condition, which yielded 140 lipid-condition pairs with which lipid distributional properties could be assessed. Across the dataset, very good measurement repeatability was observed with average relative standard deviations (RSDs) of 25% and 21% for the control and TCE conditions respectively. The average RSDs with drift correction improved to 18% and 17%. A stabilization of the repeated measurements was observed across the injection sequence for the majority of lipid species (see **Figure 3-3**). Testing the distributional properties of the drift-corrected data revealed a slight positive skew in the raw data (**Figure 3-4.A**). The log transformed data, while still spread over a range of skews, shows a roughly symmetric distribution of skew, implying log-transformation reduces skew on average (**Figure 3-4.A**). There was a similar effect in reducing excess kurtosis on average; however, in both cases log transformation results in more extreme outliers (**Figure 3-4.B**). Performing a goodness of fit test using Python's Scipy package, the observed distributions were tested for the probability a normal distribution would produce the observed results (see **Figure 3-4.C**). Overall, the log-transformed data were more consistent with an underlying normal distribution than the raw data. This indicates the data are more log-normal than normally distributed; however, neither strongly match the observed distributions (**Figure 3-4.C**). Because of this deviation from normality, the application of T-tests gives biased results.

This can be seen in **Figure 3-5** showing T-tests of bootstrapped empirical null data. This is where random subsets of replicate lipid measurements are drawn and T-tested against each other. Since these are derived from the same population, subsets should give uniform random p-values (between 0 and 1) from T-tests. However, the observed T-test results are biased toward higher p-values. This is true for both raw and the log-transformed data, indicating lipidomics data do not satisfy the statistical assumptions of T-tests. A permutational variant of the Student T-test performs better, although given its nonparametric calculation of significance, it requires at least six replicates to stabilize (**Figure 3-6**).

One could argue that the observed high p-value bias of T-tests may make the tests conservative but still useful for lipidomics data analysis. In fact, a test of expected statistical power shows replicate counts of five or more have good (>80%) power to detect a log<sub>2</sub> fold change of 1.0 (**Figure 3-7**). However, a more significant issue plaguing T-tests is that the point-null used by these tests is *a priori* false.<sup>99, 100</sup> A point null explicates that any deviation of two populations from having precisely equivalent means (to the *nth* decimal place) is a “real” difference. While true in the most technical sense possible, this is an analytically and biologically false proposition. Most significantly, mass spectrometry is sensitive to context (the chemical matrix of a sample) and a change in the concentration of one analyte could affect the measured intensity of another analyte, even if that analyte is unchanging in concentration. This is due to a dependency of the measure intensity of matrix effect, such as charge competition in the electrospray ionization process. This can be visualized by examining the internal standards. The EquiSplash deuterium-labeled internal standard mixture was added post-extraction to all of the samples at equal amounts. These standards are known, then, to be at the same concentration between the control and the solvent stress samples. However, non-equivalent fold-changes are observed with all nine internal standards (see **Figure 3-8.A**), three of which are “significant” at a p-value of < 0.01. This demonstrates the non-zero magnitude of matrix effects which need to be addressed when performing analyte-specific statistical operations. More complex samples and samples with larger real concentration differences will have larger matrix effect-induced spurious fold changes. Because the magnitude of matrix effects is experiment-dependent, a static fold-change cutoff may not generalize well. Therefore, we instead recommend using internal standards to define a null-effect range. This could be achieved in a variety of ways. The simplest option would be to add a fold-change filter defined by the largest absolute “fold-change”

observed in the internal standards, or possibly a confidence interval of the internal standard fold-changes (see **Figure 3-8.B** as an example). Other options would be to use the internal standards to create an interval null that could be used in a two one-sided test (TOST) or second-generation p-value calculation.<sup>101-103</sup> In the current experiment, only 9 internal standards were present and detectable with our method. While this does give a reasonable approximation of the null matrix effect size, using more standards may provide a more robust null range. For future experiments where analyte-specific statistics are desired, using the UltimateSPLAH ONE internal standard mixture from Avanti (with 50 internal standards likely to be measurable in negative mode) is recommended.

### *Lipidome Summary Statistics*

In many applications of lipidomics, analyte-specific information is not of significant value. Even when performing analyte-specific statistics like T-tests or ANOVAs on all observed lipids, many lipidomics studies will still try to interpret these results as a whole rather than a series of separate measurements. For example, the “significantly” changing lipids will be examined for possible trends (quantitatively or otherwise), such as a preponderance of unsaturated lipids or lipids of a (multiple) given class(es).<sup>104</sup> Without defined statistics, this analysis is by its very nature subjective and, therefore, susceptible to interpretation errors and confirmation bias. Instead, I recommend the calculation and statistical analysis of summary statistics that describe trends among lipids. Ideally, these values would be biologically interpretable and statistically robust. To that end, I propose a set of four summary statistics that have reasonable biological interpretations and together capture much about the membrane physiology. These include the average fatty acid length or unsaturation (across all lipids), the ratio of cationic (including those that tend to be zwitterionic) to anionic glycerophospholipids, and the ratio of glycerophospholipids to glycerolipids. Fatty acid length and unsaturation are immediately relevant for membrane fluidity and associated homeoviscous adaptations. The cationic to anionic ratio can indicate changes in lipid class abundances and is relevant for membrane polarity. The glycerophospholipid to glycerolipid ratio relates these two superclasses and, in most organisms, should roughly correspond to the membrane lipid to energy storage lipid ratio. Other summary statistics could and should be developed depending on the specific experiment and science question to be answered, but these specific summaries were used in the fungal lipidomics project discussed in Chapter IV. The statistical properties of these summaries

were tested to verify the applicability of statistical tests to these summaries. However, due to the absence of glycerolipids in *P. putida*, the glycerophospholipid to glycerolipid ratio summary statistic was not assessed. First, the average fatty acid length, average fatty acid unsaturation, and cationic to anionic ratios were calculated individually for all samples (after blank filtering, drift-correction, and mean-centering). The resulting values were tested for normality using Scipy's `goodness_of_fit` test (p-values are shown in **Table 3-2**). The normality tests were inconclusive from the observed data. One condition from each statistic tested generated a low p-value (0.06 or less) and one condition a high p-value (0.4 or greater). Given p-values should be uniformly random values under the null, the fact that low p-values are observed does not preclude an underlying normal distribution. The Shapiro-Wilk test of normality also generated somewhat inconclusive results, although these appear to be more in keeping with the uniform random of an underlying normal distribution for each statistic (**Table 3-2**). Quantile-quantile plots provide a graphical way of assessing the relative normality of a given distribution. These can be seen in **Figure 3-9**. While somewhat noisy, these appear to be generally consistent with an underlying normal distribution. That said, empirical null (T-testing on random subsets of replicate summary values) distributions of p-values shows some deviation from the expected uniform random at n of five replicates (see **Figure 3-10.A**). This may be negligible for the average fatty acid length and cationic to anionic glycerophospholipid ratio, but the average fatty acid unsaturation shows somewhat substantial deviation. Increasing the number of replicates to thirteen, the empirical null stabilizes very close to the theoretical for a normal distribution (**Figure 3-10.B**). That said, the precision of fatty acid summaries is high enough that a change of only a few percent between sets of three replicates is substantial enough to give highly “significant” results (this was tested by applying a pseudo-fold change to empirical null resampling). The cationic/anionic lipid summary required a change of about 20% to give similarly “significant” results. Therefore, while nonparametric tests would be ideal, these statistics (should the precision be replicated in other experiments designs) are likely sufficiently robust for Student T-tests or ANOVAs to be used without absurdity. These summary statistics are also internally normalized (that is they are ratios or relative measurements within a sample) and should, therefore, be more robust to variability in sample biomass and similar systematic errors. That said, the point-null of the T-test could still be problematic for these summary statistics. The EquiSplash internal standard mixture does not

have the fatty acid/lipid diversity to allow for the estimation of the summary statistic point-null. This could be more thoroughly addressed using the Ultimate Splash One mixture.

#### *Assessment of Quantification Method*

Across my research experience, I have previously observed that many untargeted analysis software perform peak integration rather poorly. While many examples can be found of adequate peak identification and integration, there are also many examples of poor and/or inconsistent peak detection and integration. Specific examples of MS-DIAL's failure to consistently integrate peaks is shown in **Figure 3-11**. This is not unique to MS-DIAL, and **Figure 3-12** shows similar failures by XCMS, another commonly used tool. Instead of relying on these software for peak integration, I have consistently performed manual peak integration for the quantitative analysis of lipids. It is possible, however, that automated detection of peak height may be a reasonable metric for quantitative analysis. While generally considered less analytically robust than peak area for quantification, peak height will be more robust to errors in peak boundary detection. So long as the peak apex is captured, the quantitative value will be invariant to the peak boundaries, allowing for inconsistency in boundary detection between samples.

In order to test the validity of different quantification methods, the LipoCLEAN-predicted true positive lipid identifications from MS-DIAL were used to compare MS-DIAL's reported peak area, peak height, and manual peak integration in EI-MAVEN. After combining replicate lipid species IDs and excluding a couple obviously false-positive IDs, 76 lipid species were used to compare quantitation methods. Unsurprisingly, manual integration gave the best precision between replicates, with or without drift correction (see **Table 3-3**). MS-DIAL's peak height slightly outperformed peak area with a 22% RSD compared to 25% RSD (values post drift correction). While the differences between MS-DIAL's area, height, and manual integration were smaller than expected, they did influence the distributional properties (skew and excess kurtosis) of the data as can be seen in **Figures 3-13.A and B**. This also led to decreased normality in the automated quantifications relative to manual integration (**Figure 3-13.C**).

#### **Conclusions:**

This study reports the first large-scale reproducibility experiment performed with LC-MS omics data. The large sample size (46 replicates in two conditions) allows long-run distributional properties and statistical robustness to be assessed. This study has found that lipidomics data deviates from normality whether or not a log-transformation is applied, which in turn causes T-

tests to be biased conservatively. Additionally, spike-ins of standards confirm the theoretical proposition that the point-null assumed by many parametric tests is false *a priori* for LC-MS data. Future work should investigate the utility of statistical tests that make explicit use of an interval null (such as TOST or second-generation p-values). While the distributional properties of certain proposed summary values are sufficiently normal to justify the use of parametric statistics, the average fatty acid unsaturation metric requires several replicates before satisfying the distributional assumptions of the T-test. Finally, the performance of MS-DIAL for lipid quantification is assessed and compared to manual quantification. While manual quantification performed the best, if automated quantification is required for throughput reasons, peak height should be used rather than the more common peak area. A limitation of the current work is that the internal standards used (the EquiSplash mixture) have limited diversity, which limits in turn the robustness of the interval-null estimations. This is especially problematic for the summary statistics which require several different kinds of lipid species and could therefore not be estimated. Additionally, *P. putida* has a relatively simple lipidome covering five lipid classes with only 70 lipid species reliably identified. Expanding this replicability project to a system with a more complex lipidome (such as a fungi) would allow for a deeper analysis of lipid analytes and summary statistics that covered many other lipid classes. Overall, however, this study provides a robust assessment of the data and statistical properties of the lipidomics experiments I participate in with recommendations for statistical analyses provided. This study will help support future lipidomics experiments not only through the findings described but also in providing a dataset with very high replicate counts that can be mined for further insights.

**Chapter IV - Non-model fungal lipidomics reveals lipo-chitooglucosaccharides selectively  
preserve native lipidomes**

### **Citation:**

The figures and text of this chapter are derived primarily from a draft manuscript of which I am the primary author: “Non-model fungal lipidomics reveals lipo-chitoogliosaccarides selectively preserve native lipidomes.” Matthew J. Keller, Tomás A. Rush, and Robert L. Hettich. 2024, manuscript in preparation. Tomás Rush and Robert Hettich are co-corresponding authors. My contributions to this project include performing the lipidomics experiments, analyzing the lipidomics data, drafting and editing of the entire manuscript.

### **Abstract:**

Little is known about the lipid species present in filamentous fungi like *Aspergillus fumigatus* and *Laccaria bicolor*. While frequently applied to yeasts, lipidomics has not seen wide use among non-model fungi, or filamentous fungi generally. This limits fungal biologists in the hypotheses that can be generated regarding the cell membrane and lipidome as expected lipid compositions have not yet been defined. Lipo-chitoogliosaccarides are illustrative of this as they are known to effect fungal morphology and signaling, but any impact they have on the lipidome is unknown. Here we characterize the lipidomes of *A. fumigatus* at two different temperatures and *L. bicolor* providing a rigorous definition of the default lipid profiles of these organisms. We additionally treat *A. fumigatus* and *L. bicolor* with synthetic lipo-chitoogliosaccarides and chitoogliosaccarides and measure the resulting changes in the lipidome profiles. Surprisingly, we find that although (lipo-)chitoogliosaccarides have a major effect on fungal physiology, the lipid profiles of these organisms are largely unchanged from their native state. This implies a selective preservation of the native lipidome by (lipo-)chitoogliosaccarides while at the same time remodeling the metabolome, transcriptome, proteome, and morphology of these fungi.

### **Introduction:**

#### *Aspergillus and Laccaria lipidomes*

*Aspergillus fumigatus* is an opportunistic human pathogen and an important member of soil microbial communities.<sup>105</sup> *Laccaria bicolor* is also a soilborne fungus, and has been shown to have a large effect on the growth of poplar trees which are of interest for biofuel production.<sup>106</sup> Interestingly, *L. bicolor* is also nematophagous and can form edible mushrooms.<sup>107</sup> While these organisms have been used as models for understanding pathogenicity in the infection of susceptible human lungs (*A. fumigatus*)<sup>108</sup> or their roles as a symbiont with forest trees (*L. bicolor*)<sup>106</sup>, they are relatively understudied and little is known about the lipid profiles of these

fungi. While previous lipidomics experiments have been performed on *Aspergillus fumigatus*, and the related fungus *Aspergillus flavus*, these prior studies are limited in the robustness of the approaches used to characterize the lipid species. Two such studies employed liquid chromatography-mass spectrometry, but with full-scan mass spectrometry only, no tandem mass spectrometry was employed.<sup>109, 110</sup> Exact mass matching from full-scans is unable to provide accurate identifications in lipidomics or metabolomics and therefore these studies are not of value in establishing the expected lipid profile of *Aspergillus* spp.<sup>20, 63, 66, 70, 71, 111</sup> In any case, these studies report very little of the lipidomics results and do not make their data available. One recent study involved the MALDI TOF/TOF analysis of *Aspergillus fumigatus* clinical isolates.<sup>112</sup> This study did employ tandem mass spectrometry to confirm lipid identifications, however the focus was exclusively on glycosylinositolphosphoceramides for strain characterization and by its nature MALDI TOF/TOF is not a comprehensive measurement platform. Thus, while useful, this study also does not provide a description of the native lipidome of *Aspergillus* spp. The final two prior publications of relevance both used a similar approach to study the lipidome of *Aspergillus flavus*.<sup>11, 12</sup> These utilized LC-MS untargeted lipidomics analysis using the Orbitrap Fusion MS platform. Both generated tandem MS data which was used for lipid identification with the LipidSearch software from Thermo Scientific. While the use of tandem MS with automated lipid identification is theoretically sufficient for lipid identification at the species level, there is a growing consensus in the lipidomics community that the use of current automated analysis tools without additional validation has insufficient accuracy for the confident analysis of lipids.<sup>20, 63, 64</sup> Indeed, as the Chapter II established, the poor accuracy of automated lipid identification from tandem MS data means about half of reported identifications are false positives. Additionally, neither of these papers make the raw lipidomics data available, preventing the validation of their reported results. Therefore, a fundamental characterization of the native lipidome of *Aspergillus fumigatus* (and *Aspergillus* spp. more generally) is still needed. Additionally, to the best of our knowledge, no lipidomics studies on any organisms of the *Laccaria* genus have been performed and therefore a profiling of its native lipidome is also needed.

#### *Mycology of lipo-chitooligosaccharides*

Lipo-chitooligosaccharides (LCOs) and chitooligosaccharides (COs) are amphiphilic, short-chain, chitin-derived compounds that are likely widespread microbial signals. These

molecules are produced by rhizobia bacteria<sup>113</sup>, most fungi<sup>114</sup>, and potentially nematodes.<sup>115</sup> LCOs and COs have been demonstrated to play key roles in various biological processes, including initiating organogenesis in legumes<sup>116</sup>, facilitating colonization by mycorrhizal fungi<sup>117, 118</sup>, acting as a defense mechanism against nematode colonization<sup>119</sup> (Chen et al., 2024), and functioning as fungistatic agents that modulate metabolomic, proteomic, and transcriptomic profiles in fungi.<sup>105, 114, 120, 121</sup> To further understand how LCOs and COs modulate fungal physiology, we are investigating changes in the lipidomes of *Aspergillus fumigatus* and *Laccaria bicolor* when exposed to these exogenous compounds. The lipidome is critical for maintaining membrane integrity and fluidity, facilitating signal transduction, enabling environmental adaptation, and potentially mediating symbiotic interactions. Our previous research has shown that LCO and CO treatments result in reduced fungal branching and growth, which correlates with shifts in various omics profiles. Based on the previous results observed in gene and protein regulations, and morphological changes, there is a strong expectation that the treatment of LCOs will result in changes to the lipidome of the exposed fungi.

In this study our aim is twofold. First, we characterize the lipidomes of *A. fumigatus* (at two temperatures) and *L. bicolor* to provide an understanding of the native lipid composition of these fungi under fairly standard conditions. This will contribute to our basic knowledge of lipid biology and serve as a basis for future work in these fungi. Secondly, we measure the effects of LCO and CO treatment on the lipidomes of these fungi. This will close a major knowledge gap in the physiological activity of (L)COs in fungi and support future molecular biology work in these fungi. Together this work provides insight on how fungal lipidome might change under different temperatures or exposed to microbial signal molecules, which has an influence on fungal growth, stress response, pathogenicity, and interactions with their hosts. This lipid analysis is the first high confidence profiling of lipid species on a lipidome-scale in these genera, and the effects of LCO treatment are found to be surprisingly contrary to our hypothesis.

## **Methods:**

### *Fungal Growth*

*Laccaria bicolor* strain S238N is a beneficial ectomycorrhizal fungus and a model organism for understanding nutrient exchange and genetic mechanisms involved in symbiosis with poplar (Martin et al., 2008).<sup>106</sup> *Aspergillus fumigatus* strain Afu293 is an opportunistic human pathogen and part of the causal agent that causes aspergillosis in the lungs (Latgé

1999).<sup>122</sup> This fungus is also a soilborne fungus that can regulate microbial communities (Rush et al., 2022)<sup>105</sup> and a genetic model for understanding secondary metabolite production (Brookman and Denning, 2000; Keller et al., 2019).<sup>123, 124</sup> This is the first lipidomics dataset for the *Laccaria* genus. First, we prepare fungal inoculum for the experiment. For the *L. bicolor* samples, we grew the fungus on potato dextrose agar for 21 days at 25°C in the dark.

*Laccaria bicolor* strain S238N and *Aspergillus fumigatus* strain Afu293 were used in this study. To prepare the fungi, *L. bicolor* was grown on potato dextrose agar for 21 days at 25°C in the dark, while *A. fumigatus* spores were grown on glucose minimal media without 1% thiamine<sup>125</sup> for four days at 37 °C in the dark. To collect starting inocula, a sterile cork borer was used to cut three fungal plugs from the edge of fungal growth with a diameter of 0.5 cm<sup>2</sup> for *L. bicolor*. We used 10 ml of sterile MilliQ® IQ 7000 water system and a sterile L-Shaped rod to dislodge and collect spores from the *A. fumigatus* samples and filter those spores through a 0.45µm cell strainer. Five biological replications were used per species/condition, and the fungi were axenically grown in 50 ml of potato dextrose broth (BD Difco, product number BD 254920) in 250 ml closed capped Erlenmeyer glass flasks at 25 °C at 250 rpm in a New Brunswick Excella 25 incubator shaker. We added three fungal plugs of *L. bicolor* to the 50 PDB flasks and 10<sup>6</sup> spores/ml of *A. fumigatus* to the 50 PDB flasks. The flasks were inoculated with one species of fungi, they were not mixed. *L. bicolor* and *A. fumigatus* were grown at 25 °C which represents a “field” temperature and their presence in the soil. *A. fumigatus* was also grown at 37 °C which has more relevance for its pathogenic life cycle. After 21-day incubations, fungal biomass was separated from the fungal exudate with 0.45-µm Durapore polyvinylidene fluoride membrane filter (Millipore Sigma, item number SE1M003M00). The biomass was collected off the membrane filter with a sterile spatula and placed in 15 ml Falcon tube, placed in –80 °C overnight, and then lyophilized until dried. Then, 2g of each biological replication fungal biomass was added to tubes with beads which was used to extract the lipids. Screw-cap 2 mL bead beating tubes were used which were prefilled with acid washed 800 µm and 1.4 mm zircona beads (Cole-Parmer: 2303-MM2).

#### *Lipid extraction and measurement by LC-MS*

Lipids were extracted using a Folch-based two-stage biphasic extraction.<sup>60</sup> For the *A. fumigatus* samples, 350 µL of chloroform, 175 µL of methanol and 6 µL of the EquiSplash internal standard mixture was added to each tube. This was followed by bead beating at 1500

beats per minute for 5 mins in pre-chilled ( $-20\text{ }^{\circ}\text{C}$ ) aluminum blocks. Following this,  $175\text{ }\mu\text{L}$  of water (containing  $0.2\%$  w/v NaCl) was added to each sample before vortexing. Samples were centrifuged and the lower, organic fraction was collected. An additional  $400\text{ }\mu\text{L}$  of fresh lower, organic phase solvent was added to the samples, vortexed, and shaken on ice for 40 mins. Samples were then centrifuged and the organic phase was collected and mixed with the first lipid collection. These were vacuum-centrifuged to dryness and redissolved in  $300\text{ }\mu\text{L}$  of 1:1:1 1-butanol, isopropanol, and water. A  $30\text{ }\mu\text{L}$  aliquot was taken from each sample and mixed to form a pooled QC sample. The *L. bicolor* samples were freeze dried after sample collection. These were transferred to bead-beating tubes and frozen again at  $-80\text{ }^{\circ}\text{C}$ . Before the extraction, *L. bicolor* was cryoground for 5 mins at 1500 bpm (cryogenically frozen with liquid nitrogen). Following this, chloroform and methanol were added and the extraction performed the same as with *A. fumigatus* with the following exceptions. First, the bead-beating under solvent was reduced from 5 mins to 3 mins. Second, a third lipid extract round was performed with an additional  $400\text{ }\mu\text{L}$  of lower, organic phase.

Lipid extracts were analyzed by nano liquid chromatography-tandem mass spectrometry using a Dionex UltiMate 3000 coupled to a ThermoFisher LTQ-Orbitrap Velos Pro. The LC was operated with a split-flow nano chromatography setup where a three-way T-junction connects the pump, a waste line, and a line to the nano-flow column. Although the flow rate from the pump was set to  $50\text{ }\mu\text{L}/\text{min}$ , much of this is diverted to waste at the T-junction. Backpressure produced from the waste-line likely provides flow through the column on the order of hundreds of  $\text{nL}/\text{min}$ . Nano chromatography columns ( $100\text{ }\mu\text{m}$  inner diameter) were packed in a pressure cell with C18 reversed-phase packing material from Phenomenex ( $1.7\text{ }\mu\text{m}$  C18 Kinetex with  $100\text{ }\text{\AA}$ ). A multistep gradient was employed where mobile phase A consisted of 60:40 acetonitrile:water with  $0.4\text{ g/L}$  ammonium acetate as an ionization additive and mobile phase B consisted of 90:10 isopropanol:acetonitrile with  $0.4\text{ g/L}$  ammonium acetate as an ionization additive. The chromatography method is listed in **Table 4-1**. The mass spectrometer was operated in a top-10 data dependent acquisition (DDA) mode. Full scans were performed at 30k resolution from 150-2000  $m/z$  in the Orbitrap mass analyzer. These were followed by ten DDA MS2 scans in the ion trap with a  $2\text{ }m/z$  isolation width and collisional-induced dissociation at a normalized collision energy of 33. Dynamic exclusion was performed on a single repeat count with a 40 second exclusion duration, and an expiration s/n of 4.

## Data Analysis

The lipidomics results of each dataset were analyzed with MS-DIAL<sup>48, 126</sup> followed by some manual validation as well as automated identification filtering with LipoCLEAN. MS-DIAL was run with parameters previously shown to be compatible with LipoCLEAN. Around five hundred of the *L. bicolor* lipid identifications reported by MS-DIAL were manually validated. Just over one thousand lipid identifications were manually validated in the *A. fumigatus* at 37 °C dataset. These manual annotations are useful to unambiguously confirm the presence of covered lipid species but were also used along with other datasets to build the LipoCLEAN model as discussed in Chapter II. Despite annotating several hundred lipid identifications manually, there still remained thousands unvalidated and so LipoCLEAN was employed to filter the rest of the MS-DIAL results. LipoCLEAN is a recently developed machine-learning based filter that rescores MS-DIAL identifications as likely true positives, likely false positives, and a third uncertain category. The probable true positives reported by LipoCLEAN were combined with the manually validated lipids to form the set of lipid identifications analyzed in this study. This forms the most reliable lipidomics data that can be generated by untargeted procedures. LipoCLEAN is expected to have an approximately 7% false discovery rate, as opposed to the 47% MS-DIAL natively generates. Lipid identifications were imported to EI-MAVEN<sup>94</sup> and manual integration of the extracted ion chromatograms was performed. A minority of LipoCLEAN-predicted true positives were excluded from further analysis at this stage when examination of the raw data during peak integration revealed the identifications were incorrect. Lipid peak areas were then processed in python with the following steps. First, blank filtering was performed where the condition with the highest average intensity had to be at least three times the blank average and could contain no more than two zero values. Following blank filtering, a LOWESS drift correction was performed on each lipid species using the pooled QC samples. After correction, peak areas corresponding to the same lipid species were summed together (for example, certain lipid species were measured with more than one adduct state). The data were then normalized. For the descriptive data shown in most of the figures, only sample mean-centering was performed. For data used in statistical testing, lipid class abundances were also mean-centered, with all the replicates pooled together, before sample mean centering was performed (this was done to normalize out lipid class differences, such as those resulting from ionization efficiency differences).

Summary statistics of the lipidomes were calculated from these normalized values to provide a more interpretable readout of lipid composition. Four summaries were calculated, average fatty acid length, average fatty acid unsaturation, glycerophospholipid to acylglycerol lipid ratio, and anionic to cationic (or zwitterionic) ratio. ANOVA was performed on each of these summary statistics as well as a PerMANOVA on the whole dataset (using cosine distance as the distance matrix) to provide a comprehensive view of potential lipidome alterations.

## Results:

### *Summary of Lipidomics Results Across Datasets*

After combining the results of the manual validation of lipid identifications with the predicted high confidence lipids as reported by LipoCLEAN, between 344 and 379 individual lipids species were identified in the three datasets. These represent lipid species as defined by LIPID MAPS and are reported based on the summed carbon number and degree of unsaturation of the fatty acid tails.<sup>76</sup> Thus, the number of lipid molecular species is likely substantially higher than 379 (and represent sub-forms of the detected species). By far, the triacylglycerol (TG) class had the most unique identifications with more than one-third of the total lipid species identified (**Table 4-2**). TGs are neutral lipids with three fatty acids connected to a central glycerol moiety. These are commonly understood to be energy storage lipids and are the majority class of lipid-droplets in fungi. Phosphatidylethanolamine (PE, a type of zwitterionic glycerophospholipid) lipids were the next most common class across all three datasets. Following PE, the ceramide (Cer), cardiolipid (CL), diacylglycerol (DG), phosphatidylcholine (PC), phosphatidylinositol (PI), and phosphatidylserine (PS) classes are all sizable components of the lipidome by count of identifications. Despite being analyzed in different batches, with no data leakage between the different datasets, there is very good agreement in the count of identifications per lipid class. This indicates our method is reasonably repeatable and that filamentous fungi may share similar features of their lipidomes. Within lipid classes, there were not many classes with substantively different average fatty acid lengths or degrees of unsaturation (**Table 4-3**). The ceramide (Cer) class was a notable exception, with non-negligible differences between datasets in both average unsaturation and average fatty acid length. CL lipids, while showing good agreement in fatty acid length, did had substantive differences in average unsaturation (**Figure 4-1 A and B**). To take a closer look at the changes observed in Cer profiles between *A. fumigatus* at 37 °C and *A. fumigatus* at 25 °C, shared Cer species were mean-centered and their relative fold changes

calculated (**Figure 4-2**). This showed a noticeable trend of longer ceramides at 37 °C and shorter/unsaturated ceramides at 25 °C. The relative abundance of lipid classes were also plotted, which revealed more obvious differences between datasets (**Figure 4-3**).

#### *Testing (L)CO-Modulation of Lipidomes*

In order to test for an effect of LCO or CO signaling on the lipidome of the tested fungi, ANOVAs were performed on lipid summary statistics as well as a PerMANOVA of the entire dataset. The results of these tests are summarized in **Table 4-4** and **Table 4-5**. The summary statistic ANOVAs generated p-values that are in good agreement with the null hypothesis (no effect of (L)COs on lipid composition). The lowest p-value obtained was 0.1 when testing average unsaturation in the *A. fumigatus* 37 °C dataset. These results clearly provide no evidence contrary to the null hypothesis. Performing PerMANOVA on the three datasets similarly generated data consistent with the null hypothesis (**Table 4-5**). P-values were generally high and no more than 46% of the variance was explained in any of the datasets. The *L. bicolor* PerMANOVA generated a p-value which could be considered reasonable evidence against the null (0.02), however after adjusting for multiple comparisons this increases to above 0.1. In any case, less than half of the variance can be explained by this analysis implying any effect present is less than the noise in the experiment. Principle component analyses were performed which also show no separation of the treatment groups and the control (see **Figures 4-4, 4-5, and 4-6**).

### **Discussion:**

#### *Native Lipidome of Aspergillus fumigatus and Laccaria bicolor*

One of the goals of this research was to provide a robust assessment of the native lipid compositions of *Aspergillus fumigatus* and *Laccaria bicolor*. Given there was little to no effect measured from LCO or CO treatment, all samples of a dataset (including pooled quality control samples) were averaged together to characterize the native lipidome. Quite a diversity of lipid species were observed across all three datasets, with 17, 15, and 20 different lipid classes observed for *A. fumigatus* 25 °C, *A. fumigatus* 37 °C, and *L. bicolor* respectively. Each of these datasets had 10 or more lipid classes with at least 5 different species observed. While most of these classes are glycerophospholipids, glycerolipids and sphingolipids were also observed. Interestingly, prior lipidomics work in *Candida albicans* had a similar count of lipid species identified per class (for glycerophospholipids, glycerolipids were not characterized) as the current work.<sup>127</sup> By both count of identified lipid and relative signal percent, PE, PC, PS, and PI

lipids where the major glycerophospholipid classes. Cardiolipin generally had high lipid species counts but contained a smaller proportion of the total signal. Considering all lipid classes, PE lipids were the most abundant for *A. fumigatus* 37 °C and *L. bicolor*, while TG was the most abundant by far for *A. fumigatus* 25 °C.

#### *Effect of Temperature on A. fumigatus Lipids*

While not the focus of this study, it is reasonable to compare the observed lipid compositions of *A. fumigatus* between 25 °C and 37 °C. There appear to be significant differences in the observed lipid class abundances, especially with TG lipids. This could reflect an increase in lipid droplet formation in *A. fumigatus* formation at lower temperatures. Possibly the environmental stress of suboptimal growth temperatures induces *A. fumigatus* to allocate more of its available energy to storage in the form of TG lipids rather than cell growth. Within a lipid class, however, there are fewer observable changes in the fatty acid compositions of the lipids. Lipid classes CL, DG, PC, PE, PI, PS, and TG all have nearly identical fatty acid chain lengths between the two temperatures. Only the ceramide class shows a substantive difference in chain lengths (Cer chain length is reported as the average of the fatty acid and sphingosine moieties), with an increase at 37 °C. Fatty acid unsaturation does have higher variability, but at the same time more distinct differences between temperature. Ceramide and PC lipids show a decrease in unsaturation with respect to temperature, while CL shows a large increase in unsaturation and TG a more modest increase. To more closely examine changes in ceramide fatty acid properties, the relative abundances of ceramide shared between the 25 °C and 37 °C datasets were plotted as a function of fatty acid length (**Figure 4-2**). This confirmed the presence of changes in fatty acid length and unsaturation. Overall, it is at first somewhat surprising there is not a larger effect from growth temperature on fatty acid characteristics. Cer and PC lipids show fatty acid changes in a manner that would be expected for homeoviscous adaptation; however, this is not an overall trend in the lipidome. This may be due in part to the presence of ergosterols which act to modulate membrane fluidity without requiring fatty acid changes. Unfortunately, ergosterols do not ionize readily in electrospray ionization and no ergosterols were identified. That being the case, ceramide and hexacylceramide class abundances do increase at 25 °C relative to 37 °C. These classes (especially HexCer) are thought to form lipid raft domains with ergosterols in fungi and so this increase could be indicative of homeoviscous changes by means of ergosterols.<sup>10, 14, 128</sup> In fact, having no bulk change in fatty acid unsaturation would be

consistent with previous work in *Aspergillus niger* that found no change in fatty acid unsaturation between *A. niger* grown at 26 °C and 35 °C (through a gas chromatography fatty acid methyl ester analysis).<sup>129</sup> An important caveat to these findings is that this lipidomic analysis cannot guarantee full comprehensive coverage of all lipids present in the sample. Indeed, “uncertain” lipids as reported by LipoCLEAN were excluded to trade recall in exchange for accuracy. LipoCLEAN’s “high confidence” lipids are expected to include approximately 80% of all true lipid species in the data while only allowing ~7% false discoveries. Additionally these datasets were processed, measured, and analyzed separately giving rise to the possibility of non-negligible batch effects between these datasets. Because of this, observed changes in lipidomes between datasets should not be considered as definitive but instead used as hypothesis generating data.

#### *Examining the Effect of (L)COs on Lipid Composition*

In order to test what effect, if any, treatment with LCOs or COs has on the fungal lipidomes, two complementary approaches were used. First, summary statistics of lipid composition were calculated, and each was tested in an ANOVA. These summary statistics seek to provide a concise, biologically informed method of measuring lipid remodeling. Four statistics were used, average carbon per fatty acid, average unsaturation per fatty acid, glycerophospholipid to glycerolipid ratio, and cationic/zwitterionic to anionic glycerophospholipid ratio. The average carbon and average unsaturation, as discussed above, relate to biometrics such as membrane fluidity. The glycerophospholipid to glycerolipid ratio related to the ratio of membrane lipids to energy storage lipids. The cationic/zwitterionic to anionic glycerophospholipid ratio has implications for membrane polarity and (possibly) curvature. We recommend other lipidomics practitioners implement these or similar summary statistics rather than hypothesis testing on each individual lipid quantified. Doing so will increase the interpretability of results, lower the number of multiple comparisons to be corrected for, and help avoid introducing researcher degrees of freedom.<sup>130</sup> While the LCOs and COs tested have some degree of hierarchical relationships, there is no coherent multifactorial design that can describe the relationship between the different LCO and CO treatments. Because of this, ANOVAs were performed on the summary statistics for individual datasets with each condition specified as its own experimental factor. All these tests generated results in good agreement with the null hypothesis that all group means are equal. To provide a more comprehensive view of the

lipidome at the cost of decreased interpretability, PerMANOVA (permutational multivariate ANOVA) was performed on the entirety of lipid quantifications of each dataset. The results generated by PerMANOVA also indicate there is a negligible effect, if any, of LCO or CO treatment on the lipidome. It is worth noting that these tests were decided upon after a very preliminary investigation on a subset of the data indicated there would likely be small effect sizes from LCO and CO treatment. We then settled on a predefined analysis so as to avoid bias from researcher degrees of freedom, sometimes referred to as the “garden of forking paths”.<sup>130</sup> Naturally, the power of these tests are limited by the variability of the measurements themselves. The average relative standard deviation (RSD) for all lipids in the *A. fumigatus* datasets were both in the fifty percent range. This is to some degree driven by outliers as the median RSD for both datasets were about 43%. That said, *L. bicolor* had quite tight RSDs across the dataset with an average of only 28%. These analyses do indicate then that our initial expectations of a lipid-modulatory role of LCOs and/or COs was incorrect. This is very surprising since previous analyses have shown LCOs to affect many aspects of physiology in these fungi. Morphological, transcriptomic, and proteomics alterations have all been reported for these fungi before. If such large effects are occurring, would be natural to expect concomitant changes to occur in the lipidome, especially in the case of morphological changes. It appears that the physiological roles of LCOs and COs are achieved while persevering the native lipidomes of these fungi. Several potential explanations could account for this surprising lack of lipidomic change. While unlikely, it is possible that other molecular pathways were responsible for the previously observed effects. Conversely, after 21 days of growth, the fungi may have metabolized or broken down the exogenous LCOs and COs whose effects are subsequently reversed or diluted beyond detection. Since previous experiments were conducted on agar or were otherwise spatially constrained, the liquid broth cultivation may prevent a response by the fungi to LCOs and COs either simply due to the different growth environment, or perhaps because (L)COs are adsorbed to the flask walls and are sequestered from the fungi. It is also possible that the subtle differences in the culture and growth of *A. fumigatus* and *L. bicolor* from the original publication describing their regulation by LCOs or COs (potato dextrose broth vs glucose minimal media) could make these fungi less sensitive to the presence of LCOs and COs. This might mean that in other growth conditions treatment with (L)COs could still induce a lipidome remodeling. However, at the very least, it is clear that LCO and CO treatment does not have a strong effect on the overall lipid

composition of these fungi, although it is still possible that any unidentified lipids may change in response to LCO or CO treatment.

### *Impact and Concluding Thoughts*

We consider the impact of this work to be primarily two-fold. First, this provides a detailed and robust analysis of the “default” lipidomics profile of an agriculturally and clinically relevant fungi with a dearth of prior knowledge about their lipids. This serves as an important reference for fungal physiology and expands our knowledge about this understudied kingdom. These data can be used for hypothesis generation and the raw data will be made publicly available on the Metabolomics Workbench repository. Secondly, we provide a characterization of the lipid regulatory effects of LCOs and COs and find there to be no noticeable regulation. This result is quite surprising since LCOs and COs have been found to affect many other aspects of fungal biology. Rather than *A. fumigatus* and *L. bicolor* simply not responding to (L)COs, they instead appear to remodel their physiology while selectively preserving the native lipidome. This in turn implies the native lipidome may be well tuned for fungal health under a variety of environments and conditions and require little modulation in response to various stimuli.

**Chapter V - Deuterated Lipids: A Comprehensive Examination of Lipidome-Wide  
Deuteration and the Principals of Designing Regiospecifically Deuterated Phospholipids**

This project originated in the O'Neill lab at ORNL's neutron scattering division. They were attempting to produce deuterium labeled lipids from *E. coli* for use in neutron contrast matching experiments. Their efforts were hampered by the lack of an established method for profiling the biosynthetic deuteration of phospholipids *in vivo*. In collaboration with other neutron scattering scientists at ORNL we developed methods for the quantitative analysis of *in vivo* lipid deuteration and used these to rigorously profile the incorporation of deuterium into phospholipids in *E. coli*. Based on this, we established guidelines for the rational design of regiospecifically deuterated bacterial phospholipids. We also expanded this to include the deuteration of non-native lipids produced transgenically.

The study of isotopologues and isotope labeling naturally lends itself to investigation via mass spectrometry. Different isotopes of the same element have the same general chemical properties, while varying in mass. This mass difference is easily measurable with modern MS instruments for all but large macromolecules. The generation of isotope labeled biomolecules is very important to studies utilizing neutron scattering and reflectometry, as it allows for the use of contrast matching. Contrast matching takes advantage of the fact that protium and deuterium have quite different neutron scattering densities. Selective deuteration of biomolecules can then therefore be used to make those components have a scattering density equal to that of the background, essentially making them "transparent" to neutrons. However, there were major limitations for the generation of contrast-matched lipids, which are critical for membrane-related research. In collaboration with researchers at Oak Ridge National Laboratory, we sought to address this experimental deficiency by characterizing methods for the *in-vivo* deuteration of phospholipids. Through extensive lipidomics measurements and computational modeling, I thoroughly characterized the *in vivo* deuteration of *E. coli* lipids from a variety of deuterated feedstocks. I introduced the use of a computationally derived one-number summary of deuterium labeling distributions and developed guidelines for the design of specifically deuterated phospholipids. These guidelines allow for the control of both the amount and the location of deuteration. This enables large-scale synthesis of selectively deuterated phospholipids with a native composition of lipid species. This project resulted in the publication of "Characterization of the *in vivo* deuteration of native phospholipids by mass spectrometry yields guidelines for their regiospecific customization" in *Analytical Chemistry*. This manuscript forms the first portion of this chapter. Follow-up experiments have extended the application of *in vivo* lipid

deuteration to transgenic *E. coli* that was engineered to produce non-native lipids. This allows for a better mimic of mammalian lipid profiles. While useful for basic research, such steps are also critical for research focused on human health and physiology. This work is in preparation and forms the second part of this chapter. The full impact of this work is not restricted to a specific field, but rather will be realized in the various applications and discoveries we have enabled through this work.

**Part 1. Characterization of the *in vivo* deuteration of native phospholipids by mass spectrometry yields guidelines for their regiospecific customization:**

**Citation:**

Text and figures in this section are reprinted (with some adaptation) with permission from “Characterization of the *in vivo* deuteration of native phospholipids by mass spectrometry yields guidelines for their regiospecific customization.” Matthew J. Keller, Qiu Zhang, Shuo Qian, Brian C. Sanders, Hugh M. O’Neill, and Robert L. Hettich. *Analytical Chemistry* 2024 96 (1), 212-219; DOI: 10.1021/acs.analchem.3c03750. Copyright 2023 American Chemical Society. My contribution to this work includes designing experiments, performing all LC-MS measurements, data analysis and interpretation, figure generation, drafting and editing of the manuscript.

**Abstract:**

Customization of deuterated biomolecules is vital for many advanced biological experiments, including neutron scattering. However, because it is challenging to control the proportion and regiospecificity of deuterium incorporation in live systems, often only two or three synthetic lipids are mixed together to form simplistic model membranes. This limits the applicability and biological accuracy of the results generated with these synthetic membranes. Despite some limited prior examination of deuterating *E. coli* lipids *in vivo*, this approach has not been widely implemented. Here, an extensive mass spectrometry-based profiling of *E. coli* phospholipid deuteration states with several different growth media was performed and a computational method to describe deuterium distributions with a one-number summary is introduced. The deuteration states of thirty-six lipid species were quantitatively profiled in fifteen different growth conditions and tandem mass spectrometry was used to reveal deuterium localization. Regressions were employed to enable the prediction of lipid deuteration for untested conditions. Small-angle neutron scattering was performed on select deuterated lipid samples, which validated the deuteration states calculated from the mass spectral data. Based on these

experiments, guidelines for the design of specifically deuterated phospholipids are described. This unlocks even greater capabilities from neutron-based techniques, enabling experiments that were formerly impossible.

### **Introduction:**

Stable isotope labeling is widely used for studies of biological processes and is applied to a variety of biomolecules including proteins, lipids, nucleic acids, and metabolites. For studies of lipids and membranes, incorporation of specific labels such as deuterium has proven to be very valuable in a range of applications, including Raman imaging, nuclear magnetic resonance (NMR), and neutron scattering.<sup>131-134</sup> Combining deuterium-labeled and natural isotope fatty acids in Raman microscopy has enabled simultaneous “two-color” imaging of lipids in human cell lines, and deuterium labeling of *de-novo* lipids has allowed Raman imaging of tumor boundaries.<sup>131</sup> Deuterium labeling of lipids has also enabled deuterium-NMR to probe the structural interactions between proteins and lipids in murine neural tissue as well as *E. coli* membranes.<sup>132, 133</sup> Deuterium labeled lipids are also important as standards for qualitative and quantitative analysis in lipidomics experiments.<sup>135, 136</sup>

Neutron scattering is an important technique for biochemistry, molecular biology, and biophysics investigations. The intrinsic contrast of isotopes, especially in those of hydrogen, provide a unique distinction in neutron scattering that is different than X-ray-based techniques. By changing the amount of deuterium in the solvent or in a molecule of interest, components of the sample can be selectively highlighted in a complex mixture, an approach termed “contrast variation”.<sup>137</sup> This technique provides important structural and dynamics understanding in the fundamental properties of biomembranes, membrane proteins, and lipid nanoparticles for biotechnology applications.<sup>9, 138-140</sup> In all these cases, the effective use of contrast variation requires lipids with defined deuteration states. Although some deuterated synthetic phospholipids are commercially available, they are very limited in their composition and deuterium incorporation. Generally, only two or three synthetic lipids are mixed to form model lipid membranes, which do not approach the actual complexity of even “simple” bacterial membranes. There have been recent advances in the chemical synthesis of deuterated phospholipids, but doing so on a biomimetic lipidome-wide scale has yet to be achieved.<sup>141</sup> This severely limits the relevance of these model membranes and their use in downstream applications and eliminates them entirely where *in-situ* measurements are required. There is a growing recognition in the

community that natural lipid membranes provide a more native environment for membrane proteins and can greatly impact their stability and activity when purified.<sup>142</sup> While production of deuterated biomolecules such as proteins and nucleic acids in bacteria and yeast are well established, less is known about how deuterium is incorporated into the cellular lipids.

Deuterium incorporation from labeled feedstocks into endogenous bacterial lipids has been assessed and applied to SANS only once before.<sup>9, 143</sup> In that work, the authors suggested that a controlled deuteration of different parts of a phospholipid could be achieved by varying the deuterated media supplements, although this does not appear to have been widely implemented.<sup>143</sup> That study was itself limited with only four conditions tested and four lipids profiled by mass spectrometry (using a low resolution instrument with  $\sim 0.1$   $m/z$  accuracy).<sup>143</sup>

The present work seeks to quantitatively characterize the deuteration present in the *E. coli* lipidome after culturing in various deuterated media, determine the localization of that deuterium, infer the metabolic sourcing of deuterium, and how deuteration can be controlled. While building upon prior work, this study has many more fractionally labeled conditions (14 in total), performs a much deeper lipid analysis (the deuteration states of 36 lipid species having been characterized by high-resolution accurate mass-mass spectrometry), and introduces a computational method of summarizing the deuterium distributions with one number using binomial distributions. From this, we describe guidelines for the design of location-controlled endogenously deuterated lipids that will much better simulate the actual composition and complexity of cellular membrane than a mixture of two to three synthetic standards. Additionally, the analyses and principles employed in this study should help inform other stable isotope labeling approaches of mass spectrometry.

## **Materials and Methods:**

### *Growth of E. coli with Labeled Lipids*

*E. coli* strain BL21(DE3) was used for this study. Initially, 5 mL of Luria-Bertani (LB) culture was inoculated from a glycerol stock and incubated at 37 °C overnight. The cell cultures were then adapted to grow in Enfors minimal media with different H<sub>2</sub>O/D<sub>2</sub>O ratios.<sup>144, 145</sup> A detailed description of the Enfors media used is described below, with the recipe shown in **Table 5-1**. For each adaptation step, 250  $\mu$ L of an overnight (16 h) cell culture was centrifuged at 4000 x g for 10 min and the cell pellet was resuspended into 5 mL of the appropriate fresh media and incubated at 37 °C until the cell density at 600<sub>nm</sub> (OD600<sub>nm</sub>) reached stationary phase (OD600<sub>nm</sub>

= ~5). In total, at least three transfers were done to ensure equilibration at each H<sub>2</sub>O/D<sub>2</sub>O ratio. Following this, a final culture of 30 mL was grown and samples were collected by centrifugation for LC-MS analysis.

**Table 5-2** lists the conditions tested in this study along with their proportions of labeled media. *E. coli* was grown with various combinations of deuterated water and deuterated glycerol or glucose. While most samples were analyzed in biological triplicate, a few of them (originally analyzed as a pilot study) only have one replicate. However, the differences between biological replicates were found to be negligible among triplicates (of 264 lipid-condition pairs the median relative standard deviation in the deuteration was 0.37% for biological triplicates while the average was 1.9%), thus singlets were considered to be sufficient for the analysis.

#### *Enfors media preparation*

To prepare the Enfors minimal media in H<sub>2</sub>O, weigh out desired amount of (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, Na<sub>2</sub>HPO<sub>4</sub>•7H<sub>2</sub>O, KH<sub>2</sub> PO<sub>4</sub>, (NH<sub>4</sub>)<sub>2</sub>-H-Citrate and dissolve in H<sub>2</sub>O with stirring. Then dilute media to nearly final volume before adding desired amount of MgSO<sub>4</sub>, trace metals solution, and glycerol. Finally, adjust the volume with H<sub>2</sub>O to final and filter it through 0.22 µm filtration system to make it sterile.

To prepare the Enfors minimal media in D<sub>2</sub>O, weigh out desired amount of (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, Na<sub>2</sub>HPO<sub>4</sub>•7H<sub>2</sub>O, KH<sub>2</sub> PO<sub>4</sub>, (NH<sub>4</sub>)<sub>2</sub>-H-Citrate and dissolve in minimal amount of D<sub>2</sub>O with stirring. Evaporate salts solution to dryness by using rotary evaporation. Dissolve salts in minimal amount of D<sub>2</sub>O again and repeat evaporation. Dilute to nearly final volume before adding desired amount of deuterated MgSO<sub>4</sub> (MgSO<sub>4</sub> salt was dissolved in D<sub>2</sub>O), deuterated trace metals solution, and deuterated glycerol. Finally, adjust the volume with D<sub>2</sub>O to final and filter it through 0.22 µm filtration system to make it sterile.

To prepare fractional H<sub>2</sub>O/D<sub>2</sub>O Enfors minimal media, mix hydrogenated Enfors minimal media with deuterated Enfors minimal media to the desired H<sub>2</sub>O/D<sub>2</sub>O ratio.

#### *Lipid Extraction Procedure*

For the LC-MS lipidomics, all lipid samples were extracted with modified Matyash extraction procedures.<sup>62</sup> Samples H-Gly/0, 25, 50 and 75, as well as H-Glu/100 and D-Glu/100 were extracted by adding methanol and MTBE before cellular lysis via sonication. With a probe ultrasonicator, cellular disruption was performed with 5 min of total sonication, pulsing 10 seconds of sonication followed by 10 seconds of rest. Methyl *tert*-butyl ether (MTBE) was then

added before letting lipids partition while shaking on ice for ~1 hour. The samples were then centrifuged and the upper (organic) layer was collected for analysis. The extract was then concentrated by evaporation under a nitrogen stream to about one-quarter of the original volume.

For the other lipid samples, (MTBE), methanol and water were mixed together with at the ratio of 10:3:2.5, and the resulting upper (organic-rich) phase was collected and set aside for later use. The *E. coli* pellets were disrupted via bead-beating using Zr beads (either 200  $\mu\text{m}$  beads or half-and-half 150  $\mu\text{m}$  with 200-400  $\mu\text{m}$  beads) with 150  $\mu\text{L}$  of methanol and 500  $\mu\text{L}$  of MTBE. Following this, 120  $\mu\text{L}$  of water was added to each tube before shaking on ice for 26 min. The extracts were centrifuged to facilitate phase separation and the upper phase was collected, leaving the aqueous (lower) phase and insoluble pellet, to which was added 400  $\mu\text{L}$  of fresh upper phase. This was vortexed, shaken for ~22 min and centrifuged. The resulting upper phase was mixed with the first extractive. Samples were dried under vacuum and resuspended in the “8Bu” solvent mixture (69:23:8  $\text{H}_2\text{O}$ :IPA:BuOH).<sup>58</sup>

#### *Lipidomic LC-MS Analysis*

Samples were run in positive and negative ion modes on a Dionex Ultimate-3000 reversed phase LC platform coupled online to a ThermoFisher LTQ-Orbitrap Velos Pro mass spectrometer. The method was modified from that developed by Dann-Rasche and employed a 60% ACN 40% water mobile phase A with a 90% IPA 10% ACN mobile phase B, both containing ~4.2 g/L ammonium acetate with gradient details in **Table 5-3** (to facilitate dissolution of ammonium acetate, 1% water may be included in mobile phase B).<sup>58</sup> Columns with an inner diameter of 100  $\mu\text{m}$  were used and packed in-house with Kinetex 1.7  $\mu\text{m}$  C18 (100 Å) packing material (from Phenomenex) to a length between 12-15 cm. A top 10 data-dependent acquisition method was used, where one full scan was acquired at 30,000 mass resolution with the orbitrap followed by 10 data-dependent collisional-induced dissociation (CID) scans in the ion trap at a normalized collision energy of 30. Data are available at the Metabolomics Workbench repository under study ID ST002890.<sup>52</sup>

#### *LC-MS Data Analysis*

Due to the higher complexity and unknown isotopic states of the labeled conditions, lipid identifications were performed on the unlabeled sample (H-Gly/0). Manual analysis was restricted to phosphatidylethanolamine (PE) and phosphatidylglycerol (PG) lipids with between 20 and 40 total carbons and 0-2 degrees of unsaturation, which represent the range of expected

PE and PG lipids in *E. coli*.<sup>146,147,23, 148-150</sup> This analysis included an examination of chromatographic peak shapes and fragment ions (ketene or carboxyl fatty acid neutral losses, fatty acid fragment ions, and carboxyl fatty acid/glycerol neutral losses).<sup>151</sup> Both PE and PG lipids were identified from the negative ion mode data as  $[M-H]^-$  ions, which give richly informative fragmentation spectra. An example is shown in **Figure 5-1** where all expected fragment ions are observed for PG 32:0. After tentative identification of lipids by exact mass and fragmentation pattern, lipid identifications were confirmed by considering their relative retention time (see **Figure 5-2**). Lipids of the same subtype exhibit a linear  $m/z$  vs retention time relationship which has been suggested previously to aid in lipid identification.<sup>20,77</sup> Although fatty acid (FA) fragment ions were observed, lipids were annotated following the LIPID MAPS convention at the species level (reporting the summed tail carbon and unsaturation) since individual chromatographic peaks sometimes contained FA fragments corresponding to different molecular species.<sup>76</sup> Unsaturated odd chain lipids were identified and were labeled according to their total carbons and apparent unsaturation; however, since *E. coli* is biosynthetically unable to produce odd-chain lipids, these are likely cyclopropyl FAs, which have the same mass and CID fragment ions as true odd chain unsaturated lipids.<sup>76, 152</sup>

MS-DIAL was also used for an untargeted analysis of the data, but upon validation of its phosphatidylethanolamine (PE) and phosphatidylglycerol (PG) identifications, it yielded no further accurate identifications.<sup>48</sup> Isotopologues of the identified lipids were manually quantified in the labeled samples through peak integration using El-MAVEN (PG ions were quantified as  $[M-H]^-$  species in negative mode while PE lipids were quantified as  $[M+H]^+$  species in positive mode).<sup>94</sup> Isotopologues were quantified within 10 ppm of the theoretical value; however, this value was refined for particular isotopologues based on chromatographic context to avoid artifacts resulting from near-isobars. For PE lipids, the five lowest abundance lipids in the unlabeled sample were excluded from analysis in the labeled samples, for efficiency as well as detection issues. However, these were only 0.36% of the overall PE abundance in the unlabeled sample, and their exclusion has no meaningful effect on the results or conclusions. The remaining 24 PE species were also found in the labeled samples, except for PE 37:1, which was not detected in the H-Glu/100 condition. For PG lipids, the four lowest abundance lipids in the unlabeled sample (1.5% of the total PG abundance) were excluded from analysis in the labeled samples. The remaining 12 PG species were detected in each labeled sample.

### *Lipid Extraction Procedure for SANS Experiments*

*E. coli* total lipids was extracted using a modified Bligh and Dyer method.<sup>61</sup> *E. coli* BL21(DE3) cell culture (1 L) was grown at 37 °C overnight to OD<sub>600nm</sub> ≈ 5. The cells were harvested by centrifugation at 6000 x g for 30 minutes and then washed with 1% NaCl three times followed by freeze-drying for 48 h. The final yield was ~1.8 g. Chloroform (250 mL), Methanol (500 mL), and water (200 mL) were added sequentially to the freeze dried cells to achieve a final chloroform:methanol:water ratio of 1:2:0.8 (v/v/v). The cell suspension was incubated at room temperature for 18h with stirring. This was followed by addition of 250 mL of chloroform and 250 mL of water to achieve a final chloroform:methanol:water ratio of 1:1:0.9 (v/v/v). The cell suspension/solvent mixture was then placed in a separatory funnel and allowed to phase separate. The lower chloroform rich phase that contains the total lipids was collected, chloroform and methanol were removed using rotary evaporation (Heidolph Hei-VAP Rotary Evaporator). The dry lipid pellet was then resuspended in different ratio H<sub>2</sub>O/D<sub>2</sub>O for SANS experiments.

### *Analysis of Lipid Deuteration*

Since lipids were measured with wide isotopic distributions, a method of summarizing these distributions was needed. To that end, isotopologue distributions were modeled using a binomial function (Equation 1 below) with the assumption that the distributions are generated by stochastic sampling from a pool of protium and deuterium.

$$P(x) = \frac{n!}{x! \times (n - x)!} \times (1 - p_D)^{n-x} \times p_D^x \quad (1)$$

Where  $n$  = total non-solvent exchangeable hydrogen,  $x$  = number of deuterium,  $p_D$  = probability for any non-exchangeable hydrogen to be deuterium, and  $P(x)$  is the probability of generating a lipid with  $x$  deuterium.

This was used to calculate expected lipid deuterium distributions by calculating  $P(x)$  for all values of  $x$  between 0 and  $n$  (inclusive). The probabilities were then adjusted based on the natural isotope distributions for carbon, oxygen, and nitrogen since the fine isotope structures of the lipids were not resolved. For conditions containing labeled water and unlabeled carbon source, the distributions were calculated first assuming  $p_D$  to be the %D<sub>2</sub>O present in the media (forming a “naïve distribution”). A python script was then employed to determine the value of  $p_D$  that generates a curve best matching that of the observed distribution. For this, a least-squares optimizer was used, which minimizes the root-mean-squared error between the observed and

expected distributions (scipy.optimize.least\_squares). This returns the probability of deuterium that best fits the observed data. An example code can be found at [https://github.com/m-j-keller/Lipid\\_Deuteration\\_Calculator](https://github.com/m-j-keller/Lipid_Deuteration_Calculator).

### *Small Angle Neutron Scattering (SANS)*

For SANS measurements, the lipid extracts were suspended in different H<sub>2</sub>O/D<sub>2</sub>O mixtures (expressed as 0, 20, 40, 80 and 100% (v/v) D<sub>2</sub>O) at concentrations of approximately 5 mg/mL. Subsequently, the lipid solutions were transferred into quartz circular cells with a 1 mm beam path (Hellma, Germany) for SANS measurements. The SANS measurements were performed using the Bio-SANS instrument at the High Flux Isotope Reactor, located at Oak Ridge National Laboratory.<sup>153</sup> Bio-SANS was run with a neutron wavelength ( $\lambda$ ) of 6 Å and a wavelength spread of 15%. The samples were kept at 25 °C with a dual-detector setup (sample-to-detector distances of 7 m and 1.13 m) covering a  $q$  range from 0.007 to 0.8 Å<sup>-1</sup>. The acquired data was subsequently processed using the drt-SANS software provided by the facility, resulting in 1D scattering curves (I vs.  $Q$ ) presented in absolute scale, where  $I$  is the scattering intensity arising from the contrast between the lipids and the solvent, while  $Q$  is the scattering vector.<sup>154</sup> In the determination of the contrast matching point (CMP) of the lipid mixtures, the scattering intensity was normalized by their relative concentrations. From a linear fitting of  $\sqrt{I}$  vs. D<sub>2</sub>O ratio, the CMP was calculated for each lipid.<sup>155</sup> The SANS data were further fitted by the lamellar model in SASView (sasview.org) to obtain the lipid bilayer thickness. The calculated CMP from MS data is based on  $SLD = \frac{\sum_{i=1}^n b_i}{V_m}$ , where  $b_i$  is the scattering length of atoms in a molecule with volume  $V_m$  (Table 5-4) and SLD is the scattering length density.<sup>156</sup> Then the SLD is converted to CMP in terms of D<sub>2</sub>O concentration.

## **Results and Discussion:**

### *Lipid Identifications*

After analysis of the unlabeled (H-Gly/0) data as described above, 29 PE lipid species and 16 PG species were confidently identified. These species and their peak areas are displayed in **Figure 5-3**. Several cardiolipin (CL) species were also observed, but a detailed analysis of CL lipids was not pursued. The deuteration levels of CL are expected to be very similar to PG since CL is formed via condensation of PG lipids. The results generated here are generally in accord with prior lipid analyses in *E. coli*, which show PE lipids to be the most abundant (~70-90% of total phospholipids), PG to be next most abundant (~5-15%), and a more minor contribution of

cardiolipin (~1-9%).<sup>22-25</sup> Varying reports additionally record phosphatidic acid (PA), phosphatidylserine (PS), or lysophosphatidylethanolamine (LPE) lipids.<sup>22-25</sup> However, these latter classes seem to be more minor components with PA being an intermediate in phospholipid synthesis and PS being an intermediate for PE lipids. Fatty acids identified previously range from 10:0 to 19:1, but 16:0, 18:1, and 16:1 were the most common.<sup>23, 148-150</sup>

#### *Deuterium Incorporation and Profiling*

In all conditions with labeled media, lipids were present in broad isotopic distributions. It is common to have 14-22 different isotopologues within 5% of the mode value, except when both the water and the carbon source are deuterated. An example distribution is shown in **Figure 5-4**. This distribution indicates that the range of deuterium labeling states is not negligible and needs to be considered for applications of endogenous lipid deuteration. These distributions appear to follow a Gaussian (binomial) trend and likely reflect stochastic use of H/D in lipid biosynthesis. Naïve distributions were calculated according to equation 1 assuming the  $p_D$  is equal to the %D<sub>2</sub>O in the media. These distributions did not match the observed isotopic distributions with the observed distributions showing lower deuterium incorporation with narrower distributions (**Figure 5-4.C**). After fitting the  $p_D$  based to the empirical distributions using our python script, the calculated binomial curves fit the observed data very well (**Figure 5-4.C**).

**Tables 5-5** and **5-6** show the best fit  $p_D$  values generated for all identified lipids, while **Figure 5-5** displays the results graphically. These probabilities are useful as they describe the overall deuterium incorporation of specific lipid species. However, a couple of samples (50D-Gly/0, 50D-Gly/100, D-Glu/0) had multimodal distributions which could not be accurately profiled using this method. Additionally, at 100% D<sub>2</sub>O, there was some skew from a perfect binomial, which may affect the precision of the fitted values. There is generally good agreement between the deuteration values calculated for different lipids within the same class, albeit with a few outliers. These outliers tend to be low abundance lipids, for which the data are noisier. The best fitting  $p_D$  (which corresponds to the average deuterium incorporation) is consistently lower than the percent of D<sub>2</sub>O in the media. No obvious trends of deuterium incorporation with respect to degrees of unsaturation or chain length were observed, except that longer chain PG lipids tended to have slightly higher deuteration than shorter chain lipids in the conditions containing unlabeled glycerol.

### *Lipidome Scale Deuterium Incorporation*

While knowing the deuteration states of individual lipids is powerful, it is often more useful to characterize the bulk lipidome deuteration. To do this, a weighted average  $p_D$  (average deuteration) was calculated for each lipid class using the summed isotopologue abundance of a lipid species as its weight. **Figure 5-6** shows the weighted average  $p_D$  values for PE and PG lipids with different proportions of  $D_2O$  in the media with either glucose or glycerol as the carbon source. As an estimate of the experiment-to-experiment precision of this *in-vivo* deuteration, three additional replicates of H-Glu/75 were prepared at a separate time from the three already described. PG lipids were analyzed in these samples and a weighted average PG deuteration was found to be 55.6%, which is very close to the 56.1% calculated here, and within the bootstrap-estimated confidence interval of 55.5% to 57.1% (see error bars in **Figure 5-6**). The deuteration percents for the twelve individual PG lipids are given in **Table 5-6**.

For both PG and PE lipids, the relationship between  $p_D$  and  $D_2O$  concentration could be explained by a linear fit with high  $R^2$  values (**Figure 5-7**). However, the response has a slight non-linearity to it, likely due to kinetic isotope effects. Thus, a second-degree polynomial regression was performed, which is expected to be slightly more accurate (**Figure 5-8**). This relationship can be used to predict deuteration for proportions of  $D_2O$  not tested in this study.

It is clear that the deuterium incorporation achievable for both PE and PG lipids is lower than the media  $D_2O$  percent, as shown in **Figures 5-6** and **5-8**. This is consistent with prior work showing a similar effect in proteins.<sup>144</sup> There are two possibilities that could explain this: 1) there could be a kinetic isotope bias toward protium in lipid biosynthesis, resulting in a disproportionate representation of H/D, or 2) hydrogen from the carbon source could be incorporated into the lipid. The second factor, at least, is occurring since growth at 100%  $D_2O$  with an unlabeled carbon source (where there is no meaningful amount of protiated water) yields only 79-84% deuteration in the lipids. This is also consistent with the biosynthetic mechanisms that preserve certain C-H bonds from the carbon source through acetyl-CoA to fatty acids. However, there are likely kinetic isotope effects as well. This is observable by performing a linear regression without the 100%  $D_2O$  samples (**Figure 5-7.B**). Excluding 100%  $D_2O$  yields a fit with a smaller slope and higher linearity compared to **Figure 5-7.A**, indicating that when protium is present, it is overrepresented in the resulting lipids. That being the case, the

incorporation of hydrogen from the carbon source still has a much stronger effect than do kinetic isotope effects.

### *Deuterium Localization*

In order to investigate the localization of deuterium in the lipids, fragmentation data were examined. In negative ion mode, phospholipids (more specifically glycerophospholipids) fragment primarily at the lipid tail-headgroup connections (see **Figure 5-1**). By examining the fragmentation spectra of a given isotopologue, the deuterium can be localized to either the fatty acid tails or the headgroup (considering the glycerol backbone to be part of the headgroup). For the D-Gly/0 sample, the fragmentation data consistently show 10 isotopes on the PG headgroup with a distributions of FA tail isotopes (**Figure 5-9**). This shows complete labeling of the headgroup, as PG lipids have ten non-solvent exchangeable hydrogen in the headgroup all of which are found in glycerol moieties. Conversely, the H-Gly/100 sample shows heavy tail deuteration with little to no isotopes on the headgroup of PG lipids (**Figure 5-9**). Samples 50D-Gly/0 and 50D-Gly/100 show multimodal isotopologue distributions and fragmentation patterns which are consistent with the glycerol moieties in the lipid predominately being either fully deuterated or fully protiated (data not shown).

An examination of the fragmentation patterns of lipids PG 32:0, PG 34:2, PG 36:2, PE 32:0, PE 34:2, and PE 36:2 were performed to characterize the deuteration of three of the most common fatty acid tails, FA 16:0, FA 17:1 (likely 17cy3:0), and FA 18:1. FA fragment ion relative intensities were profiled for the different isotopologues and scaled by the relative abundance of the given isotopologue. The resulting FA isotope distributions were fit to Equation 1. **Table 5-7** shows the FA percent deuteration as calculated from the observed fragment ions which were between 70% and 93% of the D<sub>2</sub>O percentage. To verify that these are reasonable estimates of the real FA distributions, the FA distributions were calculated independently for the H-glycerol conditions with varying proportions of D<sub>2</sub>O in lipids PG 32:0, PG 34:2, and PG 36:2. This was done by refitting equation 1 to the observed isotopologue distributions of the intact lipids with  $n$  and  $x$  restricted to the hydrogen present in the tails. Since in these conditions the headgroup is unlabeled, the observed distributions should reflect the convolution of the distributions of each tail. For lipids with two identical tails, the convolution of two tails has the same  $p_D$  as that of the constituent tails. Thus, the  $p_D$  as fit to PG 32:0, PG 34:2, and PG 36:2 was taken as the  $p_D$  of FA 16:0, FA 17:1, and FA 18:1, respectively. These agreed with the fragment-

calculated  $p_D$  values with relative errors of at most 3.0% and an average of 1.1% (See **Table 5-8**). This indicates the fragment-calculated values are reasonably reliable. Fitting the fatty acid-specific deuteration vs the percent  $D_2O$  for the unlabeled glycerol conditions with a second-degree polynomial gave a strong fit ( $R^2 = 0.999$ ) which should have good predictive power for any concentration of  $D_2O$  (**Figure 5-10**).

#### *Hydrogen Sourcing within Lipid Biosynthesis*

In order to understand how deuterium is incorporated during lipid biosynthesis, and thereby to better design customized lipids, it is important to consider the mechanisms involved in lipid biosynthesis. Most of the glycerophospholipid hydrogen is found in the fatty acid tails, with minimal amounts on the headgroup. In any case, headgroup deuteration can be directly controlled via feeding cells deuterated precursors. *E. coli* utilizes type II fatty acid synthesis (FAS II), which is summarized in **Figure 5-11**, starting with acetyl coenzyme A. Acetyl-CoA is transformed into malonyl-acyl carrier protein (ACP) with the loss of a hydrogen during carboxylation with no additional conversion of the remaining hydrogen.<sup>157-159</sup> Malonyl-ACP is then condensed with an additional acetyl-CoA (preserving its three carbon-hydrogen bonds), beginning chain elongation.<sup>159</sup> The terminal carbonyl is then reduced in two steps resulting in the addition of hydrogen from NAD(P)H and the replacement of an acetyl-CoA hydrogen with a water-derived hydrogen.<sup>152, 159, 160</sup> One of the enzymes in this step, FabA, however, has the ability to reversibly isomerize intermediates.<sup>152, 159, 160</sup> This isomerization has the potential to scramble hydrogen from acetyl-CoA with water and vice-versa. The growing acyl-ACP chain is condensed with additional malonyl-ACP in a very similar process until a fatty acid with the desired length is achieved. Therefore, for a saturated fatty acid (assuming there is no rearrangement of hydrogen from FabA isomerization), the terminal carbon should have 3 acetyl-CoA-derived hydrogen and the rest will have one-half NAD(P)H-derived hydrogen, one-quarter acetyl-CoA-derived hydrogen, and one-quarter water-derived hydrogen. An unsaturated fatty acid would likely have the same hydrogen composition except for the vinyl carbons, one of which has a hydrogen derived from NADPH and one from acetyl-CoA (the reaction catalyzed by FabI being skipped). The hydrogen of interest on both NAD(P)H and acetyl-CoA are not solvent-exchangeable, and so their H/D state is preserved after generation. Based on this, we can conclude that a 16:0 fatty acid will have 10, 14, and 7 hydrogen derived from acetyl-CoA, NAD(P)H, and water respectively, while a 16:1 FA would have 10, 13, and 6 hydrogen. For D-

Gly/0, none of the deuterium observed on the fatty acids can be coming from the water, thus the amount of deuterium observed on the tails indicates how much of acetyl-CoA/NAD(P)H is labeled from the carbon source. For example, FA 16:0 has 3.79 deuterium on average, coming from 24 acetyl-CoA/NAD(P)H hydrogen, for an average of 15.8%. Combining data from 16:1 and 18:1 with this, the hydrogen from acetyl-CoA and NAD(P)H that enters fatty acid synthesis seems to be derived from unexchanged hydrogen on glycerol at a rate of ~17%.

### *Prediction of Lipid Deuteration*

One of the goals of this project is to enable the design of custom deuterated lipids. In order to do so, we must be able to predict what deuteration state will result from a given media composition. In a simple case where only D<sub>2</sub>O is used as the deuteration source, overall deuteration of phospholipids can be straightforwardly derived from regressions as in **Figure 5-8**. However, for certain experiments, it may be important to distinguish between headgroup and tail deuteration. Predictions of FA-specific deuteration can be similarly derived from the regressions in **Figure 5-10**, which can be deconvolved from the overall lipid deuteration to calculate the expected headgroup deuteration. However, it is possible that merely adjusting the D<sub>2</sub>O concentration will be insufficient to produce the deuteration state desired. In this case, mixing D<sub>2</sub>O with deuterated carbon sources will be necessary. Although a detailed exploration of deuterated carbon sources was outside the scope of this work, certain observations can be made. As **Figure 5-9** demonstrates, glycerol has a high affinity for labeling the lipid headgroups while showing only minimal effect on the tails. This can be explained by the fact that phospholipids contain a glycerol moiety in the headgroup (PG has two), which may be derived from glycerol in the media with no metabolic rearrangement of carbon-hydrogen bonds. This also explains why glycerol more strongly deuterates headgroups of PG lipids than PE lipids. Conversely, D<sub>2</sub>O has a major effect on the deuteration of FA tails and a minor effect on headgroups. These complementary specificities give two approximately independent ways to adjust the deuteration (and thus neutron scattering density) differentially across the lipid. To improve this independence for PE lipids, ethanolamine may need to be added to the media to fully decouple headgroup labeling from D<sub>2</sub>O. If fractional labeling of the headgroup is desired, mixing deuterated and protiated glycerol (and ethanolamine) together in the desired proportion may not be ideal. This would make glycerol labeling a binary process with some lipids containing fully labeled glycerol and some fully unlabeled glycerol in the headgroup. Indeed, this is observed in the 50D-Gly/0

and 50D-Gly/100 samples as multimodal distributions are observed reflecting stochastic incorporation of labeled or unlabeled glycerol species. Instead, fractionally labeled glycerol ought to be used for maximum precision.

However, because FA tails are partially labeled by the carbon source, to predict deuteration with using fractional D<sub>2</sub>O and D-carbon sources simultaneously it would be best to convolve the expected individual distributions. As an example, the H-Gly/100 and 50D-Gly/0  $p_D$  values for FA 16:0 (on PG 32:0) were 0.889 and 0.028, respectively, and a convolution of the two (50D-Gly/100) would be expected to have a  $p_D$  of 0.902. This is calculated by taking the product of the protium probabilities ( $1-p_D$ ) and converting back into the deuterium probability. Calculating expected convolutions in this way gave values within 3% of the observed data for 50D-Gly/100 (average 2% error), implying this may be a useful procedure for predicting deuteration derived from mixed feedstocks (see **Table 5-7**). However, further work testing more factional D-carbon source lipidomes is needed to flesh out this convolutional approach.

#### *SANS analysis of Lipid Extracts*

Lipids that were produced from H-Glu/100, H-Gly/100, and D-Gly/100 *E. coli* culture conditions were analyzed using contrast variation SANS. Each sample set exhibited similar scattering behavior across the measured Q-range. The scattering intensity (I) decreases with the increasing D<sub>2</sub>O/H<sub>2</sub>O ratio because the contrast between the deuterated lipids and solvent is similarly decreased. The SANS profiles (**Figure 5-12**) show a unilamellar lipid bilayer without evidence of multilamellar structures.<sup>139</sup> An extended lamellar sheet-like structure with power law slope ( $P$ ) of  $-2$  can be observed at Q values  $< 0.05 \text{ \AA}^{-1}$ . In most samples, the pronounced decrease in the slope with  $P = -4$  (green lines) between Q values  $\sim 0.06 < Q < 0.2 \text{ \AA}^{-1}$  indicates a Porod region with a well-defined bilayer. In low-contrast conditions (e.g., 80 and 100% D<sub>2</sub>O), the scattering features in this region deviate from  $P = -4$  and are less well-defined, suggesting a fuzzier boundary between the bilayer slab and the surrounding solvent, due to much lower contrast. The lipid bilayer thickness was obtained by fitting the SANS data between Q = 0.04 to  $0.4 \text{ \AA}^{-1}$  using the lamellar model, as described in the methods. Overall, the bilayer thickness for the samples are comparable and between 37 to 40 Å, consistent with other studies (see **Table 5-9**).<sup>161</sup>

The contrast match point of each sample was obtained point by point across the measured Q range (**Figure 5-13**), to determine if the CMP changes at different length scales. The data

show that there is no Q-dependence to the CMP at Q values  $< 0.08 \text{ \AA}^{-1}$  in the H-Glu/100 and H-Gly/100 lipids. This demonstrates that the structure is quite uniform, implying the deuterium distribution is random or otherwise evenly distributed across the lipid chains. Samples H-Gly/100 and H-Glu/100 had CMPs around 90%, while D-Gly/100 had a CMP exceeding 100% D<sub>2</sub>O. Lipids produced at D-Gly/100, then, cannot be contrast matched in any H<sub>2</sub>O/D<sub>2</sub>O ratio. However, by mixing this lipid preparation with hydrogenated (that is, protiated) or partially deuterated lipids, it is possible to produce a sample that can be matched in 100% D<sub>2</sub>O to minimize the incoherent hydrogen scattering in a sample. With the lipid species deuteration data from **Tables 5-5** and **5-6** (as well as the relative abundances of those lipids shown in **Table 5-10**), the CMP of the PE and PG fatty acid tails were calculated based on literature lipid component volumes (**Table 5-4**). These agree well with the SANS-derived CMPs with all the MS-based CMP values falling within the uncertainty range of the SANS values.

The calculated CMPs of the three lipid preparations from MS data agree quite well with SANS measurement, giving more confidence to the MS results and demonstrating the effectiveness of described analyses. While **Tables 5-11** and **5-12** shows that the calculated headgroup CMP based on MS data are quite different from the chains, the MS-derived CMPs still agree well with SANS measurements, indicating the bilayer contrast is dominated by the hydrocarbon chains. The headgroups with much less hydrogen play insignificant roles in determining the overall CMP. Also, they have solvent exchangeable hydrogen and will change SLD according to the solvent unlike the non-exchangeable hydrogens in the fatty acid chains. This can be exploited to minimize the SLD difference between a headgroup and a chain, which sometimes causes undesired scattering contribution in a SANS contrast matching experiment.<sup>9</sup> The 2D sheet-like structure from SANS indicate the lipids form well-defined bilayers with the bilayer thickness  $\sim 40 \text{ \AA}$ . It is intriguing to see, even with very different chain lengths and degree of unsaturation, that the chain thickness is still well defined. Presumably the fluid phase of lipid maintains a single thickness without obvious lateral domain formation. Knowledge of the composition in the extracted lipids is important for an accurate calculation of the CMP. Our study here shows the MS results provide reliable data for CMP calculation that can be used to design SANS experiments with selectively deuterated lipids. With the lipid composition accurately determined from *E. coli* extracts in this study, it may be possible to perform studies in living cells as viable platforms for in-vivo studies.<sup>162</sup>

## Conclusions:

These results should enable researchers to produce native phospholipids with specifically customized deuteration states, not only at the overall deuteration level, but also with regiospecificity between the headgroup and fatty acid tails. Additionally, this ought to generalize to bacteria that are metabolically similar to *E. coli*. The data generated here were validated with SANS experimentation, which confirmed the accuracy of the MS data and modeling. The SANS data also demonstrated the applicability of this approach to neutron scattering. Overall, this work has expanded the SANS-contrast variation toolbox unlocking more powerful experiments to come, while providing insights into the flow of hydrogen through metabolism.

