Improved Liquid Chromatography-Mass Spectrometry Approaches for Characterization of Therapeutic RNA

by

Carson W. Szot

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in the University of Michigan
2023

Doctoral Committee:

Professor Kristina I. Håkansson, Chair
Assistant Professor Sarah Keane
Associate Professor Kristin Koutmou
Professor Brandon T. Ruotolo

Carson W. Szot

cwszot@umich.edu

ORCID iD: 0000-0001-5752-6220

Dedication

This dissertation is dedicated to my family, friends, and coworkers who have accompanied this journey.

Acknowledgements

I would like to acknowledge the assistance and support of the many people who have supported and guided my work: Kristina Hakansson, Kermit Murray, Revati Kumar, Kresimir Rupnik, Bijay Banstola, Kelin Wang, Fabrizio Donnarumma, Eunju Jang, Josh Salem, LeeAnne Wang, Steven DeFiglia, Nhat Le, Achala Kankanamalage, Menatallah Youssef, Tracy Hodges, Neven Mikaway, Hye Kyong Kweon, Yuri Tsybin, Ingrid Kilde, Elizabeth Tidwell, Valery Voinov, Anna Anders, Sarah Keane, Kristin Koutmou, Brandon Ruotolo, Scott Daniels, Bryan Miller, Ruwan Kurulugama, Richard Chastney, Jennifer Lippens, Zhouer Xie, Vivian Crum, Mike Hartz, Kevin Ileka, Nick Borotto, Varun Gadkari, Vitaly Pover, David Carter, James Tice, and Cameron Nobile.

I would also like to acknowledge the funding support from the National Science Foundation (CHE2004043), National Institutes of Health (RO1GM139916), and an Agilent Thought Leader Award.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

This dissertation focuses on extending and improving current applications of liquid chromatography-mass spectrometry (LC-MS) for therapeutic oligonucleotides (ONs), which continue to grow in both clinical importance and complexity, including a variety of synthetic modifications. These RNA based therapeutics represent state of the art therapeutic drugs capable of treating a wide variety of health problems. The central placement of RNA in an organism's cellular systems allows oligonucleotide therapeutics to act on virtually all biological processes of the body suggesting possibly limitless potential application. A triethylammoniumbicarbonate (TEAB)/methanol-based LC mobile phase system, free of fluorinated alcohol additives, is shown to allow sufficient chromatographic resolution and mass spectral signal in significantly more environmentally friendly workflows compared with the conventional hexafluoroisopropanol (HFIP) additive, which is both physiologically and environmentally toxic. Additionally, the use of methanol instead of acetonitrile further reduces the environmental impact of these workflows. With this type of chromatography, we address three major challenges in therapeutic oligonucleotide LC-MS analysis, including difficulties with tandem mass spectrometry (MS/MS)-based sequencing of chemically modified ONs, automated annotation of the corresponding LC-MS/MS spectra, and extension to larger nucleic acids.

In Chapter 2, we introduce OligoTap, a bioinformatics utility that allows consideration of any type of chemical modification at any ON location. To date no existing open-source software is similarly flexible. Additionally, no other open-source software allows vendor agnostic data processing. The significantly enhanced throughput of OligoTap-based automated annotation of MS/MS spectra allows for a range of charge states and collision induced dissociation (CID) voltages to be assessed for improved sequence coverage. With this approach, we reached 100% sequence coverage for a heavily modified 22mer RNA, including previously refractory 2' fluorine modifications. This novel report of full sequence coverage using environmentally friendly chemicals sets a new bar for therapeutic oligonucleotide sequencing by mass spectrometry. In Chapter 3, OligoTap is further evaluated with both CID and electron detachment dissociation (EDD) LC-MS/MS data using a larger search space containing isomeric sequence permutation decoys. The complexity of CID fragmentation remains a challenge in confident spectral annotation. This analysis illustrates the possibility of incorrect sequence assignments despite high mass accuracy and resolution. We introduce a novel quantitative approach to evaluate false discovery rates (FDRs) and demonstrate that EDD, under the utilized conditions/assumptions shows lower FDR compared with CID. We further show that EDD provides a twofold improvement in the ability to identify the correct sequence from decoys, suggesting that EDD is able to double the confidence in spectral annotations. This approach will facilitate further optimization of LC-MS/MS workflows.

Finally, in Chapter 4, we demonstrate that TEAB/methanol-based LC-MS on a 7 Tesla Fourier transform-ion cyclotron resonance mass spectrometer allows detection

and charge state deconvolution of RNA up to ~100 kDa. The wider charge state distributions generated by electrospray ionization from TEAB/methanol compared with HFIP-based solvent systems improve the input of the deconvolution algorithm, which is able to differentiate phosphate loss and iron adduction at this mass range. We also demonstrate isotopic resolution for an ~40 kDa RNA along with resolution of accompanying sodium and potassium adducts. Overall, the presented advances will aid the continuously developing area of LC-MS-based ON analysis for improved characterization of therapeutic oligonucleotides.

Chapter 1 Introduction

1.1 Importance of Therapeutic Oligonucleotides and their Analytical Characterization

Oligonucleotides (ONs) as tools for therapeutic treatments are a developing resource.[1] Common approaches for oligonucleotide therapies are gene augmentation, suppression, and editing.[2] Early studies of transcription regulation provided examples of synthetic oligonucleotide interference.[3] At that time, anion exchange chromatography with an increasing gradient of triethylammonium bicarbonate (TEAB) was employed to purify oligonucleotides from total synthesis.[3] Ribonucleic acid interference modalities, including chemically modified variants, have grown in popularity and reached various stages of approval.[4] **Figure 1.1** shows human tissue systems for which oligonucleotide interference is being developed to treat disease.[4] Recently mRNA-based vaccines have become mainstream in light of the corona virus disease pandemic.[5] These RNA based therapeutics represent the most modern therapeutic drugs capable of treating a wide variety of health problems. RNA therapeutic action is far upstream in

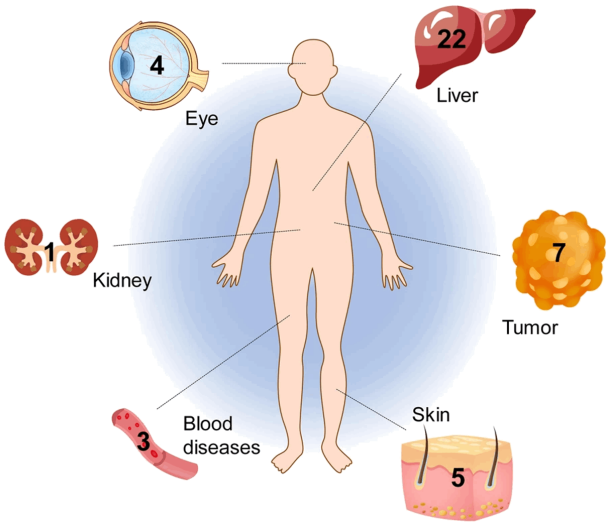


Figure 1.1 Tissue systems of the body for which siRNA therapeutics are in clinical development for disease treatment. The number in the center of the tissue indicates the number of siRNAs in development for that tissue. Figure adapted from Hu et al.[4] (2020)

regulation and metabolism. This core location in an organism's cellular systems allows oligonucleotide therapeutics to act on potentially all biological processes of the body. Therapeutic use of ONs places a high demand on their characterization and quality control. In addition, their complexity, high molecular weight, and physiochemical properties present several challenges for bioanalytical method development.[6]

Analytical tools for ON characterization include enzyme-linked immunosorbent assays, reverse transcription polymerase chain reaction, liquid chromatography-high resolution mass spectrometry (LC-HRMS), LC-tandem mass spectrometry (MS/MS),[7] Sanger sequencing, and nuclear magnetic resonance spectroscopy.[8] The former techniques require specific antibodies and primer design/validation. In addition those techniques are not ideal for detection of metabolites and degradation products.[9] Thus, LC-MS-based analysis is an important tool for superior characterization specificity.

1.2 Oligonucleotide Chromatographic Separations

Ion-pair reversed phase (IP-RP) and, increasingly, hydrophilic interaction liquid chromatography (HILIC) chromatography are both used for separating and

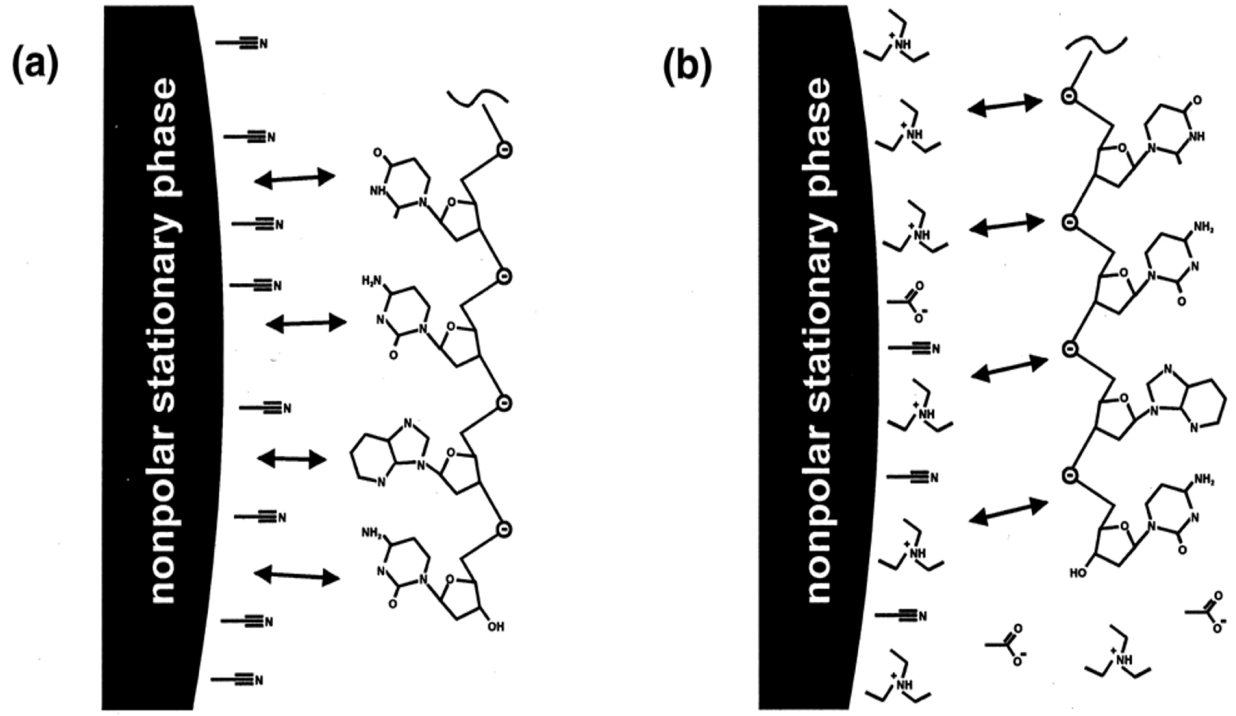


Figure 1.2 Schematic representation of the retention mechanisms in conventional (a) and ion pair (B) reverse phase liquid chromatography of oligonucleotides. Figure adapted from Huber et al.[15] (2001).

characterizing oligonucleotides from, e.g., solid state synthesis and reverse transcription.[10-12] The earliest credited account (1978) of oligonucleotide IP-RP LC used triethylammonium acetate (TEAA) as the ion pairing reagent. This approach has since been widely used and retention mechanisms have been developed, as shown in **Figure 1.2**.[9, 13-15] ONs can be retained in the absence of an ion pair (Figure 1.2 a); however, with the ion pairing reagent accumulating on the hydrophobic stationary phase, improved retention is observed between the charged portions of the IP reagent and the ON (Figure 1.2 b), resulting in improved chromatographic peak shapes. Further evolution of TEA as an ion pairing reagent has included substitution of the acetate counterion for bicarbonate (TEAB) for better ionization efficiency.[9, 16-18] A significant shift in preferred solvents for IP-RP LC-MS occurred with the introduction of hexafluoroisopropanol (HFIP) as counterion to TEA.[19] This approach showed improved

electrospray ionization (ESI) response and resulted in further exploration of a wide variety of fluoroalcohol counterions[10, 20-22] and ion pair reagents.[23] Retention mechanisms for oligonucleotide IP-RP LC have suggested that different types of nucleotides have differences in their contribution to overall retention.[24] Recently, a unified gradient theory was proposed for a more in depth understanding of IP-RP LC.[25, 26] One challenge recently emphasized in oligonucleotide RP-IP LC separation is the high degree of non-specific analyte adsorption to metal surfaces.[27] Ion mobility separation has recently been shown to allow resolution of isomeric nucleotides[28] and short oligonucleotides,[8] suggesting that a combination of IM and LC would be benifical prior to MS or MS/MS analysis.

1.3 Oligonucleotide Mass Spectrometry

Prior to the introduction of electrospray ionization, fast atom bombardment and electron ionization allowed MS analysis of chromatographically separated nucleotides and short oligonucleotides.[29] An early ESI experiment on a Sciex quadrupole mass spectrometer demonstrated analysis of a 4.3 kDa oligonucleotide.[30] The recent frontiers in mass spectrometry characterization of oligonucleotides are beginning to address larger[31-33] and heavily modified[33-35] RNA. These directions require high MS resolution/mass accuracy and improved MS/MS strategies, as well as bioinformatics solutions that can annotate highly complex spectra while accounting for fragment ion mass changes due to modifications.

Foundational bioinformatics work was performed by McCloskey and co-workers who developed software that could computationally sequence an oligonucleotide from its collision induced dissociation (CID) tandem mass spectrum, acquired following IP-RP

HPLC desalting with a TEAB-based mobile phase.[36] This early effort was expanded into the widely used MongoOligo website.[36] Recently, the importance of alternative tandem mass spectrometry techniques was reviewed in the context of characterizing synthetic single-stranded RNA.[37] Bioinformatics efforts have also been undertaken to facilitate the interpretation of tandem mass spectra from such alternative techniques, including electron detachment dissociation (EDD; LabView-based)[38], ultraviolet photodissociation and activated electron photodetachment dissociation (Python-based)[34], and negative electron transfer dissociation (C#-based).[35]

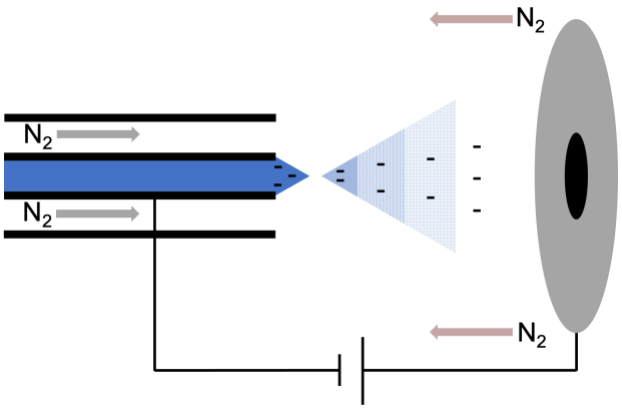


Figure 1.3 Schematic representation of electrospray ionization with LC eluent shown in dark blue inside a metal nebulizer. The mass spectrometer inlet orifice is shown in black with the gray area representing an endcap electrode.

1.4 Liquid Chromatography-Mass Spectrometry Instrumentation

Ongoing efforts to enhance the sensitivity and resolving power of mass spectrometry instrumentation have yielded closely correlated advances in oligonucleotide analysis.[39] Fourier transform ion cyclotron resonance (FT-ICR) mass analyzers have been a primary utility for characterization of nucleic acids due to their high resolution and multiple MS/MS capabilities.[40]

For oligonucleotide LC-MS, electrospray ionization is typically used in the negative ion mode. A schematic diagram of an example ESI source design is shown in **Figure 1.3**. Here, the LC eluent is directed through a grounded metal nebulizer positioned close to the mass spectrometer inlet, which is at a high positive potential with

respect to the nebulizer.[39] These high voltages are often generated by Cockroft-Walton like voltage multiplier circuits.[41] The resulting high electric field along with a nebulizing gas results in combined electrostatic/pneumatic nebulization of the liquid to form a plume of negatively charged droplets (indicated by the light blue cone in Figure 1.3). Rapid solvent evaporation, often assisted by heating, results in droplet shrinkage while retaining their charge. This process is accompanied by a sharp increase in droplet pH, which assists in generating high analyte charge states.[18] Droplet fission occurs at the Rayleigh limit where Coulomb repulsion overcomes surface tension.[42] Free oligonucleotide anions at a variety of charge states, depending on their mass, are generated after multiple droplet fission events.[42-45]

Following gas-phase oligonucleotide anion generation, these ion populations are often transmitted into the mass spectrometer vacuum[46] environment via an inlet capillary.[47] Multiinlet channel geometries have been proposed for improved transmission.[48] An example glass transfer capillary design is shown in **Figure 1.4**. In this design, the capillary outlets is encapsulated by a direct current electrode (capillary cap, in gray in Figure 1.4) that supports ion acceleration further into the mass spectrometer. This ESI source design is similar on the

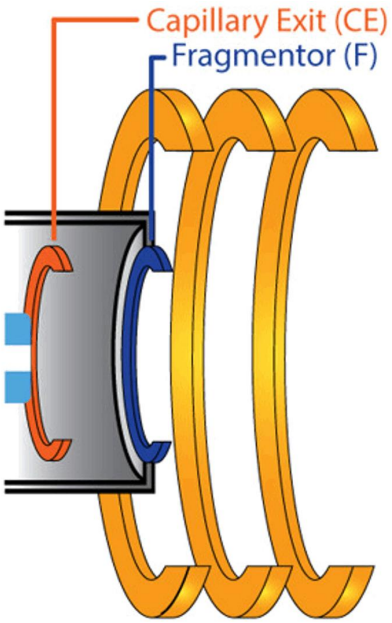


Figure 1.4 Mass spectrometer inlet glass capillary in continuity with capillary exit (light blue and orange) with additional lens (dark blue) present on the 6560c to support increased ion acceleration in this area. Ions enter an ion funnel system (gold) after exiting the inlet capillary Figure adapted from Gadkari et al. (2020).[51]

Bruker SolariX quadrupole (Q)-FT-ICR and Agilent 6560c IM-Q-TOF mass spectrometers used for the work presented in this thesis with one exception; an additional electrode element (capillary exit lens in Figure 1.4) has been implemented on the 6560c mass spectrometer to support collision induced unfolding in this region.[49, 50, 51]

Following the ion transfer capillary, both the Solarix Q-FT-ICR and Agilent 6560c IM-Q-TOF instruments use stacked ring electrodes with decreasing inner diameter, so called ion funnels, to radially focus and transmit ion packets further into the vacuum

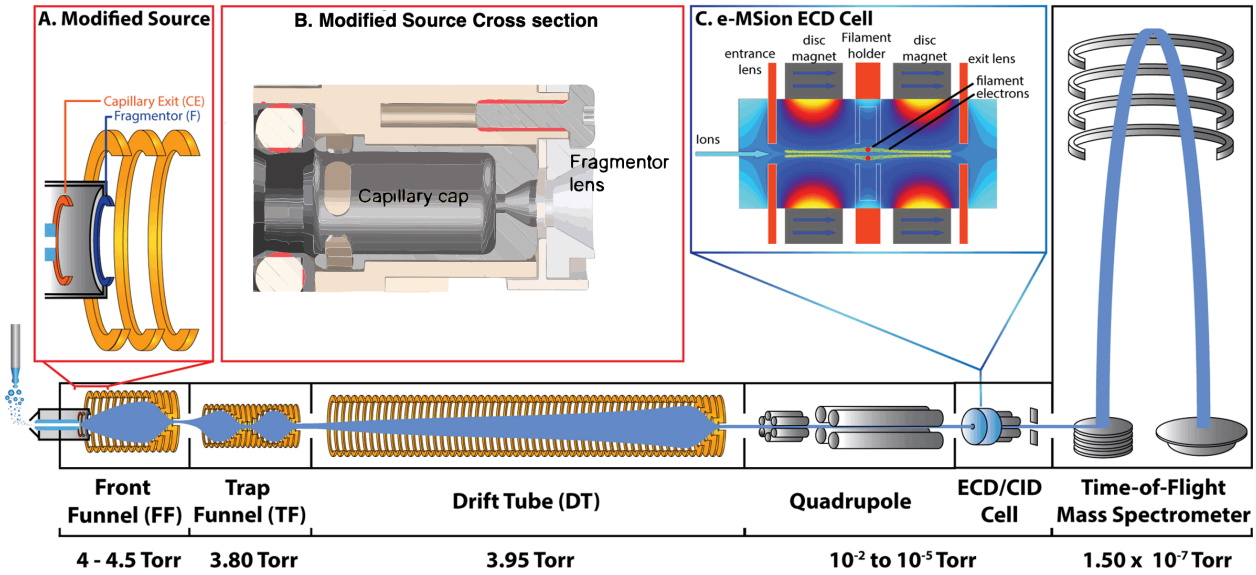


Figure 1.5 Schematic diagram of a 6560c mass spectrometer showing the ion path from the ESI source to the time of flight reflectron. Figure adapted from Gadkari et al. (2020)[51] and Kurulugama (2021)[52]

system. These ion funnels use both radio frequency and direct current voltage to accomplish this task.[53, 54] A schematic diagram of the Agilent 6560c is shown in **Figure 1.5**. This instrument utilizes a funnel trap system[55] following the first ion funnel. This trap funnel is followed by a stacked ring electrode system,[53] which can be operated as a drift tube for ion mobility applications.[56] The drift tube, in turn, is followed by a funnel and quadrupole mass filter for MS/MS precursor ion isolation.[57] For the MS/MS experiments performed in Chapter 2, the ExD cell was removed and the original hexapole collision

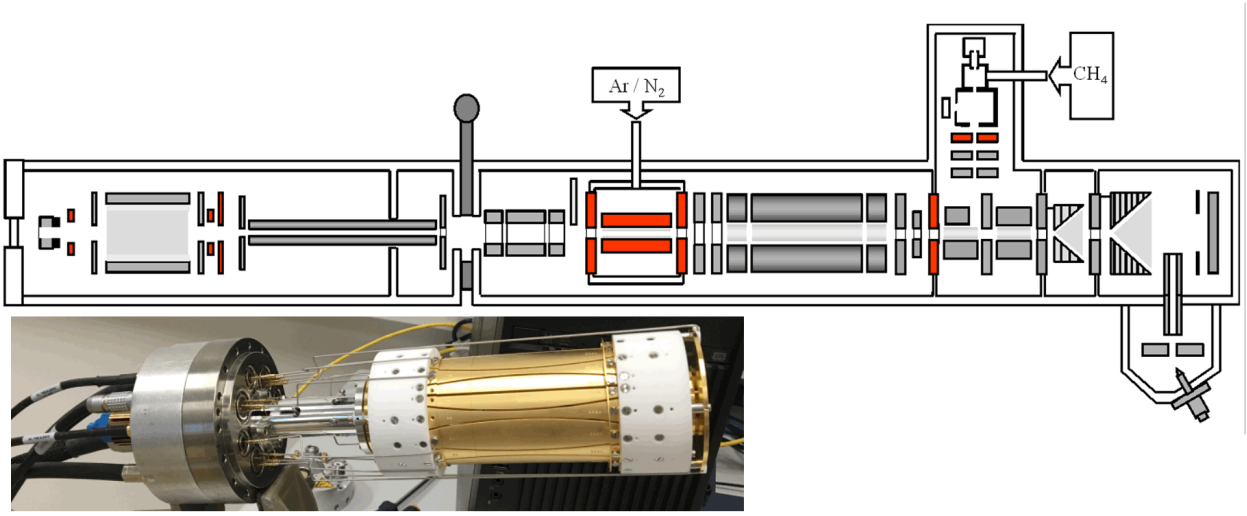


Figure 1.6 Schematic diagram of SolariX XR with ParaCell showing the path from ESI source to the ICR Trap. A image of the trap of a ParaCell is shown below the schematic. Figure was adapted from manuals.

cell was installed. Following MS/MS, an ion beam compression hexapole generates a more flat ion cloud, which is further shaped by an additional electrostatic slit prior to orthogonal acceleration and time-of-flight mass analysis with a microchannel plate detector.[58, 59]

A schematic diagram of the Bruker SolariX Q-FT-ICR instrument is shown in **Figure 1.6**. In this instrument, the inlet capillary is at a 90 degree angle with respect to the ion funnel system, which is followed by a segmented octupole for ion focusing and the possibility of introducing reagent ions for ion-ion reactions.[60] The SolariX quadrupole is followed by a hexapole[60] collision cell, which also serves as an ion trapping and accumulation device.[61] Taucher et. al have shown the potential for optimizing oligonucleotide fragmentation in such hexapole traps by adjusting the collision gas pressure.[62] Following the hexapole collision cell, a second, longer hexapole traverses the fringe field of the 7 Tesla superconducting magnet to efficiently transfer analyte ions into the dynamically harmonized[63, 64] ParaCell Penning trap[65] with a hollow cathode

electron dispenser.[66] Detection in an FT-ICR mass spectrometer occurs by amplification of the induced image current of the orbiting ions following radial excitation into coherent ion motion.[67] Application of a Fourier transform to the resulting time domain signal is typically performed in magnitude mode; however, with proper phasing (between 0 and 2π),[68] absorption and dispersion spectra can be generated from the real and imaginary components of the Fourier transform. While the signal phase is not typically known in an ion cyclotron resonance experiment, phasing algorithms and software support have promoted absorption mode processing.[69] With ad hoc phase correction becoming more accessible, the higher resolution of absorption mode spectra[68] can be realized (**Figure 1.7**).

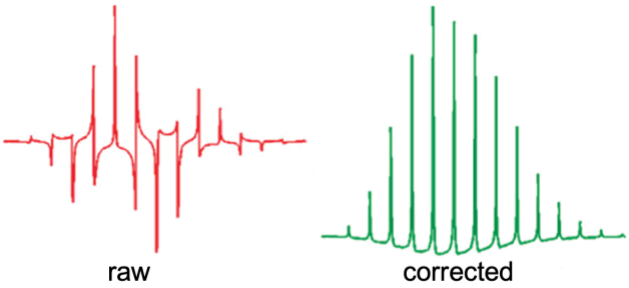


Figure 1.7 Phase correction of a signal in FT-ICR mass spectrometry signal processing. Representative unphased absorption mode signal with an asymmetric and partially negative baseline (Left, Red). Phase corrected signal (Right, green) shows a more symmetric baseline with high peak resolution. Figure adapted from Qi et al.[65] (2011).

1.5 Tandem Mass Spectrometry

Fragmentation of intact oligonucleotides above ~5kDa by tandem mass

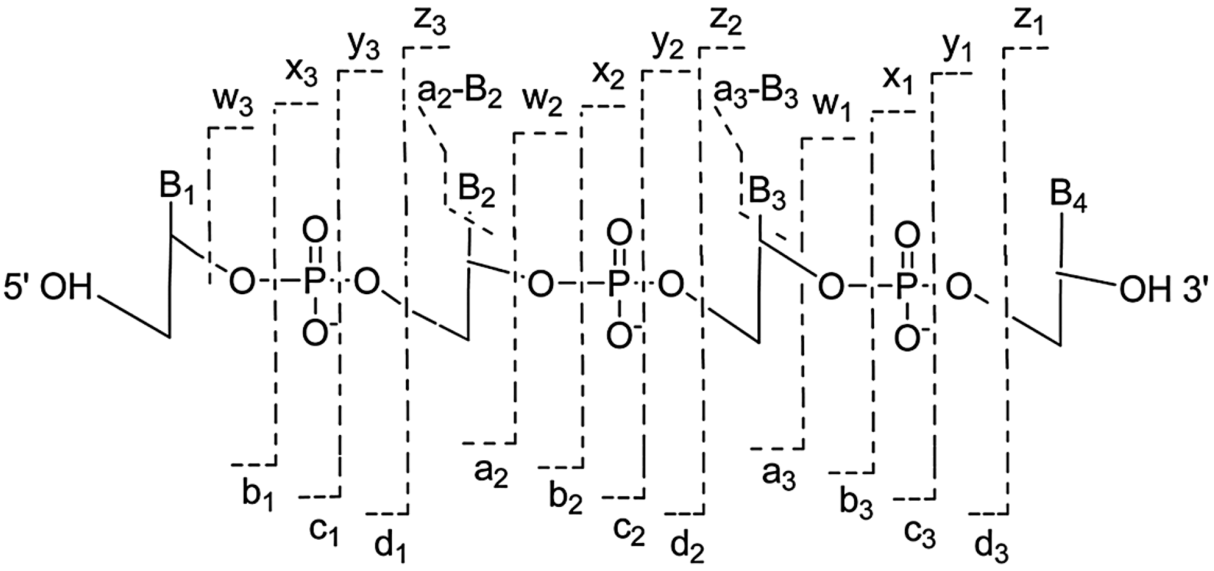


Figure 1.8 McLuckey fragment ion nomenclature for oligonucleotide tandem mass spectrometry with addition of the common nucleobase loss from 5' *a*-type fragments. Figure adapted from Brodbelt[66] (2014).

spectrometry is referred to as "top-down" analysis.[37] In such analysis, there are four possible backbone bonds that can cleave per nucleotide in the nucleic acid (see **Figure 1.8**).[70] In the McLuckey nomenclature[71], fragments containing the 5' end of the oligonucleotide are referred to as *a*, *b*, *c*, and *d* ions depending on which bond is cleaved, while fragments containing the 3' end are referred to as *w*, *x*, *y*, and *z*-type ions (Figure 1.8).

The most common MS/MS activation method is collision induced dissociation[71, 72] which majorly forms complementary (*a* – B) and *w*-type ions for DNA oligonucleotides and *c*/*y*-type ions for RNA oligonucleotides[71] (B refers to the nucleobase on the 5' side of the cleavage site). CID can be implemented in a variety of mass spectrometers and involves inelastic collisions[73-75] between analyte ions and a neutral gas such as nitrogen

or argon.[76] It has been shown that 2' modifications such as fluorination can prevent or significantly suppress backbone cleavage at adjacent backbone bonds in CID.[77] Infrared multiphoton dissociation (IRMPD)[78] is an alternative activation method that has been shown to be efficient for oligonucleotides due to the multiple phosphate chromophores at the typical IR wavelength of 10.6 μm. IRMPD can circumvent the loss of low mass ions that occur upon CID in ion traps[70] and is superior to CID inside an ICR cell[79] as no collision gas needs to be added, which deteriorates FT-ICR resolution. Radical-driven MS/MS methods such as electron detachment dissociation,[80-84] ultraviolet photodissociation,[70] and negative electron transfer dissociation[35, 77, 85, 86] have been shown to yield complementary fragmentation behavior. For example, EDD yields primarily *d/w*-type fragment ions.[84]

1.6 Dissertation Outline

This dissertation focuses on extending and improving current implementations and applications of liquid chromatography-mass spectrometry for therapeutic oligonucleotides. All experiments were performed with a mobile phase system free of fluoroalcohol additives, thus making workflows significantly more environmentally friendly. Additionally, the use of methanol instead of acetonitrile further reduces the environmental impact.

Chapter 2 describes fast chromatography to attain high quality precursor ion populations. The abundant high charge states acquired in these workflows suggest that instrument and material quality has improved since the introduction of triethylammonium bicarbonate IP-RP chromatography. The relatively high charge states obtained here are analytically more useful than the previously reported lower charge states,[14] e.g., more

efficient MS/MS can be achieved as well as improved FT-ICR MS detection. This Chapter also introduces OligoTap, a bioinformatics utility for automated annotation of MS/MS spectra from heavily modified oligonucleotides. The development of OligoTap was driven by the lack of appropriate open source software for this task. The increased analysis throughput afforded by Oligotap allows more thorough CID optimization for full sequence coverage of 22mer RNA, including heavily modiifed oligonucleotides.

Chapter 3 further evaluates OligoTap with an expanded search space containing isomeric, scrambled decoy sequences. A quantitative method is described for assessing false discovery rates (FDRs) in these analyses, which are based on the Bruker SNAP algorithm. These decoys show the possibility of false assignments in spite of high mass accuracy and resolution. Juxtaposition of LC-CID with LC-EDD data suggests that the latter activation strategy has a lower FDR. This chapter also acknowledges the current shortcomings of OligoTap in that it does not attempt to assign all possible fragmentation pathways, including radical fragment ions from EDD or internal fragments from CID.

Chapter 4 demonstrates the utility of the implemented mobile phase system for longer RNA sequences, up to 300 nt with FT-ICR MS detection. The reverse transcription polymerase chain reaction products analyzed here are of similar size as therapeutic single guide RNA. The moderate charge states generated with triethylammonium bicarbonate ion pairing reagent spread the analyte signal over a wider mass range than what would be expected with fluoroalcohol additives. This broader charge state distribution improves the input of the deconvolution algorithm and allows resolution of metal adducts and other small variations. To our knowledge, this is the largest RNA that has been analyzed with LC-FT-ICR MS to date.

Key conclusions from these chapters are summarized in Chapter 5, which also describes future directions.

1.7 References

1. Egli, M.; Manoharan, M. Chemistry, Structure and Function of Approved Oligonucleotide Therapeutics. *Nucleic Acids Res.* **2023,** *51*, 2529-2573.

2. Anguela, X. M.; High, K. A. Entering the Modern Era of Gene Therapy. *Annu. Rev. Med.* **2019,** *70*, 273-288.

3. Weber, H.; Khorana, H. G. Total Syhthesis of the Structural Gene for an Alanine Transfer Ribonucleic Acid from Yeast. Chemical Synthesis of Icosadeoxynucleotide Corresponding to the Nucleotide Sequence 21 to 40. *J. Mol. Biol.* **1972,** *72*, 219-249.

4. Hu, B.; Zhong, L.; Weng, Y.; Peng, L.; Huang, Y.; Zhao, Y.; Liang, X.-J. Therapeutic Sirna: State of the Art. *Signal Transduction Targeted Ther.* **2020,** *5*, 101.

5. Barbier, A. J.; Jiang, A. Y.; Zhang, P.; Wooster, R.; Anderson, D. G. The Clinical Progress of Mrna Vaccines and Immunotherapies. *Nat. Biotechnol.* **2022,** *40*, 840-854.

6. Sutton, J. M.; Kim, J.; Zahar, N. M. E.; Bartlett, M. G. Bioanalysis and Biotransformation of Oligonucleotide Therapeutics by Liquid Chromatography-Mass Spectrometry. *Mass Spectrom. Rev.* **2020,** *40*, 334-358.

7. Wang, L.; Meng, M.; Reuschel, S. Regulated Bioanalysis of Oligonucleotide Therapeutics and Biomarkers: Qpcr Versus Chromatographic Assays. *Bioanalysis* **2013,** *5*, 2747-2751.

8. Demelenne, A.; Nys, G.; Nix, C.; Fjeldsted, J. C.; Crommen, J.; Fillet, M. Separation of Phosphorothioated Oligonucleotide Diastereomers Using Multiplexed Drift Tube Ion Mobility Mass Spectrometry. *Anal. Chim. Acta* **2022,** *1191*, 339297.

9. Bartlett, M. G.; Chen, B.; McGinnis, A. C. Lc-Ms Bioanalysis of Oligonucleotides. In *Handbook of Lc-Ms Bioanalysis: Best Practices, Experimental Protocols, and Regulations*, LI, W.; Zhang, J.; Tse, F. L. S., Eds. John Wiley & Sons, Inc.: Hoboken, New Jersey, 2013; pp 607-627.

10. Sutton, J. M.; Guimaraes, G. J.; Annavarapu, V.; Dongen, W. D. v.; Bartlett, M. G. Current State of Oligonucleotide Characterization Using Liquid Chromatography–Mass Spectrometry: Insight into Critical Issues. *J. Am. Soc. Mass Spectrom.* **2020,** *31*, 1775-1782.

11. Demelenne, A.; Servais, A.-C.; Crommen, J.; Fillet, M. Analytical Techniques Currently Used in the Pharmaceutical Industry for the Quality Control of Rna-Based Therapeutics and Ongoing Developments. *J. Chromatogr. A* **2021,** *1651*, 462283.

12. Huang, M.; Xu, X.; Qiu, H.; Li, N. Analytical Characterization of DNA and Rna Oligonucleotides by Hydrophilic Interaction Liquid Chromatography-Tandem Mass Spectrometry. *J. Chromatogr. A* **2021,** *1648*, 462184.

13. Fritz, H.-J.; Belagaje, R.; Brown, E. L.; Fritz, R. H.; Jones, R. A.; Lees, R. G.; Khorana, H. G. High-Pressure Liquid Chromatography in Polynucleotide Synthesis. *Biochemistry* **1978,** *17*, 1257-1267.

14. Snyder, L. R.; Dolan, J. W. High-Performance Gradient Elution: The Practical Application of the Linear-Solvent-Strength Model. John Wiley & Sons: 2007.

15. Huber, C. G.; Oberacher, H. Analysis of Nucleic Acids by on-Line Liquid Chromatography–Mass Spectrometry. *Mass Spectrom. Rev.* **2001,** *20*, 310-343.

16. Huber, C. G.; Krajete, A. Analysis of Nucleic Acids by Capillary Ion-Pair Reversed-Phase Hplc Coupled to Negative-Ion Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **1999,** *71*, 3730-3739.

17. Sharma, V. K.; Glick, J.; Vouros, P. Reversed-Phase Ion-Pair Liquid Chromatography Electrospray Ionization Tandem Mass Spectrometry for Separation, Sequencing and Mapping of Sites of Basic Modifications of Isomeric Oligonucleotide Adducts Using Monolithic Column. *J. Chromatogr. A* **2012,** *1245*, 65-74.

18. Gong, L. Comparing Ion-Pairing Reagents and Counter Anions for Ion-Pair Reversed-Phase Liquid Chromatography/Electrospray Ionization Mass Spectrometry Analysis of Synthetic Oligonucleotides. *Rapid Commun. Mass Spectrom.* **2015,** *29*, 2402-2410.

19. Apffel, A.; Chakel, J. A.; Fisher, S.; Lichtenwalter, K.; Hancock, W. S. Analysis of Oligonucleotides by Hplc-Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **1997,** *69*, 1320-1325.

20. Santos, I. C.; Brodbelt, J. S. Recent Developments in the Characterization of Nucleic Acids by Liquid Chromatography, Capillary Electrophoresis, Ion Mobility, and Mass Spectrometry (2010-2020). *J. Sep. Sci.* **2021,** *44*, 340-372.

21. Basiri, B.; Hattum, H. v.; Dongen, W. D. v.; Murph, M. M.; Bartlet, M. G. The Role of Fluorinated Alcohols as Mobile Phase Modifiers for Lc-Ms Analysis of Oligonucleotides. *J. Am. Soc. Mass Spectrom.* **2017,** *28*, 190-199.

22. Liu, R.; Ruan, Y.; Zhongqiu; Gong, L. The Role of Fluoroalcohols as Counter Anions for Ion-Pairing Reversed-Phase Liquid Chromatography/High-Resolution

Electrospray Ionization Mass Spectrometry Analysis of Oligonucleotides. *Rapid Commun. Mass Spectrom.* **2019,** *33*, 697-709.

23. Li, N.; Zahar, N. M.; Saad, J. G.; Hage, E. R. E. v. d. Alkylamine Ion-Pairing Reagents and the Chromatographic Separation of Oligonucleotides. *J. Chromatogr. A* **2018,** *1580*, 110-119.

24. Gilar, M.; Fountain, K. J.; Budman, Y.; Neue, U. D.; Yardley, K. R.; Rainville, P. D.; II, R. J. R.; Gerbler, J. C. Ion-Pair Reversed-Phase High-Performance Liquid Chromatography Analysis of Oligonucleotides: Retention Prediction. *J. Chromatogr. A* **2002,** *958*, 167-182.

25. Enmark, M.; Samuelsson, J.; Fornstedt, T. Development of a Unified Gradient Theory for Ion-Pair Chromatography Using Oligonucleotide Separations as a Model Case. *J. Chromatogr. A* **2023,** *1691*, 463823.

26. Giddings, J. C. *Unified Separation Science*. Wiley: New York, 1991; Vol. New York. DOI:

27. Guimaraes, G. J.; Sutton, J. M.; Gilar, M.; Donegan, M.; Bartlett, M. G. Impact of Nonspecific Adsorption to Metal Surfaces in Ion Pair-Rp Lc-Ms Impurity Analysis of Oligonucleotides. *J. Pharm. Biomed. Anal.* **2022,** *208*, 114439.

28. Kenderdine, T.; Nemati, R.; Baker, A.; Palmer, M.; Ujma, J.; FitzGibbon, M.; Deng, L.; Royzen, M.; Langridge, J.; Fabris, D. High-Resolution Ion Mobility Spectrometry-Mass Spectrometry of Isomeric/Isobaric Ribonucleotide Variants. *J. Mass Spectrom.* **2020,** *55*, e4465.

29. Crain, P. F. Mass Spectrometric Techniques in Nucleic Acid Research. *Mass Spectrom. Rev.* **1990,** *9*, 501-575.

30. Covey, T. R.; Bonner, R. F.; Shushan, B. I.; Henion, J. The Determination of Protein, Oligonucleotide, and Peptide Molecular Weights by Ion-Spray Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **1988,** *2*, 249-256.

31. Wei, B.; Wang, J.; Cadang, L.; Goyon, A.; Chen, B.; Yang, F.; Zhang, K. Development of an Ion Pairing Reversed-Phase Liquid Chromatography-Mass Spectrometry Method for Characterization of Clustered Regularly Interspaced Short Palindromic Repeat Guide Ribonucleic Acid. *J. Chromatogr. A* **2022,** *1665*.

32. Kenderdine, T.; McIntyre, W.; Yassaghi, G.; Rollo, D.; Bunkowski, A.; Goerlach, L.; Suckau, D.; Tremintin, G.; Greig, M.; Bell, C.; Fabris, D. Integrating Internal Fragments in the Interpretation of Top-Down Sequencing Data of Larger Oligonucleotides. *J. Am. Soc. Mass Spectrom.* **2023**.

33. Crittenden, C. M.; Lanzillotti, M. B.; Chen, B. Top-Down Mass Spectrometry of Synthetic Single Guide Ribonucleic Acids Enabled by Facile Sample Clean-Up. *Anal. Chem.* **2023,** *95*, 3180-3186.

34. Santos, I. C.; Lanzillotti, M.; Shilov, I.; Basanta-Sanchez, M.; Roushan, A.; Lawler, R.; Tang, W.; Bern, M.; Brodbelt, J. S. Ultraviolet Photodissociation and Activated Electron Photodetachment Mass Spectrometry for Top-Down Sequencing of Modified Oligoribonucleotides. *J. Am. Soc. Mass Spectrom.* **2022,** *33*, 510-520.

35. Peters-Clarke, T. M.; Quan, Q.; Brademan, D. R.; Herbert, A. S.; Westphall, M. S.; Coon, J. J. Ribonucleic Acid Sequence Characterization by Negative Electron Transfer Dissociation Mass Spectrometry. *Anal. Chem.* **2020,** *92*, 4436-4444.

36. Ni, J.; Pomerantz, S. C.; Rozenski, J.; Zhang, Y.; McCloskey, J. A. Interpretation of Oligonucleotide Mass Spectra for Determination of Sequence Using Electrospray Ionization and Tandem Mass Spectrometry. *Anal. Chem.* **1996,** *68*, 1989-1999.

37. Hannauer, F.; Black, R.; Ray, A. D.; Stulz, E.; Langley, G. J.; Holman, S. W. Review of Fragmentation of Synthetic Single-Stranded Oligonucleotides by Tandem Mass Spectrometry from 2014 to 2022. *Rapid Commun. Mass Spectrom.* **2023,** *37*, e9596.

38. Karasawa, K.; Duchoslav, E.; Baba, T. Fast Electron Detachmentn Dissociation of Oligonucleotides in Electron-Nitrogen Plasma Stored in Magneto Radio-Frequency Ion Traps. *Anal. Chem.* **2022,** *94*, 15510-15517.

39. Hoffmann, E. d.; Stroobant, V. *Mass Spectrometry: Principles and Applications*. 3rd ed. ed.; John Wiley & Sons, Incorporated: 2007. DOI:

40. Breuker, K. Characterization of Ribonucleic Acids and Their Modifications by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. In *Nucleic Acids in the Gas Phase*, Gabelica, V., Ed. Springer, Berlin, Heidelberg: 2014; pp 185-202.