## Part 2. Bioproduction and small-angle neutron scattering characterization of deuterated phosphatidylcholine from *Escherichia coli*:

### Citation:

Text and figures in this section are reproduced from a manuscript in preparation: “Controlled Deuteration of Phosphatidylcholine Biosynthetically Produced from Engineered *Escherichia coli*.” Qiu Zhang, Matthew J. Keller, Wellington Leite, Robert L. Hettich, Hugh M. O’Neill, and Honghai Zhang. 2024, in preparation. My contribution to this work includes performing all LC-MS measurements, data analysis and interpretation, figure generation, drafting and editing of the manuscript.

### Abstract:

Deuterated phosphatidylcholine (PC) were produced from engineered *Escherichia coli* BL21 cells grown in deuterated media. The levels of PC deuteration could be controlled by feeding cells with different combinations of deuterated or protiated water, glycerol, and choline. PC lipids were analyzed by NMR, mass spectrometry, and small angle neutron scattering (SANS). By growing cells in deuterium oxide (D<sub>2</sub>O) in the presence of fully deuterated glycerol and choline, perdeuterated PC was produced via biological synthesis for the first time. The tail distribution of PC was analyzed by mass spectrometry, which indicated that culturing cells in deuterium environment will affect the lengths and saturation stats of the PC phospholipids. Additionally, the average neutron match points of PC isolated from the engineered *E. coli*’s lipidome were experimentally determined using SANS. For partially deuterated PC lipids, SANS showed match points of  $92 \pm 8\%$  D<sub>2</sub>O for lipids with deuterated fatty acyl chains and glycerol, and  $81 \pm 2\%$  D<sub>2</sub>O for lipids with deuterated fatty acyl chains alone, Our results will benefit

SANS studies that utilize deuterated PC lipids as biomimetic lipid bilayers or membrane-like models. This is especially important for structural studies of biomolecular systems from mammalian organisms.

### **Introduction:**

Neutron scattering is a powerful tool for biophysical research, especially with the application of deuterium labeling.<sup>137, 163, 164</sup> Deuterium labeling allows the scattering contrast of different molecular components of a sample to be modulated, thereby emphasizing or removing those components from the resulting scattering profile.<sup>165, 166</sup> In neutron scattering measurements, the distinct neutron contrast between different biomolecules, such as lipids, proteins, and nucleic acids, allows for the study of individual components in a multicomponent system. This is achieved by using mixtures of deuterium oxide (D<sub>2</sub>O) and H<sub>2</sub>O to match the neutron scattering length densities (NSLDs) of each component. This then allows scattering contribution from individual components to be deconvoluted to determine their structures within complexes.<sup>6, 167-170</sup>

Phospholipids are an important target for deuterium labeling in membrane-related studies of many biological processes, including cell recognition, signaling, selective-ion transfer, adhesion and membrane fusion, as well as in studies of membrane proteins.<sup>171-174</sup> Currently, there are three main ways to obtain deuterated phospholipids: bio-deuteration, chemical deuteration, or semi-synthetic approaches that combine both routes. Chemical deuteration has been established as an efficient method to obtain partially labelled and perdeuterated phospholipids with saturated tails, directly by the condensation of deuterated fatty acids and glycerol phosphate.<sup>175</sup> However, phospholipids with saturated fatty acyl tails are not representative of the phospholipids found in biological membranes under physiological conditions, since heteroaclys (containing one saturated fatty acyl chain and one unsaturated fatty acyl chain) are common in many organisms. Although the chemical synthesis of deuterated fatty acyl chains containing one or two unsaturated bonds has been reported, its complicated synthesis renders it impractical for routine preparation.<sup>141, 176</sup> A semi-synthetic approach combining chemical deuteration and biosynthesis has been developed for the preparation of partially deuterated PC with one saturated tail and one unsaturated tail, however, this approach still suffers from time-consuming chemical deuteration of unsaturated fatty acids.<sup>177</sup> These chemical and semi-synthetic approaches provide structurally identical phospholipids for preparing model bilayers. While such model systems are useful to investigate fundamental behaviors of cell membrane, research on native cell membranes or

native-like lipid mixtures to elucidate real biological systems has recently attracted more attention.<sup>178</sup> These attempt to replicate the lipid diversity observed in actual organisms. In this context, biological deuteration is preferred to produce deuterated total lipid extracts for some applications. The deuterated lipids can be extracted from the organism growing in minimal media in defined levels of D<sub>2</sub>O that can include deuterated carbon sources such as glycerol or glucose. Deuterated phospholipids obtained from this approach have been used in the study of interactions between phospholipids and other biomolecules, including sterols, peptides and proteins.<sup>179-181</sup>

Alongside prior work, we established a well-developed procedure for the biological synthesis of deuterated total lipid extract in part 1; however, preparation of deuterated PC is still challenging due to the lack of PC in natural prokaryotic membranes.<sup>7, 182, 183</sup> On the other hand, PC is a major component of biological membranes in many eukaryotic cells, and is an essential part to mimic native-like membranes for biophysical studies of mammals and many other eukaryotes. Because of this, many efforts have been made to produce PC of which genetically modified yeast is one of the most efficient methods. In 2014, Jouhet group reported the preparation of deuterated PC in the methylotrophic yeast *Pichia pastoris*.<sup>184</sup> In the following year, Pomorsky group successfully produced deuterated PC with *E. coli* strain AL95, albeit with little characterization of resulting lipid species.<sup>143</sup> Neither of these studies examined the deuteration of PC under fractionally deuterated conditions. Additionally, since the phospholipid composition can change in response to genetic modification or the presence of deuterium oxide, it is important to identify the isolated PC for downstream characterization studies. Herein, we report our study on the extraction and purification of deuterated PC from *E. coli* that express *Pseudomonas fluorescens* PC synthase (pf-Pcs) and its characterization by MS and SANS analysis. This can be combined with our work in part 1 and will provide a rational basis for the design of regiospecifically deuterated phospholipids (native and/or transgenic) with native-like fatty acid distributions. This expands out ability to produce biomimetic lipid bilayers for mammalian biophysical studies.

## Material and Methods:

### *Chemicals*

Ethyl oxamate, Choline, lithium aluminum deuteride ( $\text{LiAlD}_4$ ), deuterium oxide ( $\text{D}_2\text{O}$ ), methyl-d3 iodide ( $\text{CD}_3\text{I}$ ), silver oxide ( $\text{Ag}_2\text{O}$ ), sodium hydroxide ( $\text{NaOH}$ ), ethanol ( $\text{EtOH}$ ) and anhydrous tetrahydrofuran (THF) were purchased from Sigma-Aldrich and used as received without any further purification. Total lipid extract was purchased from Avanti polar lipids.

### *Strains and culture conditions*

To produce PC in *E. coli*, the pET24b plasmid containing the gene encoding pf-Pcs was transformed into *E. coli* BL21(DE3). Initially, 5 mL of Luria-Bertani (LB) culture was inoculated from a single colony and incubated at 37 °C overnight. The cell cultures were then adapted to grow in  $\text{D}_2\text{O}$  Enfors based minimal media with either H-glycerol or D-glycerol as the carbon source.<sup>183</sup> The cells underwent three exchanges of media to ensure complete adaption from  $\text{H}_2\text{O}$  to  $\text{D}_2\text{O}$  based media. For each exchange, 250  $\mu\text{L}$  of cell culture was centrifuged at 4000Xg for 10 min and the cell pellet was resuspended into 5 mL of the appropriate fresh media and incubated at 37 °C until the cell density at 600 nm ( $\text{OD}_{600\text{nm}}$ ) reached stationary phase ( $\text{OD}_{600\text{nm}} = \sim 5$ ). After the final exchange, 1 L of culture was grown in the H or D Enfors minimal media supplemented with H or D glycerol and H or D choline respectively with initial inoculation at  $\text{OD}_{600\text{nm}} = \sim 0.1$ . The cells were induced with 0.1 mM isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) to switch on expression of Pcs at  $\text{OD}_{600\text{nm}}$  of 0.8–1. Protiated or deuterated choline (1mM final concentration) was added at this time and cell growth was continued at 30 °C for 18 hours. The cells were harvested by centrifugation at 6000 x g for 30 minutes and then washed with 1% NaCl three times followed by freeze-drying for 48 h. The final dry weight was  $\sim 1.1$  g cells/L culture. The freeze-dried cells were stored at -80 °C for storage.

### *Separation of total lipid extract*

*E. coli* total lipids were extracted using the modified Bligh and Dyer method.<sup>61</sup> In brief, 250 mL chloroform, 500 mL methanol and 200 mL water were sequentially added to the freeze-dried cells (1.1 grams) to achieve a final chloroform: methanol: water ratio of 1: 2: 0.8 (v/v/v). The cells were vortexed for 15 seconds immediately following the addition of each solvent and allowed to stand for approximately 18 hours at room temperature with stirring. This was followed by addition of chloroform (250 mL) and water (250 mL) to obtain a final chloroform: methanol: water ratio of 1: 1 : 0.9 (v/v/v) to separate the aqueous and organic fractions of the

mixture. A separatory funnel was used to collect the lower chloroform phase which contains total lipid extract. The chloroform methanol mixture was then evaporated under reduced pressure using a rotovap (Buchi, R-205). The lipid residue was dried under vacuum overnight, from which 34 mg was obtained.

#### *Fraction of total lipid extract for PC purification*

Total lipids were separated by passing through silica gel column using chloroform: methanol: water (65: 25: 4) as mobile phase. The structure and deuterated level of lipids were confirmed by  $^1\text{H}$  NMR. The naming scheme for the position of the deuterium in the PC lipid is represented in the following order: tail – glycerol – head group. Non-deuterated PC grown from H-medium had a yield of 11 mg from 1.6 g of dry cell (PC-HHH). Tail-deuterated PC grown from  $\text{D}_2\text{O}$  had a yield of 9 mg from 0.56 g of dry cell (PC-DHH). Proton NMR showed an average deuteration of 72% . Tail-linker-deuterated PC grown from  $\text{D}_2\text{O}$  and D-glycerol had a yield of 6 mg from 0.46 g of dry cell (PC-DDH). Proton NMR showed an average deuteration of 82%. Head group-deuterated PC grown with D-choline showed a yield of 11 mg from 0.75 g of dry cell (PC-HHD). The average deuteration calculated from proton NMR was 12%. Per-deuterated PC grown from  $\text{D}_2\text{O}$  and D-glycerol with D-choline yielded 12 mg from 0.75 g of dry cell (PC\_ddd). Proton NMR showed only tiny peaks and were unable to be integrated, which suggested >99% average deuterium incorporation.

#### *Chemical synthesis of D-choline*

Ethyl oxamate (2.34 g, 20 mmol) was dissolved in anhydrous tetrahydrofuran (40 mL) at 0 °C.  $\text{LiAlD}_4$  (1.68 g, 40 mmol) was added portionally while the temperature was maintained at 0 °C. Then, the temperature was slowly raised up to 70 °C and the mixture was stirred overnight. Water (2 mL) was added slowly to quench the reaction. The solid was removed by filtration through celite and washed with mixture of methanol and dichloromethane (20 mL: 60 mL). The HCl solution (6 N) was added to the filtrate to adjust pH value to 5. The solvent was evaporated to form crude product for next step without any purification. Ethanolamine hydrochloride salt- $\text{d}_4$  (1.01 g, 10 mmol), iodomethane- $\text{d}_3$  (8.76 g, 60 mmol) and sodium hydroxide (2.8 g, 70 mmol) were dissolved in ethanol/water (48/18 mL) and the mixture was stirred overnight. After the reaction completed, the solvent was evaporated to dryness. Methanol (30 mL) was added to the residue and the suspension was filtered to obtain a clear solution. Methanol was removed again by evaporation. Then, water (10 mL) and silver oxide (1 g) was added to the residue and the

mixture was stirred overnight. The white solid was removed by filtration and the solvent was evaporated to dryness. Water (20 ml) was added to the residue and insoluble solid was removed by filtration to obtain a clear solution. The HCl solution (6 N) was added to adjust the pH to 4 and the mixture was kept stirring overnight. The solvent was evaporated to obtain choline- $d_{13}$  chloride salt with a final yield of 1.3 g. The average deuteration for  $d_4$ -Ethanolamine and the synthesized choline was calculated by NMR to be 99%.

#### *Mass spectrometric characterization of lipid samples*

Samples were redissolved in the “8Bu” solvent mixture (69:23:8 H<sub>2</sub>O:isopropyl alcohol (IPA):butanol (BuOH)) for analysis.<sup>58</sup> Samples were measured in positive and negative ion modes on a Dionex Ultimate-3000 reversed phase LC platform coupled online to a ThermoFisher LTQ-Orbitrap Velos Pro mass spectrometer. This method has been adapted from one developed by Dann-Rasche *et al.* and employed a 60% acetonitrile and 40% water mobile phase A with a 90% isopropanol 10% acetonitrile mobile phase B, both containing ~4.2 g/L ammonium acetate with gradient details in the table below (to facilitate dissolution of ammonium acetate, 1% water may be used in mobile phase B).<sup>58</sup> Reversed phase columns with an inner diameter of 100  $\mu$ m were packed in-house using Kinetex 1.7  $\mu$ m C18 (100 Å) stationary phase (Phenomenex—Torrance, CA) to a length of about 15 cm. The mass spectrometer was operated with one full-scan acquired at 30k resolution followed by 10 data-dependent fragmentation scans in the ion trap with a collisional-induced dissociation normalized collision energy of 30.

#### *Lipid Identification*

Lipid identifications were performed on the unlabeled samples and propagated to the labeled conditions to avoid the identification difficulties arising with more complex isotopic states. These identifications were performed through a manual analysis of the LC-MS/MS data, including chromatographic characteristics and fragmentation spectra. Lipids were annotated according to the LIPID MAPS convention at the species level (reporting the summed tail carbon and unsaturation).<sup>76</sup> Unsaturated odd chain lipids were identified and were labeled according to their total carbons and apparent unsaturation without determining if they are cyclopropyl fatty acids (FAs), which generate the same mass and fragment ions in CID fragmentation as true odd chain unsaturated lipids.<sup>76, 185</sup> The various isotopologues of the identified lipids were quantified in the samples by manual peak area integration using EI-MAVEN.<sup>94</sup>

### *Analysis of Lipid Deuteration*

Lipid deuteration was calculated following the method described in Keller *et al.* 2024.<sup>7</sup> Briefly, isotopologue distributions were modeled using a binomial function (same as equation 1 as described in part 1) to characterize the deuterium distribution of the lipids. This assumes that the distributions are generated by stochastic sampling from a mixed pool of protium and deuterium. This equation was used to generate expected lipid deuterium distributions by calculating  $P(x)$  for all values of  $x$  between 0 and  $n$  (inclusive). The probabilities are also adjusted for the natural isotope distributions for carbon, oxygen, and nitrogen since the fine isotope structures of the lipids were not resolved. This equation is combined with a python script (available at: [https://github.com/m-j-keller/Lipid\\_Deuteration\\_Calculator](https://github.com/m-j-keller/Lipid_Deuteration_Calculator)), which fits this equation to the data by optimizing the value of  $p_D$ . This returns the deuterium probability for that lipid that best fits the observed data. It corresponds to the percent deuteration of that lipid species. In the case of PC lipids with deuterated media and unlabeled choline, the hydrogen in choline was subtracted from the labelable hydrogen so that an accurate fit could be obtained using equation 1. The resulting  $p_D$  was then scaled according to the total number of non-exchangeable hydrogen. Weighted average deuteration values were calculated by weighting individual lipid species deuteration states by their relative abundance within the class.

### *Lipid Class Abundance Calculation*

For the PC/PE ratios, peak areas for the relevant species (monoisotopic for the unlabeled condition, and all reasonable isotopologues around the mode value for labeled conditions) were integrated in EI-MAVEN. The peak areas were scaled based on the ionization efficiencies for PE and PC as per the following. A serial dilution of the EquiSplash lipid standard mixture (from Avanti polar lipids) was analyzed with the same LC-MS method as the samples. The ratios of the peak areas between the PC and PE standards were obtained across the calibration curve and adjusted based on the molar concentrations of the lipids. These correspond to the relative ionization efficiencies of the different lipid species. Since this can vary at different concentrations, a linear regression was performed on the PC/PE ionization ratio vs PE standard intensity. The linear fit was used to scale the PE peak areas based on where the PE intensity was on the linear fit. For peak areas outside those used in the calibration curve, the endpoint values of the curve were used. The scaled values were summed, and the ratio of PC/PE summed

abundance was taken. The values generated are approximate estimates of the molar ratio of the two lipid classes.

#### *SANS analysis of lipid samples*

SANS measurements were performed at the Bio-SANS instrument located at the High Flux Isotope Reactor (HFIR) in the Oak Ridge National Laboratory (ORNL). A standard configuration of the dual detector system was set with the Panel Scan feature of the small-angle scattering instrument with the main detector at 15.5 m and the wide-angle curved detector at 1.13 m and rotated 1.4° to sample position. This instrument setting covered the  $Q$ -range of 0.003 Å<sup>-1</sup> to 1 Å<sup>-1</sup>, using neutron wavelength ( $\lambda$ ) of 6 Å with a relative wavelength spread ( $\Delta\lambda/\lambda$ ) of 13.2%. SANS data were corrected for instrument background, detector sensitivity, and instrument geometry using facility data reduction software, drt-SANS.<sup>154</sup> All SANS measurements were performed in 1 mm path-length cylindrical quartz cuvettes (Hellma, Müllheim, Germany) at room temperature.

### **Results and Discussion:**

#### *Characterization of PC in total lipid extract*

Initial experiments analyzed the production of phosphatidylcholine (PC) in *E. coli* cells expressing *P. fluorescens* Pcs (pf-Pcs *E. coli*). Thin layer chromatography (TLC) showed the presence of PC which was then confirmed through LC-MS analysis. As shown in **Figure 5-14.A**, the chromatographic dimension of total lipid extract from pf-Pcs-*E. coli* in protiated (HT, third spot from left) or deuterated media (DT, fourth spot from left) clearly showed one spot for PC (PC, first spot from left), while no PC spot was observed in commercial total lipid extract (CT, second spot from left). This result indicated that *E. coli* cells expressing *P. fluorescens* Pcs are able to produce PC lipids in both protiated and deuterated media. The phospholipid composition of pf-Pcs-*E. coli* that were grown in protiated or deuterated media was further analyzed by mass spectrometry. Several features corresponding to PC lipids were found in total lipid extract from pf-Pcs-*E. coli* in protiated media after manual data analysis. For example, PC 18:1\_16:1 can be identified from its fragmentation spectrum as shown in **Figure 5-14.B**. The precursor ion ( $[M+H]^+$ ) at  $m/z$  758.65 (758.5676  $m/z$  in the orbitrap full-scan) confirmed the neutral formula of C<sub>42</sub>H<sub>80</sub>NO<sub>8</sub>P at 2.5 ppm error. The ion at  $m/z$  575.57 represented the lipid with phosphocholine neutral loss, which indicated the lipid is PC. One of the lipid tails was identified by fragment ions at  $m/z$  522.49 and 504.37, which individually represented lipid with fatty acid 16:1 ketene

neutral loss and carboxylic acid neutral loss. The other tail of lipid was similarly identified as fatty acid 18:1 by fragment ions at  $m/z$  494.43 and 476.34. This identifies the feature as the PC 18:1\_16:1 lipid, although for the sake of clarity, lipids were reported at the species level (PC 34:2, in this case).

#### *Chain distribution in PC*

Deuterated (or protiated) PC was fractionated and purified by passing through a silica gel column using chloroform: methanol: water (65: 25: 4) as the mobile phase. Since the isolated PC is a mixture with different fatty acid chains and will be used as a model for mimic membranes, it is necessary to identify its composition in detail. The structure of isolated PC was first confirmed by  $^1\text{H}$  NMR. The peak at 5.36 ppm represents the two hydrogen at carbon double bonds (unsaturation) in the fatty acid chain and the peak at 5.22 represents the singular hydrogen on carbon 2 of glycerol backbone. The ratio of integration of these two peaks as 1.80 suggested that isolated PC contains an average of 0.9 double bonds per PC molecule. To obtain more details of PC composition, mass spectrometry was employed to analyze the fatty acid chain distributions. While some PC lipid species were not identified in each condition, a total of 21 PC species were identified in all conditions from the LC-MS data. The relative abundances of these species are shown in **Figure 5-15**, with the most abundant being PC 32:1, PC 34:1, PC 32:2, PC 34:2, and PC 30:1 (which can all be formed from combinations of relatively common fatty acids in *E. coli*). It is noteworthy that the deuterated growth medium, deuterated glycerol, and even deuterated choline can have a large impact on the distribution of fatty acid chains. Although the major species is PC 32:1 for each sample, other species show more dramatic changes. For example, PC 30:0 goes from the seventh most abundant species in the protiated medium to the third most abundant in the fully deuterated medium. Additionally, the average amount of unsaturated bonds in each PC molecule was 1.1 in the protiated condition but 0.87 in the deuterated medium. While this could be explained to some degree by the lack of biological replicates, this result suggests the elongase enzymes may show varied activities in protiated and deuterated environment. Further work and additional replicates are needed to confirm this, however.

#### *Control of deuterium labeling level and deuterium labeling position*

*E. coli* cells grown in deuterated media containing H-glycerol and H-choline produce partially deuterated phosphatidylcholine. NMR analysis shows a significant decrease of proton

signal on the fatty acyl chain, indicating deuterium incorporation. The most obvious change is the decrease of the large signal at 1.2 ppm corresponding to the deuteration of the majority of the hydrogens on fatty acyl chains (see **Figure 5-16**). The glycerol signals at 4.0, 4.2 and 5.2 ppm, as well as the signals at 3.4, 3.9 and 4.4 ppm belonging to the choline head group, remained the same as the signal of unlabeled PC molecule (**Figure 5-16**). Mass spectrometry of PC extracted from these cells showed that there was an average of 22 residual protium in these lipid species, which is consistent with direct incorporation of H-glycerol and choline into the lipid (for a total of 18 protium) with some glycerol protium getting incorporated into the fatty acid tails. Similar labeling trends have been observed for phospholipids in *E. coli* BL21(DE3) previously.<sup>7</sup> To investigate the carbon source effect on deuteration, *E. coli* cells were next grown in deuterium media containing D-glycerol and H-choline. PC purified from these cells showed only choline group peaks at 3.4, 3.9 and 4.4 ppm identical to that of unlabeled PC. However, the proton peaks of the glycerol linker at 4.0, 4.2 and 5.2 disappeared in this <sup>1</sup>H NMR spectrum (see **Figure 5-16**). These PC species showed an average of 13 protium in the MS data, consistent with full deuteration of the fatty acids and glycerol backbone and unlabeled choline. These results show that the proton/deuterium incorporation on PC's acyl chain derive from media, while glycerol and choline added to the growth media were directly incorporated into PC molecules without modifications. Thus, selective deuteration of PC can be controlled by employing the combination of protiated/deuterated water, glycerol, and choline.

#### *Perdeuterated choline and PC*

Although biosynthesis of deuterated total lipid extract and PC has been explored for many years, perdeuterated PC has never been prepared via this method due to the lack of perdeuterated choline. Therefore, we decided to synthesize perdeuterated choline through chemical deuteration. We adapted a modified method from Yepuri *et al.* to obtain pure product.<sup>176</sup> The deuterium incorporation level of choline was identified as 99% by NMR. To test if the perdeuterated choline synthesized in this study behaves as the same as the commercial non-deuterated choline, *E. coli* cells were next grown in protiated media containing H-glycerol and D-choline. <sup>1</sup>H-NMR analysis of this PC shows absence of proton signals from choline at 3.4, 3.9 and 4.4 ppm and presence of all other proton signals from glycerol and tails. In this context, we prepared fully deuterated PC from *E. coli* cells grown in D<sub>2</sub>O containing D-glycerol and D-choline. NMR analysis of this fully deuterated PC shows only tiny peaks (unable to be

integrated) at the same positions with the unlabeled PC, which indicated its high level of deuterium incorporation. The mass spectrometry analysis of this sample is consistent with NMR analysis result, which showed deuteration level of fully deuterated PC as 99.6%.

#### *Small-Angle Neutron Scattering Analysis*

We used Small-Angle Neutron Scattering (SANS) and contrast variation to determine the average contrast match points of the phosphatidylcholine lipids produced in this study. Using our MS data on the abundance and length of different fatty acids chains, we first calculated the average neutron scattering length density (NSLD) for each different PC lipid and expressed these values in terms of its equivalent in the NSLD of an H<sub>2</sub>O/D<sub>2</sub>O mixture - commonly expressed as H<sub>2</sub>O/D<sub>2</sub>O ratios. Our prediction of NSLDs using MS data indicates that perdeuterated PC lipids have an NSLD above the physically possible limit for contrast matching experiments (*i.e.*, 100% D<sub>2</sub>O). And predicted values for contrast match points for lipids with deuterated fatty acyl chains and deuterated fatty acyl chains plus glycerol linker were 74% D<sub>2</sub>O and 88% D<sub>2</sub>O, respectively, while fully protiated PC lipids have the lowest predicted contrast match point (28% D<sub>2</sub>O) as expected for other fully protiated phospholipids molecules. We also used SAXS to structurally characterize both protiated and deuterated PC lipids. SAXS profiles indicated that extracted PC lipid samples exhibited multilamellar structures for both H- and D-PC lipids, with multiple concentric layers (**Figure 5-17**).

Analysis of the SANS data of perdeuterated PC lipids shows an overall average NSLD of 7.18 Å<sup>-1</sup>, which is greater than the physically possible solvent contrast match with NSLD of 6.36 Å<sup>-1</sup> (100% D<sub>2</sub>O) in agreement with predicted NSLD. This indicates a complete substitution of all exchangeable and non-exchangeable hydrogens in this PC lipid form. For partially deuterated PC lipids with deuterated fatty acyl chains and glycerol (PC-DDH), and deuterated fatty acyl chains (PC-DHH), experimental average neutron contrasts were 92 ± 8% D<sub>2</sub>O and 81 ± 2% D<sub>2</sub>O, respectively. These values are similar to predicted values using our MS data and theoretical estimation (deviations of 4% D<sub>2</sub>O and 7% D<sub>2</sub>O for PC-DDH and PC-DHH, respectively). In addition, linear fits for contrast match points at different *Q*-values indicate homogeneous SLD distribution for the partially deuterated lipid forms. In PC lipids with deuterated fatty acyl plus glycerol, a slightly lower average contrast match point of 90% D<sub>2</sub>O in the low-*Q* to mid-*Q* region transitions to an increase in the high-*Q* region (*Q* > 0.02 Å<sup>-1</sup>). Conversely, for PC with deuterated fatty acyl chains, the average value remains consistent across the entire *Q*-value range,

demonstrating good agreement in linear fits and confirming a homogeneous SLD distribution (**Figure 5-18**). For the fully protiated PC, our experimental determination yields a neutron contrast match point of  $22 \pm 1\%$  D<sub>2</sub>O. This experimental value is 8% smaller than that of predicted value obtained using MS data (28% D<sub>2</sub>O) and showed no  $Q$ -dependency on the neutron contrast. **Figure 5-19** shows the SANS data and  $\sqrt{I(Q)}$  versus  $Q$  plot for determination of the neutron contrast match at different  $Q$ -values for all PC lipids forms.

This shows the contrast match points for PC lipids given defined deuteration of subcomponents of PC lipids. As characterized in part 1, fractional deuteration of these subcomponents could be performed through cultivation of *E. coli* with fractionally deuterated media/feedstocks. This would allow for the regiospecific deuteration of PC to whatever deuteration state is needed for a given experiment.

### Conclusions:

Understanding the biophysical properties of lipids is crucial for investigating diverse topics that range from fundamental studies of bilayer structure and organization, membrane protein structure and function, and advancements in drug delivery and vaccine development.<sup>186, 187</sup> While bacterial lipid production holds distinct advantages over eukaryotic systems in terms of scalability and feasibility for deuteration, differences in lipid classes and their distribution presents challenges.<sup>188</sup> Building upon previous research demonstrating the feasibility of PC production in *E. coli*,<sup>183</sup> we used this platform to engineer targeted deuterated versions of PC in *E. coli*. In this work, we describe a new method to produce deuterated phosphatidylcholine from *E. coli* BL21 cells grown in deuterated media and characterize it with the combination of NMR, mass spectrometric analysis, and small angle neutron scattering.

This biosynthetic method has several significant advantages over previous methods to produce deuterated phosphatidylcholine. First, this method can produce phosphatidylcholine with controlled deuteration from 0% to perdeuteration by feeding cells with different combinations of deuterated and non-deuterated water, glycerol, and choline. For example, fully deuterated phosphatidylcholine was produced by growing cells in deuterium oxide in the presence of fully deuterated glycerol and choline. Thus it is a versatile way to selectively target the incorporation of deuterium label into lipids, offering tailored lipid variants suitable for diverse applications. Second, this method can produce phosphatidylcholine with mixed tail length for mimic of native cell membranes to elucidate real biological systems. Combining high-

resolution MS with NMR analysis enables characterization of different lipid class distributions and the distribution within individual classes, facilitating the creation of accurate membrane models that better reflect natural systems. Third, this method addresses some of the disadvantages associated with chemical synthetic production methods because it allows for the deuteration of unsaturated acyl chains.<sup>141, 176</sup>

Overall, our study provides a robust approach for the production and characterization of an important isotopically labeled lipid, that will advance biophysical understanding and pave the way for sophisticated biomembrane-related characterization studies.

## **Chapter VI - First Report of the Covalent Incorporation of Polyfluorinated Alkyl Substances into Phospholipids**

### **Citation:**

The Figures and text of this chapter are derived primarily from a draft manuscript on which I am a coauthor: “Bacteria incorporate a polyfluoroalkyl carboxylate into the lipid bilayer.” Yongchao Xie, Diana Ramirez, Matthew J. Keller, Gao Chen, Cynthia M. Swift, Robert L. Hettich, Amanda L. May, Shawn R. Campagna, Anita Thapalia, Trent A. Key, and Frank E. Löffler. My contributions to this project include performing the lipidomics measurements, identifying the FTCA-incorporated lipids from tandem MS data, validating lipid quantification, helping draft and edit the manuscript.

### **Context:**

This chapter describes the work I have been involved in which provides the first evidence of the covalent incorporation of perfluoroalkyl substances into bacterial lipids. Yongchao Xie and the Löffler lab at UTK observed from preliminary LC-MS data what appeared to be lipids containing 8:3 FTCA from a *Pseudomonas* grown in the presence of 8:3 FTCA. 8:3 FTCA is a linear carboxylic acid with two hydrogenated carbon after the carbonyl carbon followed by eight fluorinated carbon. This is essentially a short, partially fluorinated fatty acid and its incorporation into lipids in place of a fatty acid is therefore plausible. I entered the project to provide a comprehensive lipidome analysis and confidently confirm the identification of FTCA-incorporated lipids. Since no standards exist for PFASylated lipids (8:3 FTCA or otherwise), I characterized the tandem-mass spectrometry of putative FTCA-incorporated lipids *de novo* based on observed fragment ions. Fragment  $m/z$  values were matched to the chemical formulae of plausible fragment ion structures. These were cross-checked by comparison to the tandem mass spectra of ordinary lipids and free FTCA. These results and more are described below.

### **Abstract:**

Per- and polyfluoroalkyl substances (PFAS) are global contaminants of concern to human and environmental health. Little is known about biological attenuation processes and microbiota–PFAS interactions. In *Pseudomonas* sp. strain 273, we observed covalent incorporation of 8:3 fluorotelomer carboxylate (8:3 FTCA) into phosphatidylethanolamine (PE) and phosphatidylglycerol (PG), two prominent components of lipid bilayers. Semi-quantitative lipidomic analysis estimated that 7.5% of the bacterium’s phospholipid pool contained 8:3 FTCA, resulting in a PFAS membrane incorporation factor (MIF) of 423. Synthesis of PE and PG species carrying a polyfluorocarbon (PFC) acyl tail was observed in five other axenic

bacterial cultures tested, albeit with lower MIFs. We predict that the formation of fluoromembranes is a prevalent process potentially impacting the fate, longevity, bioaccumulation, and potentially biomagnification of certain PFAS.

### **Introduction:**

PFAS are a large group of synthetic chemicals used in a broad range of consumer, industrial, and military products and processes, and have emerged as global contaminants.<sup>189</sup> PFAS have been found in rainwater, in people's blood, and at least some have the potential to cause metabolic disruption, obesity, diabetes, immune suppression, and cancer.<sup>190-194</sup> PFAS are recalcitrant in the environment, with biological processes largely limited to the transformation of some polyfluoroalkyl substances (aka precursors), leading to the formation of persistent perfluoroalkyl acids (PFAAs).<sup>195, 196</sup> Prior work with fluorinated chemicals (PFAS and otherwise), have identified situations in which biomolecules can be fluorinated. For example, 8:2 fluorotelomer acrylate was shown to be reactive with cystine residues in proteins (thiol group) and formed 8:2 fluorotelomer acrylate modifications of cystines from protein lysates *in vitro*.<sup>197</sup> Given the reactivity of acrylates, and the fact this was done with protein lysates, this was not a terribly surprising result. The relevance of this for actual organisms exposed to fluorotelomer acrylates is unclear. These fluorotelomer acrylates would have to be transported inside the cell and attack cystine residues before any neutralizing cross reactions occurred. While not PFAS, monofluorodecane and difluorodecane can be catabolized by the organohalogen degrading *Pseudomonas* species strain 273.<sup>93, 198</sup> LC-MS lipid analysis of *Pseudomonas* sp. strain 273 grown with mono/difluorodecane as the sole carbon source found lipids containing 1 to 2 fluorine in their fatty acid tails.<sup>198</sup>

In this work which follows up on the fluorodecane study, *Pseudomonas* sp. strain 273 was grown in the presence of fluorotelomer carboxylates with particular interest in the lipids resulting from this cultivation. Fluorotelomer carboxylates (FTCA), a class of PFAS, are prevalently detected in aquatic environments and considered as significant precursors to PFAAs.<sup>199, 200</sup> An in-depth investigation of microbe-PFAS interaction will provide deeper insights into the environmental fate, persistence, bioaccumulation, and potentially biomagnification of PFAS.

## Materials and Methods:

### *Cultivation*

*Pseudomonas* sp. strain 273, *Pseudomonas putida* strain GB1, *Pseudomonas* sp. strain CBS3, and *Pseudomonas citronellolis* strain P3B5 were grown in completely synthetic, defined basal salt medium as described previously.<sup>198</sup> Acetate (25 mM) was provided as the primary carbon/energy source, and 8:3 FTCA was supplied from a 10 mM methanolic stock solution to reach a final concentration of 5  $\mu$ M. *Escherichia coli* K-12 was grown in lysogeny broth (LB) medium with and without 5  $\mu$ M 8:3 FTCA. *Enterococcus faecalis* (ATCC 29212) was grown in brain heart infusion (BHI) medium with and without 5  $\mu$ M 8:3 FTCA. Cultures were incubated at 120 rpm, 30 °C for 4 days before biomass was collected by centrifugation for lipidomics analysis.<sup>198</sup> To verify that PFAS incorporation into phospholipids is not an artifact of the extraction procedure, replicate cultures of strain 273 and strain GB1 grown without 8:3 FTCA received 5  $\mu$ M 8:3 FTCA immediately before centrifugation and lipid extraction (these formed the chemistry control samples). The cell enumeration of strain 273 was performed with qPCR following established procedures.<sup>93</sup>

### *Lipid Extraction and Lipidomic Analysis*

The lipid extraction and lipidome analysis followed established protocols.<sup>198</sup> Samples were analyzed with an UltiMate 3000 liquid chromatograph (LC) coupled to an LTQ Velos Pro Orbitrap mass spectrometer. The custom-made column (18 cm length, 100  $\mu$ m inner diameter) was packed with 1.7  $\mu$ m C18 Kinetex packing material (04A- 4475, Phenomenex, Torrance, CA). The LC was operated using split-flow nano-chromatography with a reversed phase method.<sup>58</sup> The eluent consisted of (A) acetonitrile:H<sub>2</sub>O (60:40, v:v) and (B) isopropanol:acetonitrile:H<sub>2</sub>O (89:10:1, v:v:v) both containing 0.4 g L<sup>-1</sup> ammonium acetate, using a 90-min multistep gradient. The samples were analyzed in negative ion mode with a top 10 DDA method where one full scan was acquired at 30k resolution with the Orbitrap followed by 10 data-dependent collisional-induced dissociation fragmentation scans in the ion trap at 33 normalized collision energy. Since presence-absence of PFASylated lipids was particularly significant to this study, samples were run in blocks where all replicates were analyzed sequentially and replicate sets were interspersed with blank measurements to ensure minimal carryover between conditions.

## Data Processing

The mass library of PE and PG lipid classes including all possible FA tails from C6 to C20 with up to eight perfluorinated carbons in each tail was generated. The data processing followed established procedures using MAVEN and only ions with an  $m/z$  mass tolerance lower than 5 ppm were integrated for abundance analysis.<sup>198, 201</sup> The abundance data represent the average value of four biological replicates. The relative abundances of lipid species of the PE or PG classes was calculated according to equation 1.

$$\text{Relative abundance of lipid species} = \frac{\text{Peak area of the lipid species}}{\text{Sum of peak areas of lipid species from the same class}} \quad (\text{eq. 1})$$

We applied a semi-quantitative approach to estimate the relative abundances (i.e., percentages) of fluorinated lipid species based on mass signal ratios. This approach does not consider variable ionization efficiencies between fluorinated and non-fluorinated species.<sup>202</sup> Absolute quantification of fluorinated lipids would require authentic lipid standards with PFAS acyl chains, which were not available.

### Calculation of a PFAS Membrane Incorporation Factor (MIF)

In analogy to the bioconcentration factor,<sup>203, 204</sup> the PFAS membrane incorporation factor (MIF) represents the amount of PFAS covalently incorporated into the phospholipid pool relative to the PFAS concentration the bacterium is experiencing in its aqueous environment. The MIF is determined according to equation 2.

$$\text{MIF} = \frac{\text{PFAS}_{\text{cell}} (\text{w/w})}{\text{PFAS}_{\text{water}} (\text{w/w})} \quad (\text{eq. 2})$$

where  $\text{PFAS}_{\text{cell}}$  represents the ratio between the weight of PFAS covalently incorporated into the cellular phospholipid pool of one cell over the wet weight of one cell.  $\text{PFAS}_{\text{water}}$  represents the mass ratio between the weight of PFAS initially dissolved over the weight of the aqueous solution. A MIF value greater than 1 indicates that the PFAS in the lipid bilayer is concentrated relative to the PFAS in the aqueous phase (e.g., the growth medium). The  $\text{PFAS}_{\text{cell}}$  ratio is calculated according to equation 3.

$$\text{PFAS}_{\text{cell}} (\text{w/w}) = \frac{\text{PFAS per lipid (\%)} \times \text{Lipid number per cell} \times \text{MW of PFAS}}{N_a \times \text{Cell wet weight}} \quad (\text{eq. 3})$$

PFAS per lipid is determined by lipidomic analysis and represents the percentage of PFAS incorporated into PE and PG, which represent the principal lipid components in *E. coli* and *Pseudomonas* spp., accounting for more than 90% of total glycerophospholipids.<sup>205, 206</sup> The lipid number per cell assumes a generic number of  $2.2 \times 10^7$  lipid molecules per cell<sup>185</sup> and an average

cell wet weight of 1.3 pg.<sup>207-209</sup> MW of PFAS is the molecular weight of the PFAS, which is 491 g mol<sup>-1</sup> for 8:3 FTCA. Na is the Avogadro constant ( $6.02 \times 10^{23}$  mol<sup>-1</sup>). PFAS<sub>water</sub> represents the ratio between the weight of dissolved PFAS and the weight of the aqueous matrix (e.g., medium), which is calculated according to equation 4.

$$\text{PFAS}_{\text{water}}(\text{w/w}) = \frac{\text{Mass of PFAS}}{\text{Mass of medium}} \quad (\text{eq. 4})$$

In this study, the initial concentration of 8:3 FTCA supplied to the medium was 5  $\mu\text{M}$ , which corresponds to a PFAS<sub>water</sub> value of 2.46 mg PFAS per kg of water.

## Results and Discussion:

### *Identification of PFASylated lipids in Pseudomonas sp. strain 273*

The soil bacterium *Pseudomonas* sp. strain 273 metabolizes terminally fluorinated alkanes, but shows no capacity to grow with PFAS as a carbon or energy source.<sup>93, 198, 210</sup> Strain 273 was grown in minimal media with 5  $\mu\text{M}$  of the fluorotelomer 8:3 FTCA. Lipidomic analysis of this culture with liquid chromatography-tandem mass spectrometry (LC-MS/MS) revealed the surprising presence of glycerophospholipids with an 8:3 fluorocarboxylate replacing one or both of the fatty acid tails. PE 34:1 was one of the most abundant lipids in strain 273, predominately consisting of 16:0 and 18:1 fatty acids (which sum to 34:1). After growth of the bacterium in the presence of 8:3 FTCA, substantial incorporation of the intact polyfluorocarbon (PFC) compound into PE lipids was observed, with PE 8:3 FTCA\_18:1 and PE 8:3 FTCA\_16:0 accounting for  $2.3 \pm 0.8\%$  and  $2.4 \pm 0.7\%$  of the total PE signal, respectively. The fragmentation spectrum of PE 8:3 FTCA\_18:1 is shown relative to PE 34:1 in **Figure 6-1**. The fragmentation spectra are clearly consistent with that of PE glycerophospholipids,<sup>211</sup> with fatty acid neutral losses and fatty acid fragment ions as the most abundant fragmentation ions (example structures of these fragment ions are shown in **Figure 6-2**). For the putative PE 8:3 FTCA\_18:1 spectrum, the same types of fragmentation occur, but now with a seemingly intact 8:3 FTCA “fatty acid” tail. The expected structures of these fragment ions from PE 8:3 FTCA\_18:1 in **Figure 6-3**. Of note, the dominant neutral loss of the FA 18:1 tail shifts from the usual ketene mechanism to the carboxylic acid mechanism when paired with a fluorotelomer. This is likely caused by being paired with a poly fluorinated tail, although the mechanism by which this occurs is unclear. It is hard to imagine the electronegativity of fluorine drawing significant electron density from the glycerol with the intervening methylene and ester groups, possibly the expected ring closer is sterically hindered by the larger fluorinated tail. Four additional fragment ions are observed in PE 8:3 FTCA\_18:1

as compared to PE 34:1, but these are explainable as sub-fragments of 8:3 FTCA ([FTCA-CH<sub>2</sub>F<sub>2</sub>O<sub>2</sub>]<sup>-</sup>, [FTCA-CH<sub>3</sub>F<sub>3</sub>O<sub>2</sub>]<sup>-</sup>, [FTCA-CH<sub>4</sub>F<sub>4</sub>O<sub>2</sub>]<sup>-</sup>, and [FTCA-C<sub>2</sub>H<sub>4</sub>F<sub>6</sub>O<sub>2</sub>]<sup>-</sup>). Three of these were also detected in our fragmentation spectrum of free 8:3 FTCA (**Figure 6-4**) and are consistent with previously reported tandem mass spectrometry of similar PFC.<sup>212-214</sup> PE 8:3 FTCA\_8:3 FTCA, a PE species carrying two 8:3 FTCA tails, was measured in strain 273 cultures supplied with the fluorotelomer, albeit in low abundances (0.03 ± 0.01%). Dual 8:3 FTCA incorporation was verified by tandem mass analysis of PE 8:3FTCA\_8:3FTCA by identifying the 8:3 FTCA ions as well as their ketene neutral losses (**Figure 6-1.C**). These structures are shown in **Figure 6-4**. Based on these described fragmentation patterns, the presence of many other FTCA-containing lipids can be confirmed including species belonging to phosphatidylglycerol, phosphatidic acid, and cardiolipin.

To verify 8:3 FTCA incorporation and confirm the assignment of the mass spectra, *Pseudomonas* sp. strain 273 samples treated with 8:3 FTCA were reanalyzed using an Orbitrap-parallel reaction monitoring method to generate high-resolution accurate mass fragmentation spectra of the putative fluorolipids. The representative PFAS-incorporated lipids including PE 8:3 FTCA\_18:1, PE 8:3 FTCA\_17:1, PE 8:3 FTCA\_16:0, PE 8:3 FTCA\_16:1, PG 8:3 FTCA\_18:1, PG 8:3 FTCA\_17:1, PG 8:3 FTCA\_16:0, and PG 8:3 FTCA\_16:1 were targeted for Orbitrap fragmentation analysis (these can be seen in **Figures 6-4** and **6-5**). PE 8:3 FTCA\_18:1, for example, had nine fragmentation ions that matched the expected fragmentation for this putative identification, and all other peaks in the MS<sup>2</sup> had less than 2% relative abundance. The nine ions explained by the structure of PG 8:3 FTCA\_16:1 consisted of a FA 18:1 ion, an 8:3 FTCA ion, a FA 18:1 carboxylic acid neutral loss, an 8:3 FTCA carboxylic acid neutral loss, an 8:3 FTCA ketene neutral loss, and the four FTCA sub-fragments described above. These fragment ions are summarized in **Table 6-1** and matched to their expected formulae with an average *m/z* error of 0.9 ppm. The other 7 FTCA-containing species had similar fragmentation spectra, with an average *m/z* error across all seven spectra of 1.2 ppm (the spectra are shown in **Figures 6-6** and **6-7**). Overall this provides very strong evidence for the proposed identifications. These observation are also consistent with a recent study demonstrating that long-chain acyl-CoA synthetase (IACS) purified from *Gordonia* sp. strain NB4-1Y can bind exogenous fluorotelomer acid and generate fluorotelomer acyl-CoA adducts.<sup>215</sup> This study provides the enzymatic basis for the incorporation for FTCA into phospholipids as reported here.

While the exact structure of the FTCA sub-fragment ions cannot be conclusively stated, there are a few plausible options that would explain the fragment formulae. These can be rationally determined given the thermodynamics of leaving groups and charge stabilization. Carbon dioxide, for example, is known to be a common leaving group where carboxyl groups are present and carbon dioxide loss could explain the missing C and O<sub>2</sub> from [FTCA-CH<sub>2</sub>F<sub>2</sub>O<sub>2</sub>]<sup>-</sup>, [FTCA-CH<sub>3</sub>F<sub>3</sub>O<sub>2</sub>]<sup>-</sup>, and [FTCA-CH<sub>4</sub>F<sub>4</sub>O<sub>2</sub>]<sup>-</sup>. These three ions would then be explained by additional losses of two, three, or four HF leaving groups. Indeed, HF loss is considered to be “common in polyfluorinated species,”<sup>213</sup> and together, these losses explain the first three FTCA sub-fragment ions. Of note, two hypothesized structures for [FTCA-CH<sub>4</sub>F<sub>4</sub>O<sub>2</sub>]<sup>-</sup> are aromatic, which would serve to significantly stabilize the carboanion and could explain why this is the most abundant fragment ion (see **Figures 6-1, 6-4, and 6-8**). After loss of four HF groups, there are no hydrogen atom remaining in the molecule, and thus no more HF losses. The mechanism by which the final sub-fragment could be generated is harder to explain, but either possible structure for the ion shown in **Figure 6-8** would be reasonably stable and therefore not an absurd possibility.

#### *Overall FTCA incorporation*

During growth in the presence of 5 μM 8:3 FTCA, inorganic fluoride release was not apparent (fluoride detection limit 0.3 μM), suggesting no C-F bonds were broken. When 8:3 FTCA was added to strain 273 cultures immediately before sacrificial lipid extraction, fluorotelomer incorporation into PE was not observed, indicating the covalent incorporation of 8:3 FTCA into phospholipids was metabolic and not an artifact of extraction. Also, fluorinated phospholipids were not detected in extractions from control incubations (e.g., without inoculum).

In strain 273 cultures supplied with 8:3 FTCA, manual peak integration of all PE and PG lipids present found that  $5.8 \pm 1.8\%$  of the total PE signal and  $11.1 \pm 1.0\%$  of the total phosphatidylglycerol (PG) signal contain one 8:3 FTCA tail (**Figures 6-9 and 6-10**). PE and PG with two 8:3 FTCA tails accounted for  $0.03 \pm 0.01\%$  and  $0.24 \pm 0.09\%$  of the respective phospholipid signals. About 7.5% of the total phospholipid signal contained one 8:3 FTCA tail (n = 4) (**Table 6-2**). The presence of 8:3 FTCA did not substantially impair growth with acetate, and strain 273 produced  $(7.8 \pm 2.9) \times 10^8$  versus  $(9.4 \pm 1.1) \times 10^8$  cells per mL (determined with qPCR) in medium with and without the fluorotelomer, respectively. Strain 273 could be repeatedly transferred in the presence of 8:3 FTCA without apparent impact on viability.

Based on the percentage of 8:3 FTCA incorporation into PE and PG species under the defined cultivation conditions applied in this study, a membrane incorporation factor (MIF, i.e., the PFAS<sub>cell</sub> / PFAS<sub>water</sub> ratio, **Figure 6-11**) of 423 was calculated for strain 273 (**Table 6-2**). Assuming an average number of  $2.2 \times 10^7$  glycerophospholipid molecules per cell<sup>216</sup>, an average cell wet weight of 1.3 pg<sup>207, 217, 218</sup>, and a MIF of 423, we estimated that about half (i.e., 0.125  $\mu$ mol or 2.5  $\mu$ M) of the exogenously supplied 8:3 FTCA has been incorporated into the strain 273 lipidome. Note that a semi-quantitative approach was used to estimate the MIF, because authentic standards for PE and PG species carrying PFC acyl tails required for absolute quantification are not available.

#### *PFAS incorporation observed in other organisms*

To test if 8:3 FTCA incorporation is unique to *Pseudomonas* sp. strain 273 or a feature shared with other bacteria, growth experiments with *Pseudomonas putida* strain GB1, *Pseudomonas* sp. strain CBS3, *Pseudomonas citronellolis* strain P3B5, *Escherichia coli* K-12, and *Enterococcus faecalis* (ATCC 29212) were performed. In cultures of other pseudomonads grown with acetate in the presence of 5  $\mu$ M 8:3 FTCA, fluorine-containing lipid species were detected in every tested strain and the percentage of 8:3 FTCA incorporation ranged between 0.3 to 0.9% (**Table 6-2**, **Figures 6-9**, and **6-10**). The calculated MIF values for *Pseudomonas putida* strain GB1, *Pseudomonas* sp. strain CBS3 and *Pseudomonas citronellolis* strain P3B5 were 17, 49 and 27, respectively, lower than the MIF of 423 determined for *Pseudomonas* sp. strain 273. FTCA containing PE and PG species were also detected in *Escherichia coli* cells grown in the presence of 8:3 FTCA, with a calculated MIF of 97 (**Table 6-2**). In the gram-positive bacterium *Enterococcus faecalis*, 0.8% of PG incorporated 8:3 FTCA resulting in a MIF of 43 (**Table 6-2**). Apparently, the ability to incorporate 8:3 FTCA into the phospholipid pool is not unique to *Pseudomonas* sp. strain 273 and may be a shared feature among diverse bacterial taxa, with a variable incorporation efficiencies vary.

#### *Investigation of substrate range for covalent incorporation of FTCA in strain 273*

In order to test the diversity of FTCA moieties capable of being covalently incorporated into bacterial lipids, 5  $\mu$ M of FTCA with different carbon chain lengths, including 5:3 FTCA, 7:3 FTCA and 10:2 FTCA, was added to strain 273 culture grown with acetate as the primary substrate. Strain 273 was found to covalently incorporate 5:3 FTCA, 7:3 FTCA and 10:2 FTCA into both PE and PG, although 5:3 FTCA was only barely detectable while the 7:3 FTCA was

similar to 8:3 FTCA. It is possible longer chain FTCA have a higher incorporation efficiency than short chain FTCA. In culture supplemented with 10:2 FTCA, however, direct incorporation of intact 10:2 FTCA into fatty acid tail was not observed. Instead, fatty acid tail substitution with 9:3 FTCA in phospholipids was detected (e.g., PE 9:3 FTCA\_18:1), see **Figure 6-12**).

Additionally, cultures supplemented with PFOA (a perfluorinated carboxylic acid) showed no PFASylated lipids. This observation indicates bacterial incorporation of FTCA may only occur with structures containing non-fluorinated  $\alpha$ - and  $\beta$ -carbons. Tandem mass analysis of representative PFAS-incorporated PE species in strain 273 with one acyl chain substituted by FTCA with different carbon chain lengths demonstrated a consistent ionization pattern with  $[\text{FTCA-CH}_2\text{F}_2\text{O}_2]^-$ ,  $[\text{FTCA-CH}_3\text{F}_3\text{O}_2]^-$ ,  $[\text{FTCA-CH}_4\text{F}_4\text{O}_2]^-$ , and  $[\text{FTCA-C}_2\text{H}_4\text{F}_6\text{O}_2]^-$  observed in addition to the typical PE fragment ions (**Figure 6-12**). These are the same as was confirmed by high resolution measurement of PE 8:3 FTCA\_18:1.

#### *Limit of detectable FTCA incorporation*

These experiments have been performed with 5  $\mu\text{M}$  concentration of FTCA. While a relatively high concentration, this level is environmentally relevant, with 6:2 FTCA having been measured in excess of this concentration before.<sup>219</sup> To better understand the concentration limits of FTCA incorporation into lipids, *Pseudomonas* sp. strain 273 was cultured with a dilution series of 8:3 FTCA covering concentrations from 20  $\mu\text{M}$  to 150 pM. Incorporation of 8:3 FTCA into PE lipids was confirmed at 160 nM and incorporation into PG lipids was confirmed as low as 78 nM. This was achieved with the standard untargeted lipidomics method we typically use, and a targeted method would likely improve the detection limit further. This clarifies the environmental relevance of the observed FTCA incorporation, with incorporation reliably detected at 38 parts-per-billion ( $\mu\text{g/L}$ ). The signal of the FTCA-containing lipids increases linearly with the concentration of FTCA before tapering off (see **Figure 6-13**). This does not appear to be a detector saturation issue since this shape is observed even for low intensity PFASylated lipids. This shape is reminiscent of the Michaelis–Menten enzyme kinetics curve and may suggest the observed tapering could be related to enzyme saturation.

#### *Formation of Fluoromembranes Potentially Impacts Environmental PFAS Pools and Bioaccumulation*

The covalent incorporation of PFAS precursors into phospholipids has the potential to reduce their respective concentrations and mobility in groundwater, the latter now governed by

the transport of bacterial cells rather than dissolved molecules. Precursor sequestration in bacterial membranes will slow (bio)transformation reactions that act on dissolved precursors to generate more recalcitrant and stringently regulated PFAAs.<sup>195</sup> The enzymatic incorporation of 8:3 FTCA into the phospholipid pool differs from passive processes, such as the partitioning of PFAS into membrane bilayers or PFAS interactions with proteins.<sup>220</sup> Passive sorption processes have been implicated in PFAS bioaccumulation,<sup>221, 222</sup> and it is conceivable that covalent incorporation of PFAS into phospholipids also contributes to bioaccumulation along with the potential for trophic magnification through food chains. PFAS have been detected and quantified in various mammal species, including humans.<sup>223-225</sup> The quantitative analysis of mammalian tissue or blood samples for extractable organofluorine (determined as fluoride by combustion ion chromatography) and PFAS (measured by LC-MS) reported significant differences,<sup>226-228</sup> hinting at unidentified reservoirs of organofluorine in mammals. The incorporation of PFAS into phospholipids of Eukarya has yet to be demonstrated but is a possible explanation for the observed discrepancies.

The assimilation of 8:3 FTCA into the lipid bilayer represents a heretofore unrecognized sink, a finding with implications for the fate and transport of PFAS in the environment. Elevated concentrations of PFAAs and precursors are commonly measured in areas impacted by aqueous film-forming foam (AFFF) applications, and precursors often dominate the PFAS pool.<sup>229, 230</sup> Saturated aquifers can reach a bacterial cell density of  $10^9$  cells per  $\text{cm}^3$ , suggesting that a substantial mass of precursor molecules could be incorporated into bacterial membranes.<sup>231</sup> Contemporary analytical procedures do not capture PFAS covalently incorporated into phospholipids, suggesting that actual environmental PFAS loadings may be higher than groundwater analytical data suggest. The fate of PFAS covalently incorporated into phospholipids is uncertain, and information about turnover and potential release is currently lacking. The incorporation of fluorotelomer carboxylates, such as 8:3 FTCA, into bacterial biomass impacts transport and longevity of these precursors, and comprehensive understanding this process is crucial for predicting the conversion of such precursor molecules to even more recalcitrant PFAAs.

## **Future Directions of Fluorinated Lipid Research:**

### *Concluding Thoughts*

The lipidomic implications for the interaction of various microbes with fluorinated chemicals is a fascinating topic and one that is rapidly expanding. The current work opens a new domain of possibilities regarding environmental PFAS and their implications for biology. To better understand the scale of covalent PFAS incorporation, and its corresponding environmental relevance, there are at least two major areas to be addressed. These are both the physiochemical limitations of covalent PFAS incorporation (to understand what PFAS are able to be covalently incorporated into lipids), and the molar concentrations of PFASylated lipids. We have begun to address the limitations of PFAS incorporation by testing a range of different fluorinated carboxylic acids, however testing a broader range of FTCA PFAS is still needed. Additionally, future work should expand the range tested to related chemical classes such as fluorotelomer sulfonates and alcohols which could plausibly be desulfonated or dehydrogenated/oxidized respectively to form carboxylic acids before incorporation into phospholipids. In the reported study, a mass spectrometry signal ratio was used to characterize the proportion of PFASylated to non-PFASylated lipids. While this signal ratio will be proportional to the actual molar ratio, there is no guarantee of equivalence, especially since PFASylated lipids will have different chemical characteristics that will likely change the ionization efficiency of the lipids. That said, there are no fluorinated lipids that can be purchased with which the concentration of PFASylated lipids could be more rigorously elucidated. Until this occurs (possibly through custom synthesis), the proportion of PFASylated lipids in an organism cannot be conclusively established.

**Chapter VII - A Multi-Omic Investigation of the Effects of Temperature on the Bioreactor  
Cultivation of *Clostridium autoethanogenum***

### **Citation:**

The figures and text of this chapter are derived primarily from a draft manuscript on which I am a coauthor: “Multi-omic analyses detail how culture temperature impacts fermentation profiles of *Clostridium autoethanogenum*.” Payal Chirania, Richard J. Giannone, Matthew Keller, Nancy L. Engle, Timothy J. Tschaplinski, David T. Reeves, Hunter Zeleznik, Vicki Liu, Loan Tran, Michael Köpke, Steven D. Brown, and Robert L. Hettich. 2024, in preparation. My contributions to this project include performing the lipidomics experiments, analyzing the lipidomics data, helping draft the manuscript, and substantial editing of the entire manuscript.

### **Context:**

*Clostridium autoethanogenum* is a fascinating bacterium with the rare ability to ferment single carbon gaseous substrates (CO and CO<sub>2</sub>) to produce energy and organic carbon for anabolism. Ethanol is the major fermentation end product, although other compounds such as acetate and 2, 3-butandiol are also produced. This makes *C. autoethanogenum* of special interest from a carbon emissions perspective as the utilization of this organism for various biotechnological applications could achieve a zero net carbon bioprocess. Indeed, carbon monoxide (a toxic pollutant) and carbon dioxide can be used as sole carbon sources. In the case of carbon dioxide which is already oxidized, hydrogen gas (H<sub>2</sub>) is also required to provide the free energy needed for carbon fixation.<sup>232</sup> Given this unique ability, LanzaTech is using *C. autoethanogenum* as a chassis organism for biofuel production. We have a collaboration with LanzaTech to study *C. autoethanogenum*'s physiology and improve its capability for biofuel production. While other investigations of *C. autoethanogenum* have been performed in our lab previously<sup>233</sup>, of interest in this project was the response of *C. autoethanogenum* to different cultivation temperatures. Given the cell membrane is an important factor for temperature tolerance, lipidomics was performed in addition metabolomics and proteomics, which have been applied to *C. autoethanogenum* before. This makes the present study the first lipidomic investigation of *C. autoethanogenum*. This is integrated into a larger multi-omic study of the effects of temperature on *C. autoethanogenum* with a special focus on its fermentative profile.

### **Abstract:**

The fermentation of waste gases is a sustainable alternative for producing commodity chemicals and biofuels. *Clostridium autoethanogenum* is a model acetogen that can directly convert a wide range of gas mixtures containing CO, CO<sub>2</sub> and H<sub>2</sub> into non-petroleum-based fuels, including ethanol, butanol, and other industrially relevant chemicals. Growth temperature is an important factor that can alter growth rate and end-product selectivity, however the majority of *C. autoethanogenum* studies are conducted at 37°C. Here, we studied growth at either 30°C or 40°C using chemostats in conjunction with multi-omic approaches (metabolomics, proteomics, and lipidomics) to characterize the effect on product yields and overall metabolism. Growth at 40°C resulted in greater acetate selectivity (27%) compared to the lower temperature (9%) with concomitantly less ethanol produced (56% compared to ~73%). Omics analyses revealed that enzymes from both ethanol producing pathways (i.e., from acetate or acetyl-CoA) exhibited significantly reduced abundances at 40°C, consistent with fermentation product profiles. In contrast, enzyme abundances for other products (acetate, lactate, and 2,3-butanediol) increased at the higher temperature. Overall, high temperature caused a dramatic metabolic shift in the microbe, whereby >60% of the quantified proteins, metabolites, and lipids showed a significant difference in abundance. Multiple fatty acids and monoglycerides accumulated along with potential membrane remodeling at the higher temperature impacting the metabolism of fatty acids and lipids. Additionally, proteins from two different predicted bacterial microcompartment (BMC) gene clusters exhibited large but contrasting abundance differences at the two temperatures, suggesting they are temperature responsive. The characterization of this model acetogen at two temperatures provides opportunity to update a range of metabolic models while detailing the systemic changes to proteins, lipid, and metabolic product profiles to help guide future strain design.

### **Introduction:**

There is significant interest in energy and chemical production from renewable resources. This is needed so as to meet future industrial and commercial demands while aligning with stated environmental objectives and ensuring a reliable supply of feedstocks. Microbes have diverse and resilient processes that can aid in these goals and provide greener energy sources. Indeed, the Acetone Butanol Ethanol (ABE) fermentation of starches and sugars by *Clostridia* at industrial scales is significant and impactful, as reviewed recently.<sup>234</sup> However, starch and sugars are

limited as a carbon source for ABE fermentation in that they require significant agricultural investment which diverts already scarce arable land from food production.<sup>235</sup> The fermentation of waste gases is a sustainable alternative for producing commodity chemicals and biofuels that has been commercialized more recently.<sup>236</sup> One-carbon or C1 substrates such as CO<sub>2</sub> and CO are inorganic carbon sources that can be fixed into biologically available carbon and are abundantly found, inexpensive, do not compete with food, and are obtainable from agricultural, municipal, or industrial waste streams. A small subset of acetogens, such as *Clostridium autoethanogenum*, can directly convert gaseous substrates into non-petroleum based fuels, including ethanol, butanol, and other industrially relevant chemicals.<sup>236, 237</sup> *C. autoethanogenum* is a Gram-positive anaerobic bacterium that can ferment these waste gases to valuable solvents via its native Wood-Ljungdahl pathway; a feat that was recently demonstrated whereby the biocatalyst was engineered to produce non-native chemicals such as acetone and isopropanol from gas with high productivities.<sup>233</sup>

Various operating parameters such as media composition, pH, trace elements, and temperature can affect fermentation, end-product titers, and process stability, all of which are important metrics to consider when optimizing product output.<sup>238</sup> Studies on other acetogens/*Clostridium* species have reported increases in ethanol production at temperatures lower than the organism's optimum<sup>239</sup> and that both substrate consumption and solvent profiles are significantly impacted by changes in temperature. Specific to gas fermentation, several studies have shown that ethanol titers are higher when conducted at temperatures lower than the growth optimum of 37 °C. Two such studies reported enhanced ethanol production at incubation temperatures of 25 °C for *Clostridium carboxidivorans*, with the caveat that experimental groups were incubated at 37 °C for 24 h then transferred to 25 °C for the rest of cultivation<sup>240, 241</sup>. Shen *et al.* showed that at 25 °C, higher ethanol titers were achieved with even higher overall alcohol titers (ethanol, butanol, and hexanol).<sup>241</sup> Compared to the control group, the highest alcohol titer at 25 °C was 6.97 g/L – a 70% increase in alcohol production. However, these studies also reported that lower temperatures increased the lag phase of organism growth, indicating a slowdown of overall metabolism, or at least requiring a greater initial investment.<sup>240, 242, 243</sup> Interestingly, somewhat sub-optimal fermentation temperatures at 25 °C and 32 °C sometimes favor higher cell densities compared to higher temperatures.<sup>239, 243, 244</sup> Moreover, two studies showed that cell densities are almost doubled when grown at 25 °C relative to 37 °C (Max OD<sub>600</sub>

of 1.2 and 3.12 at 25 °C versus 0.55 and 1.66 at 37 °C respectively), even if the maximal growth rate is slower.<sup>240, 243</sup> These results are likely related to the “acid crash” phenomenon where fast accumulation of organic acids becomes detrimental to solvent production, which results in very low solvent titers.<sup>245</sup> The acid crash holds true for gas fermenting carboxydophilic bacteria<sup>246, 247</sup> and has also been reported in acetogenic *Clostridium ljungdahlii*.<sup>238</sup> Prevention can be achieved with pH control or by lowering fermentation temperature to slow the metabolic rate and/or change fermentation product profile.<sup>242</sup> Thus, while the lower temperature gave greater solvent yields, this was achieved over a longer period of growth relative to the optimum temperature of 37 °C.<sup>242</sup> In terms of overall fermentation productivity and/or stability, “slow and steady” may win the race.

Operating factors for gas fermentation must be controlled to lead to the highest ethanol productivity and optimized to improve stability and growth. Its relevancy in key cell processes like growth, energy maintenance, metabolite production, and membrane fluidity during fermentation warrant further investigation. To this end, we established *C. autoethanogenum* cultures at steady state conditions for two different temperatures and performed a detailed comparative analysis of the fermentation profiles, proteomes, metabolomes, and lipidomes, that resulted in informative and unexpected insights into its metabolism.

### **Materials and Methods:**

#### *Clostridium autoethanogenum* fermentations, sampling, and product determination

*C. autoethanogenum* strain DSM 10061 was obtained from the DSMZ culture collection and grown essentially as described previously.<sup>233</sup> Briefly, the strain was grown in continuous stirred tank reactors (CSTRs) via gas fermentation using a gas blend comprised of 50% CO, 30% CO<sub>2</sub>, 10% H<sub>2</sub>, and 10% N<sub>2</sub> at either 30°C or 40 °C, at pH 5.0. The medium used has also been described previously,<sup>248</sup> with 0.5 g/L yeast extract added and fed into the CSTRs at a rate of 1.0 d<sup>-1</sup>. Microbial biomass concentrations, metabolites analysis and CSTR headspace gas composition methods have been described previously.<sup>233</sup> Biomass concentrations were measured using a spectrophotometer (OD<sub>600nm</sub>), and liquid fermentation products (acetate, ethanol, lactic acid, 2,3-butanediol) were determined by using an Agilent 1200 series HPLC system, and gas composition by gas chromatography using a 490 Micro GC (Agilent). For omics analysis, sampling was conducted at steady state time points with cells harvested by centrifugation, snap frozen using liquid nitrogen and then stored at -80°C. Replication was performed via triplicate

sampling from a single reactor. This lack of reactor replication was a result of the expense and difficulty in performing these reactor cultures as well as the high stability of cultures derived from CSTRs.

#### *GC-MS based metabolomics measurements, data normalization, and analysis*

Cell pellet samples representing 3 reactor-sample replicates from both temperatures were analyzed by GC-MS metabolomics. Dried lysed samples were dissolved in silylation-grade acetonitrile and silylated by addition of N-methyl-N-(trimethylsilyl)trifluoroacetamide (MSTFA) with 1% trimethylchlorosilane (TMCS) (Thermo Scientific, Bellefonte, PA) and heated for 1 h at 70 °C to generate trimethylsilyl derivatives. For analysis, 1  $\mu$ L aliquots were injected into an Agilent Technologies 7890A GC coupled to a 5975C inert XL MS fitted with an RTX-5MS (5% diphenyl/95% dimethyl polysiloxane) 30 m  $\times$  250  $\mu$ m  $\times$  0.25  $\mu$ m film thickness capillary column with a 5 m Integra-Guard column.

A large user-created database and the commercially available Wiley Registry 10th Edition combined with the NIST 14 mass spectral database were used to identify metabolites of interest. Peaks were quantified by using extracted-ion chromatograms (EIC) rather than total ion current chromatograms, utilizing a characteristic fragment  $m/z$  for each ion, to minimize co-eluting metabolites. The EIC was scaled back to TIC using predetermined scaling factors and quantification was based on area integration and normalized to the quantity of internal standard recovered, the volume of sample processed, the derivatization volume and injection volume.

Student's t-tests (two-tailed assuming equal variances) were performed to identify metabolites that had a significant change in abundance. Metabolites with a t-test p-value  $\leq 0.05$  and an absolute fold change of  $\geq 1.4X$  were classified as significantly differential metabolites. The metabolites were classified into broad categories after manual curation based on the chemical groups present. Data Figures were generated using R packages. For metabolites with no detection, values were imputed with a constant 0.01 to represent the limit of detection for statistical analysis.

#### *Proteomics sample preparation and measurement by reversed phase chromatography and tandem mass spectrometry*

*C. autoethanogenum* cell pellets (in triplicates) from the two temperatures collected at steady-state were frozen and stored at -80 °C until proteomic analysis. Cell pellets were brought up in Tris-HCl (100 mM at pH 8.0) buffer and subjected to bead beating for cell lysis. The lysate

was adjusted to 4% SDS (sodium dodecyl sulfate), heated at 95 °C for 10 min to further denature proteins and centrifuged at 20,000 g for 5 min to separate cell debris. An aliquot of the supernatant was taken for quantification of crude protein amount in the lysate using a NanoDrop One Microvolume UV-Vis spectrophotometer (Thermo Scientific) by measuring corrected absorbance at 205 nm. The lysate was next adjusted to 10 mM DTT (DL-Dithiothreitol), followed by heat treatment at 95 °C for 10 min to denature the proteins and reduce the disulfide linkages. Samples were then adjusted to 30 mM IAA (iodoacetamide) and incubated for 20 min in darkness at room temperature to alkylate cysteine residues. Crude protein was then cleaned up using the protein aggregation capture method prior to digestion.<sup>249</sup> Briefly, 250 µg of magnetic beads (1 µm SpeedBead magnetic carboxylate modified particles; GE healthcare UK) were suspended in the same amount of crude protein lysate and aggregation of proteins onto the beads was induced by addition of acetonitrile to a final concentration of 70%, followed by a 20 min incubation. The protein-bead aggregate was then washed with 1 mL of 100% acetonitrile followed by 1 mL of 70% ethanol. Aggregated proteins were then digested with MS grade trypsin protease (1:75 w/w; Pierce-Thermo Scientific) in Tris-HCl (100 mM, pH 8.0) for 3 h at 37 °C, and again overnight at 37 °C. Tryptic peptides were then acidified with formic acid (FA; LC/MS grade) to a final concentration of 0.5% and filtered through a 10kDa MWCO centrifugal concentrator (Vivaspin500 PES; Sartorius) to remove under-digested proteins. The resulting peptides were again quantified by NanoDrop One.

Three micrograms of peptides from each sample were analyzed by reversed phase LC-MS/MS using a Vanquish UHPLC system (Thermo Scientific) coupled to an Orbitrap Q Exactive Plus mass spectrometer (Thermo Scientific). Briefly, peptides were trapped on an in-house made single frit trapping column (100 µm ID) packed with a C18 resin (5 µm Kinetex; Phenomenex) and separated by an organic gradient on an in-house pulled nanospray emitter (75 µm ID) packed with C18 resin (1.7 µm Kinetex; Phenomenex). Sample loading, trapping, and desalting was performed for 30 min in solvent A (0.1% FA in 5% acetonitrile) at 2 µL/min. For peptide elution and analytical separation, the flow rate was split to achieve ~300 nL/min. The gradient was as follows: 0–30% solvent B (0.1% FA in 70% ACN) over 185 min, to 0% solvent B over 5 min, and hold for 15 min at 100% solvent A. Each elution was followed by a column wash: 0 to 100% solvent B over 15 min, hold at 100% solvent B for 5 min, to 0% solvent B over 5 min and final equilibration with 100% solvent A for 15 min before injection of the next

sample. Total analysis time was 275 min for each peptide sample, with spectra collected during the first 240 min. Eluting peptides were measured and sequenced by the Orbitrap Q Exactive Plus mass spectrometer (operated via Xcalibur, Thermo Scientific) in data-dependent mode. Full-scan MS spectra were acquired from  $m/z$  300–1500 at a resolution of 70,000 (FWHM) with an AGC target value of 1e6. For fragmentation, the 20 most intense precursor ions were selected for MS/MS with an isolation window set to 1.8  $m/z$  and dynamic exclusion set to 30 s. HCD fragmentation was performed at normalized collision energy of 27% with an AGC target value of 1e5 and a resolution of 17,500 (FWHM).

#### *Proteomics data analysis*

For searching of peptide spectra, a custom protein database for *C. autoethanogenum* database was constructed by combining information from different sources. *C. autoethanogenum* DSM10061 protein sequences generated by Brown et al<sup>250</sup> and Humphreys et al<sup>251</sup> were downloaded from UniProt, UniParc, and RefSeq respectively. Sequences from all these sources had unique protein sequences in the files. These were first clustered at 100% using CD-HIT whereby the longer sequences were preferred for otherwise identical sequences to include possibility of sequences that could be truncated due to premature stop codons.<sup>252</sup> The resulting database contained 4,257 sequences (3,892 from the UniProt reference database, 170 from UniParc database, and 195 from RefSeq database) that could not be clustered further at 100% sequence identity.

The resulting peptide fragmentation spectra were searched against this database combined with a common protein contaminants database via a target decoy approach using the MS Amanda algorithm (v2.0) integrated in Proteome Discoverer software (version 2.3.0.523, Thermo Scientific). Peptide Spectrum matches (PSMs) were required to be at least 5 amino acids long, fully tryptic with a maximum of 2 missed cleavages, with a static modification of 57.0214 Da on cysteine residues (carbamidomethylated), a dynamic modification of 15.9949 Da on methionine residues (oxidized), and a dynamic modification of 42.0106 Da on lysine residues (acetylated). PSMs and peptides were scored and filtered at a false discovery rate (FDR) of 1% using the IMP-Elutator node in Proteome Discoverer. Peptides were quantified by chromatographic area-under-the-curve. Match between runs was enabled by performing a grouped consensus step in Proteome Discoverer on all the samples. Peptides were then mapped to their respective proteins, and their areas summed to estimate protein-level abundances.

Proteins were filtered to retain those with protein FDR of  $\leq 1\%$ , abundances log2 transformed, and distributions LOESS normalized and median centered in InfernoRDN.<sup>253</sup> The proteome dataset was further filtered to retain only those proteins that were quantified in at least 3 samples and missing values were imputed in Perseus to simulate the MS limit of detection.<sup>254</sup> Since the log-transformation was applied before calculating averages and performing statistical tests, the proteomic fold-changes are representative of differences in the geometric means of the data, and not necessarily their arithmetic means. Student's t-test (two-tailed assuming equal variances) with Benjamini Hochberg FDR correction was performed to identify proteins that had a significant change in abundance. Proteins with a q-value (adjusted t-test p-value)  $\leq 0.05$  were classified as significantly differential abundant proteins. Annotations were downloaded from UniProtKB, UniParc, and Kyoto Encyclopedia of Genes and Genomes (KEGG) using BlastKOALA.<sup>255</sup> For identifications of proteins mapping to sequences from UniParc and Refseq databases, the annotation (locus tags, accessions, and other description) were transferred from corresponding UniProt entries either based on 100% sequence clustering or based on high sequence homology ( $<100\%$ ) if present. Pathway information from KEGG and BioCyc pathways was utilized for mapping proteins to pathways. ClusterProfiler was used for identifying pathways that are overrepresented at the two temperatures.<sup>256</sup> Peptide abundances were used for examining the percentage acetylation of proteins. Plots were generated using Excel, R, and Python packages.

As complete lipid pathway information is not available for non-model organism such as *C. autoethanogenum*, we combined information from KEGG, BioCyc, and literature to construct the potential pathways for lipid synthesis in *C. autoethanogenum*. Using domain searches in NCBI and InterProScan, we also identified a potential enzyme for synthesis of plasmalogens in this organism.

#### *MTBE-based extraction of C. autoethanogenum lipids*

General use solvents methanol (MeOH), water (H<sub>2</sub>O), and acetonitrile (ACN), all LC-MS grade, were purchased from MerckMillipore (OmniSolv, Burlington, MA, USA). The ionization agent ammonium acetate (NH<sub>4</sub>Ac) and extraction solvent methyl *tert*-butyl ether (MTBE) were purchased from Sigma-Aldrich (St. Louis, MO, USA).

Cell pellets representing triplicate were stored at  $-80^{\circ}\text{C}$  prior to analysis. Before analysis, cells were washed with 1X phosphate-buffered saline (PBS) to remove any residual media before

being stored in 1.5 mL Eppendorf centrifuge tubes. A modified version of the Matyash MTBE method was used for lipid extraction.<sup>62</sup> Briefly, 150  $\mu$ L of MeOH were added to each sample, and each sample was vortexed for approximately 10 seconds to ensure cell dispersion. 500  $\mu$ L of MTBE was then added to each sample. Sonication was used for lysis – cells were pulsed for 10 seconds at 20% amplitude, followed by 10 seconds of rest, for 5 minutes while suspended in a room temperature water bath. After cell lysis, samples were lightly shaken on ice for 1 hour. To induce phase separation, 125  $\mu$ L of H<sub>2</sub>O was added to each sample followed by centrifugation for 15 minutes at 4 °C at 20,000 x g. The upper MTBE organic phase containing the nonpolar lipids was removed and dried, concentrating each 500  $\mu$ L organic partition to 100  $\mu$ L using a nitrogen air stream.

#### *Measurement of lipids by Reversed Phase-based liquid chromatography mass spectrometry*

Extracted samples were taken from the -20 °C freezer and transferred to autosampler vials. These were analyzed with an Ultimate 3000 liquid chromatograph (LC) coupled to an LTQ Velos Pro Orbitrap mass spectrometer (MS). The LC was operated using split-flow nano-chromatography with a reversed phase method adapted from a previous study.<sup>58</sup> Briefly, a 60% ACN 40% water mobile phase A with a 90% IPA 10% ACN mobile phase B, both containing ~5 mM ammonium acetate was employed. The gradient started at 1% B and was held for three minutes. Over the next 1.5 minutes, B was increased linearly to 30%. B was then ramped more slowly to 35% over 5 minutes, before increasing to 55% over 6 minutes. Over the next 35 minutes, B was increased to 70%. B was then increased to 99% in ten minutes followed by a hold for 17 minutes. This was then decreased to the starting conditions (1% B) over two minutes and the column was equilibrated for 9.5 minutes. The flow rate was set at 0.05 mL/min and the waste line provided enough backpressure such that the LC was operating at around 400 bar at 99% B. A 100  $\mu$ m inner diameter column was used packed with 1.7  $\mu$ m C18 packing material to 12 cm. The mass spectrometer was operated in negative ion mode at 2.5 kV with a top 10 DDA method where one full scan was acquired at 30k resolution with the orbitrap followed by 10 data dependent CID fragmentation scans in the ion trap at 30 normalized collision energy.

#### *Lipidomics data analysis*

For analysis of the lipidomics results, the generated LC-MS/MS files were processed using MS-DIAL<sup>48, 49</sup> and searched against a built-in lipid fragmentation database with 0.01 Da MS<sup>1</sup> tolerance and 0.3 Da MS<sup>2</sup> tolerance. Lipids identified by MS-DIAL at the MS<sup>2</sup> level where

the maximum sample was at least 5× greater signal intensity than the blank were exported for further analysis. These were manually interrogated and confirmed lipids were integrated using El-MAVEN for quantitative analysis.<sup>94</sup> Manual validation of the lipid identifications involved examining the mass accuracy of the feature as well as the MS<sup>2</sup> spectra that were available (chromatographic characteristics were also assessed). MS-DIAL annotations were generally accurate, however there were instances of misidentifications, for example, some lipids identified by MS-DIAL were found to be isotopes of other species, while certain oxidized ether phospholipids were mislabeled as isomeric standard phospholipids. A Welch T-test was performed in R on the raw intensity values (1.1, a negligibly small number, was added to the intensities to make log-transformation possible on any values with zeros). A Benjamini-Hochberg FDR control was applied to these p-values for significance calculation. Lipid fold-changes were calculated by taking the difference of the log<sub>2</sub> of the averages of each condition. These therefore represent arithmetic average fold-changes.

### **Results and Discussion:**

*The gas fermenter C. autoethanogenum shifts from ethanol to acetate production at higher growth temperature.*

*C. autoethanogenum* was grown via gas fermentation in continuous stirred-tank reactors at either 30 °C or 40 °C to steady state ( $D = \sim 1.0 \text{ d}^{-1}$ ) and sampled for main fermentation products (see Materials & Methods) and mass spectrometric analyses for each respective condition. Measurement of liquid fermentation products revealed significant differences in the metabolite profiles at the two temperatures (**Figure 7-1**). At 40 °C, ethanol selectivity was lower compared to 30 °C (56% vs. ~73%) while carbon was diverted towards acetate (27% vs. 9% at 30 °C) and somewhat toward 2,3-butanediol (2,3-BDO) (~9% vs. 7.5% at 30 °C). There was ~2.3 to 9.5% of carbon unaccounted for during steady state selectivity calculations. This is consistent with other studies on other Clostridia species that have shown an enhancement of alcohol fermentation at lower temperatures. The contribution to biomass was increased at 30°C relative to 40°C (9.9-10.4% vs. 7.5% respectively).

These results clearly indicate a remodeling of the metabolism of *C. autoethanogenum* between the two temperatures. Given the higher proportion of carbon devoted to biomass at 30 °C, it is tempting to speculate that *C. autoethanogenum* is better fit for growth at 30 °C than 40 °C, and that the less efficient growth at 40 °C may indicate the effects of temperature stress on *C.*

*autoethanogenum*. Indeed, most microbes experience a steeper decline in growth rate when exceeding rather than incedding the optimal growth temperature.<sup>257</sup> It is important to note that the growth rate must be the same. This is because *C. autoethanogenum* was cultivated in a chemostat with a defined dilution rate and the cell division rate must therefore be equal. That said, the steady-state biomass could be different as the same dilution could support different cell counts if energy efficiency is different. Whatever the reason, the temperature difference results in changes to the fermentation product profiles for *C. autoethanogenum*. At 30 °C, this results in increased ethanol production whereas at 40 °C, the bacterium produces more acetate and 2,3-BDO. A more detailed molecular understanding of these metabolic shifts may provide insights into the mechanisms driving enhanced solvent production at lower temperatures or greater acetate production at higher temperatures, both of which provide valuable leads for metabolic engineering of these microbes towards improved tolerance and/or desirable fermentation as well as development of newer models and biochemicals. To this end, *C. autoethanogenum* cells from the two temperatures were further characterized using a multi-omics approach to gain a molecular understanding of these observations.

*Proteomics and metabolomics reveal temperature-induced metabolic changes in C. autoethanogenum*

Between the two temperatures, peptides were detected that correspond to 2,416 proteins in *C. autoethanogenum*. These were filtered to include only proteins that had quantifiable values in at least three measurements. In total, 1,795 unique protein groups (essentially unique proteins) were quantified, representing ~43% of the predicted protein coding genes in *C. autoethanogenum*. Even with a 10 °C difference in growth temperature, a comparable number of proteins were quantified at both temperatures, with 1,757 (~98%) quantified across both temperatures (**Figure 7-2a**). Similar observations were obtained with metabolomics, which identified 80 metabolites in total, nearly all (97.5%) of which were detected in both temperatures (**Figure 7-2b**). Despite the overlap in proteome coverage/identification, global pairwise correlation of protein abundance revealed that only ~85% of the proteome abundance correlated between the two temperatures (**Figure 7-3**).

For examining the proteins whose abundances were changing significantly, we conducted a two-tailed student's T test for all the quantified proteins followed by Benjamini-Hochberg FDR correction for multiple testing. The analysis revealed that 1,161 proteins (~65% of quantified

proteins) were significantly differentially abundant between the two temperatures (adjusted p-value of  $\leq 0.05$ ) with 611 exhibiting a log2 fold-change (absolute value)  $> 1$  (**Figure 7-2c**).

These data corroborate the observed abundance correlation between and across temperatures and suggest temperature had a dramatic impact on the proteome. For identifying metabolites with significant abundance changes, two-tailed student's t-tests were conducted as described in Methods. Similarly, 51 metabolites representing ~64% of total quantified metabolites were differentially abundant (p-value  $\leq 0.05$ , absolute fold change  $> 1.4X$ ) between the temperatures. As the quantified proteome and metabolome both had significant perturbations, we adopted a focused strategy to highlight the major metabolic pathways that were temperature responsive.

#### *Examination of enzymes in the central metabolism pathways of C. autoethanogenum*

In *C. autoethanogenum*, ethanol and other end-products are produced after generation of acetyl-CoA through the fixation of carbon in the Wood-Ljungdahl (WL) pathway (**Figure 7-4**). To elucidate the proteomic determinants of the observed shifts in product formation, the central metabolism of *C. autoethanogenum* was visualized using Brown et al.<sup>250</sup> as a reference and the proteomic abundances were mapped onto it (**Figure 7-4a**).

All proteins mapping to the WL pathway were quantified, with many being differentially abundant (**Figure 7-4a**). Most enzymes in the reductive/methyl branch converting CO to acetyl-CoA increased in intensity at 40 °C and those converting formate to methyl exhibited on average ~3× increases. Interestingly, energy generating hydrogenases in WL-oxidative branch were not altered significantly (p  $\leq 0.05$ , fold-change  $> |1 \log_2 \text{ unit}|$  or 2×). Of the two additional monofunctional carbon monoxide dehydrogenases (CAETHG\_3005 and CAETHG\_3899) encoded by *C. autoethanogenum* to utilize CO and CO<sub>2</sub>, only CAETHG\_3005 was quantified. This was not found to be differentially abundant due to temperature. This could possibly be explained by *C. autoethanogenum* having more efficient growth at 30 °C (consistent with its greater biomass carbon fraction), thereby requiring less proteome allocation for carbon fixation per cell than at 40 °C. Interestingly there were 32 significantly changing proteins mapping to the KEGG ribosome pathway (03010), all of which were decreased in abundance at 40°C (see **Figure 7-5**). While increased ribosome mass fractions are usually associated faster growth, since the cell division rate is equal, that cannot be the case. Possibly *C. autoethanogenum* at 30 °C and 40 °C have similar absolute ribosome abundances, but at 30 °C there is less overall protein mass which gives rise to a compositional increase of ribosomes (possibly because less extraneous

proteins like chaperones are needed). Importantly, the observed fold-changes are after normalization of the samples (and injection of a constant peptide mass) and should therefore be interpreted compositionally as an allotment of the overall proteome.

After generation of acetyl-CoA, the enzymes for the different fermentation pathways (acetate, ethanol, lactate, and 2,3-BDO) were also quantified. In line with the observed liquid product selectivity (**Figure 7-1**), enzymes significantly decreased at 40 °C for both routes of ethanol production - directly via bifunctional aldehyde/alcohol dehydrogenases (~38×) or via acetate using aldehyde:ferredoxin oxidoreductases (~4× and ~6×, respectively). However, even with the observed increase in acetate levels at 40°C, abundances of only one of the two enzymes (CAETHG\_3358, Pta) converting acetyl-CoA to acetate marginally increased (~1.2×), (not “significant” based on the foldchange cutoff). Pathways for the production of pyruvate-derived fermentation products (lactate and 2,3-butandiol) show less clear results. While the abundance of Pyruvate:ferredoxin oxidoreductase (PFOR) (enzyme catalyzing acetyl-CoA to pyruvate conversion) did not change, both pyruvate utilizing branches saw “significant” increases at 40 °C with log2 fold changes of 1.6 and 1.4 for LdhA and BudA respectively. That said, the enzyme catalyzing the reversible conversion of acetoin to 2,3-BDO (following after BudA) decreased in abundance even though a relative increase in 2,3-BDO production was observed at 40 °C. The lack of clarity in proteomic abundances may mean other regulatory elements are operating on this system. Possibly temperature or allosteric regulation is affecting the reaction kinetics for the observed enzymes.

Considering the abundance stability of both Pta and PFOR across temperatures, these enzymes may be constitutively expressed at relatively constant levels sufficient for carbon flow as required central metabolism. If true, the flow of carbon would likely be dictated by post-translation regulation. A recent study in the closely related species *C. ljundahlii* has identified that acetylation plays a key role in controlling enzymatic activities.<sup>258</sup> Because of this, we also examined if changes in lysine acetylation patterns occurred with temperature variation. While no acetylation enrichment was performed, 447 of the 12,177 quantified peptides were found to be acetylated. These map to 224 proteins, many of which belong to enzymes in central metabolism (**Figure 7-4b**). Levels of lysine acetylation, in terms of total abundance within a sample, range from 0.17% to 0.53% and are considerably lower than the observations reported in *C. ljundahlii*.<sup>258</sup> Nonetheless, significant increases in the abundance of acetylation occurred at the

higher temperature with the overall percent of acetylated peptides increasing by 2.9× (40°C vs. 30°C, p value = 0.00001). Multiple enzymes in the WL pathway also exhibited increases in lysine acetylation at 40°C (**Figure 7-4b**). However, limited or sporadic increases in percent acetylation were observed in the proteins comprising hydrogenases and energy conservation in *C. autoethanogenum* (**Figure 7-4b**). A perhaps important observation was the increase in acetylation of enzymes fermenting acetyl-CoA to acetate (Pta and Ack; CAETHG\_3358 and 3359) even though they did not change substantially in total abundance. Similar observations can be made for the enzyme that interconverts acetyl-CoA to pyruvate (PFOR; CAETHG\_3029) which was also highly abundant but not significantly changing across temperatures. The enzyme responsible for the conversion of acetoin to 2,3-BDO (2,3-BDH; CAETHG\_0385) was decreased in abundance at 40°C even when 2,3-BDO production was higher. These modifications certainly could alter enzymatic activities and may help explain the lack of obvious patterns in protein abundances described above. A caveat to this is that acetylation percents were calculated from all observed acetylated peptides for a protein. A site-specific analysis was not performed (partly because of difficulties in precise localization of acetylation). Thus, if one acetylation site increases and another decreases, this effect could be blurred out. That said, changes in acetylation likely correspond to some change in proteoforms and can therefore be a useful metric to identify physiological changes.

#### *Accumulation of fatty acids and their glycerol conjugated forms occurs at 40 °C*

The central metabolic pathways described above constitute a small portion of the differentially abundant proteins observed in this study. To get a more complete picture of the metabolic adaptation to temperature, we assessed the larger set of differentially abundant proteins and metabolites. The compounds measured by GC-MS that were significantly altered with temperature belonged to four major categories: amino acids, organic acids, alcohols, and fatty acid related metabolites (**Figure 7-6a**). Additionally, 23 metabolites could not be confidently identified and were classified as unknowns, many of which significantly decreased at 40 °C (**Figure 7-6a**). Among the identified metabolites, most organic acids (e.g., succinic, malic, citramalic, 2-hydroxyglutaric acids) and amino acids (glycine, valine, beta-alanine, 5-oxo-proline, and N-acetyl-glutamic acid) significantly decreased in abundance at 40 °C. Corroborating these observations, proteomic analysis revealed that the related enzymes for some of these organic and amino acids also decreased significantly at 40 °C, suggesting that carbon is

diverted away from these amino acids and carboxylic acids when *C. autoethanogenum* is grown at the higher temperature. While the proteomics data are biomass normalized (an equivalent mass of peptides for each sample were analyzed), this was not true for the metabolomics samples. Because of this, some of these decreasing metabolites could represent changes in biomass, and not necessarily a per-cell decrease in the metabolite concentration.

In contrast, growth at 40 °C coincided with the increased abundance of meso-2,3-butanediol, RR/SS-2,3-butanediol (**Figure 7-6a**). The observed increase in the intracellular levels of 2,3-butanediol confirms the bioreactor observations described above (**Figure 7-1**) and the understanding that *C. autoethanogenum* preferentially produces these at the higher temperature. Additionally, several fatty-acid related metabolites were observed to increase significantly at 40 °C (**Figure 7-6a**). Closer inspection revealed that these consisted of fatty acids with even-numbered acyl chains (lauric acid – FA 12:0, myristic acid – FA 14:0) and their monoglycerides (glycerol-conjugated fatty acids) in saturated (1-lauroyl-glycerol, 1-myristoyl-glycerol) and mono-unsaturated (1-palmitoleoyl-glycerol, 1-myristoleoyl-glycerol) forms (**Figure 7-6b**). Interestingly, the most dramatic increase was observed for monoglycerides 1-myristoyl-glycerol (~76×), 1-myristoleoyl-glycerol (~26×), and 1-palmitoleoyl-glycerol (38×), which appeared to accumulate at 40 °C (**Figure 7-6b**).

To gain further insights into the production of these fatty acid derivatives, we next examined the fatty-acid biosynthesis pathway. As shown in **Figure 7-6c**, enzymes for almost all reactions in the fatty acid biosynthesis (FAS II) pathway were quantified. There are several enzymes that are capable of transforming acetyl-CoA to malonyl-CoA (step one of FAS II), all of which show greater intensity at 40 °C than 30 °C, although only one passes the fold-change threshold. Most of the other enzymes show marginal intensity differences, except the dehydrase FabZ (which converts hydroxyacyl-ACP to enoyl-ACP) and a thioesterase responsible for converting acyl-ACP to free fatty acid (CAETHG\_2058), which both had log2 fold-changes >2 (See **Figure 7-6C**). This could plausibly be consistent with a hypothesis that *C. autoethanogenum* upregulates its fatty acid synthesis, with initiation, dehydration, and release from acyl carrier protein possibly being the most important/rate limiting steps. Initiation could be especially important to shunt carbon to fatty acid synthesis as opposed to the myriad other uses of acetyl-CoA.

FAS II pathway enzyme abundance is one of two critical determinates of fatty acid/monoacylglycerol levels. Lipid metabolism is also intricately linked to fatty acid levels. Fatty acids are integral components of bacterial membranes, which adapt to environmental factors (temperature, pH, osmolarity, solvent stress, etc.) to maintain homeostasis, fluidity, and thus functionality through directed compositional changes. It is worth noting that while more carbon was shuttled to solvent (ethanol and butane-2,3-diol) production at 30 °C, *C. autoethanogenum* at 40 °C has the additional membrane destabilizing stress of higher temperature, which may have a synergistic effect with solvent stress. Since there was a significant increase in the thioesterase converting acyl-ACP to fatty acids, we next compared the lipidomes and related pathways in *C. autoethanogenum* at the two temperatures.

*Integrated analysis of lipidome and proteome uncovers membrane remodeling with potential glycolipid production at 40 °C*

To investigate changes in membrane composition which may link to the accumulated fatty acid/monoglyceride metabolites identified above, we performed a LC-MS/MS-based lipidomic analysis. Combining MS-DIAL's automatic lipid identification with manual curation, 108 different lipid species were identified and quantified. These species consist primarily of phosphatidylglycerol (PG) and phosphatidylethanolamine (PE) lipids as well as lyso-PG (LPG), lyso-PE (LPE) and cardiolipin (CL) (**Fig 5a**). Over half of these identifications (52%) were ether-linked phospholipids. While the lipid profile may not be totally comprehensive (limited to automated identifications that were manually curated), these data indicate that PE and PG lipids are the primary membrane lipids with a majority containing ether-linkages. To our knowledge, this is the first reported lipidomic characterization of *C. autoethanogenum*, and these results are consistent with reported lipid compositions of other Clostridial species.<sup>17</sup> The collision-induced dissociation MS<sup>2</sup> data generated in this method is not capable of distinguishing alkyl and alkenyl (plasmalogen) ether linkages due to a lack of specific fragmentation ions. Most of these are suspected to be plasmalogens, which are commonly reported in *Clostridium* species; however, there were two saturated ether lipids measured, which must necessarily be the less common alkyl ethers.<sup>259</sup>

Of the 108 quantified lipid species, 51 were differentially abundant between temperatures based on an FDR of less than 0.05 (Benjamini-Hochburg FDR derived from a two-tailed Welch's T-test) (**Figure 7-7a**). The majority of these differentially abundant lipids decreased at

40 °C and only a small subset increased (43 decreased, 8 increased). It may be that increases in select key lipid species and not in the overall phospholipid content was needed to adapt the membrane/lipid bilayer to the higher temperature. However, the asymmetry of the observation indicates that there was indeed greater biomass of 30 °C than 40 °C. Intriguingly, and quite surprisingly, lipids significantly decreasing at 40 °C generally had longer acyl chains than non-significantly changing lipids. Indeed, plotting fold change vs carbon number (for the PE and PG lipid classes) shows a clear trend where longer acyls chains have lower proportional abundance in the 40 °C samples, even for those lipids which did not reach the significance threshold (**Figure 7-7b**). Additionally, lysophospholipids (ether-LPG, LPE, and one ether-LPE) were the only lipids significantly increasing at 40 °C. This runs counter the typical homeoviscous expectations of longer-chain lipids being required at higher temperatures to maintain adequate membrane viscosity. Perhaps this indicates that the increased solvent production at 30 °C is more membrane destabilizing than the high temperature at 40 °C and the lower temperature *C. autoethanogenum* has the greater need for homeoviscous adaption with longer chains bilayer lipids.

Proteomic analysis of the lipid metabolism pathways was conducted (**Figure 7-7c**). As lipid synthesis pathways are not well defined for *C. autoethanogenum*, information from KEGG<sup>260</sup> BioCyc,<sup>261</sup> and previous studies<sup>262-264</sup> was combined to construct the pathways and identify relevant enzymes (**Figure 7-7c**). This was especially necessary for ether-phospholipids, which were found to be highly abundant in *C. autoethanogenum* but have no annotated representatives for enzymes forming them. Anaerobic plasmalogen producing enzymes have recently been identified in *C. perfringens*<sup>263</sup> and we utilized the information about the domain architecture to identify the probable homolog(s) in *C. autoethanogenum* (CAETHG\_1442, see **Figure 7-8**).

Multiple proteins could be annotated in the lipid synthesis pathways (**Figure 7-7c**); albeit enzymes directing production towards phospholipids with multiple acyl chains (PE, PC, PS, PG, cardiolipin) were either not detected, had no change in abundance, or decreased significantly at 40 °C. Remarkably, at 40 °C the enzyme PlsX (CAETHG\_3362), which has been implicated in coupling/controlling fatty acid-lipid metabolism<sup>265</sup> significantly decreased in abundance (log2 fold-change of -3.34). This enzyme is responsible for producing acylphosphate from acyl-ACP,

which is subsequently turned into phosphatidic acid (a precursor phospholipid). How this specific result fits into the overall lipid physiology of *C. autoethanogenum* is currently unclear.

Based on enzymatic abundances, it appears that conversion of lysophosphatidic acid (monoacyl) to phosphatidic acid (diacyl) increased at 40 °C (PlsC/D, CAETHG\_1773). Interestingly, a significant increase in the abundance of phospholipase C (CAETHG\_0300, ~9-fold) was also observed, which hydrolyzes diacyl phospholipids to diacylglycerols (DAG), implying *C. autoethanogenum* has differential lipidome regulation between the two temperatures. Furthermore, the abundance of processive glycosyltransferase (CAETHG\_2321, GT family 28) significantly increased at 40 °C (~33-fold). This enzyme produces glycolipids (glucosyl-DAG/diglucosyl-DAG) which leads to the hypothesis that *C. autoethanogenum* restructures its membrane at higher temperatures to include more glycolipids and less phospholipids. Although glycolipids were not observed with our measurement approach, these enzymatic abundances, combined with the bulk decrease in phospholipid abundance observed at 40 °C, provide evidence for the formation of glycolipids at the higher temperature. Depending on the exact species, these glycolipids could be providing the expected homeoviscous temperature resistance to ensure membrane integrity at 40 °C. Interestingly, our analysis also revealed that many other glycosyltransferases are present in this organism (48 distinct proteins), 21 of which were expressed and 15 are temperature responsive (**Figure 7-9**). These may play a role in production of additional glycolipids, some of which are widespread in Clostridia.<sup>259</sup>

In summary, it appears *C. autoethanogenum* adapts its lipidome to higher temperatures by reducing phospholipids in favor of glycolipids. Across the lipidome, reductions in phospholipid abundances were observed at 40 °C. Consistent with this proteomics analyses showed that at 40 °C, there were large increases in enzymes which convert diacylglycerols into glycolipids. We hypothesize that this is the compensatory mechanism by which *C. autoethanogenum* tolerates heat stressors and maintains membrane integrity.

#### *Temperature induces expression of bacterial microcompartments in C. autoethanogenum*

Along with the metabolomic and lipidomic guided assessments of the proteome as described above, we also conducted a global analysis of all the protein level changes. To this end, we performed a pathway overrepresentation analysis (**Figure 7-5**) as well as examined the proteins which exhibited the highest fold increases at both the temperatures (**Table 7-1**), many of which did not map to metabolic pathways and would otherwise have been ignored. Further

inspection revealed that many of the highly up/down regulated proteins grouped together based on their physical proximity in the genome. At both the temperatures, there was a separate close-knit group of such proteins or gene clusters that significantly and dramatically increased in abundance. At 30 °C, several top 15 proteins were present in the small genomic region spanning genes CAETHG\_3274-87 (**Table 7-1**). Similarly, multiple proteins with the highest fold increases at 40 °C belonged to a gene cluster formed by genes CAETHG\_1818-33. Intriguingly, both these sets of proteins related to separate Bacterial Microcompartments (BMC) (**Figure 7-10**). BMCs are proteinaceous, polyhedral “organelles” that encapsulate enzymes for specific reactions and are thought to be encoded by multiple established and candidate phyla. Direct examination of the *C. autoethanogenum* genome for BMCs has not been extensively reported thus far; however, *in silico* predicted information about the very similar organism (>98% sequence identity), *C. ljungdahlii*,<sup>266</sup> was recently made available.<sup>267, 268</sup> Given the substantial genetic identity, it is reasonable for *C. autoethanogenum* to harbor similar loci corresponding to BMCs. At 40 °C, multiple proteins from BMC cluster CAETHG\_3273-3290 (1,2-PD utilizing GRM3-type BMC) decreased whereas those from the larger BMC cluster CAETHG\_1810-1841 (GRM1-type BMC) increased in abundance (**Figure 7-10a-b**). While both correspond to Glycyl Radical enzyme Microcompartments (GRMs) (**Figure 7-10c-d**) that contain glycyl radical activated enzymes (GRE) as the signature pathway enzymes, limited information is available about the role of these apparently temperature responsive BMCs. There are 6 sub-groups of GRM based on function and these employ a mechanism different than the coenzyme B12 dependent pathway. Based on the signature enzymes in these two BMC loci, the one increasing in *C. autoethanogenum* at 30 °C appears to be ethanolamine metabolizing GRM (**Figure 7-10c**) whereas the BMC increased at 40 °C corresponds to choline metabolizing GRM 1 or 2, with choline trimethylamine lyase as the GRE (**Figure 7-10d**). For both these BMCs, almost all proteins in the respective cluster were identified barring select smaller shell proteins, which could be important for complete microcompartment formation and are sometimes found redundantly.

With respect to choline metabolizing BMCs (increased at 40 °C), which are generally upregulated, the presence of choline in this bacterial system (either intracellularly or from the media components) is limited at best. One possibility is that these BMCs are produced in response to sensing a host-like temperature (which, for gut microbes, would be associated with

extracellular choline) and does not directly respond to the presence or absence of choline. Alternatively, these BMCs could be related to energy or redox generation/recycling and be contributing to the increased acetate levels measured at 40°C. Additionally, it is plausible that these “super-operons” are regulated by specific proteins which are themselves temperature-responsive and whose expression turns on these BMCs. BMCs have been identified as important candidates for metabolic engineering, which has boosted research in that area. Our results here demonstrate that the commercially relevant gas fermenter, *C. autoethanogenum*, harbors BMC proteins that are actively and selectively expressed in response to environmental temperature. Further studies should be performed to confirm the assembly and activity of such microcompartments at the two temperatures.

### **Conclusions:**

The gas fermenter, *C. autoethanogenum*, can grow over a range of temperatures from 20 °C to 44 °C but has not been studied beyond its optimal temperature. As *C. autoethanogenum* is a valuable chassis organism for industrial bioproduction, information about its temperature response is critical for improving fermentation processes. Here, we provide a detailed multi-omic characterization/comparison of *C. autoethanogenum* grown at 30 °C and 40 °C. This acetogen shows extensive physiological differences with respect to temperature at the proteomic, metabolomic, and lipidomic levels (>60% of proteins, metabolites, and lipids were differentially abundant between the temperatures). Of particular relevance for its industrial application, a shift in end-product profile from ethanol to acetate (and 2,3-butanediol) was observed for *C. autoethanogenum* when going from 30 °C to 40 °C. *C. autoethanogenum* also accumulates several fatty acid-related metabolites at the higher temperature which are suggested to be related to a lipidome shift from phospholipids to glycolipids, possibly to confer enhanced thermostability. Furthermore, two bacterial microcompartments purportedly constructed by *C. autoethanogenum* are highly temperature-responsive with each comprised of proteins that exhibit some of the highest fold increases at either of the two temperatures (or else are only detected in one condition). These findings expand our understanding of this organism’s physiology, provide data for hypothesis generation, and establish new targets for future bioengineering research.

## Chapter VIII - Conclusions and future perspectives

### Summary of Research Achievements:

#### *Lipidomics Research Needs*

Lipidomics, or the large-scale measurement of lipids, can be a powerful tool for biological research. Lipids serve various roles in biological systems including structural, energetic, and signaling. While lipid assays have been used for decades, it is only relatively recently that instrumental developments have enabled the measurement of individual lipids on an omics scale. As Sanger sequencing, CRISPR gene editing, microscopy, and X-ray crystallography can all attest, biological insights are driven by (and therefore limited by) developments in analytical technology. Analytical developments have opened the new field of lipidomics, but there are still many intrinsic limitations that need to be addressed to unlock the full potential for biological research. Low accuracy lipid identification, poorly validated statistical practices, as well as a variety of application-specific issues still remain.

My doctoral research has both addressed these limitations and provided novel biological insights through the application of lipidomics to a variety of systems. While insights from specific projects are themselves of value, especially the identification of PFASylated lipids, I am more excited about the developments I have achieved in the technical aspects of lipidomics. LipoCLEAN is a major step-change in the untargeted analysis of lipidomic data. We were successful in dropping the realized false discovery rate from 45% to only 7% while retaining good recall. While further work is needed to ensure its reliable application to arbitrary datasets, LipoCLEAN has already enabled me to perform truly untargeted high throughput lipidomic analysis of unstudied fungi. This analysis was not possible at the start of my doctoral research.

Statistical interpretation of lipidomics data varies widely with many opposing assumptions made about the structure and properties of lipidomics data. My lipidomics replicability experiment was designed to address this blind spot. I have performed the largest analysis of biological replicates of any lipidomics experiment. Rather than apply statistical tests and analyses with only anecdotal support, this now allows researchers to validate their statistical approaches on actual data before applying to biological experiments. This experiment still has its own limitations. A bacterium with a relatively simple lipidome was used, and a more complex lipidome would allow better estimates of statistical power and matrix effects. A lipidome covering a wider range of lipid classes would also allow for the assessment of the properties of

lipidome summary statistics. That said, this is a very important step forward for lipidomics data analysis and one I hope may in some small way help address the replication crisis observed across many biological disciplines.

While not as widely used, neutron scattering is still an important experimental platform for biological research. Neutron scattering of membrane-bound proteins has been limited by the inability to contrast-match lipids with D<sub>2</sub>O alone. Through extensive lipidomics measurements and computational modeling, I thoroughly characterized the *in vivo* deuteration of *E. coli* lipids from a variety of deuterated feedstocks and developed guidelines for the design of specifically deuterated phospholipids. These guidelines allow for the control of both the amount and the location of deuteration, enabling contrast matching of both native and engineered lipids for neutron scattering.

In all these studies, I have innovated the technical aspects of lipidomics analysis, to unlock novel robust experimental procedures. The full impact of my research is by no means limited to this dissertation. Rather, the impact will be realized in biological experiments to come that will be more rigorous for the work I've done here. Indeed, I have provided the capability here to perform experiments that were formerly impossible.

### **Instrumental Developments:**

#### *Ion Mobility*

Looking to the future of lipidomics, there is a relatively new technological development that could reshape the field. I believe advances in ion mobility could prove to be very impactful for lipidomics. Ion mobility is a gas-phase ion separation technique that is conceptually similar to mass spectrometry but instead of profiling an ion's mass-to-charge ratio, ion mobility instead operates on the size-to-charge ratio (more precisely, "size" is the rotationally-averaged collisional cross section or CCS). Recent improvements have taken ion mobility from a relatively obscure niche technique to a common addition to LC-MS experiments. The trapped ion mobility spectrometry (TIMS) platform from Bruker is a good example of this. This allows for the determination of CCS values to within ~1% error. While this may at first sound laughable compared to the ppm-level *m/z* accuracy achievable by mass spectrometers, these CCS values can nonetheless be rather useful. First, CCS is much more related to structure than *m/z* is. For example, CCS can be used to discriminate between isomers, even in cases when tandem MS cannot. That said, CCS is not truly orthogonal to *m/z*. In fact, they correlate strongly (see **Figure**

**8-1).** This makes sense since bigger things tend to weigh more. However, ion mobility does have different selectivity than MS and is sufficiently orthogonal to help identify unknowns. Additionally, CCS can be synergistic with  $m/z$ . Ion mobility is readily adaptable to use with mass spectrometry. Mass spectrometry already generates gas phase ions, and cutting-edge time of flight instruments can measure  $m/z$  in microsecond time scales. This makes the milliseconds of ion mobility separation ideal to fit between the mass spectrometer and the chromatography which is on the scale of seconds to minutes for many peaks. This allows chromatographic peaks to be well characterized by ion mobility scans and ion mobility scan to be well covered by mass spectrometry scans.

The TIMS platform can also boost the signal in tandem MS scans. This works because of its dual cell design shown in **Figure 8-2**. An entrance ion funnel can be used to trap and retain ions while an ion packet is being separated by the analyzer section of the TIMS cell. This allows better signal due to the ion concentration in the entrance funnel of the TIMS cell. This mode of operation is known as PASEF, which stands for parallel accumulation serial fragmentation, and allows nearly 100% of the ion beam to be utilized. Other unique ion mobility instruments that have potential for significant impact in the future include the cyclic ion mobility spectrometry from Waters, and the structures for lossless ion manipulation (SLIM) technology originally developed at Pacific Northwestern National Laboratory. Thermo orbitrap-type instruments have a longer scan time which does not allow for ion mobility at chromatographic time scales. However, Thermo recently released the tribrid Astral quadrupole-orbitrap-TOF instrument which may allow for future integration of ion mobility spectrometry.

#### *Improved Fragmentation Technology*

Tandem mass spectrometry has been a major element of my dissertation research, and without it almost nothing of what I've done would be possible. There are recent developments in tandem mass spectrometry that may have the potential to improve lipidomics capabilities even further. Electron transfer dissociation (ETD) is not particularly new, having been reported as early as 2005.<sup>269</sup> However, only recently has ETD been adopted by many instrument vendors and is now widely commercially available. Because collisional induced dissociation involves multiple ion-neutral collisions that gradually build the necessary internal energy for dissociation, fragment ions generated tend to be strongly thermodynamically driven. This results in only relatively low-energy bonds being fragmented (often at heteroatoms). These fragmentation

spectra can be quite informative, but in many cases they are limited in their discriminatory power when it comes to isomers. For lipids this can clearly be seen in that CID will fragment at fatty acid tails, but not along the tail. This allows the identification of individual fatty acids based on their carbon number and degree of unsaturation, but no information regarding the location or stereochemistry (E/Z) of double bonds. ETD can fragment carbon-carbon bonds and therefore has the capability to localize double bonds, distinguish double bond stereoisomers, and fatty acid *sn*-position.<sup>270</sup> This has the potential to significantly increase the richness of the data generated by lipidomics experiments and provide much more specificity to lipid identifications. To make full use of this, however, informatic development will be needed, both to accurately identify lipids with increased specificity and also to build models that will allow for biological interpretation of the finer granularity provided. Currently, the differences between PE (16:0/18:1(9Z)), PE (18:1(9Z)/16:0), and PE (18:1(7Z)/16:0), for example, are not well understood. Additionally, ETD techniques are still limited by relatively low fragmentation efficiency, which lowers the resulting MS<sup>2</sup> signal.

### **Experimental Rigour:**

These are some of the recent instrumental developments that are likely to provide significant improvements to the capabilities of lipidomics analyses. This will propagate in turn into biological research and provide deeper and more informative experiments. However, I believe that this is in some sense a distraction from the more critical issues plaguing the field. I agree with many others in the opinion that scientific research generally, and biology in particular, needs to undergo a significant amount of proverbial soul searching. Researchers are becoming increasingly aware of what has been described as a “reproducibility crisis,” with low rates of successful replication of reported studies and decreased trust in published papers.<sup>271-275</sup> In addition to papers with actually wrong results (or the recent discovery of AI-generated papers), I have seen many studies that are conceptually flawed to the point of irrelevancy without regard to the correctness of the data. There are likely many causes for this. For one, the incentive structure of modern science strongly favors low-risk incremental advances. Publishing is everything. Promotions and funding are dependent primarily on publication numbers, which leads to a focus on publishing quickly, rather than publishing well developed studies. While this is theoretically tempered by other metrics such as citation counts and journal “impact”, there seems to be a somewhat dubious link between impact (measured in citations) and the quality of the

research.<sup>276-279</sup> As one paper put it, “As so often happens with indicators of performance, the indicator has become the target...All but forgotten in the desperation to win the game is publication as a means of communicating research findings for the public benefit.”<sup>280</sup> Additionally, to publish in the majority of journals (especially “high quality” journals), “positive” results are needed, leading to a myopic focus on those results that could possibly indicate anything of vague interest. In Omics research, this is not difficult to achieve. Measuring enough variables will inevitably result in some results that can be coerced into supporting the researchers’ preconceptions, even without intentional “p-hacking” (see the “Garden of Forking Paths” article<sup>130</sup>). This leads to an absurd situation where less rigorous science is rewarded more than detailed and reliable studies.<sup>281</sup> Indeed, based on interactions I’ve had at conferences, it seems that even those who may want to spend time building the groundwork for reliable analyses feel the need to sustain high publication rates precludes the investment needed to establish reliable experimental and interpretive paradigms.

Making matters worse, there is a perceived need to be on the cutting edge of the most impactful science possible. While I share a desire to be involved in research that has a strong potential to impact everyday life, the emphasis of impact, and more perniciously, of novelty feeds into the focus on results rather than methods. As one Nobel laureate noted, “the goal of a paper seems to have shifted from validating specific conclusions to making the broadest possible assertions.”<sup>276</sup> The need for novelty is especially dangerous in the omics fields as replication of results is almost never attempted, let alone successful. Given the statistical challenges and interpretive difficulties inherent in descriptive studies<sup>130</sup>, the best paradigm would be to perform a speculative initial investigation with untargeted omics, generate a plausible hypothesis based on the data, and confirm that with a targeted analysis on new samples, possibly with an entirely different experimental technique. Science is supposed to be based on data replicated by multiple different groups, building support for an overall model/hypothesis. However, outside of the most “important” or controversial areas there is little to no impetus among researchers to perform replication experiments, and even less to fund such work. That said, in the cancer biology field, there have been more than a few replication studies. Attempts to perform such replication experiments have found that papers generally do not include sufficient experimental details to actually attempt replication.<sup>282</sup> Even after clarifying experiments, the rate of replication experiments generating results consistent with the original study is less than 50%.<sup>275, 282</sup> A more

detailed breakdown of replication rates are shown in **Table 8-1**. This underscores the scale of the problem.

I would argue that a reset is needed in biology. Rather than throwing more money into generating lots of questionable data, we should stop and critically assess our experimental techniques, interpretive paradigms, and end goals. Instead of performing increasing complex, yet critically flawed experiments, we need to invest into building reliable rigorous analyses, that are carefully interpreted with a minimum of bias. Without such a change in focus, biological research may broadly fall into irrelevance and undermine the public trust in scientific inquiry.

## References

- (1) p-Coumaric acid. *PubChem*. National Center for Biotechnology Information. <https://pubchem.ncbi.nlm.nih.gov/compound/637542> (accessed 2024/08/20).
- (2) Cheng, T.; Zhao, Y.; Li, X.; Lin, F.; Xu, Y.; Zhang, X.; Li, Y.; Wang, R.; Lai, L. Computation of Octanol–Water Partition Coefficients by Guiding an Additive Model with Knowledge. *Journal of Chemical Information and Modeling* **2007**, *47* (6), 2140-2148. DOI: 10.1021/ci700257y.
- (3) LysoPE(0:0/14:1(9Z)). *PubChem*. National Center for Biotechnology Information. <https://pubchem.ncbi.nlm.nih.gov/compound/53480920> (accessed 2024/08/20).
- (4) Ekroos, K. Lipidomics Perspective: From Molecular Lipidomics to Validated Clinical Diagnostics. In *Lipidomics*, 2012; pp 1-19.
- (5) Han, X.; Gross, R. W. The foundations and development of lipidomics. *Journal of Lipid Research* **2022**, *63* (2), 100164. DOI: <https://doi.org/10.1016/j.jlr.2021.100164>.
- (6) van 't Hag, L.; de Campo, L.; Garvey, C. J.; Feast, G. C.; Leung, A. E.; Yepuri, N. R.; Knott, R.; Greaves, T. L.; Tran, N.; Gras, S. L.; Drummond, C. J.; Conn, C. E. Using SANS with Contrast-Matched Lipid Bicontinuous Cubic Phases To Determine the Location of Encapsulated Peptides, Proteins, and Other Biomolecules. *J Phys Chem Lett* **2016**, *7* (14), 2862-2866. DOI: 10.1021/acs.jpcclett.6b01173 From NLM.
- (7) Keller, M. J.; Zhang, Q.; Qian, S.; Sanders, B. C.; O'Neill, H. M.; Hettich, R. L. Characterization of the In Vivo Deuteration of Native Phospholipids by Mass Spectrometry Yields Guidelines for Their Regiospecific Customization. *Analytical Chemistry* **2024**, *96* (1), 212-219. DOI: 10.1021/acs.analchem.3c03750.
- (8) Li, H.-Y.; Zou, Y.; Xu, Y.; Cai, L.; Xie, S.-C.; Zhu, X.-Q.; Zheng, W.-B. Lung Lipidomic Alterations in Beagle Dogs Infected with *Toxocara canis*. *Animals* **2022**, *12* (22), 3080-3080, <https://doi.org/10.3390/ani12223080>.
- (9) Maric, S.; Skar-Gislinge, N.; Midtgaard, S.; Thygesen, M. B.; Schiller, J.; Frielinghaus, H.; Moulin, M.; Haertlein, M.; Forsyth, V. T.; Pomorski, T. G.; Arleth, L. Stealth carriers for low-resolution structure determination of membrane proteins in solution. *Acta Crystallographica Section D* **2014**, *70* (2), 317-328. DOI: doi:10.1107/S1399004713027466.

- (10) Martin, S. W.; Konopka, J. B. Lipid raft polarization contributes to hyphal growth in *Candida albicans*. *Eukaryot Cell* **2004**, 3 (3), 675-684. DOI: 10.1128/ec.3.3.675-684.2004 From NLM.
- (11) Wu, S.; Huang, W.; Wang, F.; Zou, X.; Li, X.; Liu, C.-M.; Zhang, W.; Yan, S. Integrated metabolomics and lipidomics analyses suggest the temperature-dependent lipid desaturation promotes aflatoxin biosynthesis in *Aspergillus flavus*. *Frontiers in Microbiology* **2023**, 14, Original Research. DOI: 10.3389/fmicb.2023.1137643.
- (12) Qu, C.; Li, Z.; Wang, X. UHPLC-HRMS-Based Untargeted Lipidomics Reveal Mechanism of Antifungal Activity of Carvacrol against *Aspergillus flavus*. In *Foods*, 2022; Vol. 11.
- (13) Beccaccioli, M.; Reverberi, M.; Scala, V. Fungal lipids: biosynthesis and signalling during plant-pathogen interaction. *FBL* **2019**, 24 (1), 168-181.
- (14) Rella, A.; Farnoud, A. M.; Del Poeta, M. Plasma membrane lipids and their role in fungal virulence. *Prog Lipid Res* **2016**, 61, 63-72. DOI: 10.1016/j.plipres.2015.11.003 From NLM.
- (15) Rotondo-Trivette, S.; Wang, B.; Luan, Y.; Fiehn, O.; Sun, F.; Michail, S. Reduced fecal short-chain fatty acids in hispanic children with ulcerative colitis. *Physiological Reports* **2021**, 9 (14), e14918. DOI: <https://doi.org/10.14814/phy2.14918> (accessed 2024/08/30).
- (16) Shen, T.; Oh, Y.; Jeong, S.; Cho, S.; Fiehn, O.; Youn, J. H. High-Fat Feeding Alters Circulating Triglyceride Composition: Roles of FFA Desaturation and  $\omega$ -3 Fatty Acid Availability. In *International Journal of Molecular Sciences*, 2024; Vol. 25.
- (17) Kirkwood-Donelson, K. I.; Chappel, J.; Tobin, E.; Dodds, J. N.; Reif, D. M.; DeWitt, J. C.; Baker, E. S. Investigating mouse hepatic lipidome dysregulation following exposure to emerging per- and polyfluoroalkyl substances (PFAS). *Chemosphere* **2024**, 354, 141654. DOI: <https://doi.org/10.1016/j.chemosphere.2024.141654>.
- (18) Bayir, H.; Tyurin, V. A.; Tyurina, Y. Y.; Viner, R.; Ritov, V.; Amoscato, A. A.; Zhao, Q.; Zhang, X. J.; Janesko-Feldman, K. L.; Alexander, H.; Basova, L. V.; Clark, R. S. B.; Kochanek, P. M.; Kagan, V. E. Selective early cardiolipin peroxidation after traumatic brain injury: an oxidative lipidomics analysis. *Annals of Neurology* **2007**, 62 (2), 154-169. DOI: <https://doi.org/10.1002/ana.21168>.

- (19) Iino, Y.; Naganuma, T.; Arita, M. Dysregulated ceramide metabolism in mouse progressive dermatitis resulting from constitutive activation of Jak1. *J Lipid Res* **2023**, *64* (2), 100329. DOI: 10.1016/j.jlr.2023.100329 From NLM.
- (20) Köfeler, H. C.; et al. Quality control requirements for the correct annotation of lipidomics data. *Nature Communications* **2021**, *12* (1), 4771. DOI: 10.1038/s41467-021-24984-y.
- (21) Bowden, J. A.; et al. Harmonizing lipidomics: NIST interlaboratory comparison exercise for lipidomics using SRM 1950-Metabolites in Frozen Human Plasma. *J Lipid Res* **2017**, *58* (12), 2275-2288. DOI: 10.1194/jlr.M079012 From NLM.
- (22) Furse, S.; Wienk, H.; Boelens, R.; de Kroon, A. I. P. M.; Killian, J. A. E. coli MG1655 modulates its phospholipid composition through the cell cycle. *FEBS Letters* **2015**, *589* (19PartB), 2726-2730, <https://doi.org/10.1016/j.febslet.2015.07.043>. DOI: <https://doi.org/10.1016/j.febslet.2015.07.043> (accessed 2022/12/09).
- (23) Ingram, L. O. Changes in lipid composition of Escherichia coli resulting from growth with organic solvents and with food additives. *Applied and Environmental Microbiology* **1977**, *33* (5), 1233-1236. DOI: 10.1128/aem.33.5.1233-1236.1977 (accessed 2022/12/09).
- (24) Ahlers, J.; Günther, T.; Peter, H. Phospholipids and ATPase Activity of Wild-Type and ATPase Deficient and Uncoupled Mutants of E. coli. **1975**, *30* (11-12), 832-834. DOI: doi:10.1515/znc-1975-11-1227 (accessed 2022-12-09).
- (25) Peter, H. W.; Ahlers, J. Phospholipid requirements of ATPase of Escherichia coli. *Archives of Biochemistry and Biophysics* **1975**, *170*, 169-178. DOI: [https://doi.org/10.1016/0003-9861\(75\)90108-3](https://doi.org/10.1016/0003-9861(75)90108-3).
- (26) Cajka, T.; Smilowitz, J. T.; Fiehn, O. Validating Quantitative Untargeted Lipidomics Across Nine Liquid Chromatography–High-Resolution Mass Spectrometry Platforms. *Analytical Chemistry* **2017**, *89* (22), 12360-12368. DOI: 10.1021/acs.analchem.7b03404.
- (27) Kirkwood, K. I.; Pratt, B. S.; Shulman, N.; Tamura, K.; MacCoss, M. J.; MacLean, B. X.; Baker, E. S. Utilizing Skyline to analyze lipidomics data containing liquid chromatography, ion mobility spectrometry and mass spectrometry dimensions. *Nature Protocols* **2022**, *17* (11), 2415-2430. DOI: 10.1038/s41596-022-00714-6.
- (28) Pulliam, A. N.; Gier, E. C.; Gaul, D. A.; Moore, S. G.; Fernández, F. M.; LaPlaca, M. C. Comparing Brain and Blood Lipidome Changes following Single and Repetitive Mild

Traumatic Brain Injury in Rats. *ACS Chemical Neuroscience* **2024**, *15* (2), 300-314. DOI: 10.1021/acscchemneuro.3c00603.

(29) Huang, D.; Liu, J.; Eldridge, R. C.; Gaul, D. A.; Paine, M. R. L.; Uppal, K.; MacDonald, T. J.; Fernández, F. M. Lipidome signatures of metastasis in a transgenic mouse model of sonic hedgehog medulloblastoma. *Analytical and Bioanalytical Chemistry* **2020**, *412* (25), 7017-7027. DOI: 10.1007/s00216-020-02837-9.

(30) Fuchs, B.; Süss, R.; Teuber, K.; Eibisch, M.; Schiller, J. Lipid analysis by thin-layer chromatography--a review of the current state. *J Chromatogr A* **2011**, *1218* (19), 2754-2774. DOI: 10.1016/j.chroma.2010.11.066 From NLM.

(31) Dittmer, J. C.; Lester, R. L. A Simple, Specific Spray for the Detection of Phospholipids on Thin-Layer Chromatograms. *Journal of lipid research* **1964**, *5*, 126-127. PubMed.

(32) Tamburini, S.; Anesi, A.; Ferrentino, G.; Spilimbergo, S.; Guella, G.; Jousson, O. Supercritical CO<sub>2</sub> Induces Marked Changes in Membrane Phospholipids Composition in Escherichia coli K12. *The Journal of Membrane Biology* **2014**, *247* (6), 469-477. DOI: 10.1007/s00232-014-9653-0.

(33) Li, J.; Vosegaard, T.; Guo, Z. Applications of nuclear magnetic resonance in lipid analyses: An emerging powerful tool for lipidomics studies. *Progress in Lipid Research* **2017**, *68*, 37-56. DOI: <https://doi.org/10.1016/j.plipres.2017.09.003>.

(34) Avelo, M. I.; VanRollins, M.; Horrocks, L. A. Separation and quantitation of free fatty acids and fatty acid methyl esters by reverse phase high pressure liquid chromatography. *Journal of Lipid Research* **1983**, *24* (1), 83-93. DOI: [https://doi.org/10.1016/S0022-2275\(20\)38027-5](https://doi.org/10.1016/S0022-2275(20)38027-5).

(35) Tariq, M.; Ali, S.; Ahmad, F.; Ahmad, M.; Zafar, M.; Khalid, N.; Khan, M. A. Identification, FT-IR, NMR (1H and 13C) and GC/MS studies of fatty acid methyl esters in biodiesel from rocket seed oil. *Fuel Processing Technology* **2011**, *92* (3), 336-341. DOI: <https://doi.org/10.1016/j.fuproc.2010.09.025>.

(36) Chopra, A.; Tewari, A. K.; Vatsala, S.; Kumar, R.; Sarpal, A. S.; Basu, B. Determination of Polyunsaturated Fatty Esters (PUFA) in Biodiesel by GC/GC-MS and 1H-NMR Techniques. *Journal of the American Oil Chemists' Society* **2011**, *88* (9), 1285-1296. DOI: 10.1007/s11746-011-1795-y.

(37) Berdeaux, O.; Fontagné, S.; Sémon, E.; Velasco, J.; Sébédio, J. L.; Dobarganes, C. A detailed identification study on high-temperature degradation products of oleic and linoleic acid methyl esters by GC–MS and GC–FTIR. *Chemistry and Physics of Lipids* **2012**, *165* (3), 338–347. DOI: <https://doi.org/10.1016/j.chemphyslip.2012.02.004>.

(38) Beccaria, M.; Franchina, F. A.; Nasir, M.; Mellors, T.; Hill, J. E.; Purcaro, G. Investigation of mycobacteria fatty acid profile using different ionization energies in GC-MS. *Anal Bioanal Chem* **2018**, *410* (30), 7987–7996. DOI: 10.1007/s00216-018-1421-z From NLM.

(39) Byss, M.; Tríska, J.; Elhottová, D. GC-MS-MS analysis of bacterial fatty acids in heavily creosote-contaminated soil samples. *Anal Bioanal Chem* **2007**, *387* (4), 1573–1577. DOI: 10.1007/s00216-006-1024-y From NLM.

(40) Han, X.; Gross, R. W. Global analyses of cellular lipidomes directly from crude extracts of biological samples by ESI mass spectrometry: a bridge to lipidomics. *Journal of Lipid Research* **2003**, *44* (6), 1071–1079. DOI: 10.1194/jlr.R300004-JLR200 (accessed 2024/08/07).

(41) Han, X.; Yang, K.; Gross, R. W. Multi-dimensional mass spectrometry-based shotgun lipidomics and novel strategies for lipidomic analyses. *Mass Spectrom Rev* **2012**, *31* (1), 134–178. DOI: 10.1002/mas.20342 From NLM.

(42) Gross, R. W. The evolution of lipidomics through space and time. *Biochim Biophys Acta Mol Cell Biol Lipids* **2017**, *1862* (8), 731–739. DOI: 10.1016/j.bbalip.2017.04.006 From NLM.

(43) Hartler, J.; Trötz Müller, M.; Chitraju, C.; Spener, F.; Köfeler, H. C.; Thallinger, G. G. Lipid Data Analyzer: unattended identification and quantitation of lipids in LC-MS data. *Bioinformatics* **2011**, *27* (4), 572–577. DOI: <https://doi.org/10.1093/bioinformatics/btq699>.

(44) Goracci, L.; Tortorella, S.; Tiberi, P.; Pellegrino, R. M.; Di Veroli, A.; Valeri, A.; Cruciani, G. Lipostar, a Comprehensive Platform-Neutral Cheminformatics Tool for Lipidomics. *Analytical Chemistry* **2017**, *89* (11), 6257–6264. DOI: 10.1021/acs.analchem.7b01259.

(45) Herzog, R.; Schwudke, D.; Shevchenko, A. LipidXplorer: Software for Quantitative Shotgun Lipidomics Compatible with Multiple Mass Spectrometry Platforms. *Current Protocols in Bioinformatics* **2013**, *43* (1), 14.12.11–14.12.30. DOI: <https://doi.org/10.1002/0471250953.bi1412s43> (accessed 2024/09/03).

(46) Ni, Z.; Angelidou, G.; Lange, M.; Hoffmann, R.; Fedorova, M. LipidHunter Identifies Phospholipids by High-Throughput Processing of LC-MS and Shotgun Lipidomics

Datasets. *Anal Chem* **2017**, *89* (17), 8800-8807. DOI: 10.1021/acs.analchem.7b01126 From NLM.

(47) Xu, R.; Liu, H.; Yuan, F.; Kim, S.; Kirpich, I.; McClain, C. J.; Zhang, X. Lipid Wizard: Analysis Software for Comprehensive Two-Dimensional Liquid Chromatography–Mass Spectrometry-Based Lipid Profiling. *Analytical Chemistry* **2024**, *96* (14), 5375-5383. DOI: 10.1021/acs.analchem.3c04419.

(48) Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; Arita, M. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nature Methods* **2015**, *12* (6), 523-526. DOI: 10.1038/nmeth.3393.

(49) Tsugawa, H.; Ikeda, K.; Takahashi, M.; Satoh, A.; Mori, Y.; Uchino, H.; Okahashi, N.; Yamada, Y.; Tada, I.; Bonini, P.; Higashi, Y.; Okazaki, Y.; Zhou, Z.; Zhu, Z. J.; Koelmel, J.; Cajka, T.; Fiehn, O.; Saito, K.; Arita, M.; Arita, M. A lipidome atlas in MS-DIAL 4. *Nature Biotechnology* **2020**, *38* (10), 1159–1163-1159–1163, <https://doi.org/10.1038/s41587-020-0531-2>.

(50) Schmid, R.; et al. Integrative analysis of multimodal mass spectrometry data in MZmine 3. *Nature Biotechnology* **2023**, *41* (4), 447–449-447–449, <https://doi.org/10.1038/s41587-023-01690-2>.

(51) Harzing, A. Publish or Perish. 2007.

(52) Sud, M.; Fahy, E.; Cotter, D.; Azam, K.; Vadivelu, I.; Burant, C.; Edison, A.; Fiehn, O.; Higashi, R.; Nair, K. S.; Sumner, S.; Subramaniam, S. Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res* **2016**, *44* (D1), D463-470. DOI: 10.1093/nar/gkv1042 From NLM.

(53) Haug, K.; Cochrane, K.; Nainala, V. C.; Williams, M.; Chang, J.; Jayaseelan, K. V.; O'Donovan, C. MetaboLights: a resource evolving in response to the needs of its scientific community. *Nucleic Acids Res* **2020**, *48* (D1), D440-d444. DOI: 10.1093/nar/gkz1019 From NLM.

(54) Yurekten, O.; Payne, T.; Tejera, N.; Amaladoss, F. X.; Martin, C.; Williams, M.; O'Donovan, C. MetaboLights: open data repository for metabolomics. *Nucleic Acids Research* **2023**, *52* (D1), D640-D646, <https://doi.org/10.1093/nar/gkad1045>.

- (55) Harrieder, E. M.; Kretschmer, F.; Böcker, S.; Witting, M. Current state-of-the-art of separation methods used in LC-MS based metabolomics and lipidomics. *J Chromatogr B Analyt Technol Biomed Life Sci* **2022**, *1188*, 123069. DOI: 10.1016/j.jchromb.2021.123069 From NLM.
- (56) Granafei, S.; Azzone, P.; Spinelli, V. A.; Losito, I.; Palmisano, F.; Cataldi, T. R. I. Hydrophilic interaction and reversed phase mixed-mode liquid chromatography coupled to high resolution tandem mass spectrometry for polar lipids analysis. *Journal of Chromatography A* **2016**, *1477*, 47-55. DOI: <https://doi.org/10.1016/j.chroma.2016.11.048>.
- (57) Pham, T. H.; Zaeem, M.; Fillier, T. A.; Nadeem, M.; Vidal, N. P.; Manful, C.; Cheema, S.; Cheema, M.; Thomas, R. H. Targeting Modified Lipids during Routine Lipidomics Analysis using HILIC and C30 Reverse Phase Liquid Chromatography coupled to Mass Spectrometry. *Scientific Reports* **2019**, *9* (1), 5048. DOI: 10.1038/s41598-019-41556-9.
- (58) Danne-Rasche, N.; Coman, C.; Ahrends, R. Nano-LC/NSI MS Refines Lipidomics by Enhancing Lipid Coverage, Measurement Sensitivity, and Linear Dynamic Range. *Analytical Chemistry* **2018**, *90* (13), 8093-8101. DOI: 10.1021/acs.analchem.8b01275.
- (59) Folch, J.; Ascoli, I.; Lees, M.; Meath, J.; LeBaron, F. Preparation of lipide extracts from brain tissue. *Journal of Biological Chemistry* **1951**, *191* (2), 833-841.
- (60) Folch, J.; Lees, M.; Stanley, G. H. S. A simple method for the isolation and purification of total lipides from animal tissues. *Journal of biological chemistry* **1957**, *226* (1), 497-509.
- (61) Bligh, E. G.; Dyer, W. J. A Rapid Method of Total Lipid Extraction and Purification. *Canadian Journal of Biochemistry and Physiology* **1959**, *37* (8), 911-917. DOI: 10.1139/o59-099 (accessed 2023/08/04).
- (62) Matyash, V.; Liebisch, G.; Kurzchalia, T. V.; Shevchenko, A.; Schwudke, D. Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *J Lipid Res* **2008**, *49* (5), 1137-1146. DOI: 10.1194/jlr.D700041-JLR200 From NLM.
- (63) Theodoridis, G.; Gika, H.; Raftery, D.; Goodacre, R.; Plumb, R. S.; Wilson, I. D. Ensuring Fact-Based Metabolite Identification in Liquid Chromatography–Mass Spectrometry-Based Metabolomics. *Analytical Chemistry* **2023**, *95* (8), 3909-3916. DOI: 10.1021/acs.analchem.2c05192.

- (64) Skotland, T.; Ekroos, K.; McDonald, J.; Ahrends, R.; Liebisch, G.; Sandvig, K. Pitfalls in lipid mass spectrometry of mammalian samples — a brief guide for biologists. *Nature Reviews Molecular Cell Biology* **2024**. DOI: 10.1038/s41580-024-00758-4.
- (65) Liebisch, G.; Ejsing, C. S.; Ekroos, K. Identification and Annotation of Lipid Species in Metabolomics Studies Need Improvement. *Clinical Chemistry* **2015**, *61* (12), 1542-1544. DOI: 10.1373/clinchem.2015.244830 (accessed 7/23/2024).
- (66) Koelmel, J. P.; Ulmer, C. Z.; Jones, C. M.; Yost, R. A.; Bowden, J. A. Common cases of improper lipid annotation using high-resolution tandem mass spectrometry data and corresponding limitations in biological interpretation. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **2017**, *1862* (8), 766-770. DOI: <https://doi.org/10.1016/j.bbalip.2017.02.016>.
- (67) Liebisch, G.; Ekroos, K.; Hermansson, M.; Ejsing, C. S. Reporting of lipidomics data should be standardized. *Biochim Biophys Acta Mol Cell Biol Lipids* **2017**, *1862* (8), 747-751. DOI: 10.1016/j.bbalip.2017.02.013 From NLM.
- (68) Vasilopoulou, C. G.; Sulek, K.; Brunner, A.-D.; Meitei, N. S.; Schweiger-Hufnagel, U.; Meyer, S. W.; Barsch, A.; Mann, M.; Meier, F. Reply to “Quality control requirements for the correct annotation of lipidomics data”. *Nature Communications* **2021**, *12* (1), 4772. DOI: 10.1038/s41467-021-24985-x.
- (69) Trutschel, D.; Schmidt, S.; Grosse, I.; Neumann, S. Experiment design beyond gut feeling: statistical tests and power to detect differential metabolites in mass spectrometry data. *Metabolomics* **2015**, *11* (4), 851-860. DOI: 10.1007/s11306-014-0742-y.
- (70) Mahieu, N. G.; Patti, G. J. Systems-Level Annotation of a Metabolomics Data Set Reduces 25 000 Features to Fewer than 1000 Unique Metabolites. *Analytical Chemistry* **2017**, *89* (19), 10397-10406. DOI: 10.1021/acs.analchem.7b02380.
- (71) Wang, L.; Xing, X.; Chen, L.; Yang, L.; Su, X.; Rabitz, H.; Lu, W.; Rabinowitz, J. D. Peak Annotation and Verification Engine for Untargeted LC-MS Metabolomics. *Anal Chem* **2019**, *91* (3), 1838-1846. DOI: 10.1021/acs.analchem.8b03132 From NLM.
- (72) Röst, H. L.; et al. OpenMS: a flexible open-source software platform for mass spectrometry data analysis. *Nature Methods* **2016**, *13* (9), 741–748-741–748, <https://doi.org/10.1038/nmeth.3959>.

(73) Aicheler, F.; Li, J.; Hoene, M.; Lehmann, R.; Xu, G.; Kohlbacher, O. Retention Time Prediction Improves Identification in Nontargeted Lipidomics Approaches. *Analytical Chemistry* **2015**, *87* (15), 7698–7704. <https://doi.org/10.1021/acs.analchem.5b01139>.

(74) 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. *Bioinformatics* **2022**, *38* (10), 2872–2879. <https://doi.org/10.1093/bioinformatics/btac197>.

(75) 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>.

(76) Liebisch, G.; Fahy, E.; Aoki, J.; Dennis, E. A.; Durand, T.; Ejsing, C. S.; Fedorova, M.; Feussner, I.; Griffiths, W. J.; Köfeler, H.; Merrill, A. H.; Murphy, R. C.; O'Donnell, V. B.; Oskolkova, O.; Subramaniam, S.; Wakelam, M. J. O.; Spener, F. Update on LIPID MAPS classification, nomenclature, and shorthand notation for MS-derived lipid structures. *Journal of Lipid Research* **2020**, *61* (12), 1539–1555. DOI: <https://doi.org/10.1194/jlr.S120001025>.

(77) White, J. B.; Trim, P. J.; Salagaras, T.; Long, A.; Psaltis, P. J.; Verjans, J. W.; Snel, M. F. Equivalent Carbon Number and Interclass Retention Time Conversion Enhance Lipid Identification in Untargeted Clinical Lipidomics. *Analytical Chemistry* **2022**, *94* (8), 3476–3484. DOI: 10.1021/acs.analchem.1c03770.

(78) Pulfer, M.; Murphy, R. C. Electrospray mass spectrometry of phospholipids. *Mass Spectrometry Reviews* **2003**, *22* (5), 332–364. DOI: <https://doi.org/10.1002/mas.10061> (accessed 2024/08/07).

(79) Wenk, M. R.; Lucast, L.; Di Paolo, G.; Romanelli, A. J.; Suchy, S. F.; Nussbaum, R. L.; Cline, G. W.; Shulman, G. I.; McMurray, W.; De Camilli, P. Phosphoinositide profiling in complex lipid mixtures using electrospray ionization mass spectrometry. *Nature Biotechnology* **2003**, *21* (7), 813–817. DOI: 10.1038/nbt837.

(80) Hsu, F. F.; Turk, J. Characterization of phosphatidylinositol, phosphatidylinositol-4-phosphate, and phosphatidylinositol-4,5-bisphosphate by electrospray ionization tandem mass spectrometry: a mechanistic study. *J Am Soc Mass Spectrom* **2000**, *11* (11), 986–999. DOI: 10.1016/s1044-0305(00)00172-0 From NLM.

(81) van der Hooft, J. J.; Ridder, L.; Barrett, M. P.; Burgess, K. E. Enhanced acylcarnitine annotation in high-resolution mass spectrometry data: fragmentation analysis for the classification and annotation of acylcarnitines. *Front Bioeng Biotechnol* **2015**, *3*, 26. DOI: 10.3389/fbioe.2015.00026 From NLM.

(82) Zhang, J.; Chen, Q.; Zhang, L.; Shi, B.; Yu, M.; Huang, Q.; Tang, H. Simultaneously quantifying hundreds of acylcarnitines in multiple biological matrices within ten minutes using ultrahigh-performance liquid-chromatography and tandem mass spectrometry. *Journal of Pharmaceutical Analysis* **2024**, *14* (1), 140-148. DOI: <https://doi.org/10.1016/j.jpha.2023.10.004>.

(83) McAnoy, A. M.; Wu, C. C.; Murphy, R. C. Direct qualitative analysis of triacylglycerols by electrospray mass spectrometry using a linear ion trap. *Journal of the American Society for Mass Spectrometry* **2005**, *16* (9), 1498-1509. DOI: 10.1016/j.jasms.2005.04.017.

(84) Hvattum, E. Analysis of triacylglycerols with non-aqueous reversed-phase liquid chromatography and positive ion electrospray tandem mass spectrometry. *Rapid Communications in Mass Spectrometry* **2001**, *15* (3), 187-190. DOI: 10.1002/1097-0231(20010215)15:3<187::AID-RCM211>3.0.CO;2-T

(85) Raith, K.; Neubert, R. H. H. Liquid chromatography–electrospray mass spectrometry and tandem mass spectrometry of ceramides. *Analytica Chimica Acta* **2000**, *403* (1), 295-303. DOI: [https://doi.org/10.1016/S0003-2670\(99\)00661-3](https://doi.org/10.1016/S0003-2670(99)00661-3).

(86) Hsu, F.-F. Electrospray ionization with higher-energy collision dissociation tandem mass spectrometry toward characterization of ceramides as  $[M + Li]^+$  ions: Mechanisms of fragmentation and structural identification. *Analytica Chimica Acta* **2021**, *1142*, 221-234. DOI: <https://doi.org/10.1016/j.aca.2020.09.056>.

(87) Quiroga, M.; Garay, P. G.; Alonso, J. M.; Loyola, J. M.; Martin, O. A. Bayesian additive regression trees for probabilistic programming. arXiv: 2023.

(88) Broadhurst, D.; Goodacre, R.; Reinke, S. N.; Kuligowski, J.; Wilson, I. D.; Lewis, M. R.; Dunn, W. B. Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics* **2018**, *14* (6), 72. DOI: 10.1007/s11306-018-1367-3.

- (89) Karaman, I. Preprocessing and Pretreatment of Metabolomics Data for Statistical Analysis. In *Metabolomics: From Fundamentals to Clinical Applications*, Sussulini, A. Ed.; Springer International Publishing, 2017; pp 145-161.
- (90) Thompson, J. W.; et al. International Ring Trial of a High Resolution Targeted Metabolomics and Lipidomics Platform for Serum and Plasma Analysis. *Analytical Chemistry* **2019**, *91* (22), 14407-14416. DOI: 10.1021/acs.analchem.9b02908.
- (91) Wolin, E. A.; Wolin, M. J.; Wolfe, R. S. Formation of Methane by Bacterial Extracts. *Journal of Biological Chemistry* **1963**, *238* (8), 2882-2886. DOI: [https://doi.org/10.1016/S0021-9258\(18\)67912-8](https://doi.org/10.1016/S0021-9258(18)67912-8).
- (92) Wischnak, C.; Löffler Frank, E.; Li, J.; Urbance John, W.; Müller, R. Pseudomonas sp. Strain 273, an Aerobic  $\alpha,\omega$ -Dichloroalkane Degrading Bacterium. *Applied and Environmental Microbiology* **1998**, *64* (9), 3507-3511. DOI: 10.1128/AEM.64.9.3507-3511.1998 (accessed 2024/09/13).
- (93) Xie, Y.; Chen, G.; May, A. L.; Yan, J.; Brown, L. P.; Powers, J. B.; Campagna, S. R.; Löffler, F. E. Pseudomonas sp. Strain 273 Degrades Fluorinated Alkanes. *Environmental Science & Technology* **2020**, *54* (23), 14994-15003. DOI: 10.1021/acs.est.0c04029.
- (94) Agrawal, S.; Kumar, S.; Sehgal, R.; George, S.; Gupta, R.; Poddar, S.; Jha, A.; Pathak, S. El-MAVEN: A Fast, Robust, and User-Friendly Mass Spectrometry Data Processing Engine for Metabolomics. In *High-Throughput Metabolomics: Methods and Protocols*, D'Alessandro, A. Ed.; Springer New York, 2019; pp 301-321.
- (95) Cleveland, W. S. Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association* **1979**, *74* (368), 829-836. DOI: 10.2307/2286407 (accessed 2024/09/16/).JSTOR.
- (96) Virtanen, P.; et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **2020**, *17* (3), 261-272. DOI: 10.1038/s41592-019-0686-2.
- (97) Stephens, M. A. EDF Statistics for Goodness of Fit and Some Comparisons. *Journal of the American Statistical Association* **1974**, *69* (347), 730-737. DOI: 10.1080/01621459.1974.10480196.
- (98) Harris, C. R.; et al. Array programming with NumPy. *Nature* **2020**, *585* (7825), 357-362. DOI: 10.1038/s41586-020-2649-2.

- (99) Anderson, D. R.; Burnham, K. P.; Thompson, W. L. Null Hypothesis Testing: Problems, Prevalence, and an Alternative. *The Journal of Wildlife Management* **2000**, *64* (4), 912-923. DOI: 10.2307/3803199 (accessed 2024/09/16/).JSTOR.
- (100) Meehl, P. E. Theory-Testing in Psychology and Physics: A Methodological Paradox. *Philosophy of Science* **1967**, *34* (2), 103-115. (accessed 2024/09/17/).JSTOR.
- (101) Lakens, D. Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses. *Social Psychological and Personality Science* **2017**, *8* (4), 355-362. DOI: 10.1177/1948550617697177 (accessed 2024/09/16).
- (102) Caldwell, A. R. Exploring Equivalence Testing with the Updated TOSTER R Package. PsyArXiv, 2022.
- (103) Blume, J. D.; Greevy, R. A.; Welty, V. F.; Smith, J. R.; Dupont, W. D. An Introduction to Second-Generation p-Values. *The American Statistician* **2019**, *73* (sup1), 157-167. DOI: 10.1080/00031305.2018.1537893.
- (104) Borrego, S. L.; Fahrman, J.; Hou, J.; Lin, D.-W.; Tromberg, B. J.; Fiehn, O.; Kaiser, P. Lipid remodeling in response to methionine stress in MDA-MBA-468 triple-negative breast cancer cells. *Journal of Lipid Research* **2021**, *62*, 100056. DOI: <https://doi.org/10.1016/j.jlr.2021.100056>.
- (105) Rush, T. A.; Tannous, J.; Lane, M., J.; Gopalakrishnan Meena, M.; Carrell, A. A.; Golan, J. J.; Drott, M. T.; Cottaz, S.; Fort, S.; Ané, J.-M.; Keller, N. P.; Pelletier, D. A.; Jacobson, D. A.; Kainer, D.; Abraham, P. E.; Giannone, R. J.; Labbé, J. L. Lipo-Chitooligosaccharides Induce Specialized Fungal Metabolite Profiles That Modulate Bacterial Growth. *mSystems* **2022**, *7* (6), e01052-01022. DOI: 10.1128/msystems.01052-22 (accessed 2024/07/25).
- (106) Martin, F.; et al. The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* **2008**, *452* (7183), 88-92. DOI: 10.1038/nature06556.
- (107) Klironomos, J. N.; Hart, M. M. Animal nitrogen swap for plant carbon. *Nature* **2001**, *410* (6829), 651-652. DOI: 10.1038/35070643.
- (108) Latgé, J. P. *Aspergillus fumigatus* and aspergillosis. *Clin Microbiol Rev* **1999**, *12* (2), 310-350. DOI: 10.1128/cmr.12.2.310 From NLM.
- (109) Scarpari, M.; Punelli, M.; Scala, V.; Zaccaria, M.; Nobili, C.; Ludovici, M.; Camera, E.; Fabbri, A. A.; Reverberi, M.; Fanelli, C. Lipids in *Aspergillus flavus*-maize interaction. *Frontiers in Microbiology* **2014**, *5*. DOI: 10.3389/fmicb.2014.00074.

(110) Pan, J.; Yang, X.; Hu, C.; Fu, T.; Zhang, X.; Liu, Z.; Wang, Y.; Zhang, F.; He, X.; Yu, J.-H. Functional, transcriptomic, and lipidomic studies of the choC gene encoding a phospholipid methyltransferase in *Aspergillus fumigatus*. *Microbiology Spectrum* **2023**, *12* (1), e02168-02123. DOI: 10.1128/spectrum.02168-23 (accessed 2024/07/26).

(111) Kind, T.; Fiehn, O. Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinformatics* **2006**, *7* (1), 234. DOI: 10.1186/1471-2105-7-234.

(112) Chakraborty, A.; Alsharqi, L.; Kostrzewa, M.; Armstrong-James, D.; Larrouy-Maumus, G. Intact cell lipidomics using the Bruker MBT lipid Xtract assay allows the rapid detection of glycosyl-inositol-phospho-ceramides from *Aspergillus fumigatus*. *Molecular Omics* **2024**, *20* (6), 390-396, 10.1039/D4MO00030G. DOI: 10.1039/D4MO00030G.

(113) D'Haese, W.; Holsters, M. Nod factor structures, responses, and perception during initiation of nodule development. *Glycobiology* **2002**, *12* (6), 79r-105r. DOI: 10.1093/glycob/12.6.79r From NLM.

(114) Rush, T. A.; et al. Lipo-chitoooligosaccharides as regulatory signals of fungal growth and development. *Nature Communications* **2020**, *11* (1), 3897. DOI: 10.1038/s41467-020-17615-5.

(115) Weerasinghe, R. R.; Bird, D. M.; Allen, N. S. Root-knot nematodes and bacterial Nod factors elicit common signal transduction events in *Lotus japonicus*. *Proc Natl Acad Sci U S A* **2005**, *102* (8), 3147-3152. DOI: 10.1073/pnas.0407926102 From NLM.

(116) Khokhani, D.; Carrera Carriel, C.; Vayla, S.; Irving, T. B.; Stonoha-Arther, C.; Keller, N. P.; Ané, J.-M. Deciphering the Chitin Code in Plant Symbiosis, Defense, and Microbial Networks. *Annual Review of Microbiology* **2021**, *75* (Volume 75, 2021), 583-607. DOI: <https://doi.org/10.1146/annurev-micro-051921-114809>.

(117) Maillet, F.; Poinot, V.; André, O.; Puech-Pagès, V.; Haouy, A.; Gueunier, M.; Cromer, L.; Giraudet, D.; Formey, D.; Niebel, A.; Martinez, E. A.; Driguez, H.; Bécard, G.; Dénarié, J. Fungal lipochitoooligosaccharide symbiotic signals in arbuscular mycorrhiza. *Nature* **2011**, *469* (7328), 58-63. DOI: 10.1038/nature09622 From NLM.

(118) Cope, K. R.; et al. The Ectomycorrhizal Fungus *Laccaria bicolor* Produces Lipochitoooligosaccharides and Uses the Common Symbiosis Pathway to Colonize *Populus*

Roots. *The Plant Cell* **2019**, *31* (10), 2386-2410. DOI: 10.1105/tpc.18.00676 (accessed 9/12/2024).

(119) Chen, W.; Wang, D.; Ke, S.; Cao, Y.; Xiang, W.; Guo, X.; Yang, Q. A soybean cyst nematode suppresses microbial plant symbionts using a lipochitooligosaccharide-hydrolysing enzyme. *Nature Microbiology* **2024**, *9* (8), 1993-2005. DOI: 10.1038/s41564-024-01727-5.

(120) Villalobos Solis, M. I.; Engle, N. L.; Spangler, M. K.; Cottaz, S.; Fort, S.; Maeda, J.; Ané, J.-M.; Tschaplinski, T. J.; Labbé, J. L.; Hettich, R. L.; Abraham, P. E.; Rush, T. A. Expanding the Biological Role of Lipo-Chitooligosaccharides and Chitooligosaccharides in *Laccaria bicolor* Growth and Development. *Frontiers in Fungal Biology* **2022**, *3*. DOI: 10.3389/ffunb.2022.808578.

(121) Gopalakrishnan Meena, M.; Lane, M. J.; Tannous, J.; Carrell, A. A.; Abraham, P. E.; Giannone, R. J.; Ané, J.-M.; Keller, N. P.; Labbé, J. L.; Geiger, A. G.; Kainer, D.; Jacobson, D. A.; Rush, T. A. A glimpse into the fungal metabolomic abyss: Novel network analysis reveals relationships between exogenous compounds and their outputs. *PNAS Nexus* **2023**, *2* (10), pgad322. DOI: 10.1093/pnasnexus/pgad322 (accessed 9/12/2024).

(122) Latgé, J.-P. *Aspergillus fumigatus* and Aspergillosis. *Clinical Microbiology Reviews* **1999**, *12* (2), 310-350. DOI: 10.1128/cmr.12.2.310 (accessed 2024/07/25).

(123) Brookman, J. L.; Denning, D. W. Molecular genetics in *Aspergillus fumigatus*. *Current Opinion in Microbiology* **2000**, *3* (5), 468-474. DOI: [https://doi.org/10.1016/S1369-5274\(00\)00124-7](https://doi.org/10.1016/S1369-5274(00)00124-7).

(124) Keller, N. P. Fungal secondary metabolism: regulation, function and drug discovery. *Nature Reviews Microbiology* **2019**, *17* (3), 167-180. DOI: 10.1038/s41579-018-0121-1.