41. Patel, M. K.; Sahoo, H. K.; Ghanshyam, C. High Voltage Generation for Charging of Liquid Sprays in Air-Assisted Electrostatic Nozzel Systems. *IETE J. Res.* **2016,** *62*, 424-431.

42. Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **2013,** *85*, 2-9.

43. Liigand, P.; Kaupmees, K.; Haav, K.; Liigand, J.; Leito, I.; Girod, M.; Antoine, R.; Kruve, A. Think Negative: Finding the Best Eelctrospray Ionization/Ms Mode for Your Analyte. *Anal. Chem.* **2017,** *89*, 5665-5668.

44. Fenn, M. Y. J. B. Negative Ion Production with the Electrospray Ion Source. *J. Phys. Chem.* **1984,** *88*, 4671-4675.

45. Khristenko, N.; Amato, J.; Livet, S.; Pagano, B.; Randazzo, A.; Gabelica, V. Native Ion Mobility Mass Spectrometry: When Gas-Phase Ion Structure Depends on the Electrospray Charging Process. *J. Am. Soc. Mass Spectrom.* **2019,** *30*, 1069-1081.

46. Skoog, D. A.; Holler, F. J.; Crouch, S. R. *Principles of Instrumental Analysis*. Seventh edition. ed.; Cengage Learning: 2018. DOI:

47. Lin, B.; Sunner, J. Ion Transport by Viscous Gas Flow through Capillaries. *J. Am. Soc. Mass Spectrom.* **1994,** *5*, 873-885.

48. Kim, T.; Udseth, H. R.; Smith, R. D. Improved Ion Transmission from Atmospheric Pressure Ot High Vacuum Using a Multicapillary Inlet and Electrodynamic Ion Funnel Interface. *J. Am. Soc. Mass Spectrom.* **2000,** *72*, 5014-5019.

49. Zheng, X.; Kurulugama, R. T.; Laganowsky, A.; Russell, D. H. Collision-Induced Unfolding Studies of Proteins and Protein Complexes Using Drift Tube Ion Mobility-Mass Spectrometer. *Anal. Chem.* **2020,** *92*, 7218-7225.

50. Vallejo, D. D.; Polasky, D. A.; Kurulugama, R. T.; Eschweiler, J. D.; Fjeldsted, J. C.; Ruotolo, B. T. A Modified Drift Tube Ion Mobility-Mass Spectrometer for Charge-Multiplexed Collision-Induced Unfolding. *Anal. Chem.* **2019,** *91*, 8137-8146.

51. Gadkari, V. V.; Ramírez, C. R.; Vallejo, D. D.; Kurulugama, R. T.; Fjeldsted, J. C.; Ruotolo, B. T., Enhanced Collision Induced Unfolding and Electron Capture Dissociation of Native-like Protein Ions. *Anal. Chem.* **2020,** *92*, 15489-15496. DOI: 10.1021/acs.analchem.0c03372

52. Kurulugama, R. T. In *High-Energy In-Source Ion Activation Hardware*, 0th UM/Agilent Measurement Science Symposium, Online, January 20th, 2021.

53. Kelly, R. T.; Tolmachev, A. V.; Page, J. S.; Tang, K.; Smith, R. D. The Ion Funnel: Theory, Implementations, and Applications. *Mass Spectrom. Rev.* **2009,** *29*, 294-312.

54. May, J. C.; Goodwin, C. R.; Lareau, N. M.; Leaptrot, K. L.; Morris, C. B.; Kurulugama, R. T.; Mordehai, A.; Klein, C.; Barry, W.; Darland, E.; Overney, G.; Imatani, K.; Stafford, G. C.; Fjeldsted, J. C.; McLean, J. A. Conformational Ordering of Biomolecules in the Gas Phase: Nitrogen Collision Cross Sections Measured on a Prototype High Resolution Drift Tube Ion Mobility-Mass Spectrometer. *Anal. Chem.* **2014,** *86*, 2107-2116.

55. Ibrahim, Y.; Belov, M. E.; Tolmachev, A. V.; Prior, D. C.; Smith, R. D. Ion Funnel Trap Interface for Orthogonal Time-of-Flight Mass Spectrometry. *Anal. Chem.* **2007,** *79*, 7845-7852.

56. Stow, S. M.; Causon, T. J.; Zheng, X.; Kurulugama, R. T.; Mairinger, T.; May, J. C.; Rennie, E. E.; Baker, E. S.; Smith, R. D.; McLean, J. A.; Hann, S.; Fjeldsted, J. C. An Interlaboratory Evaluation of Drift Tube Ion Mobility-Mass Spectrometry Collision Cross Section Measurements. *Anal. Chem.* **2017,** *89*, 9048-9055.

57. Campana, J. E. Elementary Theory of the Quadrupole Mass Filter. *Int. J. Mass Spectrom. Ion Phys.* **1980,** *33*, 101-117.

58. Radionova, A.; Filippov, I.; Derrick, P. J. In Pursuit of Resolution in Time-of-Flight Mass Spectrometry: A Historical Perspective. *Mass Spectrom. Rev.* **2015,** *35*, 738-757.

59. Brais, C. J.; Ibañez, J. O.; Schwartz, A. J.; Ray, S. J. Revent Advances in Instrumental Approaches to Time-of Flight Mass Spectrometry. *Mass Spectrom. Rev.* **2020,** *40*, 647-669.

60. Kaplan, D. A.; Hartmer, R.; Speir, J. P.; Stoermer, C.; Gumerov, D.; Easterling, M. L.; Brekenfeld, A.; Kim, T.; Laukien, F.; Park, M. A. Electron Transfer Dissociation in the Hexapole Collision Cell of a Hybrid Quadrupole-Hexapole Fourier Transfrom Ion Cyclotron Resonance Mass Spectrometer. *Rapid Commun. Mass Spectrom.* **2008,** *22*, 271-278.

61. Douglas, D. J.; Frank, A. J.; Mao, D. Linear Ion Traps in Mass Spectrometry. *Mass Spectrom. Rev.* **2004,** *24*, 1-29.

62. Taucher, M.; Rieder, U.; Breuker, K. Minimizing Base Loss and Internal Fragmentation in Collisionally Activated Dissociation of Multiply Deprotonated Rna. *J. Am. Soc. Mass Spectrom.* **2011,** *21*, 278-285.

63. Boldin, I. A.; Nikolaev, E. N. Fourier Transform Ion Cyclotron Resonance Cell with Dynamic Harmonization of the Electric Field in the Whole Volume by Chaping of the Excitation and Detection Electrode Assembly. *Rapid Commun. Mass Spectrom.* **2011,** *25*, 122-126.

64. Guan, S. Legacy, Current Status, and Future Challenges of Fourier Transform Ion Cyclotron Resonace Mass Spectrometry (Fticrms). *Mass Spectrom. Rev.* **2021,** *41*, 115-157.

65. Brown, L. S.; Gabrielse, G. Geonium Theory: Physics of a Single Electron or Ion in a Penning Trap. *Rev. Mod. Phys.* **1986,** *58*, 233-311.

66. Moore, J. H.; Davis, C. C.; Coplan, M. A. *Building Scientific Apparatus: A Practical Guide to Design and Construction*. 2nd ed. ed.; Addison-Wesley: Redwood City, Calif., 1989; Vol. Redwood City, Calif. DOI:

67. Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: A Primer. *Mass Spectrom. Rev.* **1998,** *17*, 1-74.

68. Qi, Y.; Thompson, C. J.; Orden, S. L. V.; O'Connor, P. B. Phase Correction of Fourier Transform Ion Cyclotron Resonance Mass Spectra Using Matlab. *J. Am. Soc. Mass Spectrom.* **2011,** *22*, 138-147.

69. Silva, M. P. D.; Kaesler, J. M.; Reemtsma, T.; Lechtenfeld, O. J. Absorption Mode Spectral Processing Improves Data Quality of Natural Organic Matter Analysis by Fourier-Transform Ion Cyclotron Resonance Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2020,** *31*, 1615-1618.

70. Brodbelt, J. S. Photodissociation Mass Spectrometry: New Tools for Characterization of Biological Molecules. *Chem. Soc. Rev.* **2014,** *43*, 2757-2783.

71. McLuckey, S. A.; Berkel, G. J. V.; Glish, G. L. Tandem Mass Spectrometry of Small, Multiply Charged Oligonucleotides. *J. Am. Soc. Mass Spectrom.* **1992,** *3*, 60-70.

72. Harris, D. C. *Quantitative Chemical Analysis*. 8th ed. ed.; W.H. Freeman and Co.: New York, 2010; Vol. New York. DOI:

73. Mayer, P. M.; Poon, C. The Mechanisms of Collisional Activation of Ions in Mass Spectrometry. *Mass Spectrom. Rev.* **2009,** *28*, 608-639.

74. McQuarrie, D. A. *Statistical Mechanics*. Harper & Row: New York, 1975; Vol. New York. DOI:

75. McQuarrie, D. A.; Simon, J. D. *Physical Chemistry: A Molecular Approach*. University Science Books: Sausalito, Calif., 1997; Vol. Sausalito, Calif. DOI:

76. Bayat, P.; Lesage, D.; Cole, R. B. Tutorial: Ion Activation in Tandem Mass Spectrometry Using Ultra-High Resolution Instrumentation. *Mass Spectrom. Rev.* **2020,** *39*, 680-702.

77. Gao, Y.; Yang, J.; Cancilla, M. T.; Meng, F.; McLuckey, S. A. Top-Down Interrogation of Chemically Modified Oligonucleotides by Negative Electron Transfer and Collision Induced Dissociation. *Anal. Chem.* **2013,** *85*, 4713-4720.

78. Gardner, M. W.; Li, N.; Ellington, A. D.; Brodbelt, J. S. Infrared Multiphoton Dissociation of Small-Interfering Rna Anions and Cations. *J. Am. Soc. Mass Spectrom.* **2010,** *21*, 580-591.

79. Little, D. P.; Speir, J. P.; Senko, M. W.; O'Connor, P. B.; McLafferty, F. W. Infrared Multiphoton Dissociation of Large Multiply Charged Ions for Biomolecule Sequencing. *Anal. Chem.* **1994,** *66*, 2809-2815.

80. Yang, J.; Mo, J.; Adamson, J. T.; Hakansson, K. Characterization of Oligodeoxynucleotides by Electron Detachment Dissociation Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2005,** *77*, 1876-1882.

81. Yang, J.; Håkansson, K. Fragmentation of Oligoribonucleotides from Gas-Phase Ion-Electron Reactions. *J. Am. Soc. Mass Spectrom.* **2006,** *17*, 1369-1375.

82. Mo, J.; Håkansson, K. Characterization of Nucleic Acid Higher Order Structure by High-Resolution Tandem Mass Spectrometry. *Anal. Bioanal. Chem.* **2006,** *386*, 675-681.

83. Kinet, C.; Gabelica, V.; Balbeur, D.; Pauw, E. D. Electron Detachment Dissociation (Edd) Pathways in Oligonucleotides. *Int. J. Mass Spectrom.* **2009,** *283*, 206-213.

84. Taucher, M.; Breuker, K. Top-Down Mass Spectrometry for Sequencing of Larger (up to 61 Nt) Rna by Cad and Edd. *J. Am. Soc. Mass Spectrom.* **2010,** *21*, 918-929.

85. Gao, Y.; McLuckey, S. A. Electron Transfer Followed by Collision-Induced Dissociation (Net-Cid) for Generating Sequence Information from Backbone-Modified Oligonucleotide Anions. *Rapid Commun. Mass Spectrom.* **2013,** *27*, 249-257.

86. Calderisi, G.; Glasner, H.; Breuker, K. Radical Transfer Dissociation for De Novo Characterization of Modified Ribonucleic Acids by Mass Spectrometry. *Angew. Chem., Int. Ed.* **2020,** *59*, 4309-4313.

Chapter 2 A Fluoroalcohol-Free LC-MS/MS Workflow with Automated Spectral Annotation to Achieve Complete Sequence Coverage for Synthetically Modified Oligonucleotides

2.1 Introduction

Oligonucleotides (ONs) of growing complexity are increasingly being incorporated into medical treatments.[1] ON-based therapeutics have been approved for treatments of cancers,[2] diseases of the eye, liver, skeletal muscle, and spinal cord,[3] as well as Sars-CoV-2[4] and have become the third major drug platform after small molecules and antibodies.[5] Synthetic ON chemical modifications for increased specificity, stability, and deliverability are often part of the drug design.[1] Such modifications may be incorporated at the sugar, nucleobase, or phosphate backbone portions of the ON with new chemical motifs continuing to be explored.[6] O-methyl, O-methoxyethyl, and fluoro substitutions at the ribose 2' position are example modifications already present in FDA-approved drugs.[1] Sugar modifications can provide greater resistance to nuclease degradation with 2'-fluoro modification also yielding improved binding affinity.[7] For example, 2'-fluoro modifications have been incorporated at the 9, 10, and 11 positions of siRNA.[7] Other chemical modifications in various stages of clinical trials or exploration include *N*-acetylgalactosamine conjugation, locked nucleic acids, and glycol nucleic acids.[7] Thus, comprehensive ON characterization techniques are required for successful therapeutic implementation.

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) fills an important role in the development, characterization, and quality control of therapeutic ONs.[8] Challenges in ON LC-MS/MS continue to be met with new technology,[9] including buffer systems[9] and column chemistry,[10] for improved chromatographic separation, tubing passivation treatments for improved detection limits,[11] and more versatile bioinformatics solutions.[12-15] In the latter area, previous data analysis utilities[12-21] (see **Table 1**) have excelled at characterizing biological ON modifications; however, the open-source space addressing synthetic, particularly 2' modifications, is limited. For example, recent initiatives, including the development of NASE,[13] NATMS,[15] and Pytheas,[14] allow for automated annotation of oligonucleotide MS/MS spectra considering a plethora of naturally occurring RNA modifications. Nevertheless, no existing data analysis utility, to our knowledge, is platform independent, i.e., allowing data formats from any instrument vendor, or able to readily incorporate novel synthetic modifications, including prediction of potentially altered fragmentation pathways. It has been reported that introduction of *O*-methyl, *O*-methoxyethyl, and fluoro substitutions at sugar 2' positions disfavors the preferred RNA backbone cleavage on the 3' side of phosphorus to yield 5' *c*-type and 3' *y*-type fragment ions upon collision induced dissociation/higher energy collision dissociation (CID/HCD). Instead cleavage of all 4 backbone positions to yield a more even distribution of complementary *a*/*w*, *b*/*x*, *c*/*y*, and *d*/*z* fragment ions occurs. [22] Thus, all fragment types should be considered when interpreting CID/HCD MS/MS spectra of 2'-modified ONs.

A wide landscape of LC solvent systems has emerged focused on ON selectivity and sensitivity based on their physiochemical properties.[23] The gold standard for ion-

pairing reverse-phase (IP-RP) LC/MS/MS includes triethylamine (TEA)/hexafluoroisopropanol (HFIP) additives but many amines and fluoroalcohols have been explored.[23, 24] One predecessor to the amine/fluorinated alcohol combinatorics is triethylammonium bicarbonate (TEAB).[23, 25, 26] Alkylammonium bicarbonate-type buffers, however, remain eclipsed by alkylammonium acetate and amine/fluoroalcohol solvent systems[25, 27] despite showing comparable chromatographic resolution.[9, 27, 28] With recent focus on the toxicity of poly- and perfluorinated compounds[29] as well as detection of HFIP with low limits of quantitation in multiple ground water samples,[30] fluoroalcohol-free solvent systems are desired. Furthermore, the quality of amines/fluorinated alcohols can have a significant effect on adduct formation in IP-RP workflows.[31]

For comprehensive therapeutic ON structural characterization, complete sequence coverage, i.e., backbone cleavage between all constituent nucleotides, is needed. While CID/HCD is the most efficient activation strategy for MS/MS, fragmentation probability varies greatly across an ON sequence and is particularly low adjacent to internal fluorinated sugars.[32, 33] Thus, to optimize sequence coverage from CID/HCD, a number of collision energies and precursor ion charge states need to be explored and combined, putting further strain on the throughput of data analysis methods. Here, we introduce OligoTap, a flexible, platform-independent bioinformatics utility for rapid, automated analysis of spectra from fluoroalcohol-free LC-MS/MS of heavily modified and unmodified ONs.

Table 2.1 Oligonucleotide bioinformatics utilities for LC/MS/MS data. First column: name or affiliation of the utility. Second column: versatility and applications of the utility. Third column: features that limit the utility's application to therapeutic RNA research.

Utility	Strengths	Challenges
SOS[16]	Biological modifications	Instrument compatibility, synthetic modifications

Adriane[17]	Biological modifications	Backbone modifications and search space
OMA and OPA[18]	High versatility	Instrument compatibility
Merck Software[12]	High versatility	Availability and automated annotation
RoboOligo[19]	High versatility	Sequence length
RNAModMapper[20]	Biological modifications	Search space, synthetic modifications
NASE[13]	Biological modifications, scoring	Instrument compatibility, synthetic modifications
Aom2s[21]	Biological modifications	Instrument compatibility, synthetic modification
NATMS[15]	Synthetic modifications, visualization, cross-ring	Instrument compatibility, limited synthetic modifications
Pytheas[14]	Biological modifications, scoring	Search space, synthetic modifications, sequence length
Thermo	High Versatility	Cost
Agilent	High Versatility	Cost
Bruker	High Versatility	Cost
Protein Metrics	High Versatility	Cost

2.2 Experimental

2.2.1 Materials

Riedel-de Haen TraceSELECT water (<5 ppb Na and <5 ppb K) in a polypropylene cubic container was purchased from Thomas Scientific (Swedesboro, NJ). Optima LC-MS grade water (≤20 ppb Na and ≤10 ppb K) and methanol (≤50 ppb Na and ≤10 ppb K) as well as Andwin 70% isopropyl alcohol (IPA) were purchased from ThermoFisher (Pittsburgh, PA). Supelco LiChropur TEAB was acquired from MilliporeSigma (St Louis, MO). RNA sequences; modified sense oligonucleotide (SO) (rU*rC*rA mGmAmA mGrArA rGrG/i2FU/ /i2FA//i2FA/rC rGrA/i2MOErG/ rUrA*rG*rG), modified antisense oligonucleotide (ASO;rC*cC*mU mAmCmU rCrGrU /i2FU//i2FA//i2FC/ rCrUrU mCmUrU /i2MOErC/rU*rG* rA), and unmodified SO (rUrCrA

rGrArA rGrArA rGrGrU rArArC rGrArG rUrArG rG) were from Integrated DNA

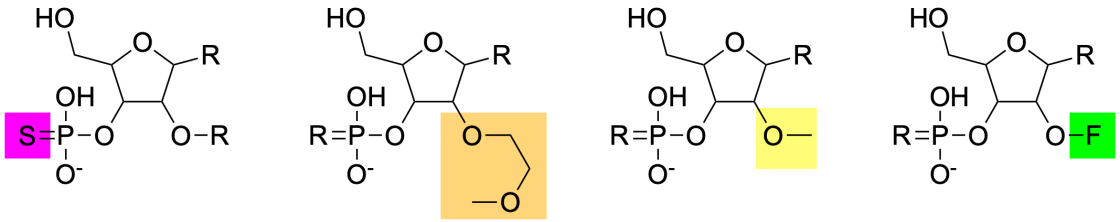


Figure 2.1 Structures of oligonucleotide modifications used in therapeutic small interfering RNA. Highlighted are the structures of phosphorothioate at the backbone, O-methyl, O-methoxyethyl, and fluoro modifications on the 2' position of the ribose, colored in magenta, cantaloupe, banana, and green, respectively.

technologies (IDT, Coralville, IA) and were shipped in their dry states. The chemical structures of the modifications are shown in **Figure 2.1**. The SO and ASO were HPLC purified at IDT, whereas the unmodified SO was subjected to standard desalting.

2.2.2 Sample Preparation

Bench space was cleaned with 70% IPA prior to opening polypropylene tube bags (Eppendorf conical sterile 5 mL and protein LoBind microcentrifuge tubes, ThermoFisher). All aliquot tubes were closed using fresh powder free nitrile exam gloves (Dot Scientific, Bourton, MI). The plastic valve spout of the TraceSelect water was purged with 5 mL prior to dispensing 5 mL water into a 5 mL conical tube. The dispensed 5 mL water was cooled to 4 ℃. x. Lyophilized oligonucleotide samples were microcentrifuged at 6,000 RPM for 30 seconds prior to opening to control for material shift during shipping. Oligonucleotides were then solubilized in the IDT tubes to a concentration of 1 mM for the unmodified SO and 0.5 mM for the modified ASO and SO by adding 250 µL and 500 µL TraceSelect water, respectively. IDT tubes were vortexed with a vortex genie-2 (Scientific Industries, Bohemia, NY) on the analog dial setting of 10 for 30-120 seconds and aliquoted into ten 25 µL or 50 µL volume aliquots,

respectively, frozen, and stored at -80 ℃. Prior to LC-MS analysis, samples were thawed, optionally diluted in water, transferred to polypropylene 0.25 mL snap cap autosampler vials with polytetrafluoroethylene caps and rubber septa (ThermoFisher) and placed in an autosampler set to 5 ℃. Vials were placed in an Agilent (Santa Clara, CA) 1100/1200 stack.

2.2.3 Liquid Chromatography

Solvent reservoir bottlehead assembly for Infinity II systems with GL45 thread cap and borosilicate frit inlet filter (Agilent, Santa Clara, CA) was used for solvent delivery to the degasser. Mobile phase A consisted of 5 mM TEAB and mobile phase B consisted of LC-MS grade methanol. Mobile phases were prepared by pouring LC-MS grade solvents into tared Corning polypropylene GL45 bottles (ThermoFisher) using an Ohaus (Parsippany, NJ) HH 320 scale calibrated with an Ohaus 200 g calibration weight. Mass of mobile phase was converted to volume at ambient conditions. TEAB was stored at 4 ℃, poured into polypropylene tubes and pipetted into the mobile phase. The LC flow path is shown in Appendix A Figure A.1 and consisted of 0.005” (red) or 0.0025” (natural) inner diameter polyether ether keytone (PEEK) tubing (IDEX, Middleboro, MA) and a stainless steel 1/16” tee with 0.25 mm through hole (Vici, Houston, TX. Gradient delay was reduced by bypassing the mixer and using a 0.005” stainless steel capillary at the binary pump and a 0.007” stainless steel capillary at the autosampler instead of calibrated tubing (Agilent). The autosampler tubing change cannot be accounted for in the system configuration to our knowledge, which induces the possibility of contaminating the metering device and causing carryover. We were unable to implement valve controlled automatic delay reduction on the autosampler,

thus the entire autosampler flow path was included in the gradient method. A 0.5 µm PEEK precolumn frit (IDEX) was placed between the autosampler and the column compartment. A six-port EHMA valve (Vici) was mounted in the flow path to direct unwanted flow to waste. Electrical connections were soldered to the valve's manual controller circuit board at the position button contacts. The wire leads were terminated to a high-density d-sub connector (Winford Engineering, Auburn, MI) and connected to a BCD/external contact board (Agilent) installed at the 1100 autosampler and controlled in the method. LC eluent was interfaced to the mass spectrometers with ES tested nebulizers (Agilent). Red PEEK tubing was replaced with natural PEEK tubing, the column compartment valve and the UV detector were bypassed for the high pH (HPH)-C18 method. An Agilent polystyrene divinylbenzene, PLRP-S 2.1x50 mm, 3 µm, 100 Å at 80 ℃ or HPH-C18 2.1x50 mm, 1.9 µm, 100 Å at 60 ℃ column was used at 0.2 and 0.1 mL/min, respectively. Activation energy optimization for the 4- and 9- charge state of each oligonucleotide was performed on the Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer (vide infra) with injection volumes of 3 µL, containing 300 pmol for the unmodified SO, 270 pmol for the ASO, and 150 pmol for the SO. For subsequent experiments on additional charge states, 2 µL injections containing 360 pmol for the unmodified sense and 180 pmol for the SO and ASO were used. LC MS/MS experiments on the 6560c ion mobility (IM)-Q-TOF instrument used 2 µL injections of stock solutions containing 2 nmol for the unmodified SO and 1 nmol for the SO and ASO. Linear gradients from 5% to 100% B over 2.5 min and 5 min at 0.2 mL/min and 0.1 mL/min were used for the PLRP-S and HPH-C18 columns, respectively. To manage run time a flow ramp from 0.1 mL/min to 0.3 mL/min at 5.5 minutes and

back down to 0.1 mL/min at 16.25 minutes during cleaning and re-equilibration was used for the HPH-C18 method. Column eluent was directed to the nebulizer of the mass spectrometer. Method reports for the LC-FT-ICR PLRP-S Appendix A Figure A.2 and LC-Q-TOF HPH-C18 Appendix A Figure A.3, respectively.

2.2.4 Mass Spectrometry

Three mass spectrometry platforms were used: a Q-Exactive Orbitrap hybrid instrument (ThermoFisher, San Jose, CA), a 6560c IM-Q-TOF (Agilent), and a SolariX Q-FT-ICR (Bruker, Billerica, MA). The metal spray shield and capillary cap components of the FT-ICR and IM-Q-TOF instruments were abrasively cleaned as needed with 8000 grit abrasive micromesh (Micro Surface, Wilton, IA) and subjected to manufacturer recommended cleaning subsequently. Resistive glass capillaries were cleaned as needed by pulling a small amount of fiber through the capillary with a nickel capillary cleaning wire (Agilent; typically not recommended for resistive capillaries), after which cleaning proceeded according to manufacturer recommendations. ES tested nebulizers were used on the SolariX and 6560c with a flush needle adjustment distance. All instruments were operated in negative ion mode. SolariX and 6560c capillary voltages were set to 3.5 kV. The SolariX source was operated at -500 V, 3 bar, 10 L/min, and 350 ℃ for the endplate offset, nebulizing pressure, dry gas flow, and dry gas temperature respectively. The 6560c source was operated at 2000 V, 35 psig, 12 L/min, and 325 ℃ for the nozzle, nebulizing pressure, dry gas flow, and dry gas temperature, respectively. Additionally, the sheath gas flow rate and temperature were 12 L/min and 350 ℃, respectively. Acquisition method reports for LC-FT-ICR Appendix A Figure A.4 and LC-IM-Q-TOF Appendix A Figure A.3 are included as Supporting Information (for

the latter instrument, LC and MS parameters are in the same report). SolariX was transmission tuned by direct infusion of an LC purified oligonucleotide via red PEEK tubing and a 22s gauge blunt cemented needle syringe (Hamilton, Reno, NV). The 6560c, operated in Q-TOF only mode, was transmission tuned by the SWARM direct infusion algorithms in the 3,200 m/z range, extended dynamic range, and slicer in high resolution position. The SolariX was set to 100-3,200 m/z range with single scans of 2 mega word datapoints and default FT processing settings for one omega detection. The 6560c MS scans were acquired at a rate of 2 Hz and the MS/MS scans were acquired at 1 Hz. The SolariX scan rate was approximately 0.9 Hz, including 4 microaccumulations during the 0.8 second accumulate during detect time in LC-capture mode. A static quadrupole isolation window of 20 m/z was used for SolariX LC-MS/MS experiments. On the 6560c a 2 minute delta retention time around the expected elution time was used with a wide quadrupole isolation window of 9 m/z. All retention times were between 4.3 and 5 minutes for the HPH-C18 method and between 2.6 and 3 minutes for the PLRP-S method. Target m/z values were manually added to acquisition methods for all oligonucleotides. These settings resulted in ~8-12 and ~12-20 MS/MS scans per target elution window for the SolariX and 6560c, respectively. One and two precursor ion charge states were fragmented per LC run for the SolariX and 6560c, respectively. Including the interwoven MS^1 scans, the data acquisition frequency of the 6560c was ~2.5-fold higher than that of the SolariX. CID energy optimization was performed on the SolariX for all oligonucleotides at 5, 10, and 15 V for the 9- charge state and at 25, 30, and 35 V for the 4- charge state (a total of 18 LC-MS/MS runs). Subsequent experiments on both the SolariX and the 6560c were performed for all

oligonucleotides at 4, 7, 10, 13, 16, 19, 22, and 25 V CID for the 11-, 10-, 9-, 8-, 7-, 6-, 5-, and 4- charge states, respectively (a total of 24 LC-MS/MS runs).

2.2.5 Data Analysis

Prior to LC/MS/MS, the mass spectrometers were externally calibrated. The 6560c and SolariX were calibrated with ESI-L low concentration tuning mix (Agilent) diluted ten-fold without the m/z 321 ion spike-in. The 6560c was calibrated to default SWARM parameters in the calibrate/check tune algorithm using the calibrant delivery system for tune mix infusion, whereas tune mix was directly infused from a syringe for the SolariX. Calibrant peaks were found within a 20 ppm window from m/z 302 to 2,284 using 9 calibrant ions and a linear calibration function, resulting in a root mean square error of 0.8 ppm.

For Bruker data, DataAnalysis 5.0 SR1 was used in three steps for each chromatogram: 1) MS^n level extracted ion chromatograms (EICs) were generated and the resulting chromatographic peaks were identified by the 3.0 compound detection algorithm. The corresponding compound MS/MS spectra were averaged and a monoisotopic MS/MS peak list was generated with the SNAP[34] algorithm over the 150-2,950 m/z range. 2) This peak list was internally recalibrated and 3) exported as Mascot generic file format (MGF). The SNAP[34] isotopic cluster finder for monoisotopic peak detection was used with a Q factor of 0.9 and a signal-to-noise (S/N) ratio of 3. For averagine-type[35, 36] of RNAs, an average repeat unit of $C_{9.5}H_{14.75}N_{3.75}O_8P_1$ was used, determined by adding an oxygen and a hydrogen to the Bruker nucleotide building block chemical formula editor, summing each element of the four nucleotides, and dividing each of the summed elements by four. This approach is slightly different from the more

accurate, previously published, RNA average repeat unit of $C_{9.5}H_{11.75}N_{3.75}O_8P_1$.[37] The compound detection algorithm typically found one compound per LC/MS/MS data file (with single oligonucleotide injection) with the exception of the unmodified SO, which was not HPLC purified at IDT). This oligonucleotide showed chromatographically resolved, low level impurities that were detected and integrated. Such impurities were removed prior to downstream processing. Subsequent, linear internal calibration was based on 3-18 typical *c*-type fragment ions from CID/HCD in a batch processing format. Following internal recalibration, a subsequent Data Analysis method was run to conserve compounds and calibration but update peak lists with a lower SNAP Q factor of 0.5. Peak assignments were made with OligoTap, using the SNAP isotope handling option, and an error of $\leq$10 ppm.

For Agilent data, the Qualitative Analysis Navigator B.10 was used with three actions: 1) MS/MS chromatograms were generated for target precursor ion charge states as product ion total ion chromatograms, 2a) the product ion total ion chromatogram was integrated using the Agile 2 integrator, and 2b) two spectra (one for each charge state targeted in the LC-MS/MS experiment) were extracted by averaging the spectra of the chromatographic peak (one peak per charge state). 3) Subsequently, an unbiased isotope model was used for isotopically resolved deconvolution of the two product ion spectra to convert all observed isotopic clusters to their corresponding neutral forms. These charge deconvoluted spectra were exported as MGF. The chromatogram extraction action was populated with 24 total ion chromatograms (8 charge states each for the three oligonucleotides) with the ‘integrate when extracted’ option selected. Integrated results were subjected to MS/MS spectral extraction

selecting scans at greater than 80% chromatographic peak height. Spectral peak identification used a maximum spike width of 2 points and a required valley of 0.70 between peaks. The unbiased isotope parameters included charge states from -1 to -11 as assignment options and a default of assigning single peak features as singly charged. Resolved isotope deconvolution used the same peak identification settings. The top 1,500 centroided, decharged MS/MS peaks were selected for export to MGF files. Peak assignments were made by OligoTap, using the isotope resolved deconvolution option (vide infra), with an error of $\leq$10 ppm.

2.3 Results and Discussion

2.3.1 OligoTap Design and Implementation

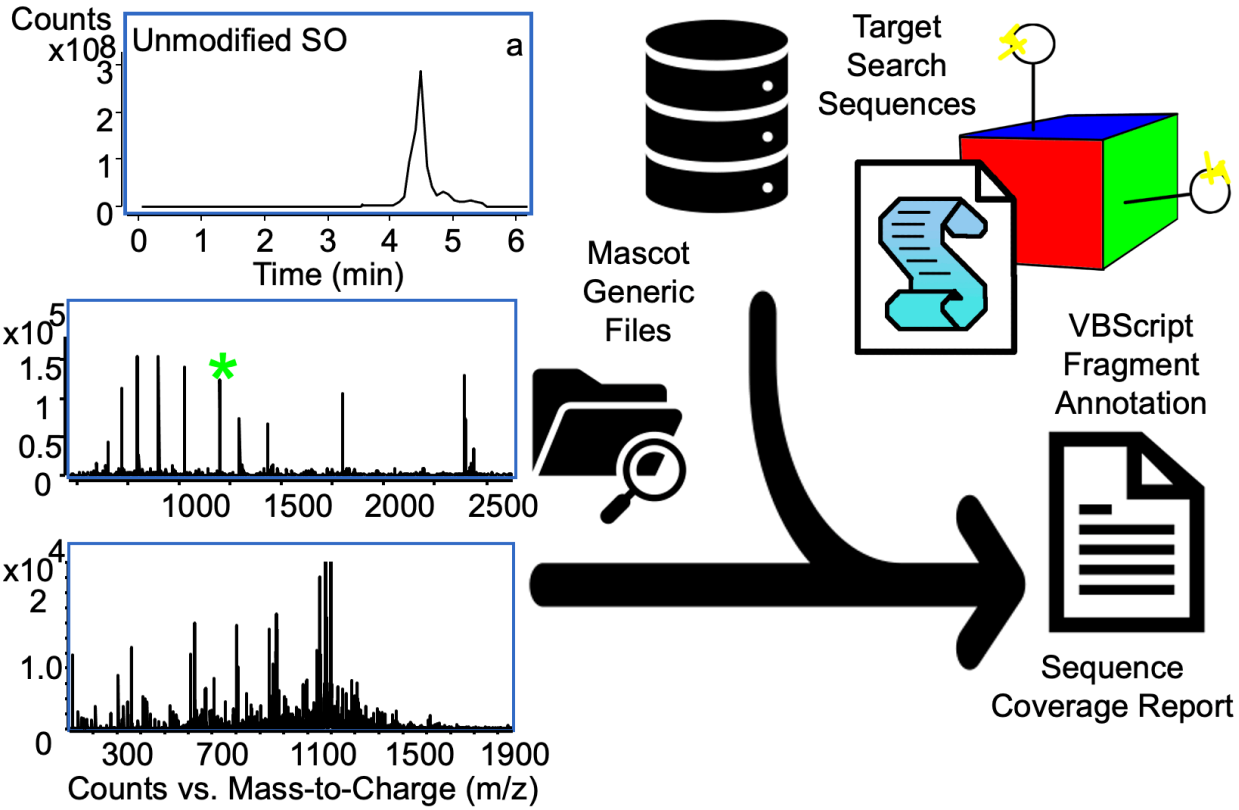


Figure 2.2: The OligoTap general workflow. LC-MS/MS data (a-c) are collected and exported to MGF format. A text file with target sequences is also generated. The Microsoft visual basic script then generates a list of theoretical backbone fragment ions for these sequences and searches the MGF files for m/z values matching these fragment ion lists. The script generates a report of matching fragments.

An overview of the OligoTap design is shown in **Figure 2.1**. In order to use this utility, .MGF files need to be generated for acquired LC-MS/MS spectra. Despite MGF being a common MS/MS data format syntax, its generation varies between instrument vendors. Example MGF files from DataAnalysis (Bruker), Qualitative Analysis (Agilent), and Proteome Discoverer (ThermoFisher) are included in the Supporting Information. All files contain m/z and abundance information for each observed peak but the differentiator between columns (e.g., space or tab character) is different. The Bruker MGF format also contains charge state information following averagine-type fitting of isotopic distributions. In addition, different numbers of digits are included for each number. These discrepancies were addressed in OligoTap by implementing an iteration

and indexing algorithm which views each character and checks whether it is an ASCII zero through nine or a period. Characters other than these are treated as delimiters between the indexed contiguous numerical values. Those contiguous series are indexed as m/z, abundance, and charge state, constituting a custom line split function. Extraction and processing from the .MGF formats was programed in OligoTap to make the spectral annotation approximately linear in time, i.e., files with a high number of lines are processed without becoming exponentially slower, by first indexing the file (vide supra) and spectral lines of interest (e.g.. 'Begin Ions' and 'End Ions') without writing all the file lines to random access memory. The identified lines are then referenced to focus the file reading commands to a single spectrum at a time. For OligoTap analysis, MGF files are placed as the only files in a directory of a Windows computer. File names cannot have extra periods or other special characters.

OligoTap also requires a text file containing targeted sequences. We chose a format with delimiters to separate nucleotides with a slash ("/"), sugars with a comma (","), nucleobases with a period ("."), termini by a hyphen ("-"), and oligonucleotide name with an equal sign ("=") on individual lines of the text file. Nucleobase options are adenine (A), cytosine (C), methylcytosine (5C), guanine (G), thymine (T), and uracil (U). Sugar options are ribose (r), deoxyribose (d), 2'-fluoro sugar (f), 2'-methoxy sugar (rm), and 2'-methoxyethoxy sugar (rmoe). Backbone linker options are phosphate (p) and phosphorothioate (s). A separate auto-formatting script was developed for conventional unmodified nucleotides. For example, the sequence GCU translates to HO-r,G.p/r,C.p/r,U-OH in this notation. The format file for our three target ONs (modified and unmodified SO, modified ASO) is shown in Appendix A Table A.1. An OligoTap output

directory can either be generated beforehand or via the OligoTap prompts. As shown in Appendix A Figure A.4, these prompts appear in the following order: 1) prompt for an output directory, 2) prompt for a search sequences file (i.e., the type of text file described above), 3) prompt for a directory with .MGF files, 4) prompt for a precursor ion m/z tolerance, 5) prompt for a fragment ion m/z error (in ppm), 6) prompt for maximum precursor ion charge state to search for, 7) prompt for maximum fragment ion charge state to search for, and 8) prompt for an isotope cluster handling strategy. The precursor ion tolerance prompt was added to accommodate differences between the center of an isolation window (reported in the .MGF format) and the calculated monoisotopic m/z value of an isolated precursor ion. The isotopic cluster handling strategy was added to account for differences in monoisotopic m/z

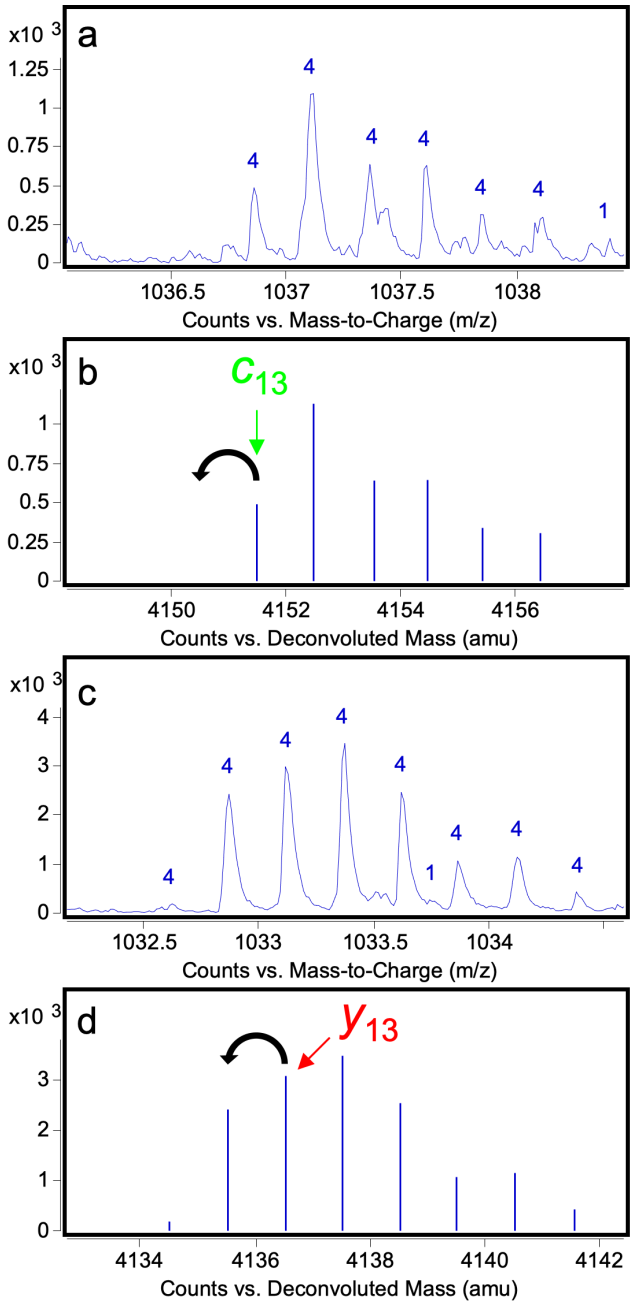


Figure 2.3: Isotopic distributions observed observed following negative ion CID of the ASO on the 6560c (a, c) identified and assigned a charge state (here 4-) by unbiased isotope cluster analysis. Charge deconvoluted, centroided isotopic clusters (b, d) exported to .MGF and analyzed by OligoTap. The c_{13} fragment ion (a, b) is assigned as a true match whereas the potential y_{13} fragment ion (c, d) is discarded as it matches the second isotopologue of the corresponding isotopic cluster.

extraction from Bruker and Agilent data as well as to allow analysis of non-deisotoped data.

For Agilent data, the unbiased isotope model identifies isotopic distributions based on peak spacings and assigns their corresponding charge state as shown in **Figures 2.2a and c** for two observed quadruply deprotonated fragment ions from the ASO oligonucleotide. Charge deconvolution to the corresponding neutral masses results in the isotopic distributions shown in **Figures 2.2b and d**. For this data type, OligoTap searches a window of 0.97-1.03 Da to the left of the assigned lowest mass peak in an isotopic cluster. If no peak is found within this window, the corresponding peak is searched as the monoisotopic mass, e.g., the assigned c_{13} ion in Figure 2b. By contrast, if a peak is found within this window, the peak in question will be discarded, e.g., the y_{13} ion in Figure 2d, which was erroneously assigned to a ^{13}C isotopologue in the absence of this strategy. For non-deisotoped data, all peaks in an isotopic cluster can potentially be matched, resulting in a high number of false positives. Default settings for the numeric prompt selections (3 m/z, 20 ppm, 4, and 3 for prompts 4-7, respectively) are auto populated and descriptions of the requested inputs are displayed in a multi-step pop-up user interface.

After answers to the prompts are provided, the script runs and displays a pop-up window, including the execution time, once the run is complete. For ~20 .MGF files searching against four sequences, the execution time is ~5 minutes. During the run, m/z values are calculated for all *abcd/wxyz* as well as common (*a* - nucleobase) backbone fragment ions for each input sequence. These calculations are based on hard coding the chemical formula for the different variable ON portions (nucleobase, sugar,

and backbone linker) into the script. These chemical formulas are accessible for modification, addition, or removal within the script. Monoisotopic atomic masses were coded to six decimal points, as shown in Appendix A Table A.2. All signals in the accepted (based on matched precursor ions) MGF file(s) are compared against all calculated theoretical fragments. The output directory contains the processed .MGF files, a text file containing the calculated m/z values for all theoretical fragment ions for each input sequence, and a report text showing identified fragment ions along with sequence coverage information. An example report for the modified SO searched against a text file containing the sequences unmodified SO, SO, ASO, and an 11mer is shown in the supporting information. OligoTap has shown reproducible performance across multiple (6) Windows 10 computers with run time correlating to the clocking frequency of the central processing unit. The script is a single core process. The OligoTap coded algorithm is available on GitHub at https://github.com/cwszot/OligoTap. Note that the variables for the chemical formulas are present in two positions; the "initialize" and "iterate construction" sections, i.e., when adding/modifying building blocks, these need to be edited in two places.