(125) Kaminskyj, S. G. Fundamentals of growth, storage, genetics and microscopy of *Aspergillus nidulans*. *Fungal Genetics Reports* **2001**, *48*. DOI: <https://doi.org/10.4148/1941-4765.1175>.

(126) Tsugawa, H.; Ikeda, K.; Takahashi, M.; Satoh, A.; Mori, Y.; Uchino, H.; Okahashi, N.; Yamada, Y.; Tada, I.; Bonini, P.; Higashi, Y.; Okazaki, Y.; Zhou, Z.; Zhu, Z.-J.; Koelmel, J.; Cajka, T.; Fiehn, O.; Saito, K.; Arita, M.; Arita, M. A lipidome atlas in MS-DIAL 4. *Nature Biotechnology* **2020**, *38* (10), 1159-1163. DOI: 10.1038/s41587-020-0531-2.

- (127) Singh, A.; Prasad, T.; Kapoor, K.; Mandal, A.; Roth, M.; Welte, R.; Prasad, R. Phospholipidome of *Candida*: each species of *Candida* has distinctive phospholipid molecular species. *Omics* **2010**, *14* (6), 665-677. DOI: 10.1089/omi.2010.0041 From NLM.
- (128) Li, S.; Du, L.; Yuen, G.; Harris, S. D. Distinct ceramide synthases regulate polarized growth in the filamentous fungus *Aspergillus nidulans*. *Mol Biol Cell* **2006**, *17* (3), 1218-1227. DOI: 10.1091/mbc.e05-06-0533 From NLM.
- (129) Suutari, M. Effect of growth temperature on lipid fatty acids of four fungi (*Aspergillus niger*, *Neurospora crassa*, *Penicillium chrysogenum*, and *Trichoderma reesei*). *Archives of Microbiology* **1995**, *164* (3), 212-216. DOI: 10.1007/BF02529973.
- (130) Gelman, A.; Loken, E. The garden of forking paths: Why multiple comparisons can be a problem , even when there is no “ fishing expedition ” or “ p-hacking ” and the research hypothesis was posited ahead of time. 2019.
- (131) Egoshi, S.; Dodo, K.; Sodeoka, M. Deuterium Raman imaging for lipid analysis. *Current Opinion in Chemical Biology* **2022**, *70*, 102181. DOI: <https://doi.org/10.1016/j.cbpa.2022.102181>.
- (132) Curatolo, W.; Jungalwala, F. B.; Sears, B.; Tuck, L.; Neuringer, L. J. Deuterium NMR spectroscopy of biosynthetically deuterated mammalian tissues. *Biochemistry* **1985**, *24* (16), 4360-4364. DOI: 10.1021/bi00337a017.
- (133) Gally, H. U.; Pluschke, G.; Overath, P.; Seelig, J. Structure of *Escherichia coli* membranes. Glycerol auxotrophs as a tool for the analysis of the phospholipid head-group region by deuterium magnetic resonance. *Biochemistry* **1981**, *20* (7), 1826-1831. DOI: 10.1021/bi00510a017.
- (134) Kang, S. Y.; Gutowsky, H. S.; Oldfield, E. Spectroscopic studies of specifically deuterium labeled membrane systems. Nuclear magnetic resonance investigation of protein-lipid interactions in *Escherichia coli* membranes. *Biochemistry* **1979**, *18* (15), 3268-3272. DOI: 10.1021/bi00582a011.
- (135) Holčápek, M.; Liebisch, G.; Ekroos, K. Lipidomic Analysis. *Analytical Chemistry* **2018**, *90* (7), 4249-4257. DOI: 10.1021/acs.analchem.7b05395.
- (136) Triebel, A.; Wenk, M. R. Analytical Considerations of Stable Isotope Labelling in Lipidomics. *Biomolecules*. **2018**, *8* (4), 151. DOI: 10.3390/biom8040151.

- (137) Krueger, S. Small-angle neutron scattering contrast variation studies of biological complexes: Challenges and triumphs. *Current Opinion in Structural Biology* **2022**, *74*, 102375. DOI: <https://doi.org/10.1016/j.sbi.2022.102375>.
- (138) Sharma, V. K.; Qian, S. Effect of an Antimicrobial Peptide on Lateral Segregation of Lipids: A Structure and Dynamics Study by Neutron Scattering. *Langmuir* **2019**, *35* (11), 4152-4160. DOI: 10.1021/acs.langmuir.8b04158.
- (139) Qian, S.; Sharma, V. K.; Clifton, L. A. Understanding the Structure and Dynamics of Complex Biomembrane Interactions by Neutron Scattering Techniques. *Langmuir* **2020**, *36* (50), 15189-15211. DOI: 10.1021/acs.langmuir.0c02516.
- (140) Midtgaard, S. R.; et al. Invisible detergents for structure determination of membrane proteins by small-angle neutron scattering. *The FEBS Journal* **2018**, *285* (2), 357-371. DOI: <https://doi.org/10.1111/febs.14345>.
- (141) Moir, M.; Yepuri, N. R.; Marshall, D. L.; Blanksby, S. J.; Darwish, T. A. Synthesis of Perdeuterated Linoleic Acid-d31 and Chain Deuterated 1-Palmitoyl-2-linoleoyl-sn-glycero-3-phosphocholine-d62. *Advanced Synthesis & Catalysis* **2022**, *364* (21), 3670-3681. DOI: <https://doi.org/10.1002/adsc.202200616> (accessed 2023/10/11).
- (142) Young, J. W. Recent advances in membrane mimetics for membrane protein research. *Biochemical Society Transactions* **2023**, *51* (3), 1405-1416. DOI: 10.1042/BST20230164 (accessed 8/4/2023).
- (143) Maric, S.; Thygesen, M. B.; Schiller, J.; Marek, M.; Moulin, M.; Haertlein, M.; Forsyth, V. T.; Bogdanov, M.; Dowhan, W.; Arleth, L.; Pomorski, T. G. Biosynthetic preparation of selectively deuterated phosphatidylcholine in genetically modified *Escherichia coli*. *Applied Microbiology and Biotechnology* **2015**, *99* (1), 241-254. DOI: 10.1007/s00253-014-6082-z.
- (144) Weiss, K. L.; Fan, Y.; Abraham, P.; Odom, M.; Pant, S.; Zhang, Q.; O'Neill, H. Chapter Nine - Fed-batch production of deuterated protein in *Escherichia coli* for neutron scattering experimentation. In *Methods in Enzymology*, Kelman, Z., O'Dell, W. B. Eds.; Vol. 659; Academic Press, 2021; pp 219-240.
- (145) Törnkvist, M.; Larsson, G.; Enfors, S. O. Protein release and foaming in *Escherichia coli* cultures grown in minimal medium. *Bioprocess Engineering* **1996**, *15* (5), 231-237. DOI: 10.1007/BF02391583.

(146) Liu, T.; Vora, H.; Khosla, C. Quantitative analysis and engineering of fatty acid biosynthesis in *E. coli*. *Metab Eng* **2010**, *12* (4), 378-386. DOI: 10.1016/j.ymben.2010.02.003 From NLM.

(147) Lu, X.; Vora, H.; Khosla, C. Overproduction of free fatty acids in *E. coli*: Implications for biodiesel production. *Metabolic Engineering* **2008**, *10* (6), 333-339. DOI: <https://doi.org/10.1016/j.ymben.2008.08.006>.

(148) Kito, M.; Aibara, S.; Kato, M.; Hata, T. Differences in fatty acid composition among phosphatidylethanolamine, phosphatidylglycerol and cardiolipin of *Escherichia coli*. *Biochim Biophys Acta* **1972**, *260* (3), 475-478. DOI: 10.1016/0005-2760(72)90062-8 From NLM.

(149) Yasuhiro, K.; Yuzuru, A.; Shoshichi, N. Composition and turnover of the phospholipids in *Escherichia coli*. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* **1967**, *144* (2), 382-390. DOI: [https://doi.org/10.1016/0005-2760\(67\)90167-1](https://doi.org/10.1016/0005-2760(67)90167-1).

(150) Siervo, A. J. D. Alterations in the Phospholipid Composition of *Escherichia coli* B During Growth at Different Temperatures. *Journal of Bacteriology* **1969**, *100*, 1342-1349.

(151) Fahy, E.; Sud, M.; Cotter, D.; Subramaniam, S. LIPID MAPS online tools for lipid research. *Nucleic Acids Res* **2007**, *35* (Web Server issue), W606-612. DOI: 10.1093/nar/gkm324 From NLM.

(152) *Escherichia coli and Salmonella: cellular and molecular biology*, 2nd ed.; Neidhardt, F. C., Ed.; ASM Press, 1996.

(153) Heller, W. T.; Urban, V. S.; Lynn, G. W.; Weiss, K. L.; O'Neill, H. M.; Pingali, S. V.; Qian, S.; Littrell, K. C.; Melnichenko, Y. B.; Buchanan, M. V.; Selby, D. L.; Wignall, G. D.; Butler, P. D.; Myles, D. A. The Bio-SANS instrument at the High Flux Isotope Reactor of Oak Ridge National Laboratory. *Journal of Applied Crystallography* **2014**, *47* (4), 1238-1246. DOI: doi:10.1107/S1600576714011285.

(154) Heller, W. T.; et al. drtsans: The data reduction toolkit for small-angle neutron scattering at Oak Ridge National Laboratory. *SoftwareX* **2022**, *19*, 101101. DOI: <https://doi.org/10.1016/j.softx.2022.101101>.

(155) Barnali Chaudhuri, I. G. M., Shuo Qian, Volker S. Urban. Biological Small Angle Scattering: Techniques, Strategies and Tips In *Advances in Experimental Medicine and Biology*,

1 ed.; Wim E. Crusio, H. D., Heinfried H. Radeke, Nima Rezaei, Ortrud Steinlein, Junjie Xiao, Ed.; Springer Singapore: 2017.

(156) Sears, V. F. Neutron scattering lengths and cross sections. *Neutron News* **1992**, *3* (3), 26-37. DOI: 10.1080/10448639208218770.

(157) Tong, L. Acetyl-coenzyme A carboxylase: crucial metabolic enzyme and attractive target for drug discovery. *Cell Mol Life Sci* **2005**, *62* (16), 1784-1803. DOI: 10.1007/s00018-005-5121-4 From NLM.

(158) Knowles, J. R. The Mechanism of Biotin-Dependent Enzymes. *Annual Review of Biochemistry* **1989**, *58* (1), 195-221. DOI: 10.1146/annurev.bi.58.070189.001211 (accessed 2022/12/09).

(159) White, S. W.; Zheng, J.; Zhang, Y.-M.; Rock, C. O. The Structural Biology of Type II Fatty Acid Biosynthesis. *Annual Review of Biochemistry* **2005**, *74* (1), 791-831. DOI: 10.1146/annurev.biochem.74.082803.133524 (accessed 2022/12/09).

(160) Moynié, L.; Leckie, S. M.; McMahon, S. A.; Duthie, F. G.; Koehnke, A.; Taylor, J. W.; Alphey, M. S.; Brenk, R.; Smith, A. D.; Naismith, J. H. Structural Insights into the Mechanism and Inhibition of the  $\beta$ -Hydroxydecanoyl-Acyl Carrier Protein Dehydratase from *Pseudomonas aeruginosa*. *Journal of Molecular Biology* **2013**, *425* (2), 365-377. DOI: <https://doi.org/10.1016/j.jmb.2012.11.017>.

(161) Lind, T. K.; Wacklin, H.; Schiller, J.; Moulin, M.; Haertlein, M.; Pomorski, T. G.; Cárdenas, M. Formation and Characterization of Supported Lipid Bilayers Composed of Hydrogenated and Deuterated *Escherichia coli* Lipids. *PLOS ONE* **2015**, *10* (12), e0144671. DOI: 10.1371/journal.pone.0144671.

(162) Semeraro, E. F.; Marx, L.; Frewein, M. P. K.; Pabst, G. Increasing complexity in small-angle X-ray and neutron scattering experiments: from biological membrane mimics to live cells. *Soft Matter* **2021**, *17* (2), 222-232, 10.1039/C9SM02352F. DOI: 10.1039/C9SM02352F.

(163) Haertlein, M.; Moulin, M.; Devos, J. M.; Laux, V.; Dunne, O.; Trevor Forsyth, V. Chapter Five - Biomolecular Deuteration for Neutron Structural Biology and Dynamics. In *Methods in Enzymology*, Kelman, Z. Ed.; Vol. 566; Academic Press, 2016; pp 113-157.

(164) Trehwella, J. Recent advances in small-angle scattering and its expanding impact in structural biology. *Structure* **2022**, *30* (1), 15-23. DOI: 10.1016/j.str.2021.09.008 (accessed 2024/08/07).

- (165) Duff, A. P.; Cagnes, M.; Darwish, T. A.; Krause-Heuer, A. M.; Moir, M.; Recsei, C.; Rekas, A.; Russell, R. A.; Wilde, K. L.; Yepuri, N. R. Chapter Three - Deuteration for biological SANS: Case studies, success and challenges in chemistry and biology. In *Methods in Enzymology*, Tainer, J. A. Ed.; Vol. 677; Academic Press, 2022; pp 85-126.
- (166) Okuda, A.; Inoue, R.; Morishima, K.; Saio, T.; Yunoki, Y.; Yagi-Utsumi, M.; Yagi, H.; Shimizu, M.; Sato, N.; Urade, R.; Kato, K.; Sugiyama, M. Deuteration Aiming for Neutron Scattering. *Biophys Physicobiol* **2021**, *18*, 16-27. DOI: 10.2142/biophysico.bppb-v18.003 From NLM.
- (167) Shah, R. S.; Senanayake, M.; Zhang, H.-H.; Pu, Y.; Biswal, A. K.; Pingali, S. V.; Davison, B.; O'Neill, H. Evidence for Lignin–Carbohydrate Complexes from Studies of Transgenic Switchgrass and a Model Lignin–Pectin Composite. *ACS Sustainable Chemistry & Engineering* **2023**, *11* (44), 15941-15950. DOI: 10.1021/acssuschemeng.3c04322.
- (168) Brady, N. G.; Qian, S.; Nguyen, J.; O'Neill, H. M.; Bruce, B. D. Small angle neutron scattering and lipidomic analysis of a native, trimeric PSI-SMALP from a thermophilic cyanobacteria. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2022**, *1863* (7), 148596. DOI: <https://doi.org/10.1016/j.bbabi.2022.148596>.
- (169) Thelen, J. L.; Leite, W.; Urban, V. S.; O'Neill, H. M.; Grishaev, A. V.; Curtis, J. E.; Krueger, S.; Castellanos, M. M. Morphological Characterization of Self-Amplifying mRNA Lipid Nanoparticles. *ACS Nano* **2024**, *18* (2), 1464-1476. DOI: 10.1021/acsnano.3c08014.
- (170) Leite, W. C.; Wu, Y.; Pingali, S. V.; Lieberman, R. L.; Urban, V. S. Change in Morphology of Dimyristoylphosphatidylcholine/Bile Salt Derivative Bicelle Assemblies with Dodecylmaltoside in the Disk and Ribbon Phases. *J Phys Chem Lett* **2022**, *13* (42), 9834-9840. DOI: 10.1021/acs.jpcl.2c02445 From NLM.
- (171) Bogojevic, O.; Nygaard, J. V.; Wiking, L.; Arevang, C.; Guo, Z. Designer phospholipids – structural retrieval, chemo-/bio- synthesis and isotopic labeling. *Biotechnology Advances* **2022**, *60*, 108025. DOI: <https://doi.org/10.1016/j.biotechadv.2022.108025>.
- (172) Thomas, G. M.; Wu, Y.; Leite, W.; Pingali, S. V.; Weiss, K. L.; Grant, A. J.; Diggs, M. W.; Schmidt-Krey, I.; Gutishvili, G.; Gumbart, J. C.; Urban, V. S.; Lieberman, R. L. SANS reveals lipid-dependent oligomerization of an intramembrane aspartyl protease from *H. volcanii*. *Biophys J* **2024**, *123* (13), 1846-1856. DOI: 10.1016/j.bpj.2024.05.029 From NLM.

- (173) Heller, W. T. Small-Angle Neutron Scattering for Studying Lipid Bilayer Membranes. In *Biomolecules*, 2022; Vol. 12.
- (174) Sebastiani, F.; Yanez Arteta, M.; Lerche, M.; Porcar, L.; Lang, C.; Bragg, R. A.; Elmore, C. S.; Krishnamurthy, V. R.; Russell, R. A.; Darwish, T.; Pichler, H.; Waldie, S.; Moulin, M.; Haertlein, M.; Forsyth, V. T.; Lindfors, L.; Cárdenas, M. Apolipoprotein E Binding Drives Structural and Compositional Rearrangement of mRNA-Containing Lipid Nanoparticles. *ACS Nano* **2021**, *15* (4), 6709-6722. DOI: 10.1021/acsnano.0c10064.
- (175) Natal'ya, A. B.; Vladimir, V. C. Methods of synthesis of deuterium-labelled lipids. *Russian Chemical Reviews* **1997**, *66* (11), 975. DOI: 10.1070/RC1997v066n11ABEH000369.
- (176) Yepuri, N. R.; Darwish, T. A.; Krause-Heuer, A. M.; Leung, A. E.; Delhom, R.; Wacklin, H. P.; Holden, P. J. Synthesis of Perdeuterated 1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine ([D(82)]POPC) and Characterisation of Its Lipid Bilayer Membrane Structure by Neutron Reflectometry. *Chempluschem* **2016**, *81* (3), 315-321. DOI: 10.1002/cplu.201500452 From NLM.
- (177) Bogojevic, O.; Leung, A. E. Enzyme-Assisted Synthesis of High-Purity, Chain-Deuterated 1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine. *ACS Omega* **2020**, *5* (35), 22395-22401. DOI: 10.1021/acsomega.0c02823.
- (178) Corucci, G.; Batchu, K. C.; Luchini, A.; Santamaria, A.; Frewein, M. P. K.; Laux, V.; Haertlein, M.; Yamaro-Botté, Y.; Botté, C. Y.; Sheridan, T.; Tully, M.; Maestro, A.; Martel, A.; Porcar, L.; Fragneto, G. Developing advanced models of biological membranes with hydrogenous and deuterated natural glycerophospholipid mixtures. *Journal of Colloid and Interface Science* **2023**, *645*, 870-881. DOI: <https://doi.org/10.1016/j.jcis.2023.04.135>.
- (179) Luchini, A.; Delhom, R.; Demé, B.; Laux, V.; Moulin, M.; Haertlein, M.; Pichler, H.; Strohmeier, G. A.; Wacklin, H.; Fragneto, G. The impact of deuteration on natural and synthetic lipids: A neutron diffraction study. *Colloids Surf B Biointerfaces* **2018**, *168*, 126-133. DOI: 10.1016/j.colsurfb.2018.02.009 From NLM.
- (180) Luchini, A.; Delhom, R.; Cristiglio, V.; Knecht, W.; Wacklin-Knecht, H.; Fragneto, G. Effect of ergosterol on the interlamellar spacing of deuterated yeast phospholipid multilayers. *Chemistry and Physics of Lipids* **2020**, *227*, 104873. DOI: <https://doi.org/10.1016/j.chemphyslip.2020.104873>.

- (181) Waldie, S.; Lind, T. K.; Browning, K.; Moulin, M.; Haertlein, M.; Forsyth, V. T.; Luchini, A.; Strohmeier, G. A.; Pichler, H.; Maric, S.; Cárdenas, M. Localization of Cholesterol within Supported Lipid Bilayers Made of a Natural Extract of Tailor-Deuterated Phosphatidylcholine. *Langmuir* **2018**, *34* (1), 472-479. DOI: 10.1021/acs.langmuir.7b02716 From NLM.
- (182) Aktas, M.; Narberhaus, F. In vitro characterization of the enzyme properties of the phospholipid N-methyltransferase PmtA from *Agrobacterium tumefaciens*. *J Bacteriol* **2009**, *191* (7), 2033-2041. DOI: 10.1128/jb.01591-08 From NLM.
- (183) Kleetz, J.; Vasilopoulos, G.; Czolkoss, S.; Aktas, M.; Narberhaus, F. Recombinant and endogenous ways to produce methylated phospholipids in *Escherichia coli*. *Appl Microbiol Biotechnol* **2021**, *105* (23), 8837-8851. DOI: 10.1007/s00253-021-11654-8 From NLM.
- (184) de Ghellinck, A.; Schaller, H.; Laux, V.; Haertlein, M.; Sferrazza, M.; Maréchal, E.; Wacklin, H.; Jouhet, J.; Fragneto, G. Production and Analysis of Perdeuterated Lipids from *Pichia pastoris* Cells. *PLOS ONE* **2014**, *9* (4), e92999. DOI: 10.1371/journal.pone.0092999.
- (185) Neidhardt, F. C.; Curtiss, R. *Escherichia Coli and Salmonella: Cellular and Molecular Biology*, 2nd ed. ed.; ASM Press, 1996.
- (186) van Meer, G.; Voelker, D. R.; Feigenson, G. W. Membrane lipids: where they are and how they behave. *Nat Rev Mol Cell Biol* **2008**, *9* (2), 112-124. DOI: 10.1038/nrm2330 From NLM.
- (187) Harayama, T.; Riezman, H. Understanding the diversity of membrane lipid composition. *Nature Reviews Molecular Cell Biology* **2018**, *19* (5), 281-296. DOI: 10.1038/nrm.2017.138.
- (188) Hwangbo, M.; Chu, K.-H. Recent advances in production and extraction of bacterial lipids for biofuel production. *Science of The Total Environment* **2020**, *734*, 139420. DOI: <https://doi.org/10.1016/j.scitotenv.2020.139420>.
- (189) Glüge, J.; Scheringer, M.; Cousins, I. T.; DeWitt, J. C.; Goldenman, G.; Herzke, D.; Lohmann, R.; Ng, C. A.; Trier, X.; Wang, Z. An overview of the uses of per- and polyfluoroalkyl substances (PFAS). *Environmental Science: Processes & Impacts* **2020**, *22* (12), 2345-2373, 10.1039/D0EM00291G. DOI: 10.1039/D0EM00291G.
- (190) Hu, X. C.; Andrews, D. Q.; Lindstrom, A. B.; Bruton, T. A.; Schaider, L. A.; Grandjean, P.; Lohmann, R.; Carignan, C. C.; Blum, A.; Balan, S. A.; Higgins, C. P.; Sunderland,

E. M. Detection of Poly- and Perfluoroalkyl Substances (PFASs) in U.S. Drinking Water Linked to Industrial Sites, Military Fire Training Areas, and Wastewater Treatment Plants.

*Environmental Science & Technology Letters* **2016**, 3 (10), 344-350. DOI: 10.1021/acs.estlett.6b00260.

(191) Sunderland, E. M.; Hu, X. C.; Dassuncao, C.; Tokranov, A. K.; Wagner, C. C.; Allen, J. G. A review of the pathways of human exposure to poly- and perfluoroalkyl substances (PFASs) and present understanding of health effects. *J Expo Sci Environ Epidemiol* **2019**, 29 (2), 131-147. DOI: 10.1038/s41370-018-0094-1 From NLM.

(192) Fenton, S. E.; Ducatman, A.; Boobis, A.; DeWitt, J. C.; Lau, C.; Ng, C.; Smith, J. S.; Roberts, S. M. Per- and Polyfluoroalkyl Substance Toxicity and Human Health Review: Current State of Knowledge and Strategies for Informing Future Research. *Environ Toxicol Chem* **2021**, 40 (3), 606-630. DOI: 10.1002/etc.4890 From NLM.

(193) Beans, C. How “forever chemicals” might impair the immune system. *Proceedings of the National Academy of Sciences* **2021**, 118 (15), e2105018118. DOI: 10.1073/pnas.2105018118 (accessed 2024/08/19).

(194) Cousins, I. T.; Johansson, J. H.; Salter, M. E.; Sha, B.; Scheringer, M. Outside the Safe Operating Space of a New Planetary Boundary for Per- and Polyfluoroalkyl Substances (PFAS). *Environmental Science & Technology* **2022**, 56 (16), 11172-11179. DOI: 10.1021/acs.est.2c02765.

(195) Dinglasan, M. J.; Ye, Y.; Edwards, E. A.; Mabury, S. A. Fluorotelomer alcohol biodegradation yields poly- and perfluorinated acids. *Environ Sci Technol* **2004**, 38 (10), 2857-2864. DOI: 10.1021/es0350177 From NLM.

(196) Kim, M. H.; Wang, N.; Chu, K. H. 6:2 Fluorotelomer alcohol (6:2 FTOH) biodegradation by multiple microbial species under different physiological conditions. *Appl Microbiol Biotechnol* **2014**, 98 (4), 1831-1840. DOI: 10.1007/s00253-013-5131-3 From NLM.

(197) Hall, D. R.; Gauthier, J.; Peng, H. Querying the In Vitro Proteome Cysteine Reactivity of 8:2 Fluorotelomer Acrylate. *Environmental Science & Technology* **2023**, 57 (35), 13015-13024. DOI: 10.1021/acs.est.3c02930.

(198) Xie, Y.; May, A. L.; Chen, G.; Brown, L. P.; Powers, J. B.; Tague, E. D.; Campagna, S. R.; Löffler, F. E. *Pseudomonas* sp. Strain 273 Incorporates Organofluorine into the

Lipid Bilayer during Growth with Fluorinated Alkanes. *Environmental Science & Technology* **2022**, 56 (12), 8155-8166. DOI: 10.1021/acs.est.2c01454.

(199) Dasu, K.; Xia, X.; Siriwardena, D.; Klupinski, T. P.; Seay, B. Concentration profiles of per- and polyfluoroalkyl substances in major sources to the environment. *Journal of Environmental Management* **2022**, 301, 113879. DOI: <https://doi.org/10.1016/j.jenvman.2021.113879>.

(200) Zhang, W.; Pang, S.; Lin, Z.; Mishra, S.; Bhatt, P.; Chen, S. Biotransformation of perfluoroalkyl acid precursors from various environmental systems: advances and perspectives. *Environmental Pollution* **2021**, 272, 115908. DOI: <https://doi.org/10.1016/j.envpol.2020.115908>.

(201) Clasquin, M. F.; Melamud, E.; Rabinowitz, J. D. LC-MS data processing with MAVEN: a metabolomic analysis and visualization engine. *Curr Protoc Bioinformatics* **2012**, Chapter 14. DOI: 10.1002/0471250953.bi1411s37 From NLM.

(202) Lauria, M. Z.; Sepman, H.; Ledbetter, T.; Plassmann, M.; Roos, A. M.; Simon, M.; Benskin, J. P.; Krueve, A. Closing the Organofluorine Mass Balance in Marine Mammals Using Suspect Screening and Machine Learning-Based Quantification. *Environmental Science & Technology* **2024**, 58 (5), 2458-2467. DOI: 10.1021/acs.est.3c07220.

(203) Chen, X.; Qadeer, A.; Liu, M.; Deng, L.; Zhou, P.; Mwizerwa, I. T.; Liu, S.; Ajmal, Z.; Xingru, Z.; Jiang, X. Chapter 13 - Bioaccumulation of emerging contaminants in aquatic biota: PFAS as a case study. In *Emerging Aquatic Contaminants*, Kumar, M., Mohapatra, S., Weber, K. Eds.; Elsevier, 2023; pp 347-374.

(204) Giesy, J.; Kannan, K. Peer Reviewed: Perfluorochemical Surfactants in the Environment. *Environmental Science & Technology* **2002**, 36 (7), 146A-152A. DOI: 10.1021/es022253t.

(205) Raetz, C. R.; Dowhan, W. Biosynthesis and function of phospholipids in *Escherichia coli*. *J Biol Chem* **1990**, 265 (3), 1235-1238. From NLM.

(206) Rühl, J.; Hein, E. M.; Hayen, H.; Schmid, A.; Blank, L. M. The glycerophospholipid inventory of *Pseudomonas putida* is conserved between strains and enables growth condition-related alterations. *Microb Biotechnol* **2012**, 5 (1), 45-58. DOI: 10.1111/j.1751-7915.2011.00286.x From NLM.

(207) Neidhardt, F. C.; Ingraham, J. L.; Schaechter, M. *Physiology of the Bacterial Cell: A Molecular Approach*, Sinauer Associates, 1990.

(208) Loferer-Krössbacher, M.; Klima, J.; Psenner, R. Determination of bacterial cell dry mass by transmission electron microscopy and densitometric image analysis. *Appl Environ Microbiol* **1998**, *64* (2), 688-694. DOI: 10.1128/aem.64.2.688-694.1998 From NLM.

(209) Glazyrina, J.; Materne, E.-M.; Dreher, T.; Storm, D.; Junne, S.; Adams, T.; Greller, G.; Neubauer, P. High cell density cultivation and recombinant protein production with *Escherichia coli* in a rocking-motion-type bioreactor. *Microbial Cell Factories* **2010**, *9* (1), 42. DOI: 10.1186/1475-2859-9-42.

(210) Xie, Y.; Ramirez, D.; Chen, G.; He, G.; Sun, Y.; Murdoch, F. K.; Löffler, F. E. Genome-Wide Expression Analysis Unravels Fluoroalkane Metabolism in *Pseudomonas* sp. Strain 273. *Environmental Science & Technology* **2023**, *57* (42), 15925-15935. DOI: 10.1021/acs.est.3c03855.

(211) Hsu, F.-F.; Turk, J. Charge-remote and charge-driven fragmentation processes in diacyl glycerophosphoethanolamine upon low-energy collisional activation: a mechanistic proposal. *Journal of the American Society for Mass Spectrometry* **2000**, *11* (10), 892-899. DOI: [https://doi.org/10.1016/S1044-0305\(00\)00159-8](https://doi.org/10.1016/S1044-0305(00)00159-8).

(212) Chen, Y. F.; Liu, T.; Hu, L. X.; Chen, C. E.; Yang, B.; Ying, G. G. Unveiling per- and polyfluoroalkyl substance contamination in Chinese paper products and assessing their exposure risk. *Environ Int* **2024**, *185*, 108540. DOI: 10.1016/j.envint.2024.108540 From NLM.

(213) Strynar, M.; McCord, J.; Newton, S.; Washington, J.; Barzen-Hanson, K.; Trier, X.; Liu, Y.; Dimzon, I. K.; Bugsel, B.; Zwiener, C.; Munoz, G. Practical application guide for the discovery of novel PFAS in environmental samples using high resolution mass spectrometry. *Journal of Exposure Science & Environmental Epidemiology* **2023**, *33* (4), 575-588. DOI: 10.1038/s41370-023-00578-2.

(214) Gremmel, C.; Frömel, T.; Knepper, T. P. HPLC–MS/MS methods for the determination of 52 perfluoroalkyl and polyfluoroalkyl substances in aqueous samples. *Analytical and Bioanalytical Chemistry* **2017**, *409* (6), 1643-1655. DOI: 10.1007/s00216-016-0110-z.

(215) Mothersole, R. G.; Mothersole, M. K.; Goddard, H. G.; Liu, J.; Van Hamme, J. D. Enzyme Catalyzed Formation of CoA Adducts of Fluorinated Hexanoic Acid Analogues using a Long-Chain acyl-CoA Synthetase from *Gordonia* sp. Strain NB4-1Y. *Biochemistry* **2024**, *63* (17), 2153-2165. DOI: 10.1021/acs.biochem.4c00336.

(216) Neidhardt, F. C. *Escherichia coli and Salmonella: Cellular and Molecular Biology*, ASM Press, 1996.

(217) Loferer-Krossbacher, M.; Klima, J.; Psenner, R. Determination of bacterial cell dry mass by transmission electron microscopy and densitometric image analysis. *Appl. Environ. Microbiol.* **1998**, *64* (2), 688-694.

(218) Glazyrina, J.; Materne, E.-M.; Dreher, T.; Storm, D.; Junne, S.; Adams, T.; Greller, G.; Neubauer, P. High cell density cultivation and recombinant protein production with *Escherichia coli* in a rocking-motion-type bioreactor. *Microb. Cell Factories* **2010**, *9*, 1-11.

(219) Zhang, Y.; Zhou, Y.; Dong, R.; Song, N.; Hong, M.; Li, J.; Yu, J.; Kong, D. Emerging and legacy per- and polyfluoroalkyl substances (PFAS) in fluorochemical wastewater along full-scale treatment processes: Source, fate, and ecological risk. *Journal of Hazardous Materials* **2024**, *465*, 133270. DOI: <https://doi.org/10.1016/j.jhazmat.2023.133270>.

(220) Ng, C. A.; Hungerbühler, K. Bioaccumulation of Perfluorinated Alkyl Acids: Observations and Models. *Environmental Science & Technology* **2014**, *48* (9), 4637-4648. DOI: 10.1021/es404008g.

(221) Fitzgerald, N. J. M.; Wargenau, A.; Sorenson, C.; Pedersen, J.; Tufenkji, N.; Novak, P. J.; Simcik, M. F. Partitioning and Accumulation of Perfluoroalkyl Substances in Model Lipid Bilayers and Bacteria. *Environmental Science & Technology* **2018**, *52* (18), 10433-10440. DOI: 10.1021/acs.est.8b02912.

(222) Forsthuber, M.; Kaiser, A. M.; Granitzer, S.; Hassl, I.; Hengstschläger, M.; Stangl, H.; Gundacker, C. Albumin is the major carrier protein for PFOS, PFOA, PFHxS, PFNA and PFDA in human plasma. *Environ Int* **2020**, *137*, 105324. DOI: 10.1016/j.envint.2019.105324 From NLM.

(223) Sundström, M.; Ehresman, D. J.; Bignert, A.; Butenhoff, J. L.; Olsen, G. W.; Chang, S. C.; Bergman, A. A temporal trend study (1972-2008) of perfluorooctanesulfonate, perfluorohexanesulfonate, and perfluorooctanoate in pooled human milk samples from Stockholm, Sweden. *Environ Int* **2011**, *37* (1), 178-183. DOI: 10.1016/j.envint.2010.08.014 From NLM.

(224) Wang, Q.; Ruan, Y.; Jin, L.; Zhang, X.; Li, J.; He, Y.; Wei, S.; Lam, J. C. W.; Lam, P. K. S. Target, Nontarget, and Suspect Screening and Temporal Trends of Per- and

Polyfluoroalkyl Substances in Marine Mammals from the South China Sea. *Environmental Science & Technology* **2021**, 55 (2), 1045-1056. DOI: 10.1021/acs.est.0c06685.

(225) De Silva, A. O.; Armitage, J. M.; Bruton, T. A.; Dassuncao, C.; Heiger-Bernays, W.; Hu, X. C.; Kärrman, A.; Kelly, B.; Ng, C.; Robuck, A.; Sun, M.; Webster, T. F.; Sunderland, E. M. PFAS Exposure Pathways for Humans and Wildlife: A Synthesis of Current Knowledge and Key Gaps in Understanding. *Environ Toxicol Chem* **2021**, 40 (3), 631-657. DOI: 10.1002/etc.4935 From NLM.

(226) Yeung, L. W. Y.; Miyake, Y.; Taniyasu, S.; Wang, Y.; Yu, H.; So, M. K.; Jiang, G.; Wu, Y.; Li, J.; Giesy, J. P.; Yamashita, N.; Lam, P. K. S. Perfluorinated Compounds and Total and Extractable Organic Fluorine in Human Blood Samples from China. *Environmental Science & Technology* **2008**, 42 (21), 8140-8145. DOI: 10.1021/es800631n.

(227) Spaan, K. M.; van Noordenburg, C.; Plassmann, M. M.; Schultes, L.; Shaw, S.; Berger, M.; Heide-Jørgensen, M. P.; Rosing-Asvid, A.; Granquist, S. M.; Dietz, R.; Sonne, C.; Rigét, F.; Roos, A.; Benskin, J. P. Fluorine Mass Balance and Suspect Screening in Marine Mammals from the Northern Hemisphere. *Environmental Science & Technology* **2020**, 54 (7), 4046-4058. DOI: 10.1021/acs.est.9b06773.

(228) Yeung, L. W. Y.; De Silva, A. O.; Loi, E. I. H.; Marvin, C. H.; Taniyasu, S.; Yamashita, N.; Mabury, S. A.; Muir, D. C. G.; Lam, P. K. S. Perfluoroalkyl substances and extractable organic fluorine in surface sediments and cores from Lake Ontario. *Environment International* **2013**, 59, 389-397. DOI: <https://doi.org/10.1016/j.envint.2013.06.026>.

(229) Schwichtenberg, T.; Bogdan, D.; Schaefer, C. E.; Stults, J.; Field, J. A. Per- and Polyfluoroalkyl Substances Enrichment in the Surface Microlayer of a Freshwater System Impacted by Aqueous Film-Forming Foams. *ACS ES&T Water* **2023**, 3 (4), 1150-1160. DOI: 10.1021/acsestwater.2c00618.

(230) Houtz, E. F.; Higgins, C. P.; Field, J. A.; Sedlak, D. L. Persistence of Perfluoroalkyl Acid Precursors in AFFF-Impacted Groundwater and Soil. *Environmental Science & Technology* **2013**, 47 (15), 8187-8195. DOI: 10.1021/es4018877.

(231) Griebler, C.; Lueders, T. Microbial biodiversity in groundwater ecosystems. *Freshwater Biology* **2009**, 54 (4), 649-677. DOI: <https://doi.org/10.1111/j.1365-2427.2008.02013.x> (accessed 2024/08/19).

- (232) Ljungdhal, L. G. The Autotrophic Pathway of Acetate Synthesis in Acetogenic Bacteria. *Annual Review of Microbiology* **1986**, *40* (1), 415-450. DeepDyve.
- (233) Liew, F. E.; et al. Carbon-negative production of acetone and isopropanol by gas fermentation at industrial pilot scale. *Nature Biotechnology* **2022**, *40* (3), 335-344. DOI: 10.1038/s41587-021-01195-w.
- (234) Jones, D. T.; Schulz, F.; Roux, S.; Brown, S. D. Solvent-Producing Clostridia Revisited. In *Microorganisms*, 2023; Vol. 11.
- (235) Rathmann, R.; Szklo, A.; Schaeffer, R. Land use competition for production of food and liquid biofuels: An analysis of the arguments in the current debate. *Renewable energy* **2010**, *35* (1), 14-22.
- (236) Fackler, N.; Heijstra, B. D.; Rasor, B. J.; Brown, H.; Martin, J.; Ni, Z.; Shebek, K. M.; Rosin, R. R.; Simpson, S. D.; Tyo, K. E.; Giannone, R. J.; Hettich, R. L.; Tschaplinski, T. J.; Leang, C.; Brown, S. D.; Jewett, M. C.; Köpke, M. Stepping on the Gas to a Circular Economy: Accelerating Development of Carbon-Negative Chemical Production from Gas Fermentation. *Annu Rev Chem Biomol Eng* **2021**, *12*, 439-470. DOI: 10.1146/annurev-chembioeng-120120-021122 From NLM.
- (237) Pavan, M.; Reinmets, K.; Garg, S.; Mueller, A. P.; Marcellin, E.; Köpke, M.; Valgepea, K. Advances in systems metabolic engineering of autotrophic carbon oxide-fixing biocatalysts towards a circular economy. *Metabolic Engineering* **2022**, *71*, 117-141. DOI: <https://doi.org/10.1016/j.ymben.2022.01.015>.
- (238) Mohammadi, M.; Najafpour, G. D.; Younesi, H.; Lahijani, P.; Uzir, M. H.; Mohamed, A. R. Bioconversion of synthesis gas to second generation biofuels: A review. *Renewable and Sustainable Energy Reviews* **2011**, *15* (9), 4255-4273. DOI: <https://doi.org/10.1016/j.rser.2011.07.124>.
- (239) Kundiya, D. K.; Wilkins, M. R.; Maddipati, P.; Huhnke, R. L. Effect of temperature, pH and buffer presence on ethanol production from synthesis gas by "Clostridium ragsdalei". *Bioresour Technol* **2011**, *102* (10), 5794-5799. DOI: 10.1016/j.biortech.2011.02.032 From NLM.
- (240) Ramió-Pujol, S.; Ganigué, R.; Bañeras, L.; Colprim, J. Incubation at 25°C prevents acid crash and enhances alcohol production in Clostridium carboxidivorans P7. *Bioresour Technol* **2015**, *192*, 296-303. DOI: <https://doi.org/10.1016/j.biortech.2015.05.077>.

(241) Shen, S.; Gu, Y.; Chai, C.; Jiang, W.; Zhuang, Y.; Wang, Y. Enhanced alcohol titre and ratio in carbon monoxide-rich off-gas fermentation of *Clostridium carboxidivorans* through combination of trace metals optimization with variable-temperature cultivation. *Bioresour Technol* **2017**, *239*, 236-243. DOI: <https://doi.org/10.1016/j.biortech.2017.04.099>.

(242) McNeil, B.; Kristiansen, B. Effect of temperature upon growth rate and solvent production in batch cultures of *Clostridium acetobutylicum*. *Biotechnology Letters* **1985**, *7* (7), 499-502. DOI: 10.1007/BF01199868.

(243) Zhang, J.; Taylor, S.; Wang, Y. Effects of end products on fermentation profiles in *Clostridium carboxidivorans* P7 for syngas fermentation. *Bioresour Technol* **2016**, *218*, 1055-1063. DOI: 10.1016/j.biortech.2016.07.071 From NLM.

(244) Maddox, I. S.; Steiner, E.; Hirsch, S.; Wessner, S.; Gutierrez, N. A.; Gapes, J. R.; Schuster, K. C. The cause of "acid-crash" and "acidogenic fermentations" during the batch acetone-butanol-ethanol (ABE-) fermentation process. *J Mol Microbiol Biotechnol* **2000**, *2* (1), 95-100. From NLM.

(245) Wang, S.; Zhang, Y.; Dong, H.; Mao, S.; Zhu, Y.; Wang, R.; Luan, G.; Li, Y. Formic acid triggers the "Acid Crash" of acetone-butanol-ethanol fermentation by *Clostridium acetobutylicum*. *Appl Environ Microbiol* **2011**, *77* (5), 1674-1680. DOI: 10.1128/aem.01835-10 From NLM.

(246) Grethlein, A. J.; Worden, R. M.; Jain, M. K.; Datta, R. Continuous production of mixed alcohols and acids from carbon monoxide. *Applied Biochemistry and Biotechnology* **1990**, *24* (1), 875-884. DOI: 10.1007/BF02920301.

(247) Phillips, J. R.; Huhnke, R. L.; Atiyeh, H. K. Syngas Fermentation: A Microbial Conversion Process of Gaseous Substrates to Various Products. In *Fermentation*, 2017; Vol. 3.

(248) Valgepea, K.; de Souza Pinto Lemgruber, R.; Meaghan, K.; Palfreyman, R. W.; Abdalla, T.; Heijstra, B. D.; Behrendorff, J. B.; Tappel, R.; Köpke, M.; Simpson, S. D.; Nielsen, L. K.; Marcellin, E. Maintenance of ATP Homeostasis Triggers Metabolic Shifts in Gas-Fermenting Acetogens. *Cell Syst* **2017**, *4* (5), 505-515.e505. DOI: 10.1016/j.cels.2017.04.008 From NLM.

(249) Batth, T. S.; Tollenaere, M. X.; Rüther, P.; Gonzalez-Franquesa, A.; Prabhakar, B. S.; Bekker-Jensen, S.; Deshmukh, A. S.; Olsen, J. V. Protein Aggregation Capture on

Microparticles Enables Multipurpose Proteomics Sample Preparation. *Mol Cell Proteomics* **2019**, *18* (5), 1027-1035. DOI: 10.1074/mcp.TIR118.001270 From NLM.

(250) Brown, S. D.; Nagaraju, S.; Utturkar, S.; De Tissera, S.; Segovia, S.; Mitchell, W.; Land, M. L.; Dassanayake, A.; Köpke, M. Comparison of single-molecule sequencing and hybrid approaches for finishing the genome of *Clostridium autoethanogenum* and analysis of CRISPR systems in industrial relevant Clostridia. *Biotechnol Biofuels* **2014**, *7*, 40. DOI: 10.1186/1754-6834-7-40 From NLM.

(251) Humphreys, C. M.; McLean, S.; Schatschneider, S.; Millat, T.; Henstra, A. M.; Annan, F. J.; Breitkopf, R.; Pander, B.; Piatek, P.; Rowe, P.; Wichlacz, A. T.; Woods, C.; Norman, R.; Blom, J.; Goesman, A.; Hodgman, C.; Barrett, D.; Thomas, N. R.; Winzer, K.; Minton, N. P. Whole genome sequence and manual annotation of *Clostridium autoethanogenum*, an industrially relevant bacterium. *BMC Genomics* **2015**, *16* (1), 1085. DOI: 10.1186/s12864-015-2287-5.

(252) Huang, Y.; Niu, B.; Gao, Y.; Fu, L.; Li, W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **2010**, *26* (5), 680-682. DOI: 10.1093/bioinformatics/btq003 (accessed 9/16/2024).

(253) Taverner, T.; Karpievitch, Y. V.; Polpitiya, A. D.; Brown, J. N.; Dabney, A. R.; Anderson, G. A.; Smith, R. D. DanteR: an extensible R-based tool for quantitative analysis of -omics data. *Bioinformatics* **2012**, *28* (18), 2404-2406. DOI: 10.1093/bioinformatics/bts449 (accessed 9/16/2024).

(254) Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M. Y.; Geiger, T.; Mann, M.; Cox, J. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* **2016**, *13* (9), 731-740. DOI: 10.1038/nmeth.3901 From NLM.

(255) Kanehisa, M.; Sato, Y.; Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol* **2016**, *428* (4), 726-731. DOI: 10.1016/j.jmb.2015.11.006 From NLM.

(256) Wu, T.; Hu, E.; Xu, S.; Chen, M.; Guo, P.; Dai, Z.; Feng, T.; Zhou, L.; Tang, W.; Zhan, L.; Fu, X.; Liu, S.; Bo, X.; Yu, G. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2021**, *2* (3), 100141. DOI: 10.1016/j.xinn.2021.100141 From NLM.

- (257) Willey, J. M.; M., S. L.; Woolverton, C. J. *Prescott's Microbiology*, 9th ed.; McGraw-Hill, 2014.
- (258) Liu, Y.; Zhang, Z.; Jiang, W.; Gu, Y. Protein acetylation-mediated cross regulation of acetic acid and ethanol synthesis in the gas-fermenting *Clostridium ljungdahlii*. *J Biol Chem* **2022**, 298 (2), 101538. DOI: 10.1016/j.jbc.2021.101538 From NLM.
- (259) Guan, Z.; Goldfine, H. Lipid diversity in clostridia. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **2021**, 1866 (9), 158966. DOI: <https://doi.org/10.1016/j.bbalip.2021.158966>.
- (260) Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research* **2017**, 45 (D1), D353-D361. DOI: 10.1093/nar/gkw1092 (accessed 9/16/2024).
- (261) Karp, P. D.; Billington, R.; Caspi, R.; Fulcher, C. A.; Latendresse, M.; Kothari, A.; Keseler, I. M.; Krummenacker, M.; Midford, P. E.; Ong, Q.; Ong, W. K.; Paley, S. M.; Subhraveti, P. The BioCyc collection of microbial genomes and metabolic pathways. *Briefings in Bioinformatics* **2019**, 20 (4), 1085-1093. DOI: 10.1093/bib/bbx085 (accessed 9/16/2024).
- (262) Goldfine, H. The anaerobic biosynthesis of plasmalogens. *FEBS Letters* **2017**, 591 (18), 2714-2719. DOI: <https://doi.org/10.1002/1873-3468.12714> (accessed 2024/09/16).
- (263) Jackson, D. R.; Cassilly, C. D.; Plichta, D. R.; Vlamakis, H.; Liu, H.; Melville, S. B.; Xavier, R. J.; Clardy, J. Plasmalogen Biosynthesis by Anaerobic Bacteria: Identification of a Two-Gene Operon Responsible for Plasmalogen Production in *Clostridium perfringens*. *ACS Chem Biol* **2021**, 16 (1), 6-13. DOI: 10.1021/acscchembio.0c00673 From NLM.
- (264) Sohlenkamp, C.; Geiger, O. Bacterial membrane lipids: diversity in structures and pathways. *FEMS Microbiol Rev* **2016**, 40 (1), 133-159. DOI: 10.1093/femsre/fuv008 From NLM.
- (265) Sastre, D. E.; Pulschen, A. A.; Basso, L. G. M.; Benites Pariente, J. S.; Marques Netto, C. G. C.; Machinandiarena, F.; Albanesi, D.; Navarro, M. V. A. S.; de Mendoza, D.; Gueiros-Filho, F. J. The phosphatidic acid pathway enzyme PlsX plays both catalytic and channeling roles in bacterial phospholipid synthesis. *Journal of Biological Chemistry* **2020**, 295 (7), 2148-2159. DOI: <https://doi.org/10.1074/jbc.RA119.011147>.
- (266) Annan, F. J.; Al-Sinawi, B.; Humphreys, C. M.; Norman, R.; Winzer, K.; Köpke, M.; Simpson, S. D.; Minton, N. P.; Henstra, A. M. Engineering of vitamin prototrophy in

*Clostridium ljungdahlii* and *Clostridium autoethanogenum*. *Applied Microbiology and Biotechnology* **2019**, *103* (11), 4633-4648. DOI: 10.1007/s00253-019-09763-6.

(267) Axen, S. D.; Erbilgin, O.; Kerfeld, C. A. A Taxonomy of Bacterial Microcompartment Loci Constructed by a Novel Scoring Method. *PLOS Computational Biology* **2014**, *10* (10), e1003898. DOI: 10.1371/journal.pcbi.1003898.

(268) Zarzycki, J.; Erbilgin, O.; Kerfeld Cheryl, A. Bioinformatic Characterization of Glycyl Radical Enzyme-Associated Bacterial Microcompartments. *Applied and Environmental Microbiology* **2015**, *81* (24), 8315-8329. DOI: 10.1128/AEM.02587-15 (accessed 2024/09/16).

(269) Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* **2004**, *101* (26), 9528-9533. DOI: 10.1073/pnas.0402700101 From NLM.

(270) *Complete structural elucidation of lipids in a single experiment using electron activated dissociation (EAD)*; Sciex, 2021. <https://sciex.com/tech-notes/life-science-research/lipidomics/complete-structural-elucidation-of-lipids-in-a-single-experiment0>.

(271) Fanelli, D. Is science really facing a reproducibility crisis, and do we need it to? *Proceedings of the National Academy of Sciences* **2018**, *115* (11), 2628-2631.

(272) Miyakawa, T. No raw data, no science: another possible source of the reproducibility crisis. *Molecular Brain* **2020**, *13* (1), 24. DOI: 10.1186/s13041-020-0552-2.

(273) Peng, R. The Reproducibility Crisis in Science: A Statistical Counterattack. *Significance* **2015**, *12* (3), 30-32. DOI: 10.1111/j.1740-9713.2015.00827.x (accessed 10/15/2024).

(274) Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **2016**, *533* (7604), 452-454. DOI: 10.1038/533452a.

(275) Errington, T. M.; Mathur, M.; Soderberg, C. K.; Denis, A.; Perfito, N.; Iorns, E.; Nosek, B. A. Investigating the replicability of preclinical cancer biology. *eLife* **2021**, *10*, e71601. DOI: 10.7554/eLife.71601.

(276) Kaelin Jr, W. G. Publish houses of brick, not mansions of straw. *Nature* **2017**, *545* (7655), 387-387. DOI: 10.1038/545387a.

(277) Lawrence, P. A. The mismeasurement of science. *Current Biology* **2007**, *17* (15), R583-R585. DOI: <https://doi.org/10.1016/j.cub.2007.06.014>.

- (278) Lawrence, P. A. Lost in publication: how measurement harms science. *Ethics in Science and Environmental Politics* **2008**, 8 (1), 9-11.
- (279) Adler, N. J.; Harzing, A.-W. When Knowledge Wins: Transcending the Sense and Nonsense of Academic Rankings. *Academy of Management Learning & Education* **2009**, 8 (1), 72-95. DOI: 10.5465/amle.2009.37012181 (accessed 2024/10/10).
- (280) Macdonald, S.; Kam, J. Aardvark et al.: quality journals and gamesmanship in management studies. *Journal of Information Science* **2007**, 33 (6), 702-717. DOI: 10.1177/0165551507077419 (accessed 2024/10/10).
- (281) Grimes, D. R.; Bauch, C. T.; Ioannidis, J. P. A. Modelling science trustworthiness under publish or perish pressure. *Royal Society Open Science* **2018**, 5 (171511). DOI: <https://doi.org/10.1098/rsos.171511>.
- (282) Errington, T. M.; Denis, A.; Perfito, N.; Iorns, E.; Nosek, B. A. Challenges for assessing replicability in preclinical cancer biology. *eLife* **2021**, 10, e67995. DOI: 10.7554/eLife.67995.
- (283) White, T.; Bursten, S.; Federighi, D.; Lewis, R. A.; Nudelman, E. High-Resolution Separation and Quantification of Neutral Lipid and Phospholipid Species in Mammalian Cells and Sera by Multi-One-Dimensional Thin-Layer Chromatography. *Analytical Biochemistry* **1998**, 258 (1), 109-117. DOI: <https://doi.org/10.1006/abio.1997.2545>.
- (284) Wang, K.; Zhao, Y.; Xu, L.; Liao, X.; Xu, Z. Health outcomes of fruit juice and fruity beverage: a comparative analysis of gut microbiota and metabolomics in rats. 2022.
- (285) Marra, J. Direct measurement of the interaction between phosphatidylglycerol bilayers in aqueous electrolyte solutions. *Biophysical Journal* **1986**, 50 (5), 815-825. DOI: [https://doi.org/10.1016/S0006-3495\(86\)83522-6](https://doi.org/10.1016/S0006-3495(86)83522-6).
- (286) Armen, R. S.; Uitto, O. D.; Feller, S. E. Phospholipid Component Volumes: Determination and Application to Bilayer Structure Calculations. *Biophysical Journal* **1998**, 75 (2), 734-744. DOI: [https://doi.org/10.1016/S0006-3495\(98\)77563-0](https://doi.org/10.1016/S0006-3495(98)77563-0).
- (287) 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. DOI: 10.1021/acs.analchem.7b01709.



## Appendix

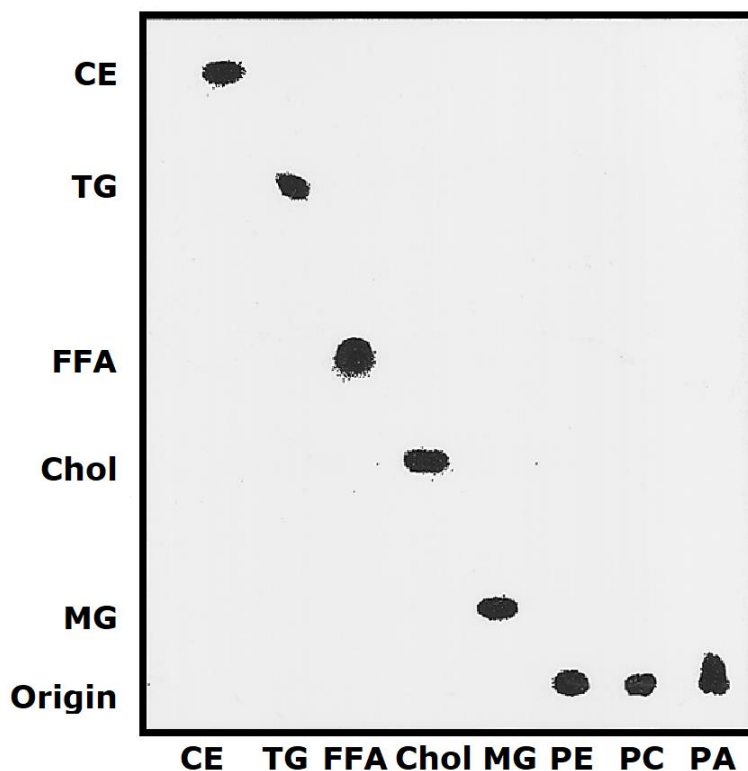


**Figure 1-1. Representative 2-D Lipid TLC.**

Lipid class abbreviations as follows: CE: cholesteryl ester, TG: triacylglycerol, FFA: free fatty acids, Chol: cholesterol, MG: monoacylglycerol, PE: phosphatidylethanolamine, PC: phosphatidylcholine, PA: phosphatidic acid. Figure is reproduced with marginal adaptation from article<sup>283</sup> with permission from Elsevier.

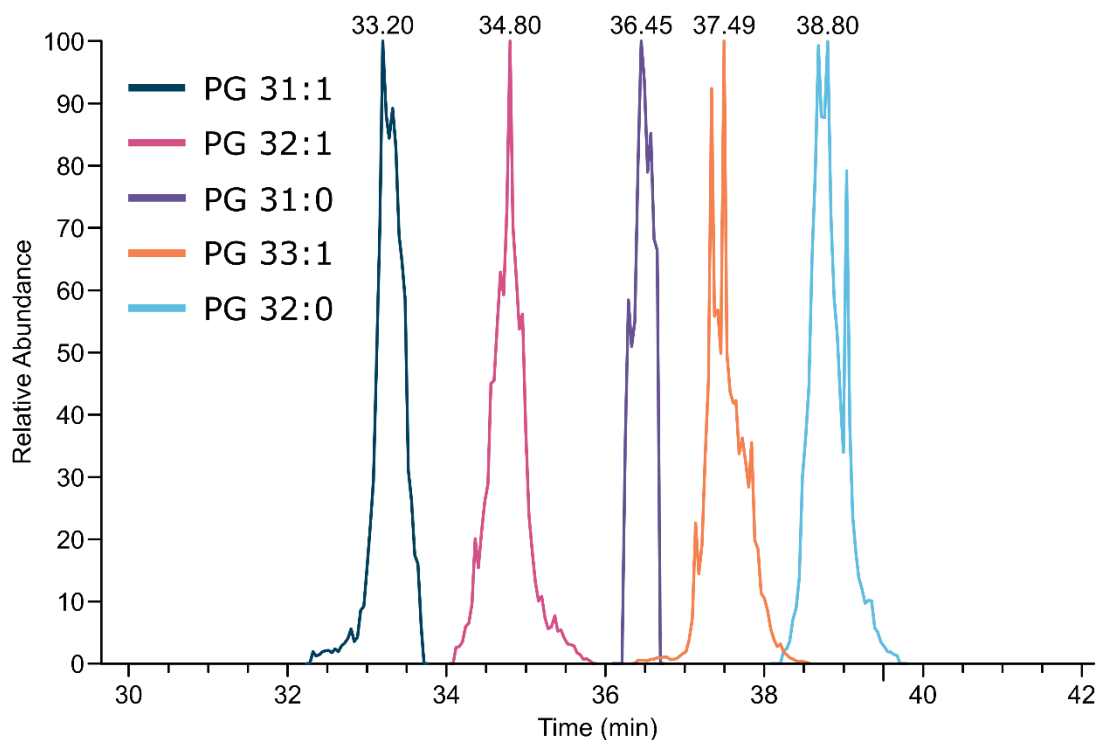


**Figure 1-2. Elution profiles of related phosphatidylglycerol lipids in reversed phase chromatography.**

The extracted ion chromatograms for five phosphatidylglycerol lipids are shown, demonstrating the good separation of reversed phase chromatography and high methylene selectivity of C18 stationary phase.

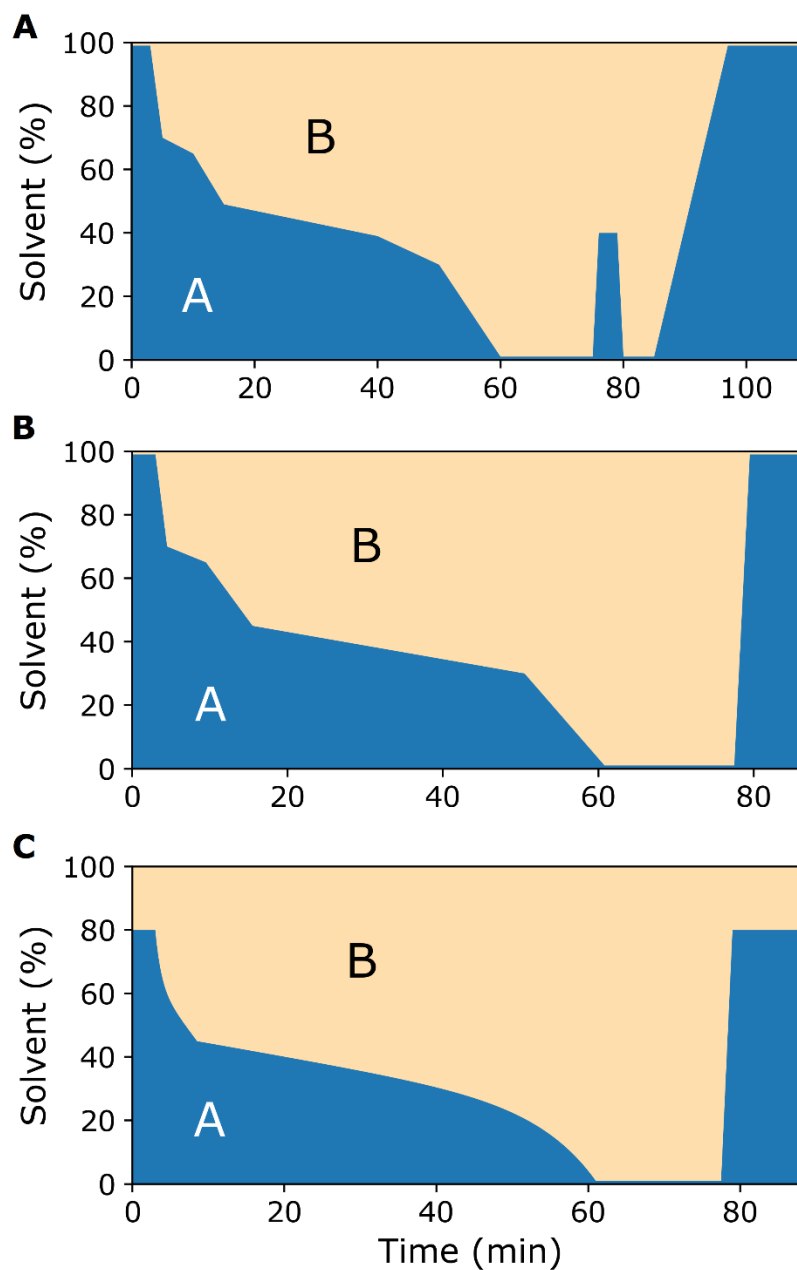


**Figure 1-3. Evolution of RPLC Gradients Employed.**

Mobile phase solvent gradients for the LC methods. On the graphs “B” shows the organic solvent and “A” shows the aqueous-rich solvent. A. Original LC gradient developed by Danne-Rasche *et al.* B. Gradient for the shortened LC method. C. The simplified LC method.

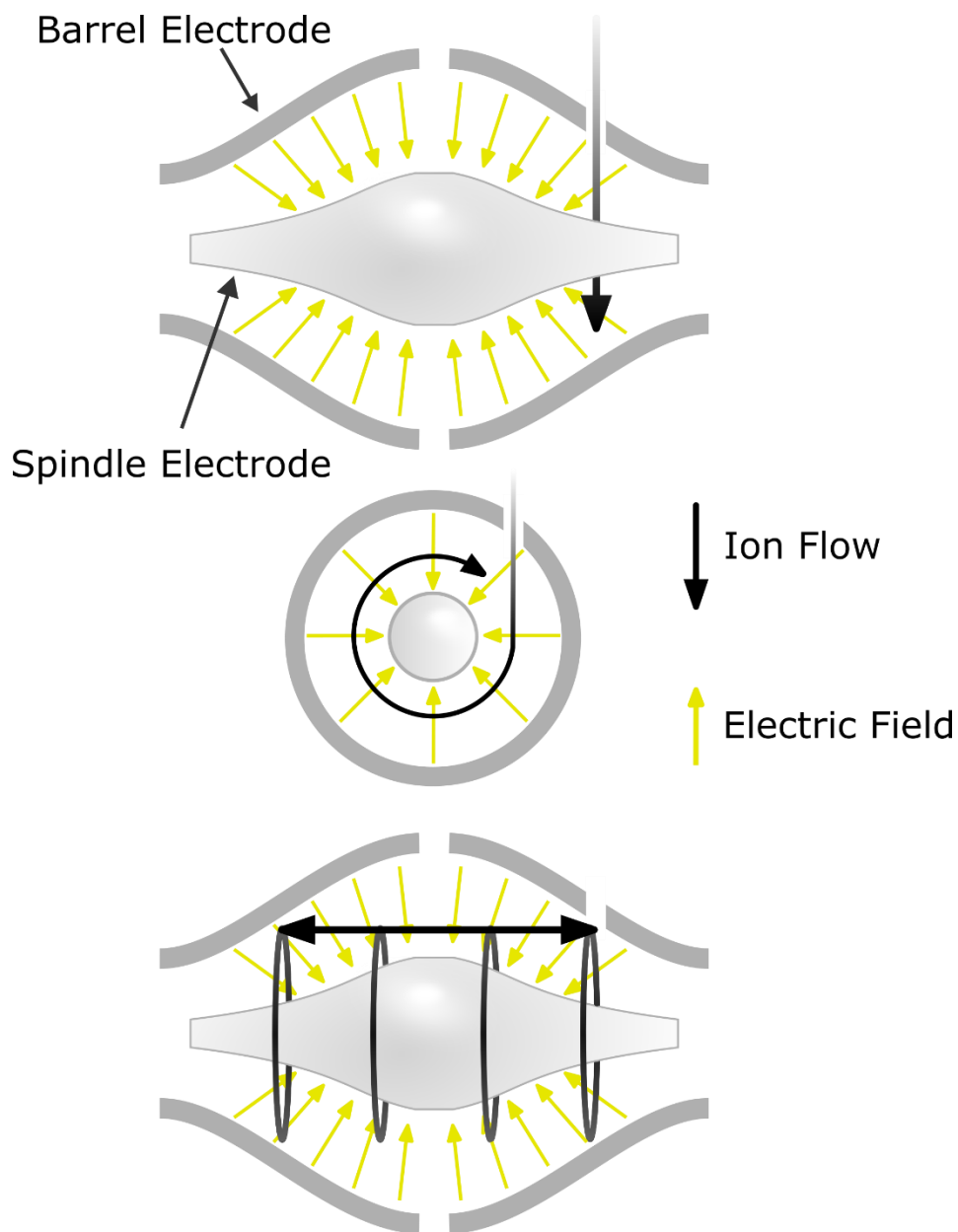


**Figure 1-4. Cross section of an orbitrap mass analyzer.**

After collisional cooling in the C-trap, ions are injected into the orbitrap. The injection happens off center with respect to both the transverse and the lengthwise dimensions. The voltage difference between the barrel and spindle electrodes and the angular momentum result in the ions being caught in a circular (centripetal) orbit. The curve of the electrodes also results in ions being pushed inward from either end, thereby forming a radial oscillation the frequency of which is inversely proportional to  $m/z$ .

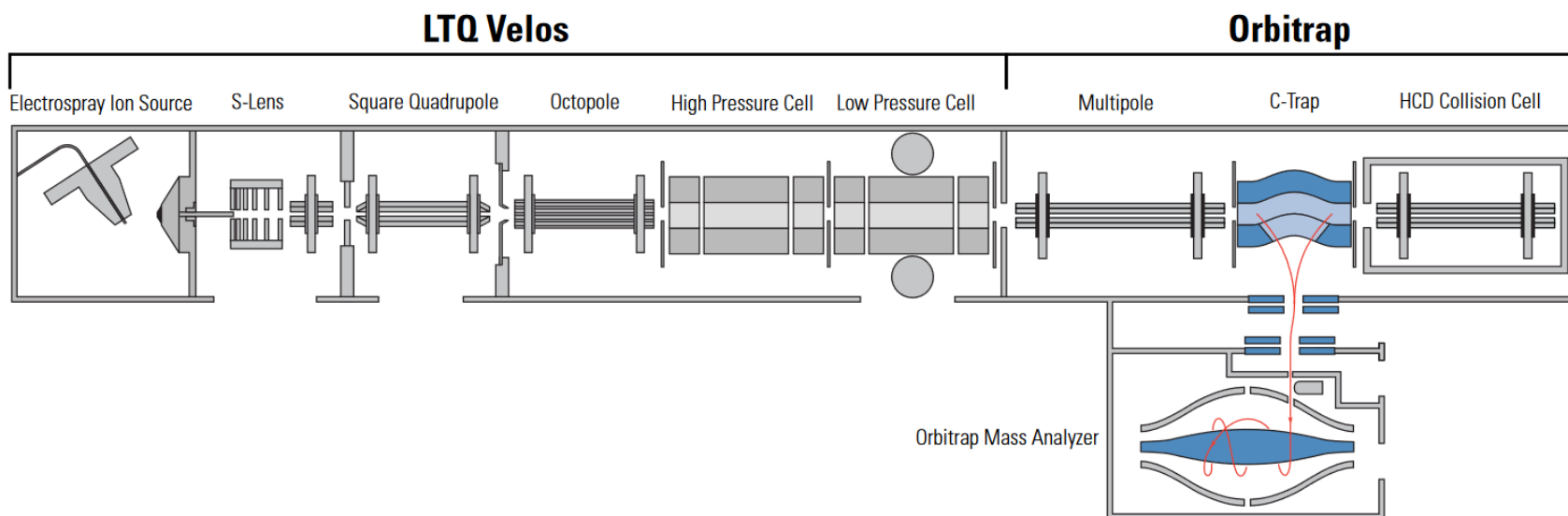


**Figure 1-5. LTQ-Orbitrap Block Diagram.**

A schematic showing the major components of an LTQ-Orbitrap. Graphic taken from the LTQ Orbitrap Pro manual from Thermo Scientific. The square quadrupole is curved so as to eliminate neutrals that enter the mass spectrometer. The ion trap has a high pressure cell used for CID fragmentation and a low pressure cell used for mass analysis.

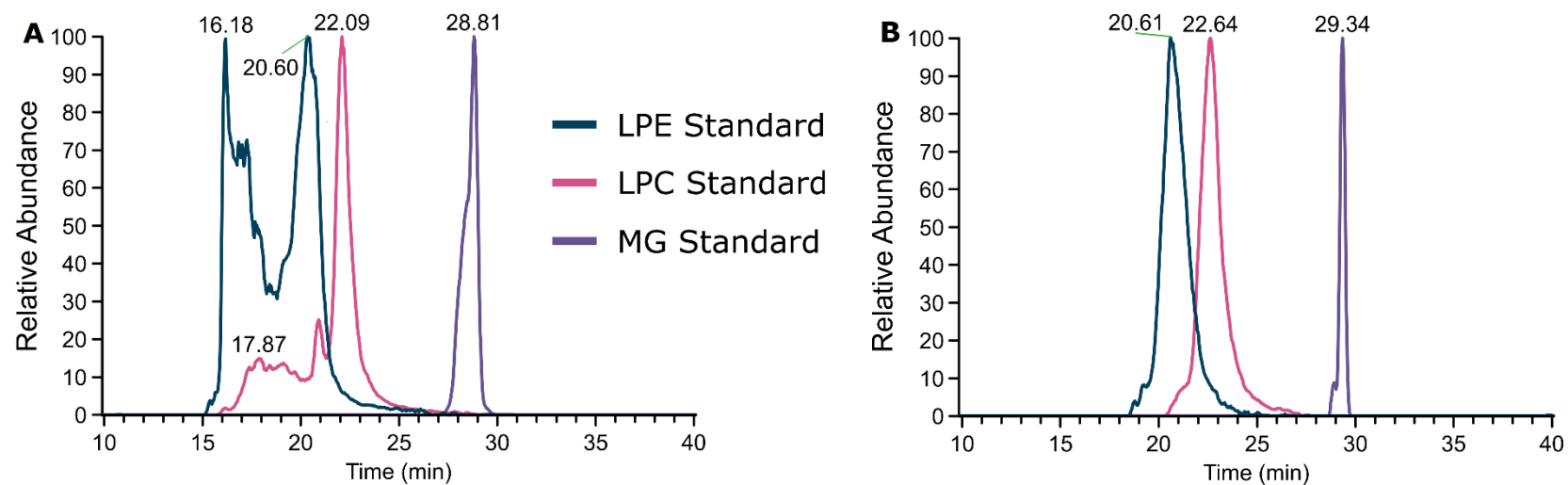


**Figure 1-6. Injection Volume Solvent Overload**

Extracted ion chromatograms (smoothed) showing degradation of peak shape due to injection volume overwhelming the column. A. a 10 µL injection of lipids from 1:1:1 butanol:isopropanol:water. B. a 2 µL injection of the same sample.

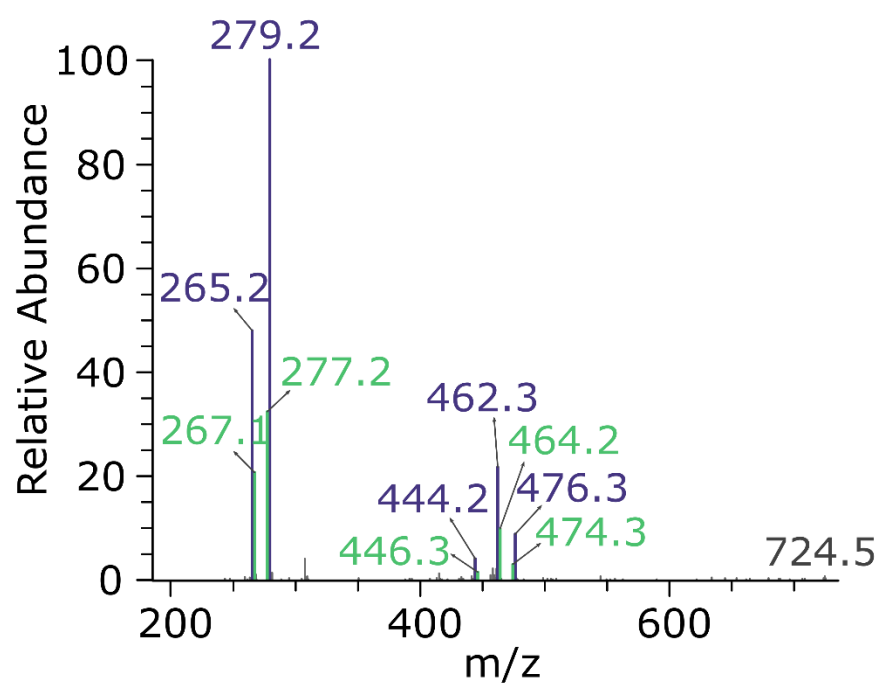
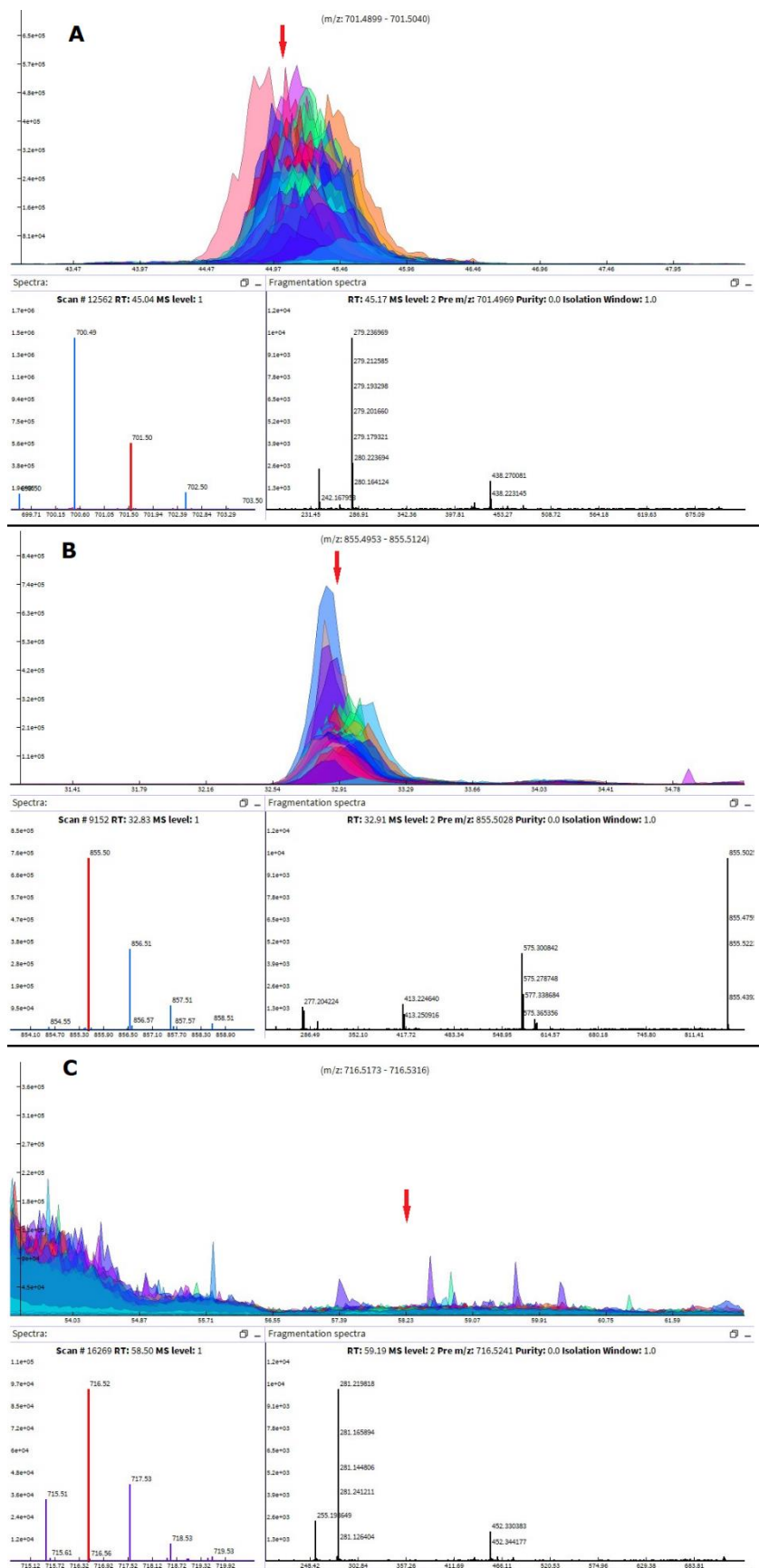


**Figure 2-1. PE 35:4 fragmentation spectrum.**

The fragmentation pattern observed for lipid PE 35:4 ( $m/z$  724.492) in the *Aspergillus* pooled quality control sample (dataset 2). This single chromatographic peak has fragment ions coming from both PE 17:2\_18:2 (purple) and PE 17:1\_18:3 (green) molecular species.

**Figure 2-2. Manual validation of MS-DIAL lipid IDs.**

Three examples of manually validated lipid identifications from MS-DIAL. Each panel shows the extracted ion chromatogram (top subpanel), the isotope packet observed in the full-scan (bottom left subpanel), and the fragmentation spectra observed (bottom right subpanel). The extracted ion chromatograms are zoomed to region containing the retention time reported by MS-DIAL (marked with a red arrow). A. This feature was identified as CL 68:2|CL 16:0\_16:0\_18:0\_18:2 by MSDIAL as the [M-2H]<sup>2-</sup> ion. Manual inspection of the isotope packet reveals the ion is singly charged and itself an isotope of m/z 700.5. Despite the presence of glycerophospholipid-type ions, this annotation is a false positive. This feature is most likely a PE 33:2 [M-H]<sup>-</sup> ion. B. This feature was identified as PI 36:5|PI 18:2\_18:3 by MS-DIAL as the [M-H]<sup>-</sup> ion. Manual validation of this feature confirmed the PI 36:5 annotation. The feature had a good peak shape with an isotope pattern and mass accuracy that was consistent with PI 36:5. A detailed inspection of the fragmentation pattern gave strong evidence to confirm the PI 36:5 identity with fatty acid ions, fatty acid/headgroup neutral losses, and headgroup fragment ions. C. This feature was identified as PE 34:1|PE 16:0\_18:1 by MS-DIAL as the [M-H]<sup>-</sup> ion. After generating the extracted ion chromatogram, it was clear that there was no real peak associated with this feature. The feature instead corresponded to the noise region tailing after the actual peak (to the left of the zoomed chromatogram region).  
(Figure displayed on following page)



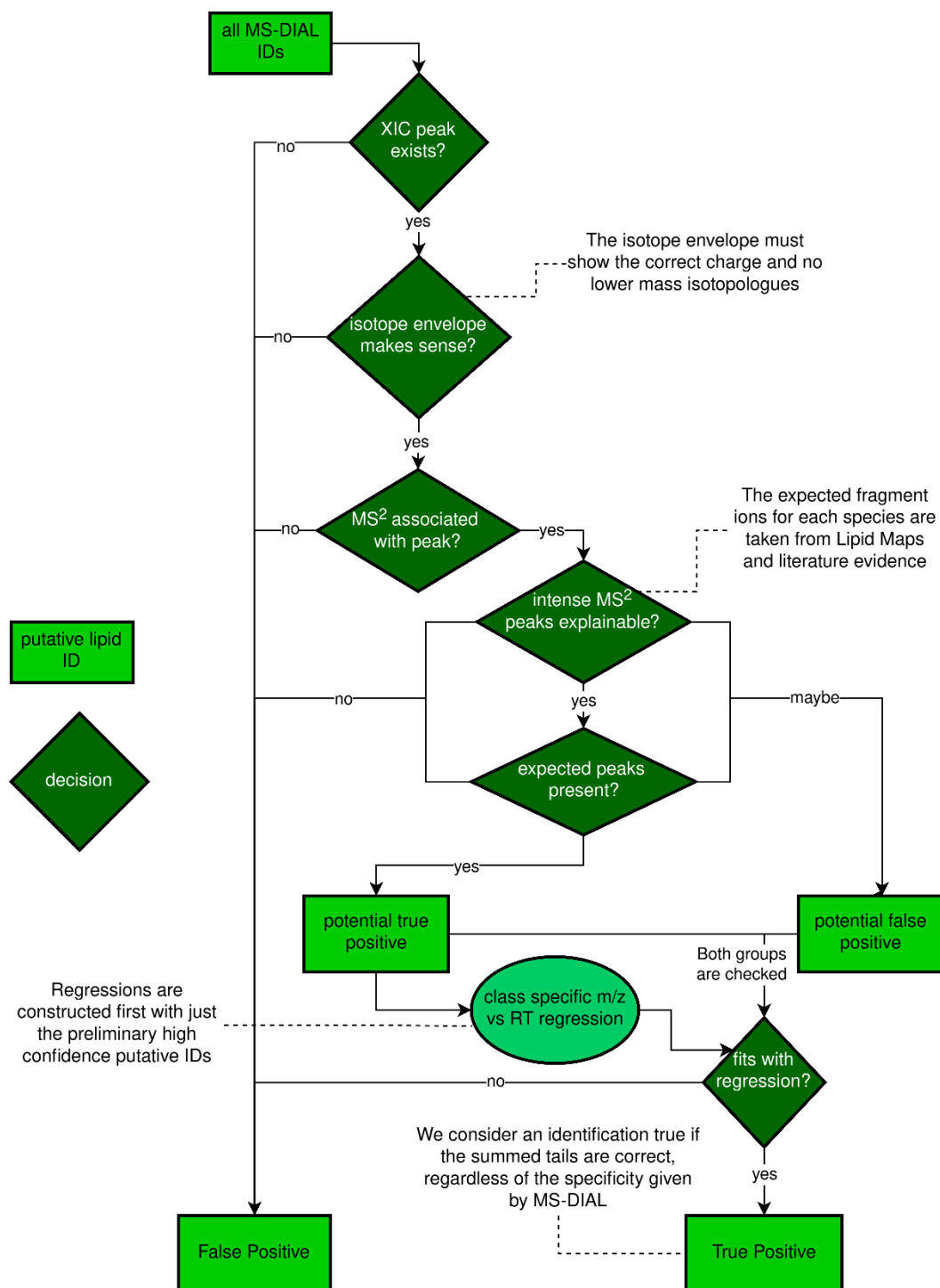


**Figure 2-3. The manual annotation workflow.**

The workflow used during manual validation of reported MS-DIAL lipid IDs. In practice most decisions were not rigidly binary, but rather a holistic accounting of the evidence was used for deciding whether a putative ID was a reliable true positive. This workflow provides the most reliable lipid identifications achievable without standards.

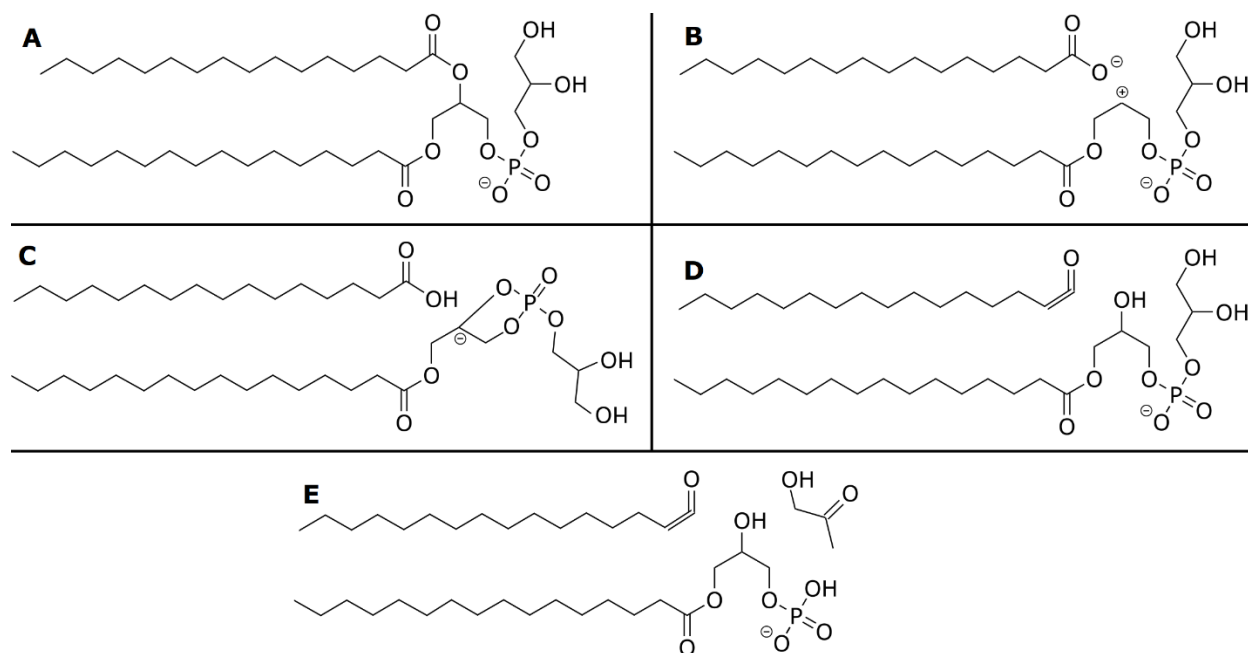


**Figure 2-4. Fragment ions of PG lipids.**

Example fragmentation ion structures for a PG 16:0\_16:0 lipid. A.  $[M-H]^-$  ion. B. Fatty acid carboxylate ion. C. Carboxylic acid fatty acid neutral loss. D. Ketene fatty acid neutral loss. E. Combination ketene fatty acid neutral loss and headgroup loss.

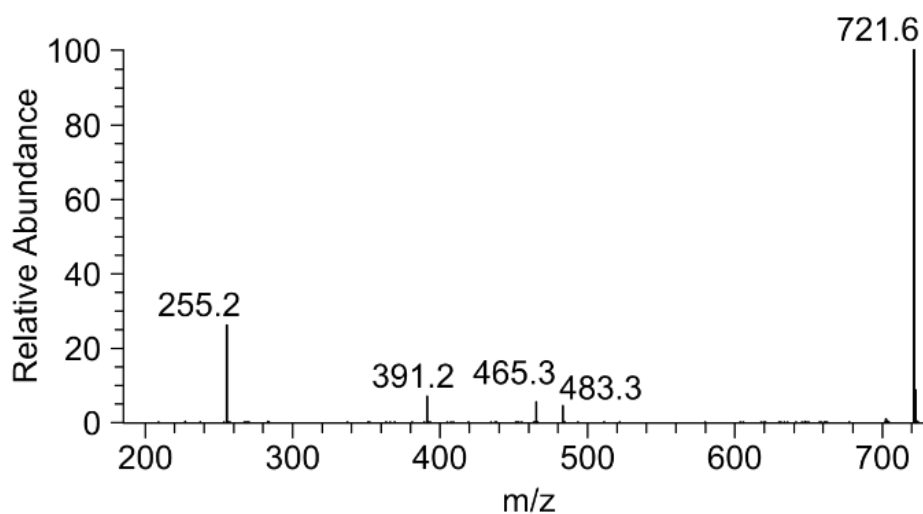


**Figure 2-5. Example fragmentation spectrum of PG 16:0\_16:0.**

An experimental  $MS^2$  spectrum showing the fragment ions described in Figure 2-4 for PG 16:0\_16:0.

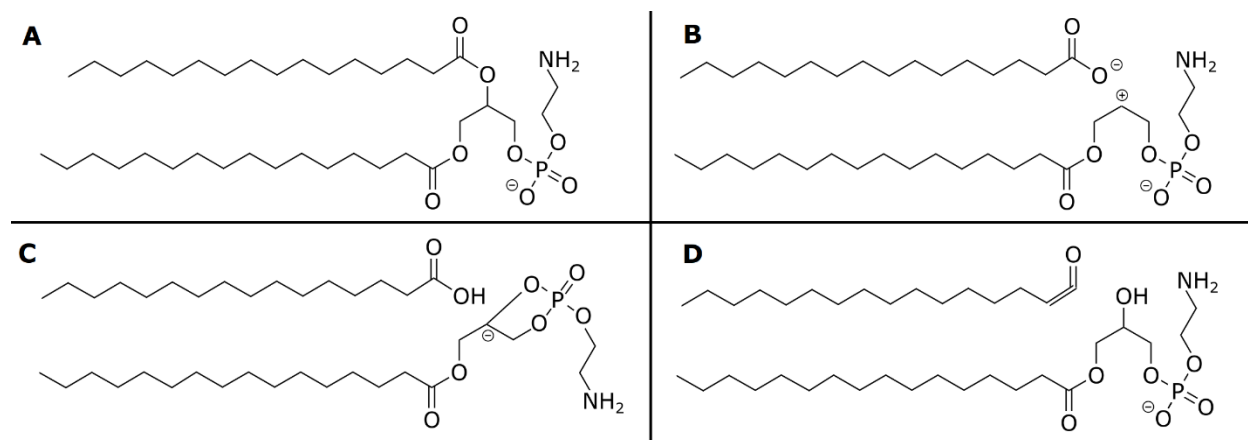


**Figure 2-6. Fragment ions of PE lipids.**

Example fragmentation ion structures for a PE 16:0\_16:0 lipid. A.  $[M-H]^-$  ion. B. Fatty acid carboxylate ion. C. Carboxylic acid fatty acid neutral loss. D. Ketene fatty acid neutral loss.

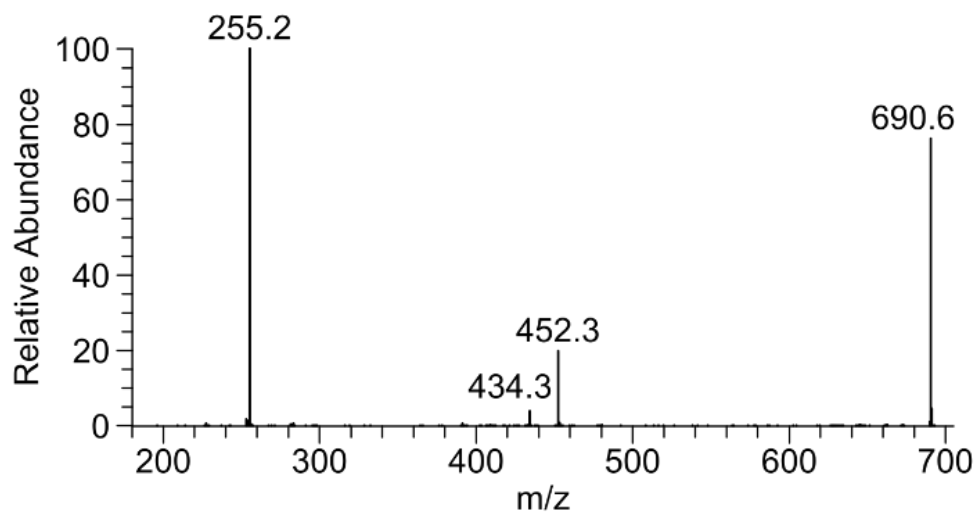


**Figure 2-7. Example fragmentation spectrum of PE 16:0\_16:0**

An experimental  $MS^2$  spectrum showing the fragment ions described in Figure 2-6 for PE 16:0\_16:0.

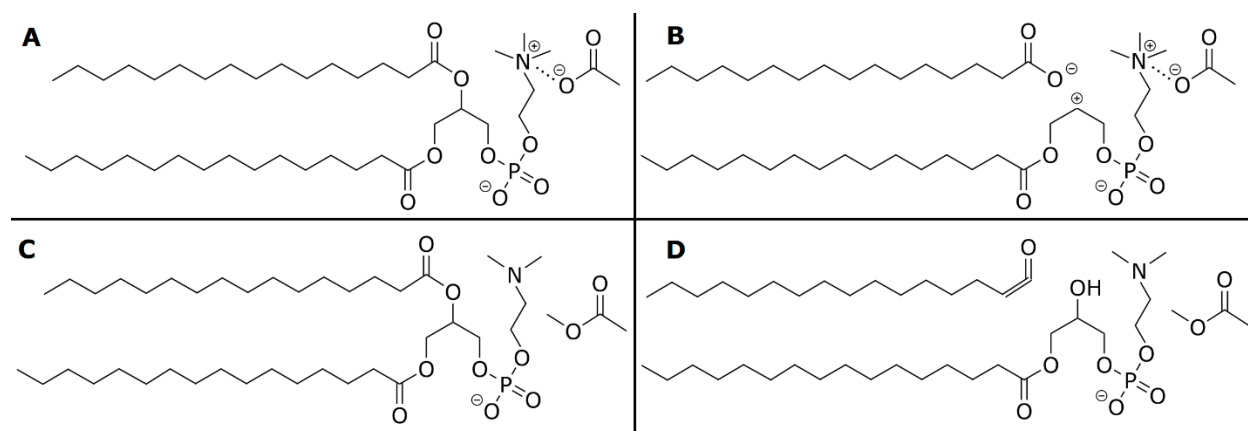


**Figure 2-8. Fragment ions of PC lipids.**

Example fragmentation ion structures for a PE 16:0\_16:0 lipid. A.  $[M+Acetate]^-$  ion. B. Fatty acid carboxylate ion. C. Methyl acetate neutral loss. D. Ketene fatty acid and methyl acetate neutral loss.

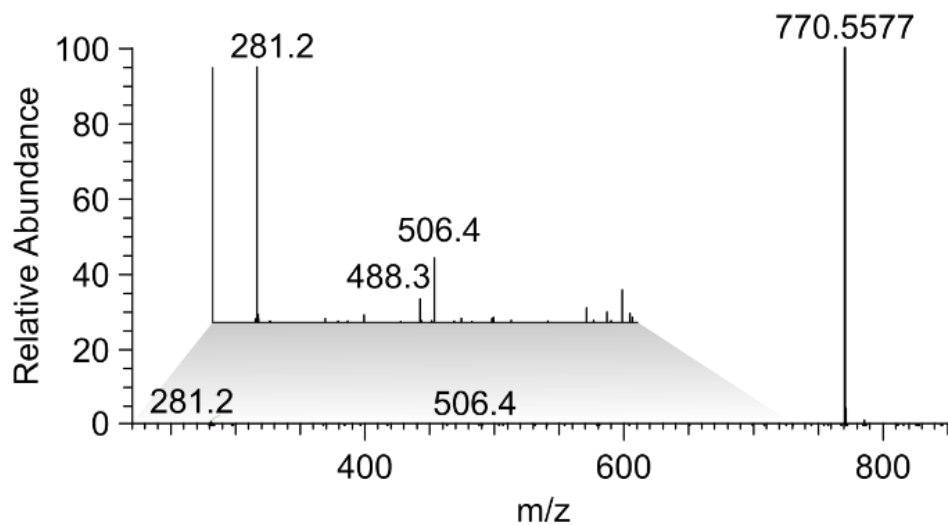


**Figure 2-9. Example fragmentation spectrum of PC 18:1\_18:1.**

An experimental fragmentation spectrum of an example PC lipid. The lower  $m/z$  range has been blown up so low relative abundance fragment ions can be observed.

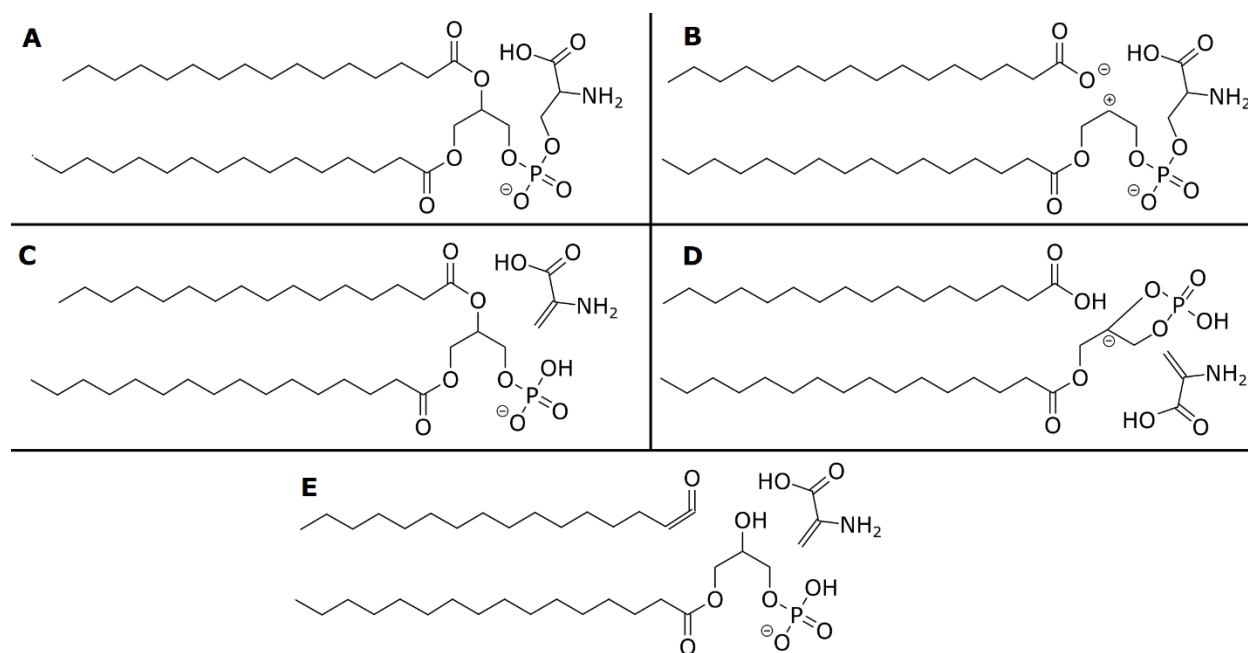


**Figure 2-10. Fragment ions of PS lipids.**

Example fragmentation ion structures for a PS 16:0<sub>16:0</sub> lipid. A.  $[M-H]^-$  ion. B. Fatty acid carboxylate ion. C. Serine headgroup neutral loss. D. Combination carboxylic acid fatty acid neutral loss and headgroup loss. E. Combination ketene fatty acid neutral loss and headgroup loss.

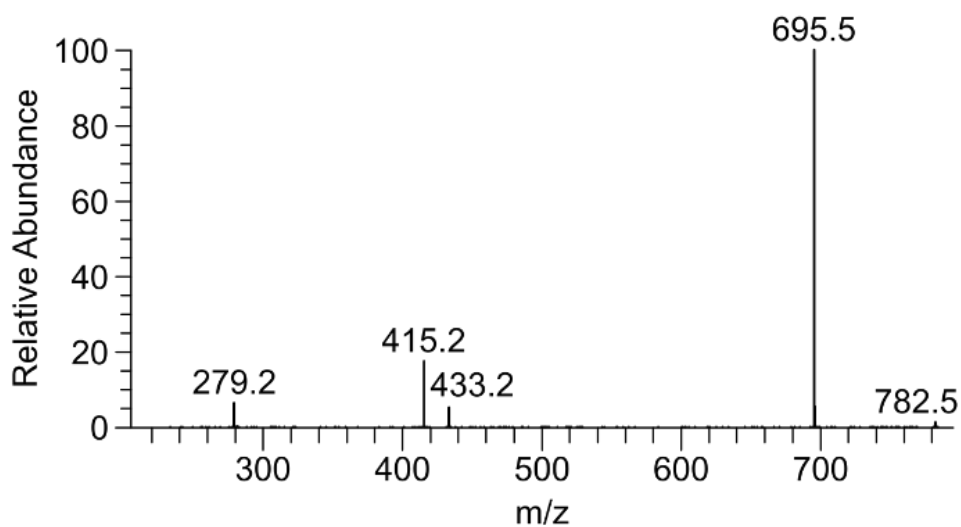


**Figure 2-11. Example fragmentation spectrum of PS 18:2<sub>18:2</sub>.**

An experimental MS<sup>2</sup> showing the ions described in Figure 2-10 for lipid PS 18:2<sub>18:2</sub>.

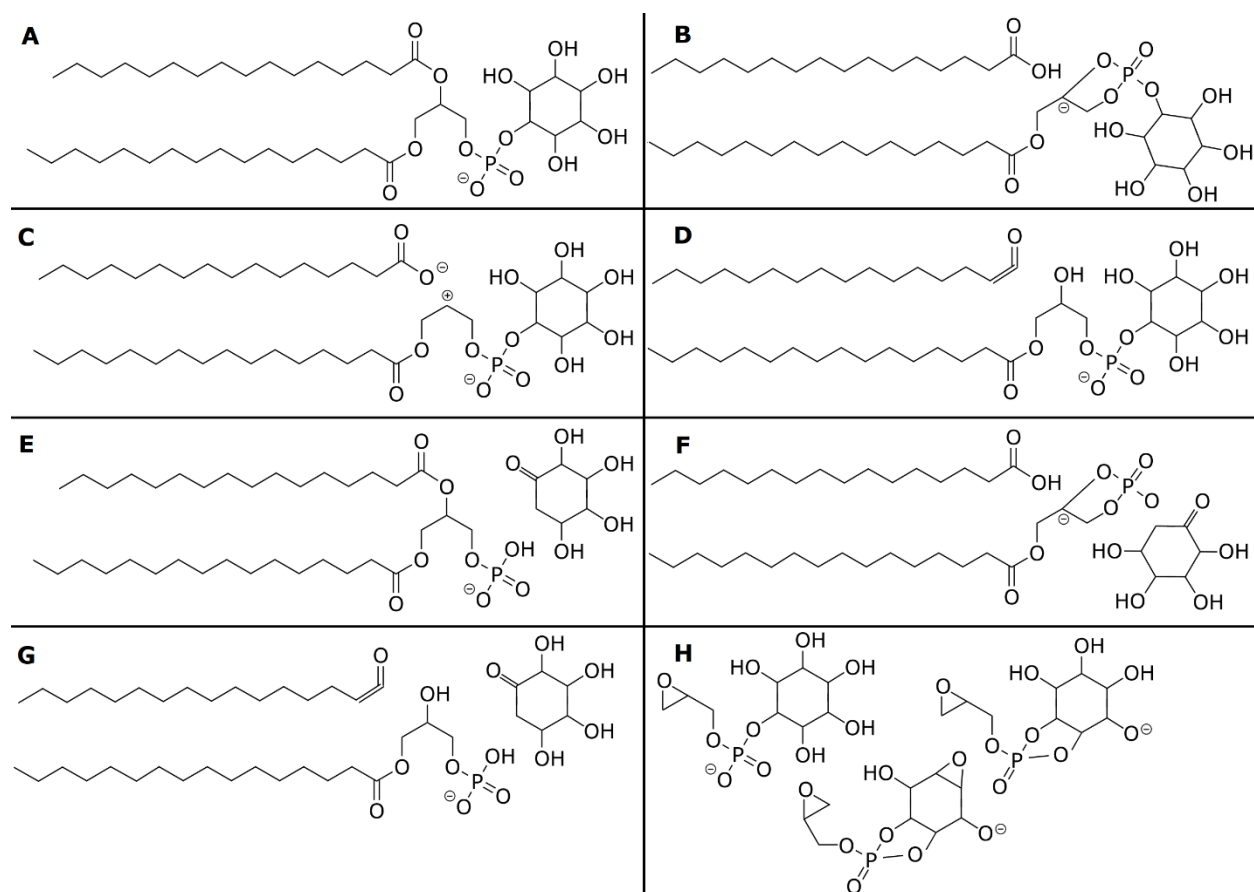


**Figure 2-12. Fragment ions of PI lipids.**

Example fragmentation ion structures for a PS 16:0\_16:0 lipid. A.  $[M-H]^-$  ion. B. Fatty acid carboxylic neutral loss C. Fatty acid carboxylate ion. D. Fatty acid ketene neutral loss. E. Inositol neutral loss. F. Carboxylic acid fatty acid and inositol combined neutral losses. G. Ketene fatty acid and inositol combined neutral losses. H. Three different glycerophosphate-inositol fragment ions.

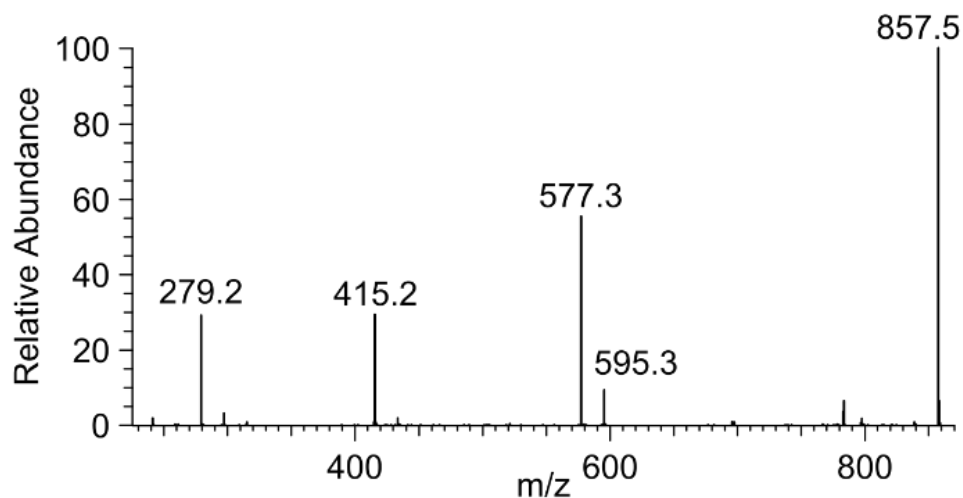


**Figure 2-13. Example fragmentation spectrum of PI 18:2\_18:2.**

An experimental MS<sup>2</sup> spectrum showing most the ions described in Figure 2-12 for the PI 18:2\_18:2 lipid.

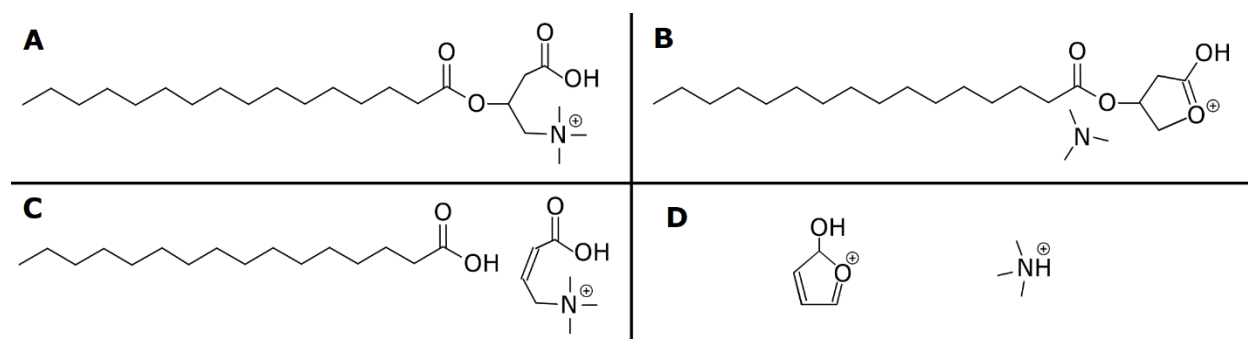


**Figure 2-14. Example fragment ions of a carnitine lipid.**

Example fragmentation of a Car 16:0 lipid. A shows the pseudo  $[M+H]^+$  ion. B shows the trimethylammonium neutral loss. C shows the carnitine fragment ion. D shows additional fragment ions of the carnitine moiety.

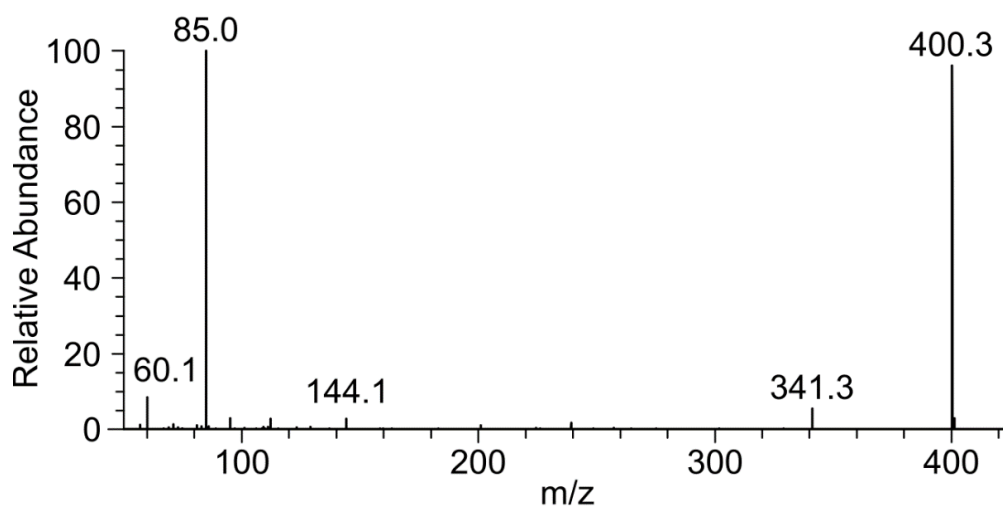


**Figure 2-15. QE Fragmentation Spectrum of Car 16:0.**

An experimental MS<sup>2</sup> spectrum of Car 16:0 showing the peaks described in Figure 2-14.

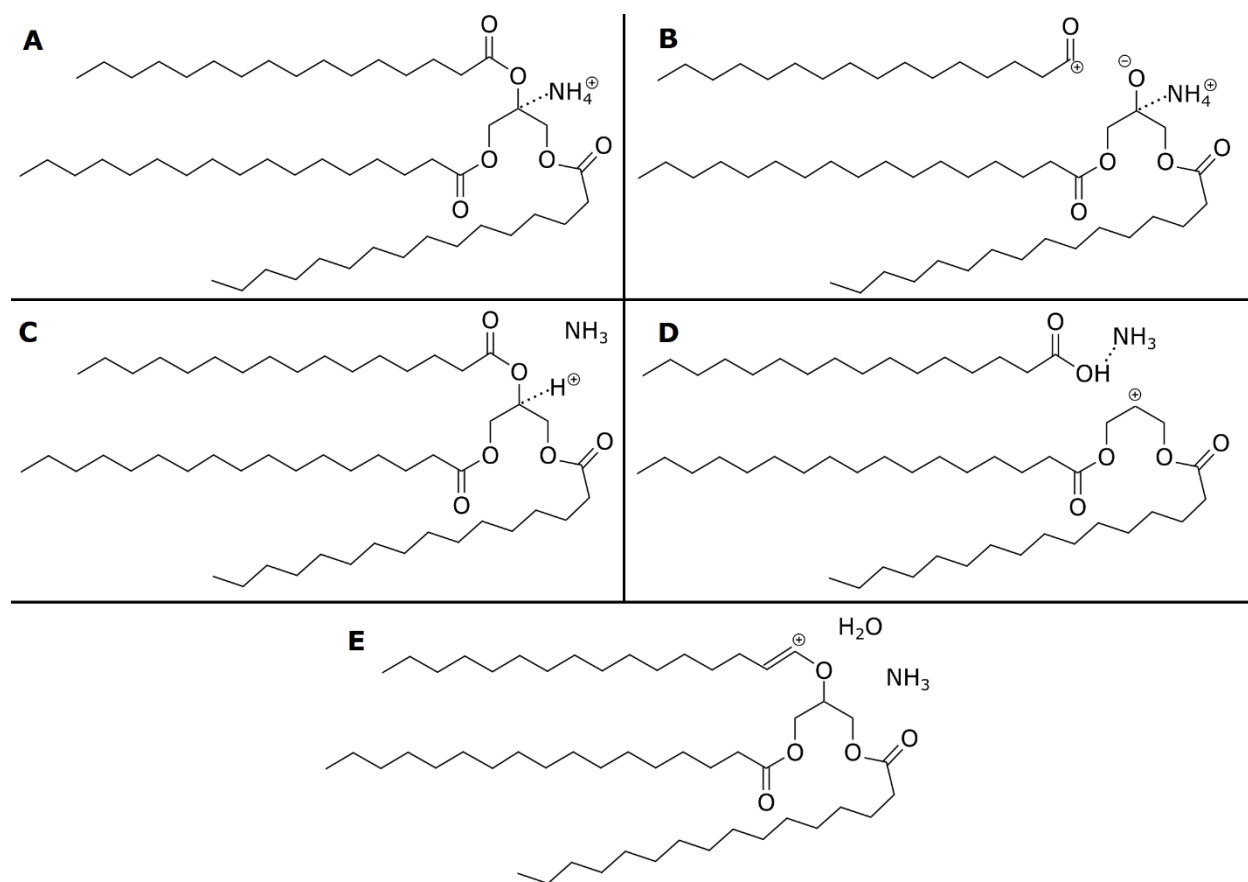


**Figure 2-16. Example Fragmentation Ions of a TAG 16:0\_16:0\_16:0.**

A shows the ammonium adduct of TAG. B shows the expected ion structure of the fatty acid fragment ion. C shows the ammonia neutral loss. D shows the combination fatty acid-ammonia neutral loss. E shows a combination ammonia-water neutral loss.

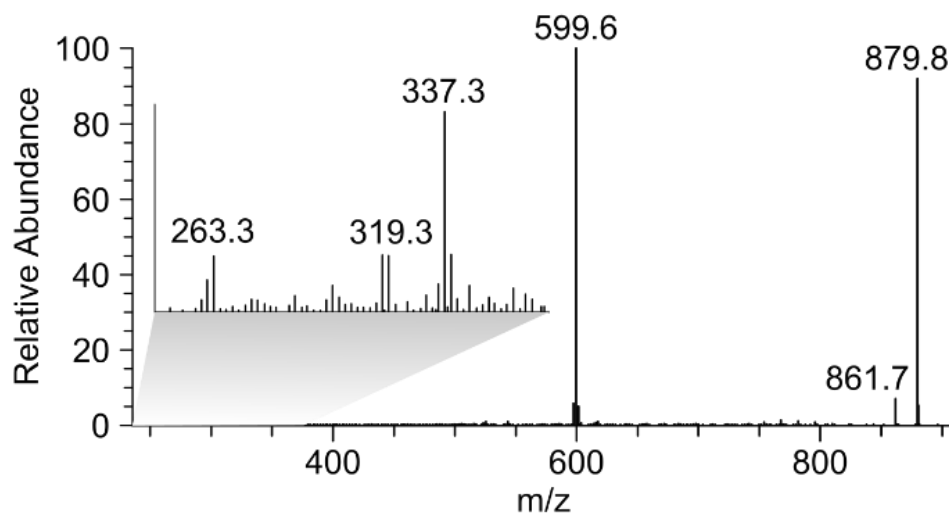


**Figure 2-17. Fragmentation Spectrum of a TAG Lipid.**

The experimental fragmentation spectrum of a TAG 18:2\_18:2\_18:2 lipid. Some amount of TAG 18:1\_18:2\_18:3 can also be observed in the spectrum. The lower  $m/z$  range has been blown up so low relative abundance fragment ions can be observed.

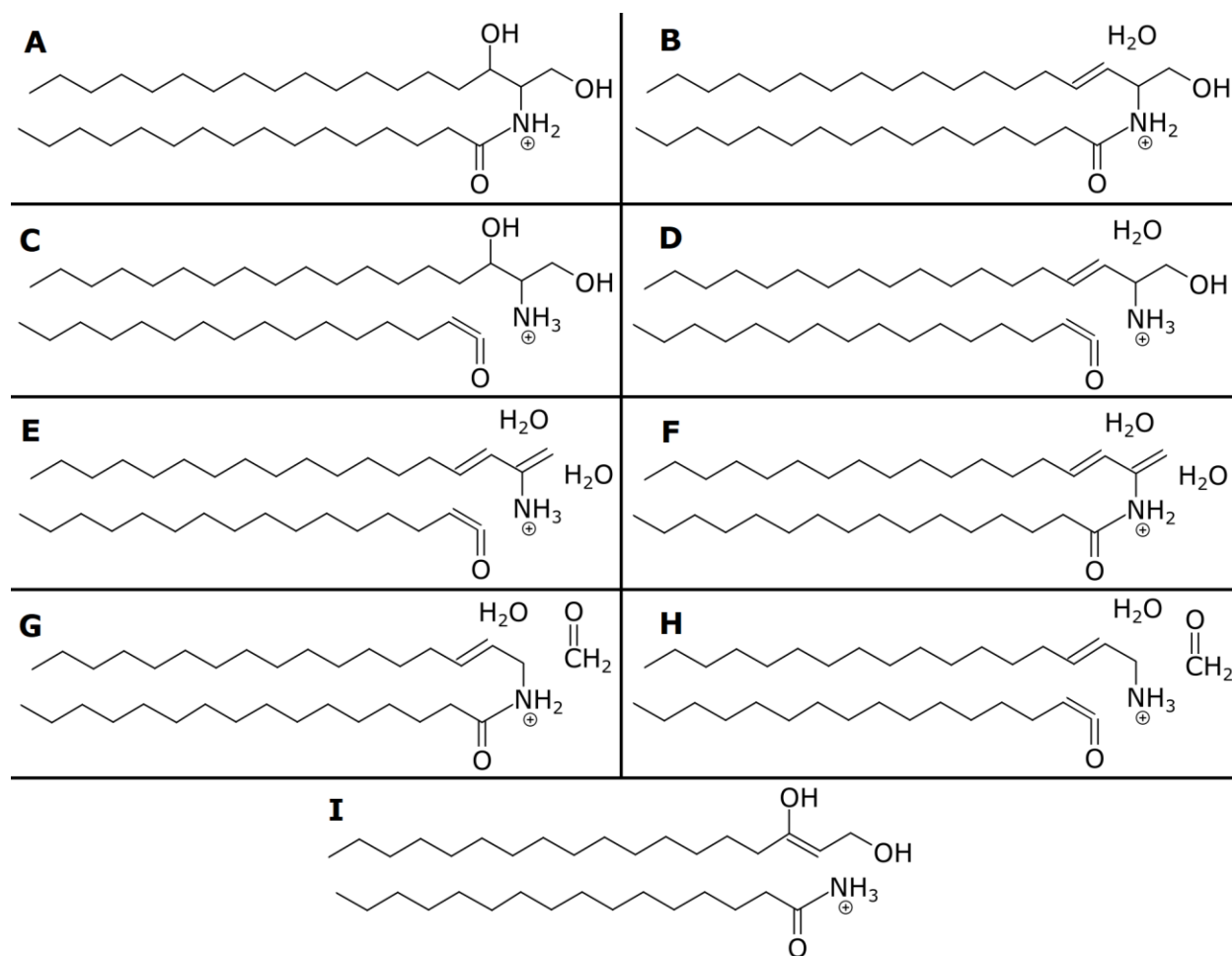


**Figure 2-18. Fragmentation Ion Structures for Cer Lipids.**

Suggested ion structures for fragments from Cer\_NDS and/or Cer\_NS lipids. This is based primarily on personal experience and should be interpreted loosely. A shows the  $[M+H]^+$  ion. B shows a water neutral loss. C shows a ketene neutral loss of the fatty acid moiety. D shows a combination ketene-water neutral loss. E shows shows a combination ketene-double water neutral loss. F shows a double water neutral loss. G shows a combination water acetone neutral loss. H shows a combination water, acetone, and ketene neutral loss. I shows a fatty acid fragment ion.

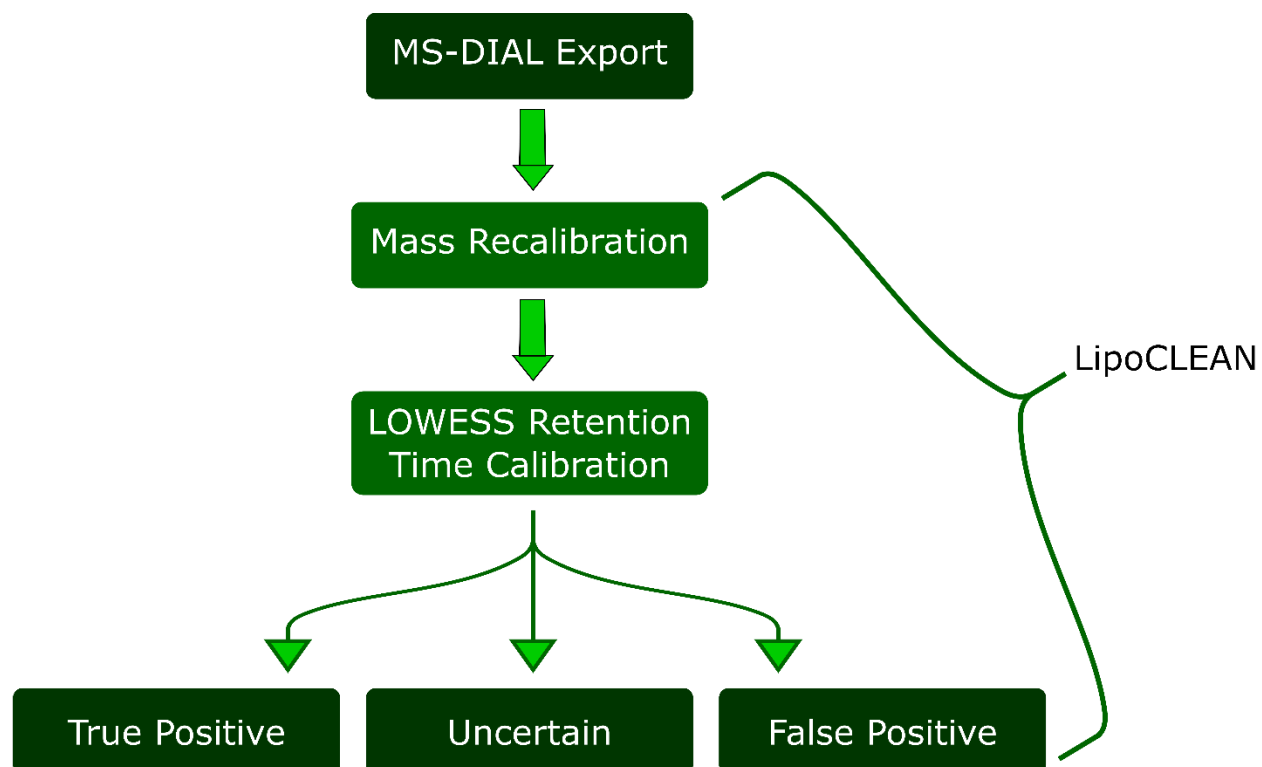


**Figure 2-19. The architecture of the tool LipoCLEAN.**

Dark green icons are external text files while all other icons represent internal processes within the tool. The model first identifies a subset of lipids with which to correct  $m/z$ . Then a subset of lipids are identified to build a retention time reference. The final model rescores lipid identifications into three categories. Predicted positives can be used for downstream analysis without substantial manual filtering. Predicted negatives can be safely ignored (these contained 8% annotated positives in the assessed model). Finally, the reanalyze category requires manual investigation to be used with confidence.

**Table 2-1. The datasets that were manually annotated for model training and validation.** Count refers to the number of putative lipid identifications that were manually annotated. Datasets will be referenced by ID number throughout the document. All datasets were run in both positive and negative mode.

| ID | Source                   | Instrument         | Count | FDR | Use        | Organism                     |
|----|--------------------------|--------------------|-------|-----|------------|------------------------------|
| 1  | in-house                 | Orbitrap Velos Pro | 526   | 62% | validation | <i>Laccaria bicolor</i>      |
| 2  | in-house                 | Orbitrap Velos Pro | 1015  | 35% | training   | <i>Aspergillus fumigatus</i> |
| 3  | MTBLS5583 <sup>8</sup>   | Q-Exactive         | 742   | 46% | training   | <i>Canis familiaris</i>      |
| 4  | MTBLS4108 <sup>284</sup> | TripleTOF 6600     | 1125  | 47% | training   | <i>Rattus norvegicus</i>     |

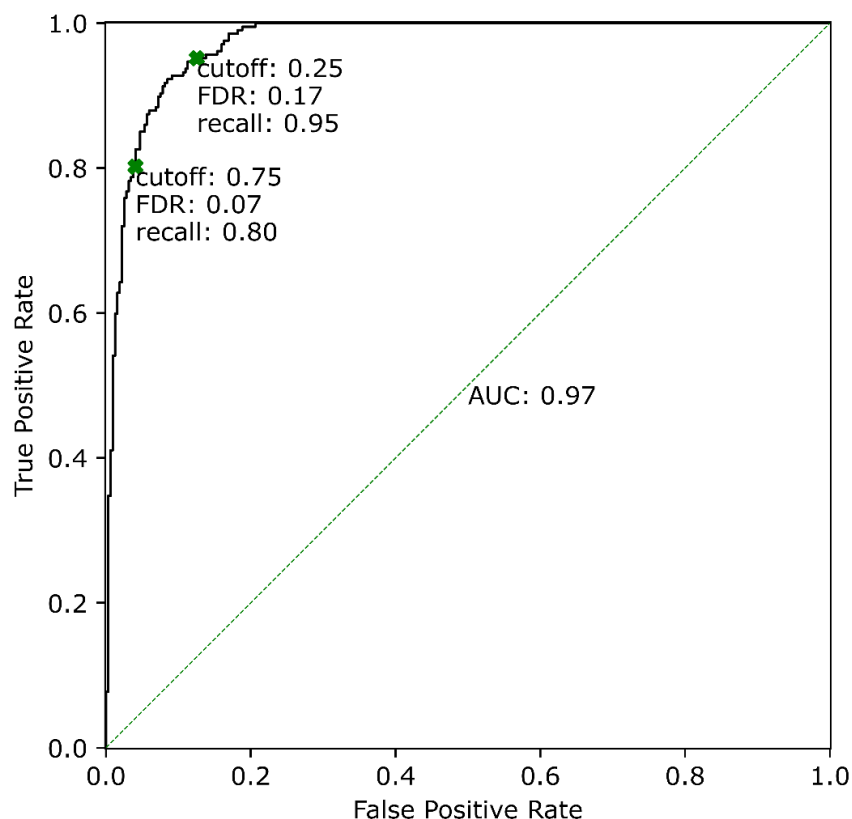


**Figure 2-20. Model performance on the validation dataset.**

Here we show the receiver operating characteristic (ROC) curve for the model when used on dataset 1 (table 1). The two cutoffs are shown as green Xs and performance metrics at those cutoffs are annotated. The theoretical performance of a random model is shown in the dashed green line.

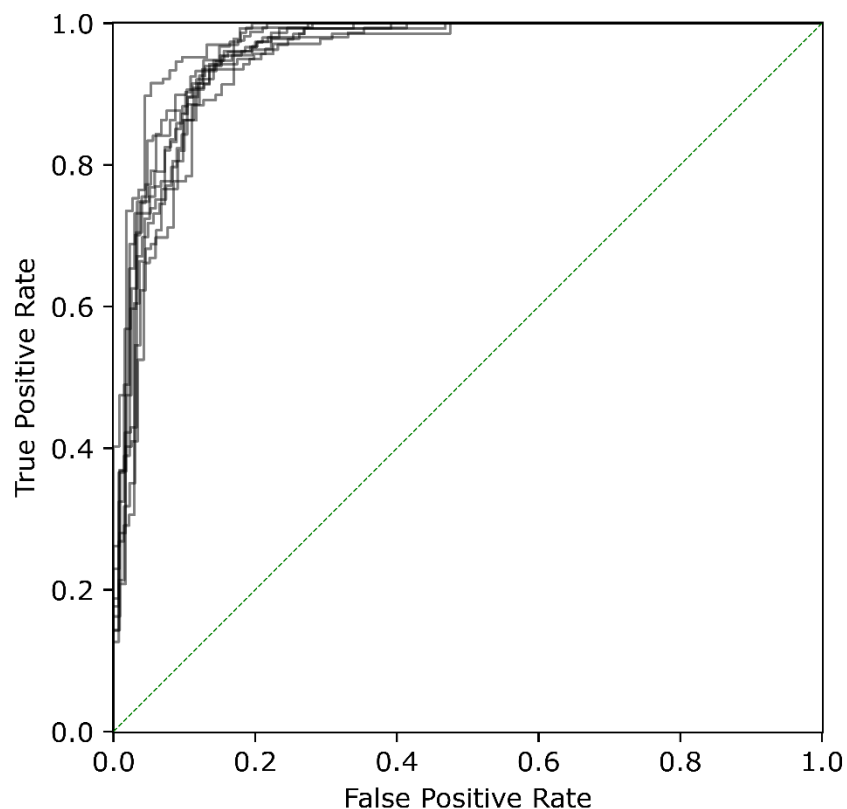


**Figure 2-21. Assessment of the repeatability of the model training process.**

We generated 32 random 80-20 train-test splits of datasets 2 and 3 (Table 1). A model was trained on each of the training data subsets. The model was evaluated on its associated test subset and the ROC curves for each of these tests are plotted. The training process shows strong consistency which means that performance metrics calculated for a particular model on a hold out dataset are likely to be similar to what would be obtained from a different model trained on slightly different data.

**Table 2-2: The number of manually annotated putative lipid IDs per class within each dataset.**

| <b>Lipid Class</b> | <b>Dataset 1</b> | <b>Dataset 2</b> | <b>Dataset 3</b> | <b>Dataset 4</b> | <b>Total</b> |
|--------------------|------------------|------------------|------------------|------------------|--------------|
| <b>TG</b>          | 239              | 243              | 41               | 241              | 764          |
| <b>CL</b>          | 107              | 61               | 60               | 313              | 541          |
| <b>PE</b>          | 108              | 90               | 85               | 146              | 429          |
| <b>PC</b>          | 0                | 26               | 219              | 24               | 269          |
| <b>PG</b>          | 0                | 10               | 86               | 107              | 203          |
| <b>PS</b>          | 0                | 100              | 39               | 54               | 193          |
| <b>PI</b>          | 0                | 44               | 55               | 79               | 178          |
| <b>DG</b>          | 52               | 18               | 16               | 41               | 127          |
| <b>OxPE</b>        | 0                | 0                | 69               | 0                | 69           |
| <b>CAR</b>         | 0                | 11               | 17               | 37               | 65           |
| <b>LPC</b>         | 0                | 2                | 0                | 45               | 47           |
| <b>Cer_NS</b>      | 0                | 7                | 38               | 0                | 45           |
| <b>LPE</b>         | 12               | 0                | 0                | 25               | 37           |
| <b>OxTG</b>        | 0                | 35               | 0                | 0                | 35           |
| <b>Cer_AP</b>      | 0                | 35               | 0                | 0                | 35           |
| <b>HexCer_HS</b>   | 15               | 0                | 0                | 0                | 15           |
| <b>LPG</b>         | 0                | 0                | 0                | 13               | 13           |
| <b>EtherPE</b>     | 0                | 13               | 0                | 0                | 13           |
| <b>PA</b>          | 0                | 0                | 8                | 0                | 8            |
| <b>Cer_NDS</b>     | 0                | 0                | 8                | 0                | 8            |
| <b>HexCer_AP</b>   | 7                | 0                | 0                | 0                | 7            |
| <b>HexCer_HDS</b>  | 3                | 0                | 0                | 0                | 3            |
| <b>HexCer_NS</b>   | 2                | 0                | 0                | 0                | 2            |
| <b>EtherOxPE</b>   | 0                | 2                | 0                | 0                | 2            |
| <b>OxPG</b>        | 0                | 0                | 1                | 0                | 1            |

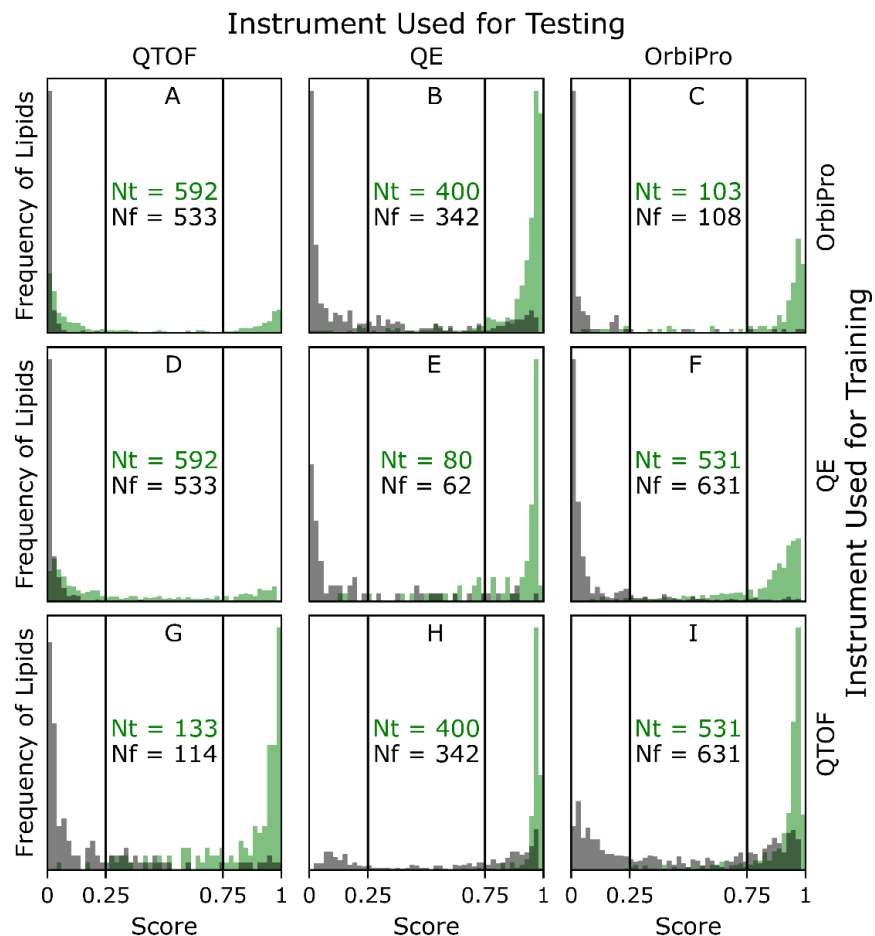


**Figure 2-22. Assessing generalization to new instruments.**

A new model was trained on each dataset and its performance was evaluated on all datasets individually. Datasets 2, 3, and 4 (Table 1) are listed as OrbiPro (Orbitrap Velos Pro), QE (Q Exactive), and QTOF (TripleTOF 6600), respectively. A hold-out subset of each dataset was used for evaluating the performance of the model trained on that dataset (C, E, G). Nt is the number of annotated true lipid identifications and Nf is the number of annotated false identifications.

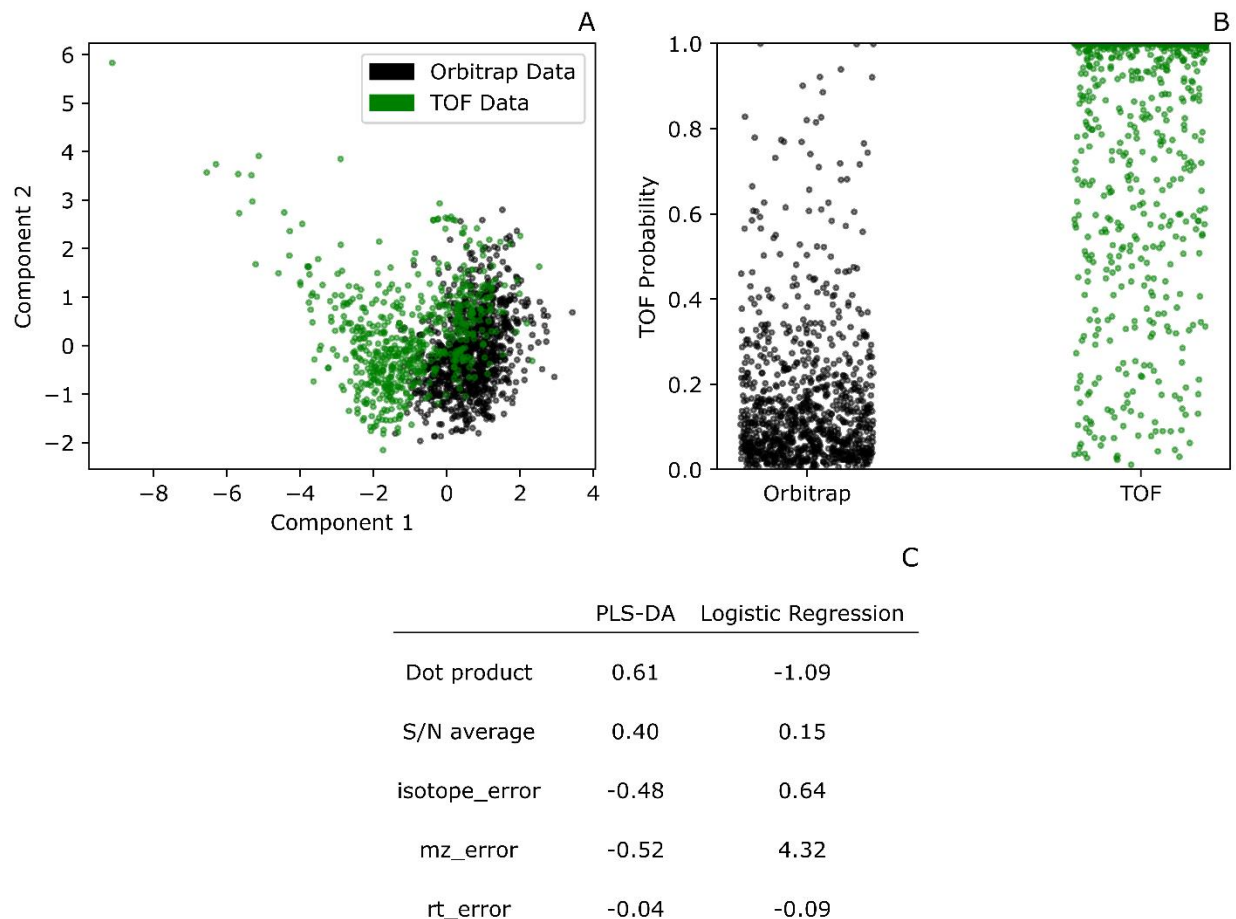


**Figure 2-23. Investigating the incompatibility between TOF and Orbitrap data.**

These analyses use the predictors available to the final model (Figure 1C) to separate Orbitrap data (datasets 1-3, Table 1) from TOF data (dataset 4). Only annotated positive examples are used. A) PLS-DA ordination partially separates the two classes. B) The predicted probabilities from a logistic regression show a noticeable difference between groups. C) The parameters of the above models. The PLS-DA parameters are the factor loadings on component 1. It is not expected that the parameters should correlate between models but large absolute magnitudes are indicative of separability along that dimension.

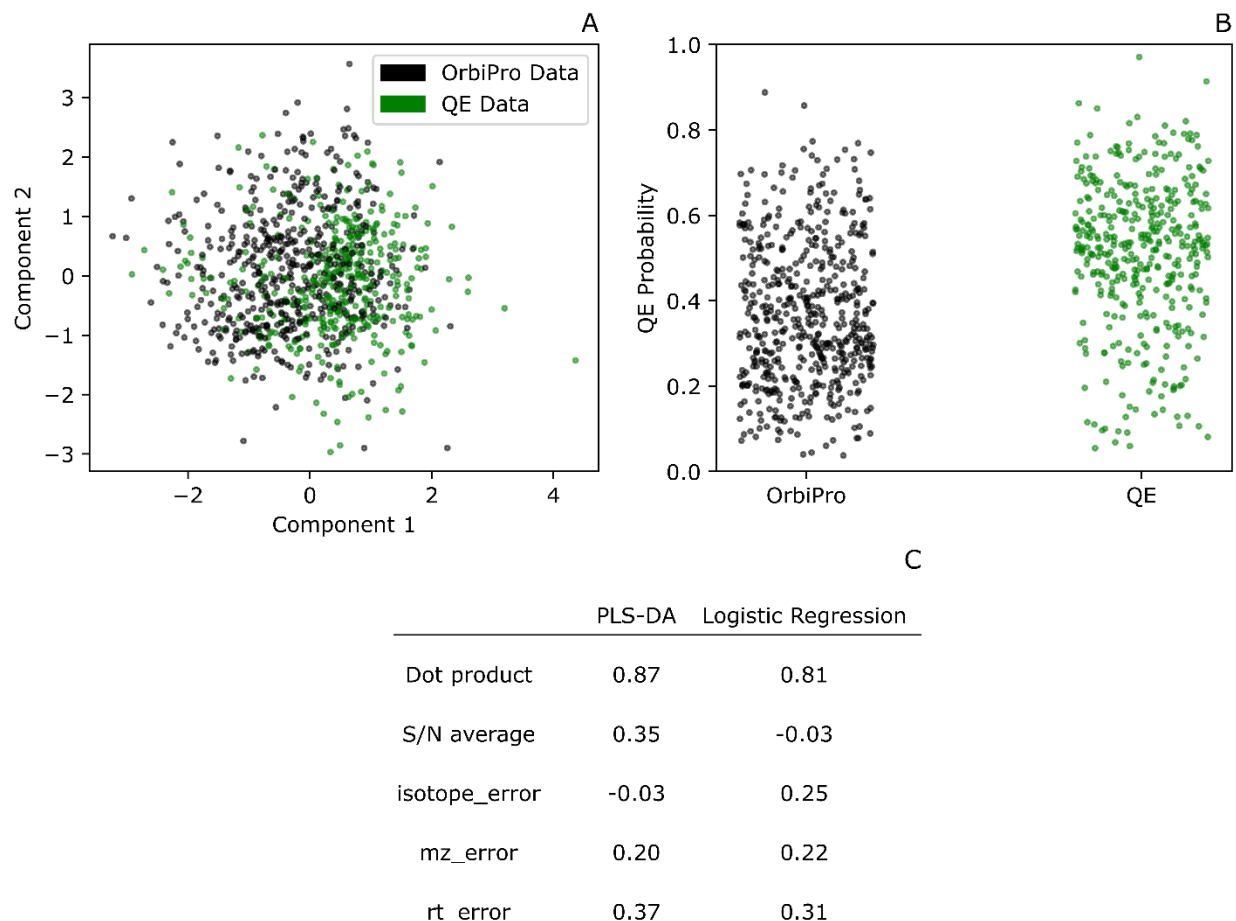


**Figure 2-24. Investigating the compatibility between Q Exactive and Orbitrap Velos Pro data.**

These analyses use the predictors available to the final model (Figure 1C) to separate Orbitrap Velos Pro data (datasets 1-2, Table 1) from Q Exactive data (dataset 3). Only annotated positive examples are used. A) PLS-DA ordination fails to separate the two classes, although slight differences are visible. B) The predicted probabilities from a logistic regression show a marginal difference between groups. C) The parameters of the above models. The PLS-DA parameters are the factor loadings on component 1. It is not expected that the parameters should correlate between models but large absolute magnitudes are indicative of separability along that dimension. There is no bias towards 0 on the loadings in PLS-DA, unlike logistic regression parameters, so in the absence of separability the loadings are expected to be random.

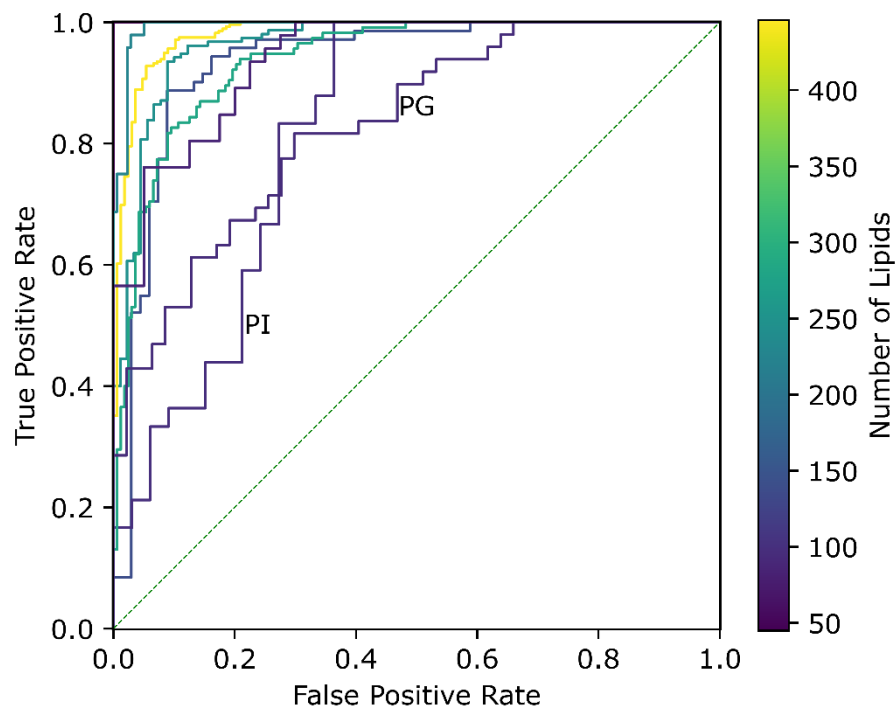


**Figure 2-25. Assessing generalization to new lipid classes.**

Each lipid class in turn was held out from datasets 1-3 (Table 1), a new model was trained on these censored data and its performance assessed on the held-out lipid class data. The ROC curves for the two poorly performing models are labeled with their corresponding test lipid class, PI and PG. Number of lipids refers to the total number of training entries in the held-out class.

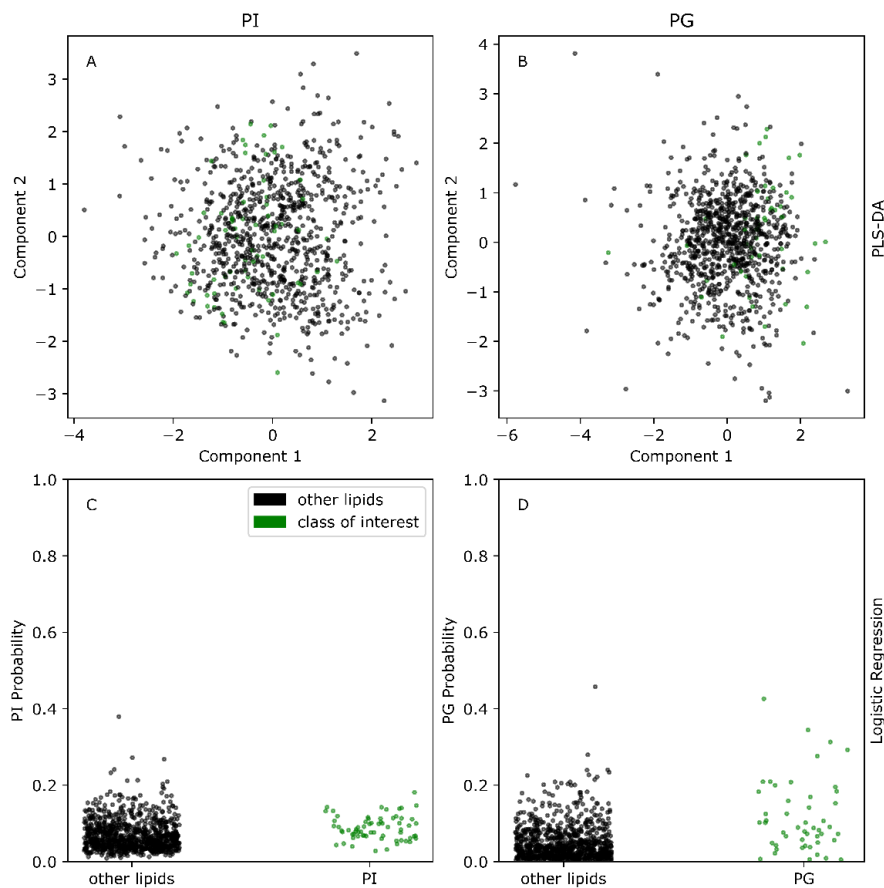


**Figure 2-26. Investigating poor generalization performance.**

We examined whether the poor cross validation performance (Figure 4) for PI (A and C) and PG (B and D) lipids was due to differences in the predictor distributions. Only annotated true positives from datasets 2 and 3 (Table 1) are included in this analysis. We sought to determine if the lipid classes that displayed poor generalization performance in Figure 4 occupy a substantially different part of the predictor space than other lipid classes as this would explain why their absence from the training dataset produced a model that failed to appropriately predict their quality. A and B show the results of PLS-DA ordination which fails to separate either query class from the rest of the lipids. The features for this analysis are the same predictors used in the final classifier (Figure 1C). C and D show the results of a logistic regression model that was trained to predict the query class using the same set of predictors as above, with the exception that the absolute values of  $m/z$  and RT error were used because logistic regression does not work well with nonmonotonic response behavior. This model also failed to find a relationship between query lipid class and the predictors. Taken together these results indicate that a distribution shift is unlikely to be the cause of poor generalization observed in Figure 4.

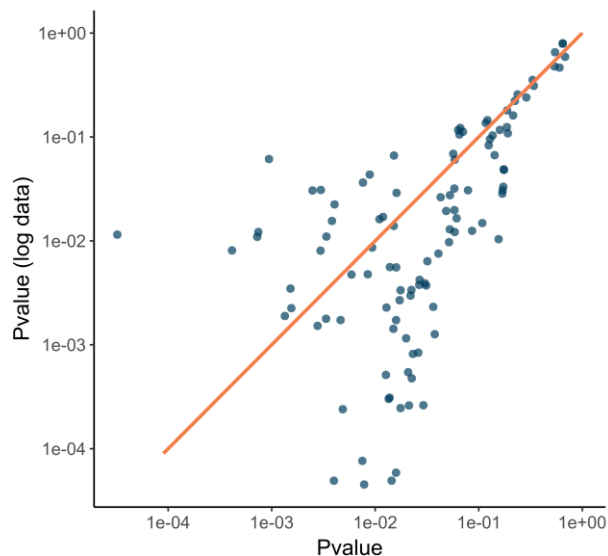


**Figure 3-1. P-values obtained with and without log-transformation of the raw data.** Data are from a study of *Clostridium autoethanogenum* with 108 manually quantified lipid species. Each dot represents the p-value of a given lipid obtained with and without log-transformation.

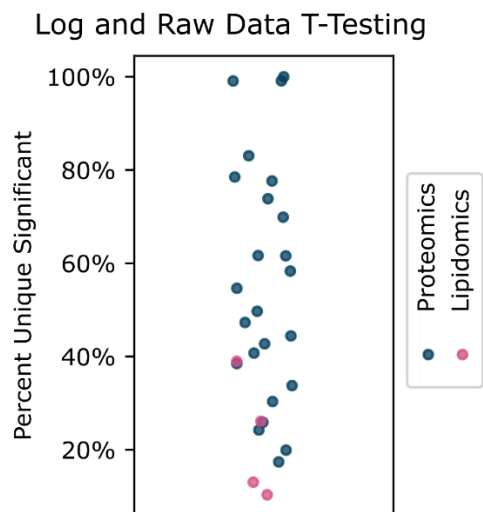


**Figure 3-2. Proportion of significant analytes that were only significant in one operation.** Horizontal scatter was applied to better visualize overlapping experiments and has no other meaning. In certain experiments, there were no shared significant proteins between log and no log.

**Table 3-1. Chromatography conditions used in the lipidomics replicability experiment.** Curve refers to the relative curvature applied to the mobile phase gradient. A curve of 5 is a linear interpolation while values <5 are convex (increasing rate of change across interpolation) while values >5 are concave.

| Time (min) | 0  | 3  | 8.5 | 61 | 77.5 | 79 | 89 |
|------------|----|----|-----|----|------|----|----|
| % A        | 80 | 80 | 45  | 1  | 1    | 80 | 80 |
| % B        | 20 | 20 | 55  | 99 | 99   | 20 | 20 |
| Curve      | 5  | 5  | 3   | 7  | 5    | 5  | 5  |

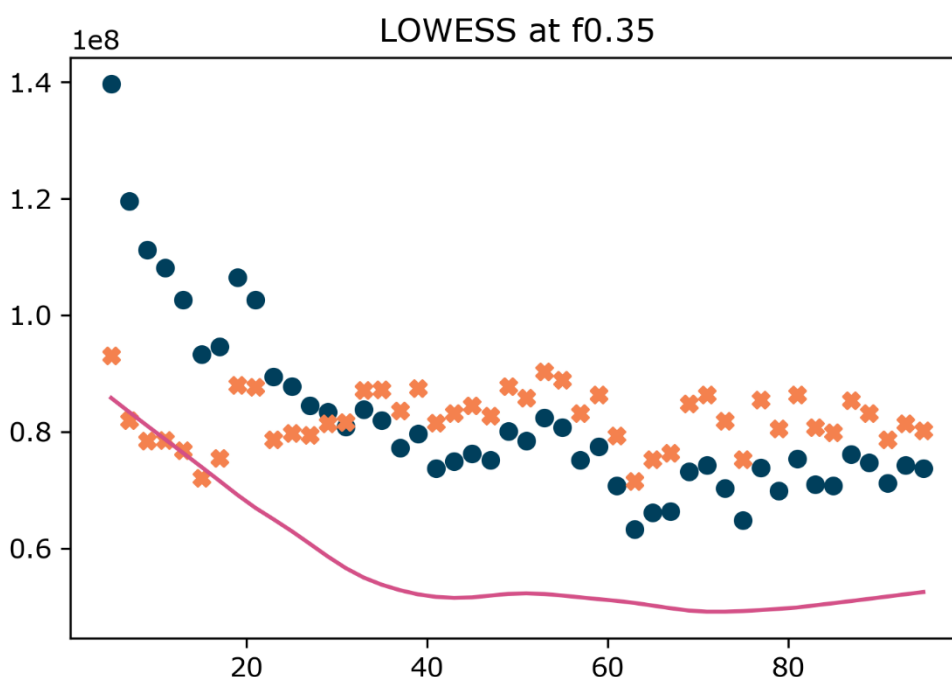


**Figure 3-3. Example of instrumental drift and LOWESS correction.**

The blue dots show the manually quantified peak areas of control replicates across the measurements. The pink curve represents the LOWESS regression of the solvent stress condition. The orange “X”s represent the LOWESS-corrected control values.