2.3.2 Fluoroalcohol-Free LC/MS

The chosen fluoroalcohol-free TEAB/methanol-based LC solvent system was first implemented on a SolariX Q-FT-ICR mass spectrometer at the University of Michigan (UM) and optimized for separation of the three chosen 22-mer oligonucleotides. These sequences include a number of common synthetic modifications and have a size typical of siRNA therapeutics.[7] We found that the LC flow rate had a significant effect on the observed MS^1 charge state distributions (**Figure 2.3**). Because several charge states

may need to be sampled to maximize ON sequence coverage, we chose to work at 0.1 mL/min at which higher and lower charge states showed more similar abundance

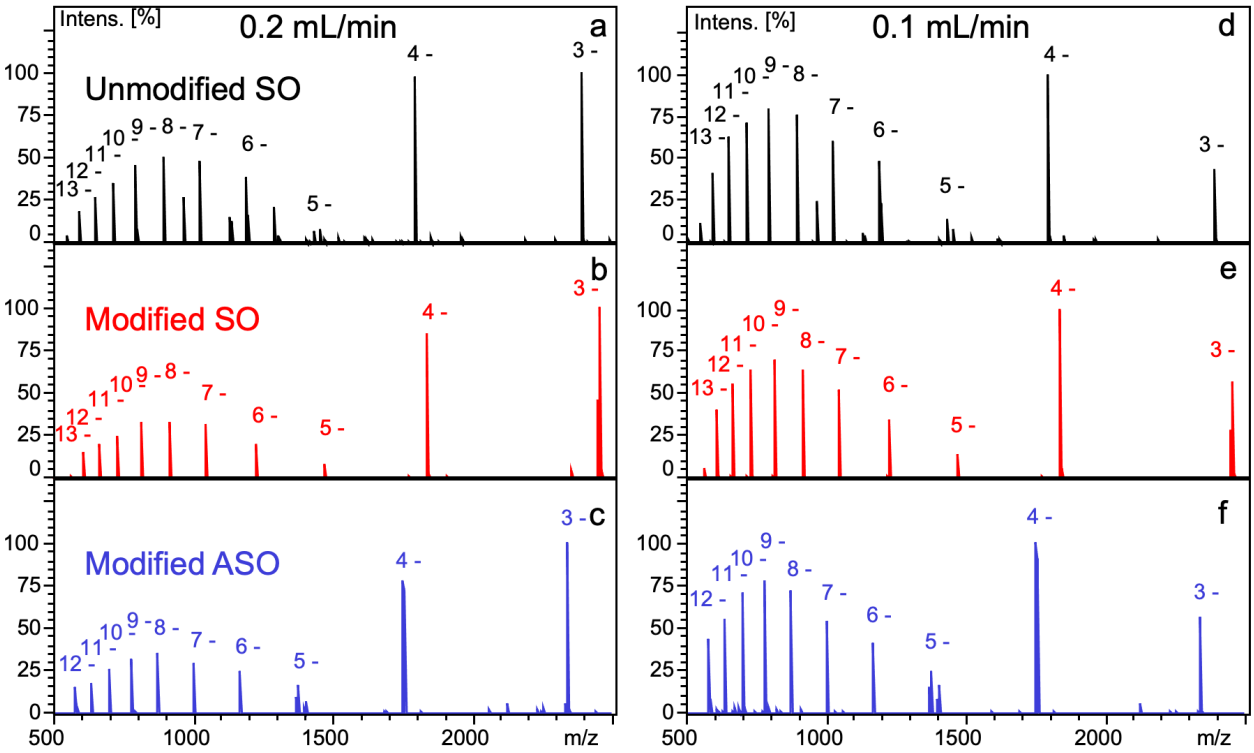


Figure 2.4: LC/MS spectra of three 22-mer oligonucleotides with a TEAB/methanol-based solvent system on a SolariX Q-FT-ICR mass spectrometer. Averaged mass spectra across each elution time showed more even abundance of high and low charge states at 0.1 mL/min (d,e,f) compared with 0.2 mL/min(a,b,c).

(Figure 3, Right). This effect was similar for all three sequences with detected charge states from 3- to 13-, similar to the higher charge states typically observed with fluoroalcohol-based solvents. [24, 38, 39]

2.3.3 Fluoroalcohol-Free LC/MS/MS

Despite the moderate length of the three chosen ONs, CID/HCD MS/MS spectra of the ~7 kDa precursor ions acquired on all three instruments (SolariX, Agilent 6560c, and Q-Exactive) for both low and high charge states are complex in nature. The region below m/z 1,200 is particularly crowded with fragment ions of varying sizes and charge states. Thus, high resolution and mass accuracy is a necessity. **Figure 2.4a** shows an

example zoomed-in view of a convoluted low-abundance region of the CID MS/MS spectrum from the IM-Q-TOF, which had a resolving power of ~15-20k under the chosen conditions. Manual interpretation of this region is challenging. However, upon subjecting these data to the Agilent unbiased isotope cluster algorithm in Qualitative Analysis Navigator software, two isotopic clusters (labeled "66" and "67" in **Figure 2.4b**) are identified. The assigned corresponding charge states are shown in **Figure 2.4c** as 5- and 1-, respectively. **Figure 2.4d** shows the same m/z region on the FT-ICR mass spectrometer with noticeably higher resolving power (~50-90k). The Bruker SNAP algorithm assigns a 5- species with a monoisotopic m/z value (indicated by the prime in front of the number) of 771.8872, corresponding to a d_{12} fragment ion from the ASO. The isotopic pattern calculated from the RNA averagine repeat unit, $C_{9.5}H_{14.75}N_{3.75}O_8P_1$, is shown in **Figure 2.4e**. For this ion, the

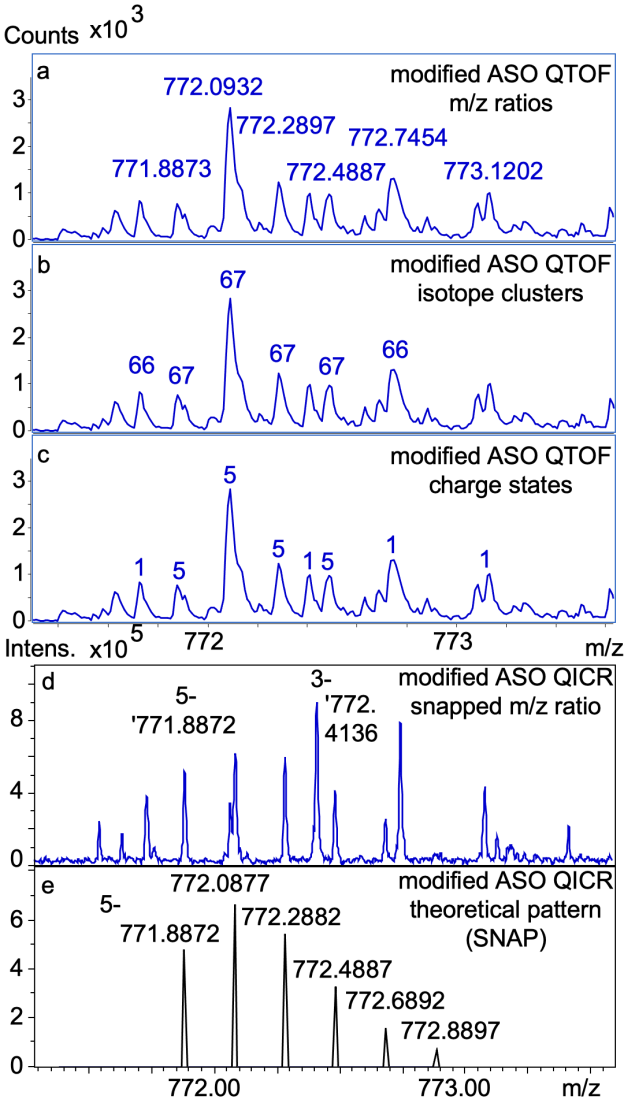


Figure 2.5: Isotopacially resolved signals in CID tandem mass spectra of the ASO. Zoomed-in IM-Q-TOF spectrum with centroided m/z ratios assigned by Qualitative Analysis Navigator (a). Isotope cluster assignments (number 66 and 67) from an unbiased isotope model (b). Charge states of the assigned overlapping isotope clusters (five and one, respectively; c). The same spectral window following FT-ICR analysis under otherwise similar conditions (d). The 5- isotopic cluster was also assigned in DataAnalysis with the SNAP algorithm (d). Theoritical isotope cluster calculated from an RNA repeat unit (e).

Agilent and Bruker algorithms agree. However, the cluster 66 assigned on the Q-TOF is not assigned in the FT-ICR data, likely because the isotopic cluster relative abundance does not match the averagine for a singly charged ion at this m/z ratio. In addition,

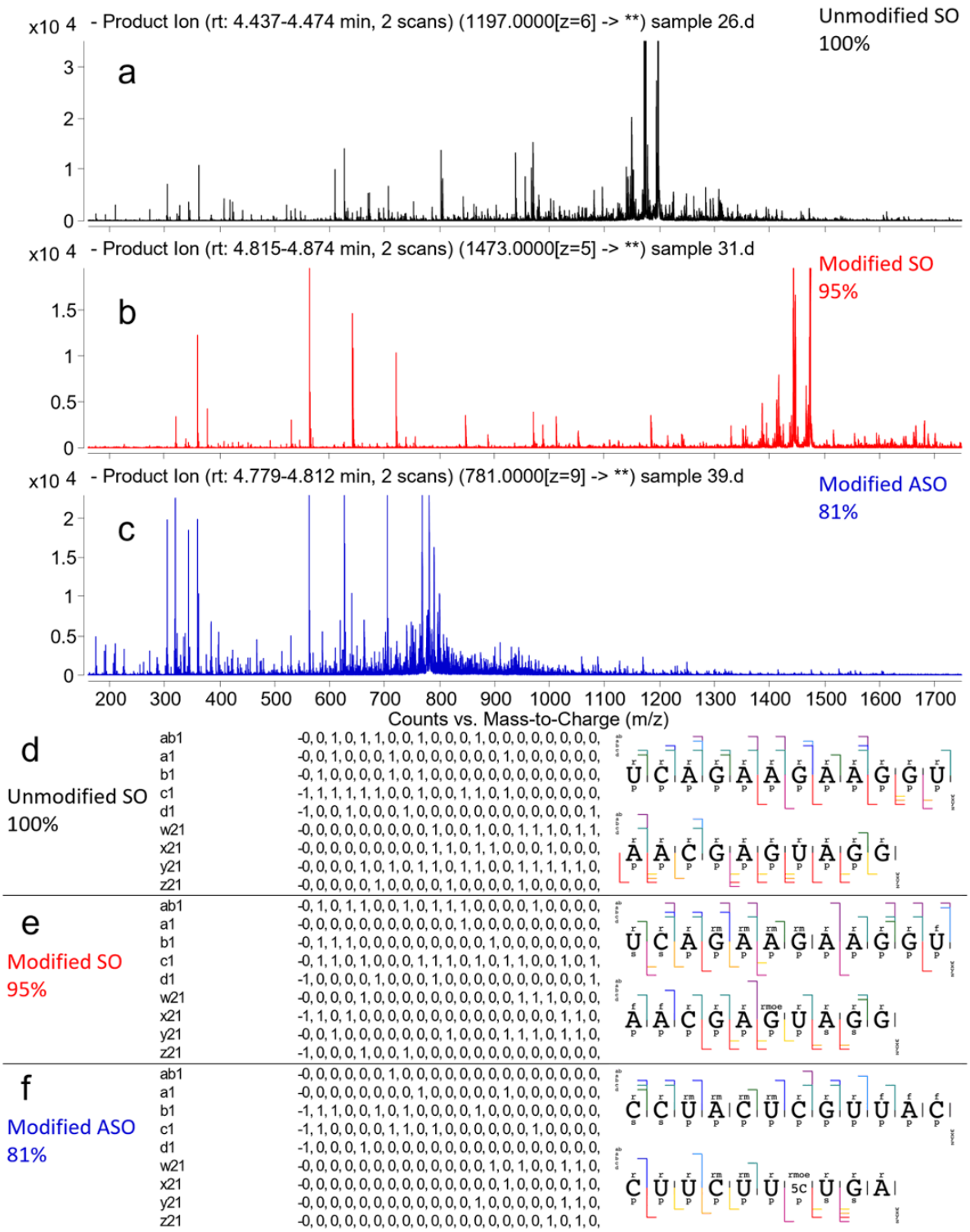


Figure 2.6: CID tandem mass spectra ad resulting sequence coverage for three 22-mer oligonucleotides on an Agilent 6560c. CID of the 6- charge state (a), the 5- charge state (b), and the 9- charge state (c) provided the highest sequence coverage for the unmodified SO, the modified SO, and the modified ASO, respectively. Boolean sequence coverage map from OligoTap (d-f, Left) and visualization from the in-house written Terrapin (d-f, Right).in a second in house written Python Turtle script. The same outputs are shown for the modified sense and modified antisense in figure part e and f.

SNAP assigns a 3- ion at m/z 772.4136 not identified in the Agilent data, likely due to

the overlapping peaks at m/z 772.6884 (the second isotopologue of the 3- cluster) and m/z 772.6892 (the fifth isotopologue of the assigned 5- cluster). A shoulder is noted in the lower resolution data; however, the FT-ICR is able to resolve these two peaks for accurate assignment. On the other hand, in order to achieve the corresponding high resolving power, the total acquisition time of the FT-ICR is significantly longer than for the IM-Q-TOF. Consequently, with the FT-ICR, only one precursor ion charge state can be analyzed with high MS/MS spectral quality per LC run. By contrast, the IM-Q-TOF was able to gather data for two charge states without any loss in data quality.

Overall, MS/MS spectra were acquired on both the IM-Q-TOF and Q-FT-ICR platforms for the eight most abundant charge states of the three ON 22-mers. Example MS/MS spectra are shown in **Figures 2.5a-c** for IM-Q-TOF data for the charge states that provided the highest sequence coverage for each ON. The unmodified ON reached 100% sequence coverage in many cases, a representative spectrum is shown for the 6- precursor ion (Figure 5a). As expected, sequence coverage was lower for the heavily modified ONs, specifically in the middle around the 2'fluorinated sugars. The difficulty of achieving backbone cleavage next to such modifications has been previously reported.[33] For the modified SO, the 5- charge state provided the highest sequence coverage, 95%, with only one backbone position missing (Figure 5b), whereas the modified ASO was more challenging with the highest sequence coverage being 81% (four missed backbone positions) for the 9- charge state. All data were ran through OligoTap with the resulting Boolean sequence coverage maps for Figures 5 a-c being shown in **Figures 2.5 d-f** (Left). Because these outputs are not visually appealing, an inhouse Python Turtle script, Terrapin, was developed to represent the data as shown in

Figures d-f (Right). Terrapin allows all modifications to be indicated in a compact format and color codes each fragment ion type occurring at each backbone cleavage position.

2.3.4 Procedure for Attaining 100% Sequence Coverage

To achieve full sequence coverage of the modified sense and antisense oligonucleotides, additional LC/MS/MS data were collected on the Solarix instrument with varying collision energy. The three oligonucleotides were first subjected to six different CID voltages, 5, 10, and 15 V for a high charge state (z = -9) and 25, 30, and 35 V for a low charge state (z = -4), i.e., 18 LC/MS/MS runs using the HPH-C18 method (a total of 8.1 hours). Assigned sequence coverages annotated by OligoTap for externally calibrated data at 20 ppm mass tolerance are shown in Appendix A Figure A.6. A SNAP quality factor of 0.25 and S/N ratio of 1 was used for this local maximization step. Linear extrapolation of optimum CID voltage for maximum sequence coverage of other charge states was then performed from the optimum voltages found for z = -4 and -9. A value of 10 V was used for the 9- charge state; however, 25 V rather than 30 V was used for the 4- charge state to reduce the risk of over fragmenting the samples. The resulting linear fit is shown in Figure S7d (the initially measured 4- and 9- charge states are labeled with red data points). LC-MS/MS spectra were then collected for the 4- to 11- charge states at the extrapolated optimum CID energies, internally calibrated, analyzed via OligoTap, and mined for missing fragments. Both FT-ICR and Q-TOF data were gathered. For SolariX data, only signals with a SNAP quality score of ≥0.5 and S/N ratio of ≥3 were considered. The quintuply charged d_{12} ion shown in Figure 4 is one of the fragments added to the sequence coverage following these all-charge-state experiments. This approach resulted in full sequence coverage for all three

ONs on the SolariX instrument (~8 hours). The same experiments on the 6560c showed full sequence coverage for the unmodified and modified SO but only 90% for the ASO in half the time (4 hours) at an intensity threshold of 150. With no intensity threshold, full sequence coverage is supported for all three ONs. Overall, OligoTap was key to rapid, automated mining of these different conditions to maximize sequence coverage. For this deep annotation, high resolution data is a great advantage.

2.4 Conclusion

We present support for full sequence coverage of heavily modified oligonucleotides based on stringent, automated deep mining of LC-CID MS/MS spectra from different conditions and instrument platforms/data formats with the developed bioinformatics utility OligoTap. A critical step in data processing, highlighted in this work, is appropriate isotopic clustering of fragment ions. We utilized two strategies: the Bruker SNAP algorithm and the Agilent unbiased grouping algorithm with or without additional support from OligoTap. The use of mascot generic format files for processing oligonucleotide data appears to be a viable option with the bottleneck being adequate isotope clustering and monoisotopic peak detection prior to generation of the .MGF file. We also show that fluoroalcohol-free LC solvent systems can provide a reproducible means of generating high quality spectra with limited adduction and chromatographic resolution of impurities. Lower flow rates appeared to increase oligonucleotide average charge states while maintaining a wide range for sequence coverage optimization. The detection of fragments supporting sequence coverage is based on low stringency 3:1 signal to noise ratio, suggesting that fragments assigned here are near the limit of

detection. The use of triethylammonium bicarbonate produces a wide range of charge states in standard and heated electrospray ionization.

2.5 References

1. Lundin, K. E.; Gissberg, O.; Smith, C. I. Oligonucleotide Therapies: The Past and the Present. *Human Gene Therapy* **2015,** *26*, 475-485.

2. Xiong, H.; Veedu, R. N.; Diermeier, S. D. Recent Advances in Oligonucleotide Therapeutics in Oncology. *Int. J. Mol. Sci.* **2021,** *22*, 3295.

3. Roberts, T. C.; Langer, R.; Wood, M. J. A. Advances in Oligonucleotide Drug Delivery. *Nat. Rev. Drug Discovery* **2020,** *19*, 673-694.

4. Forni, G.; Mantovani, A. Covid-19 Vaccines: Where We Stand and Challenges Ahead. *Cell Death Differ.* **2021,** *28*, 626-639.

5. McKenzie, L. K.; El-Khoury, R.; Thorpe, J. D.; Damha, M. J.; Hollenstein, M. Recent Progress in Non-Native Nucleic Acid Modifications. *Chem. Soc. Rev.* **2021,** *50*, 5126-5164.

6. Khvorova, A.; Watts, J. K. The Chemical Evolution of Oligonucleotide Therapies of Clinical Utility. *Nat. Biotechnol.* **2017,** *35*, 238-248.

7. Hu, B.; Zhong, L.; Weng, Y.; Peng, L.; Huang, Y.; Zhao, Y.; Liang, X.-J. Therapeutic Sirna: State of the Art. *Signal Transduction Targeted Ther.* **2020,** *5*, 101.

8. Demelenne, A.; Servais, A.-C.; Crommen, J.; Fillet, M. Analytical Techniques Currently Used in the Pharmaceutical Industry for the Quality Control of Rna-Based Therapeutics and Ongoing Developments. *J. Chromatogr. A* **2021,** *1651*, 462283.

9. Hannauer, F.; Black, R.; Ray, A. D.; Stulz, E.; Langley, G. J.; Holman, S. W. Advancements in the Characterisation of Oligonucleotides by High Performance Liquid Chromatography-Mass Spectrometry in 2021: A Short Review. *Anal. Sci. Adv.* **2022,** *3*, 90-102.

10. Sutton, J. M.; Guimaraes, G. J.; Annavarapu, V.; Dongen, W. D. v.; Bartlett, M. G. Current State of Oligonucleotide Characterization Using Liquid Chromatography–Mass Spectrometry: Insight into Critical Issues. *J. Am. Soc. Mass Spectrom.* **2020,** *31*, 1775-1782.

11. Guimaraes, G. J.; Sutton, J. M.; Gilar, M.; Donegan, M.; G.Bartlett, M. Impact of Nonspecific Adsorption to Metal Surfaces in Ion Pair-Rp Lc-Ms Impurity Analysis of Oligonucleotides. *J. Pharm. Biomed. Anal.* **2022,** *208*, 114439.

12. Yang, J.; Leopold, P.; Helmy, R.; Parish, C.; Arvary, B.; Mao, B.; Meng, F. Design and Application of an Easy to Use Oligonucleotide Mass Calculation Program. *J. Am. Soc. Mass Spectrom.* **2013,** *24*, 1315-1318.

13. Wein, S.; Andrews, B.; Sachsenberg, T.; Santos-Rosa, H.; Kohlbacher, O.; Kouzarides, T.; Garcia, B. A.; Weisser, H. A Computational Platform for High-Throughput Analysis of Rna Sequences and Modifications by Mass Spectrometry. *Nat. Commun.* **2020,** *11*, 926.

14. D'Ascenzo, L.; Popova, A. M.; Abernathy, S.; Sheng, K.; Limbach, P. A.; Williamson, J. R. Pytheas: A Software Package for the Automated Analysis of Rna Sequences and Modifications Via Tandem Mass Spectrometry. *Nat. Commun.* **2022,** *13*, 2424.

15. Santos, I. C.; Lanzillotti, M.; Shilov, I.; Basanta-Sanchez, M.; Roushan, A.; Lawler, R.; Tang, W.; Bern, M.; Brodbelt, J. S. Ultraviolet Photodissociation and Activated Electron Photodetachment Mass Spectrometry for Top-Down Sequencing of Modified Oligoribonucleotides. *J. Am. Soc. Mass Spectrom.* **2022,** *33*, 510-520.

16. Rozenski, J.; McCloskey, J. A. Sos: A Simple Interactive Program for Ab Initio Oligonucleotide Sequencing by Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2002,** *13*, 200-203.

17. Nakayama, H.; Akiyama, M.; Taoka, M.; Yamauchi, Y.; Nobe, Y.; Ishikawa, H.; Takahashi, N.; Isobe, T. Ariadne: A Database Search Engine for Identification and Chemical Analysis of Rna Using Tandem Mass Spectrometry Data. *Nucleic Acids Res.* **2009,** *37*, e47.

18. Nyakas, A.; Blum, L. C.; Stucki, S. R.; Reymond, J.-L.; Schürch, S. Oma and Opa—Software-Supported Mass Spectra Analysis of Native and Modified Nucleic Acids. *J. Am. Soc. Mass Spectrom.* **2013,** *24*, 249-256.

19. Sample, P. J.; Gaston, K. W.; Alfonzo, J. D.; Limbach, P. A. Robooligo: Software for Mass Spectrometry Data to Support Manual and De Novo Sequencing of Post-Transcriptionally Modified Ribonucleic Acids. *Nucleic Acids Res.* **2015,** *43*, e64.

20. Yu, N.; Lobue, P. A.; Cao, X.; Limbach, P. Rnamodmapper: Rna Modification Mapping Software for Analysis of Liquid Chromatography Tandem Mass Spectrometry Data. *Anal. Chem.* **2017,** *89*, 10744-10752.

21. Ortiz, D.; Gasilova, N.; Sepulveda, F.; Patiny, L.; Dyson, P. J.; Menin, L. Aom2s: A New Web-Based Application for DNA/Rna Tandem Mass Spectrometry Data Interpretation. *Rapid Commun. Mass Spectrom.* **2020,** *34*, e8927.

22. Gao, Y.; McLuckey, S. A. Fragmentation Reactions of Nucleic Acid

Ions in the Gas Phase. In *Nucleic Acids in the Gas Phase*, 1 ed.; Gabelica, V., Ed. Springer: Berlin, Heidelberg, 2014; pp 131-182.

23. Santos, I. C.; Brodbelt, J. S. Recent Developments in the Characterization of Nucleic Acids by Liquid Chromatography, Capillary Electrophoresis, Ion Mobility, and Mass Spectrometry (2010-2020). *J. Sep. Sci.* **2021,** *44*, 340-372.

24. Apffel, A.; Chakel, J. A.; Fischer, S.; Lichtenwalter, K.; Hancock, W. S. Analysis of Oligonucleotides by Hplc-Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **1997,** *69*, 1320-1325.

25. Huber, C. G.; Oberacher, H. Analysis of Nucleic Acids by on-Line Liquid Chromatography–Mass Spectrometry. *Mass Spectrom. Rev.* **2001,** *20*, 310-343.

26. Gong, L. Comparing Ion-Pairing Reagents and Counter Anions for Ion-Pair Reversed-Phase Liquid Chromatography/Electrospray Ionization Mass Spectrometry Analysis of Synthetic Oligonucleotides. *Rapid Commun. Mass Spectrom.* **2015,** *29*, 2402-2410.

27. Huber, C. G.; Krajete, A. Analysis of Nucleic Acids by Capillary Ion-Pair Reversed-Phase Hplc Coupled to Negative-Ion Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **1999,** *71*, 3730-3739.

28. Bartlett, M. G.; Chen, B.; McGinnis, A. C. Lc-Ms Bioanalysis of Oligonucleotides. In *Handbook of Lc-Ms Bioanalysis: Best Practices, Experimental Protocols, and Regulations*, LI, W.; Zhang, J.; Tse, F. L. S., Eds. John Wiley & Sons, Inc.: Hoboken, New Jersey, 2013; pp 607-627.

29. 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 Humanhealth Review: Current State of Knowledge and Strategiesfor Informing Future Research. *Environ. Toxicol. Chem.* **2021,** *40*, 606-630.

30. Neuwald, I. J.; Hübner, D.; Wiegand, H. L.; Valkov, V.; Borchers, U.; Nödler, K.; Scheurer, M.; Hale, S. E.; Arp, H. P. H.; Zahn, D. Ultra-Short-Chain Pfass in the Sources of German Drinking Water: Prevalent, Overlooked, Difficult to Remove, and Unregulated. *Environ. Sci. Technol.* **2022,** *55*, 6380-6390.

31. Birdsall, R. E.; Gilar, M.; Shion, H.; Yu, Y. Q.; Chen, W. Reduction of Metal Adducts in Oligonucleotide Mass Spectra in Ion-Pair Reversed-Phase Chromatography/Mass Spectrometry Analysis. *Rapid Commun. Mass Spectrom.* **2016,** *30*, 1667-1679.

32. Gao, Y.; McLuckey, S. A. Collision-Induced Dissociation of Oligonucleotide Anions Fully Modified at the 2'-Position of the Ribose: 2'-F/-H and 2'-F/-H/-Ome Mix-Mers. *J. Mass Spectrom.* **2012,** *47*, 364-369.

33. Gao, Y.; Yang, J.; Cancilla, M. T.; Meng, F.; McLuckey, S. A. Top-Down Interrogation of Chemically Modified Oligonucleotides by Negative Electron Transfer and Collision Induced Dissociation. *Anal. Chem.* **2013,** *85*, 4713-4720.

34. Köster, C. Mass Spectrometry Method for Accurate Mass Determination of Unknown Ions. US 6 188 064 B1, 1999.

35. Goldfarb, D.; Lafferty, M. J.; Herring, L. E.; Wang, W.; Major, M. B. Approximating Isotope Distributions of Biomolecule Fragments. *ACS Omega* **2018,** *3*, 11383-11391.

36. Senko, M. W.; Beu, S. C.; McLafferty, F. W. Determination of Monoisotopic Masses and Ion Populations for Large Biomolecules from Resolved Isotopic Distributions. *J. Am. Soc. Mass Spectrom.* **1995,** *6*, 229-233.

37. Taucher, M.; Rieder, U.; Breuker, K. Minimizing Base Loss and Internal Fragmentation in Collisionally Activated Dissociation of Multiply Deprotonated Rna. *J. Am. Soc. Mass Spectrom.* **2010,** *21*, 278-285.

38. Fountain, K. J.; Gilar, M.; Gebler, J. C. Analysis of Native and Chemically Modified Oligonucleotides by Tandem Ion-Pair Reversed-Phase High-Performance Liquid Chromatography/Electrospray Ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2003,** *17*, 646-653.

39. Wong, D.; Rye, P.; Madden, S.; Slysz, G.; Cody, C. In *Comprehensive, Automated, and Integrated Software for Oligonucleotide Characterization and Sequence Confirmation*, Proceedings of the 70th ASMS Conference on Mass Spectrometry and Allied Topics, Minneapolis, MN, June 5-9, 2022; p WP 386.

Chapter 3 False Discovery Rates in Automated Annotation of Oligonucleotide Liquid Chromatography-Tandem Mass Spectra: Collision Induced vs. Electron Detachment Dissociation

3.1 Introduction

Tandem mass spectrometry (MS/MS)-based sequencing is well established for peptides and proteins,[1, 2] including bioinformatics utilities[3-5] and their false discovery rates (FDRs)[1, 6] in complex mixture liquid chromatography-mass spectrometry (LC-MS) analysis.[7] However, automated annotation of oligonucleotide (ON) tandem mass spectra continues to be a developing space.[8-12] Recently introduced bioinformatics utilities have been catalyzed by the accelerating development of ON-based therapeutics[13, 14] and the accompanying need for analytical characterization, including liquid chromatography (LC)-MS/MS.[15]

Collision induced dissociation (CID) is the most common activation method for LC-MS/MS.[16, 17] In oligonucleotide CID, typically performed in negative ion mode,[18] there are four possible bond cleavage positions per nucleotide with 5' fragments termed *a*, *b*, *c*, *d* ions and 3' fragments termed *w*, *x*, *y*, *z* ions.[19] DNA ONs show preferred formation of (*a* – B) and their complementary *w*-type ions in which "B" refers to neutral loss of the nucleobase 3' to the backbone cleavage site.[20] By contrast, RNA ONs show preferred formation of complementary *c*/*y* ion pairs.[18] Thus, the chemical constituent at the 2' sugar position plays a major role in the fragmentation mechanism.[21] Because therapeutic ONs are often heavily chemically modified to improve efficacy,[22] including at

the 2' sugar position, their CID MS/MS spectra are complex.[17, 19] In Chapter 2, we introduced OligoTap, a bioinformatics utility that allows any modification to be considered in automated CID MS/MS spectral annotation. However, similar to other biopolymers,[23] CID-based MS/MS sequence coverage often decreases with increasing ON length.[24] Furthermore, 2' modifications such as fluorination can significantly decrease the cleavage probability of adjacent backbone bonds.[19]

Alternative activation methods,[24-28] such as electron detachment dissociation (EDD)[25, 29] have been shown to provide high ON sequence coverage with different backbone cleavage preferences compared with CID. For example, EDD majorly yields non-complementary *d*- and *w*-type fragments for both DNA and RNA.[24-26] EDD on an LC time scale was first demonstrated by Kjeldsen et al for phosphopeptides.[30] Recent MS/MS technologies have shown promise for fast and effective EDD[31, 32] including LC-EDD in the 10-30 eV range.[32]

The complementary nature of alternative MS/MS methods[33, 34] such as EDD compared with CID adds further complexity to spectral annotation and analysis algorithms. Recently introduced tools such as the NucleicAcidSearchEngine (NASE) have begun to tackle this complex assignment space, including the associated FDRs.[10, 14] Other efforts have focused on the development of activation strategies that provide less complex ON MS/MS spectra with accompanying lower FDR.[35] One approach to evaluating FDR is to introduce scrambled sequences, or decoys, into the search space.[1, 35, 36] Here, we use the latter strategy to evaluate sequence coverage and FDR in automated OligoTap annotation of LC-CID vs. LC-EDD tandem mass spectra for a 19mer RNA.

3.2 Experimental Procedure

3.2.1 Materials

The 19mer RNA CGCGCCGACUGCGGCGCG**U** was purchased from Integrated DNA technologies (IDT, Coralville, IA). The sample label did not include the final nucleotide of the sequence but did include the molecular weight. The molecular weight was confirmed by mass spectrometry indicating that the final nucleotide contained uracil. This nucleotide identity was additionally confirmed via MS/MS and matching with all four nucleobase possibilities using OligoTap. LiChropur trimethylammonium bicarbonate TEAB was procured from Sigma (St. Louis, MO). Optima LC-MS grade water (≤20 ppb Na and ≤10 ppb K) and methanol (≤50 ppb Na and ≤10 ppb K) were purchased from ThermoFisher (Pittsburgh, PA).

3.2.2 LC-FT-ICR MS/MS

Ion pair-reverse phase (IP-RP) liquid chromatography at a flow rate of 0.050-0.300 mL/min was performed with an Agilent 1100 HPLC coupled to a SolariX XR Q-FT-ICR (7T; Bruker, Billerica, MA) mass spectrometer. The sample concentration was not available as this RNA had been frozen in water for an extended time period. The sample was first tested with the method described in Chapter 2. However, the signal abundance was lower than for previous LC-EDD experiments.[37] Thus, the LC method was adjusted to a lower flow rate (0.050 mL/min) and the ion accumulation time was set to 0.4 s to approach an LC-EDD MS/MS total ion chromatogram intensity of 2e9. [37] One challenge in transitioning the workflow was the inability of the SolariX platform to perform microaccumulations (Chapter 2) when using an in-cell activation technique. An

Agilent HPH-C18 2.1x50 mm, 1.9 µm, 100 Å, stainless steel column was used at 60 ˚C with a 5 mM TEAB/methanol linear gradient from 5% to 100% B over 10 min. To manage run time, a flow ramp from 0.050 mL/min to 0.300 mL/min at 10 minutes and back down to 0.05 mL/min at 20.75 minutes during cleaning and re-equilibration was implemented. Column eluent was directed to an Agilent ES standard nebulizer at a standard needle extension flush with the nebulizer. This nebulizer was attached to the mass spectrometer ion source, which was operated in negative ion mode at 3,500 V. The sprayshield was set to 500 V lower than the inlet capillary, the nebulizer pressure to 3 bar, and the drying gas temperature and flow rate to 350 ˚ C and 10 L/min, respectively. LC data were acquired in serial mode with 1 scan per mass spectrum, 0.4 second external accumulation time, and 1 megaword acquisition size. Excitation/detection was performed in broadband mode with the frequency sweep in the decreasing frequency direction for excite calculation version 1 in Bruker ftmsControl, v2.3.0.

A static quadrupole isolation window of 20 m/z was used for SolariX LC-MS/MS experiments. Abundant precursor ion charge states, ranging from 3- to 12-, were obtained in an LC-MS run prior to MS/MS experiments. The 8- charge state of the RNA was the second most abundant following the 3- charge state. The higher 8- charge state was chosen for both CID and EDD. The collision voltage for CID was 11 V and EDD was performed with a hollow cathode heating current of 1.5 A, a bias voltage (internal ionization energy) of -30 V, and a lens voltage of +20 V. The irradiation time was 1 second.

3.2.3 Data Analysis

Bruker's DataAnalysis 5.0 SR1 was used in three steps for each chromatogram: 1) MS^n level extracted ion chromatograms (EICs) were generated and the resulting chromatographic peaks were identified by the 3.0 compound detection algorithm. The corresponding compound MS/MS spectra were averaged and a monoisotopic MS/MS peak list was generated with the SNAP[21] algorithm over the 100-3000 m/z range. 2) This peak list was internally recalibrated, and 3) exported as Mascot generic file format (MGF). The SNAP isotopic cluster finder for monoisotopic peak detection was used with a Q factor of 0.9 and a signal-to-noise (S/N) ratio of 3. For averagine-type fitting of RNAs, an average repeat unit of $C_{9.5}H_{14.75}N_{3.75}O_8P_1$ was used, determined by adding an oxygen and a hydrogen to the Bruker nucleotide building block chemical formula editor, summing each element of the four nucleotides, and dividing each of the summed elements by four. This approach is slightly different from the more accurate, previously published, RNA average repeat unit of $C_{9.5}H_{11.75}N_{3.75}O_8P_1$. The compound detection algorithm typically found one compound per LC/MS/MS data file (with single oligonucleotide injection). Subsequent, linear internal calibration was based on >3 typical *c*-type and *d*-type fragment ions from CID and EDD, respectively, in a batch processing format. Following internal recalibration, a subsequent Data Analysis method was run to conserve compounds and calibration but update peak lists with a lower SNAP Q factor of 0.5. Peak assignments were made with OligoTap (see Chapter 2), using the SNAP isotope handling option, and an error of $\leq$10 ppm.

Permutations of oligonucleotide sequence segments, 2-5 nucleotides in length (see **Figure 3.3a**), were generated in Python. The resulting lists of permutated sequences were processed in Excel. Duplicate sequence appear due to the same

nucleotide being placed in the same position across multiple possible rearrangements of the sequence. Such duplicates were removed from the permutation lists. The non-degenerate permutated sequences, all 19 nucleotides in length, were placed in a text file. A VBScript was used to copy these sequence permutations into a different text file with the necessary delimiters for OligoTap 1.0.33 to read. The permutated sequences were named according to the length of the permutated sequence element. Percent sequence coverage assigned by OligoTap for the different permutation sites and lengths (i. e. 3', 5', middle, and 2,3,4, and 5) were extracted from the OligoTap results files with a Python script. The extracted results were used to generate plots like those shown in **Figure 3.4**.

3.3 Results and Discussion

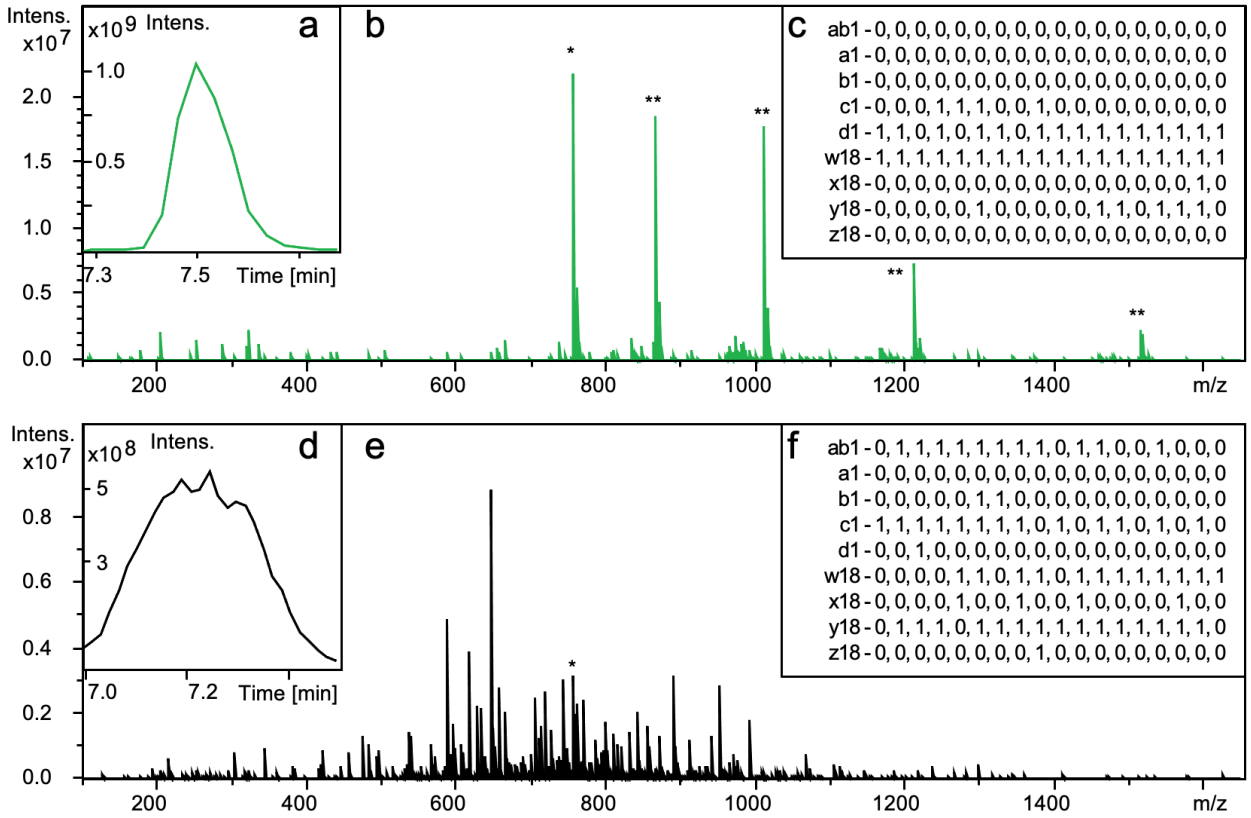


Figure 3.1 LC-MS/MS of a 19mer RNA. Total ion chromatogram (a) MS/MS spectrum (b) and OligoTap spectral annotation (c) following EDD. Total ion chromatogram (d), MS/MS spectrum (3) and OligoTap spectral annotation (f) following CID.

3.3.1 LC-CID FT-ICR MS/MS of 19 mer RNA

A 5 mM TEAB/methanol mobile phase at 0.050 mL/min allowed efficient desalting to yield an abundant peak in the EDD MS^2 total ion chromatogram (**Figure 3.1 a**) with an elution window of ~15 s for a 19mer RNA. However, the utilized chromatographic method appeared to lack sufficient reequlibration time based on the gradual broadening of chromatographic peaks in subsequent runs (e.g., ~30 s elution window in **Figure 3.1 d**). This chromatographic peak broadening resulted in ~2.5-fold reduction in maximum ion abundance. On the other hand, a wider window is available for signal averaging. For LC-CID MS/MS, the abundant 8- charge state (m/z 759) was

chosen as the precursor ion, due to its higher charge state compared to the most abundant 3- charge state. A collision voltage of 11 V was deemed appropriate as the precursor ion was depleted to a similar level as the abundance of nearby fragment ions. Across the CID-MS/MS spectrum (**Figure 3.1 e**) both complementary *c/y* and complementary (*a* –B)/*w*-type ions are observed. OligoTap annotation of this spectrum (**Figure 3.1 f)** supports full sequence coverage based on 11 observed (*a* – B) ions, 13 *c* ions, 12 *w* ions, and 15 *y* ions.

3.3.2 LC-EDD FT-ICR MS/MS of 19 mer RNA

EDD has previously been demonstrated[25, 26, 29, 30, 38, 39] with FT-ICR mass spectrometers equipped with conventional ICR cells. Our SolariX Q-FT-ICR instrument was recently upgraded with a dynamically harmonized[40] "ParaCell", which changes the conditions for EDD[41]. As schematically shown in **Figure 3.2**, such ICR cells add a radial electric field through "shimming" electrodes interspersed between the excitation and detection areas. This cell design allows for enhanced dynamic range,[42] i.e., electrons can be injected along with an anion population at large abundance and small abundance signals can still be detected. In addition, we hypothesize that higher electron energy may be needed due to the increased electric field from a higher anion

population. The latter condition is exacerbated with LC analysis in which analytes are entering the cell as a relatively concentrated ensemble.

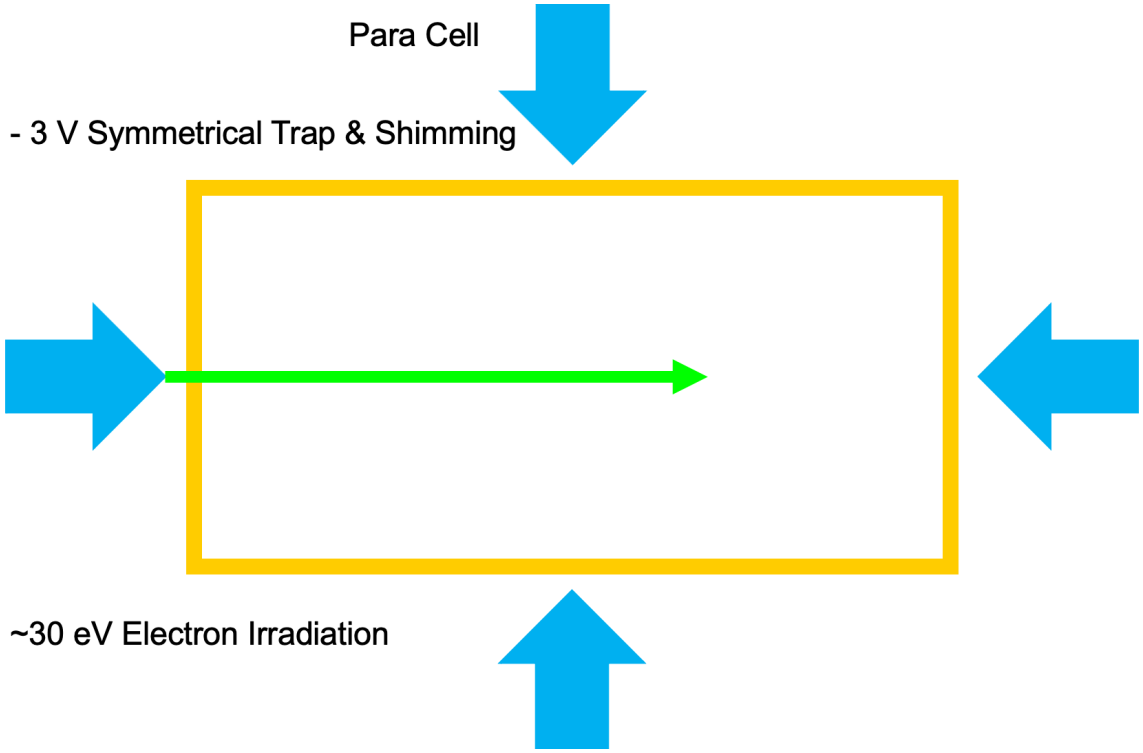


Figure 3.2 Schematic depiction of the ParaCell with the ion source to the right. For anion trapping, - 3 V is applied to both the front and rear trapping plates as well as the orthogonal shimming electrodes, creating a symmetrical electric field (indicated by the yellow lines. For EDD, an electron beam (indicated in green) generated by an indirectly heated hollow cathode is injected into the cell with an accelerating potential of -30 V.

Efficient EDD of the 19mer RNA was observed at 30 V cathode bias, which is higher than the previously reported maximum performance at 19 V with an Infinity ICR cell.[25] However, this electron energy regime is expected to need further tuning for optimum performance, including characterization of precursor anion population effects. Nevertheless, under the current EDD conditions (**Figure 3.1 b**), 15 *d* ions and all 18 possible *w* ions, along with 4 *c* ions and 6 *y* ions were annotated by OligoTap (**Figure 3.1 c**), corresponding to full sequence coverage of the 19mer RNA. As typical in EDD, the precursor ion (z = 8-,m/z = 759) remains as the most abundant signal followed by

the 7-, 6-, 5-, and 4- oxidized species resulting from detachment of 1-4 electrons, respectively, with no fragmentation.

3.3.3 OligoTap FDR Analysis for CID and EDD LC-MS/MS Spectra

a

CGCGCCGACUGCGGCGCGU

……..……..$C_5G_4C_3G_2U_1$

$C_1G_2C_3G_4C_5$…….....……..

…….$A_4C_2U_1G_3C_5$……

b

#	3’	5’	Middle
2	2	2	2
3	6	3	6
4	12	6	24
5	30	10	60

c

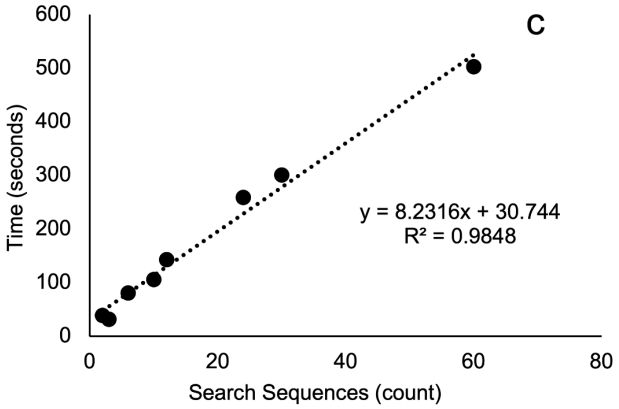


Figure 3.3 Selected permutation regions of the target 19mer RNA sequence (a). Up to five nucleotides were permutated with consecutive elongation as indicated by the subscripted numbers in the three chosen regions. The number of unique , isomeric permutated sequences for 2-5 nt sequence stretches in each sequence segment (b). The OligoTap 1.0.33 execution time as function of the number of sequences in the search space (c).

In order to begin to address FDRs in automated OligoTap sequence annotations it would be ideal to generate all isomeric permutations of a sequence of interest; however, this approach is computationally expensive. Thus, we chose to generate permutations of up to five nucleotides of the 19mer RNA sequence at each end as well as in the middle of the sequence (**Figure 3.3 a**). While permutations at only one end may have been sufficient for this evaluation, inclusion of both ends, provides an opportunity to study the effects of two and three types of nucleotides for the 5’ and 3’ end, respectively (CGCGC vs. CGCGU) while the middle portion contains all four nucleotides (ACUGC). Nucleotides were added sequentially to the permutated sequence by either progressively moving

inward into the sequence from the ends or symmetrically moving outwards from the middle portion of the sequence (as indicated by the subscripted numbers in Figure 3.3 a). The number of non-degenerate, isomeric permutated sequences ranged from 2 to 60 with the lowest number corresponding to two nucleotide length permutated segments and the highest number corresponding to a five nucleotide length permutated segment from the middle of the sequence (**Figure 3.2b**). These sequence permutations resulted in a total of 12 sequence pools, corresponding to 2, 3, 4, and 5 permutated nucleotides for each of the three sequence regions.

OligoTap 1.0.33 was used to annotate both the LC-CID and LC-EDD MS/MS spectra (Figure 3.1 b, e) against each of the twelve sequence permutation ensembles, which all contained the true sequence. Because we had not previously challenged OligoTap with a larger search space, we were pleased to observe that the run time was linear with respect to the number of sequences in the search space (**Figure 3.3 c**) with ~8 seconds added to the execution time for each additional sequence in the search sequence file. In these analyses, the “searched data” file population remained constant at two MS/MS spectra (one CID, one EDD). There are ~400 million permutations for a 12 nucleotide sequence, i.e., the currently assessed OligoTap1.0.33 would be incapable of practically searching all permutations on a single computer.

The assigned sequence coverage versus length of the permutated sequence segment for the 5', 3', and middle sequence regions, respectively, are plotted in **Figure 3.4** for both the CID and EDD MS/MS data. All six plots (three sequence regions for both CID and EDD data) show a maximum sequence coverage of 100% (red bars in Figure 3.4), independently of the number of permutated sequences in the search, because the true sequence is present in all searches. However, the minimum sequence coverage (blue bars in Figure 3.4) decreases with the length of the permutated

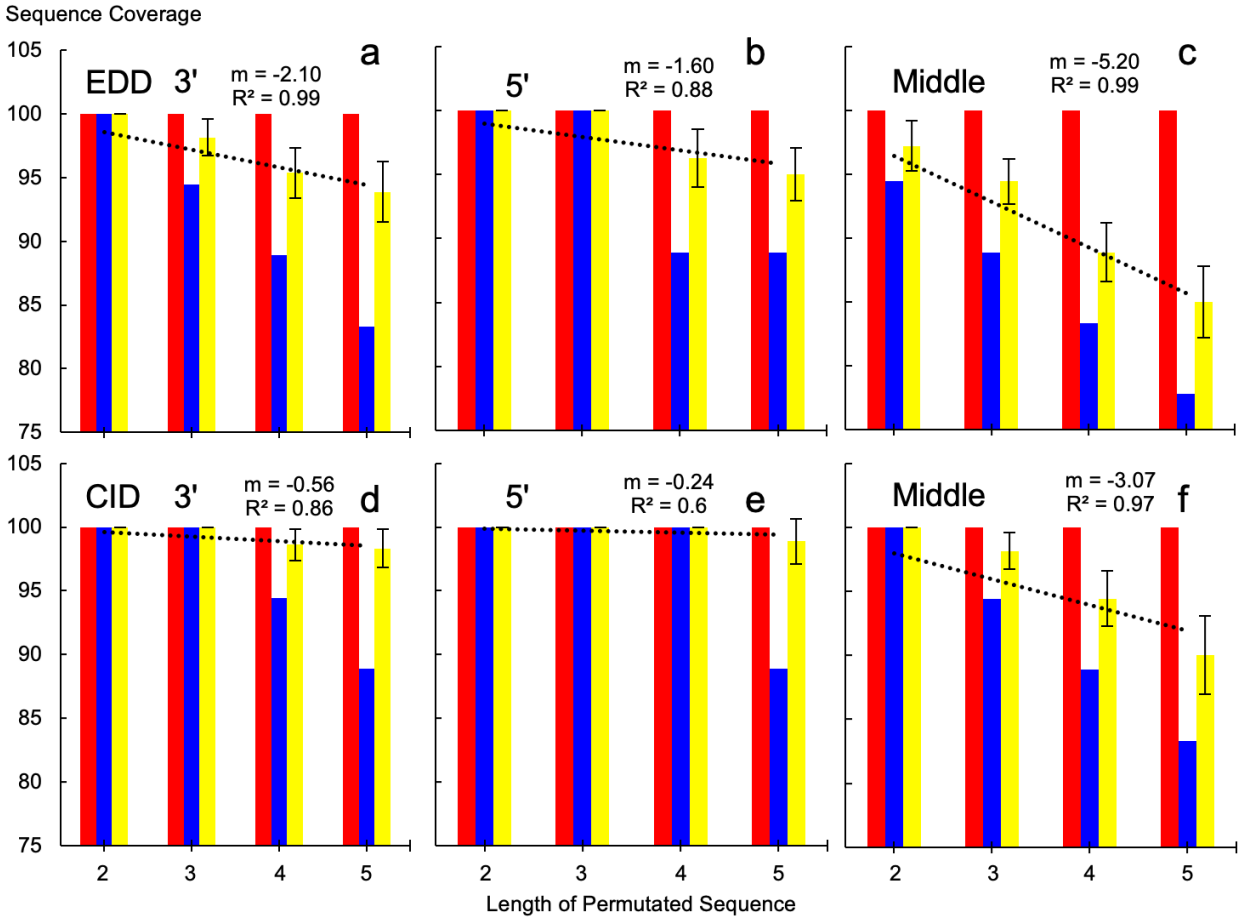


Figure 3.4 Sequence coverage annotated by OligoTap as function of the length of the permutated sequence region for the 5' (b,e), 3' (a,d)) and middle (c,f) portions of the 19mer RNA sequence for both CID (d-f) and EDD (a-c) MS/MS data. Each plot shows maximum sequence coverage (red bars), minimum sequence coverage (blue bars) and calculated average sequence coverage (yellow) for all isomeric sequences in the corresponding search space. The average sequence coverage bars include error bars at one standard deviation. The linear regression equation had the intercept set to100 at the shortest permutation sequence length of 2 nucleotides. Each plot includes the correlation coefficient and slope (m) of the linear trendline.

sequence region. The latter observation is explained by a higher number of differentiating theoretical fragment ions between the true and the scrambled sequences when more nucleotides are permutated. In order to quantitatively assess OligoTap's ability to differentiate between the true and decoy sequences the average sequence coverage for all sequences in the corresponding search space (see Figure 3.3 b) was calculated for each permutated sequence region and MS/MS technique. The corresponding numbers are plotted as yellow bars in **Figure 3.4**. Based on this analysis, it is apparent that the EDD data generally produced a greater gap between the maximum and minimum sequence coverage assignments. Linear trendlines assigned to the average sequence coverages also show greater negative slopes in the EDD data compared with the corresponding CID data. These graphs thus suggest that EDD has superior ability, compared with CID, to differentiate the true sequence from the decoy sequences However, it should be noted that OligoTap, version 1.0.33 does not calculate possible internal fragments from CID, nor possible odd electron *a*• and *z*• fragments from EDD. Both of these experimental methods may also generate additional unaccounted for fragmentation pathways which contribute to spectral complexity. Thus, further analysis is needed to support the current conclusion that EDD results in lower FDR than CID.

3.4 Conclusion

The presented data show that OligoTap is capable of automatically annotating both CID and EDD MS/MS spectra with a moderately sized sequence database. A linear increase in run time as function of the number of sequences in the search space was observed. In an attempt to evaluate FDR in CID vs. EDD data, scrambled isomeric

sequences were generated and EDD was shown to better differentiate the true sequence from such decoy sequences. Based on the average slope of the three sequence regions EDD showed twofold improvement in the ability to differentiate sequence permutations. However, further optimization is needed to validate this result, including evaluation of a range of collision energies for CID and a range of electron energies for EDD. A larger array of RNA sequences should also be characterized, including a number of common synthetic chemical modifications. In addition, a more comprehensive assessment of the permutational landscape is warranted.

3.5 References

1. Choi, H.; Nesvizhskii, A. I. False Discovery Rates and Related Statistical Concepts in Mass Spectrometry-Based Proteomics. *J. Proteome Res.* **2008,** *7*, 47-50.

2. Angel, T. E.; Aryal, U. K.; Hengel, S. M.; Baker, E. S.; Kelly, R. T.; Robinson, E. W.; Smith, R. D. Mass Spectrometry-Based Proteomics: Existing Capabilities and Futire Directions. *Chem. Soc. Rev.* **2012,** *41*, 3912-3928.

3. Hernandez, P.; Müller, M.; Appel, R. D. Automated Protein Identification by Tandem Mass Spectrometry: Issues and Strategies. *Mass Spectrom. Rev.* **2005,** *25*, 235-254.

4. Noor, Z.; Ahn, S. B.; Baker, M. S.; Ranganathan, S.; Mohamedali, A. Mass Spectrometry-Based Protein Identification in Proteomics-a Review. *Briefings Bioinf.* **2021,** *22*, 1620-1638.

5. Bereman, M. S. Tools for Monitoring System Suitability in Lc Ms/Ms Centric Proteomic Experiments. *Proteomics* **2014,** *15*, 891-902.

6. Aggarwal, S.; Yadav, A. K. False Discovery Rate Estimation in Proteomics. In *Statistical Analysis in Proteomics*, Jung, K., Ed. Springer New York: New York, NY, 2016; pp 119-128.

7. Nesvizhskii, A. I. A Survey of Computational Methods and Error Rate Estimation Procedures for Peptide and Protein Identification in Shotgun Proteomics. *J. Proteomics* **2010,** *73*, 2092-2123.

8. Kenderdine, T.; McIntyre, W.; Yassaghi, G.; Rollo, D.; Bunkowski, A.; Goerlach, L.; Suckau, D.; Tremintin, G.; Greig, M.; Bell, C.; Fabris, D. Integrating Internal Fragments in the Interpretation of Top-Down Sequencing Data of Larger Oligonucleotides. *J. Am. Soc. Mass Spectrom.* **2023**.

9. D'Ascenzo, L.; Popova, A. M.; Abernathy, S.; Sheng, K.; Limbach, P. A.; Williamson, J. R. Pytheas: A Software Package for the Automated Analysis of Rna Sequences and Modifications Via Tandem Mass Spectrometry. *Nat. Commun.* **2022,** *13*, 2424.

10. Wein, S.; Andrews, B.; Sachsenberg, T.; Santos-Rosa, H.; Kohlbacher, O.; Kouzarides, T.; Garcia, B. A.; Weisser, H. A Computational Platform for High-Throughput Analysis of Rna Sequences and Modifications by Mass Spectrometry. *Nat. Commun.* **2020,** *11*, 926.

11. Largy, E.; Ranz, M. Oligor: A Native Hdx/Ms Data Processing Application Dedicated to Oligonucleotides. *Anal. Chem.* **2023,** *95*, 10.1021/acs.analchem.3c01321.

12. Crittenden, C. M.; Lanzillotti, M. B.; Chen, B. Top-Down Mass Spectrometry of Synthetic Single Guide Ribonucleic Acids Enabled by Facile Sample Clean-Up. *Anal. Chem.* **2023,** *95*, 3180-3186.

13. Kulkarni, J. A.; Witzigmann, D.; Thomson, S. B.; Chen, S.; Leavitt, B. R.; Cullis, P. R.; Meel, R. v. d. The Current Landscape of Nucleic Acid Therapeutics. *Nat. Nanotechnol.* **2021,** *16*, 630-643.

14. Guimaraes, G. J.; Kim, J.; Bartlett, M. G. Characterization of Mrna Therapeutics. *Mass Spectrometry Reviews n/a*.

15. Sutton, J. M.; Guimaraes, G. J.; Annavarapu, V.; Dongen, W. D. v.; Bartlett, M. G. Current State of Oligonucleotide Characterization Using Liquid Chromatography–Mass Spectrometry: Insight into Critical Issues. *J. Am. Soc. Mass Spectrom.* **2020,** *31*, 1775-1782.

16. Schürch, S. Characterization of Nucleic Acids by Tandem Mass Spectrometry - the Second Decade (2004-2013): From DNA to Rna and Modified Sequences. *Mass Spectrom. Rev.* **2014,** *35*, 483-523.

17. Hannauer, F.; Black, R.; Ray, A. D.; Stulz, E.; Langley, G. J.; Holman, S. W. Review of Fragmentation of Synthetic Single-Stranded Oligonucleotides by Tandem Mass Spectrometry from 2014 to 2022. *Rapid Commun. Mass Spectrom.* **2023,** *37*, e9596.

18. Huang, T.-y.; Kharlamova, A.; Liu, J.; McLuckey, S. A. Ion Trap Collision-Induced Dissociation of Multiply Deprotonated Rna: C/Y-Ions Versus (a-B)/W-Ions. *J. Am. Soc. Mass Spectrom.* **2008,** *18*.

19. Gao, Y.; Yang, J.; Cancilla, M. T.; Meng, F.; McLuckey, S. A. Top-Down Interrogation of Chemically Modified Oligonucleotides by Negative Electron Transfer and Collision Induced Dissociation. *Anal. Chem.* **2013,** *85*, 4713-4720.

20. Wu, J.; McLuckey, S. A. Gas-Phase Fragmentation of Oligonucleotide Ions. *Int. J. Mass Spectrom.* **2004,** *237*, 197-241.

21. Gao, Y.; McLuckey, S. A. Fragmentation Reactions of Nucleic Acid Ions in the Gas Phase. In *Nucleic Acids in the Gas Phase*, 1 ed.; Gabelica, V., Ed. Springer: Berlin, Heidelberg, 2014; pp 131-182.

22. Hu, B.; Zhong, L.; Weng, Y.; Peng, L.; Huang, Y.; Zhao, Y.; Liang, X.-J. Therapeutic Sirna: State of the Art. *Signal Transduction Targeted Ther.* **2020,** *5*, 101.

23. Coon, J. J.; Syka, J. E. P.; Shabanowitz, J.; Hunt, D. F. Tandem Mass Spectrometry for Peptide and Protein Sequence Analysis. *BioTechniques* **2018,** *38*, 519-523.

24. Taucher, M.; Breuker, K. Top-Down Mass Spectrometry for Sequencing of Larger (up to 61 Nt) Rna by Cad and Edd. *J. Am. Soc. Mass Spectrom.* **2010,** *21*, 918-929.

25. Yang, J.; Mo, J.; Adamson, J. T.; Hakansson, K. Characterization of Oligodeoxynucleotides by Electron Detachment Dissociation Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2005,** *77*, 1876-1882.

26. Kinet, C.; Gabelica, V.; Balbeur, D.; Pauw, E. D. Electron Detachment Dissociation (Edd) Pathways in Oligonucleotides. *Int. J. Mass Spectrom.* **2009,** *283*, 206-213.

27. Santos, I. C.; Lanzillotti, M.; Shilov, I.; Basanta-Sanchez, M.; Roushan, A.; Lawler, R.; Tang, W.; Bern, M.; Brodbelt, J. S. Ultraviolet Photodissociation and Activated Electron Photodetachment Mass Spectrometry for Top-Down Sequencing of Modified Oligoribonucleotides. *J. Am. Soc. Mass Spectrom.* **2022,** *33*, 510-520.

28. Peters-Clarke, T. M.; Quan, Q.; Brademan, D. R.; Herbert, A. S.; Westphall, M. S.; Coon, J. J. Ribonucleic Acid Sequence Characterization by Negative Electron Transfer Dissociation Mass Spectrometry. *Anal. Chem.* **2020,** *92*, 4436-4444.

29. Budnik, B. A.; Haselmann, K. F.; Zubarev, R. A. Electron Detachment Dissociation of Peptide Di-Anions: An Electron-Hole Recombination Phenomenon. *Chemical Physics Letters* **2001,** *342*, 299-302.

30. Kjeldsen, F.; Hørning, O. B.; Jensen, S. S.; Giessing, A. M. B.; Jensen, O. N. Towards Liquid Chromatography Time-Scale Peptide Sequencing and Characterization of Post-Translational Modifications in the Negative-Ion Mode Using Electron

Detachment Dissociation Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2008,** *19*, 1156-1162.

31. Papanastasiou, D.; Kounadis, D.; Lekkas, A.; Orfanopoulos, I.; Mpozatzidis, A.; Smyrnakis, A.; Panagiotopoulos, E.; Kosmopoulou, M.; Reinhardt-Szyba, M.; Fort, K.; Makarov, A.; Zubarev, R. A. The Omnitrap Platform: A Versatile Segmented Linear Ion Trap for Multidimensional Multiple-Stage Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2022,** *33*, 1990-2007.

32. Karasawa, K.; Duchoslav, E.; Baba, T. Fast Electron Detachmentn Dissociation of Oligonucleotides in Electron-Nitrogen Plasma Stored in Magneto Radio-Frequency Ion Traps. *Anal. Chem.* **2022,** *94*, 15510-15517.

33. Zubarev, R. A.; Zubarev, A. R.; Savitski, M. M. Electron Capture/Transfer Versus Collisionally Activated/Induced Dissociations: Solo or Duet? *J. Am. Soc. Mass Spectrom.* **2008,** *19*, 753-761.

34. Zhao, B.; Reilly, C. P.; Davis, C.; Matouschek, A.; Reilly, J. P. Use of Multiple Ion Fragmentation Methods to Identify Protein Cross-Links and Facilitate Comparision of Data Interpretation Algorithms. *J. Proteome Res.* **2020,** *19*, 2758-2771.

35. Jr., J. D. H. W. M. J. P. S. C. M. S. R. K. P. Y. J. E. S. J. L. S. In *Optimized Fragmentation of Oligonucleotides Supresses Undesired Fragmentation Products and Enables Confident Sequence Assignment*, Proceedings of the 71th ASMS Conference on Mass Spectrometry and Allied Topics, Huston, TX, June 4-8, 2023.

36. Colaert, N.; Degroeve, S.; Helsens, K.; Martens, L. Analysis of the Resolution Limitations of Peptide Identification Algorithms. *J. Proteome Res.* **2011,** *10*, 5555-5561.

37. Hakansson, C. W. S. K. In *High Temperature Fluoroalcohol-Free Liquid Chromatography-Electron Detachment Dissociation Tandem Mass Spectrometry of Rna*, Proceedings of the 71th ASMS Conference on Mass Spectrometry and Allied Topics, Huston, TX, June 4-8, 2023.

38. McFarland, M. A.; Marshall, A. G.; Hendrickcon, C. L.; Nilsson, C. L.; Fredman, P.; Månsson, J.-E. Structural Characterization Os the Gm1 Ganglioside by Infrared Multiphoton Dissociation, Electron Capture Dissociation, and Electron Detachment Dissociation Electrospary Ionization Ft-Icr Ms/Ms. *J. Am. Soc. Mass Spectrom.* **2005,** *16*, 752-762.

39. Wolff, J. J.; Amster, I. J.; Chi, L.; Linhardt, R. J. Electron Detachment Dissociaiton of Glycosaminoglycan Tetrasaccharides. *J. Am. Soc. Mass Spectrom.* **2007,** *18*, 234-244.

40. Boldin, I. A.; Nikolaev, E. N. Fourier Transform Ion Cyclotron Resonance Cell with Dynamic Harmonization of the Electric Field in the Whole Volume by Chaping of

the Excitation and Detection Electrode Assembly. *Rapid Commun. Mass Spectrom.* **2011,** *25*, 122-126.

41. Pepi, L. E.; Sasiene, Z. J.; Mendis, P. M.; Jackson, G. P.; Amster, I. J. Structural Characterization of Sulfated Glycosaminoglycans Uisng Charge-Transfer Dissociation. *J. Am. Soc. Mass Spectrom.* **2020,** *31*, 2143-2153.

42. Nikolaev, E. N. Some Notes About Ft Icr Mass Spectrometry. *Int. J. Mass Spectrom.* **2015,** *377*, 421-431.

Chapter 4 High Temperature Fluoroalcohol-Free Liquid Chromatography-FT-ICR Mass Spectrometry of RNA up to 100 kDa

4.1 Introduction

Electrospray ionization mass spectrometry (ESI-MS) has the ability to form and detect gaseous ions via desolvation directly from solution for large biomolecular analytes[1], including biomolecule-ligand complexes.[2, 3] However, in contrast to the widespread application of ESI-MS towards protein analysis, nucleic acid analysis has historically lagged behind due to a number of additional challenges. These challenges include preferred ionization in negative ion mode, which involves increased electronic noise in many mass spectrometer types;[4] the higher gas-phase lability of the DNA/RNA backbone;[5] and the higher propensity for salt adduction, which complicates mass spectral interpretation, dilutes available signal, and increases demand for high resolution MS.[6] The challenge of non-volatile cation adduction to the phosphate backbone is exacerbated with increasing oligonucleotide length[7].

Nevertheless, numerous examples of moderate to large molecular weight DNA/RNA analysis by ESI-MS have been reported with many involving Fourier transform ion cyclotron resonance (FT-ICR) mass analyzers, which offer the highest possible resolution. For example, direct confirmation of RNA-ligand binding in a multiplexed assay was demonstrated by Ibis Therapeutics with ESI-FT-ICR MS.[2] Targeting of RNA with small molecule ligands continue to be an important strategy for implementation in disease therapeutics[8] with linezolid[9] and neomycin[10] being recent

examples. The latter example extended the application of direct infusion ESI-FT-ICR MS to a 13 kDa RNA-ligand complex.[10] Muddiman and co-workers showed that nanoelectrospray ionization via a fused silica capillary emitter with an inner diameter in the range of 10 to 50 μm[11] allowed FT-ICR MS analysis of an ~300 kDa PCR product, showing adequate charge state resolution.[12]

For quality control applications, liquid chromatography (LC) is typically interfaced with MS analysis to improve quantitation and detection of low level impurities.[13] Hyphenated LC-MS analysis also allows for improved control of cation adduction in oligonucleotide workflows.[14] A wide variety of mobile phase additives have been investigated for ion-pairing reverse phase (IP-RP) liquid chromatography of oligonucleotides,[14] including triethylammonium acetate (TEAA) with a polystyrene divinylbenzene stationary phase for 110 nucleotide RNA.[15] Recent investigations have shown direct detection and characterization of ~100 nucleotide RNA, including synthetic guide RNA, via size exclusion like desalting/C18 chromatography.[16] The more conventional 15 mM triethylamine (TEA)/400 mM hexafluoroisopropanol (HFIP)[17] mobile phase has also been extended towards larger RNA (~32 kDa).[18] Birdsall et al. demonstrated that, in such workflows, metal adduct mitigation strategies should consider both sample preparation (e.g., ion exchange, molecular cut-off filters, metal chelators) and LC conditions.[6] The latter considerations include elimination of trace metals in mobile phase solvents (even LC/MS grade), mobile phase additives, liquid reservoirs, and flow paths. These authors showed that a low pH reconditioning step following each LC/MS run can avoid metal ion build-up over time.[6] Fabris and co-workers showed that cation mitigation may also be achieved by reduction of ESI emitter

diameter, due to reduced cation population available in smaller droplets.[19] With salt management strategies and improved bioinformatic utilities detection and sequencing of RNA in the 100-300 nucleotide range is becoming more routine.[16, 20, 21]

While orbitrap[18] and time-of-flight[22] (TOF) mass analyzers are widely available and can offer high resolution, large nucleic acid analysis is pushing their limits. The higher acquisition rate of TOF-based instruments allows improved sampling across chromatographic peaks; however, typically at lower resolution than with frequency-based measurements, such as orbitrap and FT-ICR. Phase correction of time-domain signals for absorption-mode processing can improve resolution of Fourier transform mass spectrometry,[23] including large biomolecule analysis.[24] Thus, reduced acquisition time is needed for a given resolution, which improves chromatographic compatability.[25] Another solution to maximize MS resolution is to purposely widen chromatographic peaks to meet the required sampling frequency, so called “peak parking”.[26] For frequency-based instruments, the required time-domain transient length for a given resolution can be computationally predicted.[27] Nevertheless, challenges remain in isotopically resolving large nucleic acids on an LC time scale.

In addition, despite approaches such as the convex gradient program,[28] LC conditions for larger nucleic acids have further room for improvement in ion exchange, hydrophilic interaction[29] and ion pairing implementations. The progression of oligonucleotide LC-MS solvents from TEAA, triethyl ammonium bicarbonate (TEAB)[30] to TEA-HFIP[31] has shown increasing sensitivity and reported accessible charge states[32] but at the cost of decreasing environmental compatibility. The latter issue remains with alternative fluoroalcohols.[33] Here, we demonstrate that in vitro-transcribed RNA up to

~100 kDa can be detected via fluoroalcohol-free (TEAB-based) LC/FT-ICR MS with isotopic resolution at 40 kDa.

4.2 Methods

4.2.1 Materials and Sample Preparation

RNA (112 and 300 nucleotides (nts), respectively) was in vitro transcribed with T7 polymerase and purified by strong anion ion exchange chromatography and size exclusion chromatography by the Koutmous group. Prior to LC analysis, RNA samples were further desalted via MWCO (Sigma, St. Louis, MO) filtration with either LiChropur ammonium acetate or ammonium formate (<= 5 mg/kg) (Sigma) prepared at 100 mM in polypropylene 15 mL tubes. Optima LC-MS grade water (≤20 ppb Na and ≤10 ppb K; ThermoFisher (Pittsburgh, PA) or Riedel-de Haen TraceSELECT water (<5 ppb Na and <5 ppb K)) in a polypropylene cubic container purchased from Thomas Scientific (Swedesboro, NJ) was used for the 112 nt and 300 nt samples respectively. Molecular weight cut off filter desalting was performed five times sequentially in the same filter.

4.2.2 LC-UV and LC-FT-ICR MS Analysis

LC separation was performed with an Agilent 1100 HPLC at a flow rate of 0.100-0.400 mL/min (0.300 mL/min for LC-UV). An Agilent PRLP-S polystyrene-divinylbenzene 2.1x50 mm, 5 µm, 1000 Å, stainless steel column was used at 80 ˚C with a 100 or 5 mM LiChropur TEAB (Sigma)/methanol (ThermoFisher Pittsburgh, PA) gradient. For UV detection an Agilent 1100 variable wavelength detector with a micro flow cell (Santa Clara, CA) was used. When separating the 112 and 300 nt samples for chromatographic resolution at ion pair concentration of 100 mM TEAB a 2 µL an

injection of a mixture loaded 0.4 and 0.2 µg 112 and 300 nt on the column. For LC-FT-ICR MS analysis, eluent was directed through an Agilent ES standard nebulizer at a standard needle position flush with the nebulizer on a SolariX Q-FT-ICR (Bruker, Billerica, MA). When acquiring quadrupole isolation data of the 112 nt and 300 nt samples 5 µg and 2 µg of material was loaded on column, respectively. The ESI source was operated in negative ion mode at 4.0 kV with an endplate offset of -500 V at 350 ˚C drying gas temperature, 10 L/min drying gas flow rate, and 3 bar nebulizer pressure. Data were acquired in serial mode with 1 (for 112 nt RNA) or 4 (for 300 nt RNA) scan averaging, 0.4 second (for 112 nt RNA) and 4 second (for 300 nt RNA) external accumulation time, and 2 megaword acquisition size. To limit space charge in the FT-ICR cell, a quadrupole isolation window of 35 and 20 m/z was used, resulting in detection of one charge state (35-) for the 112 nt RNA and three charges states (95-, 96-, and 97-) for the 300 nt RNA. Excitation was performed in broadband mode with the frequency sweep in either increasing or decreasing frequency direction (for comparison) for excite calculation version 1 or version 2 in Bruker ftmsControl v2.3.0.

4.2.3 Data Analysis

The total ion chromatograms were integrated by the version 3.0 MS algorithm in Bruker DataAnalysis 5.0 with sensitivity at 95%, area threshold at 20%, and the mass spectrum calculation set to generate profile spectra only, which populated the compound spectra of the analysis tree. The resulting profile mass spectra were smoothed by a Gaussian algorithm with a width of 0.05 m/z for the 112 nt RNA and 0.02 m/z for the 300 nt RNA. The “Apex” peak finder was used with signal-to-noise set to 10 and relative intensity set to 20% by changing the instrument type in the settings. Charge deconvolution was

implemented with the “related-ion” option with “protein/large molecule” selected to generate a true mass spectrum.

For the 112 nt RNA, the “resolved ion” deconvolution algorithm was also run using automated charge detection. Acquired LC-MS data were processed in Bruker ftmsProcessing 2.3. The five most abundant scans (scans 323-327 for the 112 nt RNA) were selected for post processing. The “batch sum serial free induction decays” (FIDs) tool was used to co-add the corresponding time-domain transients into a single transient. Sine apodization is the default in Compass MRMS 4.0 prior to magnitude mode processing. A half sine apodization was also applied to the FID prior to absorption mode processing (AMP). Using “AMP control 3.3” (within ftmsControl) the selectively co-added FIDs were phased using an AMP center mass of 1040 m/z, corresponding to the 35- charge state. The software then searched the FID and applied phase windows to the data for the absorption mode fast Fourier transform to generate the absorption mode mass spectrum.

4.3 Results

4.3.1 Fluoroalcohol-free LC-UV Optimization

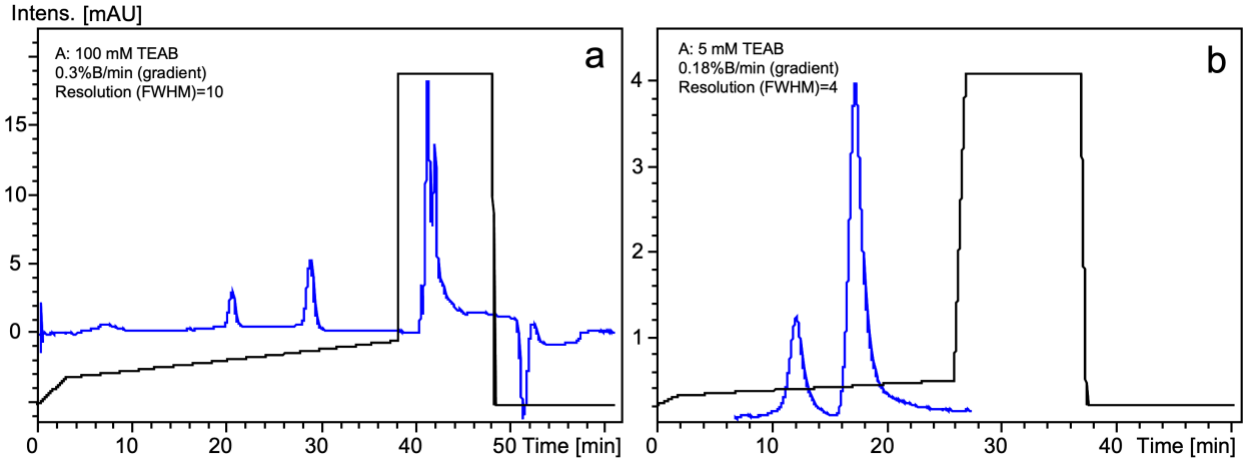


Figure 4.1: LC-UV analyses of a 40 kDa (112 nt) and a 100 kDa (300 nt) RNA with a 2.1x50 mm, 5 µm, 1000 Å PLRP-S column and a TEAB/methanol gradient at 80 ℃. The blue lines show the 260 nm wavelength signal and the black lines show the %B gradient for which the minimum was 5% B and the maximum was 100% B. Solvent A was either 100 mM TEAB (a) or 5 mM TEAB (b).

To explore whether fluoroalcohol-free LC conditions would be suitable for larger RNA separations, we first applied the common IP-RP approach of 100 mM ion pairing reagent (TEAB) and a shallow gradient with a front end ramp for LC/UV analysis. Under these conditions, we observed a FWHM resolution of 10 for the 40 kDa (112 nt) and 100 kDa (300 nt) RNAs, eluting at ~20.8 and ~29.0 minutes, respectively (**Figure 4.1a)**. The FWHM of the analytes at 100 mM TEAB was 0.7 and 0.8 min for the 112 and 300 nt respectively.

Because 100 mM TEAB may lead to significant salt adduction and accompanying signal dilution with MS detection, along with an increased rate of ion source contamination, we also explored the lower IP concentration of 5 mM TEAB. At this lower ion pair concentration, the slope of the gradient had to be reduced by a factor of ~2 to obtain comparable resolution, 4, between the 40 and 100 kDa RNAs (**Figure 4.1b**). Under these conditions, the two RNAs eluted at ~12.3 and ~17.4 minutes, respectively with fairly broad elution windows (1.2 minutes FWHM for both). These data suggest that

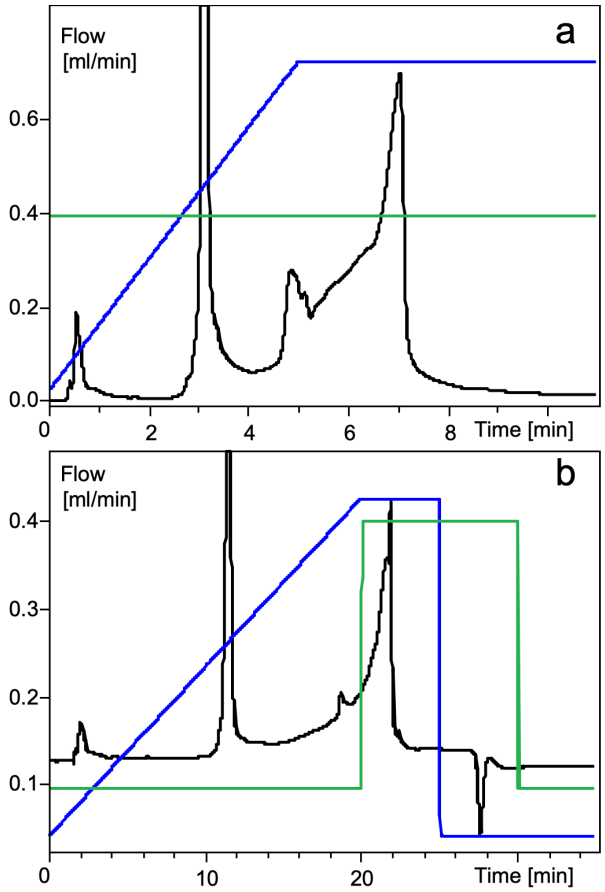


Figure 4.2: LC-UV analyses of a 112 nt RNA. A scouting gradient at 5 mM TEAB/methanol resulted in a narrow chromatographic peak width (a). A four fold flow rate reduction showed similar chromatographic results (b) but allows ~3 fold longer sampling time for high resolution MS due to the wider elution window. Black lines show the 260 nm wavelength signal, blue lines show the %B gradient, and green lines show the mobile phase flow rate.

adequate LC separation can be achieved at the more MS-friendly 5 mM TEAB concentration.

To further optimize the 5 mM TEAB conditions for an overall faster method, a steeper scouting gradient over a shorter time, 5 minutes, was applied for the 112 nt RNA (**Figure 4.2.a**). This gradient resulted in a significantly earlier elution time, 2.8 min, with a narrow, 0.06 minutes (FWHM), elution window. However, as discussed above, a wider elution window may be desired to improve spectral quality with frequency-based MS detection. To provide more time for higher resolution MS scans, a flow rate drop from 400 to 0.100 mL/min and a four-fold gradient slope decrease was applied. These conditions resulted in a three-fold wider LC peak (FWHM) of 0.18 minutes at an elution time 11.5 min (**Figure 4.2 b**). The flow rate was increased to previous levels for washing and reequlibration in order to not significantly lengthen the time between sample injections. However, the flow rate jump does add pressurization equilibration time before and after the wash. These times need to be calibrated for in the corresponding method. We found that depressurization equilibration

required ~4 minutes. Based on the chromatogram in **Figure 4.2 b**, an ~2.5 minute hold at 100% B appeared sufficient to adequately wash the system's flow path for subsequent experiments.

4.3.2 Fluoroalcohol-Free LC-FT-ICR MS of 112 nt RNA

LC separation of the 112 nt RNA, as depicted in **Figure 4.2 b**, was performed and the eluent was directed to a 7 T FT-ICR mass spectrometer operating in negative ion mode. The mass spectrum resulting from the ten most abundant scans across the LC elution time is shown in **Figure 4.3 a**. The corresponding charge state distribution

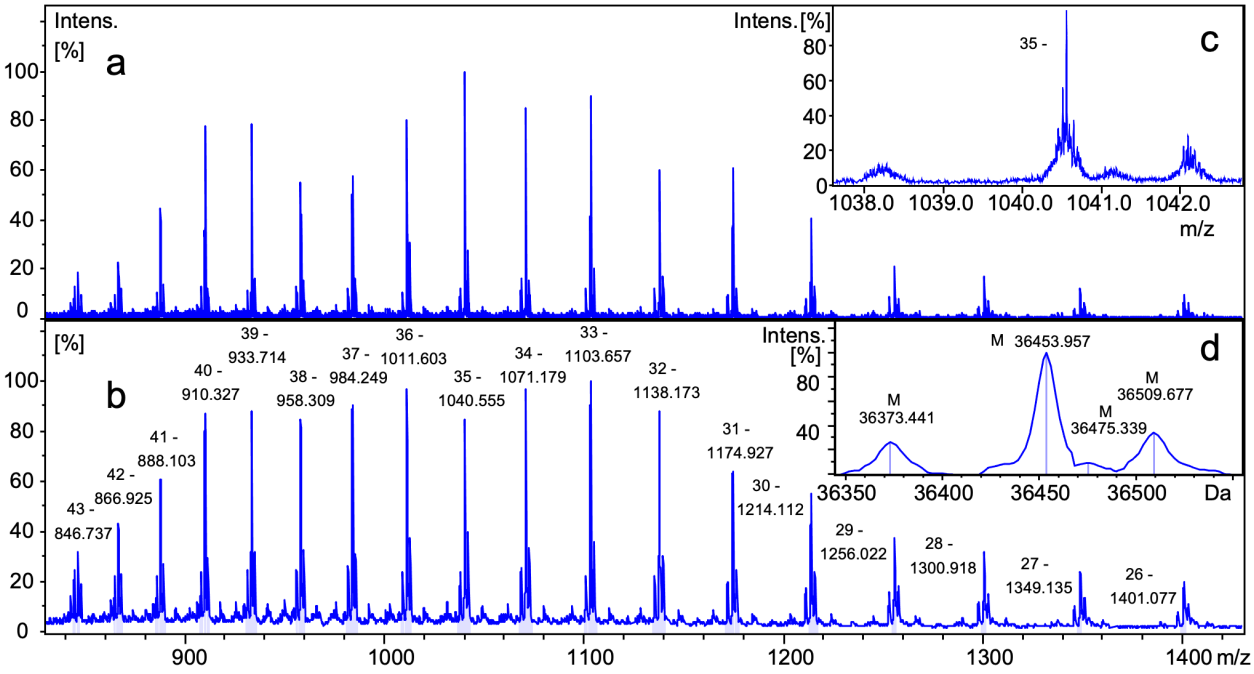


Figure 4.3: Charge state distribution of a 40 kDa RNA following LC-FT-ICR MS with a 5 mM TEAB/methanol gradient. An average of ~10 scans of unprocessed mass spectra (a). Average spectrum following Gaussian smoothing and charge state assignment (b). Zoomed in view of the 35- charge state prior to smoothing (c), and deconvoluted mass spectrum (d) showing 4 primary analyte signals.

contains a broad range of charge states above 20% relative abundance in the m/z range 830-1430. Each charge state shows 3-4 resolved isotope clusters (Figure 3.3 c for the 35- charge state); however, the corresponding isotopic distributions were poorly resolved. The unresolved isotopologues are likely due to space charge effects in the

FT-ICR cell with too many charges present during excitation and detection.[34, 35] Thus, the broadband FT-ICR spectrum was subjected to smoothing (**Figure 4.3 b**) for more accurate low resolution charge state deconvolution. The deconvoluted mass spectrum is shown in **Figure 4.3 d**. The “related ion” deconvolution algorithm assigned charge states from 26- to 43- for signals >20% relative abundance, corresponding to a mass ~36 kDa. The four features observed for the charge states -33 through -39 are present in the deconvoluted spectrum (**Figure 4.3 d**). The mass differences between these peaks suggest the primary signal appears at 36,454 Da with a relatively low abundance sodium adduct at +22 Da, an iron adduct at +55 Da, and a phosphate loss at -80 Da. The formation of iron adducts is a process known to occur in liquid chromatography flow paths.[36-38] The phosphate loss may be from a terminal phosphate on the RNA. A shoulder on the potential iron adduct peak may correspond to a potassium adduct (+38 Da).