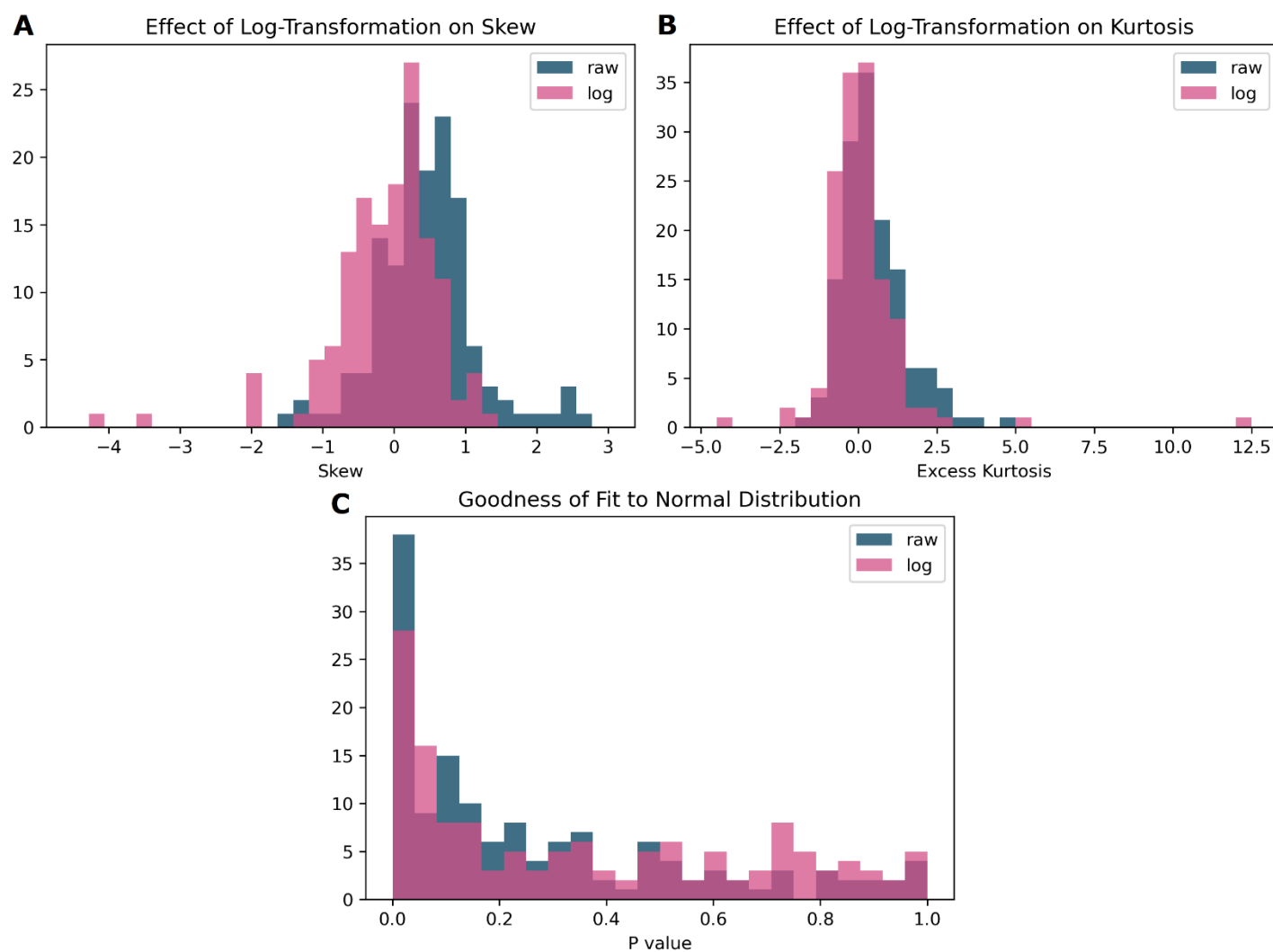


**Figure 3-4. The effect of log-transformation on distributional properties of lipidomic data.**

Panel A. shows the skew of lipid replicates. B. shows excess kurtosis of replicates. C. shows the p-values resulting from a goodness of fit test (compared to a normal distribution).

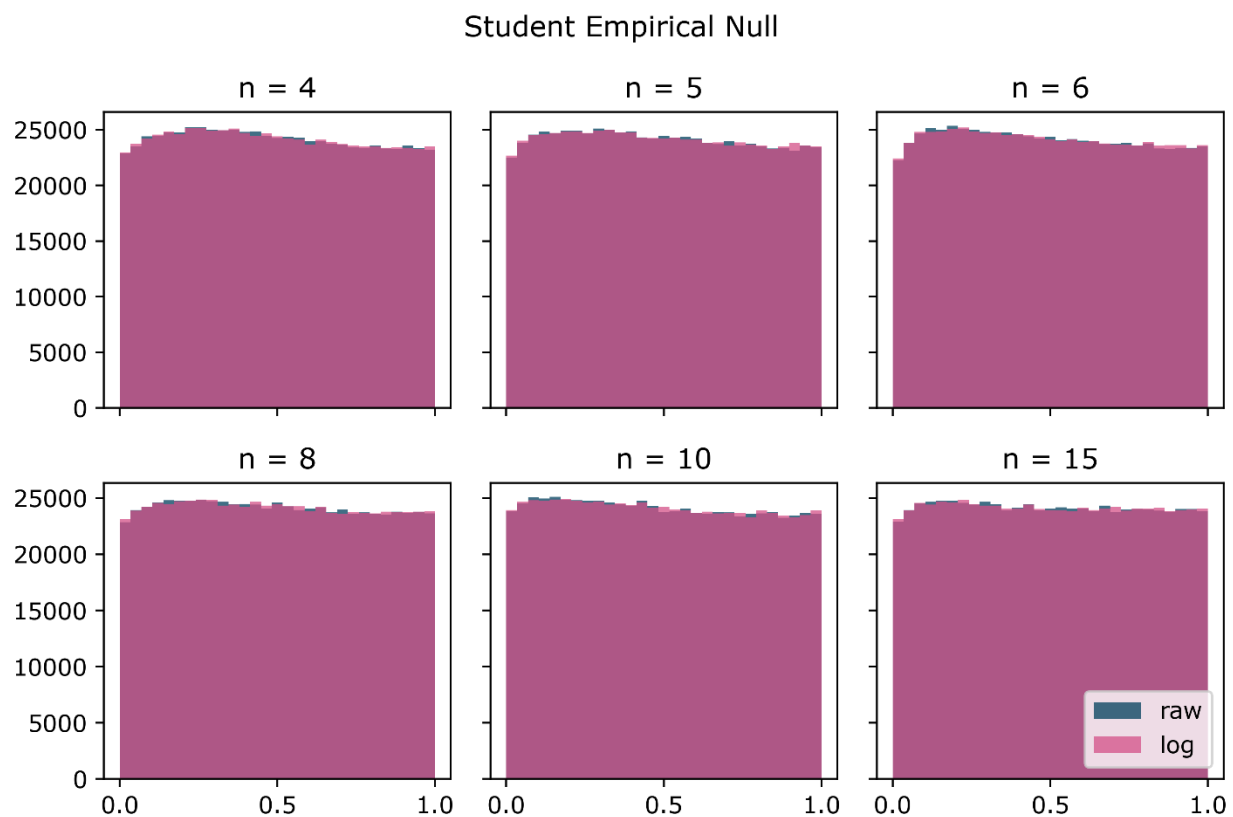


**Figure 3-5. P-value distribution obtained from lipidomics T-test empirical null**  
P-values are calculated from bootstrapped null comparisons using a Student two-tailed T-test.

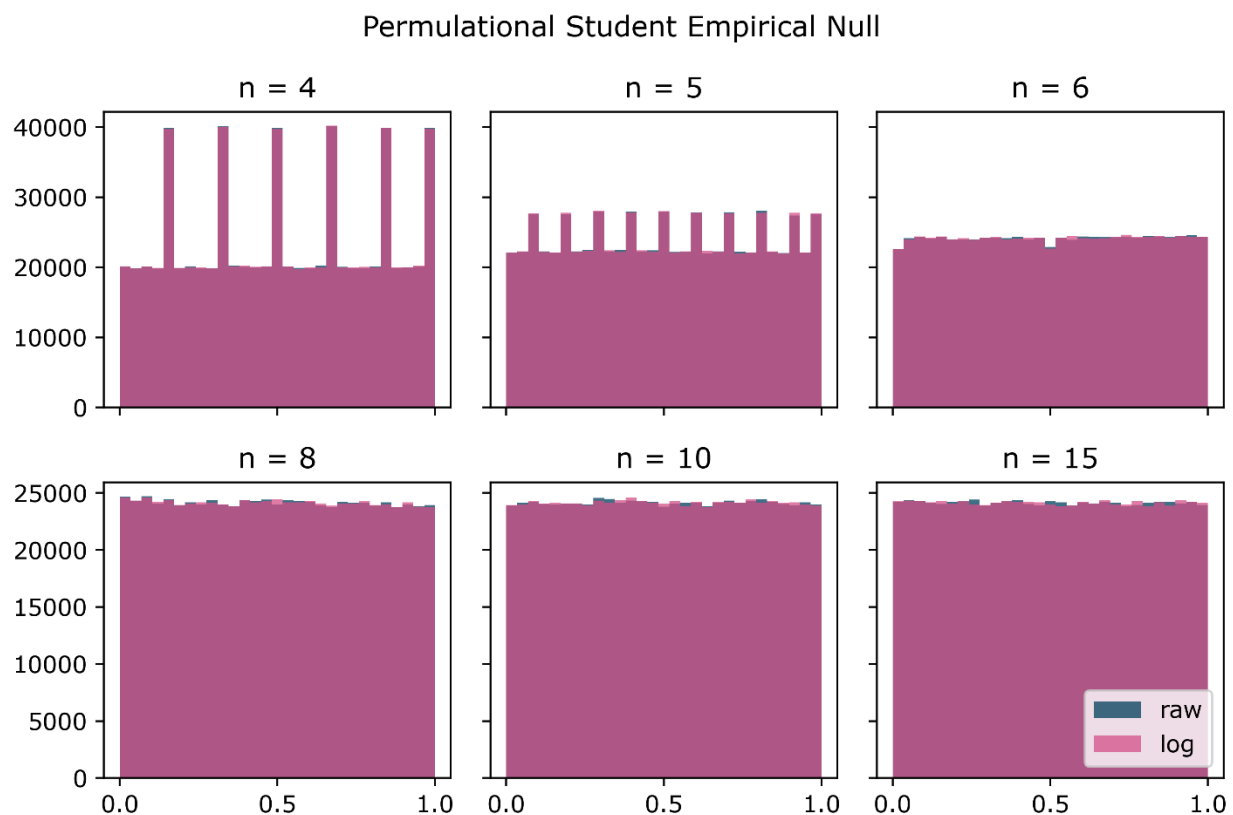


**Figure 3-6. P-value distribution obtained from lipidomics permutational T-test empirical null**

P-values are calculated from bootstrapped null comparisons using a permutational Student two-tailed T-test. For permutational calculation of significance, each test was limited to 5000 permutations (exact tests were performed when less than 5000 permutations were calculatable).

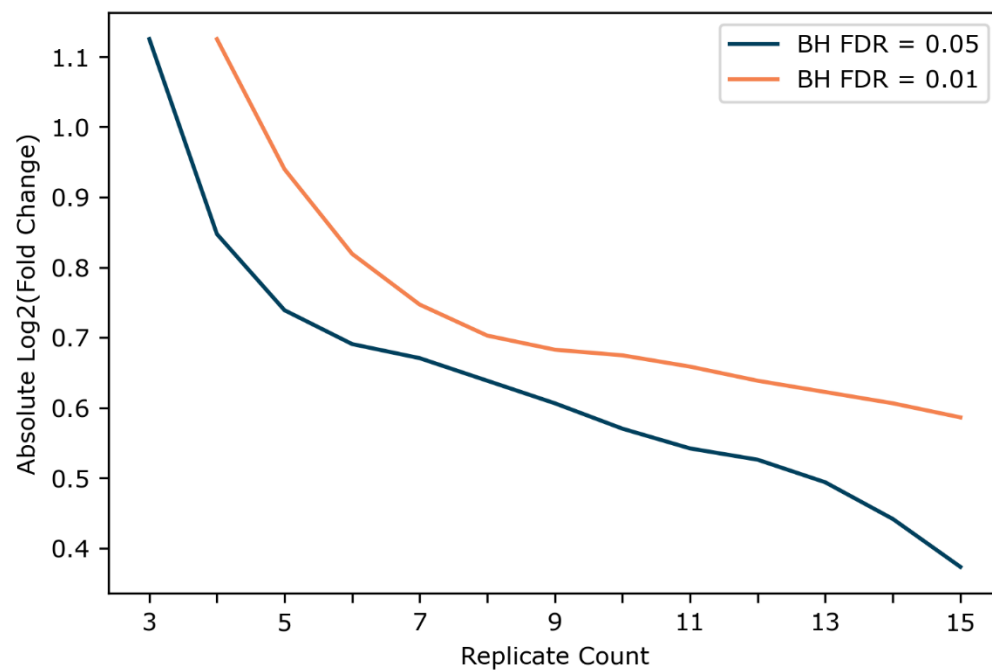


**Figure 3-7. Power based on LOWESS of FC vs frequency of significant plot at given BH cutoff**

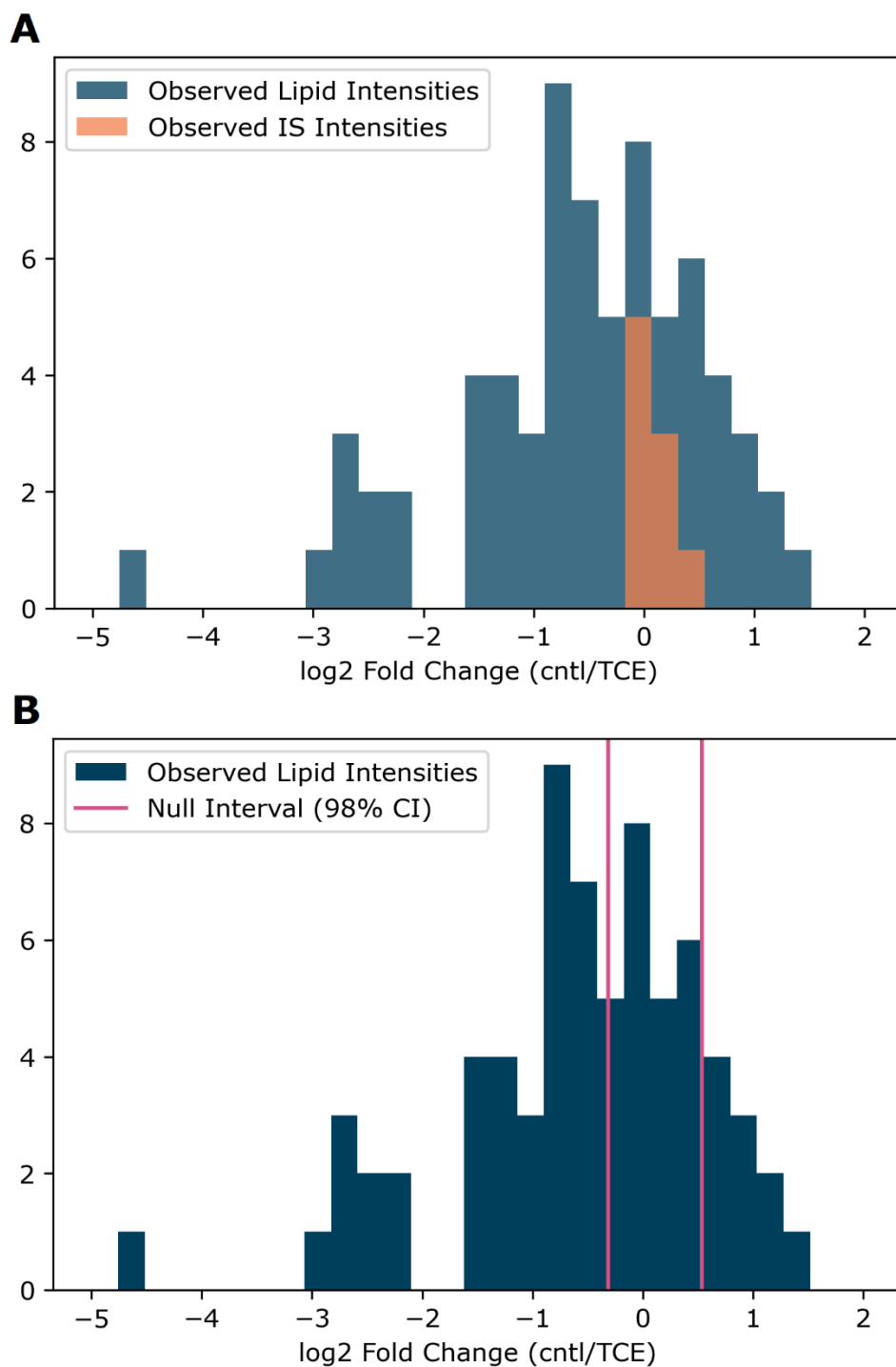


**Figure 3-8. Observed fold changes (control/solvent stress) from *P. putida*.**

A. Shows the internal standard fold changes (orange) overlaid on the lipid fold changes. B. Shows the lipid fold changes with a 98% confidence interval (parametric based on a t-distribution) constructed from the observed internal standard fold-changes (pink lines).

**Table 3-2. Normality test of lipid summary statistics**

| <b>Statistic</b>       | <b>Condition</b> | <b>Normal Goodness of Fit Pvalue</b> | <b>Shapiro Pvalue</b> |
|------------------------|------------------|--------------------------------------|-----------------------|
| <b>FA length</b>       | Control          | 0.854                                | 0.867                 |
| <b>FA length</b>       | TCE              | 0.0551                               | 0.0173                |
| <b>FA unsaturation</b> | Control          | 0.0234                               | 0.0699                |
| <b>FA unsaturation</b> | TCE              | 0.463                                | 0.310                 |
| <b>Cat/An</b>          | Control          | 0.0769                               | 0.127                 |
| <b>Cat/An</b>          | TCE              | 0.946                                | 0.934                 |

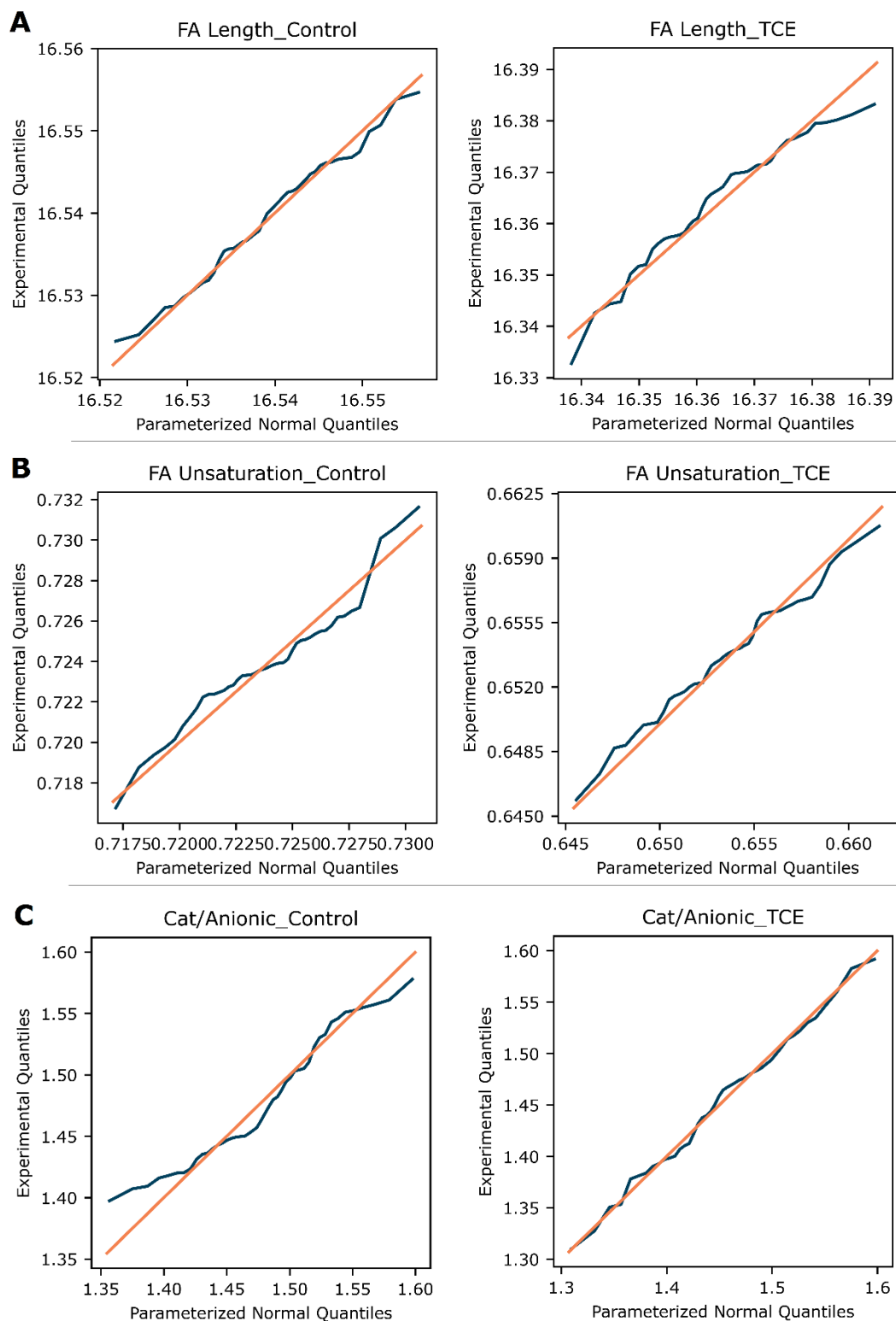


**Figure 3-9. Summary Statistic Q-Q Plots.**

Q-Q plots of the fatty acid length (A), fatty acid unsaturation (B), and cationic/anionic phospholipid ratio (C) summary statistics for control and solvent stress conditions.

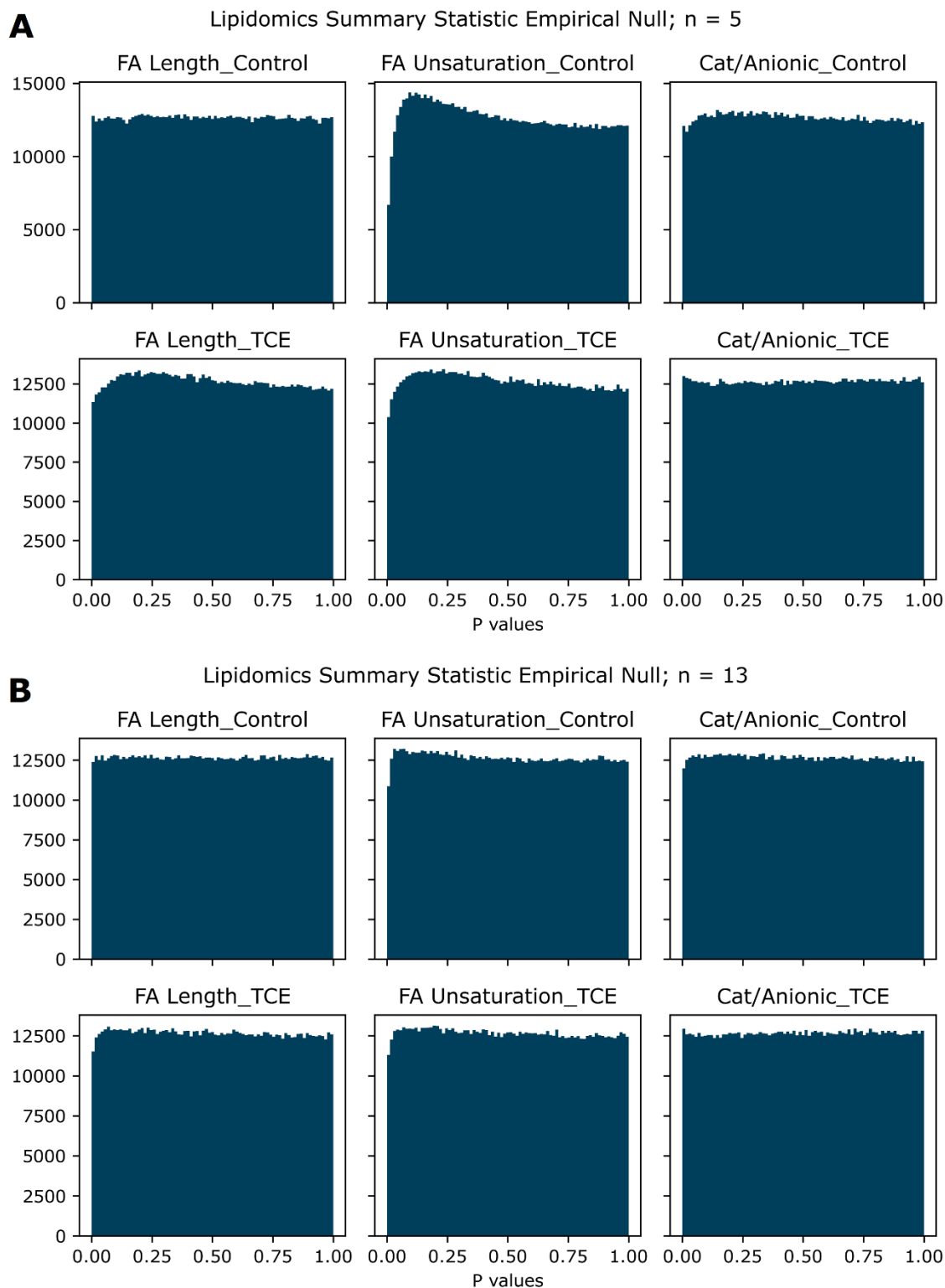


**Figure 3-10. Empirical null for summary statistics.**  
Histogram of summary stat empirical null p-values with five (A) and thirteen (B) replicates.

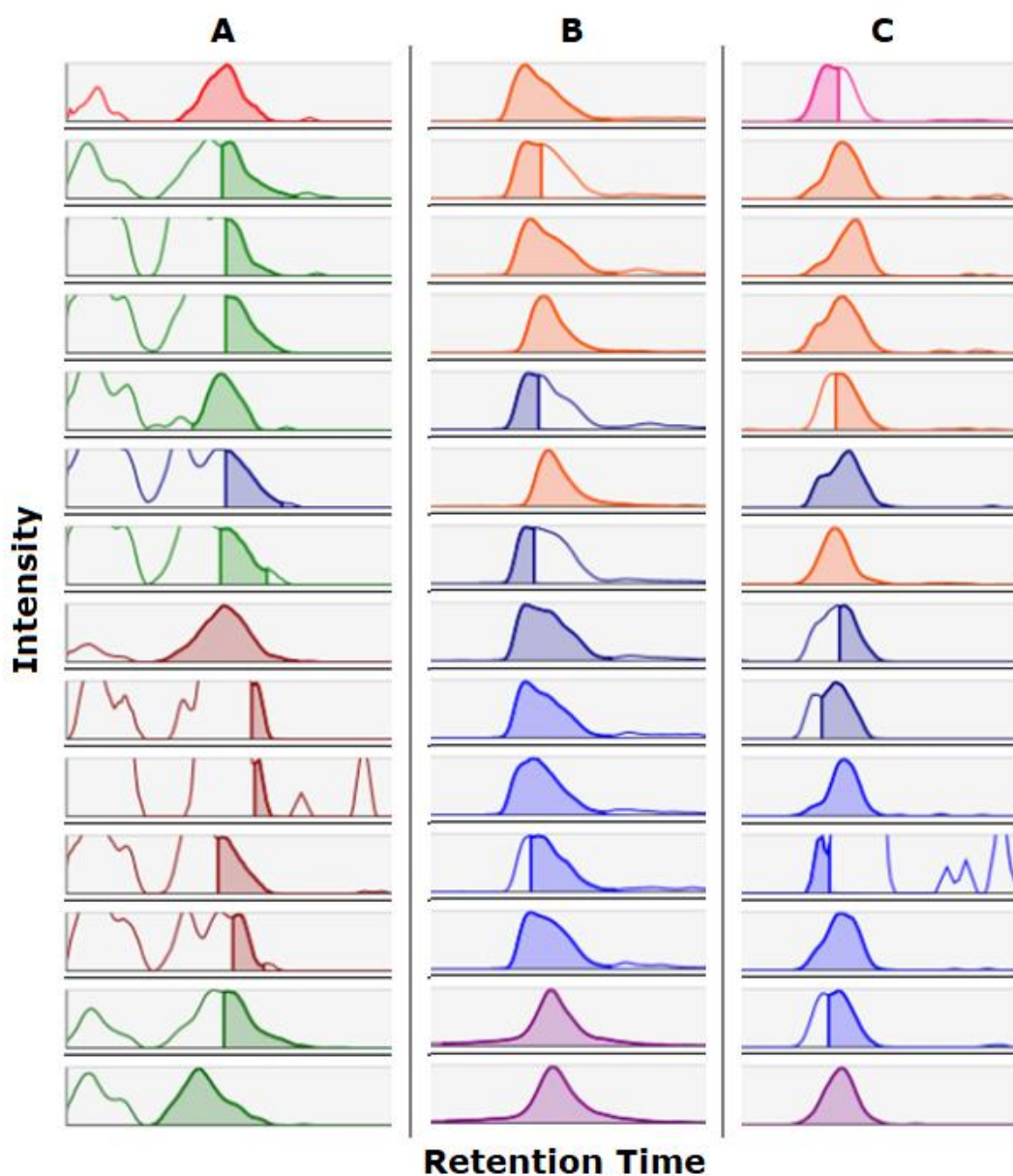


**Figure 3-11. Examples of Misintegration Performed by MS-DIAL**

Each column (A, B, and C) shows a feature reported by MS-DIAL with the peak integrations performed in different samples. These show different examples of inaccurate and/or inconsistent peak integration performed by MS-DIAL.

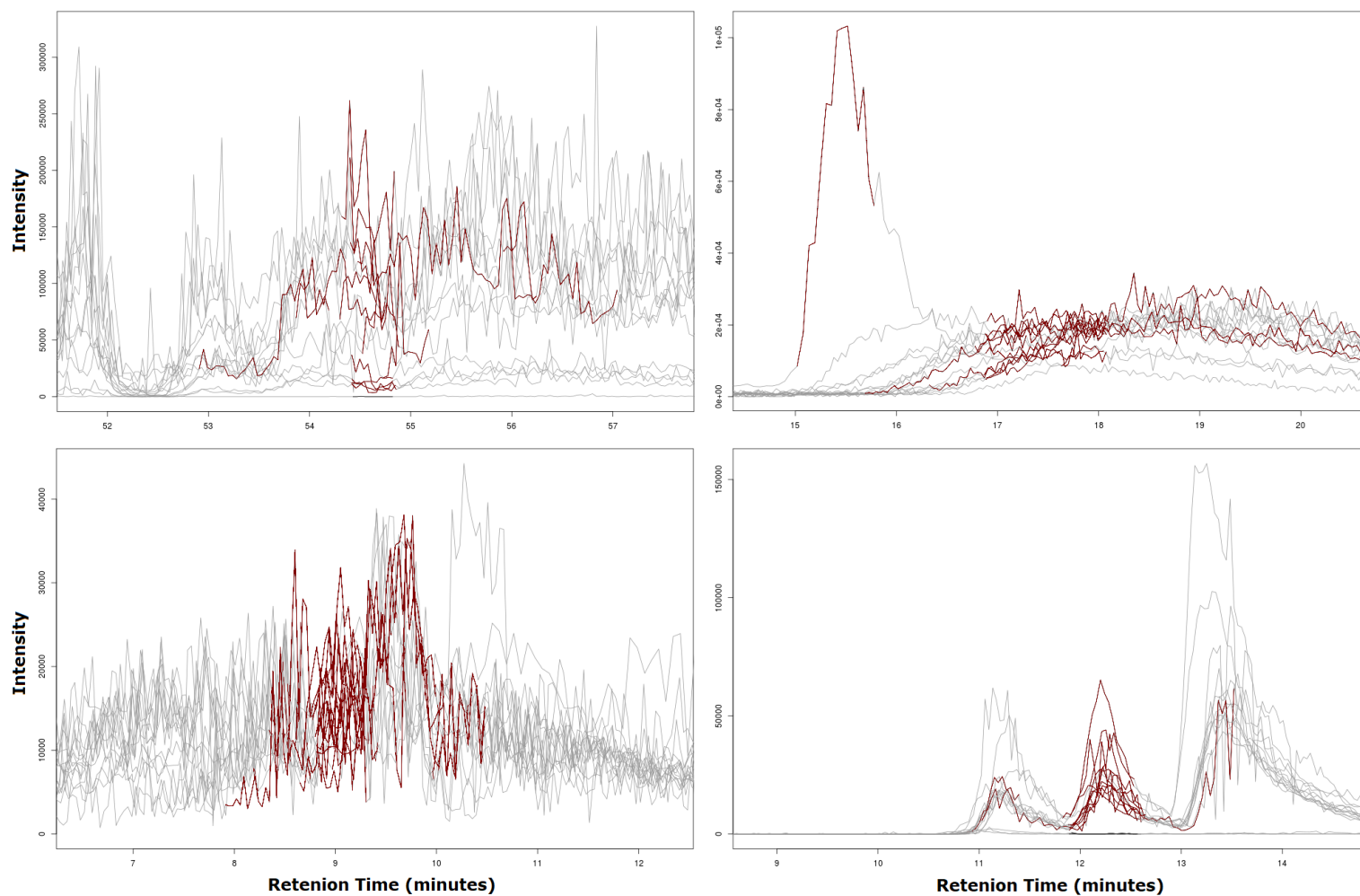


**Figure 3-12. Examples of Misintegration Performed by XCMS**

Features detected by XCMS with samples overlaid. The red traces show the regions integrated by XCMS. These show different examples of inaccurate and/or inconsistent peak integration performed by XCMS.

**Table 3-3. Comparison of quantification method precision.**

Data represent the 76 lipid species derived from MS-DIAL and LipoCLEAN before blank filtering. Cardiolipin identified and quantified in separate adduct states were treated independently for this analysis.

| <b>Condition Avg RSDs</b> | <b>El-MAVEN Manual</b> | <b>MS-DIAL Area</b> | <b>MS-DIAL Height</b> |
|---------------------------|------------------------|---------------------|-----------------------|
| <b>Control</b>            | 26.04%                 | 30.02%              | 29.19%                |
| <b>Control-Drift Corr</b> | 19.77%                 | 25.39%              | 23.28%                |
| <b>TCE</b>                | 20.34%                 | 24.15%              | 23.65%                |
| <b>TCE-Drift Corr</b>     | 18.61%                 | 24.09%              | 21.53%                |

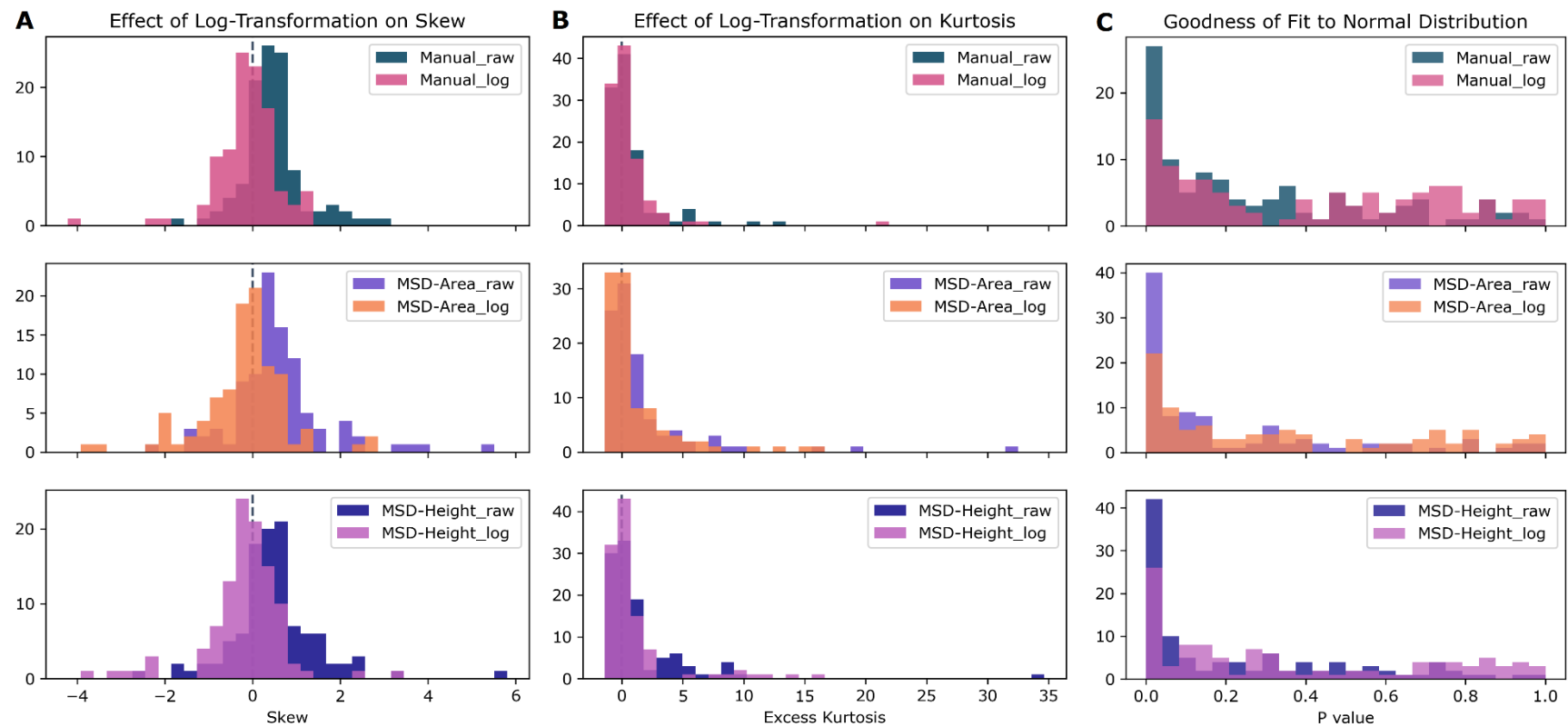


**Figure 3-13. Distributional properties from different quantification methods.**  
Observed skew (A), kurtosis (B), and normality (C) with and without log-transformation.

**Table 4-1. Chromatography conditions used in the lipidomics replicability experiment.**

| Time (min) | 0  | 3  | 4.5 | 9.5 | 15.5 | 50.5 | 60.8 | 77.5 | 79.5 | 89 |
|------------|----|----|-----|-----|------|------|------|------|------|----|
| % A        | 99 | 99 | 70  | 65  | 45   | 30   | 1    | 1    | 99   | 99 |
| % B        | 1  | 1  | 30  | 35  | 55   | 70   | 99   | 99   | 1    | 1  |

**Table 4-2. Number of Lipid Species per Lipid Class**

\*Indicates that the presence of lipid species corresponding to the given class have been manually confirmed. Other classes are based fully on LipoCLEAN results

| Lipid Class    | <i>A. fumigatus</i> 25 °C | <i>A. fumigatus</i> 37 °C | <i>L. bicolor</i> |
|----------------|---------------------------|---------------------------|-------------------|
| <b>Cer</b>     | 27                        | 26*                       | 33                |
| <b>CL</b>      | 29                        | 28*                       | 14*               |
| <b>DG</b>      | 13                        | 20*                       | 26*               |
| <b>FAHFA</b>   | 2                         | 0                         | 2                 |
| <b>HexCer</b>  | 6                         | 9                         | 17*               |
| <b>MG</b>      | 1                         | 1                         | 1                 |
| <b>NAE</b>     | 2                         | 0                         | 0                 |
| <b>LPC</b>     | 3                         | 8                         | 2                 |
| <b>PC</b>      | 21                        | 37*                       | 27                |
| <b>etherPC</b> | 8                         | 0                         | 15                |
| <b>LPE</b>     | 5                         | 7                         | 7*                |
| <b>PE</b>      | 34                        | 40*                       | 41*               |
| <b>etherPE</b> | 1                         | 1                         | 1                 |
| <b>PEtOH</b>   | 0                         | 0                         | 1                 |
| <b>PG</b>      | 5                         | 10*                       | 2                 |
| <b>LPI</b>     | 0                         | 1                         | 0                 |
| <b>PI</b>      | 17                        | 24*                       | 11                |
| <b>PMeOH</b>   | 0                         | 0                         | 2                 |
| <b>PS</b>      | 15                        | 17*                       | 11                |
| <b>etherPS</b> | 0                         | 0                         | 2                 |
| <b>SG</b>      | 0                         | 0                         | 3                 |
| <b>TG</b>      | 157                       | 150*                      | 126*              |
| <b>Total</b>   | 346                       | 379                       | 344               |

**Table 4-3. Average Carbon Number and Unsaturation by Lipid Class**

| Sample:        | <i>A. fumigatus</i> 25 °C |       | <i>A. fumigatus</i> 37 °C |       | <i>L. bicolor</i> |       |
|----------------|---------------------------|-------|---------------------------|-------|-------------------|-------|
| Lipid Class    | Carbon                    | Unsat | Carbon                    | Unsat | Carbon            | Unsat |
| <b>Cer</b>     | 14.2                      | 0.64  | 18.8                      | 0.24  | 20.1              | 0.13  |
| <b>CL</b>      | 17.6                      | 1.29  | 17.9                      | 1.76  | 18.0              | 1.94  |
| <b>DG</b>      | 17.8                      | 1.51  | 17.7                      | 1.49  | 17.9              | 1.65  |
| <b>FAHFA</b>   | 17.8                      | 1.93  | -                         | -     | 15.4              | 1.34  |
| <b>HexCer</b>  | 18.5                      | 1.43  | 18.5                      | 1.31  | 18.6              | 1.15  |
| <b>MG</b>      | 15.0                      | 2.00  | 15.0                      | 2.00  | 15.0              | 2.00  |
| <b>NAE</b>     | 24.6                      | 4.00  | -                         | -     | -                 | -     |
| <b>LPC</b>     | 18.0                      | 2.00  | 16.9                      | 0.92  | 18.0              | 1.93  |
| <b>PC</b>      | 17.8                      | 1.68  | 17.7                      | 1.42  | 17.8              | 1.64  |
| <b>etherPC</b> | 17.2                      | 1.47  | -                         | -     | 17.3              | 2.01  |
| <b>LPE</b>     | 17.4                      | 1.04  | 17.1                      | 0.91  | 17.8              | 1.69  |
| <b>PE</b>      | 17.5                      | 1.24  | 17.6                      | 1.28  | 17.5              | 1.46  |
| <b>etherPE</b> | 8.0                       | 0.00  | 21.0                      | 1.00  | 18.0              | 3.50  |
| <b>PEtOH</b>   | -                         | -     | -                         | -     | 18.0              | 2.00  |
| <b>PG</b>      | 17.1                      | 0.86  | 17.2                      | 1.15  | 17.2              | 1.15  |
| <b>LPI</b>     | -                         | -     | 16.0                      | 0.00  | -                 | -     |
| <b>PI</b>      | 17.4                      | 1.22  | 17.3                      | 1.15  | 17.5              | 1.44  |
| <b>PMeOH</b>   | -                         | -     | -                         | -     | 17.8              | 1.80  |
| <b>PS</b>      | 17.2                      | 1.10  | 17.3                      | 1.12  | 17.4              | 1.47  |
| <b>etherPS</b> | -                         | -     | -                         | -     | 17.9              | 1.89  |
| <b>SG</b>      | -                         | -     | -                         | -     | 14.1              | 0.87  |
| <b>TG</b>      | 17.7                      | 1.27  | 17.8                      | 1.41  | 17.9              | 1.47  |
| <b>Total</b>   |                           |       |                           |       |                   |       |

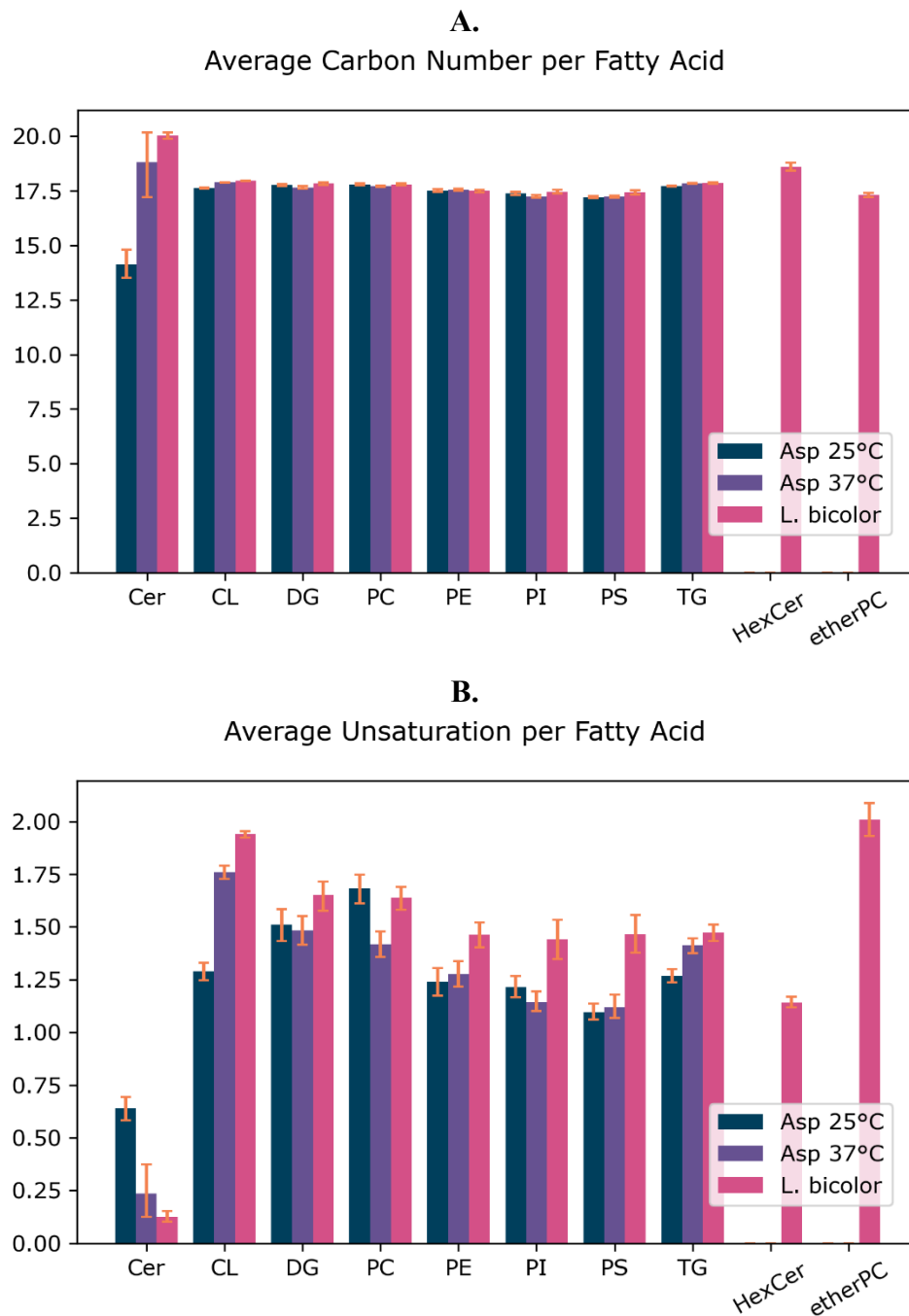


**Figure 4-1. Lipid Class Fatty Acid Characteristics**

A: Average fatty acid length for common lipid classes. B: Average fatty acid unsaturation for common lipid classes. For A and B only lipid classes with  $\geq 10$  species measured are shown. Error bars are based on bootstrapped confidence intervals with 10k iterations 2<sup>nd</sup> and 98<sup>th</sup> bootstrap percentiles are shown.

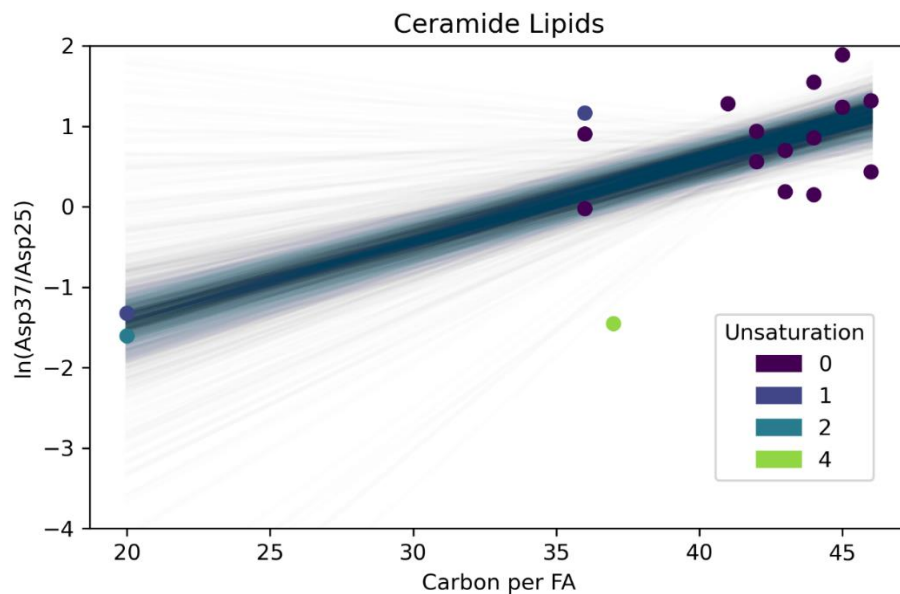


**Figure 4-2. Ceramide lipids between different temperatures.**

Natural logarithm of arithmetic fold-change vs carbon number for shared ceramide lipids in *A. fumigatus*. Shared ceramide lipids were mean-center across Asp25 and Asp37 for better relative comparison. Foldchange vs carbon number regression was performed 2k times on bootstrapped data. Resulting regressions lines are overlayed on the plot.

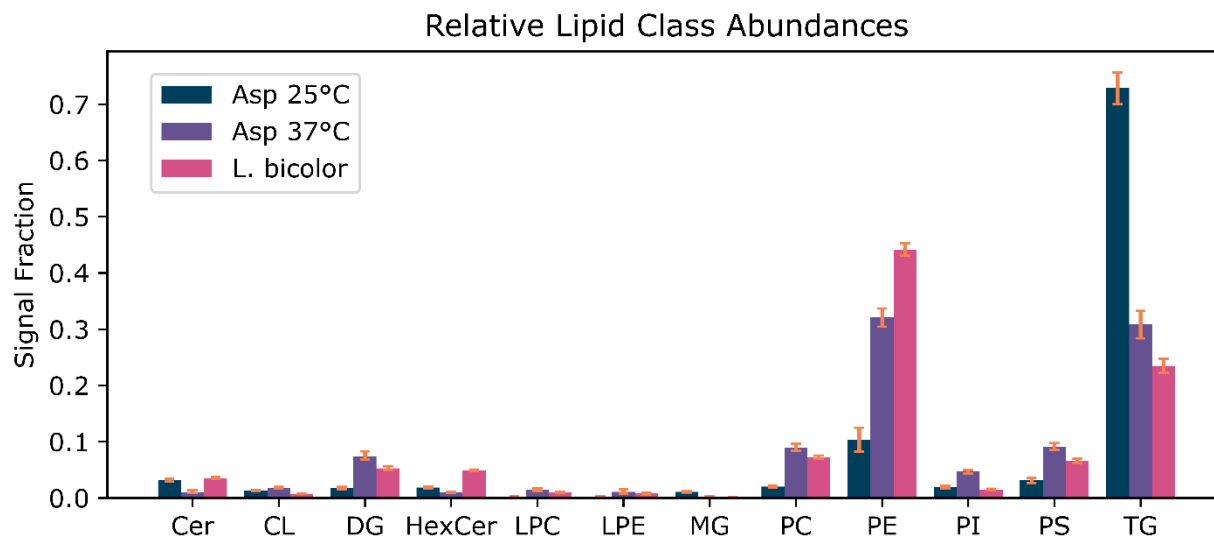


**Figure 4-3. Proportions of different lipid classes in the three datasets.** Only the top 12 lipid classes are shown. Error bars are based on bootstrapped confidence intervals (1<sup>st</sup> and 99<sup>th</sup> percentiles) with 10k iterations.

**Table 4-4. P-values obtained from ANOVA tests on lipidome summary statistics**

| Metric Tested                                 | A. fumigatus 25 °C | A. fumigatus 37 °C | L. bicolor |
|-----------------------------------------------|--------------------|--------------------|------------|
| Avg fatty acid length                         | 0.56               | 0.24               | 0.29       |
| Avg fatty acid unsaturation                   | 0.69               | 0.10               | 0.12       |
| Glycerophospholipid/glycerolipid              | 0.40               | 0.25               | 0.45       |
| Cationic/anionic<br>glycerophospholipid ratio | 0.27               | 0.39               | 0.72       |

**Table 4-5. Results from PERMANOVA.**

First value represents variance explained (%); second value shows associated p-values

| Condition      | A. fumigatus 25 °C | A. fumigatus 37 °C | L. Bicolor |
|----------------|--------------------|--------------------|------------|
| Overall        | 40%; 0.14          | 19%; 0.59          | 46%; 0.028 |
| 16nS           | 12%; 0.060         | 2%; 0.28           | 6%; 0.13   |
| 16S            | 1.6%; 0.46         | 0.6%; 0.82         | 32%; 0.31  |
| 18nS           | 0.3%; 0.70         | 1%; 0.65           | 8%; 0.06   |
| 18S            | 11%; 0.075         | 4%; 0.16           | 13%; 0.013 |
| CO4            | 3%; 0.29           | 7%; 0.1            | 4%; 0.22   |
| CO5            | 0.7%; 0.92         | 3%; 0.23           | 3%; 0.32   |
| CO8            | 12%; 0.046         | 0.1%; 0.95         | 8%; 0.07   |
| EtOH (Control) | 0.7%; 0.61         | 0.1%; 0.93         | 0.8%; 0.71 |

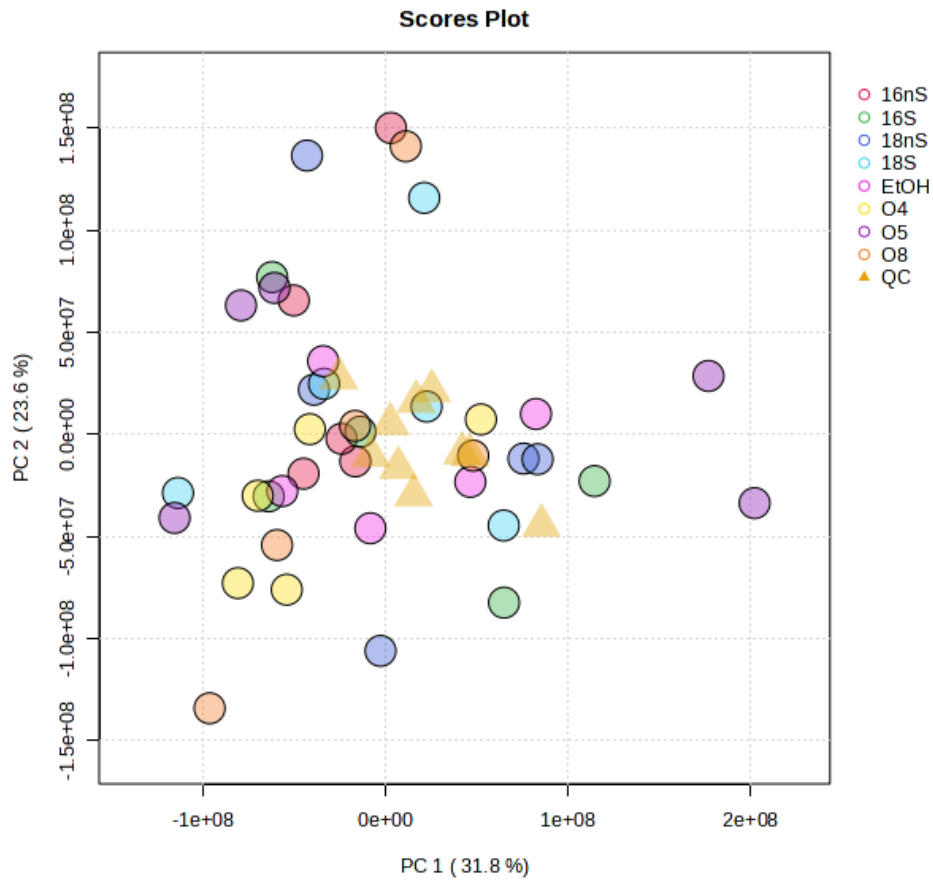


**Figure 4-4. PCA of the lipidomics data for *L. bicolor***

All quantified lipids from *L. bicolor* were used in a principal component analysis. The two most discriminatory dimensions are shown below. There are no observable differences between the classes that exceeds the intra-condition variability.

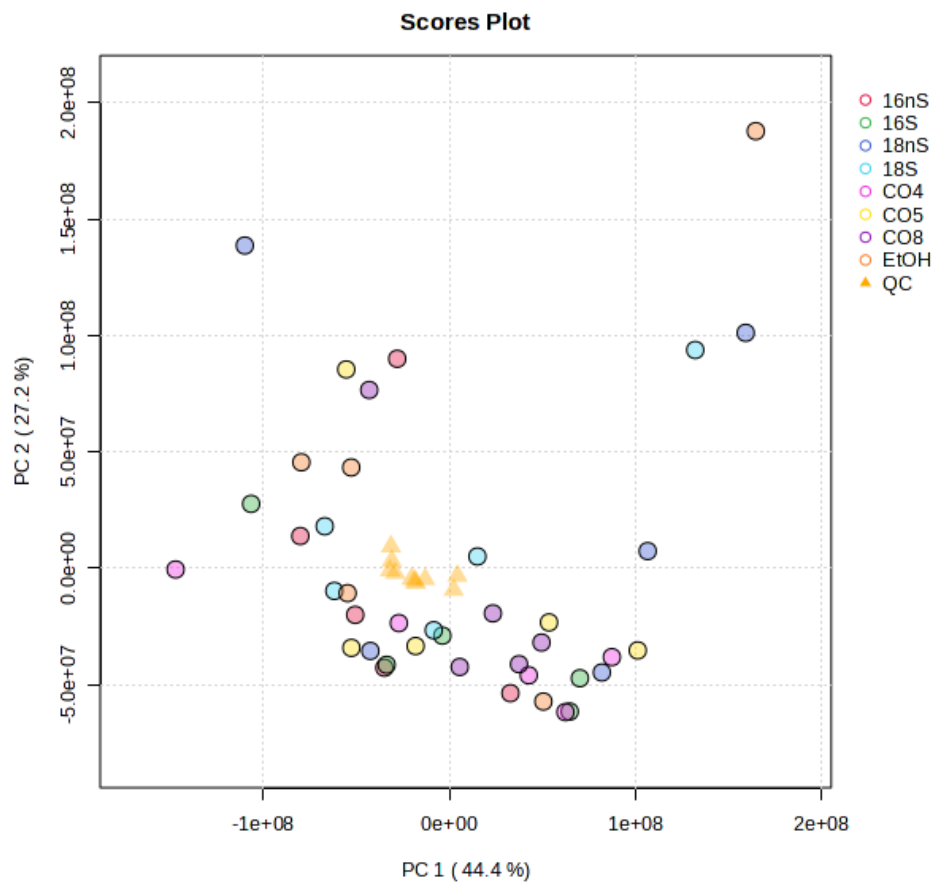


**Figure 4-5. PCA of the lipidomics data for *A. fumigatus* at 37 °C**

All quantified lipids from *A. fumigatus* grown at 37 °C were used in a principal component analysis. The two most discriminatory dimensions are shown below. There are no observable differences between the classes that exceeds the intra-condition variability.

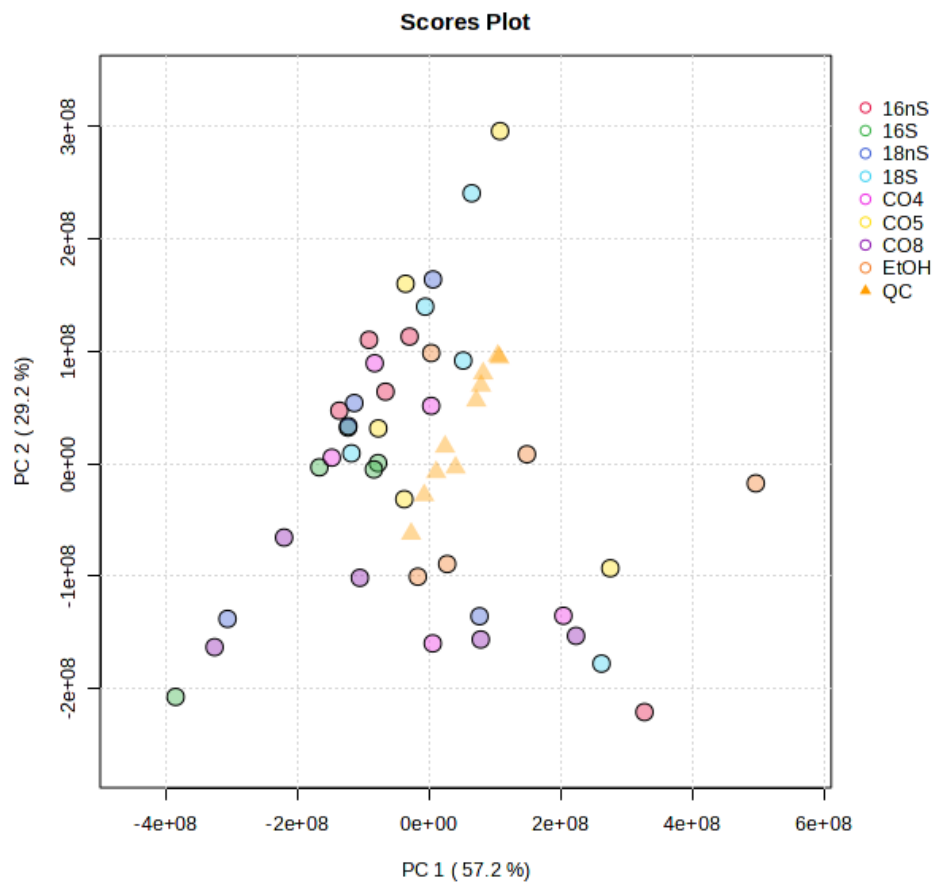


**Figure 4-6. PCA of the lipidomics data for *A. fumigatus* at 25 °C**

All quantified lipids from *A. fumigatus* grown at 25 °C were used in a principal component analysis. The two most discriminatory dimensions are shown below. There are no observable differences between the classes that exceeds the intra-condition variability.

**Table 5-1. Recipe for one liter of Enfors media.**

| Batch Recipe | Salts                                               | MWT    | mM Final    | g/L media  |
|--------------|-----------------------------------------------------|--------|-------------|------------|
|              | (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>     | 132.14 | 51.9        | 6.86       |
|              | Na <sub>2</sub> HPO <sub>4</sub> •7H <sub>2</sub> O | 268.07 | 36.6        | 9.81       |
|              | KH <sub>2</sub> PO <sub>4</sub>                     | 136.09 | 11.5        | 1.57       |
|              | (NH <sub>4</sub> ) <sub>2</sub> -H-Citrate          | 226.18 | 2.17        | 0.49       |
|              | MgSO <sub>4</sub> •7H <sub>2</sub> O                | 246.47 | 1           | 1ml @ 1M   |
|              | Trace Metals                                        |        |             | 1 ml       |
|              | Glycerol                                            |        |             | 5          |
| Traces       | Salts                                               | MWT    | final conc. | g/L traces |
|              | CaCl <sub>2</sub> •2H <sub>2</sub> O                | 147.01 | 3.4 µM      | 0.5        |
|              | CoCl <sub>2</sub>                                   | 129.84 | 0.75 µM     | 0.098      |
|              | CuSO <sub>4</sub> •5H <sub>2</sub> O                | 249.69 | 0.64 µM     | 0.16       |
|              | FeCl <sub>3</sub> •6H <sub>2</sub> O                | 270.3  | 62 µM       | 16.7       |
|              | MnSO <sub>4</sub> •H <sub>2</sub> O                 | 169.02 | 0.67 µM     | 0.114      |
|              | Na <sub>2</sub> EDTA•2H <sub>2</sub> O              | 372.24 | 54 µM       | 20.1       |
|              | ZnSO <sub>4</sub> •7H <sub>2</sub> O                | 287.56 | 0.63 µM     | 0.18       |

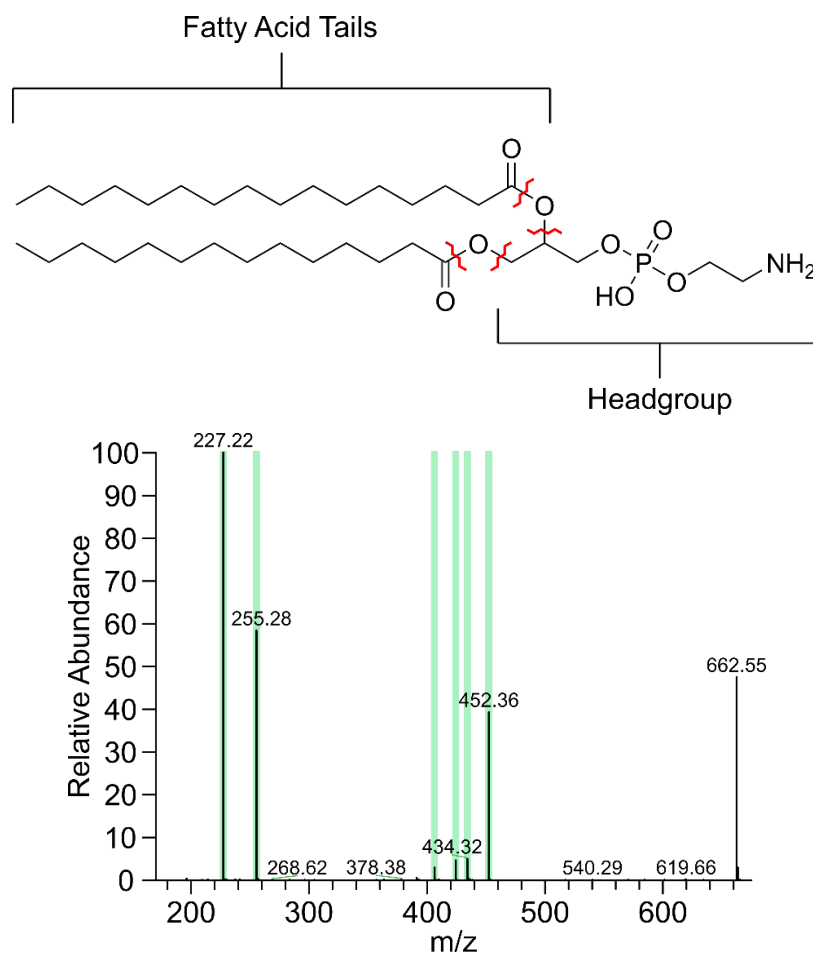
**Table 5-2. Sample ID and growth conditions.**

| Sample ID   | %D <sub>2</sub> O | Carbon             | n |
|-------------|-------------------|--------------------|---|
| H-Gly/0     | 0                 | H-Glycerol         | 1 |
| H-Gly/25    | 25                | H-Glycerol         | 1 |
| H-Gly/50    | 50                | H-Glycerol         | 1 |
| H-Gly/75    | 75                | H-Glycerol         | 1 |
| H-Gly/100   | 100               | H-Glycerol         | 3 |
| 50D-Gly/0   | 0                 | 50/50 H/D-Glycerol | 3 |
| DGly/0      | 0                 | D-Glycerol         | 3 |
| 50D-Gly/100 | 100               | 50/50 H/D-Glycerol | 3 |
| D-Gly/100   | 100               | D-Glycerol         | 3 |
| D-Glu/0     | 0                 | D-Glucose          | 3 |
| H-Glu/25    | 25                | H-Glucose          | 3 |
| H-Glu/50    | 50                | H-Glucose          | 3 |
| H-Glu/75    | 75                | H-Glucose          | 3 |
| H-Glu/100   | 100               | H-Glucose          | 1 |
| D-Glu/100   | 100               | D-Glucose          | 1 |

**Table 5-3. LC gradient.**

Concentrations of mobile phases across the length of the gradient.

| Time (min)    | 0  | 1  | 2.5 | 7.5 | 13.5 | 48.5 | 58.5 | 75.5 | 77.5 | 87 |
|---------------|----|----|-----|-----|------|------|------|------|------|----|
| Solvent A (%) | 99 | 99 | 70  | 65  | 45   | 30   | 1    | 1    | 99   | 99 |
| Solvent B (%) | 1  | 1  | 30  | 45  | 55   | 70   | 99   | 99   | 1    | 1  |

**Figure 5-1. Structure and fragmentation of example PE lipid.**

The structure of a glycerophospholipid (PE 30:0) is shown. Red marks indicate collisional-induced dissociation fragmentation locations. The lower panel shows the fragmentation spectra resulting from such a lipid. Green highlights indicate expected fragment ions.

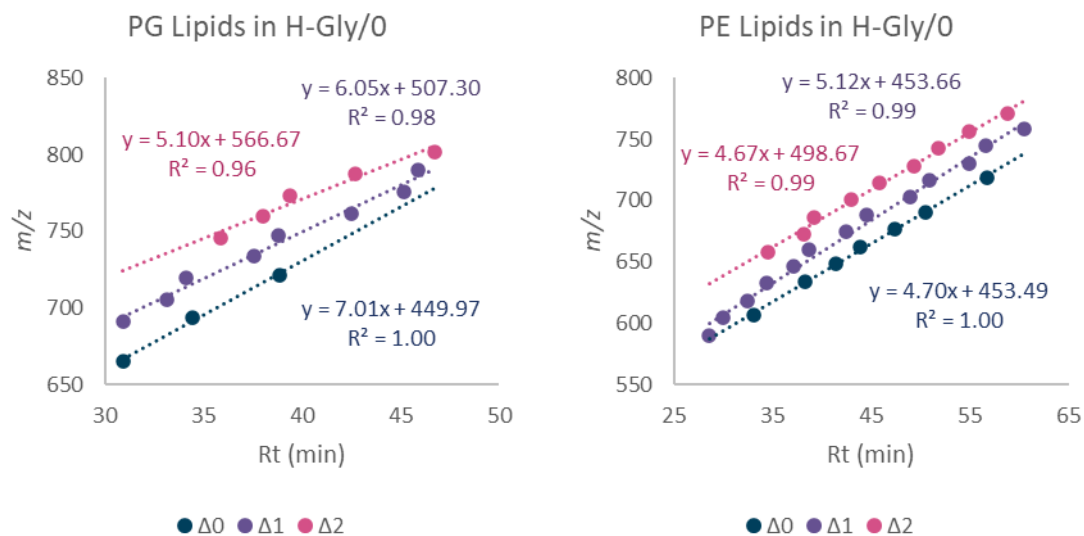


**Figure 5-2. Retention time regression of identified lipids.**

Mass to charge vs retention time plots demonstrate the linear relationship that supports lipid identifications.

**Table 5-4. Average lipid component volumes used for SLD calculation.**

| Group           | Volume/ $\text{\AA}^3$ <sup>285, 286</sup> |
|-----------------|--------------------------------------------|
| CH <sub>3</sub> | 52.7                                       |
| CH <sub>2</sub> | 28.1                                       |
| C=C             | 45.0                                       |
| PG Headgroup    | 300                                        |
| PE Headgroup    | 284.8                                      |

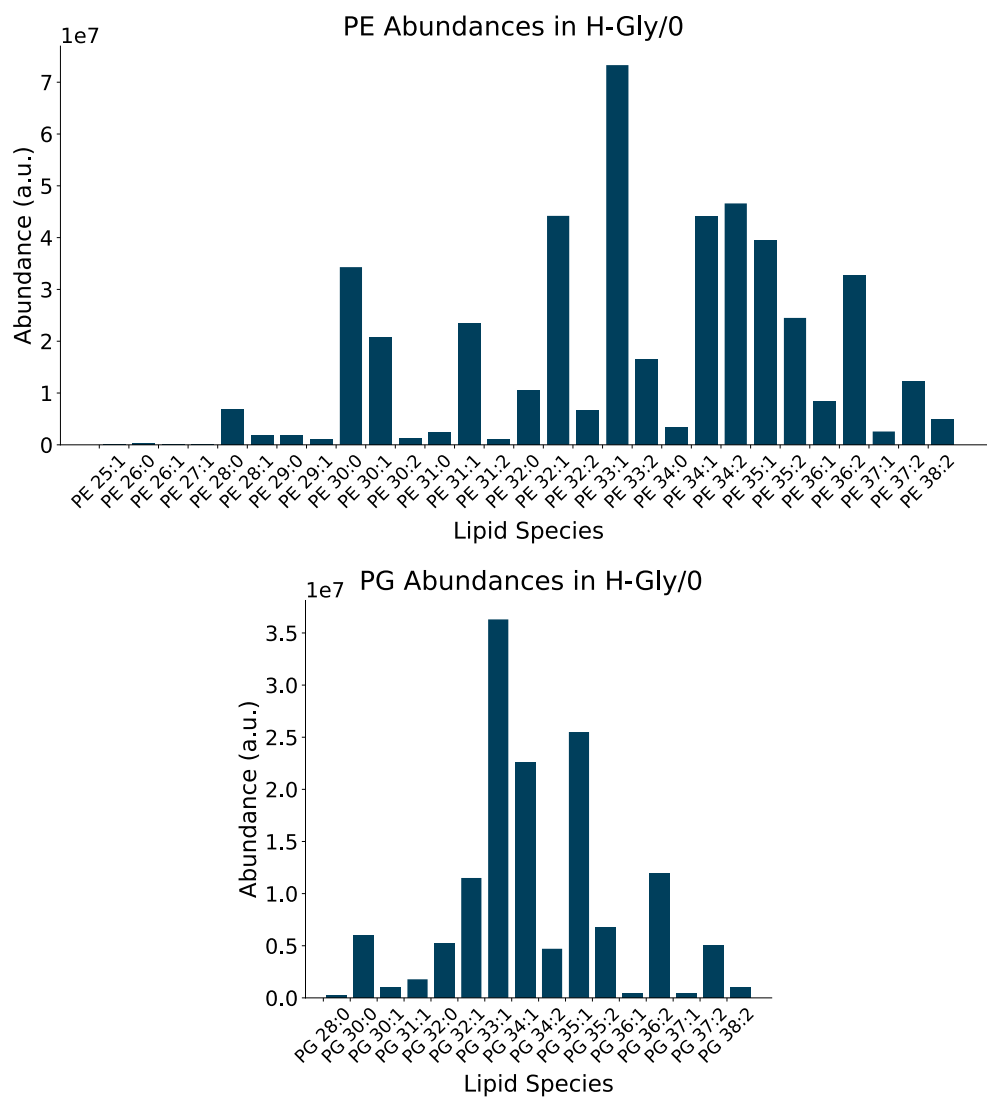


**Figure 5-3. Lipid abundances of PE and PG lipids in the unlabeled sample.**

Peak areas for PG and PE lipids identified in the unlabeled sample. Values are in arbitrary units (a.u.) and represent the integration of the extracted ion chromatogram for each lipid.

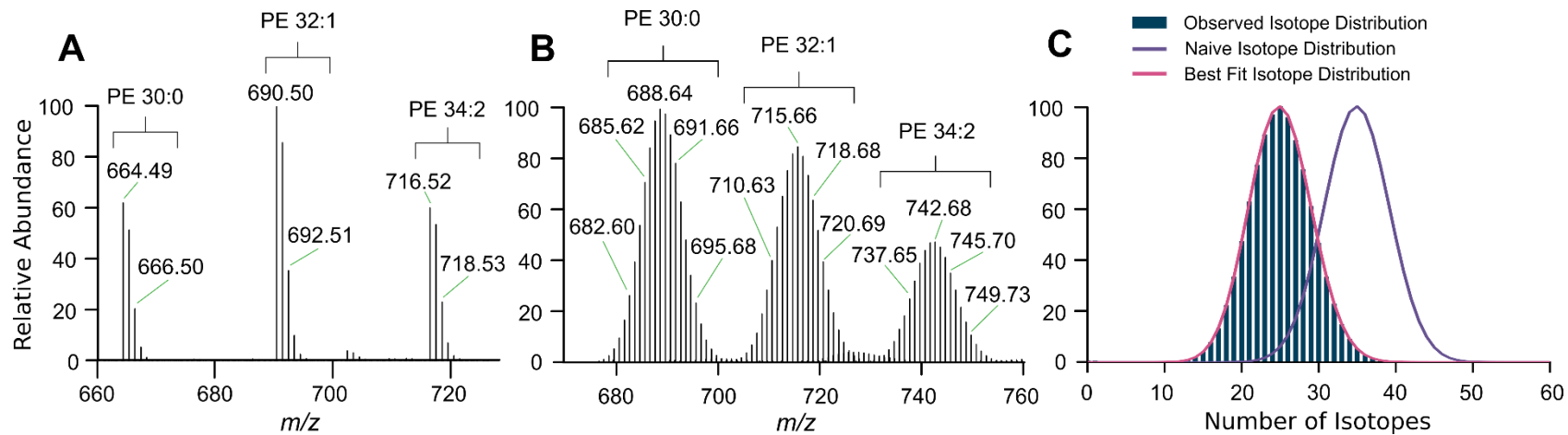


**Figure 5-4. Example mass spectra and isotope fitting.**

A and B are averaged mass spectra from H-Gly/0 (A) and H-Gly/50 (B) showing the isotope packets for the coeluting lipids PE 30:0, PE 32:1, and PE 34:2. (C) shows PE 32:2 lipid isotope data from H-Gly/50 with the binomial model fit to the observed data.

**Table 5-5. Best fit deuterium probabilities (p<sub>D</sub>) for PE lipids.**

|                | H-Gly<br>25%D2O | H-Glu<br>25%D2O | H-Gly<br>50%D2O | H-Glu<br>50%D2O | H-Gly<br>75%D2O | H-Glu<br>75%D2O | H-Gly<br>100%D2O | H-Glu<br>100%D2O | D-Gly<br>100%D2O | D-Glu<br>100%D2O |
|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|------------------|------------------|------------------|------------------|
| <b>PE 28:1</b> | 16.6%           | 15.9%           | 35.1%           | 34.8%           | 55.0%           | 57.3%           | 83.7%            | 82.2%            | 99.2%            | 98.7%            |
| <b>PE 28:0</b> | 16.7%           | 15.6%           | 35.0%           | 34.6%           | 55.2%           | 59.3%           | 82.7%            | 82.5%            | 99.2%            | 98.7%            |
| <b>PE 29:1</b> | 16.3%           | 15.4%           | 34.5%           | 34.1%           | 54.9%           | 56.7%           | 82.0%            | 83.4%            | 99.2%            | 98.6%            |
| <b>PE 29:0</b> | 17.1%           | 9.8%            | 35.1%           | 23.9%           | 54.9%           | 57.1%           | 83.5%            | 84.3%            | 99.1%            | 98.5%            |
| <b>PE 30:2</b> | 16.3%           | 15.6%           | 34.5%           | 33.6%           | 54.8%           | 56.2%           | 82.4%            | 82.4%            | 99.2%            | 98.6%            |
| <b>PE 30:1</b> | 16.6%           | 15.9%           | 34.8%           | 34.3%           | 55.2%           | 56.9%           | 82.7%            | 82.1%            | 99.2%            | 98.6%            |
| <b>PE 30:0</b> | 17.3%           | 15.6%           | 36.0%           | 34.9%           | 56.8%           | 57.5%           | 84.3%            | 82.8%            | 99.2%            | 98.7%            |
| <b>PE 31:1</b> | 16.6%           | 15.5%           | 35.2%           | 34.1%           | 55.7%           | 56.4%           | 82.5%            | 82.6%            | 99.2%            | 98.6%            |
| <b>PE 31:0</b> | 17.0%           | 10.5%           | 35.7%           | 34.3%           | 54.8%           | 57.4%           | 56.2%            | 82.3%            | 99.0%            | 98.5%            |
| <b>PE 32:2</b> | 16.7%           | 15.8%           | 35.0%           | 34.1%           | 55.8%           | 56.7%           | 84.1%            | 82.8%            | 99.3%            | 98.7%            |
| <b>PE 32:1</b> | 17.0%           | 15.9%           | 35.7%           | 34.5%           | 56.5%           | 56.6%           | 84.5%            | 82.5%            | 99.2%            | 98.7%            |
| <b>PE 32:0</b> | 17.4%           | 15.0%           | 36.2%           | 34.8%           | 57.2%           | 57.3%           | 84.9%            | 82.2%            | 99.2%            | 98.6%            |
| <b>PE 33:2</b> | 16.5%           | 15.7%           | 34.9%           | 33.9%           | 55.5%           | 56.3%           | 83.5%            | 82.6%            | 99.2%            | 98.7%            |
| <b>PE 33:1</b> | 17.1%           | 15.9%           | 35.7%           | 34.5%           | 56.6%           | 57.1%           | 83.5%            | 82.4%            | 99.2%            | 98.7%            |
| <b>PE 34:2</b> | 16.5%           | 15.5%           | 35.0%           | 34.0%           | 55.7%           | 56.2%           | 84.1%            | 82.3%            | 99.2%            | 98.7%            |
| <b>PE 34:1</b> | 17.3%           | 16.0%           | 36.1%           | 34.9%           | 57.2%           | 57.1%           | 85.1%            | 82.3%            | 99.2%            | 98.7%            |
| <b>PE 34:0</b> | 17.0%           | 12.1%           | 36.2%           | 32.7%           | 56.5%           | 55.7%           | 85.2%            | 82.3%            | 99.2%            | 98.7%            |
| <b>PE 35:2</b> | 17.0%           | 15.8%           | 35.8%           | 34.4%           | 56.6%           | 57.1%           | 84.9%            | 82.7%            | 99.2%            | 98.7%            |
| <b>PE 35:1</b> | 17.4%           | 15.6%           | 36.4%           | 34.7%           | 57.4%           | 57.3%           | 84.3%            | 82.4%            | 99.2%            | 98.7%            |
| <b>PE 36:2</b> | 17.0%           | 15.7%           | 36.0%           | 34.5%           | 57.0%           | 56.7%           | 85.3%            | 82.5%            | 99.2%            | 98.7%            |
| <b>PE 36:1</b> | 17.1%           | 14.9%           | 36.0%           | 34.5%           | 57.4%           | 57.2%           | 86.0%            | 82.6%            | 99.2%            | 98.7%            |
| <b>PE 37:2</b> | 17.1%           | 15.7%           | 36.3%           | 34.8%           | 57.5%           | 57.4%           | 86.0%            | 82.7%            | 99.2%            | 98.7%            |
| <b>PE 37:1</b> | 16.9%           | 14.7%           | 35.3%           | 34.6%           | 58.6%           | 57.5%           | 85.1%            | ND               | 99.2%            | 98.8%            |

**Table 5-6. Best fit deuterium probabilities (pd) for PG lipids.**

|                    | H-Gly<br>25%D2O | H-Glu<br>25%D2O | H-Gly<br>50%D2O | H-Glu<br>50%D2O | H-Gly<br>75%D2O | H-Glu<br>75%D2O | H-Glu<br>75%D2O<br>Repeat | H-Gly<br>100%<br>D2O | H-Glu<br>100%<br>D2O | D-Gly<br>100%<br>D2O | D-Glu<br>100%<br>D2O |
|--------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|---------------------------|----------------------|----------------------|----------------------|----------------------|
| <b>PG<br/>30:1</b> | 15.5%           | 15.7%           | 32.7%           | 33.9%           | 52.1%           | 57.7%           | 55.3%                     | 77.5%                | 82.7%                | 99.2%                | 98.6%                |
| <b>PG<br/>30:0</b> | 16.2%           | 15.8%           | 33.8%           | 34.5%           | 53.3%           | 58.3%           | 56.0%                     | 78.1%                | 83.7%                | 99.2%                | 98.7%                |
| <b>PG<br/>31:1</b> | 15.4%           | 15.4%           | 32.7%           | 33.7%           | 51.9%           | 51.2%           | 55.3%                     | 77.2%                | 82.7%                | 99.3%                | 99.6%                |
| <b>PG<br/>32:1</b> | 15.7%           | 15.7%           | 33.2%           | 34.1%           | 53.3%           | 56.6%           | 55.8%                     | 79.4%                | 82.1%                | 99.2%                | 98.7%                |
| <b>PG<br/>32:0</b> | 16.0%           | 15.9%           | 33.9%           | 34.7%           | 53.3%           | 56.2%           | 56.3%                     | 79.1%                | 81.9%                | 99.3%                | 98.7%                |
| <b>PG<br/>33:1</b> | 15.7%           | 15.4%           | 33.3%           | 33.5%           | 52.7%           | 55.6%           | 55.2%                     | 78.4%                | 82.1%                | 99.3%                | 98.8%                |
| <b>PG<br/>34:2</b> | 15.5%           | 15.5%           | 32.8%           | 33.7%           | 52.4%           | 55.9%           | 55.3%                     | 79.5%                | 82.2%                | 99.3%                | 98.6%                |
| <b>PG<br/>34:1</b> | 16.1%           | 15.9%           | 34.2%           | 34.3%           | 54.2%           | 56.3%           | 56.0%                     | 81.0%                | 82.3%                | 99.2%                | 98.7%                |
| <b>PG<br/>35:2</b> | 15.6%           | 15.4%           | 33.2%           | 33.5%           | 53.1%           | 55.6%           | 55.2%                     | 79.8%                | 82.2%                | 99.1%                | 98.6%                |
| <b>PG<br/>35:1</b> | 16.2%           | 15.5%           | 34.2%           | 34.0%           | 54.3%           | 56.0%           | 55.8%                     | 80.2%                | 81.8%                | 99.3%                | 98.7%                |
| <b>PG<br/>36:2</b> | 15.9%           | 15.6%           | 33.7%           | 34.1%           | 54.0%           | 56.1%           | 55.8%                     | 81.1%                | 82.5%                | 99.3%                | 98.5%                |
| <b>PG<br/>37:2</b> | 16.2%           | 15.7%           | 34.3%           | 34.1%           | 54.9%           | 56.1%           | 55.9%                     | 81.7%                | 82.4%                | 99.1%                | 98.7%                |

**Table 5-7. Deuterium incorporation (best fitting pD distribution) specific to fatty acids**

| <b>Condition</b>                               | <b>FA 16:0<br/>(PG 32:0)</b> | <b>FA 16:0<br/>(PE 32:0)</b> | <b>FA 17:1<br/>(PG 34:2)</b> | <b>FA 17:1<br/>(PE 34:2)</b> | <b>FA 18:1<br/>(PG 36:2)</b> | <b>FA 18:1<br/>(PE 36:2)</b> |
|------------------------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|
| <b>H-Gly/25</b>                                | 18.8%                        | 19.0%                        | 18.4%                        | 17.8%                        | 18.0%                        | 18.7%                        |
| <b>H-Gly/50</b>                                | 39.7%                        | 39.1%                        | 38.1%                        | 38.0%                        | 38.0%                        | 39.2%                        |
| <b>H-Gly/75</b>                                | 61.5%                        | 61.2%                        | 59.8%                        | 60.0%                        | 61.6%                        | 62.0%                        |
| <b>H-Gly/100</b>                               | 89.9%                        | 89.9%                        | 89.7%                        | 88.9%                        | 93.1%                        | 91.7%                        |
| <b>H-Glu/100</b>                               | 86.4%                        | 87.0%                        | 84.9%                        | 85.0%                        | 85.0%                        | 85.6%                        |
| <b>D-Glu/0</b>                                 | 8.3%                         | 14.6%                        | 23.6%                        | 22.1%                        | 16.2%                        | 17.6%                        |
| <b>D-Gly/0</b>                                 | 12.0%                        | 12.4%                        | 16.6%                        | 17.9%                        | 14.1%                        | 14.6%                        |
| <b>50D-Gly/0</b>                               | 2.8%                         | 3.6%                         | 7.5%                         | 7.9%                         | 4.7%                         | 5.2%                         |
| <b>50D-Gly/100</b>                             | 93.0%                        | 92.3%                        | 92.5%                        | 92.5%                        | 94.0%                        | 93.0%                        |
| <b>Convolution of 50D-Gly/0 with H-Gly/100</b> | 90.2%                        | 90.3%                        | 90.5%                        | 89.8%                        | 93.4%                        | 92.1%                        |

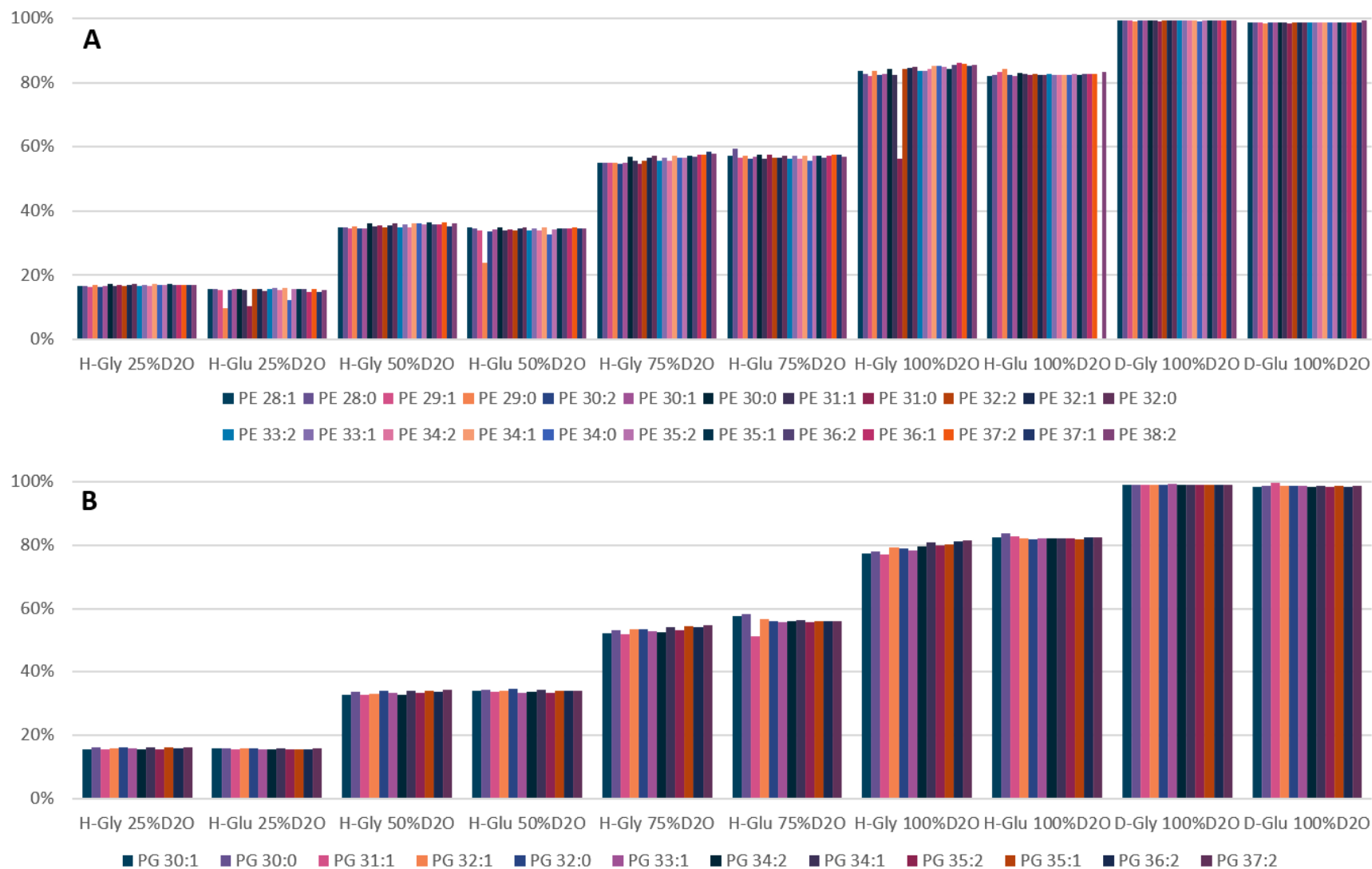


**Figure 5-5. Best fit deuterium percent for individual lipids.**

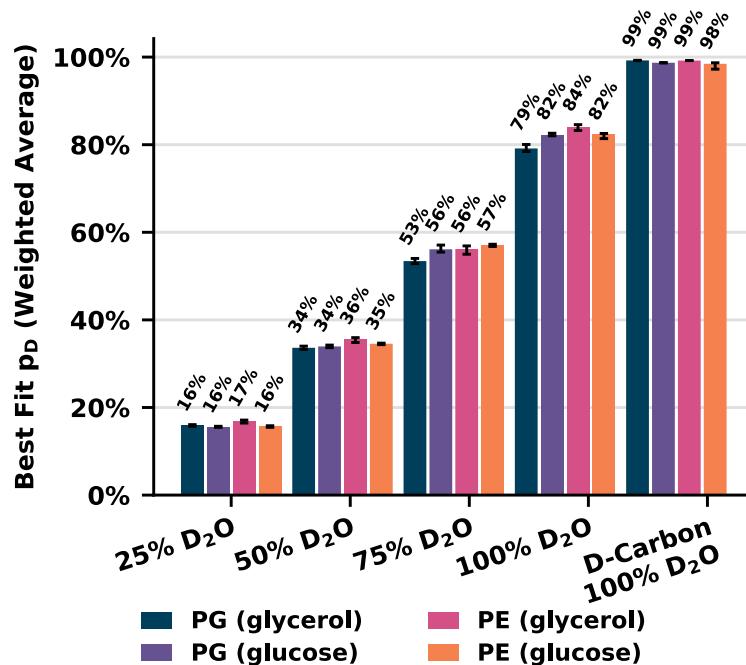


**Figure 5-6. Weighted average deuterium incorporations.**

Bar chart showing the weighted average deuterium incorporation ( $p_D$ ) for PG and PE lipids in samples with varying D<sub>2</sub>O in the media. All samples were grown with unlabeled carbon sources except for the last set of columns. Error bars are based on a bootstrap resampling of the individual lipid  $p_D$  values and weights.

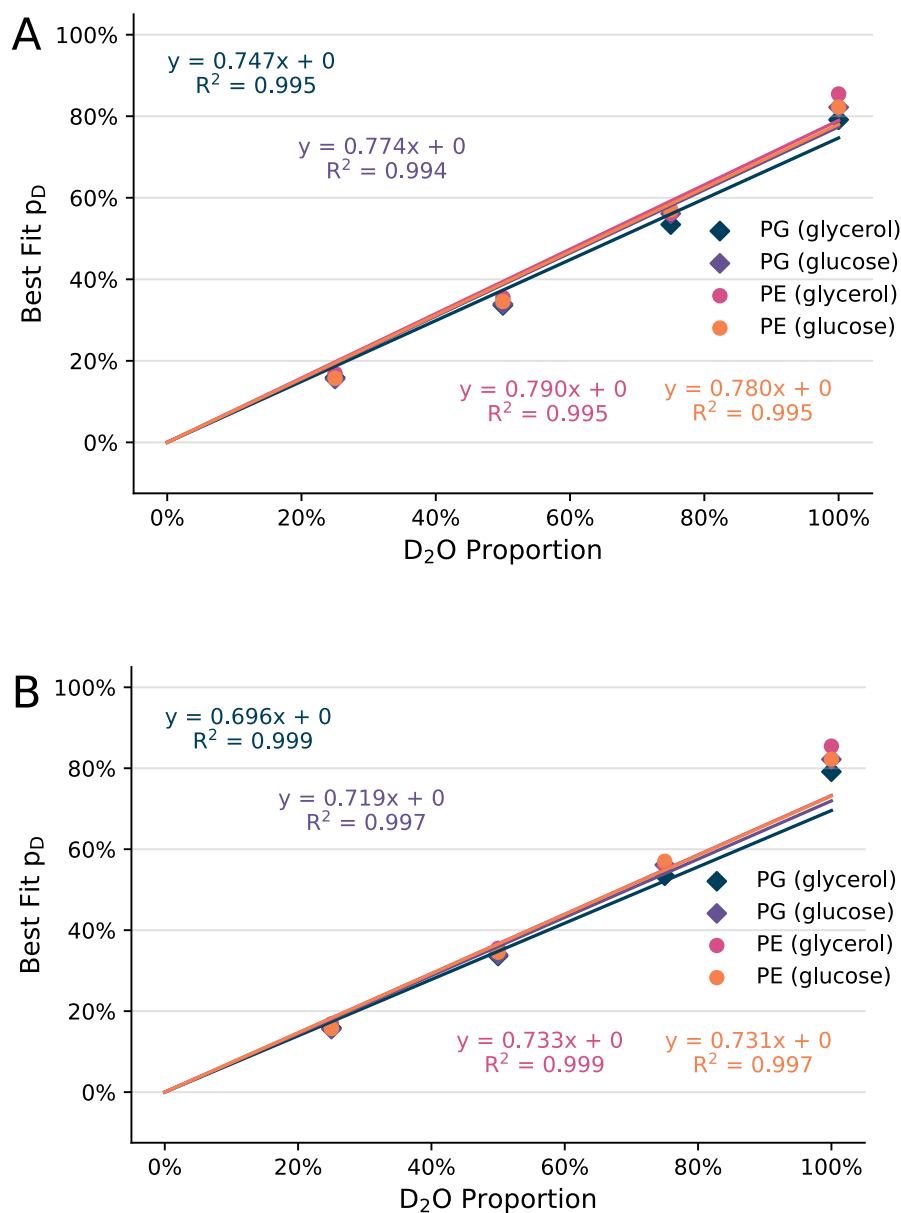


**Figure 5-7. Lipid deuteration linear regressions.**

Weighted average deuterium incorporation  $p_D$  vs D<sub>2</sub>O for the conditions with unlabeled carbon sources. Trendlines are shown as linear regressions with equations and  $R^2$  values shown on the chart. Panel A shows the linear regression including the 100% D<sub>2</sub>O condition while panel B shows the linear regression without including the 100% D<sub>2</sub>O condition.

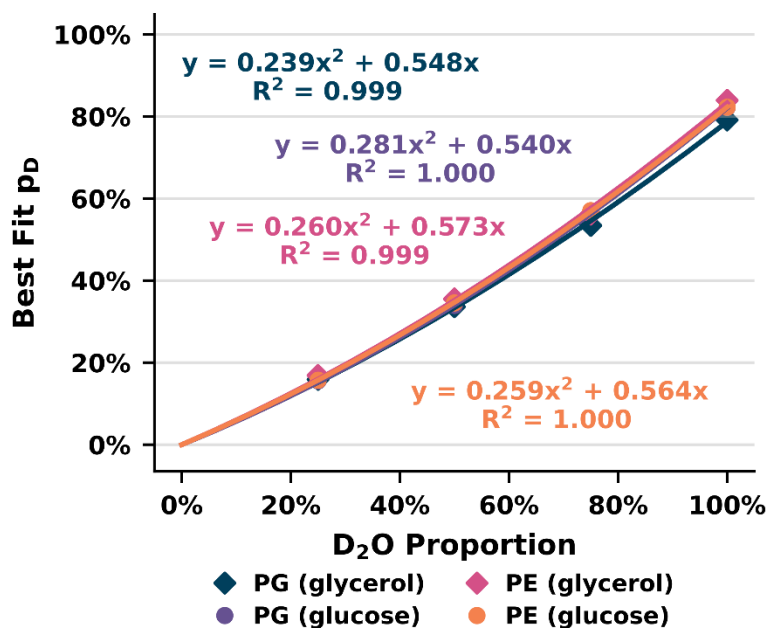


**Figure 5-8. Deuterium incorporation polynomial regressions.**

Weighted average deuterium incorporation ( $p_D$ ) is plot vs %D<sub>2</sub>O demonstrating the strong relationship between the two. Equations refer to second degree polynomials which are constrained to intercept 0.

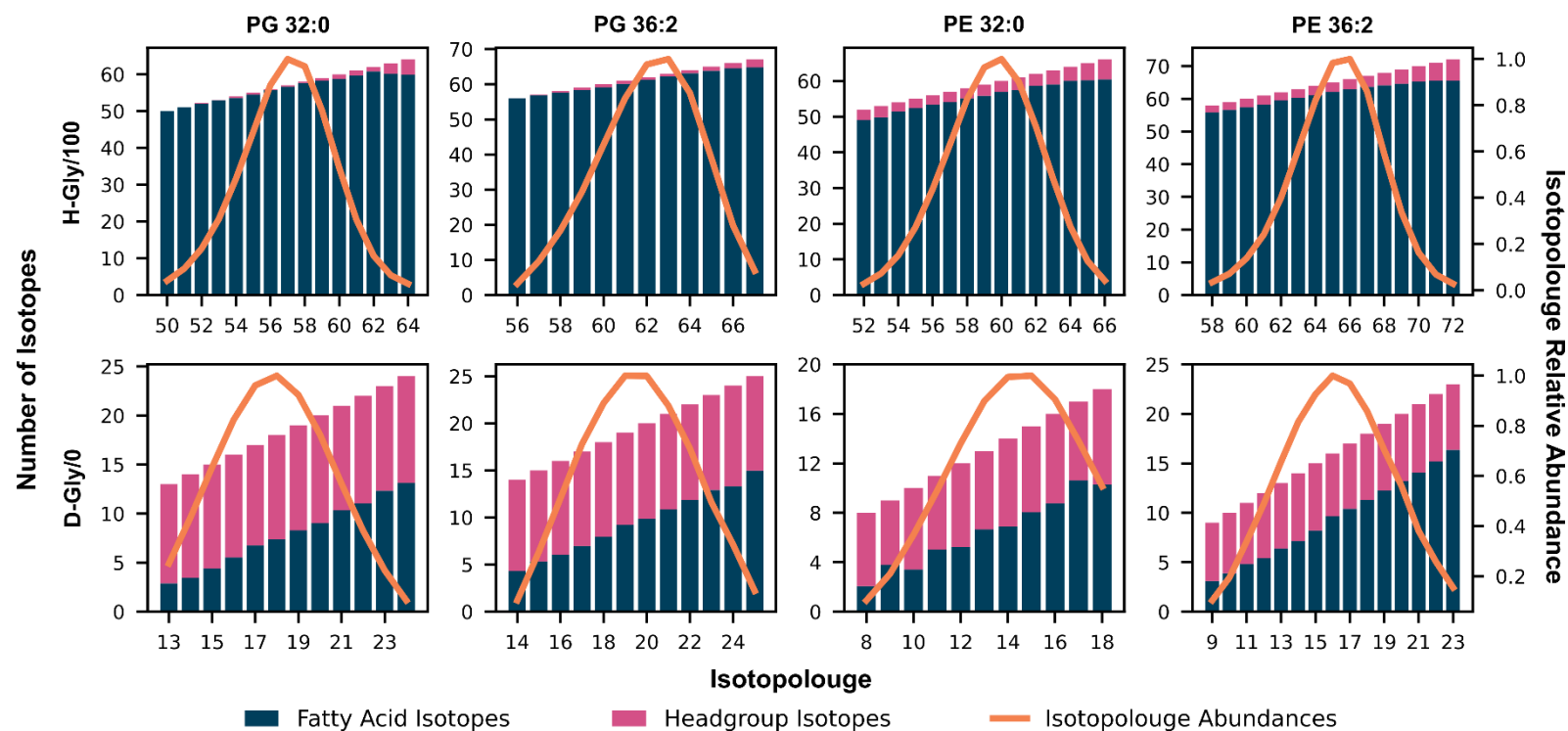


**Figure 5-9. PG Isotope Localization for Selected Lipids in H-Gly/100 and D-Gly/0.**

The x-axes on these charts indicate specific isotopologues with the localization of the isotope split between headgroup and FA tails (y-axes). The right-hand y-axis corresponds to the orange curve showing the relative abundance of the observed isotopologues.

**Table 5-8. Comparison of fatty acid specific deuterium incorporation calculation.**

| Condition             | H-Gly/25 | H-Gly/50 | H-Gly/75 | H-Gly/100 |
|-----------------------|----------|----------|----------|-----------|
| FA 16:0 (FA avg fit)  | 18.9%    | 39.4%    | 61.3%    | 89.9%     |
| FA 16:0 (Lipid fit)   | 18.5%    | 39.3%    | 61.8%    | 91.5%     |
| FA 16:0 Percent Error | 1.7%     | 0.03%    | 0.87%    | 1.8%      |
| FA 17:1 (FA avg fit)  | 18.1%    | 38.0%    | 59.9%    | 89.3%     |
| FA 17:1 (Lipid fit)   | 17.9%    | 38.0%    | 60.8%    | 92.1%     |
| FA 17:1 Percent Error | 1.1%     | 0.02%    | 1.5%     | 3.0%      |
| FA 18:1 (FA avg fit)  | 18.4%    | 38.6%    | 61.8%    | 92.4%     |
| FA 18:1 (Lipid fit)   | 18.3%    | 38.7%    | 60.8%    | 93.3%     |
| FA 18:1 Percent Error | 0.54%    | 0.31%    | 1.6%     | 0.96%     |

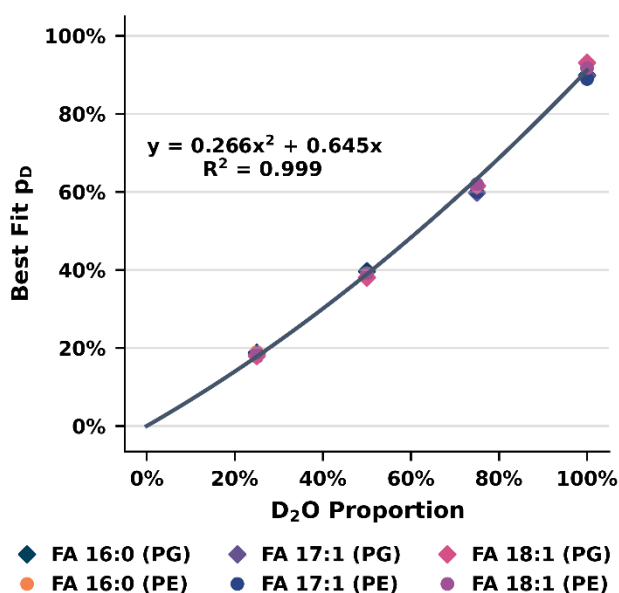


**Figure 5-10. Polynomial regression of D<sub>2</sub>O and fatty acid deuteration.**

A scatterplot showing the relation between fatty acid-specific deuteration and the D<sub>2</sub>O concentration (with H-glycerol). A polynomial regression was performed including all the datapoints and was constrained to intercept Y at 0.

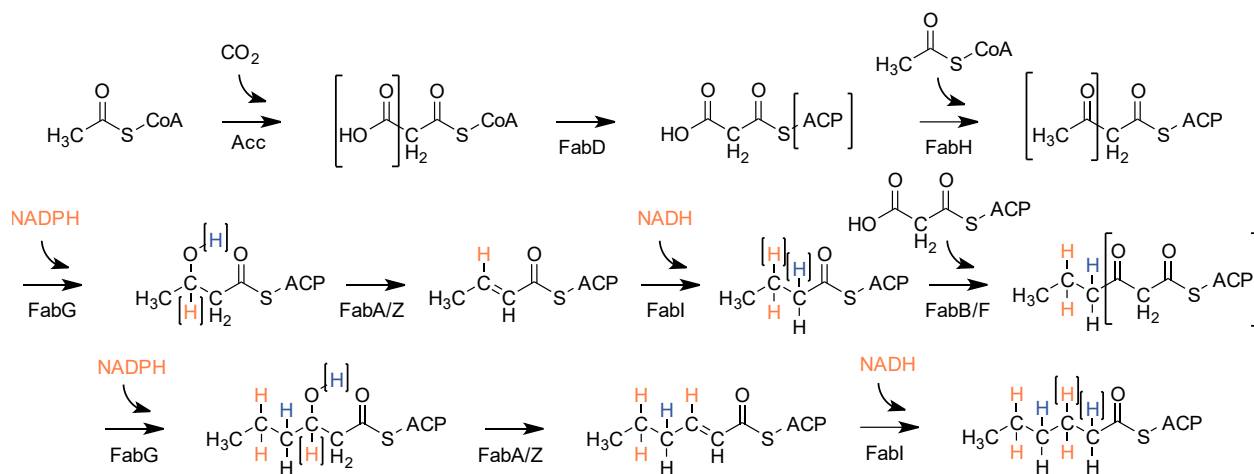


**Figure 5-11. Fatty acid II synthesis scheme.**

**Table 5-9. Lipid bilayer thickness at different D<sub>2</sub>O concentrations.**

Lipid bilayer thickness (in Å) in different solvents from SANS data fitting. The results from 80% and 100% D<sub>2</sub>O are less reliable due to much lower contrast.

|                                           | 0% D <sub>2</sub> O | 20% D <sub>2</sub> O | 40% D <sub>2</sub> O | 60% D <sub>2</sub> O | 80% D <sub>2</sub> O | 100% D <sub>2</sub> O |
|-------------------------------------------|---------------------|----------------------|----------------------|----------------------|----------------------|-----------------------|
| <b>H-Glucose<br/>100% D<sub>2</sub>O</b>  | 39.92±0.14          | 40.80±0.17           | 39.89±0.25           | 40.79±0.38           | 45.25±1.21           | 35.43±1.64            |
| <b>H-Glycerol<br/>100% D<sub>2</sub>O</b> | 38.36±0.10          | 38.51±0.13           | 37.22±0.19           | 37.13±0.28           | 34.01±0.85           | 49.44±2.96            |
| <b>D-Glycerol<br/>100% D<sub>2</sub>O</b> | 38.43±0.05          | 38.70±0.08           | 38.66±0.11           | NA                   | 40.44±0.21           | NA                    |

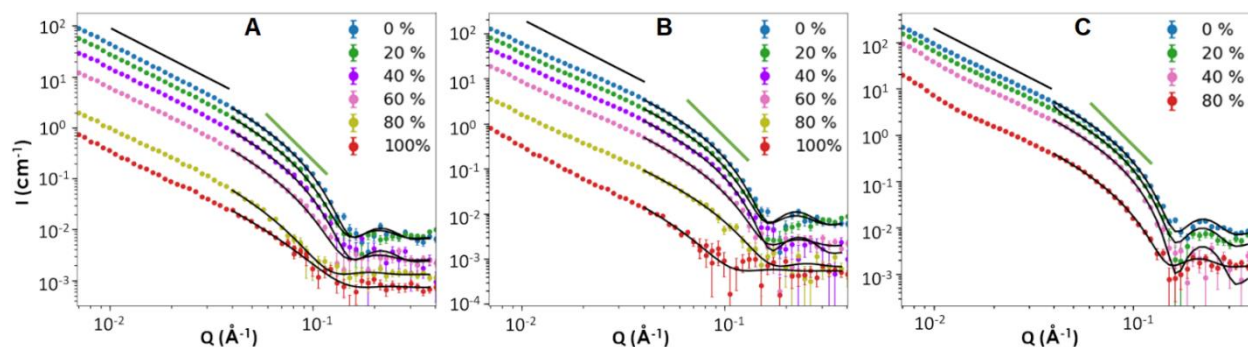


**Figure 5-12. SANS contrast series of *E. coli* at three different conditions.**

A. H-glucose 100% D<sub>2</sub>O (from top to bottom are SANS data measured in 0, 20, 40, 60, 80, 100% D<sub>2</sub>O solvent, respectively.) B. H-glycerol 100% D<sub>2</sub>O (from top to bottom are SANS data measured in 0, 20, 40, 60, 80, 100% D<sub>2</sub>O solvent, respectively.) C. D-glycerol 100% D<sub>2</sub>O (from top to bottom are SANS data measured in 0, 20, 40, 80% D<sub>2</sub>O solvent, respectively). The black line represents  $P = -2$ , and the green line represents  $P = -4$ . The lamellar model fittings are overlaid on top of the data as continuous curves.

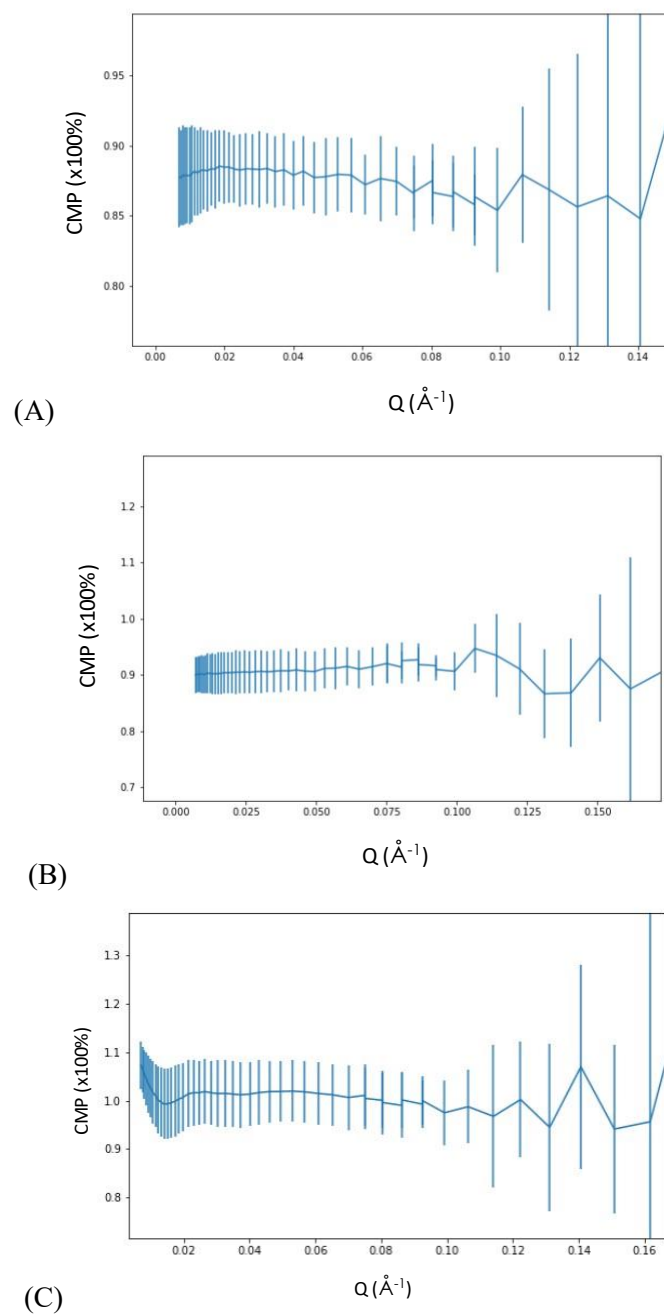


**Figure 5-13. Contrast matching point calculated over the Q range.**  
A. H-Glucose 100% D<sub>2</sub>O. B. H-Glycerol 100% D<sub>2</sub>O. C. D-Glycerol 100% D<sub>2</sub>O

**Table 5-10. Relative abundance of phospholipids used for CMP calculation.**  
Shown as intensity fraction of lipid class.

| <b>Lipid</b>   | <b>D-glycerol/D<sub>2</sub>O</b> | <b>H-glycerol/D<sub>2</sub>O</b> | <b>H-glucose/D<sub>2</sub>O</b> |
|----------------|----------------------------------|----------------------------------|---------------------------------|
| <b>PE 28:0</b> | 3.19E-02                         | 2.94E-02                         | 7.22E-03                        |
| <b>PE 28:1</b> | 8.35E-03                         | 7.06E-03                         | 2.41E-03                        |
| <b>PE 29:0</b> | 3.45E-03                         | 5.57E-03                         | 1.06E-03                        |
| <b>PE 29:1</b> | 4.78E-03                         | 2.04E-03                         | 7.33E-04                        |
| <b>PE 30:0</b> | 7.42E-02                         | 8.58E-02                         | 6.35E-02                        |
| <b>PE 30:1</b> | 4.51E-02                         | 6.47E-02                         | 3.61E-02                        |
| <b>PE 30:2</b> | 8.93E-04                         | 2.27E-03                         | 3.33E-04                        |
| <b>PE 31:0</b> | 1.51E-03                         | 8.84E-03                         | 2.34E-03                        |
| <b>PE 31:1</b> | 8.60E-02                         | 6.46E-02                         | 3.40E-02                        |
| <b>PE 32:0</b> | 1.77E-02                         | 4.43E-02                         | 1.84E-02                        |
| <b>PE 32:1</b> | 7.26E-02                         | 1.45E-01                         | 1.78E-01                        |
| <b>PE 32:2</b> | 8.79E-03                         | 1.33E-02                         | 5.56E-03                        |
| <b>PE 33:1</b> | 2.56E-01                         | 2.12E-01                         | 2.82E-01                        |
| <b>PE 33:2</b> | 1.15E-02                         | 2.60E-02                         | 1.77E-02                        |
| <b>PE 34:0</b> | 3.50E-03                         | 8.08E-03                         | 6.66E-03                        |
| <b>PE 34:1</b> | 4.73E-02                         | 1.01E-01                         | 1.44E-01                        |
| <b>PE 34:2</b> | 7.98E-02                         | 5.32E-02                         | 4.72E-02                        |
| <b>PE 35:1</b> | 1.10E-01                         | 6.25E-02                         | 7.79E-02                        |
| <b>PE 35:2</b> | 2.02E-02                         | 1.38E-02                         | 1.74E-02                        |
| <b>PE 36:1</b> | 1.04E-02                         | 1.09E-02                         | 2.09E-02                        |
| <b>PE 36:2</b> | 5.95E-02                         | 2.52E-02                         | 2.88E-02                        |
| <b>PE 37:1</b> | 1.45E-02                         | 3.08E-03                         | 3.46E-06                        |
| <b>PE 37:2</b> | 1.21E-02                         | 8.64E-03                         | 7.78E-03                        |
| <b>PE 38:2</b> | 2.00E-02                         | 3.39E-03                         | 1.30E-03                        |
| <b>PG 30:1</b> | 2.84E-02                         | 4.88E-02                         | 6.06E-03                        |
| <b>PG 30:0</b> | 9.97E-02                         | 1.02E-01                         | 4.93E-02                        |
| <b>PG 31:1</b> | 6.07E-02                         | 5.09E-02                         | 7.94E-03                        |
| <b>PG 32:1</b> | 6.33E-02                         | 1.39E-01                         | 2.19E-01                        |
| <b>PG 32:0</b> | 7.40E-02                         | 7.42E-02                         | 6.22E-02                        |
| <b>PG 33:1</b> | 3.47E-01                         | 2.57E-01                         | 2.68E-01                        |
| <b>PG 34:2</b> | 4.51E-02                         | 4.89E-02                         | 2.77E-02                        |
| <b>PG 34:1</b> | 6.19E-02                         | 9.62E-02                         | 2.28E-01                        |
| <b>PG 35:2</b> | 1.92E-02                         | 1.61E-02                         | 1.76E-02                        |
| <b>PG 35:1</b> | 1.38E-01                         | 1.23E-01                         | 7.56E-02                        |
| <b>PG 36:2</b> | 5.00E-02                         | 3.26E-02                         | 3.35E-02                        |
| <b>PG 37:2</b> | 1.33E-02                         | 1.10E-02                         | 5.32E-03                        |

**Table 5-11. Contrast match point from experimental SANS data and MS data.**

|                                           | CMP averaged<br>from $<0.08 \text{ \AA}^{-1}$ | CMP determined<br>from $Q_{\min}$ | Chain CMP calculated from MS data                                    |
|-------------------------------------------|-----------------------------------------------|-----------------------------------|----------------------------------------------------------------------|
| <b>H-Glucose<br/>100% D<sub>2</sub>O</b>  | $88.0\% \pm 2.8\%$                            | $87.8\% \pm 3.6\%$                | PE 87.7% PG 88.1% (molar ratio 8.70/1)<br>Weighted average CMP=87.7% |
| <b>H-Glycerol<br/>100% D<sub>2</sub>O</b> | $90.6\% \pm 3.6\%$                            | $89.9\% \pm 3.3\%$                | PE 89.1% PG 85.0% (molar ratio 9.42/1)<br>Weighted average CMP=88.7% |
| <b>D-Glycerol<br/>100% D<sub>2</sub>O</b> | $101.7\% \pm 6.5\%$                           | $107.3\% \pm 4.9\%$               | PE 105.4% PG 105.6%<br>Regardless PE/PG ratio CMP ~105.4%            |

**Table 5-12. Calculated CMP at different solvents based on headgroup deuteration level.**

Deuteration derived from MS data. Due to exchangeable H among headgroups, the CMP varies at different D<sub>2</sub>O solvents.

|                                           |    | 0%<br>D <sub>2</sub> O | 20%<br>D <sub>2</sub> O | 40%<br>D <sub>2</sub> O | 60%<br>D <sub>2</sub> O | 80%<br>D <sub>2</sub> O | 100%<br>D <sub>2</sub> O |
|-------------------------------------------|----|------------------------|-------------------------|-------------------------|-------------------------|-------------------------|--------------------------|
| <b>H-Glucose 100%<br/>D<sub>2</sub>O</b>  | PE | 69.7%                  | 71.9%                   | 74.9%                   | 77.8%                   | 81.6%                   | 85.7%                    |
|                                           | PG | 71.0%                  | 73.2%                   | 75.9%                   | 78.6%                   | 82.1%                   | 86.0%                    |
| <b>H-Glycerol 100%<br/>D<sub>2</sub>O</b> | PE | 60.1%                  | 62.4%                   | 65.1%                   | 68.3%                   | 72.0%                   | 76.2%                    |
|                                           | PG | 43.3%                  | 45.5%                   | 48.0%                   | 51.0%                   | 54.4%                   | 58.2%                    |
| <b>D-Glycerol 100%<br/>D<sub>2</sub>O</b> | PE | 88.5%                  | 90.6%                   | 93.5%                   | NA                      | 100.0%                  | NA                       |
|                                           | PG | 90.6%                  | 92.6%                   | 95.3%                   | NA                      | 101.4%                  | NA                       |

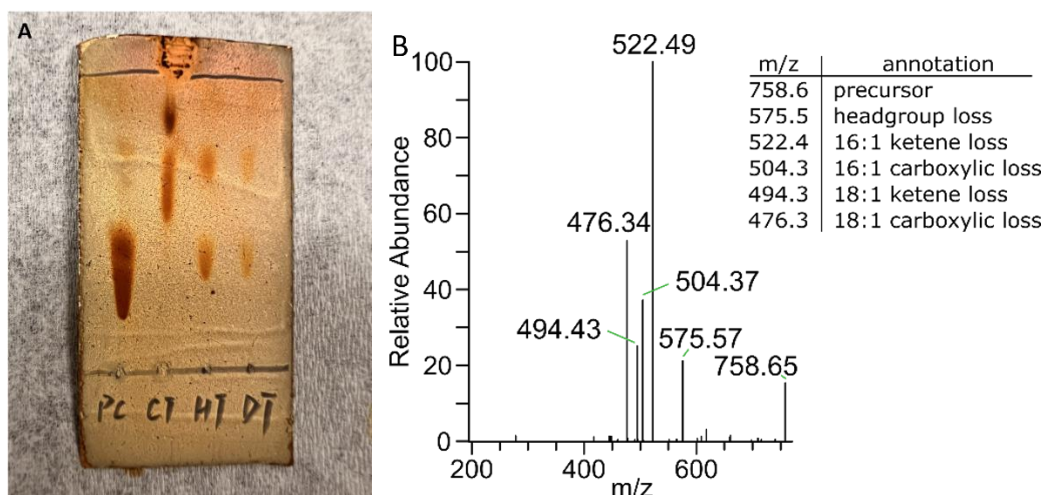


**Figure 5-14. Demonstration of PC production from engineered *E. coli*.**

A. TLC analysis of PC (first spot from left), commercial *E. coli* total lipid extract (CT, second spot from left), total lipid extract from H<sub>2</sub>O culture (HT, third spot from left), total lipid extract from D<sub>2</sub>O-based culture (DT, fourth spot from left). The total lipid extract was produced and purified following the general method described in Materials and Methods. B. Example mass spectrum of identified PC lipids. This lipid is PC 18:1\_16:1 identified in the protiated total lipid extract.

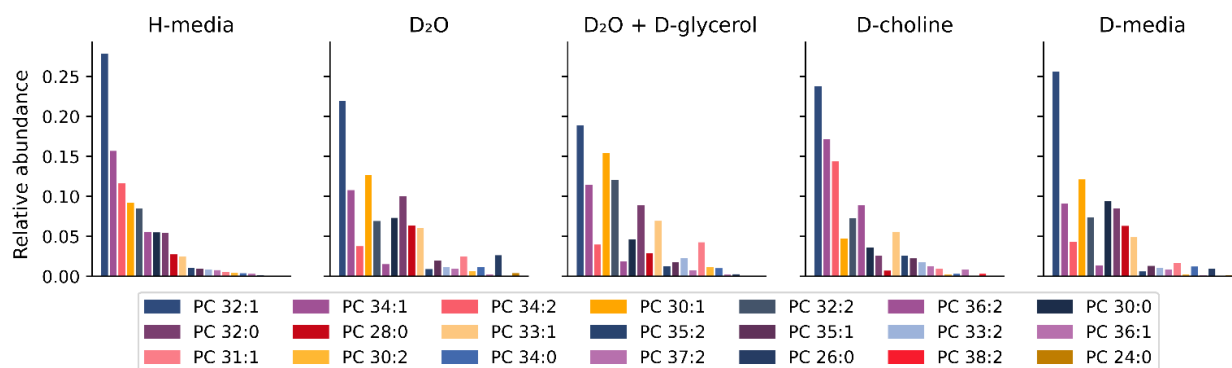


**Figure 5-15. Relative abundance of PC lipids identified in all culture conditions.**

Values correspond to normalized peak areas associated with each lipid species. Each value was divided by the sum of the individual intensities of each shared lipid.

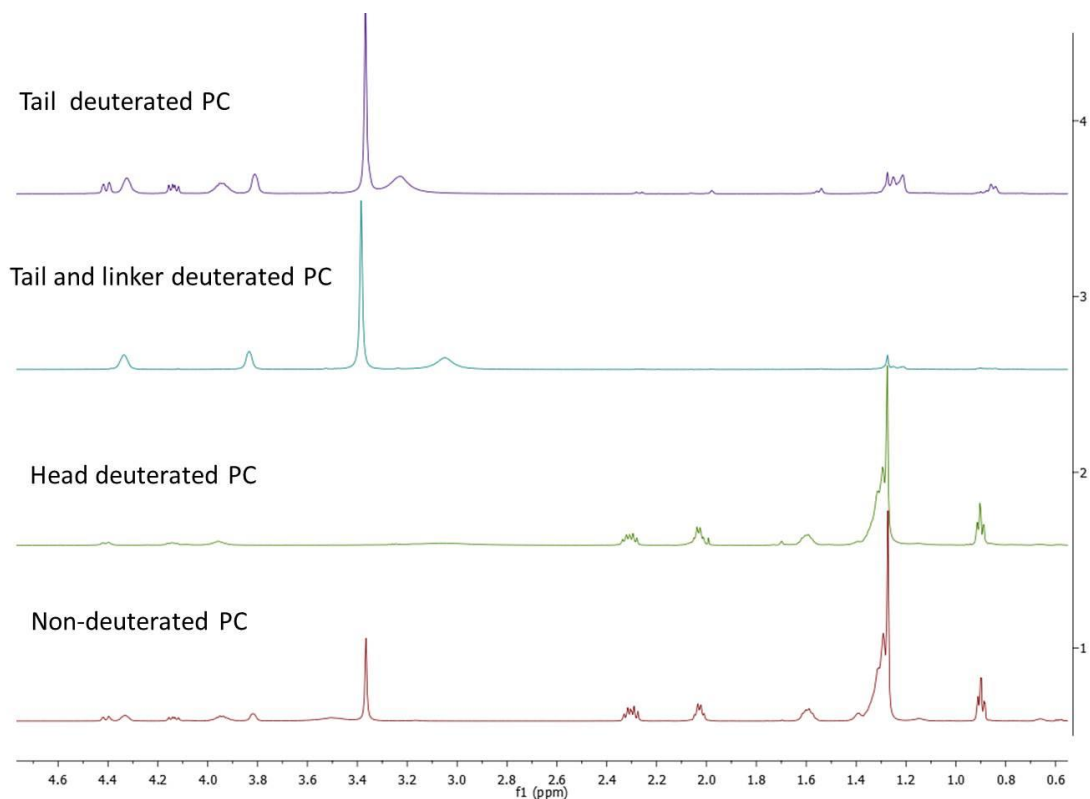


**Figure 5-16. Proton NMR spectra of isolated PC with different deuteration levels.**

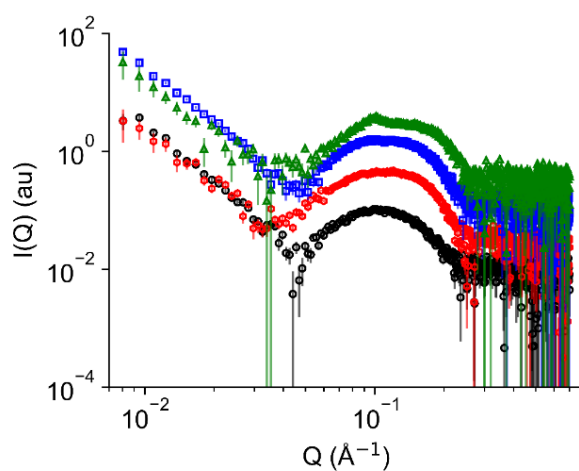


**Figure 5-17. SAXS data for deuterium labeled PC lipids.**

(black) protiated lipid; (red) Tail-deuterated PC, where only the fatty acyl chains are deuterated; (blue) Tail and linker-deuterated PC, with both the fatty acyl chains and glycerol backbone deuterated; and (green) Perdeuterated PC.

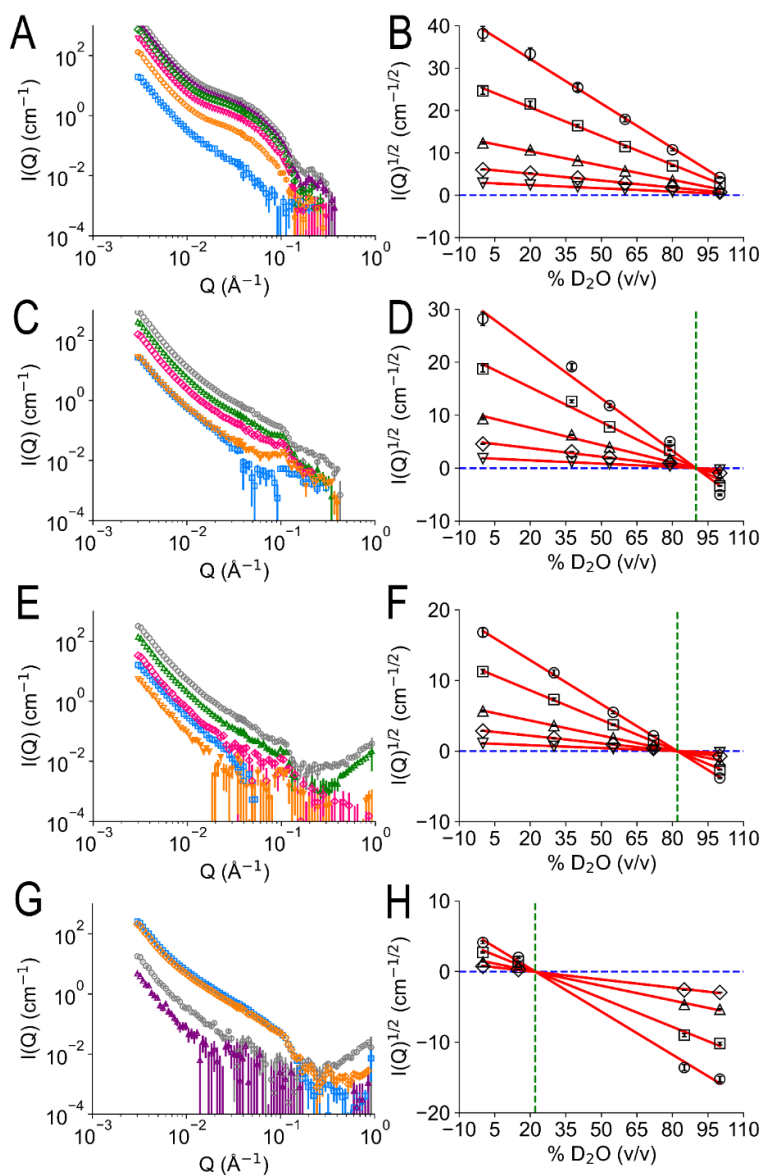


**Figure 5-18. SANS data and contrast match point analysis for deuterium labeled PC lipids.** SANS profiles and contrast match point determination for perdeuterated PC lipids (PC-DDD) (A, B), tail and linker deuterated PC lipids (PC-DDH) (C,D), tail deuterated PC lipids (PC-DHH) (E,F), and fully protiated PC lipids (PC-HHH) (G,H), respectively. SANS profiles were measured in 100% D<sub>2</sub>O (blue), 80 or 85% D<sub>2</sub>O (orange), 55 or 50 % D<sub>2</sub>O (pink), 40 or 30 % D<sub>2</sub>O (green), 20 or 15 % D<sub>2</sub>O purple, and 0% D<sub>2</sub>O (gray). The black circles, squares, triangles-up, diamonds, and triangles-down corresponds to square-root of the scattering intensity at Q points 0.0032 Å<sup>-1</sup>, 0.0039 Å<sup>-1</sup>, 0.0056 Å<sup>-1</sup>, 0.0085 Å<sup>-1</sup>, and 0.0172 Å<sup>-1</sup>, respectively. Intersection of dash blue and red lines indicates the neutron contrast match point for the non- and isotopically labeled phospholipids forms.

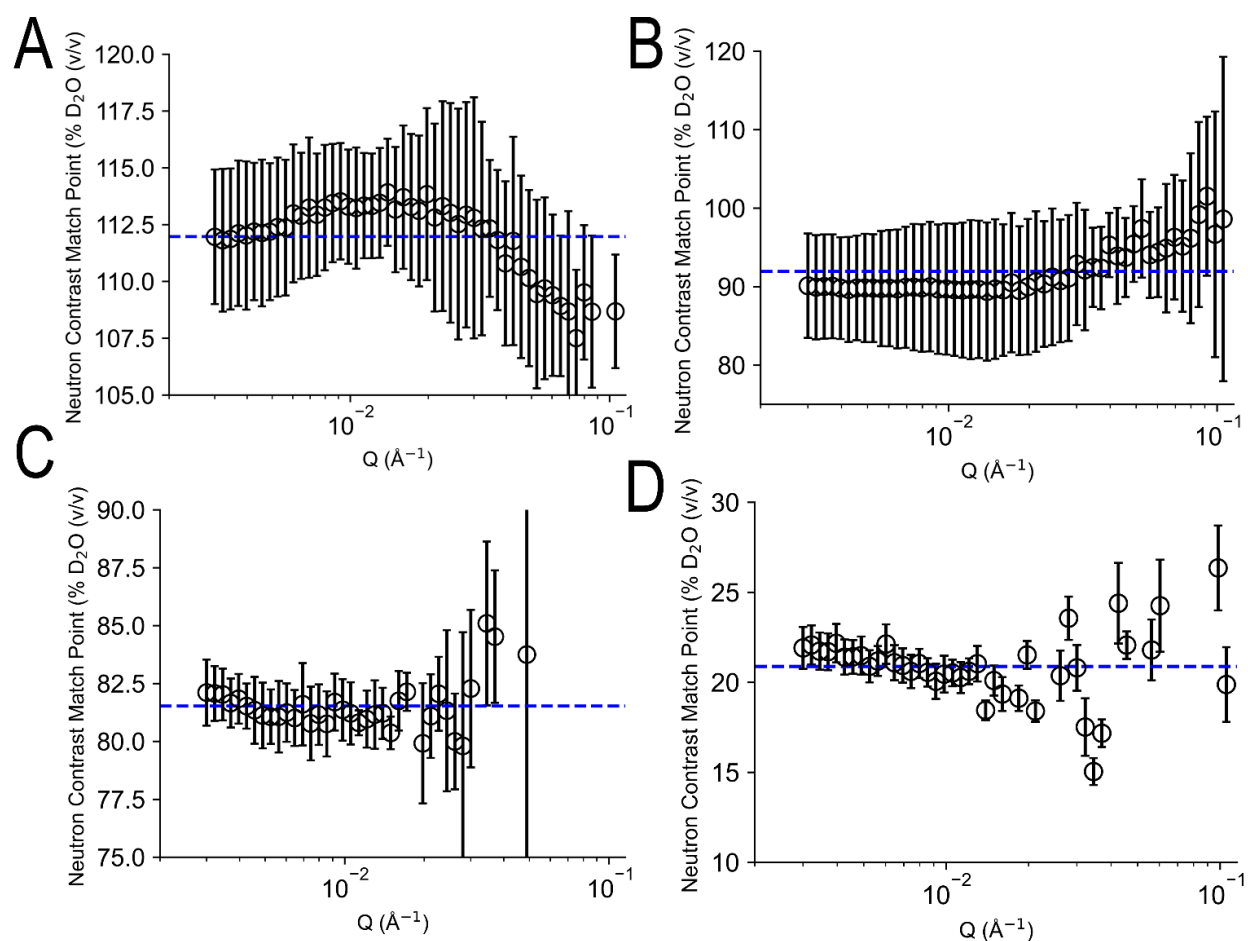


**Figure 5-19.  $Q$ -dependency of neutron contrast match point for non- and isotopically labeled phospholipids forms.**

A. perdeuterated PC lipids. B. tail and linker deuterated PC lipids. C. tail deuterated PC lipids. D. fully protiated PC lipids, respectively. The average value was estimated in  $Q$ -range from 0.003 Å<sup>-1</sup> to 0.1 Å<sup>-1</sup> and are shown as a dash blue line, and the  $Q$ -dependency of neutron contrast match point for each lipid form are shown as black open circles.

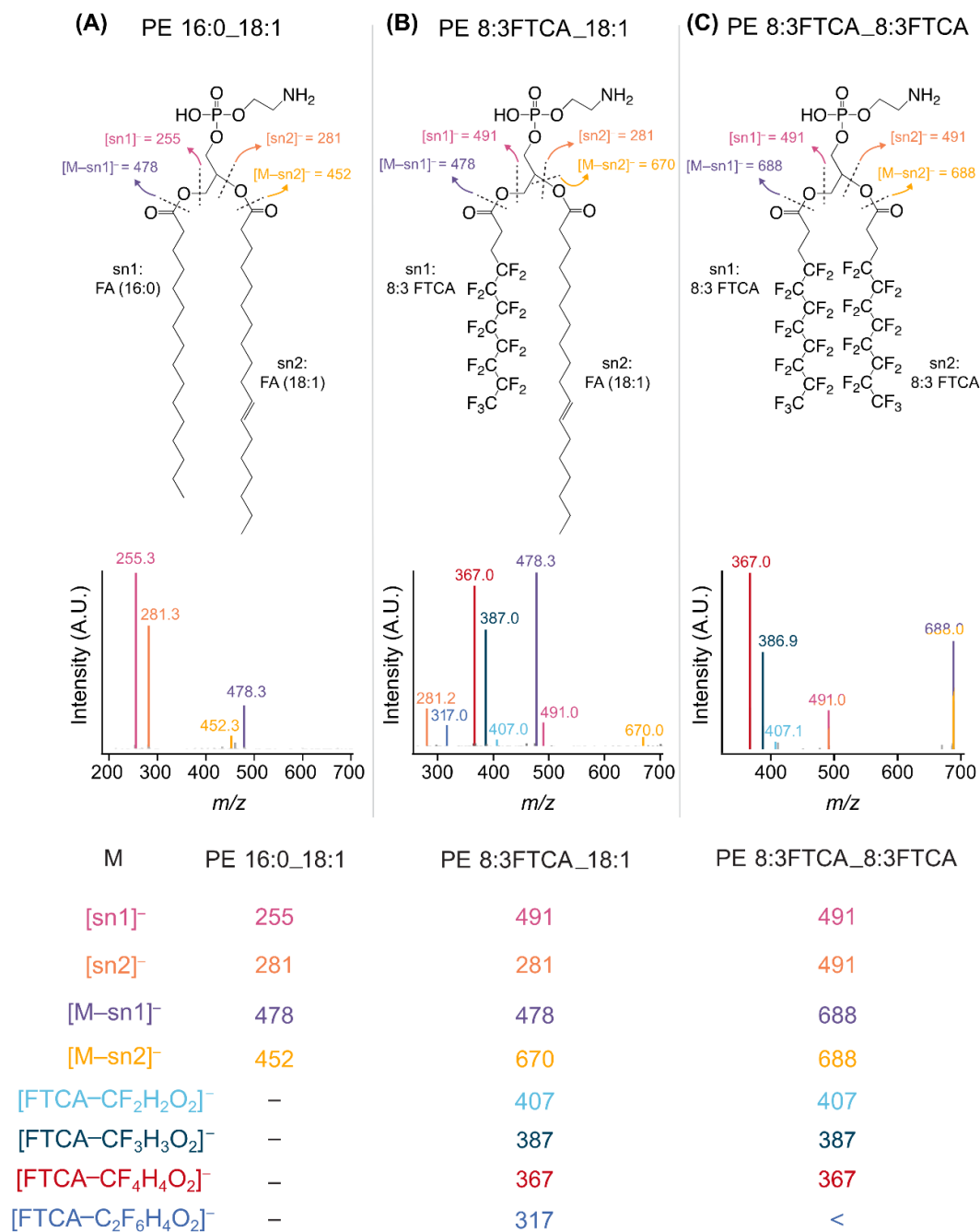


**Figure 6-1. Structure and fragmentation of representative PE lipids containing up to two 8:2 FTCA as FA tails**

The molecular structure, associated ion fragments, and the tandem mass spectra of the target molecules, including PE 34:1 (A), PE 29:1-8PFC (B), and PE 22:0-16PFC (C) are shown. The fatty acid tail positions (i.e., sn1 or sn2) are arbitrarily designated. The mass spectra are normalized to the intensity of the ion fragment with the highest intensity in the selected  $m/z$  range.

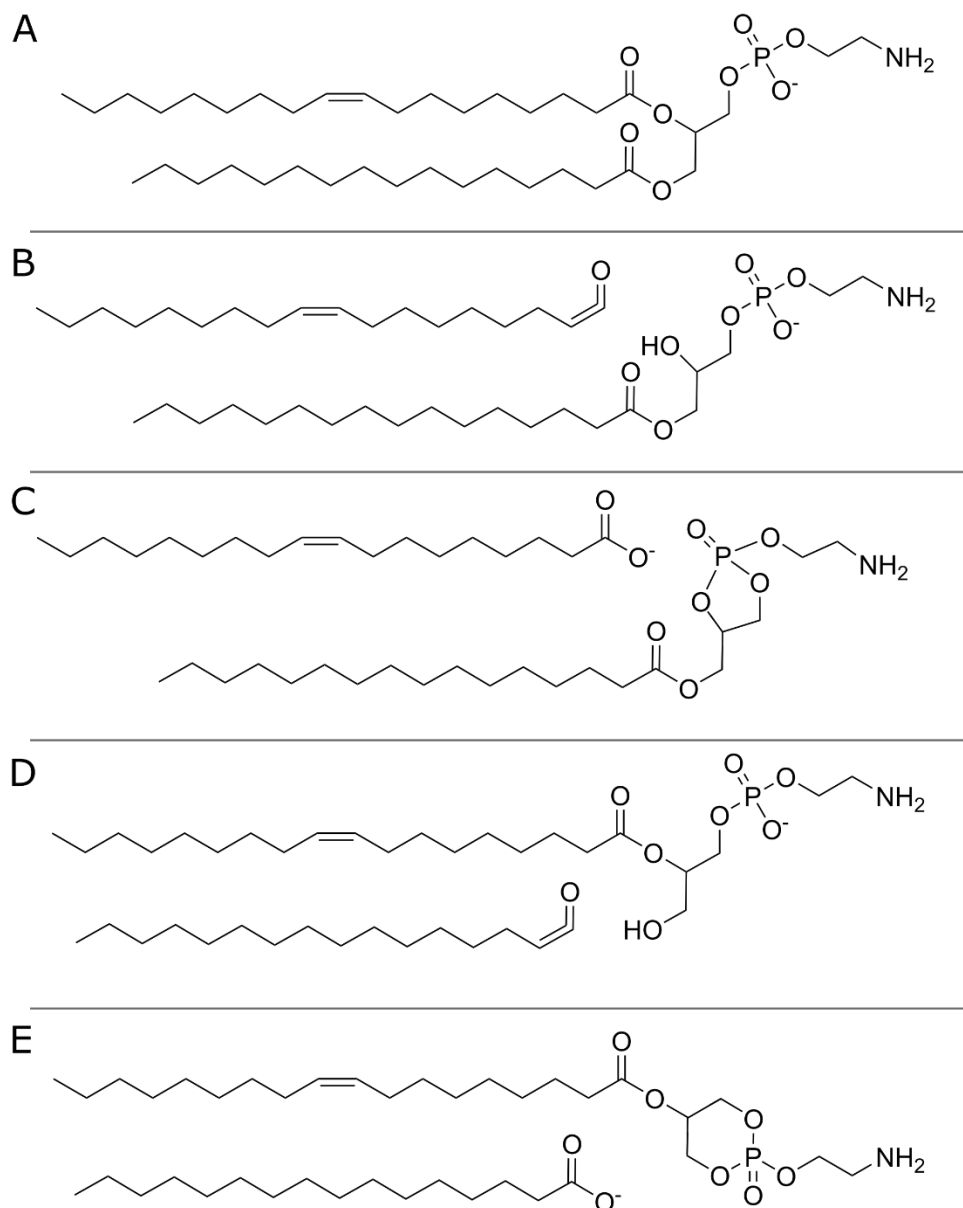


**Figure 6-2. Expected ion structures for the tandem MS of PE 34:1**

A. Shows the precursor ion. B. Shows the 18:1 ketene neutral loss. C. Shows the 18:1 fatty acid ion. D. Shows the 16:0 ketene neutral loss. E. Shows the 16:0 fatty acid ion.

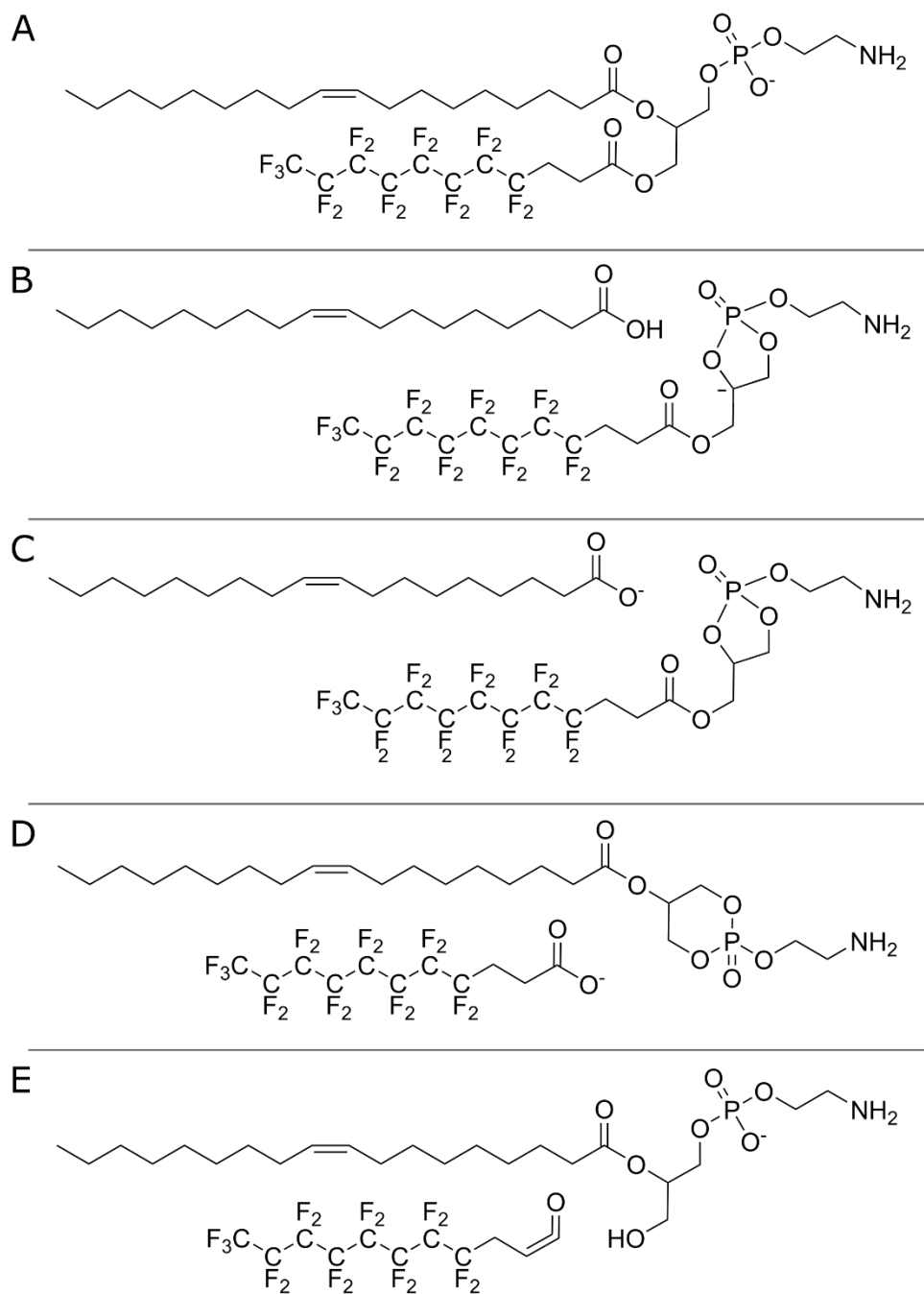


**Figure 6-3. Expected ion structures for the tandem MS of PE 18:1\_8:3 FTCA**

A. Shows the precursor ion. B. Shows the 18:1 carboxylic acid neutral loss. C. Shows the 18:1 fatty acid ion. D. Shows the 8:3 FTCA fatty acid ion. E. Shows the 8:3 FTCA ketene neutral loss.

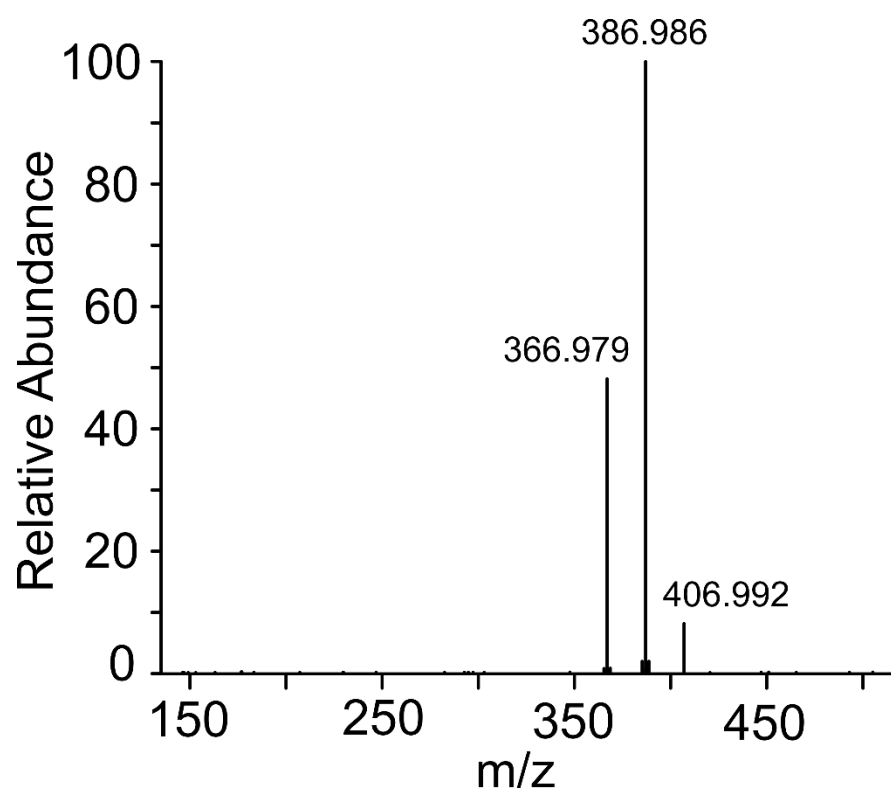


**Figure 6-4. HRAM tandem MS spectrum of free 8:3 FTCA**  
High resolution accurate mass (orbitrap) tandem MS of free 8:2 FTCA.

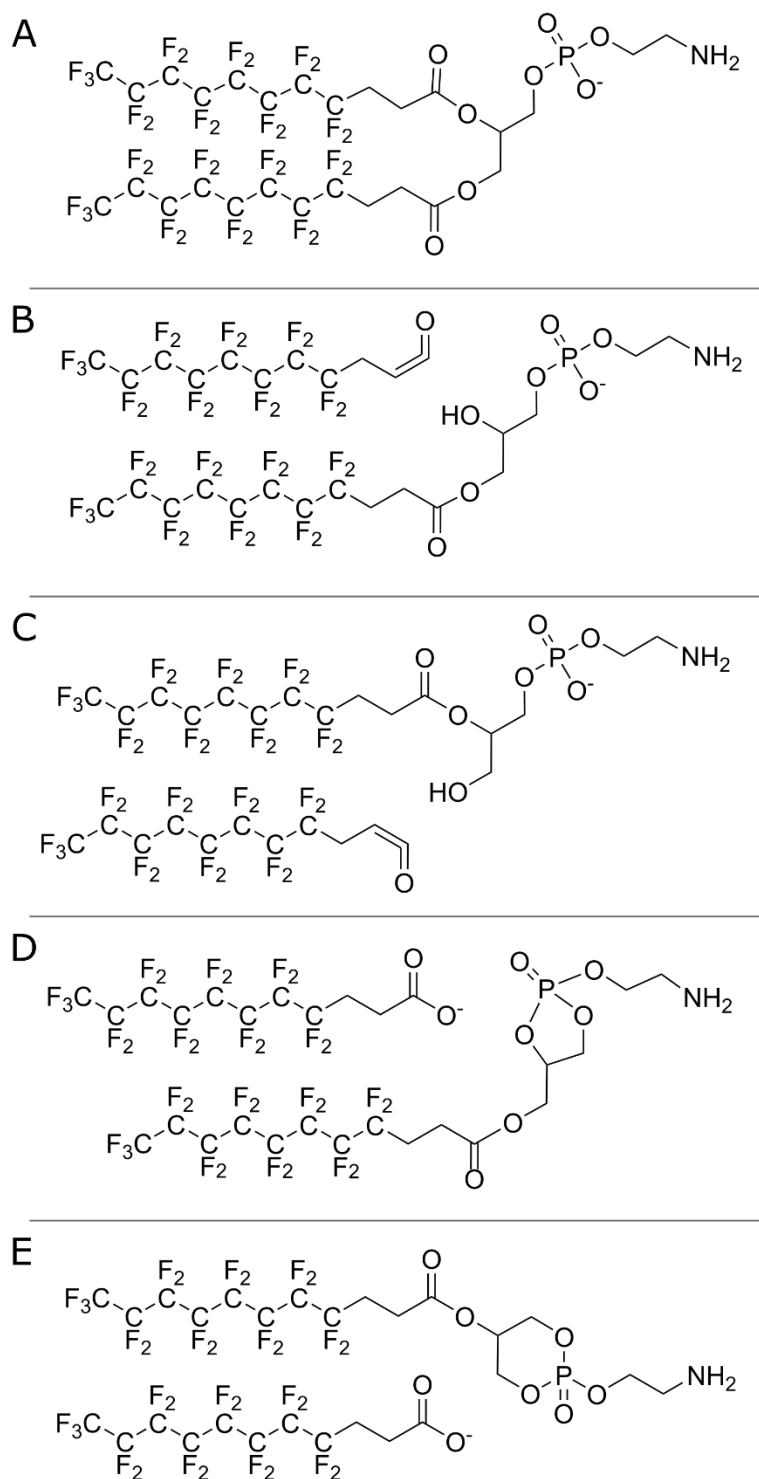


**Figure 6-5. Expected ion structures for the tandem MS of PE 18:1\_8:3 FTCA**

A. Shows the precursor ion. B and C. Show the 8:3 FTCA ketene neutral losses. D and E. Show the 8:3 FTCA fatty acid ions.