In order to improve MS resolution for this RNA, the same chromatographic conditions were applied and the eluent was directed into the 7 T FT-ICR instrument with the quadrupole set to isolate the 35- charge state, which was abundant in the broadband spectrum (**Fig 4.3 a**). The resulting total ion chromatogram is shown in

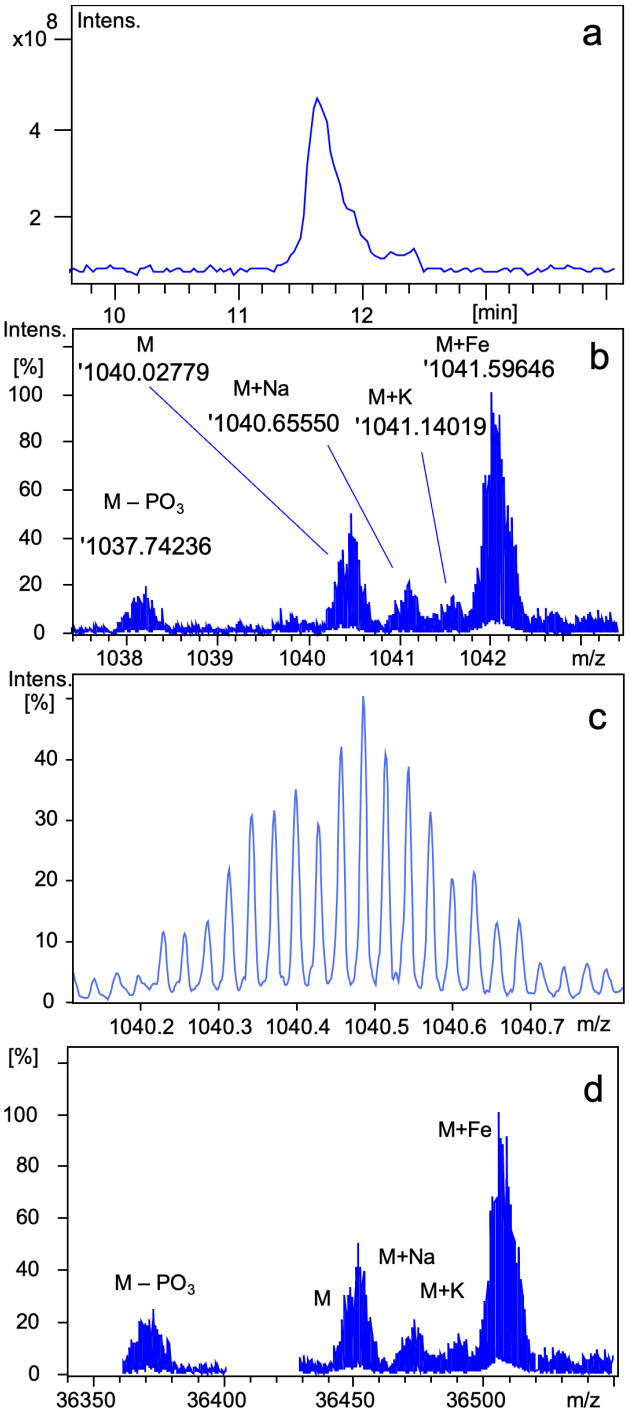


Figure 4.4 a and the mass spectrum from the integrated chromatogram is shown in **Figure 4.4 b** A zoomed-in view of the 36,454 Da species (**Figure 4.4 c** reveals nearly baseline isotopologue resolution at a resolving power ~110k. With this greatly improved spectral quality, the adduct features at +22, +39, and +55 Da as well as the potential phosphate loss at -80 Da are clearly resolved and detected with the deisotoping tool SNAP[39] to provide the corresponding monoisotopic mass.. The potassium adduct was not resolved in the "related ion" deconvolution following spectral smoothing (Fig. 3.3 d). The spacing between the isotopologue peaks above 8% relative intensity in Fig. 3.4c is 0.0285±6, which supports the assigned charge state of 35-. Isotope-resolved deconvolution is shown in **Figure 4.4 d**. An inter-run comparison of broadband detection (Fig. 3.3 d) and detection following quadrupole isolation (Fig. 3.4 d) shows a

Figure 4.4 LC-FT-ICR MS with quadrupole isolation of the 35- charge state of a 40 kDa RNA with a 5 mM TEAB/methanol gradient. Total ion chromatogram (a), corresponding high resolution mass spectrum (b), zoomed in view of the signal with identified monoisotopic m/z 1040.03 (labeled "M"; c), and mass spectrum following isotopically resolved deconvolution (d).

relative abundance increase of the iron adduct. These results from a 40 kDa RNA demonstrate some of the challenges associated with making high resolution MS measurements of larger nucleic acids, including the development of a customized chromatographic method with quadrupole isolation of a single charge state. Alternative approaches may include FT-ICR MS at higher magnetic field strength,[40] higher order image current detection geometry (e.g., two omega),[41, 42] and/or improved ion population control[35]. However, without changing any experimental conditions, absorption mode processing of FT-ICR MS data is another route to improved resolution.

4.3.3 Absorption Mode Processing and Extension to 100 kDa RNA

Absorption mode processing was explored for the time domain FT-ICR MS data of the 40 kDa RNA. Five transients were co-added to generate the summed FID in **Figure 4.5 a**. The exponentially decaying isotopic beat pattern of the isolated 35- charge state is clearly observed with five isotopic beats over the 2 s detection time. The default magnitude mode setting in the Bruker software applies a sine apodization function, **Figure 4.5 b**, to mitigate edge artifacts in the subsequent discreet fast Fourier transform (FFT). The resulting magnitude mode frequency domain spectrum is shown in Figure 3.5 e (black line). One advantage of a magnitude mode FFT is that the resulting frequency spectrum lacks negative components below the zero baseline. However, an absorption mode FFT can yield such negative features due to

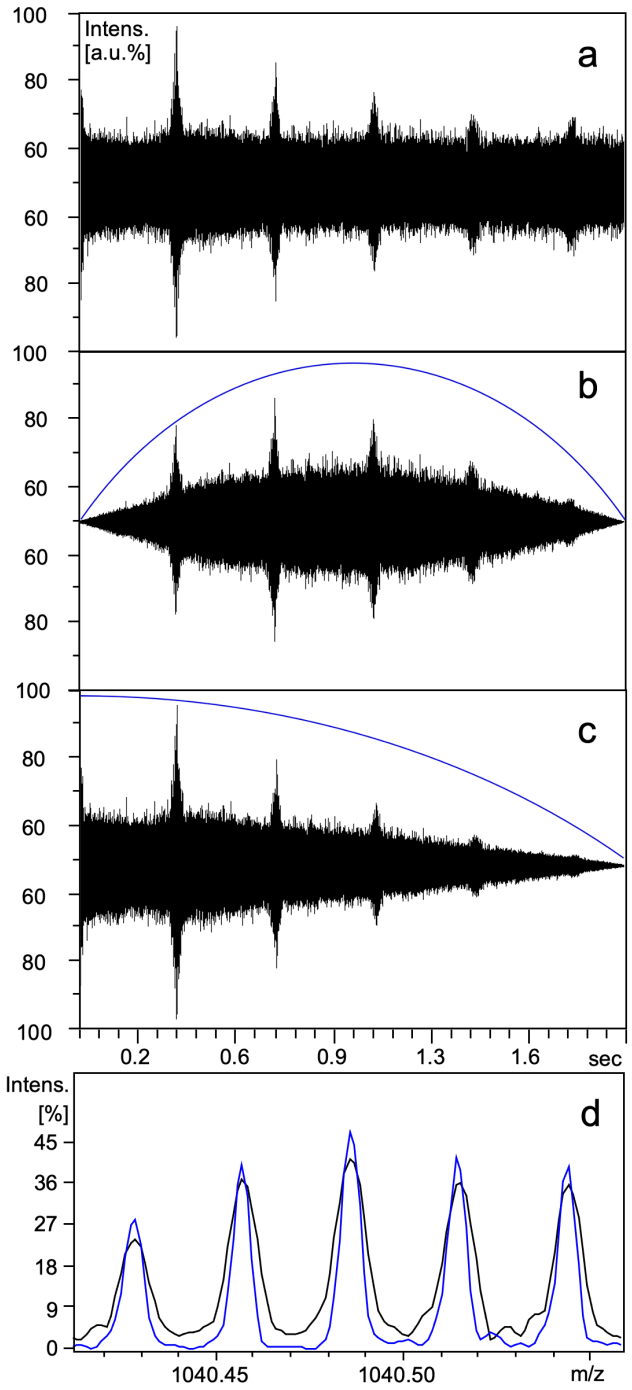


Figure 4.5 Time domain transient of selectively co-added FIDs acquired in serial mode acquisition LC-MS of a 112 nt RNA (a). A representative default magnitude mode “full sine” apodization function overlayed on the apodized transient (b). A representative default absorption mode “half-sine” apodization function overlayed on the apodized transient (c). The resulting mass spectra following magnitude mode processing (black line) and absorption mode processing (blue line), respectively (d) with absorption mode showing increased resolving power.

improper phasing of the time-domain data.[43] The default absorption mode half-sine apodization function of the time domain data is shown in **Figure 4.5 c**. The resulting frequency domain spectrum (**Figure 4.5d**, blue line) shows increased resolving power ~180k-190k, compared with ~110k-120k in magnitude mode with only minor negative spectral features. A higher SNAP quality factor, 0.96, was noted for isotopic distribution fitting to the absorption mode data compared with 0.91 for the magnitude mode spectrum. An increased peak height is also visually apparent in absorption mode; however, the SNAP algorithm computes a decreased signal-to-noise ratio from 95 to 85 for the magnitude and absorption mode spectra, respectively. This discrepancy may be due to the absence of additional background subtraction.

The larger 300 nucleotide RNA was subjected to LC-FT-ICR MS with the same mobile phase as in Figure 3.4 , i.e., 5 mM TEAB and methanol. The resulting broadband mass spectrum across the elution window is shown in Figure 3.6 a. Similar to the 112 nt RNA, resolving power is lower than desired in broadband mode although individual charge states from 53 to 125 are resolved. Thus, the analysis was repeated with a 20 m/z quadrupole isolation window centered on the 96- charge state. This isolation width resulted in co-isolation of the neighboring 97- and 95- charge states. However, contrary to the 112 nt RNA, isotopic resolution was not achieved. Thus, Gaussian smoothing was performed at a width of 0.02 m/z. The corresponding smoothed mass spectrum, revealing three molecular features per charge state, is shown in Figure 3.6 b. The "related ion" deconvolution process resulted in the true mass spectrum shown in **Figure 4.6 c**. Observed mass differences suggest that the left feature corresponds to phosphate loss (-80 Da) and the right feature corresponds to an iron adduct (+55),

similar to the 112 nt RNA while potential sodium and potassium adducts are unresolved. To explore whether isotopic resolution is feasible at 7 Tesla, Spectroswiss' FTMS Simulator was used to generate a theoretical time-domain signal for the 80- charge state at m/z 1,210.[27] This simulation showed the presence of four isotopic beats in four seconds (**Figure 4.6 d**). This transient length corresponds to 4 megaword acquisition size, which is possible with standard instrument settings for 7T dipolar detection at 4 megaword acquisition size.

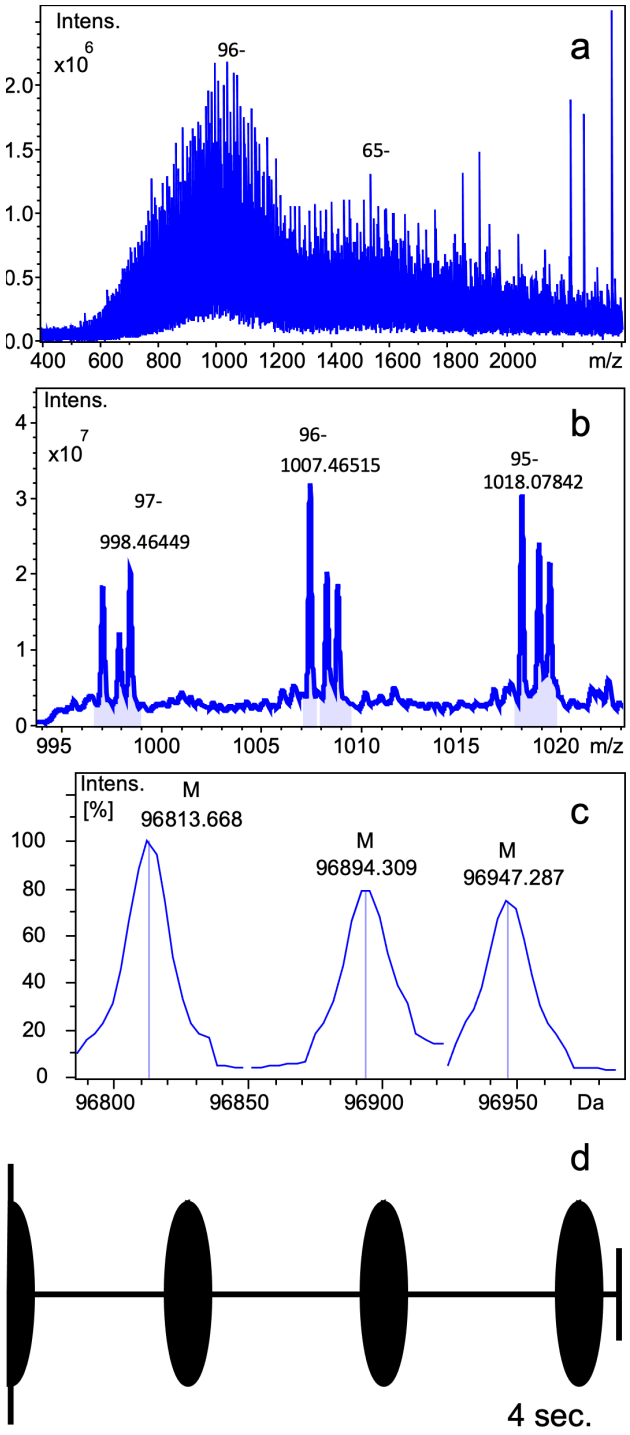


Figure 4.6 LC-FT-ICR MS of a 100 kDa RNA with a 5 mM TEAB/methanol gradient. An average of 10 scans of unprocessed broadband mass spectra (a), mass spectrum following quadrupole isolation of the 95-, 96-, and 97- charge states (b), deconvoluted mass spectrum (c), and simulated time domain isotopic beat pattern for an 80- charge state at 7 Tesla (d).

4.4 Conclusion

While we are excited about the achieved data, there is still room to optimize various experimental parameters, i.e. , isotopic resolution for the larger RNA at 7 Tesla is not out of reach, as indicated in Figure 4.6 d. The current sample preparation and LC

methods yield a low abundance of cation adducts; however, the common iron adduct was observed to grow with time. The main metallic surface area in the flow path is the interior of the column. The utilized column type has an option with a PEEK lined surface, which may be worth exploring. On the other hand, the column surface as the major source of iron does not fully explain the interun increase in iron adduction. We hypothesize that this observed increase originates from the post-column flow path, consisting of a post UV detector stainless steel valve and the stainless steel ESI needle. This portion of the flow path is stagnant when flow is diverted to waste during column washing. In addition, electrolytic activity and high temperature during ESI may further contribute. This portion of the flow path may need more constant flushing between runs. Use of a smaller emitter inner diameter, e.g. the Agilent micronebulizer, may also provide a solution, however, such implementation would require a flow rate decrease.

On the mass spectrometry side, selection of a higher charge state, e.g., 110- with lower m/z ratio where quadrupole isolation efficiency is improved may allow for overall less charge to enter the ICR cell as co-isolation of neighboring charge states may be avoided. Such reduced space charge would decrease the decay rate of the time domain signal towards the goal of a longer detection time. The shimming parameters of the ParaCell were also not explored in depth, i.e., such optimization may further reduce the transient ring down time. With higher quality data, absorption mode processing for longer RNA may also be feasible for further increase in resolution at a given transient length.

4.5 References

1. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospary Ionization for Mass Spectrometry of Large Biomolecules. *Science* **1989,** *246*, 64-71.

2. Sannes-Lowery, K. A.; Drader, J. J.; Griffey, R. H.; Hofstadler, S. A. Fourier Transform Ion Cyclotron Resonance Mass Spectrometry as a High Throughput Affinity Screen to Identify Rna Binding Ligands. *TrAC, Trends Anal. Chem.* **2000,** *19*, 481-491.

3. Thomas, J. R.; Hergenrother, P. J. Targeting Rna with Small Molecules. *Chem. Rev.* **2008,** *108*, 1171-1224.

4. Liigand, P.; Kaupmees, K.; Haav, K.; Liigand, J.; Leito, I.; Girod, M.; Antoine, R.; Kruve, A. Think Negative: Finding the Best Eelctrospray Ionization/Ms Mode for Your Analyte. *Anal. Chem.* **2017,** *89*, 5665-5668.

5. Pourshahian, S. Therapeutic Oligonucleotides, Impurities, Degradants, and Their Characterization by Mass Spectrometry. *Mass Spectrom. Rev.* **2019,** *40*, 73-157.

6. Birdsall, R. E.; Gilar, M.; Shion, H.; Yu, Y. Q.; Chen, W. Reduction of Metal Adducts in Oligonucleotide Mass Spectra in Ion-Pair Reversed-Phase Chromatography/Mass Spectrometry Analysis. *Rapid Commun. Mass Spectrom.* **2016,** *30*, 1667-1679.

7. Muddiman, D. C.; Smith, R. D. Sequencing and Characterization of Larger Oligonucleotides by Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Rev. Anal. Chem.* **1998,** *17*, 1-68.

8. Arjmand, F.; Afsan, Z.; Sharma, S.; Parveen, S.; Yousuf, I.; Sartaj, S.; Siddique, H. R.; Tabassum, S. Recent Advances in Matallodrug-Like Molecules Targeting Non-Coding Rnas in Cancer Chemotherapy. *Coordination Chemistry Reviews* **2019,** *387*, 47-59.

9. Warner, K. D.; Hajdin, C. E.; Weeks, K. M. Principals for Targeting Rna with Drug-Like Small Molecules. *Nat. Rev. Drug Discovery* **2018,** *17*, 547-558.

10. Heel, S. V.; Bartosik, K.; Juen, F.; Kreutz, C.; Micura, R.; Breuker, K. Native Top-Down Mass Spectrometry Uncovers Two Distinct Binding Botifs of a Functional Neomycin-Sensing Riboswitch Aptamer. *J. Am. Chem. Soc.* **2023,** *145*, 15284-15294.

11. Hannis, J. C.; Muddiman, D. C. Nanoelectrospray Mass Spectrometry Using Non-Metalized, Tapered (50 → 10 Um) Fused-Silica Capillaries. *Rapid Commun. Mass Spectrom.* **1998,** *12*, 443-448.

12. Muddiman, D. C.; Null, A. P.; Hannis, J. C. Precise Mass Measurement of a Double-Stranded 500 Base-Pair (309) Kda Polymerase Chain Reaction Product by Negative Ion Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **1999,** *13*, 1201-1204.

13. Sutton, J. M.; Guimaraes, G. J.; Annavarapu, V.; Dongen, W. D. v.; Bartlett, M. G. Current State of Oligonucleotide Characterization Using Liquid Chromatography–Mass Spectrometry: Insight into Critical Issues. *J. Am. Soc. Mass Spectrom.* **2020,** *31*, 1775-1782.

14. Goyon, A.; Yehl, P.; Zhang, K. Characterization of Therapeutic Oligonucleotides by Liquid Chromatography. *J. Pharm. Biomed. Anal.* **2020,** *182*, 113105.

15. Shiba, Y.; Masuda, H.; Watanabe, N.; Ego, T.; Takagaki, K.; Ishiyama, K.; Ohgi, T.; Yano, J. Chemical Synthesis of a Very Long Oligoribonucleotide with 2-Cyanoethoxymethyl (Cem) as the 2'-O-Protecting Group: Structural Identification and Biological Activity of a Synthetic 110mer Precursor-Microrna Candidate. *Nucleic Acids Res.* **2007,** *35*, 3287-3296.

16. Crittenden, C. M.; Lanzillotti, M. B.; Chen, B. Top-Down Mass Spectrometry of Synthetic Single Guide Ribonucleic Acids Enabled by Facile Sample Clean-Up. *Anal. Chem.* **2023,** *95*, 3180-3186.

17. Apffel, A.; Chakel, J. A.; Fisher, S.; Lichtenwalter, K.; Hancock, W. S. Analysis of Oligonucleotides by Hplc-Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **1997,** *69*, 1320-1325.

18. Wei, B.; Wang, J.; Cadang, L.; Goyon, A.; Chen, B.; Yang, F.; Zhang, K. Development of an Ion Pairing Reversed-Phase Liquid Chromatography-Mass Spectrometry Method for Characterization of Clustered Regularly Interspaced Short Palindromic Repeat Guide Ribonucleic Acid. *J. Chromatogr. A* **2022,** *1665*.

19. Kenderdine, T.; Xia, Z.; Williams, E. R.; Fabris, D. Submicrometer Nanospray Emitters Provide New Insights into the Mechanism of Cation Adduction to Anionic Oligonucleotides. *Anal. Chem.* **2018,** *90*, 13541-13548.

20. Kenderdine, T.; Yassaghi, G.; Daniele, R.; Bunkowski, A.; Pengelley, S.; Suckau, D.; Tremintin, G.; Greg, M.; Bell, C.; Fabris, D. In *Evaluating the Contribution of Internal Fragments to the Sequence Characterization of Progressively Larger Oligonucleotides and Rnas by Top-Down Mass Spectrometry*, Proceedings of the 70th ASMS Conference on Mass Spectrometry and Allied Topics, Minneapolis, MN, June 5-9, 2022.

21. Wong, D.; Rye, P.; Madden, S.; Slysz, G.; Cody, C. In *Comprehensive, Automated, and Integrated Software for Oligonucleotide Characterization and Sequence*

Confirmation, Proceedings of the 70th ASMS Conference on Mass Spectrometry and Allied Topics, Minneapolis, MN, June 5-9, 2022; p WP 386.

22. Hendel, A.; Bak, R. O.; Clark, J. T.; Kennedy, A. B.; Ryan, D. E.; Roy, S.; Steinfeld, I.; Lunstad, B. D.; Kaiser, R. J.; Wilkens, A. B.; Bacchetta, R.; Tsalenko, A.; Dellinger, D.; Bruhn, L.; Porteus, M. H. Chemically Modified Guide Rnas Enhance Crispr-Cas Genome Editing in Human Primary Cells. *Nat. Biotechnol.* **2015,** *35*, 985-989.

23. Qi, Y.; Thompson, C. J.; Orden, S. L. V.; O'Connor, P. B. Phase Correction of Fourier Transform Ion Cyclotron Resonance Mass Spectra Using Matlab. *J. Am. Soc. Mass Spectrom.* **2011,** *22*, 138-147.

24. Campuzano, I. D. G.; Nshanian, M.; Spahr, C.; Lantz, C.; Netirojjanakul, C.; Li, H.; Wongkongkathep, P.; Wolff, J. J.; Loo, J. A. High Mass Analysis with a Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: From Inorganic Salt Clusters to Antibody Conjugates and Beyond. *J. Am. Soc. Mass Spectrom.* **2020,** *31*, 1155-1162.

25. Shaw, J. B.; Lin, T.-Y.; III, F. E. L.; Tolmachev, A. V.; Tolic, N.; Robinson, E. W.; Koppenaal, D. W.; Pasa-Tolic, L. 21 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometer Greatly Expands Mass Spectrometry Toolbox. *J. Am. Soc. Mass Spectrom.* **2016,** *27*, 1929-1936.

26. Li, F.; Maguigad, J.; Pelzer, M.; Jiang, X.; Ji, Q. C. A Novel 'Peak Parking' Strategy for Ultra-Performance Liquid Chromatography/Tandem Mass Spectrometric Detection for Enhanced Performance of Bioanalytical Assays. *Rapid Commun. Mass Spectrom.* **2008,** *22*, 486-494.

27. Nagornov, K. O.; Kozhinov, A. N.; Gasilova, N.; Menin, L.; Tsybin, Y. O. Transient-Mediated Simulations of Ftms Isotopic Distributions and Mass Spectra to Guide Experiment Design and Data Analysis. *J. Am. Soc. Mass Spectrom.* **2020,** *31*, 1927-1942.

28. Snyder, L. R.; Dolan, J. W. High-Performance Gradient Elution: The Practical Application of the Linear-Solvent-Strength Model. John Wiley & Sons: 2007.

29. Goyon, A.; Blevins, M. S.; Napolitano, J. G.; Nguyen, D.; Goel, M.; Scott, B.; Wang, J.; Koenig, S. G.; Chen, T.; Zhang, K. Characterization of Antisense Oligonucleotides and Guide Ribonucleic Acid Diastereomers by Hydrophilic Interaction Liquid Chromatography Coupled to Mass Spectrometry. *J. Chromatogr. A* **2023,** *1708*, 464327.

30. Huber, C. G.; Krajete, A. Analysis of Nucleic Acids by Capillary Ion-Pair Reversed-Phase Hplc Coupled to Negative-Ion Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **1999,** *71*, 3730-3739.

31. Lin, Z. J.; Li, W.; Dai, G. Application of Lc-Ms for Quantatative Analysis and Metabolite Identification of Therapeutic Oligonucleotides. *J. Pharm. Biomed. Anal.* **2007,** *44*, 330-341.

32. Huber, C. G.; Oberacher, H. Analysis of Nucleic Acids by on-Line Liquid Chromatography–Mass Spectrometry. *Mass Spectrom. Rev.* **2001,** *20*, 310-343.

33. Santos, I. C.; Brodbelt, J. S. Recent Developments in the Characterization of Nucleic Acids by Liquid Chromatography, Capillary Electrophoresis, Ion Mobility, and Mass Spectrometry (2010-2020). *J. Sep. Sci.* **2021,** *44*, 340-372.

34. Nikolaev, E. N.; Kostyukevich, Y. I.; Vladimirov, G. N. Fourier Transform Ion Cyclotron Resonance (Ft Icr) Mass Spectrometry: Theory and Simulations. *Mass Spectrom. Rev.* **2016,** *35*, 10.1002/mas.21422.

35. Mitchell, D. W.; Smith, R. D. Cyclotron Motion of Two Coulombically Interacting Ion Clouds with Implications to Fourier-Transform Ion Cyclotron Resonance Mass Spectrometry. *Phys. Rev. E* **1995,** *52*, 4366-4386.

36. Guimaraes, G. J.; Sutton, J. M.; Gilar, M.; Donegan, M.; Bartlett, M. G. Impact of Nonspecific Adsorption to Metal Surfaces in Ion Pair-Rp Lc-Ms Impurity Analysis of Oligonucleotides. *J. Pharm. Biomed. Anal.* **2022,** *208*, 114439.

37. Mowery, R. A. The Corrosion of 316 Stainless Steel in Process Liquid Chromatography with Acetonitrile or Methanol Carriers. *J. Chromatogr. Sci.* **1985,** *23*, 22-29.

38. Ivleva, V. B.; Yu, Y.-Q.; Gilar, M. Ultra-Performance Liquid Chromatogtaphy/Tandem Mass Spectrometry (Uplc/Ms/Ms) and Uplc/Mse Analysis of Rna Oligonucleotides. *Rapid Commun. Mass Spectrom.* **2010,** *24*, 2631-2640.

39. Köster, C. Mass Spectrometry Method for Accurate Determination of Unknown Ions. US 6 188 064 B1, 1999.

40. Marshall, A. G.; Guan, S. Advantages of High Magnetic Field for Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **1996,** *10*, 1819-1823.

41. Nagornov, K. O.; Gorshkov, M. V.; Kozhinov, A. N.; Tsybin, Y. O. High-Resolution Fourier Transform Ion Cyclotron Resonance Mass Spectrometry with Increased Throughput for Biomolecular Analysis. *Anal. Chem.* **2014,** *86*, 9020-9028.

42. Nikolaev, E.; Lioznov, A. Evaluations of Major Historical Icr Cell Designs Using Electric Field Simulations. *Mass Spectrom. Rev.* **2022,** *41*, 262-283.

43. Marshall, A. G.; Comisarow, M. B.; Parisod, G. Relaxation and Spectral Line Shape in Fourier Transform Ion Resonance Spectroscopy. *J. Chem. Phys.* **1979,** *71*, 4434-4444.

Chapter 5 Conclusions and Future directions

5.1 Dissertation Summary

This dissertation extended the known applications of liquid chromatography mass spectrometry for therapeutic oligonucleotides. In this work a mobile phase system free of fluoroalcohol additives makes these workflows environmentally friendly. Additionally, the use of methanol instead of acetonitrile further reduces the environmental. Overall these results show improvements to known experimental capabilities of the platforms using these mobile phase components, suggesting acetonitrile and fluoroalcohol additive may not be necessary for these experimental spaces and should be avoided if amenable.

Chapter 2 described fast chromatography to attain high quality precursor ion populations. The abundant high charge states accessed in these workflows suggest that instrument and material quality has improved since the introduction of triethylammonium bicarbonate IP-RP chromatography. The relatively high charge states obtained here are analytically more useful than the historic charge states obtained in the literature.[2] The applied green mobile phase is shown to be capable of attaining full sequence coverage of the difficult to characterize heavily modified oligonucleotides.

Chapter 3 demonstrated the utility of the mobile phase for longer RNA sequences. The polymerase chain reaction products investigated here were similar in character to therapeutic single guide RNA. The moderate charge states generated by

triethylammonium bicarbonate spread signal over a wider mass range than what would be expected with fluoroalcohol additives. The broader charge state distribution simplifies algorithm deconvolution.

Chapter 4 described a method of numerically comparing the attainable sequence confirmation resolution for a given set of data analysis model and data. This part critically assessed the sequence fidelity of the OligoTap model for the Bruker SNAP workflow under harsh conditions of attempting to assign isobaric sequence permutations. These decoys show the possibility of false assignments data sets in spite of high accuracy and resolution. The juxtaposition of the CID tandem technique to EDD highlighted selection of tandem method by selecting a high sequence resolution method. This chapter also acknowledged the shortcomings of the OligoTap model in that it does not attempt to assign many radical tandem mass spectrometry fragment types, OligoTap does attempt to assign many collision induced dissociation fragmentation pathways. We hypothesized in this chapter there a local maxima for sequence coverage and sequence resolution yet to be identified in CID tandem mass spectrometry.

5.2 Future Directions

5.2.1 OligoTap Charge Carrier Correction

An important adjustment should be made to the OligoTap program. The algorithm uses hydrogen as its charge carrier which is one electron to large in mass. The single charge chemical formula files generated by OligoTap provide one means of avoiding this issue. One could use the chemical formulas for the fragments as a

theoretical mass list starting point for their application. The mass defect introduced for a theoretical precursor or fragment approaches 5 ppm mass error when considering the tenth charge state of the analyte. (**Figure 5.1**) Neutral mass fragments such as those produced by neutral isotopically resolved deconvolution should be in agreement with the chemical formula calculated by OligoTap, such as those neutral masses generated by the 6560c workflow described in Chapter 2. The SNAPed workflow described in Chapter 2 is subject to this mass defect. The correction of the mass defect would need to be addressed for two of the three different isotope handling methods in OligoTap. This also implies the defect needs to be changed for two polarities addressed in those two isotope models as OligoTap addresses both positive and negative polarity. The chemical formula representation which is based on integers to represent the number of atoms can remain unchanged in both charges for all three isotope models as standard chemical formula notation does not distinguish between protons and hydrogens in the text report output. On the FT-ICR platform internal calibration is typically performed. The selection of internal calibration points is important. In the work of this thesis the calibrant points are taken from the OligoTap algorithm which suggests the multiply charged internal calibration points have the mass defect introduced to them. Generally the assignments

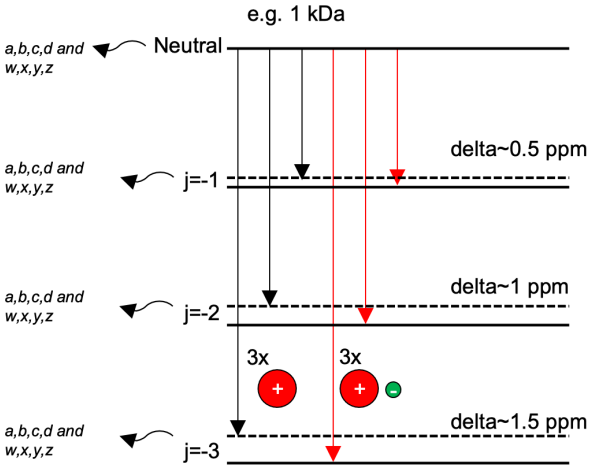


Figure 5.1 The mass defect introduced by electron delta for a multiply negatively charged analyte. The black arrows represent the theoretically correct reduction in mass of an analyte introduced by the removal of a proton. The red arrows show the reduction in mass of an analyte by the removal of hydrogen which is the calculation preformed by OligoTap 1.0.33.

from OligoTap would have a positive mass bias due but considering the internal calibration points have the negative mass bias of the electron the assignments in this work would have a slightly lower positive assignment bias than what would be theoretically expected. In this work the maximum charge considered is -11 thus the positive bias of assignments should theistically be somewhere between approximately 0 (zero) and 5.5 ppm mass error.

5.2.2 Additional OligoTap Directions

The OligoTap program takes text file and MGF inputs and outputs text files. Many programs and programming languages are able to parse and create text files. In the case of OligoTap two major computational duratives of the OligoTap operation have already been demonstrated Chapter 2 and Chapter 4. The report document of OligoTap has shown to be parsable by Python script to help facilitate sequence coverage visualization by Steven D. In Chapter 4 it is shown generation sequence decoys can be preformed programmatically and run. The formatting of the search sequences file to have an indicator (e.g. “=myoligo”) delimited was done for organization and with the intention of scaling up the search sequence strategy implemented in OligoTap. The modifications available in OligoTap should provide means of implementing OligoTap for processing oligonucleotide digestion workflows. The major limitation being the currently available modified bases in OligoTap 1.0.33. The program was tested with the AutoMSMS data acquisition modes available in the Bruker, Agilent, and Thermo acquisition softwares.

OligoTap should be applicable to fragmentation of longer RNA as is but the correction of the mass defect would facilitate higher precision assignments. Additional

programming of OligoTap to make assignments of radical ions described by Karasawa et al.[3] would improve the applicability to electron detachment dissociation. Wysobusine and glycan modified oligonucleotides[4] are good targets for demonstrating the utility of EDD for sequencing targets with labile modifications.

5.2.3 Further Critical review of the EDD Mechanism

The sequence coverage attained in Chapter 4 for the EDD fragmentation of a 19mer standard oligonucleotide and more rigorous analysis of the work presented by Szot et al.[5] of a different 19mer RNA sequence usages the current standing mechanism for EDD of oligonucleotides does not currently treat the mechanism of the 3' end. The proposed method[6] suggests the final *d*-type ion towards the 3' end is a forbidden possibility based on the proposed curling of the 3' end for the mechanisms progression. It may be an alternative to consider the complementary radical ions in an updated mechanistic proposal.[3] Additional oligonucleotides and different lengths would be good support for an updated mechanism.

5.2.4 6560c Based Measurements

Oligonucleotide modifications introduced in this work (phosphothioate) have been shown to introduce detectable collisional cross sectional differences upon implementation of drift tube multiplexing.[7] Similar 5-mer sequences with 2′-F modifications may show confirmations resolvable by the 6560c. The 2′-Fluoro-arabinonucleic Acid (2'F-ANA)[8] has the

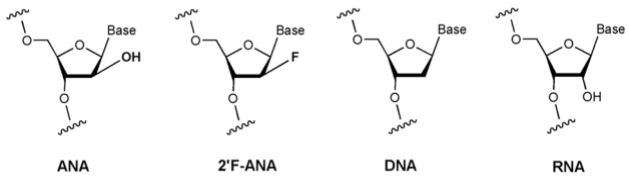


Figure 5.2 The steriochemistry of the ANA and 2'F ANA oligonucleotide structures compared to DNA and RNA.[1]

2’F in opposite stereochemistry than the 2’F RNA which might have implication for the relative efficiency of fragmentation. The 2’F-ANA structure has been acknowledged for its possible utility in probing interactions.[9] The 2’F RNA substitution increased binding affinity[10] differential duplex unfolding may be detectable between an unmodified duplex and the modified duplex composed of the modified sequences in Chapter 2. The duplexes would need to be prepared in similar workflows.

5.2.5 High Temperature Fluoroalcohol-Free Liquid Chromatography-FT-ICR Mass Spectrometry of RNA

Different experimental parameters were suggested in “Chapter 4 High Temperature Fluoroalcohol-Free Liquid Chromatography-FT-ICR Mass Spectrometry of RNA up to 100 kDa” . Addressing these parameters may prove fruitful for analyzing long RNA. Integrated DNA technologies and collaborators are able to produce RNA in the ~120 nt range. Iron in the flow path of the LC and possibly other sources should be addressed. The inter run increase in iron is a surprising finding of the work, possibly injection of EDTA from a sample vial could also help mitigate the iron presence. The volume of mobile phase between the post UV valve and needle may have contributed. Perhaps further flushing of this volume in the method valve program would reduce this possible iron contribution. Narrower quadrupole isolation at a higher charge state e.g. 110 may improve quadrupole isolation efficiency and reduce required free induction decay time. Narrower quadrupole isolation to one charge state should improve the space charge environment of the cell. Improving the trapping parameters for the ParaCell may improve the achievable transient ring down time. Absorption mode processing may allow the sacrifice of transient collection for additional scan averaging.

Narrower emitter inner diameter e.g. Agilent micronebulizer may provide a reduction in cation adduction, this implementation would require further flow rate decrease. Possibly better ParaCell shimming could improve transient acquisition.

5.3 References

1. Watts, J. K.; Katolik, A.; Viladoms, J.; Damha, M. J. Studies on the Hydrolytic Stability of 2'-Fluoroarabinonucleic Acid (2'f-Ana). *Org. Biomol. Chem.* **2009,** *7*, 1904-1910.

2. Huber, C. G.; Oberacher, H. Analysis of Nucleic Acids by on-Line Liquid Chromatography–Mass Spectrometry. *Mass Spectrom. Rev.* **2001,** *20*, 310-343.

3. Karasawa, K.; Duchoslav, E.; Baba, T. Fast Electron Detachmentn Dissociation of Oligonucleotides in Electron-Nitrogen Plasma Stored in Magneto Radio-Frequency Ion Traps. *Anal. Chem.* **2022,** *94*, 15510-15517.

4. Hu, B.; Zhong, L.; Weng, Y.; Peng, L.; Huang, Y.; Zhao, Y.; Liang, X.-J. Therapeutic Sirna: State of the Art. *Signal Transduction Targeted Ther.* **2020,** *5*, 101.

5. Hakansson, C. W. S. K. In *High Temperature Fluoroalcohol-Free Liquid Chromatography-Electron Detachment Dissociation Tandem Mass Spectrometry of Rna*, Proceedings of the 71th ASMS Conference on Mass Spectrometry and Allied Topics, Huston, TX, June 4-8, 2023.

6. Taucher, M.; Breuker, K. Top-Down Mass Spectrometry for Sequencing of Larger (up to 61 Nt) Rna by Cad and Edd. *J. Am. Soc. Mass Spectrom.* **2010,** *21*, 918-929.

7. Demelenne, A.; Nys, G.; Nix, C.; Fjeldsted, J. C.; Crommen, J.; Fillet, M. Separation of Phosphorothioated Oligonucleotide Diastereomers Using Multiplexed Drift Tube Ion Mobility Mass Spectrometry. *Anal. Chim. Acta* **2022,** *1191*, 339297.

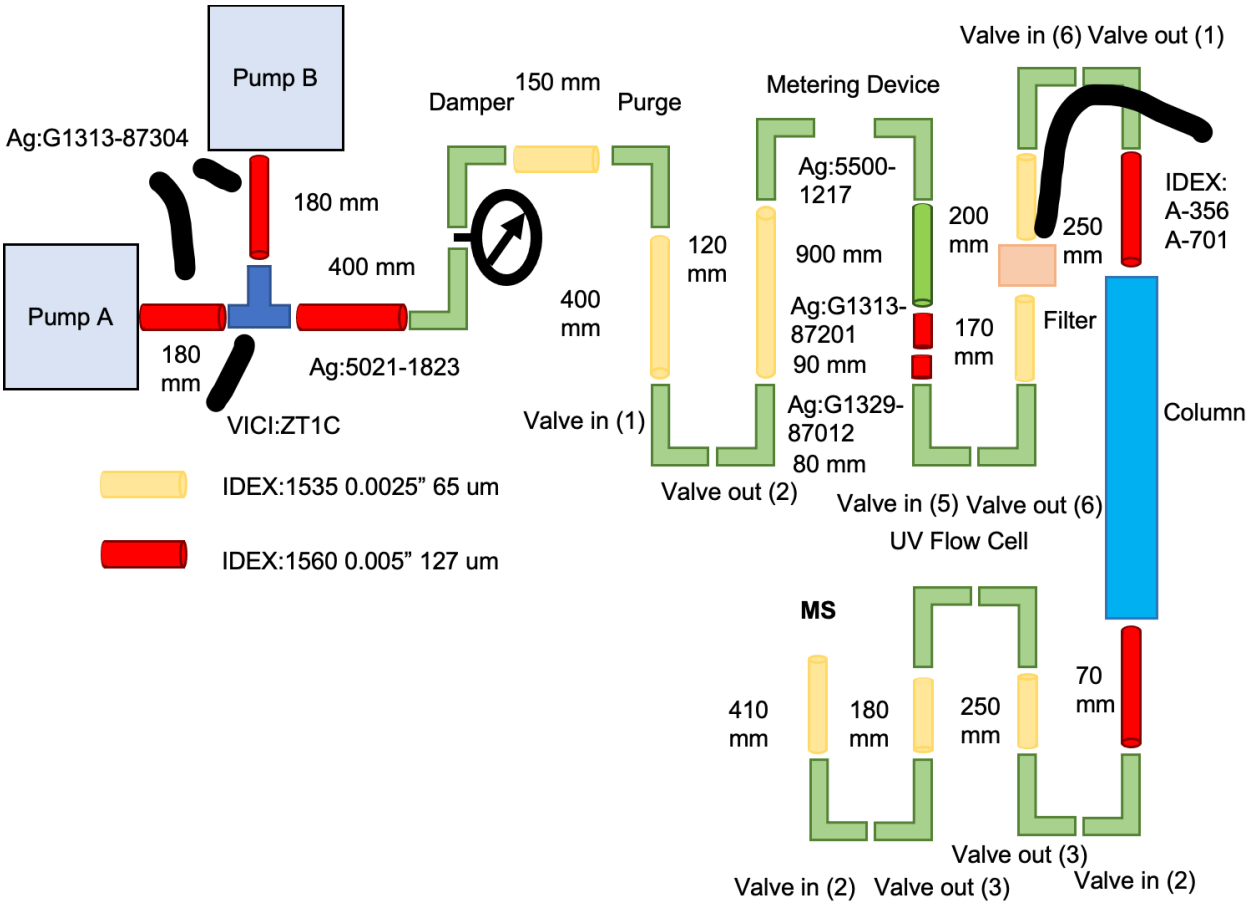
8. El-Khoury, R.; Damha, M. J. 2'fluoro-Arabinonucleoc Acid (Fana): A Versatile Tool for Probing Biomolecular Interactions. *Acc. Chem. Res* **2021,** *54*, 2287-2297.

9. Khvorova, A.; Watts, J. K. The Chemical Evolution of Oligonucleotide Therapies of Clinical Utility. *Nat. Biotechnol.* **2017,** *35*, 238-248.

Appendices

Appendix A: siRNA LC-MS/MS Additional Information

Appendix Figure A.1 Diagram showing some of the key components of the LC flow path on the Agilent 1100. Natural and red colored peek are colored and part numbers for specific non-inhouse cut/assembled parts. The damper of the binary pump is marked and a cartoon of a pressure reading is included, modifications upstream of this point can be hazardous as the mechanical specifications may not be within expected system tolerances and may cause hardware fault. The gren tubing after the metering device is smaller than the standard software user interfaces expect which may cause sample to be drawn into the metering device plunger, possibly causing damage and/or carryover. The LC mixer was bypassed to reduce gradient delay time/volume. Recommended pressure limit of tubing directly after the damper (~340 bar) was not exceeded in typical operation. Valve positions before and after the column are the column compartment valve when in operation. The valve after the UV flow cell was a VICI valve described in the text.



Appendix Figure A.3 FT-ICR method report for PLRP-S column showing the reported metadata associated with the method. The external contact board was wired to the VICI valve and the valve contact time of 0.05 min was deemed to be adequate for triggering the switch during experimentation.

Method Set: D:\Methods\Carson\20230427\2.5 min grad sample acq trim_flush adjust.m

Separation Method: 2.5 min grad sample acq trim_flush adjust.m

Run Times

Total run time:	18 min
Acquisition time:	1.1 min
Start delay time:	2.4 min

Segment ICF System

Binary Pump

Flow:	0.200 mL/min	**Low Pressure Limit:**	0.00 bar
High Pressure Limit:	300.00 bar	**Maximum Flow Gradient:**	100.000 mL/min²

Stroke A

Automatic Stroke Calculation A:	Yes

Stroke B

Automatic Stroke Calculation B:	Yes

Compress A

Compressibility Mode A:	Compressibility Value Set	**Compressibility A:**	50 10e-6/bar

Compress B

Compressibility Mode B:	Compressibility Value Set	**Compressibility B:**	115 10e-6/bar

Stop Time

Stoptime Mode:	Time set	**Stoptime:**	18.00 min

Post Time

Posttime Mode:	Off

Solvent Composition

Channel	Used	Percent
A	Yes	95.0
B	Yes	5.0

Timetable

Time	A	B	C	D	Flow	Pressure
2.50	0.0	100.0			---	---
5.50	0.0	100.0			---	---
6.25	95.0	5.0			---	---
18.00	95.0	5.0			---	---

Sampler

Auxiliary

Draw Speed:	200 µL/min	**Eject Speed:**	200 µL/min
Draw Position Offset:	0.0 mm		

Injection

Injection Mode:	Injection with needle wash	**Injection Volume:**	2.00 µL

High throughput

Stop Time

Stoptime Mode:	As pump/No limit

Post Time

Posttime Mode:	Off

External Contacts

Contact A:	Open	**Contact B:**	Open
Contact C:	Open	**Contact D:**	Open

Timetable

Time	Function	Parameter
2.00	Change Contacts	Switch contact C to closed
2.05	Change Contacts	Switch contact C to open
4.00	Change Contacts	Switch contact D to closed
4.05	Change Contacts	Switch contact D to open
5.00	Change Contacts	Switch contact C to closed
5.05	Change Contacts	Switch contact C to open
5.50	Change Contacts	Switch contact D to closed
5.55	Change Contacts	Switch contact D to open

Needle Wash

Needle Wash Location:	Wash Vial	**Wash Location:**	Vial 100

Overlapped Injection

Enable Overlapped Injection:	No

Column Comp.

Valve Position:	Position 2 (Port 1 -> 6)

Left Temperature Control

Temperature Control Mode:	Temperature Set	**Temperature:**	80.0 °C

Right Temperature Control

Right temperature Control Mode:	Combined

Stop Time

Stoptime Mode:	As pump/injector

Post Time

Posttime Mode:	Off

Timetable

Enable Analysis Left Temperature

Enable Analysis Left Temperature On:	Yes	**Enable Analysis Left Temperature Value:**	0.8 °C

Enable Analysis Right Temperature

Enable Analysis Right Temperature On:	Yes	**Enable Analysis Right Temperature Value:**	0.8 °C

VWD

Signal Peakwidth:	> 0.1 min (2 s resp. time) (3.43 Hz)	**Analog Output Source Channel:**	1
Signal Polarity:	Positive (+)	**Lamp on required for aquisition:**	Yes
Acquire Signal without Reference:	No	**Acquire Reference only:**	No

Analog Output

Analog Zero Offset:	5 %	**Analog Attenuation:**	1000 mAU

Signals

Prepare Mode

Margin for negative Absorbance:	100 mAU

Autobalance

Autobalance Prerun:	Yes	**Autobalance Postrun:**	No

Scan Variables

Scan Range From:	190 nm	**Scan Range To:**	400 nm
Scan Range Step:	2 nm		

Stoptime

Stoptime Mode:	As pump/injector

Posttime

Posttime Mode:	Off

Timetable

Signal

Acquire Signal:	Yes	**Signal ID:**	Signal A
Signal Wavelength:	260 nm		

Appendix Figure A.5 Method report of HPH-C18 Acquisition method on 6560c IM-Q-TOF mass spectrometer.

Acquisition Method Report

Agilent Technologies

Acquisition Method Info

Method Name	NS_6_7.m
Method Path	D:\Projects\OligosiRNA\Methods\20230519\NS_6_7.m (Version: NA)
Method Description	

Device List

- Binary Pump
- Sampler
- Column Comp.
- VWD
- Q-TOF

TOF/Q-TOF Mass Spectrometer

Component Name	MS Q-TOF	**Component Model**	G6560B
Ion Source	Dual AJS ESI	**Stop Time (min)**	No Limit/As Pump
Can wait for temp.	Disable	**Fast Polarity**	False
MS Abs. threshold	200	**MS Rel. threshold(%)**	0.010
MS/MS Abs. threshold	5	**MS/MS Rel. threshold(%)**	0.010
Ion Mobility Mode	QTOF Only		

Time Segments

Time Segment #	Start Time (min)	Diverter Valve State	Storage Mode	Ion Mode
1	0	Waste	Both	Dual AJS ESI

Time Segment 1

Acquisition Mode TargetedMS2

MS Min Range (m/z)	100
MS Max Range (m/z)	3000
MS Scan Rate (spectra/sec)	2.00
MS/MS Min Range (m/z)	100
MS/MS Max Range (m/z)	3000
MS/MS Scan Rate (spectra/sec)	1.00
Max Time Between MS (sec)	Disabled
Decision Engine	Disabled
Use Fixed Collision Energies	0.00

Targeted Mass Table

Start Mass	Z	Ret. Time (min)	Delta Ret. Time (min)	Isolation Width	Collision Energy	Acq. Time (ms/spec)
1026	7	4.6	2	Wide (~9 amu)	16	
1197	6	4.6	2	Wide (~9 amu)	19	
1026	7	4.6	2	Wide (~9 amu)	16	
1197	6	4.6	2	Wide (~9 amu)	19	

Instrument Parameters

Parameter	Value
Gas Temp (°C)	325
Gas Flow (l/min)	12
Nebulizer (psig)	35
SheathGasTemp	350
SheathGasFlow	12

Scan Segments

Scan Seg #	Ion Polarity
1	Negative

Scan Segment 1

Scan Source Parameters

Parameter	Value
VCap	3500
Nozzle Voltage (V)	2000
Fragmentor	400
Skimmer1	65
OctopoleRFPeak	750

ReferenceMasses

Ref Mass Enabled	Disabled
Ref Nebulizer (psig)	

Chromatograms

Chrom Type	Label	Expt. Type	Polarity Type	Offset	Y-Range
TIC	TIC	Both	Both	15	10000000
EIC	EIC	MSMS	Both	15	10000000
EIC	EIC	MS	Both	15	10000000

Name: Binary Pump Model: G1312A

Flow	0.100 mL/min
Low Pressure Limit	0.00 bar
High Pressure Limit	340.00 bar
Maximum Flow Gradient	100.000 mL/min²
Stroke A	
Automatic Stroke Calculation A	Yes
Stroke B	
Automatic Stroke Calculation B	Yes
Compress A	
Compressibility Mode A	Compressibility Value Set
Compressibility A	50 10e-6/bar
Compress B	
Compressibility Mode B	Compressibility Value Set
Compressibility B	115 10e-6/bar
Stoptime	
Stoptime Mode	Time set
Stoptime	20.25 min
Posttime	
Posttime Mode	Off

Solvent Composition

	Channel	Name 1	Used	Percent
1	A		Yes	95.0 %
2	B		Yes	5.0 %

Timetable

	Time	A	B	Flow	Pressure
1	Start. Cond. min	95.0 %	5.0 %	0.100 mL/min	340.00 bar
2	5.00 min	0.0 %	100.0 %	0.100 mL/min	--- bar
3	5.50 min	0.0 %	100.0 %	0.100 mL/min	--- bar
4	5.51 min	0.0 %	100.0 %	0.300 mL/min	--- bar
5	7.75 min	0.0 %	100.0 %	0.300 mL/min	--- bar
6	8.25 min	95.0 %	5.0 %	0.300 mL/min	--- bar
7	16.25 min	95.0 %	5.0 %	0.300 mL/min	--- bar
8	16.26 min	95.0 %	5.0 %	0.100 mL/min	--- bar
9	20.25 min	95.0 %	5.0 %	0.100 mL/min	--- bar

Name: Sampler Model: G1329A

Auxiliary

Draw Speed	200 µL/min
Eject Speed	200 µL/min
Draw Position Offset	0.0 mm

Injection

Injection Mode	Injection with needle wash
Injection Volume	2.00 µL

Needle Wash

Needle Wash Location	Wash Vial
Wash Location	Vial 100

High throughput

Overlapped Injection

Enable Overlapped Injection	No

Stoptime

Stoptime Mode	As Pump/No Limit

Posttime

Posttime Mode	Off

External Contacts

Contact A	Open
Contact B	Open
Contact C	Open
Contact D	Open

Timetable

Use Injector Program

Use Injector Program	No

Timetable

Timetable

	Time	Function	Parameter
1	3.50 min	Change Contacts	Switch contact C to closed
2	3.55 min	Change Contacts	Switch contact C to open
3	5.50 min	Change Contacts	Switch contact D to closed
4	5.55 min	Change Contacts	Switch contact D to open
5	7.50 min	Change Contacts	Switch contact C to closed
6	7.55 min	Change Contacts	Switch contact C to open
7	7.75 min	Change Contacts	Switch contact D to closed
8	7.80 min	Change Contacts	Switch contact D to open

Name: Column Comp. Model: G1316A

Valve Position	Use current

Left Temperature Control

Temperature Control Mode	Temperature Set
Temperature	60.0 °C

Enable Analysis Left Temperature

Enable Analysis Left Temperature On	Yes
Enable Analysis Left Temperature Value	0.8 °C

Right Temperature Control

Right temperature Control Mode	Combined

Enable Analysis Right Temperature

Enable Analysis Right Temperature On	Yes
Enable Analysis Right Temperature Value	0.8 °C

Stoptime

Stoptime Mode	As Pump/Injector

Posttime

Posttime Mode	Off

Timetable

Name: VWD **Model:** G1314A

Signal Peakwidth	> 0.1 min (2 s resp. time) (3.43 Hz)
Analog Output Source Channel	1
Signal Polarity	Positive (+)
Lamp on required for acquisition	Yes
Acquire Signal without Reference	No
Acquire Reference only	No
Analog Output	
Analog Zero Offset	5 %
Analog Attenuation	1000 mAU
Signals	
Signal	
Acquire Signal	Yes
Signal ID	Signal A
Signal Wavelength	260 nm
Prepare Mode	
Margin for negative Absorbance	100 mAU
Autobalance	
Autobalance Prerun	Yes
Autobalance Postrun	No
Scan Variables	
Scan Range From	190 nm
Scan Range To	400 nm
Scan Range Step	2 nm
Stoptime	
Stoptime Mode	As Pump/Injector
Posttime	
Posttime Mode	Off
Timetable	

Appendix Figure A.7 FT-ICR acquisition method parameter report showing the recorded metadata of the instrument method.

ftmsControl

818-10

Chromatography Traces

Name	ftmsControl	hystar	Color
TIC	true	true	#FF0000
BPC	true	true	#404040
Capillary	true	true	#0000FF

[Main]

Polarity

Polarity:	Negative	**Alternate Polarity:**	off

Chromatography

LC Mode:	HyStar_LC-MS	**LC Run Time:**	4.0 min
Auto MS/MS:	off	**ESI High Voltage:**	on
Source Quench:	on		

[Mode]

Acquisition Mass Control

Detection Mode:	Broadband	**Data Acquisition Size:**	2097152
Data Processing Size (SI):	4194304	**Broadband Low Mass:**	100.4 m/z
Broadband High Mass:	3000.0 m/z	**Pulse Program:**	basic_ADD

Data Storage

Perform Data Reduction:	off	**Quantile:**	0.0 %
Save Full Profile Spectrum:	on	**Save FID File:**	on

Accumulation

NS:	1	**Source Accumulation:**	0.000 sec
Ion Accumulation Time:	0.800 sec	**Ion cooling time:**	0.000 sec
Time of Flight to Detector:	0.001 sec	**Selective Accumulation:**	off
CASI Segmented Acquisition:	off	**Accumulation During Detection:**	on
LC Capture Mode:	on		

LC Capture Mode

Individual Fill Time:	0.2 sec/accum	**Individual Quench Time:**	0.04341748 sec
Number of Fills:	4	**Sampling Rate:**	4.108168 Accum/sec

Auto Calibration

On-line Calibration:	off

[Processing]

Algorithms

Peak Picking Algorithm:	Statistical_CPlusPlus	**Noiselevel Algorithm:**	Segemented_Java

[Source Info]

API Source

Source:	ESI	**Capillary:**	3500.0 V
Spray Shield:	-500.0 V	**Drying Gas Flow Rate:**	10.0 L/min

[Ion Transfer]

Source Optics

Capillary Exit:	-260.0 V	**Deflector Plate:**	-240.0 V
Skimmer 1:	-15.0 V	**Funnel RF Amplitude:**	150.0 Vpp

Octopole

Octopole Frequency:	1.0	**Octopole RF Amplitude:**	350.0 Vpp

Quadrupole

Q1 Mass:	818.0 m/z

Collision Cell

Collision Voltage (Entrance):	12.0 V	**DC Extract Bias (Entrance):**	-2.3 V
		Collision Cell RF:	2000.0 Vpp
Collision Cell Frequency:	1.0	**Transfer Line RF:**	350.0 Vpp
Transfer Line Frequency:	1.0		

Gas Control

Collision Gas Flow Rate:	72.0 %	**Collision Gas:**	on

Time-of-Flight

Time of Flight to Detector:	0.001 sec

[Analyzer]

Para Cell

Transfer Exit Lens:	20.0 V	**Analyzer Entrance:**	10.0 V
Sidekick:	-0.0 V	**Sidekick Offset:**	1.5 V
Front Trap Plate:	-3.0 V	**Back Trap Plate:**	-3.0 V
Excitation Power (Sweep/ Shot):	7.1 dB	**Sweep Direction:**	Decreasing
No. of Cell Fills:	1		

[Source MS]

In Source Fragmentation

InSource CID:	off	**InSource Collision Energy:**	90.0 V
MALDI-ISD:	off		

Quadrupole MS/MS

Q1 Isolate:	on	**Q1 CID:**	on

Q1 CID Energy:	10.0 V	**Q1 Mass:**	818.0 m/z
Q1 Isolation Window:	20.0	**Collision Cell RF:**	2000.0 Vpp
Collision Cell Frequency:	1.0		

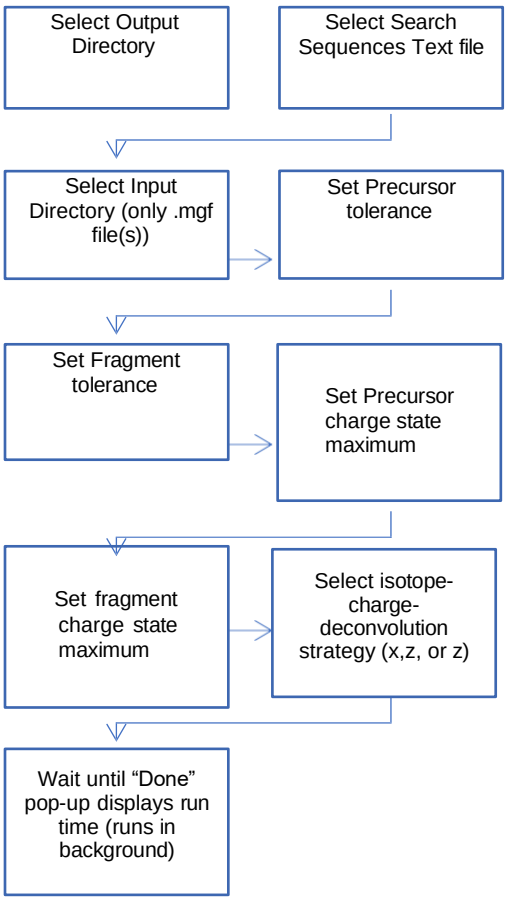
ETD

ETD Switch:	off

Appendix Table A.1 The format of the search sequences for the siRNA oligonucleotides. Font size is 6.5 to depict how the search sequences text file has one sequence per line.

HO-r,C.s/r,C.s/rm,U.p/rm,A.p/rm,C.p/rm,U.p/r,C.p/r,G.p/r,U.p/f,U.p/f,A.p/f,C.p/r,C.p/r,U.p/r,U.p/rm,C.p/rm,U.p/r,U.p/rmoe,5C.p/r ,U.s/r,G.s/r,A-OH=MAS
HO-r,U.s/r,C.s/r,A.p/rm,G.p/rm,A.p/rm,A.p/rm,G.p/r,A.p/r,A.p/r,G.p/r,G.p/f,U.p/f,A.p/f,A.p/r,C.p/r,G.p/r,A.p/rmoe,G.p/r,U.p/r,A.s/r,G.s/r,G-OH=modifiedsense
HO-r,U.p/r,C.p/r,A.p/r,G.p/r,A.p/r,A.p/r,G.p/r,A.p/r,A.p/r,G.p/r,G.p/r,U.p/r,A.p/r,A.p/r,C.p/r,G.p/r,A.p/r,G.p/r,U.p/r,A.p/r,G.p/r,G-OH=sense

Appendix Figure A.9 OligoTap prompts flow illustrating the inputs asked of a user in pop-up format



Appendix Table A.2 Elemental masses used in OligoTap for calculating chemical formula from input sequences.

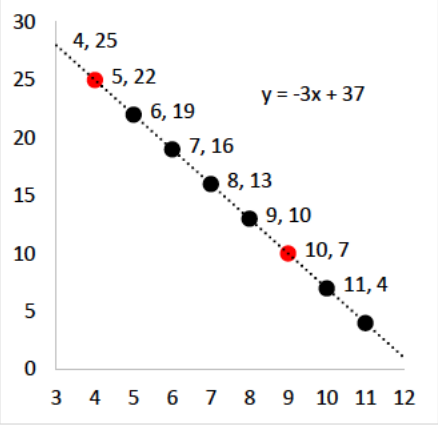
C_ = 12
H = 1.007825
F = 18.998403
N = 14.003074
O = 15.994915
P = 30.973763
S = 31.972072

Appendix Figure A.11 Linear optimization and approximation of collision energy and approximation for multiple charge states.

	unmodified SO		modified SO		modified ASO		composite	
CID (V)	z=-9	z=-4	z=-9	z=-4	z=-9	z=-4	z=-9	z=-4
5	95		81		81		85.667	
10	100		81		76		85.667	
15	86		43		52		60.333	
25		85		81		62		76
30		95		90		62		82.333
35		81		100		52		77.667
9	10			9	10			4	9
4	25			4	25			25	10
Slope	-3								
Intercep	37								