**Table 6-1. Orbitrap tandem MS of PE 8:3 FTCA\_18:1**Theoretical  $m/z$  corresponds to the expected  $m/z$  given the proposed ion formula.

| Expected Fragment                           | Fragment Ion Formula | Theoretical $m/z$ | Observed $m/z$ | $m/z$ error (ppm) |
|---------------------------------------------|----------------------|-------------------|----------------|-------------------|
| <b>FA 18:1</b>                              | C18H33O2-            | 281.2486          | 281.2483       | 1.01              |
| <b>[FTCA-C2H4F6O2]–</b>                     | C9F11-               | 316.983           | 316.9827       | 0.959             |
| <b>[FTCA-CH4F4O2]–</b>                      | C10F13-              | 366.9798          | 366.9794       | 0.834             |
| <b>[FTCA-CH3F3O2]–</b>                      | C10F14H-             | 386.986           | 386.9857       | 0.809             |
| <b>[FTCA-CH2F2O2]–</b>                      | C10F15H2-            | 406.9922          | 406.9918       | 1.03              |
| <b>FTCA Carboxylic Acid Neutral Loss</b>    | C23H43O6NP-          | 460.2834          | 460.283        | 0.895             |
| <b>FTCA Ketene Neutral Loss</b>             | C23H45O7NP-          | 478.2939          | 478.2938       | 0.276             |
| <b>FTCA-</b>                                | C11F17O2H4-          | 490.9945          | 490.9941       | 0.967             |
| <b>FA 18:1 Carboxylic Acid Neutral Loss</b> | C16H14F17O6N1P-      | 670.0293          | 670.0286       | 1.03              |

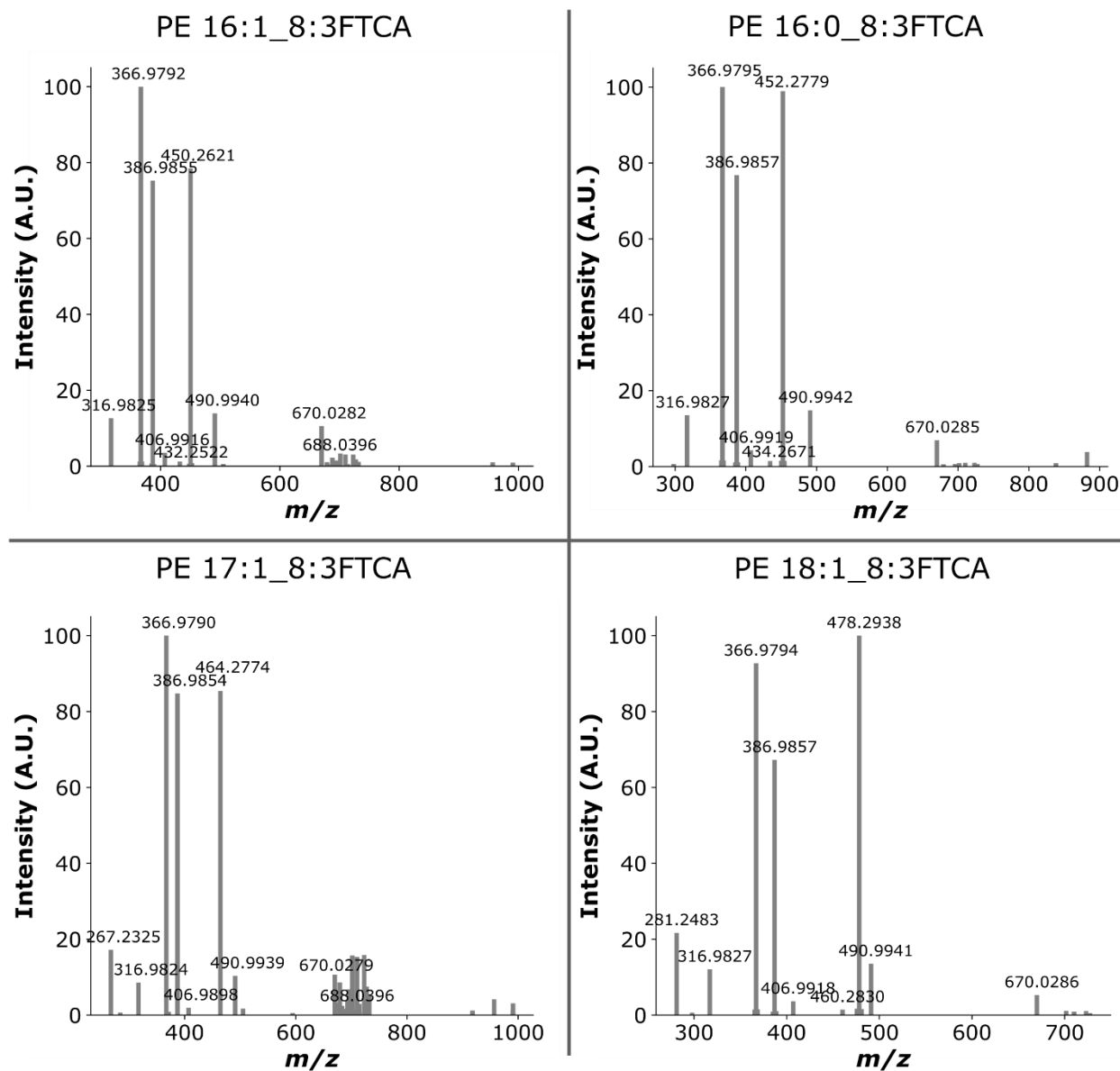


**Figure 6-6. HRAM tandem MS of FTCA-containing PE Lipids**  
 High resolution accurate mass (orbitrap) tandem MS of PFASylated PE lipids.

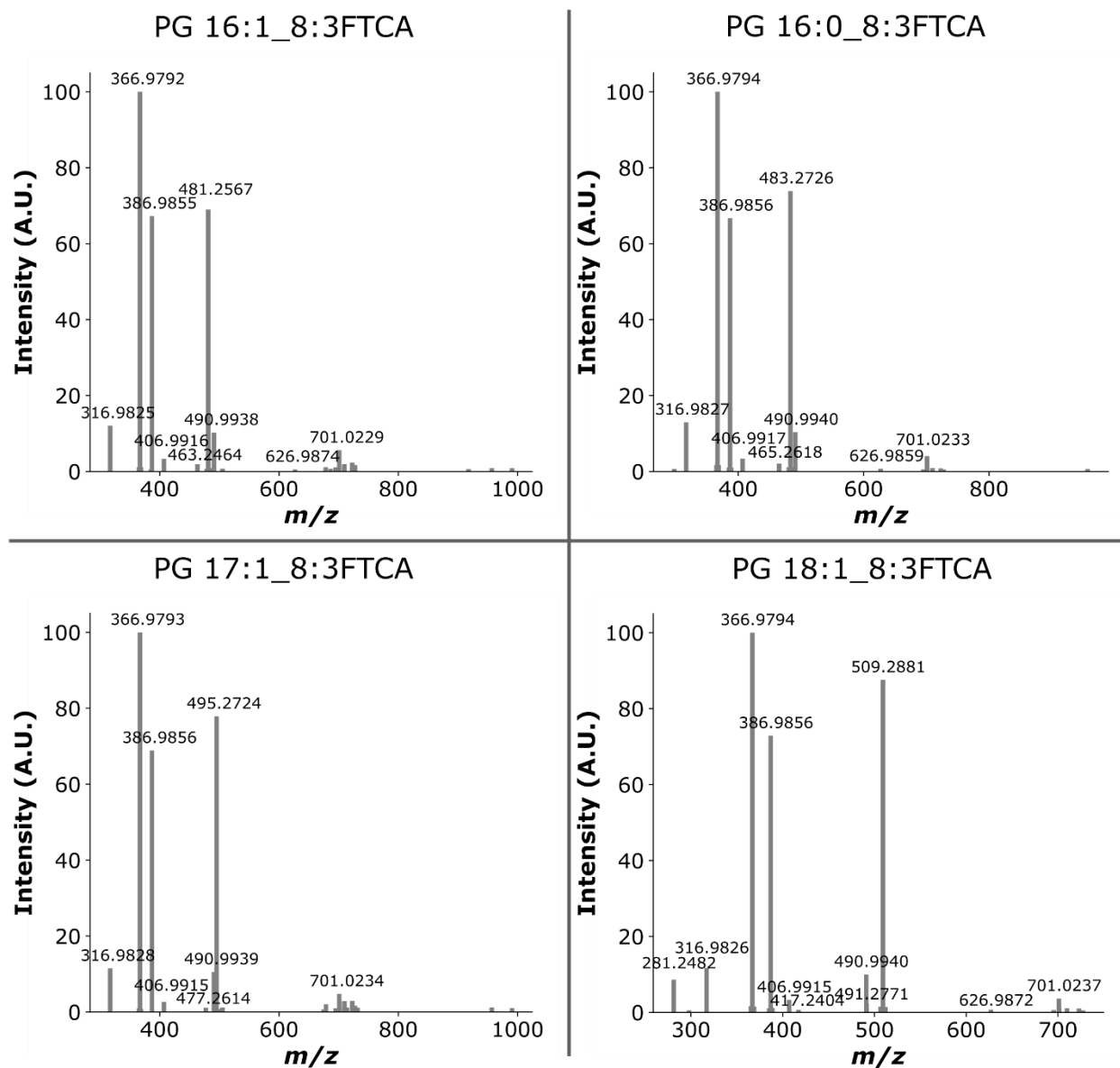


**Figure 6-7. HRAM tandem MS of FTCA-containing PG Lipids**  
High resolution accurate mass (orbitrap) tandem MS of PFASylated PE lipids.

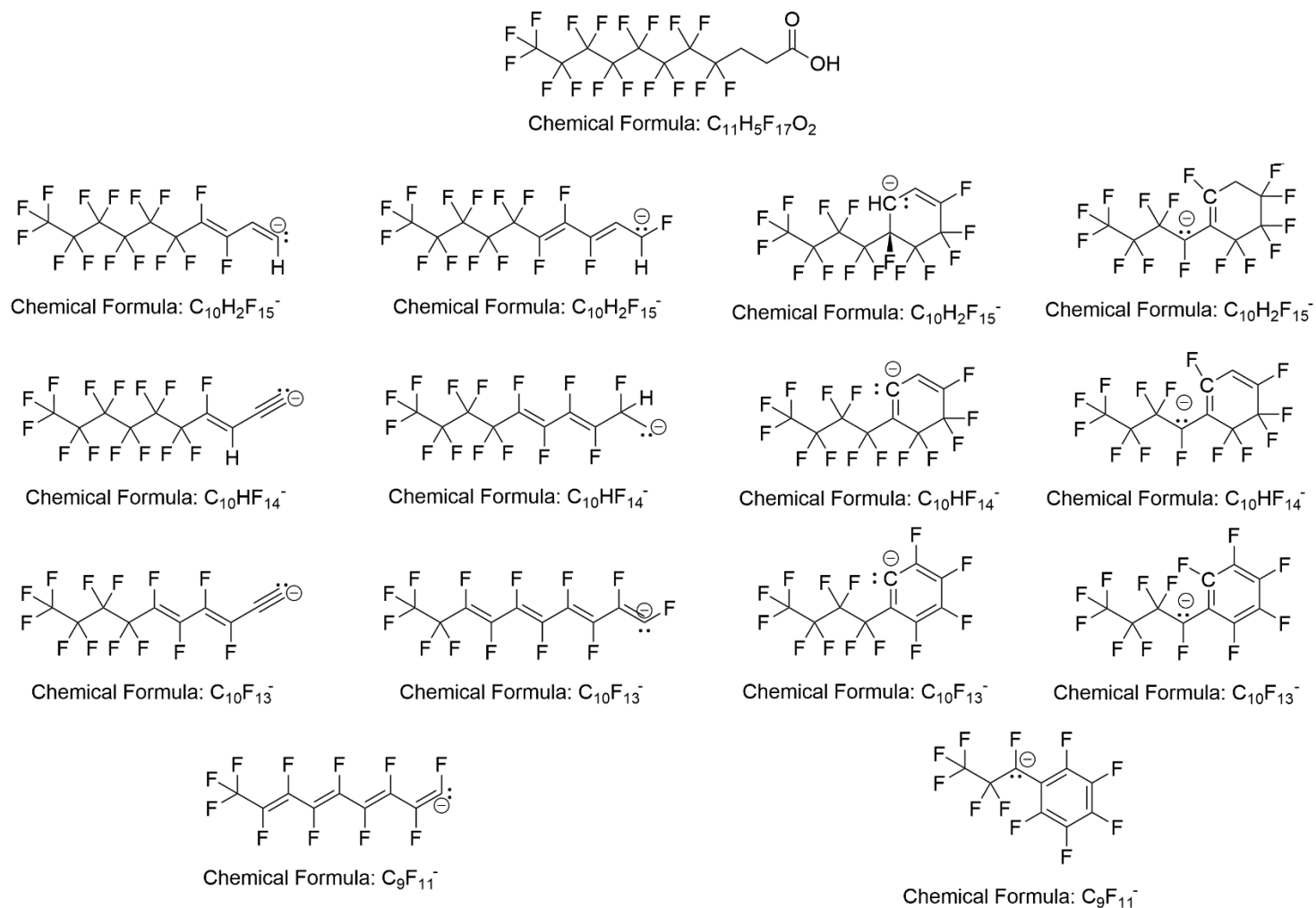


**Figure 6-8. Possible structures for the FTCA fragment ions observed in the tandem MS spectra.**

The top-most structure is the neutral 8:3 FTCA. Each subsequent row corresponds to a fragment ion observed in the FTCA-incorporated lipid spectra. Each fragment has multiple possible structures it could correspond to, and additional experimentation is required to confirm any of these structures.

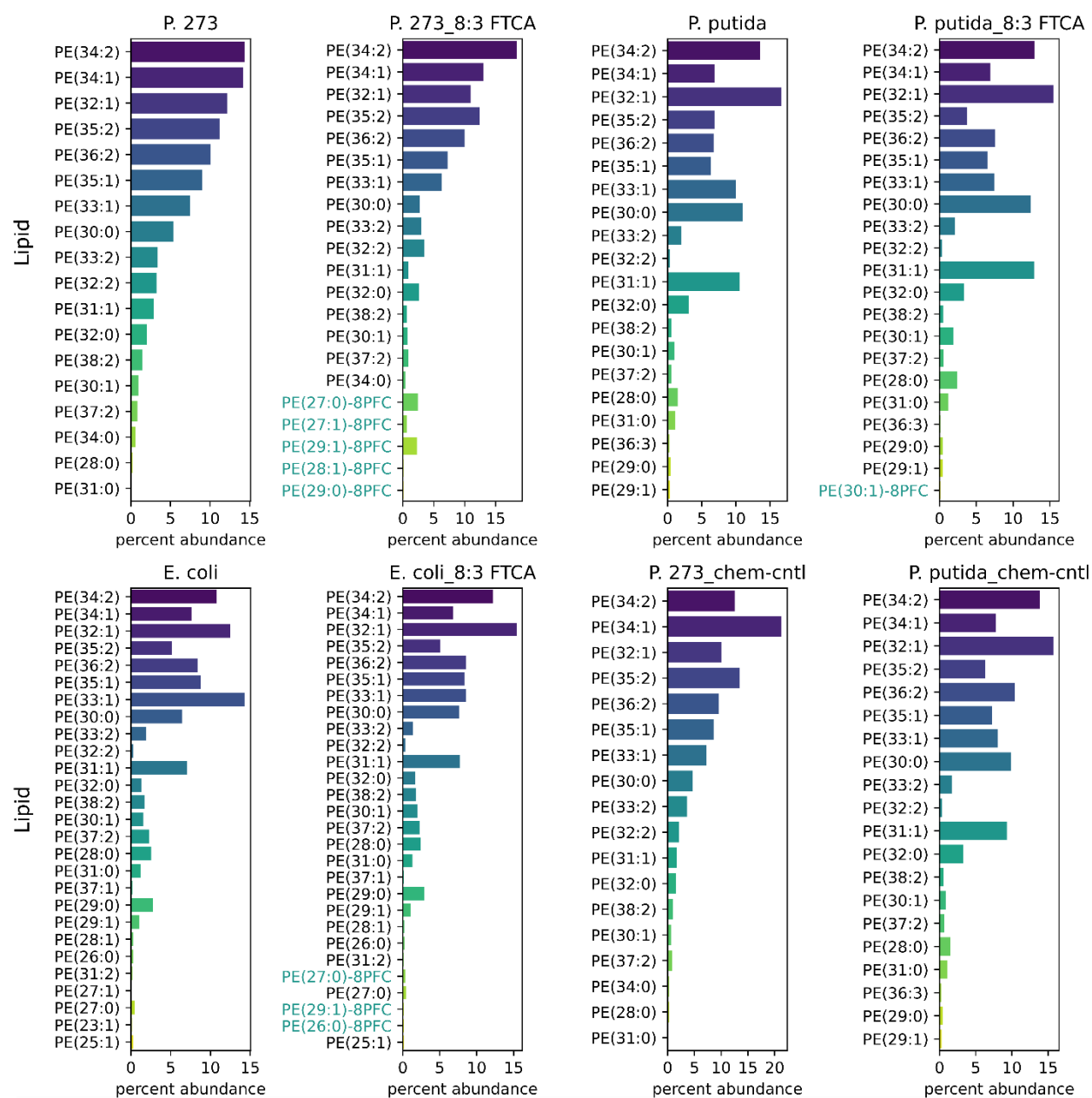


**Figure 6-9. Relative compositions of PE lipids.**

The relative abundance (percent of total signal) of all PE lipids exceeding 0.1% of the total PE signal for the given condition are shown. FTCA-containing lipid names are highlighted in cyan.

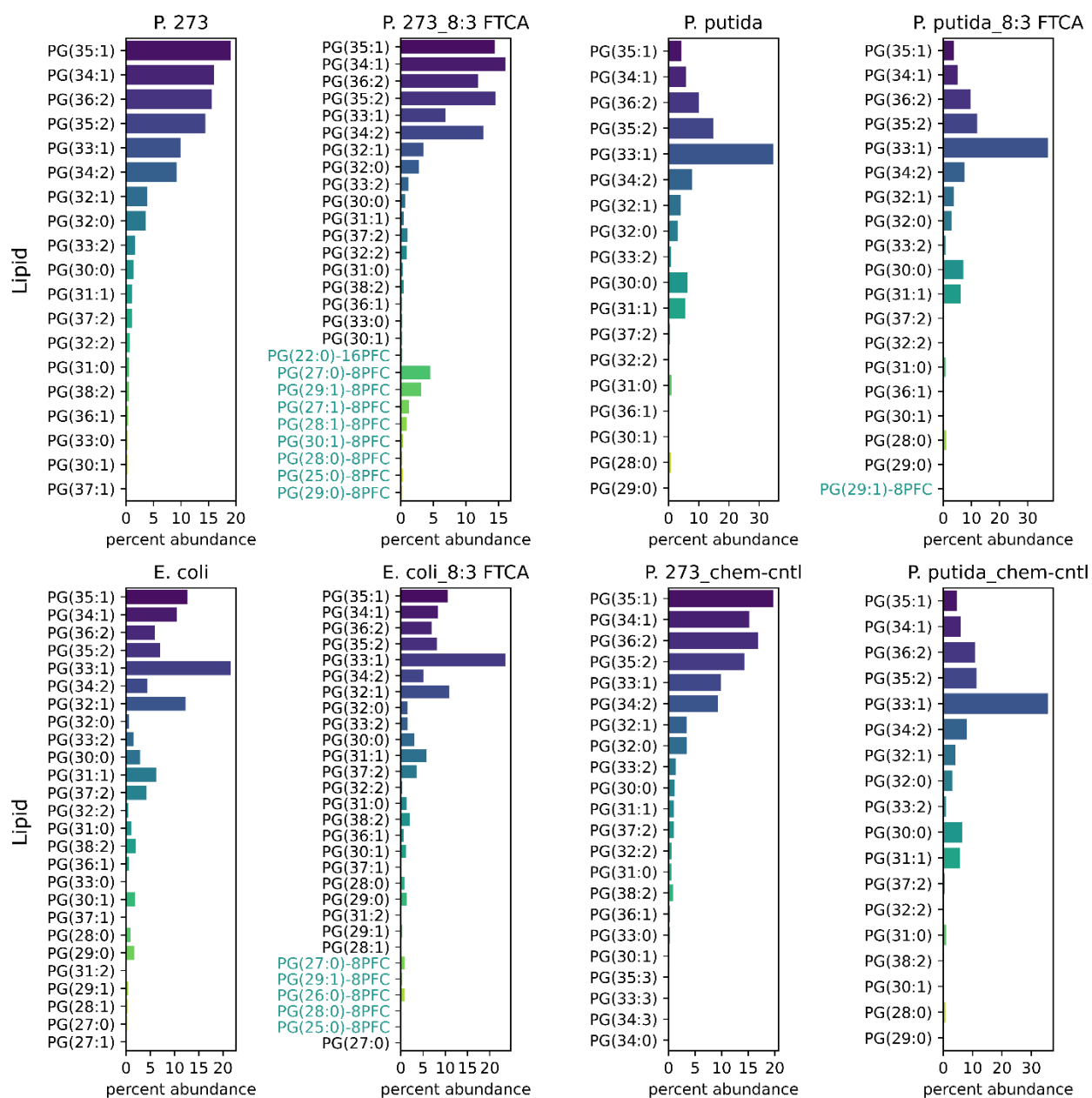


**Figure 6-10. Relative compositions of PG lipids.**

The relative abundance (percent of total signal) of all PG lipids exceeding 0.1% of the total PG signal for the given condition are shown. FTCA-containing lipid names are highlighted in cyan.

**Table 6-2. Comparison of PFAS membrane incorporation factors determined for different bacterial taxa cultivated in the presence of 5  $\mu$ M 8:3 FTCA.**

|                                                                                | <i>Pseudomonas</i><br>sp. strain 273 | <i>Pseudomonas</i><br><i>putida</i> strain GB1 | <i>Pseudomonas</i><br>sp. strain CBS3 | <i>Pseudomonas citronellolis</i><br>strain P3B5 | <i>E. coli</i> | <i>Enterococcus</i><br><i>faecalis</i> |
|--------------------------------------------------------------------------------|--------------------------------------|------------------------------------------------|---------------------------------------|-------------------------------------------------|----------------|----------------------------------------|
| <b>8:3 FTCA in PE (%)</b>                                                      | 5.9                                  | 0.2                                            | 0.4                                   | 0.3                                             | 0.8            | -                                      |
| <b>8:3 FTCA in PG (%)</b>                                                      | 11.6                                 | 0.4                                            | 1.3                                   | 0.8                                             | 2.6            | 0.8                                    |
| <b>PE/PG ratio<sup>a</sup></b>                                                 | 2.5                                  | 0.9                                            | 1.6                                   | 2.4                                             | 0.9            | -                                      |
| <b>8:3 FTCA in lipid (%)<sup>b</sup></b>                                       | 7.5                                  | 0.3                                            | 0.9                                   | 0.5                                             | 1.7            | 0.8                                    |
| <b>8:3 FTCA<sub>cell</sub> (mg PFAS<br/>per gram of biomass)<sup>c,d</sup></b> | 1.04                                 | 0.04                                           | 0.12                                  | 0.07                                            | 0.24           | 0.1                                    |
| <b>MIF</b>                                                                     | 423                                  | 17                                             | 49                                    | 27                                              | 97             | 43                                     |

a The PE/PG ratio represents the ratio between the total abundance of PE and PG based on the measured peak areas.

b The percentage of PFAS in lipids represents the weighted percentage of 8:3 FTCA in PE and in PG, calculated from the PE/PG ratio, except for *Enterococcus faecalis*, which contains 5 PG as the dominant lipid class.

c The calculation assumed an average cell wet weight of 1.3 pg.

d The same initial concentration of 8:3 FTCA (5  $\mu$ M) was supplied to all cultures, resulting in the same initial PFAS/water ratio (i.e., PFAS<sub>water</sub> = 0.00246 mg of 8:3 FTCA per gram of water).

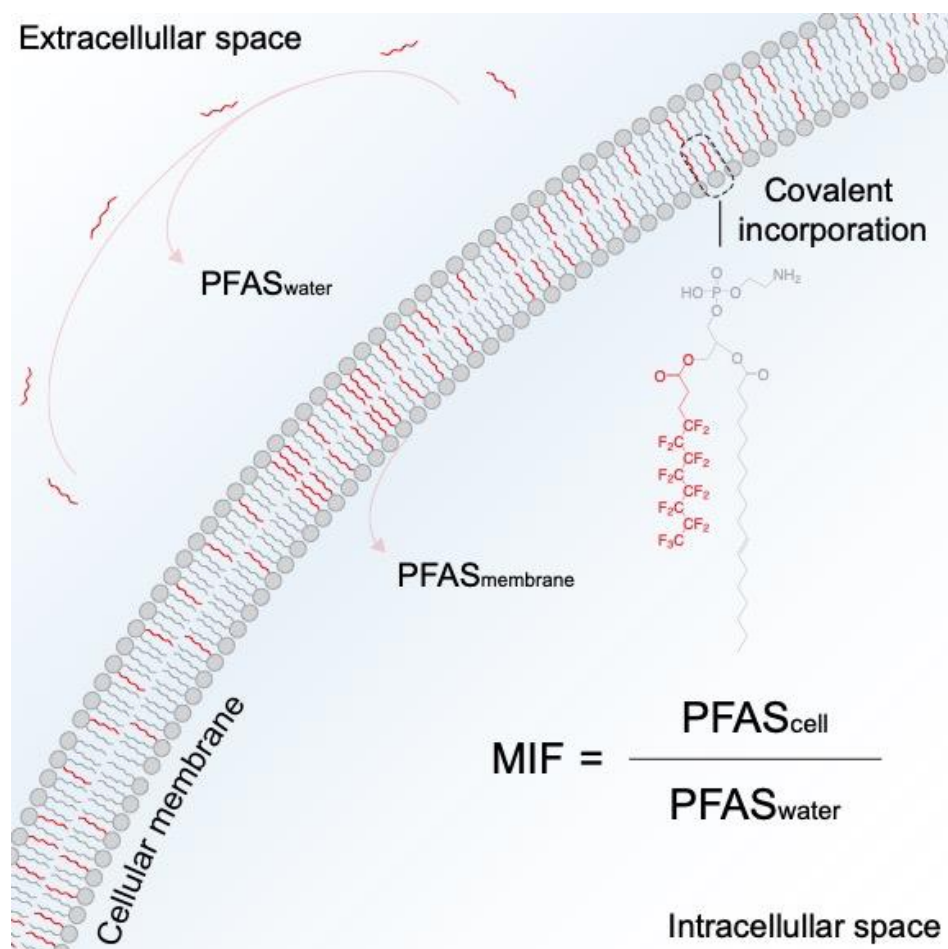


**Figure 6-11. Schematic illustrating the covalent incorporation of aqueous PFAS (i.e., 8:3 FTCA) into the bacterial lipid bilayer.**

The representative fluorinated phospholipid PE 29:1-8PFC (i.e., PE 8:3FTCA\_18:1) is displayed with covalent incorporation of acyl-8:3-FTCA in sn1 position. The membrane incorporation factor (MIF) represents the ratio between PFAS (e.g., 8:3 FTCA) concentration in the membrane (i.e., PFAS<sub>cell</sub> over the dissolved PFAS (e.g., 8:3 FTCA) concentration in the aqueous phase (i.e., PFAS<sub>water</sub>)).

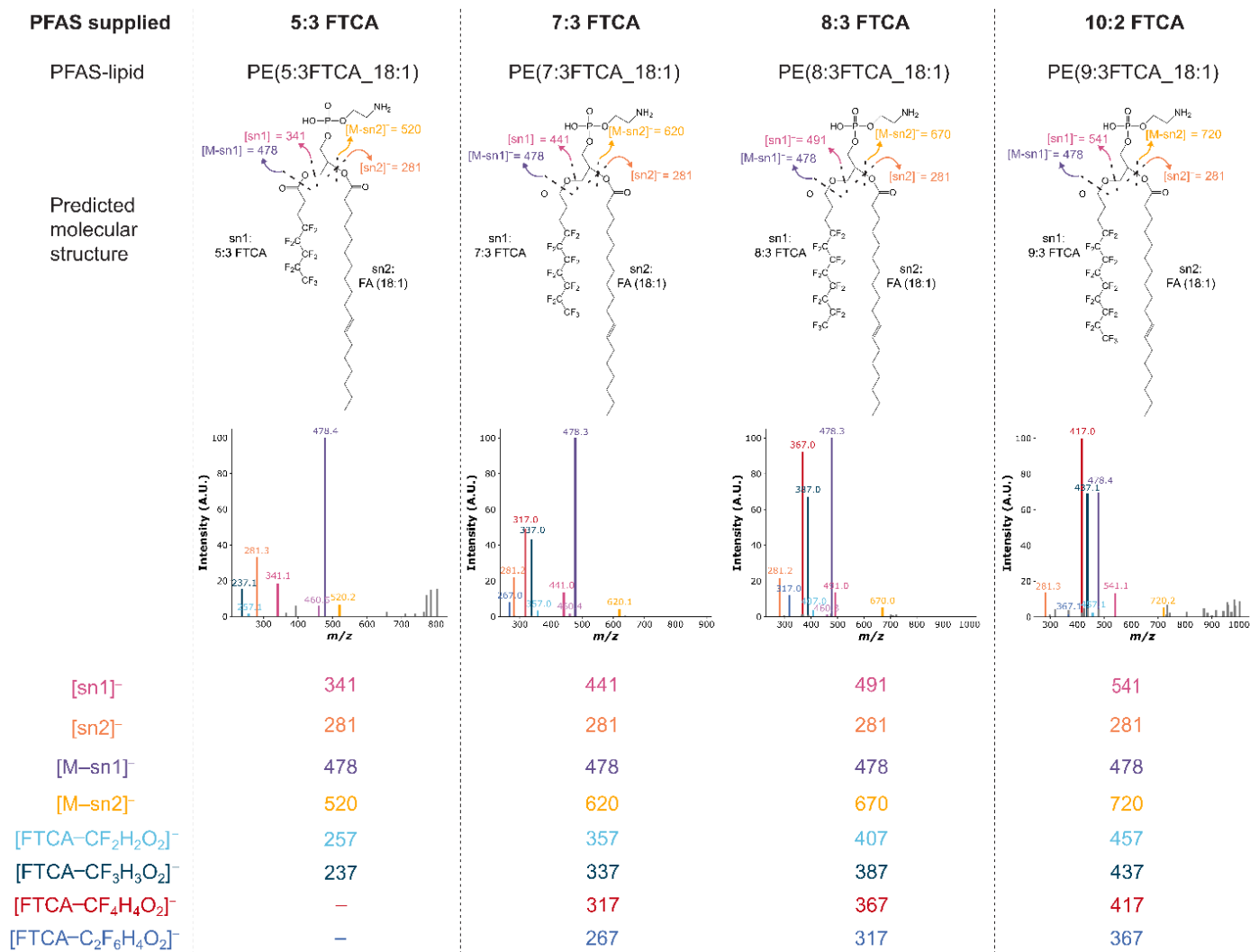


**Figure 6-12. Tandem MS of PFASylated lipids with 5:3, 7:3, 8:3, and 10:2 FTCA.**

Fragmentation spectra confirming incorporation of 5:3, 7:3, 8:3, and 9:3 FTCA from the 5:3, 7:3, 8:3, and 10:2 treated conditions, respectively.

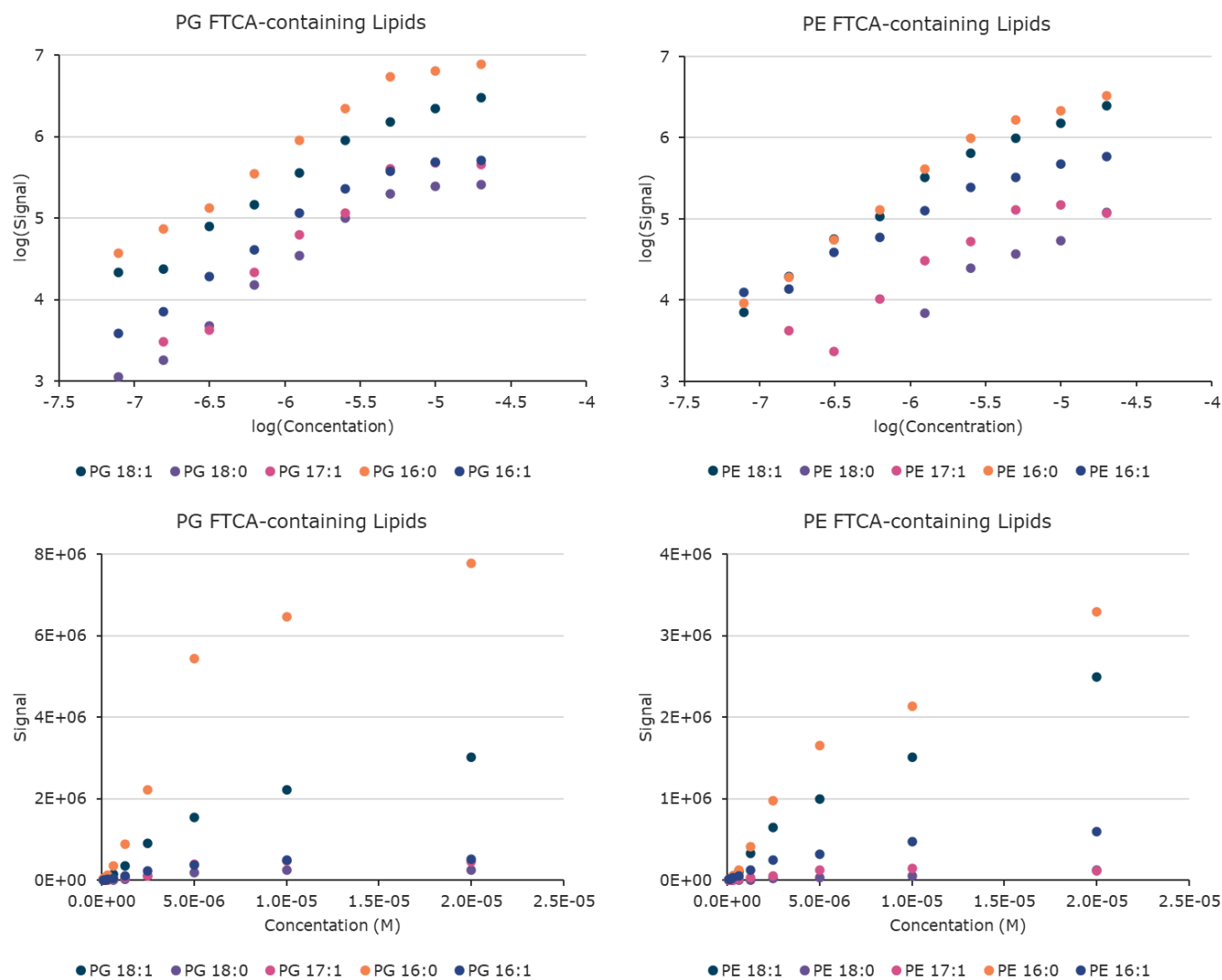


**Figure 6-13. PFASylated lipid signal as a function of FTCA concentration**

Lipid species consist of an 8:3 FTCA moiety and the FA listing in the legend. Log-log and raw plots are shown.

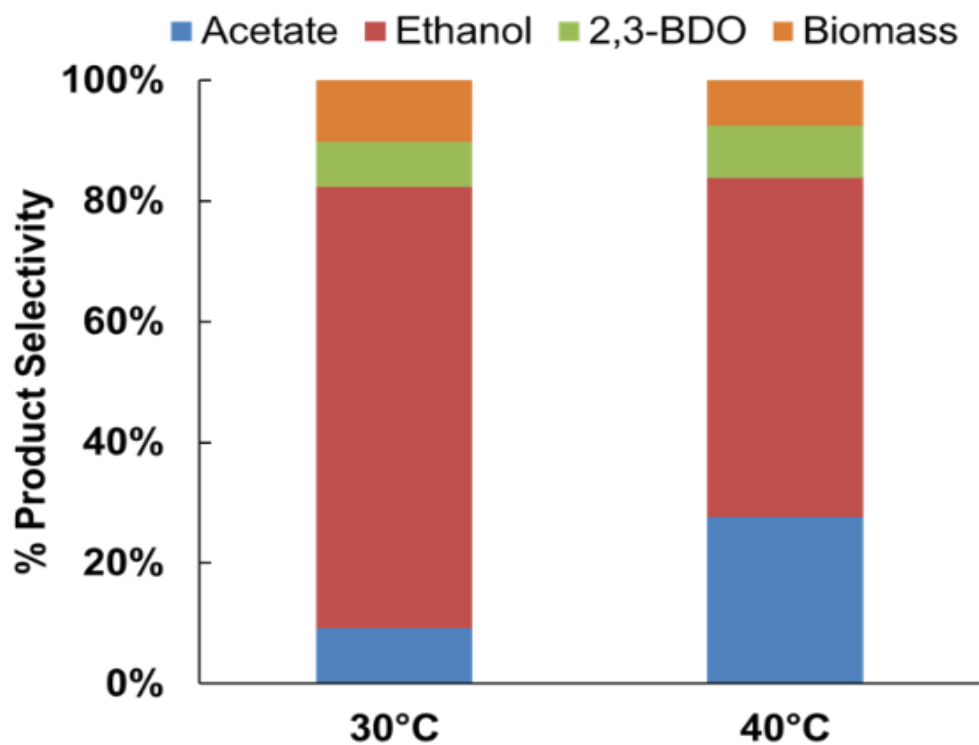


**Figure 7-1. End fermentation-product profile of *C. autoethanogenum* at 30 and 40 °C.** Selectivity of end-products at steady-state conditions for the *C. autoethanogenum* grown at the two temperatures. Lactate was not detected at these time points. ~8% and ~2% carbon could not be accounted for at 30 °C and 40 °C, respectively.

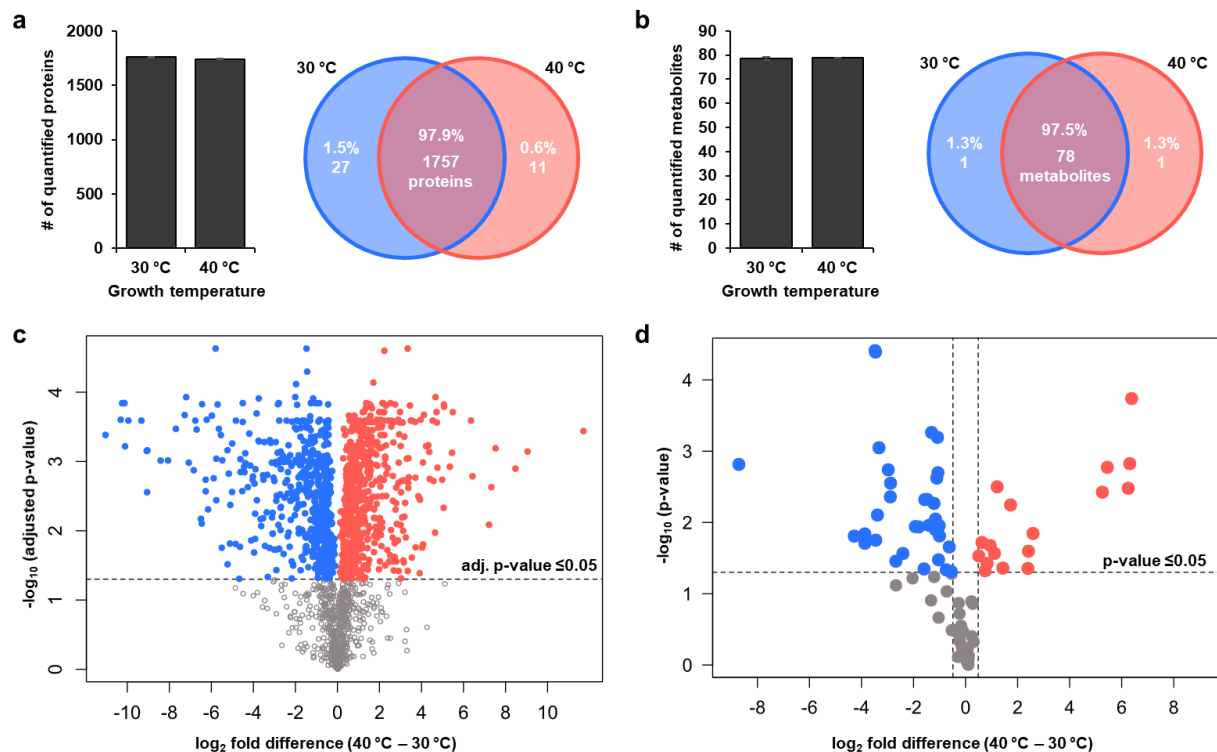


**Figure 7-2. Abundance changes in the proteome and metabolome drive growth at the two temperatures.**

(a) The diversity (count and identity) of proteins is comparable with growth temperature but abundance changes. (b) The diversity (count and identity) of metabolites is comparable with growth temperature but abundance changes. (c) Protein volcano plot. (d) Metabolite volcano plot. Each dot represents either a protein or a metabolite with the colors highlighting whether the proteins or metabolites significantly change in abundance and the direction of change. Red: high at 40 °C, Blue: low at 40 °C Grey: not significant. Data represent mean  $\pm$  standard deviation of 3 replicates.

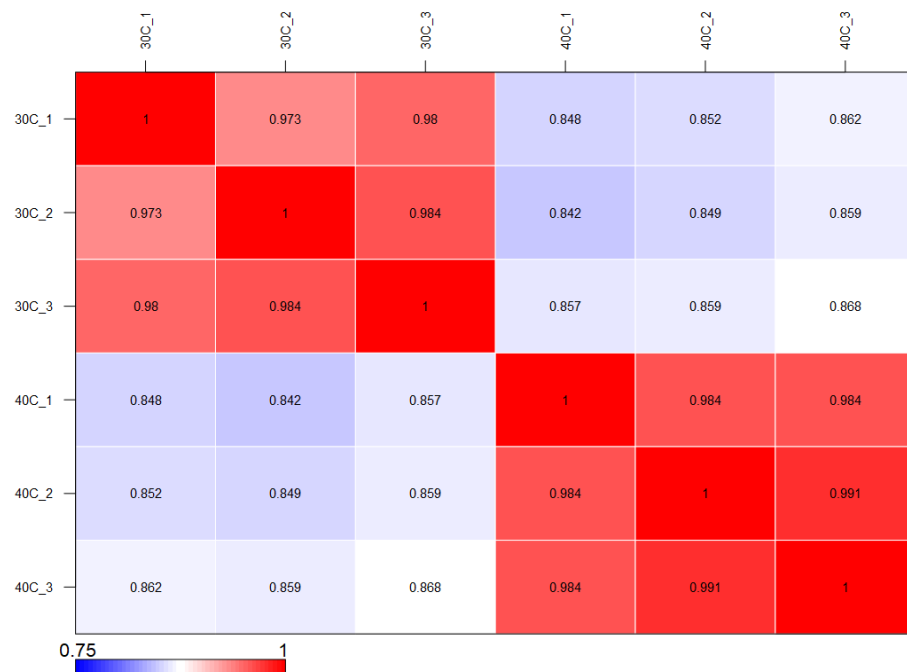


**Figure 7-3. Proteome correlation heatmap.**

Heatmap showing the correlation of the proteomes among all the samples at the two temperatures.

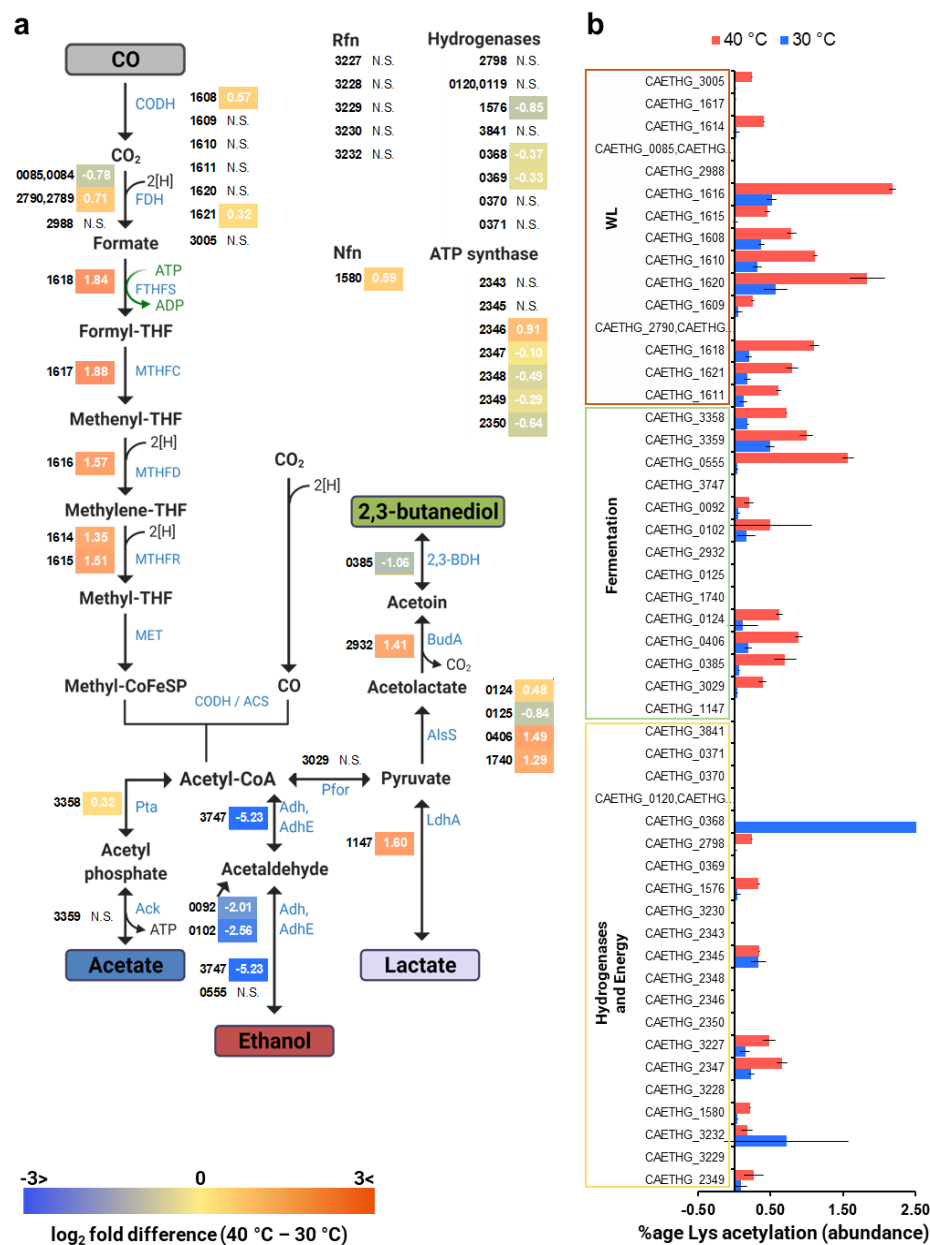


**Figure 7-4. Central metabolic pathways in *C. autoethanogenum*.**

(a) Proteome coverage of the WL and fermentation pathways in *C. autoethanogenum* along with the hydrogenases and energy conservation complexes. (b) Abundance contributed by peptides acetylated at lysine residues for pathway enzymes. The number annotations of the proteins should be preceded by “CAETHG\_” to generate locus tags. Data represent mean ± standard deviation of 3 replicates. The full list of proteins belonging to the central metabolism is provided in SI, here the quantified proteins have been marked.

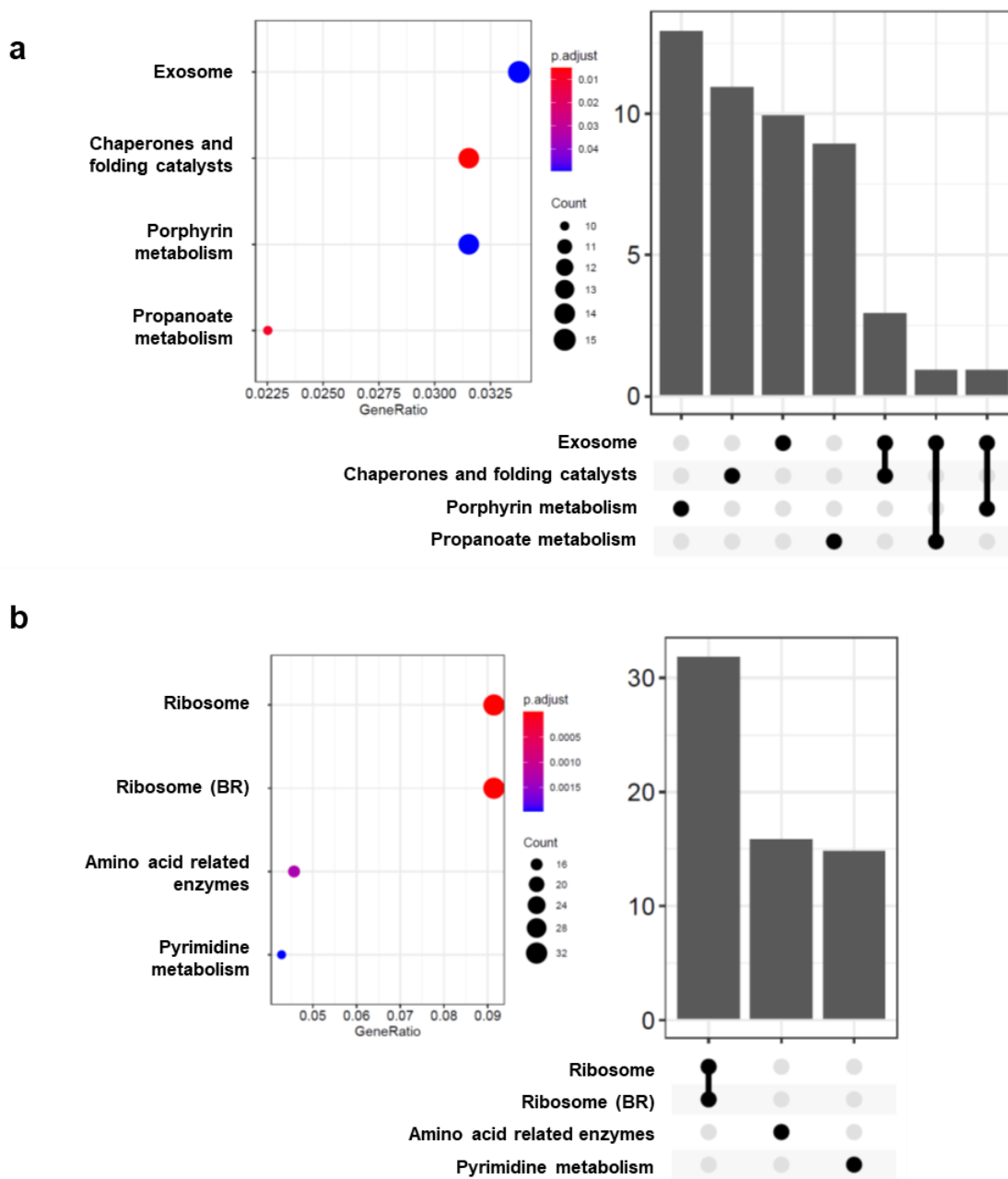


**Figure 7-5. Proteome KEGG pathway enrichment analysis.**

Pathway overrepresentation analysis for proteins for differentially abundant due to change in temperature. KEGG pathways upregulated at 40 °C are shown in (a) and those found to be downregulated are shown in (b). ClusterProfiler was used for identifying overrepresented pathways based on KEGG Orthology terms.

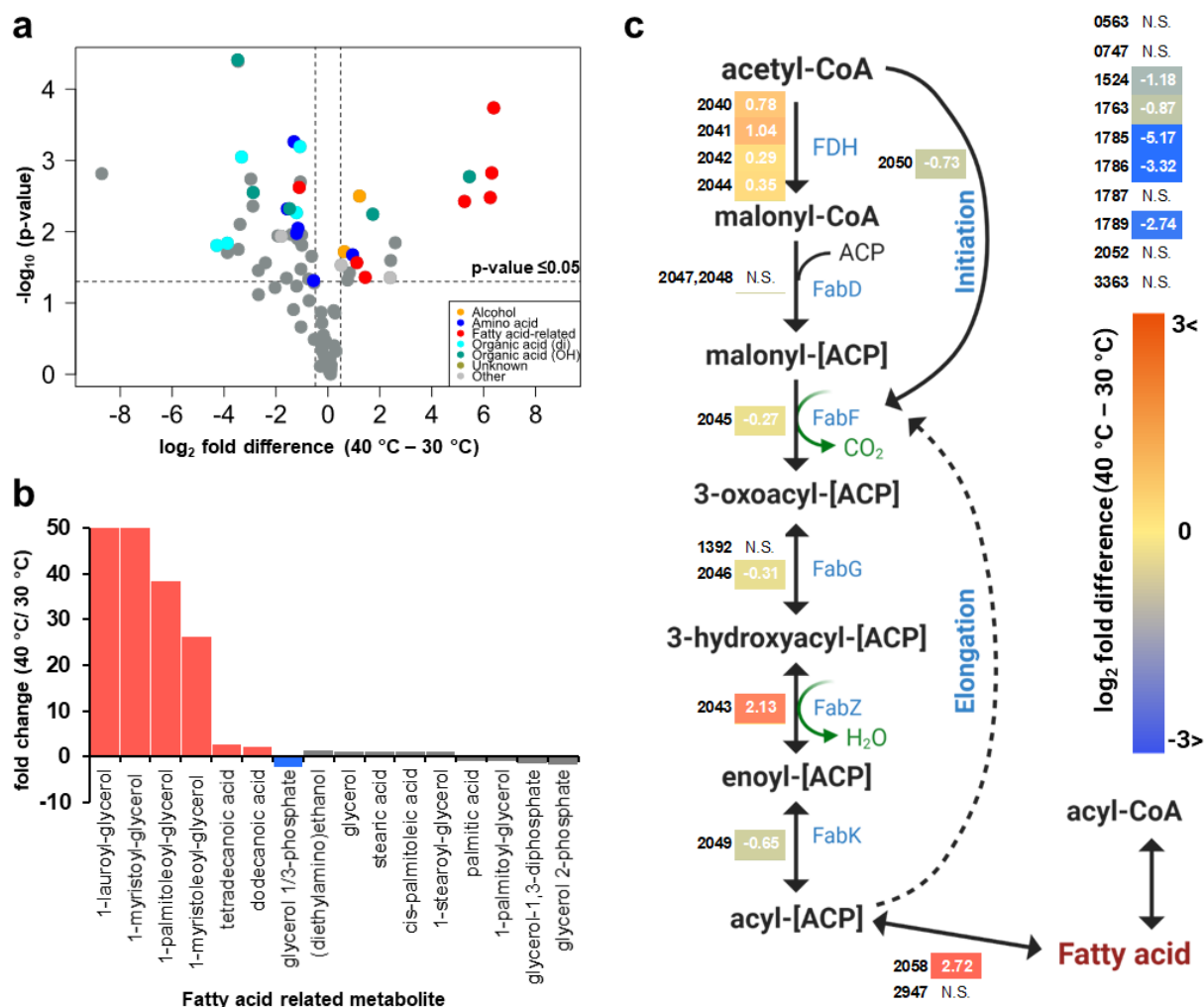


**Figure 7-6. Fatty acids accumulate at high temperature with changes in fatty acid metabolism.**

(a) Volcano plot of metabolites quantified by GC-MS. (b) Changes in quantified fatty-acid related metabolites with temperature. (Significantly increased at 40 °C – red, decreased – blue, not changing – grey) (c) Predicted metabolic pathway saturated fatty acid biosynthesis in *C. autoethanogenum* with quantified proteins mapped. The number annotations of the proteins should be preceded by “CAETHG\_” to generate locus tags. Data represent the mean of 3 replicates

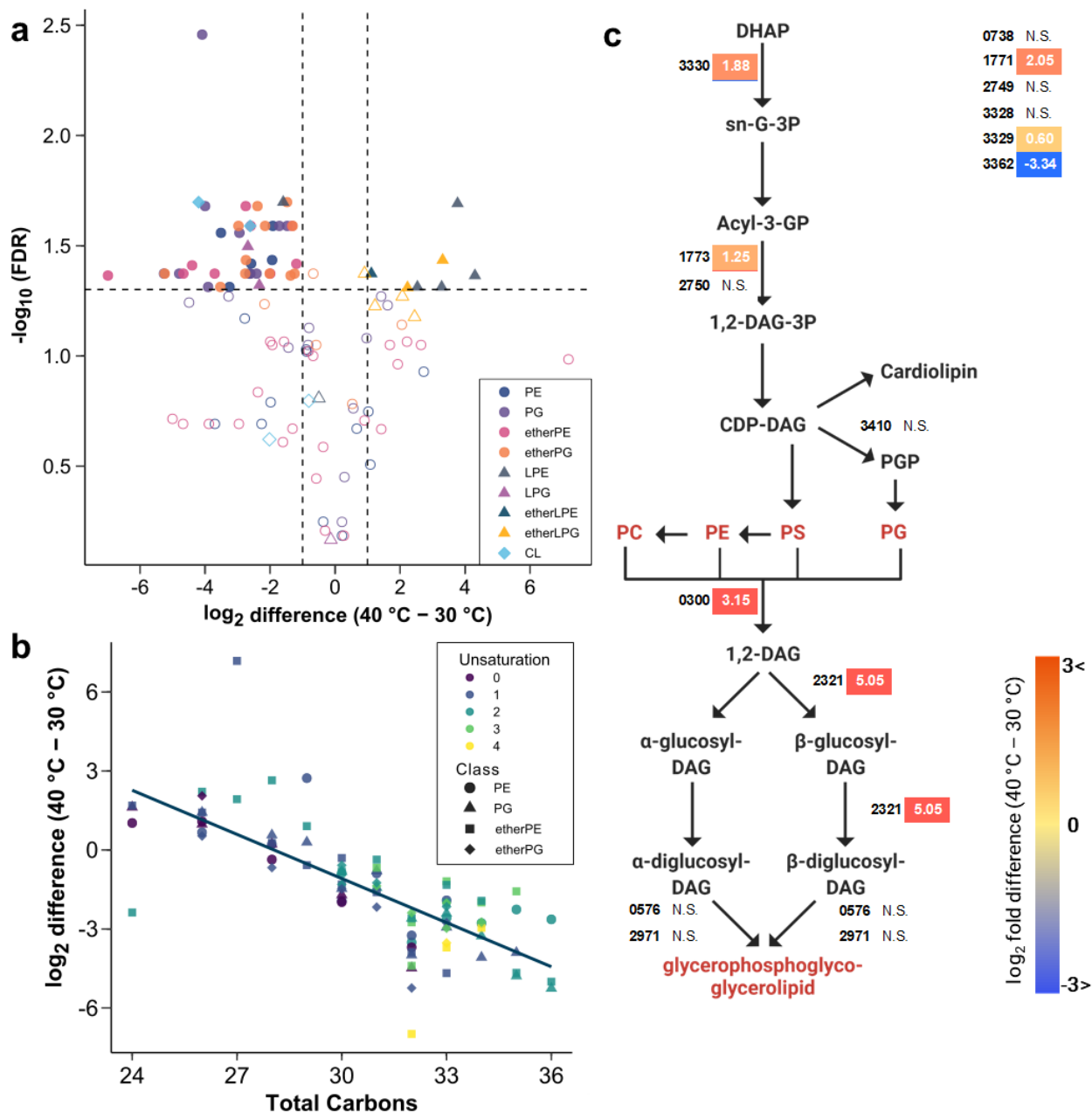


**Figure 7-7. Lipidomics results and lipid proteome fold-changes.**

a) Volcano plot of quantified lipids. Horizontal and vertical dashed lines represent 0.05 FDR (Benjamini-Hochberg) and two-fold change (1  $\log_2$  difference) cutoffs respectively. (b) Ester and ether PE and PG lipids' foldchanges plot against total fatty acid carbon. (c) Lipid metabolism pathway (incomplete) with quantified proteins mapped. The number annotations of the proteins should be preceded by "CAETHG\_" to generate locus tags.



**Figure 7-8. Anaerobic plasmalogen homolog.**

Domain analysis of protein CAETHG\_1442. All predicted domains for plasmalogen producing PlsA/R enzyme were present in this protein.

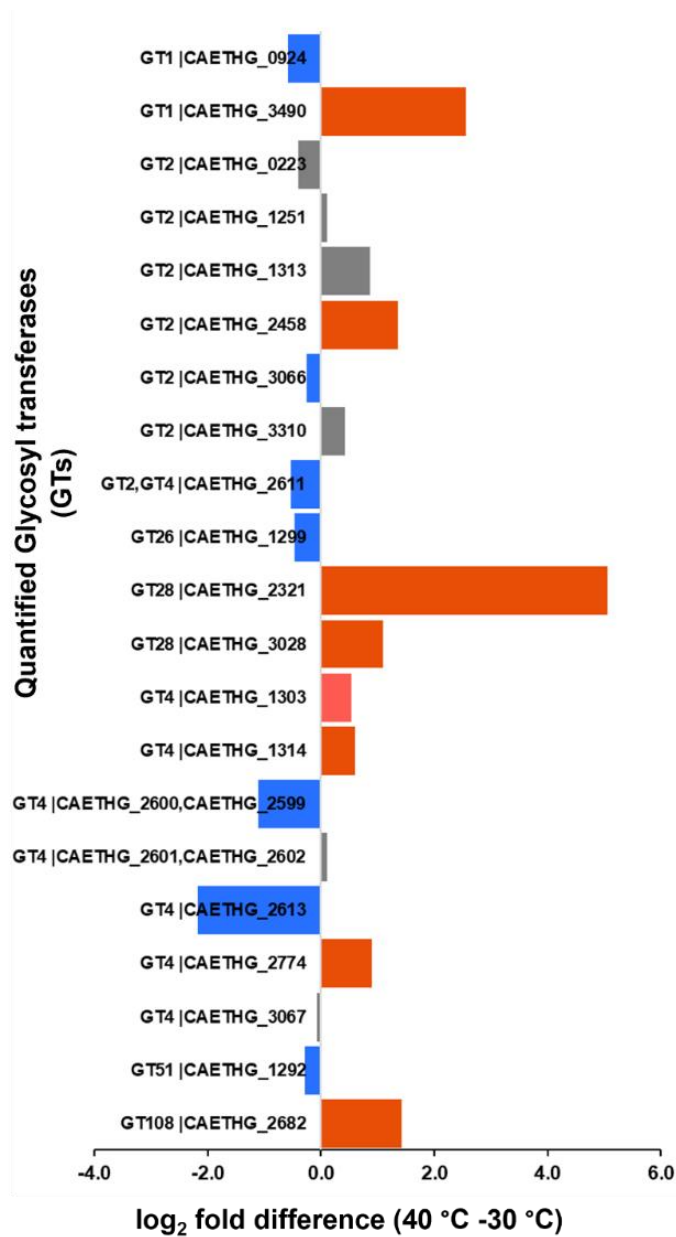


**Figure 7-9. Glycosyl transferases fold-changes.**

Glycosyl transferases that were quantified across the two temperatures. The fold changes are represented by the length of the bars – high at 40 °C (red) and low at 40 °C (blue).

**Table 7-1. Proteins with the highest fold changes with temperature.**

| Accession     | Locus tag   | Description                                           | Direction     | log2 FC |
|---------------|-------------|-------------------------------------------------------|---------------|---------|
| <b>U5RVB9</b> | CAETHG_1078 | Alcohol dehydrogenase                                 | Low in 40 °C  | -8.41   |
| <b>U5RW61</b> | CAETHG_1373 | Phosphoglycerate dehydrogenase                        | Low in 40 °C  | -10.13  |
| <b>U5RVM6</b> | CAETHG_1374 | Dimethylmenaquinone methyltransferase                 | Low in 40 °C  | -10.25  |
| <b>U5RSA1</b> | CAETHG_1375 | Major facilitator superfamily MFS_1                   | Low in 40 °C  | -9.33   |
| <b>U5RWM2</b> | CAETHG_1718 | SGNH_hydro domain-containing protein                  | Low in 40 °C  | -11.04  |
| <b>U5RU95</b> | CAETHG_2081 | Carbamate kinase                                      | Low in 40 °C  | -10.31  |
| <b>U5RU62</b> | CAETHG_2082 | Putrescine carbamoyltransferase (PTCase)              | Low in 40 °C  | -7.69   |
| <b>U5RYY6</b> | CAETHG_2565 | Peptidase M16C associated domain protein              | Low in 40 °C  | -9.04   |
| <b>U5RY34</b> | CAETHG_3274 | Pyruvate formate-lyase PFL                            | Low in 40 °C  | -7.27   |
| <b>U5S1K1</b> | CAETHG_3275 | Glycyl-radical enzyme activating protein family       | Low in 40 °C  | -9.06   |
| <b>U5RY39</b> | CAETHG_3279 | Alcohol dehydrogenase                                 | Low in 40 °C  | -10.10  |
| <b>U5S1K6</b> | CAETHG_3280 | MIP family channel protein                            | Low in 40 °C  | -10.16  |
| <b>U5S127</b> | CAETHG_3281 | Uncharacterized protein                               | Low in 40 °C  | -8.06   |
| <b>U5S1K9</b> | CAETHG_3285 | Ethanolamine utilization EutQ family protein          | Low in 40 °C  | -9.95   |
| <b>U5RXP3</b> | CAETHG_3287 | Acetaldehyde dehydrogenase (Acetylating)              | Low in 40 °C  | -9.08   |
| <b>U5RT73</b> | CAETHG_0322 | Xylose isomerase domain-containing protein TIM barrel | High in 40 °C | 5.07    |
| <b>U5RUC3</b> | CAETHG_0714 | Uncharacterized protein                               | High in 40 °C | 4.99    |
| <b>U5RWJ3</b> | CAETHG_1515 | Uncharacterized protein                               | High in 40 °C | 5.49    |
| <b>U5RXF4</b> | CAETHG_1819 | Acetaldehyde dehydrogenase (Acetylating)              | High in 40 °C | 5.08    |
| <b>U5RTF4</b> | CAETHG_1822 | Microcompartments protein                             | High in 40 °C | 7.53    |
| <b>U5RTF7</b> | CAETHG_1827 | Uncharacterized protein                               | High in 40 °C | 5.35    |
| <b>U5RTW9</b> | CAETHG_1828 | Choline trimethylamine-lyase activating enzyme (CutD) | High in 40 °C | 7.33    |
| <b>U5RTJ3</b> | CAETHG_1831 | Microcompartments protein                             | High in 40 °C | 8.47    |
| <b>U5RTG1</b> | CAETHG_1832 | Microcompartments protein                             | High in 40 °C | 9.05    |
| <b>U5RTX4</b> | CAETHG_1833 | PadR domain-containing protein                        | High in 40 °C | 5.45    |
| <b>U5RU74</b> | CAETHG_2094 | Heat shock protein Hsp20                              | High in 40 °C | 7.21    |
| <b>U5RUL4</b> | CAETHG_2095 | Heat shock protein Hsp20                              | High in 40 °C | 11.70   |
| <b>U5RVB1</b> | CAETHG_2321 | Monogalactosyldiacylglycerol synthase                 | High in 40 °C | 5.05    |
| <b>U5RV44</b> | CAETHG_2389 | Quinolinate phosphoribosyl transferase                | High in 40 °C | 6.43    |
| <b>U5S274</b> | CAETHG_3661 | C_GCAxxG_C_C family protein                           | High in 40 °C | 6.36    |

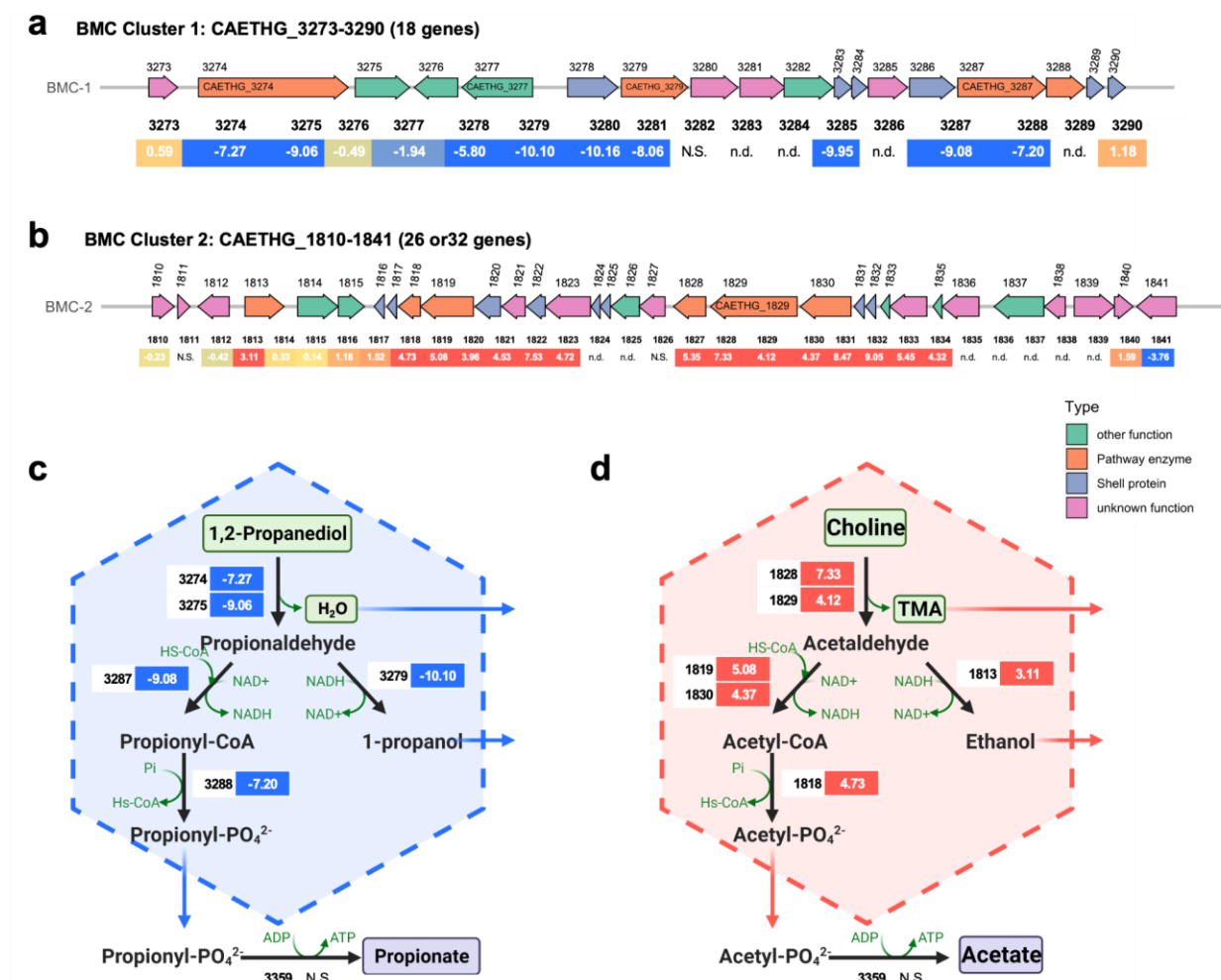


**Figure 7-10. Bacterial Microcompartment related proteins.**

(a) Structure of BMC cluster downregulated at 40 °C. (b) Structure of BMC cluster upregulated at 40 °C. (c) Predicted metabolic pathway for BMC cluster downregulated at 40 °C. (d) Predicted metabolic pathway of BMC cluster upregulated at 40 °C. The number annotations of the proteins should be preceded by “CAETHG\_” to generate locus tags. The fold changes are depicted as log2 values.

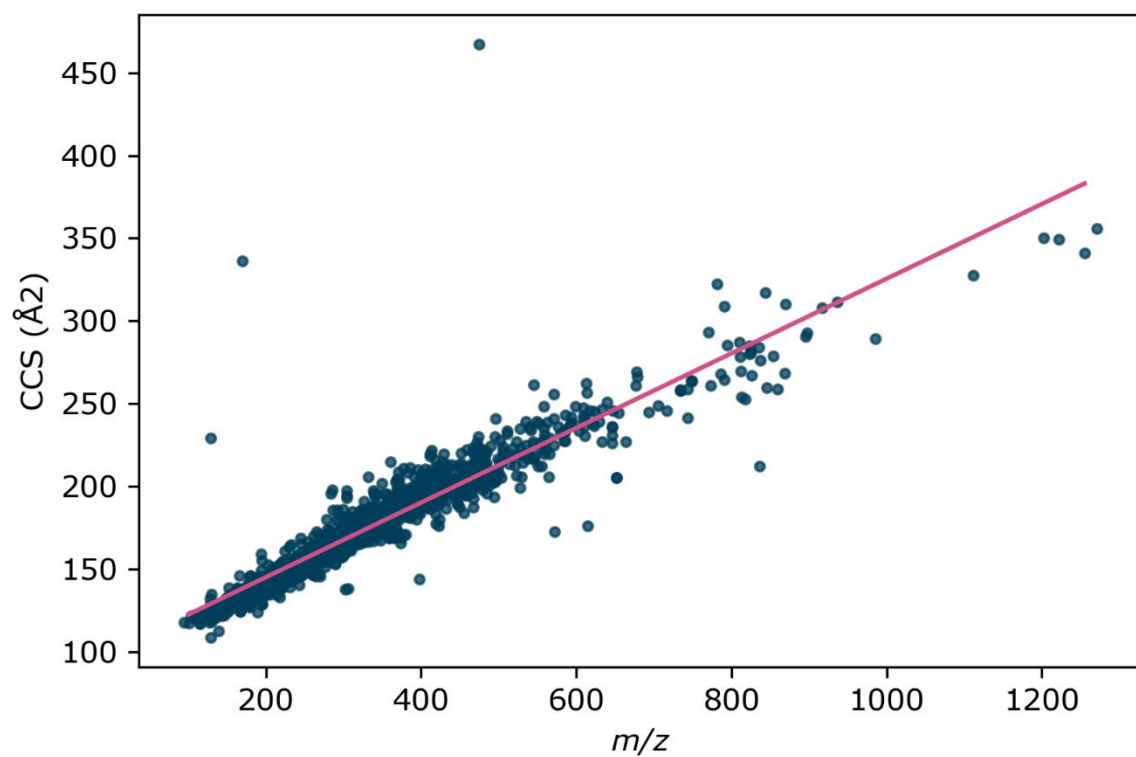


**Figure 8-1. Mass-to-charge collisional cross section relationship.**

Plot of CCS vs  $m/z$  for a large set of drug-like molecules. Data from Hines *et al.*<sup>287</sup> The pink lines shows the correlation (linear regression) between  $m/z$  and CCS.

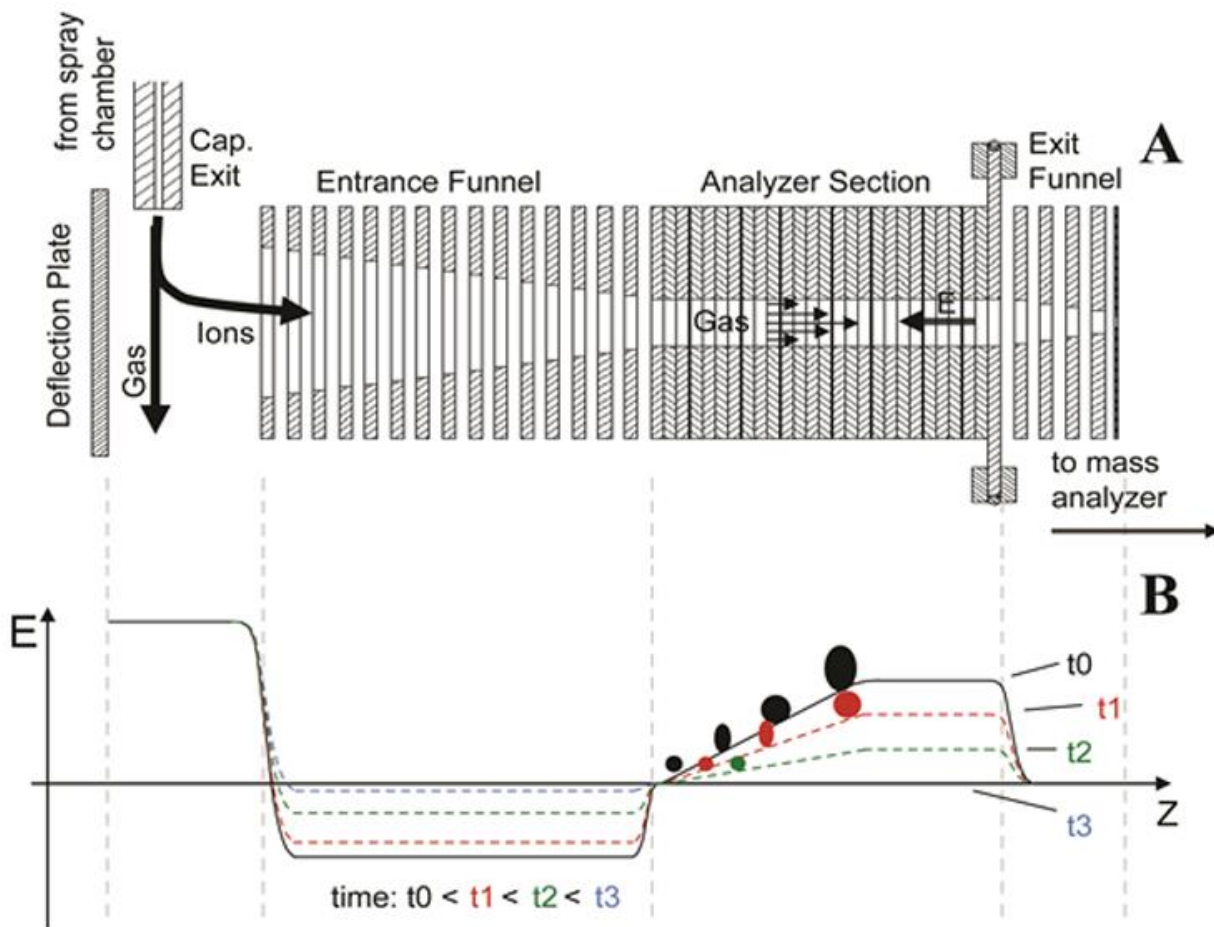


**Figure 8-2. TIMS cell diagram**

In the TIMS cell, gas flow pushes ions forward (right) with a force proportional to their CCS. A voltage bias opposes this force, allowing ions to be trapped in the entrance funnel. Ions in the analyzer will exit to the mass analyzer when the electric field strength is lower than the push provided by the gas flow. Figure adapted from the Boston University School of Medicine's Fourier Transform Mass Spectrometry Lab. At <https://www.bumc.bu.edu/ftms/research/tims/>.

**Table 8-1. Replication Rates of Cancer Biology Research.**

Tabular representation of replication experiments from 23 preclinical cancer biology papers. ES stands for effect size; CI stands for 95% confidence interval. Data reproduced from Errington *et al.*<sup>275</sup>

| <b>Outcome Scale</b>                      | <b>Papers</b> | <b>Experiments</b> | <b>Effects</b> | <b>All outcomes</b> |
|-------------------------------------------|---------------|--------------------|----------------|---------------------|
| <b>Count</b>                              | 23            | 50                 | 158            | 188                 |
| <b>Same ES direction</b>                  | 17/19 (89%)   | 26/35 (74%)        | 80/101 (79%)   | 95/116 (82%)        |
| <b>Same ES direction and significance</b> | 8/19 (42%)    | 17/33 (52%)        | 42/97 (43%)    | 44/112 (39%)        |
| <b>Original ES in Replication CI</b>      | 5/19 (26%)    | 3/33 (9%)          | 17/97 (18%)    | 26/112 (23%)        |
| <b>Replicate ES in Original CI</b>        | 5/19 (26%)    | 11/33 (33%)        | 42/97 (43%)    | 50/112 (45%)        |
| <b>Replicate ES &lt; Original ES</b>      | 18/19 (95%)   | 32/33 (97%)        | 94/97 (97%)    | 109/112 (97%)       |

### **Vita**

Matthew Keller was born in California and raised in Connecticut and Tennessee. He graduated Summa Cum Laude from the University of Tennessee, Chattanooga with a B.S. in Chemistry with a minor in Biology. He also achieved an M.S. degree in Analytical Chemistry from the University of Tennessee, Knoxville before joining the Genome Science and Technology Ph.D. program. Scientific interests include mass spectrometry, all forms of chromatography, statistical analysis, molecular biology, and experimental reproducibility.