	4	5	6	7	8	9	10	11
z	25	22	19	16	13	10	7	4

11	4
10	7
9	10
8	13
7	16
6	19
5	22
4	25



Appendix B: OligoTap 1.0.33

OligoTap1.0.33 which is available on Github at https://github.com/cwszot/OligoTap is reported with line numbers directly before the line. Lines are wrapped in subsequent lines until the next line number is marked on the next line.

```
1 'inputs for "varriable bases" (vb) 1-3
2 vb1C=0
3 vb1O=0
4 vb1H=1
5 vb1N=0
6 vb1P=0
7 vb1S=0
8 vb1F=0
9 '-----
10 vb2C=0
11 vb2O=0
12 vb2H=1
13 vb2N=0
14 vb2P=0
15 vb2S=0
16 vb2F=0
17 '-----
18 vb3C=0
19 vb3O=0
20 vb3H=1
21 vb3N=0
22 vb3P=0
23 vb3S=0
24 vb3F=0
25 '-----
26
27
28 'the interface code whichsets the user options
29 '-----------------------------------------------
30 'modified nucleosides match the 2021 MODOMICS data base
31 Dim objFolder, objItem, objShell, fragfilepath, fragsfilepaths(), filevar(), files(),
filenames(),filepaths(), tempfile(), filelines(), mgf_check(), begin_ion_indices(),
end_ion_indices(), oligofragments(), sequences(), oligo(), oligo_1(), filevar_m,oligoname
(), tempfile_report, oligoprecursors()
32 'begin calculator definitions
33 Dim oligofragment_zs(),oligofragment_ints()
34 dim alpha, beta, o_array0, o_array1, nucleotide, nucleotides, sugar, link, base, var,
var1, a(), var2, var3, var4, var5, var6, test
35 dim aC(), aH(), aF(), aN(), aO(), aP(), a_S()
```

```
dim abC(), abH(), abF(), abN(), abO(), abP(), ab_S()
dim message
dim small_dalton, big_dalton
small_dalton=0.97
big_dalton=1.03
'end calculator definitions
' error handling
On Error Resume Next
SelectFolder = vbNull
' Create a dialog object
Set objShell = CreateObject( "Shell.Application" )
Set objFolder = objShell.BrowseForFolder( 0, "Select Folder for the script to generate
theoritical fragment tables and a report text file", 0, myStartFolder )
' Return the path of the selected folder
If IsObject( objfolder ) Then SelectFolder = objFolder.Self.Path
fragfilepath = objFolder.Self.Path
' Standard housekeeping
objShell.Close
Set objFolder = Nothing
Set objshell = Nothing
On Error Goto 0
Set wShell=CreateObject("WScript.Shell")
Set oExec=wShell.Exec("mshta.exe ""about:<input type=file
id=FILE><script>FILE.click();new
ActiveXObject('Scripting.FileSystemObject').GetStandardStream(1).WriteLine(FILE.value);cl
ose();resizeTo(0,0);</script>""")
oligosfile = oExec.StdOut.ReadLine
Set oExec=Nothing
Set wShell = Nothing
Set oExec = Nothing
Dim fsoligo, MyoligoFile
Set fsoligo = CreateObject("Scripting.FileSystemObject")
Set MyoligoFile= fsoligo.OpenTextFile(oligosfile, 1)
MyoligoFile.ReadAll
ij=MyoligoFile.Line
MyoligoFile.Close
Set MyoligoFile = Nothing
ij_1=ij-1
ReDim oligos(ij_1)
ReDim oligo_1(ij_1)
ReDim oligoname(ij_1)
ReDim filevar(ij_1)
ReDim fragsfilepaths(ij_1)
Set MyoligoFile= fsoligo.OpenTextFile(oligosfile, 1)
c=0
For Each oligo in oligos
oligos(c)=MyoligoFile.ReadLine
oligo_2=Split(oligos(c),"=")
oligo_1(c)=oligo_2(0)
oligoname(c)=oligo_2(1)
'MsgBox oligoname(c)
fragsfilepaths(c)=fragfilepath
c=c+1
```

```
85 Next
86 MyoligoFile.Close
87 Set MyoligoFile=Nothing
88 Set fsoligo=Nothing
89 Dim fso, f3, f1, fc, s, f2, f4, f5, f6
90 Set fso = CreateObject("Scripting.FileSystemObject")
91 Set objShell = CreateObject( "Shell.Application" )
92 Set f4 = objShell.BrowseForFolder( 0, "Select Folder containing .mgf files", 0,
myStartFolder )
93 Set fc = fso.GetFolder(f4.Self.Path).Files
94 'possibly input a user input option
95
96
97 filelength = 0
98 For Each f1 in fc
99 s = s & f1.name
100 s = s & "<BR>"
101 varriable=Split(f1,"/")
102 mgf_check_length=0
103 for each element in varriable
104 mgf_check_length=mgf_check_length+1
105 Next
106 mgf_check_length_1=0
107 mgf_check_length_1=mgf_check_length-1
108 if varriable(mgf_check_length_1) = "temp/" then
109 else
110 final_mgf_element=Split(varriable(mgf_check_length_1),".")
111 if final_mgf_element(1) = "mgf" then
112 filelength=filelength+1
113 else
114 end if
115 end if
116 Next
117 d=filelength-1
118 ReDim files(d)
119 ReDim filenames(d)
120 ReDim filepaths(d)
121 ReDim tempfile(d)
122 ReDim filelines(d)
123 If (fso.FolderExists(fragfilepath&"\temp")) Then
124 Else
125 fso.CreateFolder (fragfilepath&"\temp")
126 End If
127 c7=0
128 For Each f1 in fc
129 If f1 = fragfilepath&"\temp" Then
130 Else
131 files(n)=fragfilepath&"\"&f1.name
132 filenames(n)=f1.name
133 Set f2 = fso.GetFile(f1)
134 f2.Copy (fragfilepath&"\temp\"&f1.name)
135 tempfile(c7)=fragfilepath&"\temp\"&f1.name
136 End If
```

```
137 c7=c7+1
138 s=c
139 Next
140 UserInput_precursor = InputBox("Enter acceptable m/z range of precursor isolation(s):"&
Chr(13)+"acceptable m/z range"&Chr(13)+"e.g:3","m/z tolerance","3",default)
141 ppmrange_precursor=CDbl(UserInput_precursor)
142 ppmrange2_precursor=CDbl(ppmrange_precursor/CDbl(2))
143 UserInput = InputBox("Enter ppm tolerance of fragment assignments:"&Chr(13)+"acceptable
ppm range"&Chr(13)+"e.g:20","m/z tolerance","20",default)
144 ppmrange=CDbl(UserInput)
145 ppmrange2=ppmrange/2000000
146 UserInput_max_charge_per_base = InputBox("Enter max charge of database precursor(s) to
search for:"&Chr(13)+"max charge"&Chr(13)+"e.g:5","Charge","5",default)
147 max_charge_per_base=CDbl(UserInput_max_charge_per_base)
148 UserInput=InputBox("Enter max charge of fragment to search for:","fragment charge states"
,"3",default)
149 max_charge_fragment=UserInput
150
151
152 charge_comparison=InputBox("Select charge-isotope strategy"&Chr(13)+"y=monoisotopic
(SNAP) w/charge column"&Chr(13)+"x=charge neutral resolved isotope deconvolution"&Chr(13
)+"Anyother input leads to no isotope handling","Isotope Strategy","x",default)
153 'charge_comparison="y"
154
155 skipcount=0
156 c8=0
157 Mytime1=Time
158 For Each mgf in tempfile
159 'MsgBox "Mgf read event"
160 Set fso = CreateObject("Scripting.FileSystemObject")
161 Set MyFile= fso.OpenTextFile(tempfile(c8), 1)
162 MyFile.ReadAll
163 i=MyFile.Line
164 MyFile.Close
165 Set MyFile=Nothing
166 skipcounter=i-1
167 Set MyFile= fso.OpenTextFile(tempfile(c8), 1)
168 begin_ion_count=0
169 end_ion_count=0
170 skipcount=0
171 For j=1 to skipcounter
172 lineinfo=MyFile.ReadLine
173 if lineinfo = "BEGIN IONS" then
174 begin_ion_index=MyFile.Line
175 ReDim Preserve begin_ion_indices(begin_ion_count)
176 begin_ion_indices(begin_ion_count)=MyFile.Line
177 begin_ion_count=begin_ion_count+1
178 elseif lineinfo = "END IONS" then
179 end_ion_index=MyFile.Line
180 ReDim Preserve end_ion_indices(end_ion_count)
181 end_ion_indices(end_ion_count)=MyFile.Line
182 end_ion_count=end_ion_count+1
183 else
```

```
End If
skipcount=skipcount+1
Next
MyFile.Close
Set MyFile=Nothing
skipcount=0
c6=0
c4=0
oligoprecursor=0
For each MSMS in begin_ion_indices
'MsgBox "msms read evet and c4="&c4
oligofragmentcount=0
Set MyFile= fso.OpenTextFile(tempfile(c8), 1)
k=begin_ion_indices(c6)-1
For j=1 to k
MyFile.SkipLine
Next
ion_lines=end_ion_indices(c6)-begin_ion_indices(c6)
For j=1 to ion_lines
info=MyFile.ReadLine
infoarray=Split(info,"=")
items=0
For each item in infoarray
items=items+1
Next
if infoarray(0)="PEPMASS" then
midstring=Mid(info,1,18)
oligomassarray=Split(midstring,"=")
oligoprecursor=CDbl(Mid(oligomassarray(1),1,9))
'MsgBox oligoprecursor
'ReDim Preserve oligoprecursors()
else
End If
If items=1 and j<ion_lines and info<>"END IONS" then
'split_info=Split(info," ")
info_length=len(info)
'if split_info_length>9 then
'c13 is meant to be the counter indicating column of the ion lines e.g.
mz, intensity, etc
c13=0
For i=1 to info_length
if mid(info,i,1)="0" or mid(info,i,1)="1" or mid(info,i,1)="2" or mid
(info,i,1)="3" or mid(info,i,1)="4" or mid(info,i,1)="5" or mid(info,
i,1)="6" or mid(info,i,1)="7" or mid(info,i,1)="8" or mid(info,i,1)=
"9" or mid(info,i,1)="." then
if c13=0 then
m_z=m_z&mid(info,i,1)
'MsgBox "carry="&carry
'oligofragment=CDbl(midstring)
'oligofragmentcount=oligofragmentcount+1
'ReDim Preserve
oligofragments(oligofragmentcount)
'oligofragments(oligofragmentcount)=oligofragment
```

```
232 elseif c13=1 then
233 m_z_int=m_z_int&mid(info,i,1)
234 elseif c13=2 then
235 m_z_z=m_z_z&mid(info,i,1)
236 else
237 'oligofragment=carry
238 'carry=nothing
239 'c13=c13+1
240 end If
241 'elseif mid(info,i,1)="0" or mid(info,i,1)="1" or mid(info,i,1)="2"
or mid(info,i,1)="3" or mid(info,i,1)="4" or mid(info,i,1)="5" or
mid(info,i,1)="6" or mid(info,i,1)="7" or mid(info,i,1)="8" or
mid(info,i,1)="9" or mid(info,i,1)="." then
242 ' carry=carry&mid(split_info(0),i,1)
243 else
244 c13=c13+1
245 end if
246 Next
247 'midstring=Mid(info,1,9)
248 oligofragment=CDbl(m_z)
249 m_z=""
250 oligofragment_int=CDbl(m_z_int)
251 m_z_int=""
252 if charge_comparison="y" then
253 If IsNumeric(m_z_z) Then
254 else
255 '...do something
256 MsgBox "Common issue is inappropriate peak picking, check
spectra are SNAPed with charge assignemnt"&Chr(13)&"all ions
must assigned charge in MGF file(s)"
257 WScript.Quit()
258 WSH.Quit()
259 End If
260 'MsgBox CStr(m_z_z)
261 oligofragment_z=CDbl(m_z_z)
262 m_z_z=""
263 Else
264 end if
265 oligofragmentcount=oligofragmentcount+1
266 ReDim Preserve oligofragments(oligofragmentcount)
267 oligofragments(oligofragmentcount)=oligofragment
268 ReDim Preserve oligofragments_ints(oligofragmentcount)
269 oligofragments_ints(oligofragmentcount)=oligofragment_int
270 ReDim Preserve oligofragments_zs(oligofragmentcount)
271 oligofragments_zs(oligofragmentcount)=oligofragment_z
272 else
273 End If
274 Next
275 'only for the first MSMS of the mgf file set generate text files of the oligo
database for subsequent MSMS dont open/close write the files
agian
276 ccc=0
277 For each oligonucleotide in oligos
```

```
278 'ccc=0
279 'uniquestructureset=Inputbox("Enter chemical formula for variable base cut
at the sugar 1' position. e.g:r,U2.p (MODOMICS Code U2)","Variable Base mods
input","C-4,H-3,F-0,N-2,O-1,P-0,S-1",default)
280 'MsgBox "ccc="&ccc
281 filevar(ccc) = fragfilepath+"\"+"oligo_"+oligoname(ccc)+" "+Cstr(ccc)+".txt"
282 'the source code which translates the oligo string to a text
file
283 '-------------------------------------------
284 'define arrays for the number of constituent atoms in the w
fragments?
285 'difine array for a-b fragment to hold the number of each element in the
fragment
286 'oligonucleotide with delimiters to devide the oligo at the building
blocks
287 'the 5' and 3' ends are separated by "-"
288 'the ends can be OH , me or ph for oh , methyl or phosphate
end
289 'the sugar group is separatred by "," and the r denotes ribose, 2'F ribose
as f, 2'o-methyl ribose as rm, 2' o-ethyl-o-methyl ribose as
rmoe
290 'the linker is separated by "." and will include phosphorous and sulfur as
the core of the phosphate backbone and/or phosphothiorate
291 'the nucleobase is separated by / and the bases will be
definied
292 oligo = oligo_1(ccc)
293 'define the molecule as having zero atoms of each type
initially
294 MC=0
295 MH=0
296 MF=0
297 MN=0
298 MO=0
299 MP=0
300 MS=0
301 MC1=0
302 MH1=0
303 MF1=0
304 MN1=0
305 MO1=0
306 MP1=0
307 MS1=0
308
'-----------------------------------------------------------------------'
309 'split the user input string describing the oligonucleotide
sequence
310 o_array0 = Split(oligo,"-",-1,1)
311 'a test message box to print the sequence with teh ends cut
off
312 o_array1 = Split(o_array0(1),"/",-1,1)
313 'count eac of the nucleotides in the sequence
314 i=0
315 For Each nucleotide In o_array1
```

```
316 i=i+1
317 Next
318 max_charge_index=CInt(max_charge_per_base)
319 j=i-1
320 'update the fragment array with the new found length of the
sequence
321 '---------- alpha ions (used to construct other ions)
322 ReDim negchargestates(i)
323 ReDim poschargestates(i)
324 ReDim aC(i)
325 ReDim aO(i)
326 ReDim aH(i)
327 ReDim aF(i)
328 ReDim aN(i)
329 ReDim aP(i)
330 ReDim a_S(i)
331 ReDim a_mz(i)
332 '---------- [a- base] ions
333 ReDim abC(i)
334 ReDim abH(i)
335 ReDim abF(i)
336 ReDim abN(i)
337 ReDim abO(i)
338 ReDim abP(i)
339 ReDim ab_S(i)
340 ReDim ab_mz(i)
341 '---------- b ions
342 ReDim bC(i)
343 ReDim bH(i)
344 ReDim bF(i)
345 ReDim bN(i)
346 ReDim bO(i)
347 ReDim bP(i)
348 ReDim bS(i)
349 ReDim b_mz(i)
350 '---------- c ions
351 ReDim cC(i)
352 ReDim cH(i)
353 ReDim cF(i)
354 ReDim cN(i)
355 ReDim cO(i)
356 ReDim cP(i)
357 ReDim cS(i)
358 ReDim c_mz(i)
359 '---------- d ions
360 ReDim dC(i)
361 ReDim dH(i)
362 ReDim dF(i)
363 ReDim dN(i)
364 ReDim d_O(i)
365 ReDim dP(i)
366 ReDim dS(i)
367 ReDim d_mz(i)
```

```
'---------- w ions
ReDim wC(i)
ReDim wH(i)
ReDim wF(i)
ReDim wN(i)
ReDim wO(i)
ReDim wP(i)
ReDim wS(i)
ReDim w_mz(i)
'---------- x ions
ReDim xC(i)
ReDim xH(i)
ReDim xF(i)
ReDim xN(i)
ReDim xO(i)
ReDim xP(i)
ReDim xS(i)
ReDim x_mz(i)
'---------- y ions
ReDim yC(i)
ReDim yH(i)
ReDim yF(i)
ReDim yN(i)
ReDim yO(i)
ReDim yP(i)
ReDim yS(i)
ReDim y_mz(i)
'---------- z ions
ReDim zC(i)
ReDim zH(i)
ReDim zF(i)
ReDim zN(i)
ReDim zO(i)
ReDim zP(i)
ReDim zS(i)
ReDim z_mz(i)
ReDim banana(i)
'----------a ions
ReDim a1C(i)
ReDim a1H(i)
ReDim a1F(i)
ReDim a1N(i)
ReDim a1O(i)
ReDim a1P(i)
ReDim a1S(i)
ReDim a1_mz(i)
'----------'
ReDim percent(j)
c11=0
For Each location in percent
percent(c11)=0
c11=c11+1
Next
```

```
421 'define boolean array for semi-visual sequence mapping
422 ReDim ab_location(j)
423 ReDim a_location(j)
424 ReDim b_location(j)
425 ReDim c_location(j)
426 ReDim d_location(j)
427 ReDim w_location(j)
428 ReDim x_location(j)
429 ReDim y_location(j)
430 ReDim z_location(j)
431 'fill the boolean array with zeros by default
432 c14=0
433 For Each location in ab_location
434 ab_location(c14)=0
435 a_location(c14)=0
436 b_location(c14)=0
437 c_location(c14)=0
438 d_location(c14)=0
439 w_location(c14)=0
440 x_location(c14)=0
441 y_location(c14)=0
442 z_location(c14)=0
443 c14=c14+1
444 Next
445 c14=0
446 'MsgBox c11
447 'define an array one less than the length of the sequence
448
'------------------------------------------------------------------------'
449 'add the mass of the 5' end to the structure
450 if o_array0(0) = "HO" then
451 'this is intentionally left blank to correct fot eh sugar construction
at the 0 index
452 MH=0
453 MC=0
454 MF=0
455 MN=0
456 MO=0
457 MP=0
458 MS=0
459 elseif o_array0(0) = "ph" then
460 MH=1
461 MO=3
462 MP=1
463 MC=0
464 MF=0
465 MN=0
466 MS=0
467 'dual biotin
468 elseif o_array0(0)= "bio2" then
469 MH=65
470 MO=13
471 MP=2
```

```
MC=36
MF=0
MN=6
MS=2
else o_array0(0) = "me"
MH=2
MC=1
MF=0
MN=0
MO=0
MP=0
MS=0
End if

'-------------------------------------------------------------------'
'add the mass of the 3' end to the structure
if o_array0(2) = "OH" then
MH1=1
MC1=0
MF1=0
MN1=0
MO1=1
MP1=0
MS1=0
'MsgBox MH
elseif o_array0(2) = "ph" then
MH1=2
MO1=4
MP1=1
MC1=0
MF1=0
MN1=0
MS1=0
'dual biotin
elseif o_array0(2)= "bio2" then
MH1=66
MO1=13
MP1=2
MC1=36
MF1=0
MN1=6
MS1=2
else o_array0(2) = "me"
MH1=3
MC1=1
MF1=0
MN1=0
MO1=0
MP1=0
MS1=0
End if

'-------------------------------------------------------------------'
```

```
523 c=0
524 For Each sugar In o_array1
525 var = Split(o_array1(c),",",-1,1)
526 'add the atoms of the sugar to the fragments
527 if c=0 then
528 'addition of the 5' end to the structure I think this needs to be
done before anyother fragments are developed from this ion
series
529 aC(c) = aC(c)+MC
530 aH(c) = aH(c)+MH
531 aF(c) = aF(c)+MF
532 aN(c) = aN(c)+MN
533 aO(c) = aO(c)+MO
534 aP(c) = aP(c)+MP
535 a_S(c) = a_S(c)+MS
536 bC(c)=bC(c)+MC
537 bH(c)=bH(c)+MH
538 bO(c)=bO(c)+MO
539 bF(c)=bF(c)+MF
540 bN(c)=bN(c)+MN
541 bP(c)=bP(c)+MP
542 bS(c)=bS(c)+MS
543 if var(0) = "r" then
544 aC(c)=aC(c)+5
545 aH(c)=aH(c)+3
546 aO(c)=aO(c)+3
547 aF(c)=aF(c)+0
548 aN(c)=aN(c)+0
549 aP(c)=aP(c)+0
550 a_S(c)=a_S(c)+0
551 bC(c)=bC(c)+5
552 bH(c)=bH(c)+7
553 bO(c)=bO(c)+4
554 bF(c)=bF(c)+0
555 bN(c)=bN(c)+0
556 bP(c)=bP(c)+0
557 bS(c)=bS(c)+0
558 cC(c)=bC(c)
559 cH(c)=bH(c)
560 cO(c)=bO(c)
561 cF(c)=bF(c)
562 cN(c)=bN(c)
563 cP(c)=bP(c)
564 cS(c)=bS(c)
565 ElseIF var(0) = "d" then
566 aC(c)=aC(c)+5
567 aH(c)=aH(c)+3
568 aO(c)=aO(c)+2
569 aF(c)=aF(c)+0
570 aN(c)=aN(c)+0
571 aP(c)=aP(c)+0
572 a_S(c)=a_S(c)+0
573 bC(c)=bC(c)+5
```

```
574 bH(c)=bH(c)+7
575 bO(c)=bO(c)+3
576 bF(c)=bF(c)+0
577 bN(c)=bN(c)+0
578 bP(c)=bP(c)+0
579 bS(c)=bS(c)+0
580 cC(c)=bC(c)
581 cH(c)=bH(c)
582 cO(c)=bO(c)
583 cF(c)=bF(c)
584 cN(c)=bN(c)
585 cP(c)=bP(c)
586 cS(c)=bS(c)
587 ElseIF var(0) = "f" then
588 aC(c)=aC(c)+5
589 aH(c)=aH(c)+2
590 aO(c)=aO(c)+2
591 aF(c)=aF(c)+1
592 aN(c)=aN(c)+0
593 aP(c)=aP(c)+0
594 a_S(c)=a_S(c)+0
595 bC(c)=bC(c)+5
596 bH(c)=bH(c)+6
597 bO(c)=bO(c)+3
598 bF(c)=bF(c)+1
599 bN(c)=bN(c)+0
600 bP(c)=bP(c)+0
601 bS(c)=bS(c)+0
602 cC(c)=bC(c)
603 cH(c)=bH(c)
604 cO(c)=bO(c)
605 cF(c)=bF(c)
606 cN(c)=bN(c)
607 cP(c)=bP(c)
608 cS(c)=bS(c)
609 ElseIF var(0) = "rm" then
610 aC(c)=aC(c)+6
611 aH(c)=aH(c)+5
612 aO(c)=aO(c)+3
613 aF(c)=aF(c)+0
614 aN(c)=aN(c)+0
615 aP(c)=aP(c)+0
616 a_S(c)=a_S(c)+0
617 bC(c)=bC(c)+6
618 bH(c)=bH(c)+9
619 bO(c)=bO(c)+4
620 bF(c)=bF(c)+0
621 bN(c)=bN(c)+0
622 bP(c)=bP(c)+0
623 bS(c)=bS(c)+0
624 cC(c)=bC(c)
625 cH(c)=bH(c)
626 cO(c)=bO(c)
```

```
627 cF(c)=bF(c)
628 cN(c)=bN(c)
629 cP(c)=bP(c)
630 cS(c)=bS(c)
631 Else var(0) = "rmoe"
632 aC(c)=aC(c)+8
633 aH(c)=aH(c)+9
634 aO(c)=aO(c)+4
635 aF(c)=aF(c)+0
636 aN(c)=aN(c)+0
637 aP(c)=aP(c)+0
638 a_S(c)=a_S(c)+0
639 bC(c)=bC(c)+8
640 bH(c)=bH(c)+13
641 bO(c)=bO(c)+4
642 bF(c)=bF(c)+0
643 bN(c)=bN(c)+0
644 bP(c)=bP(c)+0
645 bS(c)=bS(c)+0
646 cC(c)=bC(c)
647 cH(c)=bH(c)
648 cO(c)=bO(c)
649 cF(c)=bF(c)
650 cN(c)=bN(c)
651 cP(c)=bP(c)
652 cS(c)=bS(c)
653 End if
654 'for fragments other than the a-base the base of the first
nucleotide needs to be added
655 var5 = Split(o_array1(c),".",-1,1)
656 var7 = Split(var5(0),",",-1,1)
657 if var7(1) = "A" then
658 bC(c)=bC(c)+5
659 bH(c)=bH(c)+4
660 bO(c)=bO(c)+0
661 bF(c)=bF(c)+0
662 bN(c)=bN(c)+5
663 bP(c)=bP(c)+0
664 bS(c)=bS(c)+0
665 cC(c)=bC(c)
666 cH(c)=bH(c)
667 cO(c)=bO(c)
668 cF(c)=bF(c)
669 cN(c)=bN(c)
670 cP(c)=bP(c)
671 cS(c)=bS(c)
672 ElseIF var7(1) = "C" then
673 bC(c)=bC(c)+4
674 bH(c)=bH(c)+4
675 bO(c)=bO(c)+1
676 bF(c)=bF(c)+0
677 bN(c)=bN(c)+3
678 bP(c)=bP(c)+0
```

```
679 bS(c)=bS(c)+0
680 cC(c)=bC(c)
681 cH(c)=bH(c)
682 cO(c)=bO(c)
683 cF(c)=bF(c)
684 cN(c)=bN(c)
685 cP(c)=bP(c)
686 cS(c)=bS(c)
687 ElseIF var7(1) = "5C" then
688 bC(c)=bC(c)+5
689 bH(c)=bH(c)+7
690 bO(c)=bO(c)+1
691 bF(c)=bF(c)+0
692 bN(c)=bN(c)+3
693 bP(c)=bP(c)+0
694 bS(c)=bS(c)+0
695 cC(c)=bC(c)
696 cH(c)=bH(c)
697 cO(c)=bO(c)
698 cF(c)=bF(c)
699 cN(c)=bN(c)
700 cP(c)=bP(c)
701 cS(c)=bS(c)
702 ElseIF var7(1) = "G" then
703 bC(c)=bC(c)+5
704 bH(c)=bH(c)+4
705 bO(c)=bO(c)+1
706 bF(c)=bF(c)+0
707 bN(c)=bN(c)+5
708 bP(c)=bP(c)+0
709 bS(c)=bS(c)+0
710 cC(c)=bC(c)
711 cH(c)=bH(c)
712 cO(c)=bO(c)
713 cF(c)=bF(c)
714 cN(c)=bN(c)
715 cP(c)=bP(c)
716 cS(c)=bS(c)
717 ElseIF var7(1) = "22G" then
718 bC(c)=bC(c)+5+2
719 bH(c)=bH(c)+4+4
720 bO(c)=bO(c)+1
721 bF(c)=bF(c)+0
722 bN(c)=bN(c)+5
723 bP(c)=bP(c)+0
724 bS(c)=bS(c)+0
725 cC(c)=bC(c)
726 cH(c)=bH(c)
727 cO(c)=bO(c)
728 cF(c)=bF(c)
729 cN(c)=bN(c)
730 cP(c)=bP(c)
731 cS(c)=bS(c)
```

```
732 ElseIF var7(1) = "3483G" then
733 'this is the first mod creatd in this way 20220305, the second
base mod added after 5c where the delta between the standard is
added after the standard
734 bC(c)=bC(c)+5+11
735 bH(c)=bH(c)+4+15
736 bO(c)=bO(c)+1+4
737 bF(c)=bF(c)+0
738 bN(c)=bN(c)+5+1
739 bP(c)=bP(c)+0
740 bS(c)=bS(c)+0
741 cC(c)=bC(c)
742 cH(c)=bH(c)
743 cO(c)=bO(c)
744 cF(c)=bF(c)
745 cN(c)=bN(c)
746 cP(c)=bP(c)
747 cS(c)=bS(c)
748 ElseIF var7(1) = "T" then
749 bC(c)=bC(c)+5
750 bH(c)=bH(c)+5
751 bO(c)=bO(c)+2
752 bF(c)=bF(c)+0
753 bN(c)=bN(c)+2
754 bP(c)=bP(c)+0
755 bS(c)=bS(c)+0
756 cC(c)=bC(c)
757 cH(c)=bH(c)
758 cO(c)=bO(c)
759 cF(c)=bF(c)
760 cN(c)=bN(c)
761 cP(c)=bP(c)
762 cS(c)=bS(c)
763 Else var7(1) = "U"
764 bC(c)=bC(c)+4
765 bH(c)=bH(c)+3
766 bO(c)=bO(c)+2
767 bF(c)=bF(c)+0
768 bN(c)=bN(c)+2
769 bP(c)=bP(c)+0
770 bS(c)=bS(c)+0
771 cC(c)=bC(c)
772 cH(c)=bH(c)
773 cO(c)=bO(c)
774 cF(c)=bF(c)
775 cN(c)=bN(c)
776 cP(c)=bP(c)
777 cS(c)=bS(c)
778 End If
779 if var5(1) = "p" then
780 'each atomic addition here represents the structure fo the
~typically~ phosphate backbone link between sugars
781 'here an standard phosphate backbone and phosphothiorate are
```

```
included
782 cC(c)=cC(c)+0
783 cH(c)=cH(c)+0
784 cO(c)=cO(c)+2
785 cF(c)=cF(c)+0
786 cN(c)=cN(c)+0
787 cP(c)=cP(c)+1
788 cS(c)=cS(c)+0
789 Else var5(1) = "s"
790 'this is the phosphothiorate option
791 cC(c)=cC(c)+0
792 cH(c)=cH(c)+0
793 cO(c)=cO(c)+1
794 cF(c)=cF(c)+0
795 cN(c)=cN(c)+0
796 cP(c)=cP(c)+1
797 cS(c)=cS(c)+1
798 End if
799 'beep boop bop at index zero addition of one proton fufils
structure, many would not ocnsider it a real fragmetn but i think it
makes sense ot fiish it here outside the for loop
800 abC(0) = aC(0)+0
801 'addition of proton to the 3' position to resolve where a link would
go for downstream fragments the proton is added at the base posiiton
in the base calculation section above
802 abH(0) = aH(0)+1
803 abF(0) = aF(0)+0
804 abN(0) = aN(0)+0
805 abO(0) = aO(0)+0
806 abP(0) = aP(0)+0
807 ab_S(0) = a_S(0)+0
808 'at this point the a-b(0) is complete to construct the other
fragments addition of the base will be nessicary
809 'add the hydrogen at the end repeat unit
810 bH(0) = bH(0)+1
811 bF(0) = bF(0)+0
812 bN(0) = bN(0)+0
813 bO(0) = bO(0)+0
814 bP(0) = bP(0)+0
815 bS(0) = bS(0)+0
816
'------------------------------------------------------------
817 else
818 var = Split(o_array1(c),",",-1,1)
819 if var(0) = "r" then
820 d=c-1
821 aC(c)=aC(d)+5
822 aH(c)=aH(d)+3
823 'additonal hydrogens to saturate the previous sugar can be added
here
824 aH(c)=aH(c)+4
825 aO(c)=aO(d)+3
826 aF(c)=aF(d)+0
```

```
827 aN(c)=aN(d)+0
828 aP(c)=aP(d)+0
829 a_S(c)=a_S(d)+0
830 'additional hydrogens to saturate previous sugar not needed for the
b ions
831 bC(c)=bC(d)+5
832 bH(c)=bH(d)+7
833 bO(c)=bO(d)+4
834 bF(c)=bF(d)+0
835 bN(c)=bN(d)+0
836 bP(c)=bP(d)+0
837 bS(c)=bS(d)+0
838 cC(c)=bC(c)
839 cH(c)=bH(c)
840 cO(c)=bO(c)
841 cF(c)=bF(c)
842 cN(c)=bN(c)
843 cP(c)=bP(c)
844 cS(c)=bS(c)
845 ElseIF var(0) = "d" then
846 d=c-1
847 aC(c)=aC(d)+5
848 aH(c)=aH(d)+3
849 'additonal hydrogens to saturate the previous sugar can be added
here
850 aH(c)=aH(c)+4
851 aO(c)=aO(d)+2
852 aF(c)=aF(d)+0
853 aN(c)=aN(d)+0
854 aP(c)=aP(d)+0
855 a_S(c)=a_S(d)+0
856 bC(c)=bC(d)+5
857 bH(c)=bH(d)+7
858 bO(c)=bO(d)+3
859 bF(c)=bF(d)+0
860 bN(c)=bN(d)+0
861 bP(c)=bP(d)+0
862 bS(c)=bS(d)+0
863 cC(c)=bC(c)
864 cH(c)=bH(c)
865 cO(c)=bO(c)
866 cF(c)=bF(c)
867 cN(c)=bN(c)
868 cP(c)=bP(c)
869 cS(c)=bS(c)
870 ElseIF var(0) = "f" then
871 d=c-1
872 aC(c)=aC(d)+5
873 aH(c)=aH(d)+2
874 'additonal hydrogens to saturate the previous sugar can be added
here
875 aH(c)=aH(c)+4
876 aO(c)=aO(d)+2
```

```
aF(c)=aF(d)+1
aN(c)=aN(d)+0
aP(c)=aP(d)+0
a_S(c)=a_S(d)+0
bC(c)=bC(d)+5
bH(c)=bH(d)+6
bO(c)=bO(d)+3
bF(c)=bF(d)+1
bN(c)=bN(d)+0
bP(c)=bP(d)+0
bS(c)=bS(d)+0
cC(c)=bC(c)
cH(c)=bH(c)
cO(c)=bO(c)
cF(c)=bF(c)
cN(c)=bN(c)
cP(c)=bP(c)
cS(c)=bS(c)
ElseIF var(0) = "rm" then
d=c-1
aC(c)=aC(d)+6
aH(c)=aH(d)+5
'additonal hydrogens to saturate the previous sugar can be added
here
aH(c)=aH(c)+4
aO(c)=aO(d)+3
aF(c)=aF(d)+0
aN(c)=aN(d)+0
aP(c)=aP(d)+0
a_S(c)=a_S(d)+0
bC(c)=bC(d)+6
bH(c)=bH(d)+9
bO(c)=bO(d)+4
bF(c)=bF(d)+0
bN(c)=bN(d)+0
bP(c)=bP(d)+0
bS(c)=bS(d)+0
cC(c)=bC(c)
cH(c)=bH(c)
cO(c)=bO(c)
cF(c)=bF(c)
cN(c)=bN(c)
cP(c)=bP(c)
cS(c)=bS(c)
Else var(0) = "rmoe"
d=c-1
aC(c)=aC(d)+8
aH(c)=aH(d)+9
'additonal hydrogens to saturate the previous sugar can be added
here
aH(c)=aH(c)+4
aO(c)=aO(d)+4
aF(c)=aF(d)+0
```

```
928 aN(c)=aN(d)+0
929 aP(c)=aP(d)+0
930 a_S(c)=a_S(d)+0
931 bC(c)=bC(d)+8
932 bH(c)=bH(d)+13
933 bO(c)=bO(d)+4
934 bF(c)=bF(d)+0
935 bN(c)=bN(d)+0
936 bP(c)=bP(d)+0
937 bS(c)=bS(d)+0
938 cC(c)=bC(c)
939 cH(c)=bH(c)
940 cO(c)=bO(c)
941 cF(c)=bF(c)
942 cN(c)=bN(c)
943 cP(c)=bP(c)
944 cS(c)=bS(c)
945 End if
946 d=c-1
947 var5 = Split(o_array1(d),".",-1,1)
948 'final isolation of the oligonucleotide element from the string by
isolation with the other delimiter on the other side of the
element
949 'adding the previous linker d ot the current fragment (this section of
the code is in the index>=1 so it should be correct)[remembering that
indices start at zero]
950 if var5(1) = "p" then
951 'each atomic addition here represents the structure fo the
~typically~ phosphate backbone link between sugars
952 'here an standard phosphate backbone and phosphothiorate are
included
953 aC(c)=aC(c)+0
954 aH(c)=aH(c)+1
955 aO(c)=aO(c)+3
956 aF(c)=aF(c)+0
957 aN(c)=aN(c)+0
958 aP(c)=aP(c)+1
959 a_S(c)=a_S(c)+0
960 bC(c)=aC(c)
961 bH(c)=aH(c)
962 bO(c)=aO(c)
963 bF(c)=aF(c)
964 bN(c)=aN(c)
965 bP(c)=aP(c)
966 bS(c)=a_S(c)
967 'account for the previous linker
968 cC(c)=bC(c)
969 cH(c)=bH(c)
970 cO(c)=bO(c)
971 cF(c)=bF(c)
972 cN(c)=bN(c)
973 cP(c)=bP(c)
974 cS(c)=bS(c)
```

```
Else var5(1) = "s"
'this is the phosphothiorate option
aC(c)=aC(c)+0
aH(c)=aH(c)+1
aO(c)=aO(c)+2
aF(c)=aF(c)+0
aN(c)=aN(c)+0
aP(c)=aP(c)+1
a_S(c)=a_S(c)+1
bC(c)=aC(c)
bH(c)=aH(c)
bO(c)=aO(c)
bF(c)=aF(c)
bN(c)=aN(c)
bP(c)=aP(c)
bS(c)=a_S(c)
'account for the previous linker
cC(c)=bC(c)
cH(c)=bH(c)
cO(c)=bO(c)
cF(c)=bF(c)
cN(c)=bN(c)
cP(c)=bP(c)
cS(c)=bS(c)
'MsgBox "s tree"
End if
'taking the previous base and add it to the a fragment
var7 = Split(var5(0),",",-1,1)
if var7(1) = "A" then
aC(c)=aC(c)+5
aH(c)=aH(c)+4
aO(c)=aO(c)+0
aF(c)=aF(c)+0
aN(c)=aN(c)+5
aP(c)=aP(c)+0
a_S(c)=a_S(c)+0
ElseIF var7(1) = "C" then
aC(c)=aC(c)+4
aH(c)=aH(c)+4
aO(c)=aO(c)+1
aF(c)=aF(c)+0
aN(c)=aN(c)+3
aP(c)=aP(c)+0
a_S(c)=a_S(c)+0
ElseIF var7(1) = "5C" then
aC(c)=aC(c)+5
aH(c)=aH(c)+7
aO(c)=aO(c)+1
aF(c)=aF(c)+0
aN(c)=aN(c)+3
aP(c)=aP(c)+0
a_S(c)=a_S(c)+0
ElseIF var7(1) = "G" then
```

```
aC(c)=aC(c)+5
aH(c)=aH(c)+4
aO(c)=aO(c)+1
aF(c)=aF(c)+0
aN(c)=aN(c)+5
aP(c)=aP(c)+0
a_S(c)=a_S(c)+0
ElseIF var7(1) = "22G" then
aC(c)=aC(c)+5+2
aH(c)=aH(c)+4+4
aO(c)=aO(c)+1
aF(c)=aF(c)+0
aN(c)=aN(c)+5
aP(c)=aP(c)+0
a_S(c)=a_S(c)+0
ElseIF var7(1) = "3483G" then
aC(c)=aC(c)+5+11
aH(c)=aH(c)+4+15
aO(c)=aO(c)+1+4
aF(c)=aF(c)+0
aN(c)=aN(c)+5+1
aP(c)=aP(c)+0
a_S(c)=a_S(c)+0
ElseIF var7(1) = "T" then
aC(c)=aC(c)+5
aH(c)=aH(c)+5
aO(c)=aO(c)+2
aF(c)=aF(c)+0
aN(c)=aN(c)+2
aP(c)=aP(c)+0
a_S(c)=a_S(c)+0
Else var7(1) = "U"
aC(c)=aC(c)+4
aH(c)=aH(c)+3
aO(c)=aO(c)+2
aF(c)=aF(c)+0
aN(c)=aN(c)+2
aP(c)=aP(c)+0
a_S(c)=a_S(c)+0
End If
'take the current base and add to the current b fragment
var5 = Split(o_array1(c),".",-1,1)
var7 = Split(var5(0),",",-1,1)
if var7(1) = "A" then
bC(c)=aC(c)+5
bH(c)=aH(c)+4
'additonal hydrogens to saturate the previous sugar possibly belongs
here
bH(c)=bH(c)+4
bO(c)=aO(c)+0
bF(c)=aF(c)+0
bN(c)=aN(c)+5
bP(c)=aP(c)+0
```

```
1080 bS(c)=a_S(c)+0
1081 'the base addition is the same for the b and c ions
1082 cC(c)=bC(c)
1083 cH(c)=bH(c)
1084 cO(c)=bO(c)
1085 cF(c)=bF(c)
1086 cN(c)=bN(c)
1087 cP(c)=bP(c)
1088 cS(c)=bS(c)
1089 ElseIF var7(1) = "C" then
1090 bC(c)=aC(c)+4
1091 bH(c)=aH(c)+4
1092 'additonal hydrogens to saturate the previous sugar possibly belongs
here
1093 bH(c)=bH(c)+4
1094 bO(c)=aO(c)+1
1095 bF(c)=aF(c)+0
1096 bN(c)=aN(c)+3
1097 bP(c)=aP(c)+0
1098 bS(c)=a_S(c)+0
1099 'the base addition is the same for the b and c ions
1100 cC(c)=bC(c)
1101 cH(c)=bH(c)
1102 cO(c)=bO(c)
1103 cF(c)=bF(c)
1104 cN(c)=bN(c)
1105 cP(c)=bP(c)
1106 cS(c)=bS(c)
1107 ElseIF var7(1) = "5C" then
1108 bC(c)=aC(c)+5
1109 bH(c)=aH(c)+7
1110 'additonal hydrogens to saturate the previous sugar possibly belongs
here
1111 bH(c)=bH(c)+4
1112 bO(c)=aO(c)+1
1113 bF(c)=aF(c)+0
1114 bN(c)=aN(c)+3
1115 bP(c)=aP(c)+0
1116 bS(c)=a_S(c)+0
1117 'the base addition is the same for the b and c ions
1118 cC(c)=bC(c)
1119 cH(c)=bH(c)
1120 cO(c)=bO(c)
1121 cF(c)=bF(c)
1122 cN(c)=bN(c)
1123 cP(c)=bP(c)
1124 cS(c)=bS(c)
1125 ElseIF var7(1) = "G" then
1126 bC(c)=aC(c)+5
1127 bH(c)=aH(c)+4
1128 'additonal hydrogens to saturate the previous sugar possibly belongs
here
1129 bH(c)=bH(c)+4
```

```
1130 bO(c)=aO(c)+1
1131 bF(c)=aF(c)+0
1132 bN(c)=aN(c)+5
1133 bP(c)=aP(c)+0
1134 bS(c)=a_S(c)+0
1135 'the base addition is the same for the b and c ions
1136 cC(c)=bC(c)
1137 cH(c)=bH(c)
1138 cO(c)=bO(c)
1139 cF(c)=bF(c)
1140 cN(c)=bN(c)
1141 cP(c)=bP(c)
1142 cS(c)=bS(c)
1143 ElseIF var7(1) = "22G" then
1144 bC(c)=aC(c)+5+2
1145 bH(c)=aH(c)+4+4
1146 'additonal hydrogens to saturate the previous sugar possibly belongs
here
1147 bH(c)=bH(c)+4
1148 bO(c)=aO(c)+1
1149 bF(c)=aF(c)+0
1150 bN(c)=aN(c)+5
1151 bP(c)=aP(c)+0
1152 bS(c)=a_S(c)+0
1153 'the base addition is the same for the b and c ions
1154 cC(c)=bC(c)
1155 cH(c)=bH(c)
1156 cO(c)=bO(c)
1157 cF(c)=bF(c)
1158 cN(c)=bN(c)
1159 cP(c)=bP(c)
1160 cS(c)=bS(c)
1161 ElseIF var7(1) = "3483G" then
1162 bC(c)=aC(c)+5+11
1163 bH(c)=aH(c)+4+15
1164 'additonal hydrogens to saturate the previous sugar possibly belongs
here
1165 bH(c)=bH(c)+4
1166 bO(c)=aO(c)+1+4
1167 bF(c)=aF(c)+0
1168 bN(c)=aN(c)+5+1
1169 bP(c)=aP(c)+0
1170 bS(c)=a_S(c)+0
1171 'the base addition is the same for the b and c ions
1172 cC(c)=bC(c)
1173 cH(c)=bH(c)
1174 cO(c)=bO(c)
1175 cF(c)=bF(c)
1176 cN(c)=bN(c)
1177 cP(c)=bP(c)
1178 cS(c)=bS(c)
1179 ElseIF var7(1) = "T" then
1180 bC(c)=aC(c)+5
```

```
1181 bH(c)=aH(c)+5
1182 'additonal hydrogens to saturate the previous sugar possibly belongs
here
1183 bH(c)=bH(c)+4
1184 bO(c)=aO(c)+2
1185 bF(c)=aF(c)+0
1186 bN(c)=aN(c)+2
1187 bP(c)=aP(c)+0
1188 bS(c)=a_S(c)+0
1189 'the base addition is the same for the b and c ions
1190 cC(c)=bC(c)
1191 cH(c)=bH(c)
1192 cO(c)=bO(c)
1193 cF(c)=bF(c)
1194 cN(c)=bN(c)
1195 cP(c)=bP(c)
1196 cS(c)=bS(c)
1197 Else var7(1) = "U"
1198 bC(c)=aC(c)+4
1199 bH(c)=aH(c)+3
1200 'additonal hydrogens to saturate the previous sugar possibly belongs
here
1201 bH(c)=bH(c)+4
1202 bO(c)=aO(c)+2
1203 bF(c)=aF(c)+0
1204 bN(c)=aN(c)+2
1205 bP(c)=aP(c)+0
1206 bS(c)=a_S(c)+0
1207 'the base addition is the same for the b and c ions
1208 cC(c)=bC(c)
1209 cH(c)=bH(c)
1210 cO(c)=bO(c)
1211 cF(c)=bF(c)
1212 cN(c)=bN(c)
1213 cP(c)=bP(c)
1214 cS(c)=bS(c)
1215 End If
1216 'since the base addition is the last event in the construction of the b
ion it would seem the final addition to complete the c ion must be done
after the b ion
1217 'add the current linker
1218 var5 = Split(o_array1(c),".",-1,1)
1219 if c<j then
1220 if var5(1) = "p" then
1221 'each atomic addition here represents the structure fo the
~typically~ phosphate backbone link between
sugars
1222 'here an standard phosphate backbone and phosphothiorate are
included
1223 cC(c)=cC(c)+0
1224 cH(c)=cH(c)+0
1225 cO(c)=cO(c)+3
1226 cF(c)=cF(c)+0
```

```
cN(c)=cN(c)+0
cP(c)=cP(c)+1
cS(c)=cS(c)+0
Else var5(1) = "s"
'this is the phosphothiorate option
cC(c)=cC(c)+0
cH(c)=cH(c)+0
cO(c)=cO(c)+2
cF(c)=cF(c)+0
cN(c)=cN(c)+0
cP(c)=cP(c)+1
cS(c)=cS(c)+1
'MsgBox "s tree"
End if
else
End If
'addition of two protons to cap the two carbons where additonal bonds
will be made in the subsequent fragment
abC(c) = aC(c)+0
abH(c) = aH(c)+2
abF(c) = aF(c)+0
abN(c) = aN(c)+0
abO(c) = aO(c)+0
abP(c) = aP(c)+0
ab_S(c) = a_S(c)+0
if c<j then
'add to the 3' position the O and H which matched my structure
design i think
bC(c) = bC(c)+0
bH(c) = bH(c)+1
bF(c) = bF(c)+0
bN(c) = bN(c)+0
bO(c) = bO(c)+1
bP(c) = bP(c)+0
bS(c) = bS(c)+0
Else c=j
MC=bC(c)+MC1
MH=bH(c)+MH1
MF=bF(c)+MF1
MN=bN(c)+MN1
MO=bO(c)+MO1
MP=bP(c)+MP1
MS=bS(c)+MS1
Mstr="M - H, C"+CStr(MC)+"H"+CStr(MH)+"F"+CStr(MF)+"N"+CStr(MN
)+"O"+CStr(MO)+"P"+CStr(MP)+"S"+CStr(MS)+" "
End If
End If
dC(c) = cC(c)+0
dH(c) = cH(c)+2
dF(c) = cF(c)+0
dN(c) = cN(c)+0
d_O(c) = cO(c)+1
dP(c) = cP(c)+0
```

```
dS(c) = cS(c)+0
a1C(c) = bC(c)-0
a1H(c) = bH(c)-0
a1F(c) = bF(c)-0
a1N(c) = bN(c)-0
a1O(c) = bO(c)-1
a1P(c) = bP(c)-0
a1S(c) = bS(c)-0
c=c+1
Next
j=c
c=0
d=j-1
For Each sugar In o_array1
wC(c) = MC-a1C(d)
wH(c) = MH-a1H(d)+1
wF(c) = MF-a1F(d)
wN(c) = MN-a1N(d)
wO(c) = MO-a1O(d)
wP(c) = MP-a1P(d)
wS(c) = MS-a1S(d)
xC(c) = wC(c)
xH(c) = wH(c)-2
xF(c) = wF(c)
xN(c) = wN(c)
xO(c) = wO(c)-1
xP(c) = wP(c)
xS(c) = wS(c)
yC(c) = MC-cC(d)
yH(c) = MH-cH(d)-1
yF(c) = MF-cF(d)
yN(c) = MN-cN(d)
yO(c) = MO-cO(d)
yP(c) = MP-cP(d)
yS(c) = MS-cS(d)
zC(c) = yC(c)
zH(c) = yH(c)-1
zF(c) = yF(c)
zN(c) = yN(c)
zO(c) = yO(c)-1
zP(c) = yP(c)
zS(c) = yS(c)
d=d-1
c=c+1
Next
d=c-1
wC(d) = wC(d)
wH(d) = wH(d)
wF(d) = wF(d)
wN(d) = wN(d)
wO(d) = wO(d)
wP(d) = wP(d)
wS(d) = wS(d)
```

```
'-------------------------------------------------'
c=0
C_ = 12
H = 1.007825
F = 18.998403
N = 14.003074
O = 15.994915
P = 30.973763
S = 31.972072
Mmstr=Cstr(MC*C_+MH*H+MF*F+MN*N+MO*O+MP*P+MS*S)
q=1
c=0
For Each z in negchargestates
negchargestates(c)=Cstr((MC*C_+(MH+1-q)*H+MF*F+MN*N+MO*O+MP*P+MS*S)/q)
poschargestates(c)=Cstr((MC*C_+(MH+1+q)*H+MF*F+MN*N+MO*O+MP*P+MS*S)/q)
q=q+1
c=c+1
Next
q=0
c=0
For Each sugar In o_array1
a1_mz(c) =CStr(a1C(c)*C_+a1H(c)*H+a1F(c)*F+a1N(c)*N+a1O(c)*O+a1P(c)*P+
a1S(c)*S)
ab_mz(c) =CStr(abC(c)*C_+abH(c)*H+abF(c)*F+abN(c)*N+abO(c)*O+abP(c)*P+
ab_S(c)*S)
b_mz(c) =CStr(bC(c)*C_+bH(c)*H+bF(c)*F+bN(c)*N+bO(c)*O+bP(c)*P+bS(c)*
S)
c_mz(c) =CStr(cC(c)*C_+cH(c)*H+cF(c)*F+cN(c)*N+cO(c)*O+cP(c)*P+cS(c)*
S)
d_mz(c) =CStr(dC(c)*C_+dH(c)*H+dF(c)*F+dN(c)*N+d_O(c)*O+dP(c)*P+dS(c
)*S)
w_mz(c) =CStr(wC(c)*C_+wH(c)*H+wF(c)*F+wN(c)*N+wO(c)*O+wP(c)*P+wS(c)*
S)
x_mz(c) =CStr(xC(c)*C_+xH(c)*H+xF(c)*F+xN(c)*N+xO(c)*O+xP(c)*P+xS(c)*
S)
y_mz(c) =CStr(yC(c)*C_+yH(c)*H+yF(c)*F+yN(c)*N+yO(c)*O+yP(c)*P+yS(c)*
S)
z_mz(c) =CStr(zC(c)*C_+zH(c)*H+zF(c)*F+zN(c)*N+zO(c)*O+zP(c)*P+zS(c)*
S)
c=c+1
Next
if c4=0 then
Dim f
Set fso = CreateObject("Scripting.FileSystemObject")
Set f = fso.OpenTextFile(filevar(ccc), 2, True)
c=0
q=c+1
f.Write oligo&Chr(13)
f.Write Mmstr&Chr(13)
f.Write Mstr&Chr(13)
f.Write "molecular charge states"&Chr(13)
c=0
g=c+1
```

```
1374 For Each z in negchargestates
1375 f.Write negchargestates(c)&", charge=-"&g&Chr(13)
1376 f.Write poschargestates(c)&", charge=+"&g&Chr(13)
1377 c=c+1
1378 g=g+1
1379 Next
1380 c=0
1381 f.Write "all charge states are (1-)"&Chr(13)
1382 For Each sugar In o_array1
1383 If q=1 then
1384 ElseIf q<i then
1385 f.Write "a-b"+CStr(q)+Chr(44)+Chr(9)+ab_mz(c)+Chr(44)+Chr(9)+"C"+
CStr(abC(c))+"H"+CStr(abH(c))+"F"+CStr(abF(c))+"N"+CStr(abN(c))+
"O"+CStr(abO(c))+"P"+CStr(abP(c))+"S"+CStr(ab_S(c))+" "&Chr(13)
1386 End If
1387 c=c+1
1388 q=q+1
1389 Next
1390 c=0
1391 q=c+1
1392 For Each sugar In o_array1
1393 if q<i then
1394 f.Write "a"+CStr(q)+Chr(44)+Chr(9)+a1_mz(c)+Chr(44)+Chr(9)+"C"+
CStr(a1C(c))+"H"+CStr(a1H(c))+"F"+CStr(a1F(c))+"N"+CStr(a1N(c))+
"O"+CStr(a1O(c))+"P"+CStr(a1P(c))+"S"+CStr(a1S(c))+" "&Chr(13)
1395 End If
1396 c=c+1
1397 q=q+1
1398 Next
1399 c=0
1400 q=c+1
1401 For Each sugar In o_array1
1402 if q<i then
1403 f.Write "b"+CStr(q)+Chr(44)+Chr(9)+b_mz(c)+Chr(44)+Chr(9)+"C"+
CStr(bC(c))+"H"+CStr(bH(c))+"F"+CStr(bF(c))+"N"+CStr(bN(c))+"O"+
CStr(bO(c))+"P"+CStr(bP(c))+"S"+CStr(bS(c))+" "&Chr(13)
1404 End If
1405 c=c+1
1406 q=q+1
1407 Next
1408 c=0
1409 q=c+1
1410 For Each sugar In o_array1
1411 if q<i then
1412 f.Write "c"+CStr(q)+Chr(44)+Chr(9)+c_mz(c)+Chr(44)+Chr(9)+"C"+
CStr(cC(c))+"H"+CStr(cH(c))+"F"+CStr(cF(c))+"N"+CStr(cN(c))+"O"+
CStr(cO(c))+"P"+CStr(cP(c))+"S"+CStr(cS(c))+" "&Chr(13)
1413 End If
1414 c=c+1
1415 q=q+1
1416 Next
1417 c=0
1418 q=c+1
```

```
For Each sugar In o_array1
if q<i then
f.Write "d"+CStr(q)+Chr(44)+Chr(9)+d_mz(c)+Chr(44)+Chr(9)+"C"+
CStr(dC(c))+"H"+CStr(dH(c))+"F"+CStr(dF(c))+"N"+CStr(dN(c))+"O"+
CStr(d_O(c))+"P"+CStr(dP(c))+"S"+CStr(dS(c))+" "&Chr(13)
End If
c=c+1
q=q+1
Next
c=1
For Each sugar In o_array1
if c<i then
f.Write "w"+CStr(c)+Chr(44)+Chr(9)+w_mz(c)+Chr(44)+Chr(9)+"C"+
CStr(wC(c))+"H"+CStr(wH(c))+"F"+CStr(wF(c))+"N"+CStr(wN(c))+"O"+
CStr(wO(c))+"P"+CStr(wP(c))+"S"+CStr(wS(c))+" "&Chr(13)
End If
c=c+1
Next
c=1
For Each sugar In o_array1
if c<i then
f.Write "x"+CStr(c)+Chr(44)+Chr(9)+x_mz(c)+Chr(44)+Chr(9)+"C"+
CStr(xC(c))+"H"+CStr(xH(c))+"F"+CStr(xF(c))+"N"+CStr(xN(c))+"O"+
CStr(xO(c))+"P"+CStr(xP(c))+"S"+CStr(xS(c))+" "&Chr(13)
End If
c=c+1
Next
c=1
For Each sugar In o_array1
if c<i then
f.Write "y"+CStr(c)+Chr(44)+Chr(9)+y_mz(c)+Chr(44)+Chr(9)+"C"+
CStr(yC(c))+"H"+CStr(yH(c))+"F"+CStr(yF(c))+"N"+CStr(yN(c))+"O"+
CStr(yO(c))+"P"+CStr(yP(c))+"S"+CStr(yS(c))+" "&Chr(13)
End If
c=c+1
Next
c=1
For Each sugar In o_array1
if c<i then
f.Write "z"+CStr(c)+Chr(44)+Chr(9)+z_mz(c)+Chr(44)+Chr(9)+"C"+
CStr(zC(c))+"H"+CStr(zH(c))+"F"+CStr(zF(c))+"N"+CStr(zN(c))+"O"+
CStr(zO(c))+"P"+CStr(zP(c))+"S"+CStr(zS(c))+" "&Chr(13)
End If
c=c+1
Next
f.Close
else
End If
ccc=ccc+1
'---
'insert peak comparison here
Set fso = CreateObject("Scripting.FileSystemObject")
Set f6 = fso.OpenTextFile(fragfilepath+"\"+"report.txt", 8, True)
```

```
1462 '---
1463 c5=0
1464 'MsgBox max_charge_index
1465 For precursor=0 to max_charge_index
1466 oligoprecursor_H=CDbl(negchargestates(c5))+CDbl(ppmrange2_precursor)
1467 oligoprecursor_L=CDbl(negchargestates(c5))-CDbl(ppmrange2_precursor)
1468 if CDbl(oligoprecursor)<CDbl(oligoprecursor_H) and CDbl(oligoprecursor
)>CDbl(oligoprecursor_L) then
1469 f6.Write CStr(tempfile(c8))&Chr(13)
1470 f6.Write "neg"+Chr(44)+Chr(9)+"exp="+CStr(oligoprecursor)+Chr(44)+Chr
(9)+"theo="+CStr(negchargestates(c5))+Chr(44)+Chr(9)+oligoname(ccc-1
)+Chr(44)+Chr(9)+"charge="+CStr(precursor+1)&Chr(13)
1471 c9=0
1472 d1=0
1473 c12=0
1474 For Each location in percent
1475 c12=c12+1
1476 Next
1477 For Each sugar In o_array1
1478 d1=c9+1
1479 'a1_mz_n(c)=CDbl(a1_mz(c))
1480 'ok i think if i reconstruct the high low ppm range value in the
onditional line at 1418 I can have the multiple charge state
looped in the comparison
1481 'a1_mz_L
=CDbl(a1_mz(c9))-CDbl(a1_mz(c9))*ppmrange2
1482 'a1_mz_H
=CDbl(a1_mz(c9))+CDbl(a1_mz(c9))*ppmrange2
1483 'ab_mz_L
=CDbl(ab_mz(c9))-CDbl(ab_mz(c9))*ppmrange2
1484 'ab_mz_H
=CDbl(ab_mz(c9))+CDbl(ab_mz(c9))*ppmrange2
1485 'b_mz_L
=CDbl(b_mz(c9))-CDbl(b_mz(c9))*ppmrange2
1486 'b_mz_H
=CDbl(b_mz(c9))+CDbl(b_mz(c9))*ppmrange2
1487 'c_mz_L
=CDbl(c_mz(c9))-CDbl(c_mz(c9))*ppmrange2
1488 'c_mz_H
=CDbl(c_mz(c9))+CDbl(c_mz(c9))*ppmrange2
1489 'd_mz_L
=CDbl(d_mz(c9))-CDbl(d_mz(c9))*ppmrange2
1490 'd_mz_H
=CDbl(d_mz(c9))+CDbl(d_mz(c9))*ppmrange2
1491 'w_mz_L
=CDbl(w_mz(c9))-CDbl(w_mz(c9))*ppmrange2
1492 'w_mz_H
=CDbl(w_mz(c9))+CDbl(w_mz(c9))*ppmrange2
1493 'x_mz_L
=CDbl(x_mz(c9))-CDbl(x_mz(c9))*ppmrange2
1494 'x_mz_H
=CDbl(x_mz(c9))+CDbl(x_mz(c9))*ppmrange2
1495 'y_mz_L
```

```
=CDbl(y_mz(c9))-CDbl(y_mz(c9))*ppmrange2
1496 'y_mz_H
=CDbl(y_mz(c9))+CDbl(y_mz(c9))*ppmrange2
1497 'z_mz_L
=CDbl(z_mz(c9))-CDbl(z_mz(c9))*ppmrange2
1498 'z_mz_H
=CDbl(z_mz(c9))+CDbl(z_mz(c9))*ppmrange2
1499 'MsgBox a1_mz_n_L
1500 'this is probably the place to put annother for loop to itterate
through the fragmtn ion signals
1501 if charge_comparison="y" then
1502 'c10 is the counter to step through all the read fragment signal
lines from the mgf file
1503 c10=0
1504 For Each signal In oligofragments
1505 For z=1 to max_charge_fragment
1506 if oligofragments(c10)<CDbl((a1_mz(c9)-(z-1)*H)/z)+
CDbl((a1_mz(c9)-(z-1)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((a1_mz(c9)-(z-1)*H)/z)-CDbl
((a1_mz(c9)-(z-1)*H)/z)*ppmrange2 and d1<c12 and CDbl
(oligofragments_zs(c10))=z then
1507 'annotate here
1508 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
1509 f6.Write "a"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr(
9)+Chr(45)+CStr(z)+Chr(44)+Chr(9)+CStr((a1_mz(c9
)-(z-1)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(a1C(c9))+
"H"+CStr(a1H(c9)-(z-1))+"F"+CStr(a1F(c9))+"N"+
CStr(a1N(c9))+"O"+CStr(a1O(c9))+"P"+CStr(a1P(c9
))+"S"+CStr(a1S(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((a1_mz(c9)-(z-1)*H)/z))/((
a1_mz(c9)-(z-1)*H)/z)*1000000),4))&Chr(13)
1510 percent(d1)=1
1511 a_location(d1)=1
1512 else
1513 End if
1514 if oligofragments(c10)<CDbl((ab_mz(c9)-(z-1)*H)/z)+
CDbl((ab_mz(c9)-(z-1)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((ab_mz(c9)-(z-1)*H)/z)-CDbl
((ab_mz(c9)-(z-1)*H)/z)*ppmrange2 and d1<c12 and CDbl
(oligofragments_zs(c10))=z then
1515 'Cmpd(2).Annotations.AddAnnotation ab_mz_L,
prime_int*1.2, ab_mz_H, prime_int*1.2,
prime_int*1.6, "a-b"+CStr(d), True
1516 'annotate here
1517 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
1518 f6.Write "ab"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr
(9)+Chr(45)+CStr(z)+Chr(44)+Chr(9)+CStr((ab_mz(c9
)-(z-1)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(abC(c9))+
"H"+CStr(abH(c9)-(z-1))+"F"+CStr(abF(c9))+"N"+
CStr(abN(c9))+"O"+CStr(abO(c9))+"P"+CStr(abP(c9
))+"S"+CStr(ab_S(c9))+Chr(44)+Chr(9)+CStr(Round
(((oligofragments(c10)-((ab_mz(c9)-(z-1)*H)/z
))/((ab_mz(c9)-(z-1)*H)/z)*1000000),4))&Chr(13)
```

```
percent(d1)=1
ab_location(d1)=1
else
end if
if oligofragments(c10)<CDbl((b_mz(c9)-(z-1)*H)/z)+
CDbl((b_mz(c9)-(z-1)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((b_mz(c9)-(z-1)*H)/z)-CDbl((
b_mz(c9)-(z-1)*H)/z)*ppmrange2 and d1<c12 and CDbl(
oligofragments_zs(c10))=z then
'annotate here
f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
f6.Write "b"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr(
9)+Chr(45)+CStr(z)+Chr(44)+Chr(9)+CStr((b_mz(c9
)-(z-1)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(bC(c9))+"H"
+CStr(bH(c9)-(z-1))+"F"+CStr(bF(c9))+"N"+CStr(bN(
c9))+"O"+CStr(bO(c9))+"P"+CStr(bP(c9))+"S"+CStr(
bS(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((b_mz(c9)-(z-1)*H)/z))/((
b_mz(c9)-(z-1)*H)/z)*1000000),4))&Chr(13)
percent(d1)=1
b_location(d1)=1
else
end if
if oligofragments(c10)<CDbl((c_mz(c9)-(z-1)*H)/z)+
CDbl((c_mz(c9)-(z-1)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((c_mz(c9)-(z-1)*H)/z)-CDbl((
c_mz(c9)-(z-1)*H)/z)*ppmrange2 and d1<c12 and CDbl(
oligofragments_zs(c10))=z then
f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
f6.Write "c"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr(
9)+Chr(45)+CStr(z)+Chr(44)+Chr(9)+CStr((c_mz(c9
)-(z-1)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(cC(c9))+"H"
+CStr(cH(c9)-(z-1))+"F"+CStr(cF(c9))+"N"+CStr(cN(
c9))+"O"+CStr(cO(c9))+"P"+CStr(cP(c9))+"S"+CStr(
cS(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((c_mz(c9)-(z-1)*H)/z))/(c_mz
(c9))*1000000),4))&Chr(13)
percent(d1)=1
c_location(d1)=1
else
end if
if oligofragments(c10)<CDbl((d_mz(c9)-(z-1)*H)/z)+
CDbl((d_mz(c9)-(z-1)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((d_mz(c9)-(z-1)*H)/z)-CDbl((
d_mz(c9)-(z-1)*H)/z)*ppmrange2 and d1<c12 and CDbl(
oligofragments_zs(c10))=z then
f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
f6.Write "d"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr(
9)+Chr(45)+CStr(z)+Chr(44)+Chr(9)+CStr((d_mz(c9
)-(z-1)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(dC(c9))+"H"
+CStr(dH(c9)-(z-1))+"F"+CStr(dF(c9))+"N"+CStr(dN(
c9))+"O"+CStr(d_O(c9))+"P"+CStr(dP(c9))+"S"+CStr(
dS(c9))+Chr(44)+Chr(9)+CStr(Round(((
```

```
oligofragments(c10)-((d_mz(c9)-(z-1)*H)/z))/((
d_mz(c9)-(z-1)*H)/z)*1000000),4))&Chr(13)
1541 percent(d1)=1
1542 d_location(d1)=1
1543 else
1544 end if
1545 if oligofragments(c10)<CDbl((w_mz(c9)-(z-1)*H)/z)+
CDbl((w_mz(c9)-(z-1)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((w_mz(c9)-(z-1)*H)/z)-CDbl((
w_mz(c9)-(z-1)*H)/z)*ppmrange2 and d1<=c12 and CDbl(
oligofragments_zs(c10))=z then
1546 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
1547 f6.Write "w"+Chr(44)+Chr(9)+CStr(c9)+Chr(44)+Chr(
9)+Chr(45)+CStr(z)+Chr(44)+Chr(9)+CStr((w_mz(c9
)-(z-1)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(wC(c9))+"H"
+CStr(wH(c9)-(z-1))+"F"+CStr(wF(c9))+"N"+CStr(wN(
c9))+"O"+CStr(wO(c9))+"P"+CStr(wP(c9))+"S"+CStr(
wS(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((w_mz(c9)-(z-1)*H)/z))/((
w_mz(c9)-(z-1)*H)/z)*1000000),4))&Chr(13)
1548 'MsgBox "c12="&c12
1549 'MsgBox "c9="&c9
1550 percent(c12-c9)=1
1551 w_location(c12-c9)=1
1552 else
1553 end if
1554 if oligofragments(c10)<CDbl((x_mz(c9)-(z-1)*H)/z)+
CDbl((x_mz(c9)-(z-1)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((x_mz(c9)-(z-1)*H)/z)-CDbl((
x_mz(c9)-(z-1)*H)/z)*ppmrange2 and d1<=c12 and CDbl(
oligofragments_zs(c10))=z then
1555 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
1556 f6.Write "x"+Chr(44)+Chr(9)+CStr(c9)+Chr(44)+Chr(
9)+Chr(45)+CStr(z)+Chr(44)+Chr(9)+CStr((x_mz(c9
)-(z-1)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(xC(c9))+"H"
+CStr(xH(c9)-(z-1))+"F"+CStr(xF(c9))+"N"+CStr(xN(
c9))+"O"+CStr(xO(c9))+"P"+CStr(xP(c9))+"S"+CStr(
xS(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((x_mz(c9)-(z-1)*H)/z))/((
x_mz(c9)-(z-1)*H)/z)*1000000),4))&Chr(13)
1557 percent(c12-c9)=1
1558 x_location(c12-c9)=1
1559 else
1560 end if
1561 if oligofragments(c10)<CDbl((y_mz(c9)-(z-1)*H)/z)+
CDbl((y_mz(c9)-(z-1)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((y_mz(c9)-(z-1)*H)/z)-CDbl((
y_mz(c9)-(z-1)*H)/z)*ppmrange2 and d1<=c12 and CDbl(
oligofragments_zs(c10))=z then
1562 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
1563 f6.Write "y"+Chr(44)+Chr(9)+CStr(c9)+Chr(44)+Chr(
9)+Chr(45)+CStr(z)+Chr(44)+Chr(9)+CStr((y_mz(c9
)-(z-1)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(yC(c9))+"H"
```

```
+CStr(yH(c9)-(z-1))+"F"+CStr(yF(c9))+"N"+CStr(yN(
c9))+"O"+CStr(yO(c9))+"P"+CStr(yP(c9))+"S"+CStr(
yS(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((y_mz(c9)-(z-1)*H)/z))/((
y_mz(c9)-(z-1)*H)/z)*1000000),4))&Chr(13)
1564 percent(c12-c9)=1
1565 y_location(c12-c9)=1
1566 else
1567 end if
1568 if oligofragments(c10)<CDbl((z_mz(c9)-(z-1)*H)/z)+
CDbl((z_mz(c9)-(z-1)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((z_mz(c9)-(z-1)*H)/z)-CDbl((
z_mz(c9)-(z-1)*H)/z)*ppmrange2 and d1<=c12 and CDbl(
oligofragments_zs(c10))=z then
1569 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
1570 f6.Write "z"+Chr(44)+Chr(9)+CStr(c9)+Chr(44)+Chr(
9)+Chr(45)+CStr(z)+Chr(44)+Chr(9)+CStr((z_mz(c9
)-(z-1)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(zC(c9))+"H"
+CStr(zH(c9)-(z-1))+"F"+CStr(zF(c9))+"N"+CStr(zN(
c9))+"O"+CStr(zO(c9))+"P"+CStr(zP(c9))+"S"+CStr(
zS(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((z_mz(c9)-(z-1)*H)/z))/((
z_mz(c9)-(z-1)*H)/z)*1000000),4))&Chr(13)
1571 percent(c12-c9)=1
1572 z_location(c12-c9)=1
1573 else
1574 end if
1575 Next
1576 'c=c+1
1577 c10=c10+1
1578 Next
1579 Elseif charge_comparison="x" then
1580 c10=0
1581 For Each signal In oligofragments
1582 'For z=1 to max_charge_fragment
1583 if oligofragments(c10)<CDbl((a1_mz(c9)+(1*H)))+CDbl((
a1_mz(c9)+(1*H)))*ppmrange2 and oligofragments(c10)>CDbl
((a1_mz(c9)+(1*H)))-CDbl((a1_mz(c9)+(1*H)))*ppmrange2 and
d1<c12 then
1584 'the second if statmetn below is to check if the
signal in the oligofragments has a signal which is
between 0.97 and 1.03 daltons lower mass than the
signal itself, if false this peak is not a
descernable monooisotopic peak in this workflow
1585 'else if the peak is a descernable monoisotopic peak
it may be assigned
1586 light_logic=true
1587 light_upper_window=oligofragments(c10)-small_dalton
1588 light_lower_window=oligofragments(c10)-big_dalton
1589 for lightest_cluster_check_counter=1 to
oligofragmentcount
1590 light_check=oligofragments(
lightest_cluster_check_counter)
```

```
1591 if light_check<light_upper_window and light_check
>light_lower_window then
1592 light_logic=false
1593 else
1594 End If
1595 Next
1596 if light_logic=true then
1597 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
1598 f6.Write "a"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr(
9)+Chr(45)+CStr(0)+Chr(44)+Chr(9)+CStr((a1_mz(c9
)+(1*H)))+Chr(44)+Chr(9)+"C"+CStr(a1C(c9))+"H"+
CStr(a1H(c9)+1)+"F"+CStr(a1F(c9))+"N"+CStr(a1N(c9
))+"O"+CStr(a1O(c9))+"P"+CStr(a1P(c9))+"S"+CStr(
a1S(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-(a1_mz(c9)+(1*H)))/(a1_mz(c9
)+(1*H))*1000000),4))&Chr(13)
1599 percent(d1)=1
1600 a_location(d1)=1
1601 Else
1602 End if
1603 else
1604 End if
1605 if oligofragments(c10)<CDbl((ab_mz(c9)+(1*H)))+CDbl((
ab_mz(c9)+(1*H)))*ppmrange2 and oligofragments(c10)>CDbl
((ab_mz(c9)+(1*H)))-CDbl((ab_mz(c9)+(1*H)))*ppmrange2 and
d1<c12 then
1606 light_logic=true
1607 light_upper_window=oligofragments(c10)-small_dalton
1608 light_lower_window=oligofragments(c10)-big_dalton
1609 for lightest_cluster_check_counter=1 to
oligofragmentcount
1610 light_check=oligofragments(
lightest_cluster_check_counter)
1611 if light_check<light_upper_window and light_check
>light_lower_window then
1612 light_logic=false
1613 else
1614 End If
1615 Next
1616 if light_logic=true then
1617 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
1618 f6.Write "ab"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr
(9)+Chr(45)+CStr(0)+Chr(44)+Chr(9)+CStr((ab_mz(c9
)+(1*H)))+Chr(44)+Chr(9)+"C"+CStr(abC(c9))+"H"+
CStr(abH(c9)+1)+"F"+CStr(abF(c9))+"N"+CStr(abN(c9
))+"O"+CStr(abO(c9))+"P"+CStr(abP(c9))+"S"+CStr(
ab_S(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-(ab_mz(c9)+(1*H)))/(ab_mz(c9
)+(1*H))*1000000),4))&Chr(13)
1619 percent(d1)=1
1620 ab_location(d1)=1
1621 Else
1622 End if
```

```
1623 else
1624 end if
1625 if oligofragments(c10)<CDbl((b_mz(c9)+(1*H)))+CDbl((b_mz(
c9)+(1*H)))*ppmrange2 and oligofragments(c10)>CDbl((b_mz(
c9)+(1*H)))-CDbl((b_mz(c9)+(1*H)))*ppmrange2 and d1<c12
then
1626 light_logic=true
1627 light_upper_window=oligofragments(c10)-small_dalton
1628 light_lower_window=oligofragments(c10)-big_dalton
1629 for lightest_cluster_check_counter=1 to
oligofragmentcount
1630 light_check=oligofragments(
lightest_cluster_check_counter)
1631 if light_check<light_upper_window and light_check
>light_lower_window then
1632 light_logic=false
1633 else
1634 End If
1635 Next
1636 if light_logic=true then
1637 'annotate here
1638 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
1639 f6.Write "b"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr(
9)+Chr(45)+CStr(0)+Chr(44)+Chr(9)+CStr((b_mz(c9
)+(1*H)))+Chr(44)+Chr(9)+"C"+CStr(bC(c9))+"H"+
CStr(bH(c9)+1)+"F"+CStr(bF(c9))+"N"+CStr(bN(c9))+
"O"+CStr(bO(c9))+"P"+CStr(bP(c9))+"S"+CStr(bS(c9
))+Chr(44)+Chr(9)+CStr(Round(((oligofragments(c10
)-(b_mz(c9)+(1*H))))/(b_mz(c9)+(1*H))*1000000),4
))&Chr(13)
1640 percent(d1)=1
1641 b_location(d1)=1
1642 Else
1643 End If
1644 else
1645 end if
1646 if oligofragments(c10)<CDbl((c_mz(c9)+(1*H)))+CDbl((c_mz(
c9)+(1*H)))*ppmrange2 and oligofragments(c10)>CDbl((c_mz(
c9)+(1*H)))-CDbl((c_mz(c9)+(1*H)))*ppmrange2 and d1<c12
then
1647 light_logic=true
1648 light_upper_window=oligofragments(c10)-small_dalton
1649 light_lower_window=oligofragments(c10)-big_dalton
1650 for lightest_cluster_check_counter=1 to
oligofragmentcount
1651 'MsgBox
"m/z="&CStr(oligofragments(lightest_cluster_check
_counter))&Chr(13)&"light_upper_window="&CStr(lig
ht_upper_window)&Chr(13)&"c_mz="&CStr((c_mz(c9)+(
1*H)))&Chr(13)&"light_lower_window="&CStr(light_l
ower_window)
1652 light_check=oligofragments(
lightest_cluster_check_counter)
```

```
1653 if light_check<light_upper_window and light_check
>light_lower_window then
1654 light_logic=false
1655 else
1656 End if
1657 Next
1658 if light_logic=true then
1659 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
1660 f6.Write "c"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr(
9)+Chr(45)+CStr(0)+Chr(44)+Chr(9)+CStr((c_mz(c9
)+(1*H)))+Chr(44)+Chr(9)+"C"+CStr(cC(c9))+"H"+
CStr(cH(c9)+1)+"F"+CStr(cF(c9))+"N"+CStr(cN(c9))+
"O"+CStr(cO(c9))+"P"+CStr(cP(c9))+"S"+CStr(cS(c9
))+Chr(44)+Chr(9)+CStr(Round(((oligofragments(c10
)-(c_mz(c9)+(1*H))))/(c_mz(c9)+(1*H))*1000000),4
))&Chr(13)
1661 percent(d1)=1
1662 c_location(d1)=1
1663 Else
1664 End if
1665 else
1666 end if
1667 if oligofragments(c10)<CDbl((d_mz(c9)+(1*H)))+CDbl((d_mz(
c9)+(1*H)))*ppmrange2 and oligofragments(c10)>CDbl((d_mz(
c9)+(1*H)))-CDbl((d_mz(c9)+(1*H)))*ppmrange2 and d1<c12
then
1668 light_logic=true
1669 light_upper_window=oligofragments(c10)-small_dalton
1670 light_lower_window=oligofragments(c10)-big_dalton
1671 for lightest_cluster_check_counter=1 to
oligofragmentcount
1672 light_check=oligofragments(
lightest_cluster_check_counter)
1673 if light_check<light_upper_window and light_check
>light_lower_window then
1674 light_logic=false
1675 else
1676 End If
1677 Next
1678 if light_logic=true then
1679 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
1680 f6.Write "d"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr(
9)+Chr(45)+CStr(0)+Chr(44)+Chr(9)+CStr((d_mz(c9
)+(1*H)))+Chr(44)+Chr(9)+"C"+CStr(dC(c9))+"H"+
CStr(dH(c9)+1)+"F"+CStr(dF(c9))+"N"+CStr(dN(c9))+
"O"+CStr(d_O(c9))+"P"+CStr(dP(c9))+"S"+CStr(dS(c9
))+Chr(44)+Chr(9)+CStr(Round(((oligofragments(c10
)-(d_mz(c9)+(1*H))))/(d_mz(c9)+(1*H))*1000000),4
))&Chr(13)
1681 percent(d1)=1
1682 d_location(d1)=1
1683 Else
1684 end If
```

```
1685 else
1686 end if
1687 if oligofragments(c10)<CDbl((w_mz(c9)+(1*H)))+CDbl((w_mz(
c9)+(1*H)))*ppmrange2 and oligofragments(c10)>CDbl((w_mz(
c9)+(1*H)))-CDbl((w_mz(c9)+(1*H)))*ppmrange2 and d1<=c12
then
1688 light_logic=true
1689 light_upper_window=oligofragments(c10)-small_dalton
1690 light_lower_window=oligofragments(c10)-big_dalton
1691 for lightest_cluster_check_counter=1 to
oligofragmentcount
1692 light_check=oligofragments(
lightest_cluster_check_counter)
1693 if light_check<light_upper_window and light_check
>light_lower_window then
1694 light_logic=false
1695 else
1696 End If
1697 Next
1698 if light_logic=true then
1699 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
1700 f6.Write "w"+Chr(44)+Chr(9)+CStr(c9)+Chr(44)+Chr(
9)+Chr(45)+CStr(0)+Chr(44)+Chr(9)+CStr((w_mz(c9
)+(1*H)))+Chr(44)+Chr(9)+"C"+CStr(wC(c9))+"H"+
CStr(wH(c9)+1)+"F"+CStr(wF(c9))+"N"+CStr(wN(c9))+
"O"+CStr(wO(c9))+"P"+CStr(wP(c9))+"S"+CStr(wS(c9
))+Chr(44)+Chr(9)+CStr(Round(((oligofragments(c10
)-(w_mz(c9)+(1*H)))/(w_mz(c9)+(1*H))*1000000),4
))&Chr(13)
1701 percent(c12-c9)=1
1702 w_location(c12-c9)=1
1703 Else
1704 End if
1705 else
1706 end if
1707 if oligofragments(c10)<CDbl((x_mz(c9)+(1*H)))+CDbl((x_mz(
c9)+(1*H)))*ppmrange2 and oligofragments(c10)>CDbl((x_mz(
c9)+(1*H)))-CDbl((x_mz(c9)+(1*H)))*ppmrange2 and d1<=c12
then
1708 light_logic=true
1709 light_upper_window=oligofragments(c10)-small_dalton
1710 light_lower_window=oligofragments(c10)-big_dalton
1711 for lightest_cluster_check_counter=1 to
oligofragmentcount
1712 light_check=oligofragments(
lightest_cluster_check_counter)
1713 if light_check<light_upper_window and light_check
>light_lower_window then
1714 light_logic=false
1715 else
1716 End If
1717 Next
1718 if light_logic=true then
```

```
1719 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
1720 f6.Write "x"+Chr(44)+Chr(9)+CStr(c9)+Chr(44)+Chr(
9)+Chr(45)+CStr(0)+Chr(44)+Chr(9)+CStr((x_mz(c9
)+(1*H)))+Chr(44)+Chr(9)+"C"+CStr(xC(c9))+"H"+
CStr(xH(c9)+1)+"F"+CStr(xF(c9))+"N"+CStr(xN(c9))+
"O"+CStr(xO(c9))+"P"+CStr(xP(c9))+"S"+CStr(xS(c9
))+Chr(44)+Chr(9)+CStr(Round(((oligofragments(c10
)-(x_mz(c9)+(1*H)))/(x_mz(c9)+(1*H))*1000000),4
))&Chr(13)
1721 percent(c12-c9)=1
1722 x_location(c12-c9)=1
1723 Else
1724 end if
1725 else
1726 end if
1727 if oligofragments(c10)<CDbl((y_mz(c9)+(1*H)))+CDbl((y_mz(
c9)+(1*H)))*ppmrange2 and oligofragments(c10)>CDbl((y_mz(
c9)+(1*H)))-CDbl((y_mz(c9)+(1*H)))*ppmrange2 and d1<=c12
then
1728 light_logic=true
1729 light_upper_window=oligofragments(c10)-small_dalton
1730 light_lower_window=oligofragments(c10)-big_dalton
1731 for lightest_cluster_check_counter=1 to
oligofragmentcount
1732 MsgBox "m/z="&CStr(oligofragments(
lightest_cluster_check_counter))&Chr(13)&
"light_upper_window="&CStr(light_upper_window)&Chr(13
)&"y_mz="&CStr((y_mz(c9)+(1*H)))&Chr(13)&
"light_lower_window="&CStr(light_lower_window)
1733 light_check=oligofragments(
lightest_cluster_check_counter)
1734 if light_check<light_upper_window and light_check
>light_lower_window then
1735 light_logic=false
1736 else
1737 End If
1738 Next
1739 if light_logic=true then
1740 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
1741 f6.Write "y"+Chr(44)+Chr(9)+CStr(c9)+Chr(44)+Chr(
9)+Chr(45)+CStr(0)+Chr(44)+Chr(9)+CStr((y_mz(c9
)+(1*H)))+Chr(44)+Chr(9)+"C"+CStr(yC(c9))+"H"+
CStr(yH(c9)+1)+"F"+CStr(yF(c9))+"N"+CStr(yN(c9))+
"O"+CStr(yO(c9))+"P"+CStr(yP(c9))+"S"+CStr(yS(c9
))+Chr(44)+Chr(9)+CStr(Round(((oligofragments(c10
)-(y_mz(c9)+(1*H)))/(y_mz(c9)+(1*H))*1000000),4
))&Chr(13)
1742 percent(c12-c9)=1
1743 y_location(c12-c9)=1
1744 Else
1745 End if
1746 else
1747 end if
```

```
1748 if oligofragments(c10)<CDbl((z_mz(c9)+(1*H)))+CDbl((z_mz(
c9)+(1*H)))*ppmrange2 and oligofragments(c10)>CDbl((z_mz(
c9)+(1*H)))-CDbl((z_mz(c9)+(1*H)))*ppmrange2 and d1<=c12
then
1749 light_logic=true
1750 light_upper_window=oligofragments(c10)-small_dalton
1751 light_lower_window=oligofragments(c10)-big_dalton
1752 for lightest_cluster_check_counter=1 to
oligofragmentcount
1753 light_check=oligofragments(
lightest_cluster_check_counter)
1754 if light_check<light_upper_window and light_check
>light_lower_window then
1755 light_logic=false
1756 else
1757 End If
1758 Next
1759 if light_logic=true then
1760 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
1761 f6.Write "z"+Chr(44)+Chr(9)+CStr(c9)+Chr(44)+Chr(
9)+Chr(45)+CStr(0)+Chr(44)+Chr(9)+CStr((z_mz(c9
)+(1*H)))+Chr(44)+Chr(9)+"C"+CStr(zC(c9))+"H"+
CStr(zH(c9)+1)+"F"+CStr(zF(c9))+"N"+CStr(zN(c9))+
"O"+CStr(zO(c9))+"P"+CStr(zP(c9))+"S"+CStr(zS(c9
))+Chr(44)+Chr(9)+CStr(Round(((oligofragments(c10
)-(z_mz(c9)+(1*H)))/(z_mz(c9)+(1*H))*1000000),4
))&Chr(13)
1762 percent(c12-c9)=1
1763 z_location(c12-c9)=1
1764 else
1765 End If
1766 else
1767 end if
1768 'Next
1769 'c=c+1
1770 c10=c10+1
1771 Next
1772 Else
1773 c10=0
1774 For Each signal In oligofragments
1775 For z=1 to max_charge_fragment
1776 if oligofragments(c10)<CDbl((a1_mz(c9)-(z-1)*H)/z)+
CDbl((a1_mz(c9)-(z-1)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((a1_mz(c9)-(z-1)*H)/z)-CDbl
((a1_mz(c9)-(z-1)*H)/z)*ppmrange2 and d1<c12 then
1777 'annotate here
1778 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
1779 f6.Write "a"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr(
9)+Chr(45)+CStr(z)+Chr(44)+Chr(9)+CStr((a1_mz(c9
)-(z-1)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(a1C(c9))+
"H"+CStr(a1H(c9)-(z-1))+"F"+CStr(a1F(c9))+"N"+
CStr(a1N(c9))+"O"+CStr(a1O(c9))+"P"+CStr(a1P(c9
))+"S"+CStr(a1S(c9))+Chr(44)+Chr(9)+CStr(Round(((
```

```
oligofragments(c10)-((a1_mz(c9)-(z-1)*H)/z))/((
a1_mz(c9)-(z-1)*H)/z)*1000000),4))&Chr(13)
1780 percent(d1)=1
1781 a_location(d1)=1
1782 else
1783 End if
1784 if oligofragments(c10)<CDbl((ab_mz(c9)-(z-1)*H)/z)+
CDbl((ab_mz(c9)-(z-1)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((ab_mz(c9)-(z-1)*H)/z)-CDbl
((ab_mz(c9)-(z-1)*H)/z)*ppmrange2 and d1<c12 then
1785 'Cmpd(2).Annotations.AddAnnotation ab_mz_L,
prime_int*1.2, ab_mz_H, prime_int*1.2,
prime_int*1.6, "a-b"+CStr(d), True
1786 'annotate here
1787 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
1788 f6.Write "ab"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr
(9)+Chr(45)+CStr(z)+Chr(44)+Chr(9)+CStr((ab_mz(c9
)-(z-1)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(abC(c9))+
"H"+CStr(abH(c9)-(z-1))+"F"+CStr(abF(c9))+"N"+
CStr(abN(c9))+"O"+CStr(abO(c9))+"P"+CStr(abP(c9
))+"S"+CStr(ab_S(c9))+Chr(44)+Chr(9)+CStr(Round
(((oligofragments(c10)-((ab_mz(c9)-(z-1)*H)/z
))/((ab_mz(c9)-(z-1)*H)/z)*1000000),4))&Chr(13)
1789 percent(d1)=1
1790 ab_location(d1)=1
1791 else
1792 end if
1793 if oligofragments(c10)<CDbl((b_mz(c9)-(z-1)*H)/z)+
CDbl((b_mz(c9)-(z-1)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((b_mz(c9)-(z-1)*H)/z)-CDbl((
b_mz(c9)-(z-1)*H)/z)*ppmrange2 and d1<c12 then
1794 'annotate here
1795 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
1796 f6.Write "b"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr(
9)+Chr(45)+CStr(z)+Chr(44)+Chr(9)+CStr((b_mz(c9
)-(z-1)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(bC(c9))+"H"
+CStr(bH(c9)-(z-1))+"F"+CStr(bF(c9))+"N"+CStr(bN(
c9))+"O"+CStr(bO(c9))+"P"+CStr(bP(c9))+"S"+CStr(
bS(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((b_mz(c9)-(z-1)*H)/z))/((
b_mz(c9)-(z-1)*H)/z)*1000000),4))&Chr(13)
1797 percent(d1)=1
1798 b_location(d1)=1
1799 else
1800 end if
1801 if oligofragments(c10)<CDbl((c_mz(c9)-(z-1)*H)/z)+
CDbl((c_mz(c9)-(z-1)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((c_mz(c9)-(z-1)*H)/z)-CDbl((
c_mz(c9)-(z-1)*H)/z)*ppmrange2 and d1<c12 then
1802 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
1803 f6.Write "c"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr(
9)+Chr(45)+CStr(z)+Chr(44)+Chr(9)+CStr((c_mz(c9
)-(z-1)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(cC(c9))+"H"
```

```
+CStr(cH(c9)-(z-1))+"F"+CStr(cF(c9))+"N"+CStr(cN(
c9))+"O"+CStr(cO(c9))+"P"+CStr(cP(c9))+"S"+CStr(
cS(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((c_mz(c9)-(z-1)*H)/z))/(c_mz
(c9))*1000000),4))&Chr(13)
1804 percent(d1)=1
1805 c_location(d1)=1
1806 else
1807 end if
1808 if oligofragments(c10)<CDbl((d_mz(c9)-(z-1)*H)/z)+
CDbl((d_mz(c9)-(z-1)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((d_mz(c9)-(z-1)*H)/z)-CDbl((
d_mz(c9)-(z-1)*H)/z)*ppmrange2 and d1<c12 then
1809 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
1810 f6.Write "d"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr(
9)+Chr(45)+CStr(z)+Chr(44)+Chr(9)+CStr((d_mz(c9
)-(z-1)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(dC(c9))+"H"
+CStr(dH(c9)-(z-1))+"F"+CStr(dF(c9))+"N"+CStr(dN(
c9))+"O"+CStr(d_O(c9))+"P"+CStr(dP(c9))+"S"+CStr(
dS(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((d_mz(c9)-(z-1)*H)/z))/((
d_mz(c9)-(z-1)*H)/z)*1000000),4))&Chr(13)
1811 percent(d1)=1
1812 d_location(d1)=1
1813 else
1814 end if
1815 if oligofragments(c10)<CDbl((w_mz(c9)-(z-1)*H)/z)+
CDbl((w_mz(c9)-(z-1)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((w_mz(c9)-(z-1)*H)/z)-CDbl((
w_mz(c9)-(z-1)*H)/z)*ppmrange2 and d1<=c12 then
1816 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
1817 f6.Write "w"+Chr(44)+Chr(9)+CStr(c9)+Chr(44)+Chr(
9)+Chr(45)+CStr(z)+Chr(44)+Chr(9)+CStr((w_mz(c9
)-(z-1)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(wC(c9))+"H"
+CStr(wH(c9)-(z-1))+"F"+CStr(wF(c9))+"N"+CStr(wN(
c9))+"O"+CStr(wO(c9))+"P"+CStr(wP(c9))+"S"+CStr(
wS(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((w_mz(c9)-(z-1)*H)/z))/((
w_mz(c9)-(z-1)*H)/z)*1000000),4))&Chr(13)
1818 'MsgBox "c12="&c12
1819 'MsgBox "c9="&c9
1820 percent(c12-c9)=1
1821 w_location(c12-c9)=1
1822 else
1823 end if
1824 if oligofragments(c10)<CDbl((x_mz(c9)-(z-1)*H)/z)+
CDbl((x_mz(c9)-(z-1)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((x_mz(c9)-(z-1)*H)/z)-CDbl((
x_mz(c9)-(z-1)*H)/z)*ppmrange2 and d1<=c12 then
1825 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
1826 f6.Write "x"+Chr(44)+Chr(9)+CStr(c9)+Chr(44)+Chr(
9)+Chr(45)+CStr(z)+Chr(44)+Chr(9)+CStr((x_mz(c9
)-(z-1)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(xC(c9))+"H"
```

```
+CStr(xH(c9)-(z-1))+"F"+CStr(xF(c9))+"N"+CStr(xN(
c9))+"O"+CStr(xO(c9))+"P"+CStr(xP(c9))+"S"+CStr(
xS(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((x_mz(c9)-(z-1)*H)/z))/((
x_mz(c9)-(z-1)*H)/z)*1000000),4))&Chr(13)
1827 percent(c12-c9)=1
1828 x_location(c12-c9)=1
1829 else
1830 end if
1831 if oligofragments(c10)<CDbl((y_mz(c9)-(z-1)*H)/z)+
CDbl((y_mz(c9)-(z-1)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((y_mz(c9)-(z-1)*H)/z)-CDbl((
y_mz(c9)-(z-1)*H)/z)*ppmrange2 and d1<=c12 then
1832 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
1833 f6.Write "y"+Chr(44)+Chr(9)+CStr(c9)+Chr(44)+Chr(
9)+Chr(45)+CStr(z)+Chr(44)+Chr(9)+CStr((y_mz(c9
)-(z-1)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(yC(c9))+"H"
+CStr(yH(c9)-(z-1))+"F"+CStr(yF(c9))+"N"+CStr(yN(
c9))+"O"+CStr(yO(c9))+"P"+CStr(yP(c9))+"S"+CStr(
yS(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((y_mz(c9)-(z-1)*H)/z))/((
y_mz(c9)-(z-1)*H)/z)*1000000),4))&Chr(13)
1834 percent(c12-c9)=1
1835 y_location(c12-c9)=1
1836 else
1837 end if
1838 if oligofragments(c10)<CDbl((z_mz(c9)-(z-1)*H)/z)+
CDbl((z_mz(c9)-(z-1)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((z_mz(c9)-(z-1)*H)/z)-CDbl((
z_mz(c9)-(z-1)*H)/z)*ppmrange2 and d1<=c12 then
1839 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
1840 f6.Write "z"+Chr(44)+Chr(9)+CStr(c9)+Chr(44)+Chr(
9)+Chr(45)+CStr(z)+Chr(44)+Chr(9)+CStr((z_mz(c9
)-(z-1)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(zC(c9))+"H"
+CStr(zH(c9)-(z-1))+"F"+CStr(zF(c9))+"N"+CStr(zN(
c9))+"O"+CStr(zO(c9))+"P"+CStr(zP(c9))+"S"+CStr(
zS(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((z_mz(c9)-(z-1)*H)/z))/((
z_mz(c9)-(z-1)*H)/z)*1000000),4))&Chr(13)
1841 percent(c12-c9)=1
1842 z_location(c12-c9)=1
1843 else
1844 end if
1845 Next
1846 'c=c+1
1847 c10=c10+1
1848 Next
1849 End if
1850 c9=c9+1
1851 Next
1852 c11=0
1853 sum=0
1854 'coverage=0
```

```
For Each location in percent
sum=sum+percent(c11)
c11=c11+1
Next
'sequence coverage table writing
'
''
f6.Write "ab1"+Chr(9)+"-"
c13=1
For location=1 to (c11-1)
f6.Write CStr(ab_location(c13))+Chr(44)+Chr(32)
c13=c13+1
Next
f6.Write Chr(13)
''
''
f6.Write "a1"+Chr(9)+"-"
c13=1
For location=1 to (c11-1)
f6.Write CStr(a_location(c13))+Chr(44)+Chr(32)
c13=c13+1
Next
f6.Write Chr(13)
''
''
f6.Write "b1"+Chr(9)+"-"
c13=1
For location=1 to (c11-1)
f6.Write CStr(b_location(c13))+Chr(44)+Chr(32)
c13=c13+1
Next
f6.Write Chr(13)
''
''
f6.Write "c1"+Chr(9)+"-"
c13=1
For location=1 to (c11-1)
f6.Write CStr(c_location(c13))+Chr(44)+Chr(32)
c13=c13+1
Next
f6.Write Chr(13)
''
''
f6.Write "d1"+Chr(9)+"-"
c13=1
For location=1 to (c11-1)
f6.Write CStr(d_location(c13))+Chr(44)+Chr(32)
c13=c13+1
Next
f6.Write Chr(13)
''
'''
''
```

```
f6.Write "w"+CStr(c11-1)+Chr(9)+"-"
c13=1
For location=1 to (c11-1)
f6.Write CStr(w_location(c13))+Chr(44)+Chr(32)
c13=c13+1
Next
f6.Write Chr(13)
''
''
f6.Write "x"+CStr(c11-1)+Chr(9)+"-"
c13=1
For location=1 to (c11-1)
f6.Write CStr(x_location(c13))+Chr(44)+Chr(32)
c13=c13+1
Next
f6.Write Chr(13)
''
''
f6.Write "y"+CStr(c11-1)+Chr(9)+"-"
c13=1
For location=1 to (c11-1)
f6.Write CStr(y_location(c13))+Chr(44)+Chr(32)
c13=c13+1
Next
f6.Write Chr(13)
''
''
f6.Write "z"+CStr(c11-1)+Chr(9)+"-"
c13=1
For location=1 to (c11-1)
f6.Write CStr(z_location(c13))+Chr(44)+Chr(32)
c13=c13+1
Next
f6.Write Chr(13)
''
'
f6.Write "percent coverage(%)="+CStr(sum/(c11-1)*100)&Chr(13)
else
'f6.Write oligoprecursor&Chr(13)
End If
c11=0
For Each location in percent
percent(c11)=0
c11=c11+1
Next
'redefine boolean array for semi-visual sequence mapping
'fill the boolean array with zeros by default
c14=0
For Each location in ab_location
ab_location(c14)=0
a_location(c14)=0
b_location(c14)=0
c_location(c14)=0
```

```
1961 d_location(c14)=0
1962 w_location(c14)=0
1963 x_location(c14)=0
1964 y_location(c14)=0
1965 z_location(c14)=0
1966 c14=c14+1
1967 Next
1968 c14=0
1969 'MsgBox negchargestates(c5)
1970 'MsgBox oligoprecursor oligoprecursor
1971 c5=c5+1
1972 Next
1973 c5=0
1974 For precursor=0 to max_charge_index
1975 oligoprecursor_H=CDbl(poschargestates(c5))+CDbl(ppmrange2_precursor)
1976 oligoprecursor_L=CDbl(poschargestates(c5))-CDbl(ppmrange2_precursor)
1977 if oligoprecursor<oligoprecursor_H and oligoprecursor>oligoprecursor_L
then
1978 f6.Write CStr(tempfile(c8))&Chr(13)
1979 f6.Write "pos"+Chr(44)+Chr(9)+"exp="+CStr(oligoprecursor)+Chr(44)+Chr
(9)+"theo="+CStr(poschargestates(c5))+Chr(44)+Chr(9)+oligoname(ccc-1
)+Chr(44)+Chr(9)+"charge="+CStr(precursor+1)&Chr(13)
1980 c9=0
1981 d1=0
1982 c12=0
1983 For Each location in percent
1984 c12=c12+1
1985 Next
1986 For Each sugar In o_array1
1987 d1=c9+1
1988 'a1_mz_n(c)=CDbl(a1_mz(c))
1989 'a1_mz_L
=CDbl(a1_mz(c9)+2*H)-CDbl(a1_mz(c9)+2*H)*ppmrange2
1990 'a1_mz_H
=CDbl(a1_mz(c9)+2*H)+CDbl(a1_mz(c9)+2*H)*ppmrange2
1991 'ab_mz_L
=CDbl(ab_mz(c9)+2*H)-CDbl(ab_mz(c9)+2*H)*ppmrange2
1992 'ab_mz_H
=CDbl(ab_mz(c9)+2*H)+CDbl(ab_mz(c9)+2*H)*ppmrange2
1993 'b_mz_L
=CDbl(b_mz(c9)+2*H)-CDbl(b_mz(c9)+2*H)*ppmrange2
1994 'b_mz_H
=CDbl(b_mz(c9)+2*H)+CDbl(b_mz(c9)+2*H)*ppmrange2
1995 'c_mz_L
=CDbl(c_mz(c9)+2*H)-CDbl(c_mz(c9)+2*H)*ppmrange2
1996 'c_mz_H
=CDbl(c_mz(c9)+2*H)+CDbl(c_mz(c9)+2*H)*ppmrange2
1997 'd_mz_L
=CDbl(d_mz(c9)+2*H)-CDbl(d_mz(c9)+2*H)*ppmrange2
1998 'd_mz_H
=CDbl(d_mz(c9)+2*H)+CDbl(d_mz(c9)+2*H)*ppmrange2
1999 'w_mz_L
=CDbl(w_mz(c9)+2*H)-CDbl(w_mz(c9)+2*H)*ppmrange2
```

```
2000 'w_mz_H
=CDbl(w_mz(c9)+2*H)+CDbl(w_mz(c9)+2*H)*ppmrange2
2001 'x_mz_L
=CDbl(x_mz(c9)+2*H)-CDbl(x_mz(c9)+2*H)*ppmrange2
2002 'x_mz_H
=CDbl(x_mz(c9)+2*H)+CDbl(x_mz(c9)+2*H)*ppmrange2
2003 'y_mz_L
=CDbl(y_mz(c9)+2*H)-CDbl(y_mz(c9)+2*H)*ppmrange2
2004 'y_mz_H
=CDbl(y_mz(c9)+2*H)+CDbl(y_mz(c9)+2*H)*ppmrange2
2005 'z_mz_L
=CDbl(z_mz(c9)+2*H)-CDbl(z_mz(c9)+2*H)*ppmrange2
2006 'z_mz_H
=CDbl(z_mz(c9)+2*H)+CDbl(z_mz(c9)+2*H)*ppmrange2
2007 'MsgBox a1_mz_n_L
2008 'this is probably the place to put annother for loop to itterate
through the fragmtn ion signals
2009 if charge_comparison="y" then
2010 c10=0
2011 For Each signal In oligofragments
2012
2013 For z=1 to max_charge_fragment
2014 if oligofragments(c10)<CDbl((a1_mz(c9)+(1+z)*H)/z)+
CDbl((a1_mz(c9)+(1+z)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((a1_mz(c9)+(1+z)*H)/z)-CDbl
((a1_mz(c9)+(1+z)*H)/z)*ppmrange2 and d1<c12 and CDbl
(oligofragments_zs(c10))=z then
2015 'annotate here
2016 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
2017 f6.Write "a"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr(
9)+Chr(43)+CStr(z)+Chr(44)+Chr(9)+CStr((a1_mz(c9
)+(1+z)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(a1C(c9))+
"H"+CStr(a1H(c9)+(z+1))+"F"+CStr(a1F(c9))+"N"+
CStr(a1N(c9))+"O"+CStr(a1O(c9))+"P"+CStr(a1P(c9
))+"S"+CStr(a1S(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((a1_mz(c9)+(1+z)*H)/z))/((
a1_mz(c9)+(1+z)*H)/z)*1000000),4))&Chr(13)
2018 percent(d1)=1
2019 a_location(d1)=1
2020 else
2021 End if
2022 if oligofragments(c10)<CDbl((ab_mz(c9)+(1+z)*H)/z)+
CDbl((ab_mz(c9)+(1+z)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((ab_mz(c9)+(1+z)*H)/z)-CDbl
((ab_mz(c9)+(1+z)*H)/z)*ppmrange2 and d1<c12 and CDbl
(oligofragments_zs(c10))=z then
2023 'Cmpd(2).Annotations.AddAnnotation ab_mz_L,
prime_int*1.2, ab_mz_H, prime_int*1.2,
prime_int*1.6, "a-b"+CStr(d), True
2024 'annotate here
2025 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
2026 f6.Write "ab"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr
(9)+Chr(43)+CStr(z)+Chr(44)+Chr(9)+CStr((ab_mz(c9
```

```
)+(1+z)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(abC(c9))+
"H"+CStr(abH(c9)+(z+1))+"F"+CStr(abF(c9))+"N"+
CStr(abN(c9))+"O"+CStr(abO(c9))+"P"+CStr(abP(c9
))+"S"+CStr(ab_S(c9))+Chr(44)+Chr(9)+CStr(Round
(((oligofragments(c10)-((ab_mz(c9)+(1+z)*H)/z
))/((ab_mz(c9)+(1+z)*H)/z)*1000000),4))&Chr(13)
2027 percent(d1)=1
2028 ab_location(d1)=1
2029 else
2030 end if
2031 if oligofragments(c10)<CDbl((b_mz(c9)+(1+z)*H)/z)+
CDbl((b_mz(c9)+(1+z)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((b_mz(c9)+(1+z)*H)/z)-CDbl((
b_mz(c9)+(1+z)*H)/z)*ppmrange2 and d1<c12 and CDbl(
oligofragments_zs(c10))=z then
2032 'annotate here
2033 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
2034 f6.Write "b"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr(
9)+Chr(43)+CStr(z)+Chr(44)+Chr(9)+CStr((b_mz(c9
)+(1+z)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(bC(c9))+"H"
+CStr(bH(c9)+(z+1))+"F"+CStr(bF(c9))+"N"+CStr(bN(
c9))+"O"+CStr(bO(c9))+"P"+CStr(bP(c9))+"S"+CStr(
bS(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((b_mz(c9)+(1+z)*H)/z))/((
b_mz(c9)+(1+z)*H)/z)*1000000),4))&Chr(13)
2035 percent(d1)=1
2036 b_location(d1)=1
2037 else
2038 end if
2039 if oligofragments(c10)<CDbl((c_mz(c9)+(1+z)*H)/z)+
CDbl((c_mz(c9)+(1+z)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((c_mz(c9)+(1+z)*H)/z)-CDbl((
c_mz(c9)+(1+z)*H)/z)*ppmrange2 and d1<c12 and CDbl(
oligofragments_zs(c10))=z then
2040 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
2041 f6.Write "c"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr(
9)+Chr(43)+CStr(z)+Chr(44)+Chr(9)+CStr((c_mz(c9
)+(1+z)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(cC(c9))+"H"
+CStr(cH(c9)+(z+1))+"F"+CStr(cF(c9))+"N"+CStr(cN(
c9))+"O"+CStr(cO(c9))+"P"+CStr(cP(c9))+"S"+CStr(
cS(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((c_mz(c9)+(1+z)*H)/z))/((
c_mz(c9)+(1+z)*H)/z)*1000000),4))&Chr(13)
2042 percent(d1)=1
2043 c_location(d1)=1
2044 else
2045 end if
2046 if oligofragments(c10)<CDbl((d_mz(c9)+(1+z)*H)/z)+
CDbl((d_mz(c9)+(1+z)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((d_mz(c9)+(1+z)*H)/z)-CDbl((
d_mz(c9)+(1+z)*H)/z)*ppmrange2 and d1<c12 and CDbl(
oligofragments_zs(c10))=z then
2047 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
```

```
2048 f6.Write "d"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr(
9)+Chr(43)+CStr(z)+Chr(44)+Chr(9)+CStr((d_mz(c9
)+(1+z)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(dC(c9))+"H"
+CStr(dH(c9)+2)+"F"+CStr(dF(c9))+"N"+CStr(dN(c9
))+"O"+CStr(d_O(c9))+"P"+CStr(dP(c9))+"S"+CStr(dS
(c9))+Chr(44)+Chr(9)+CStr(Round(((oligofragments(
c10)-((d_mz(c9)+(1+z)*H)/z))/((d_mz(c9)+(1+z)*H)/
z)*1000000),4))&Chr(13)
2049 percent(d1)=1
2050 d_location(d1)=1
2051 else
2052 end if
2053 if oligofragments(c10)<CDbl((w_mz(c9)+(1+z)*H)/z)+
CDbl((w_mz(c9)+(1+z)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((w_mz(c9)+(1+z)*H)/z)-CDbl((
w_mz(c9)+(1+z)*H)/z)*ppmrange2 and d1<=c12 and CDbl(
oligofragments_zs(c10))=z then
2054 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
2055 f6.Write "w"+Chr(44)+Chr(9)+CStr(c9)+Chr(44)+Chr(
9)+Chr(43)+CStr(z)+Chr(44)+Chr(9)+CStr((w_mz(c9
)+(1+z)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(wC(c9))+"H"
+CStr(wH(c9)+(z+1))+"F"+CStr(wF(c9))+"N"+CStr(wN(
c9))+"O"+CStr(wO(c9))+"P"+CStr(wP(c9))+"S"+CStr(
wS(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((w_mz(c9)+(1+z)*H)/z))/((
w_mz(c9)+(1+z)*H)/z)*1000000),4))&Chr(13)
2056 'MsgBox "c12="&c12
2057 'MsgBox "c9="&c9
2058 percent(c12-c9)=1
2059 w_location(c12-c9)=1
2060 else
2061 end if
2062 if oligofragments(c10)<CDbl((x_mz(c9)+(1+z)*H)/z)+
CDbl((x_mz(c9)+(1+z)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((x_mz(c9)+(1+z)*H)/z)-CDbl((
x_mz(c9)+(1+z)*H)/z)*ppmrange2 and d1<=c12 and CDbl(
oligofragments_zs(c10))=z then
2063 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
2064 f6.Write "x"+Chr(44)+Chr(9)+CStr(c9)+Chr(44)+Chr(
9)+Chr(43)+CStr(z)+Chr(44)+Chr(9)+CStr((x_mz(c9
)+(1+z)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(xC(c9))+"H"
+CStr(xH(c9)+(z+1))+"F"+CStr(xF(c9))+"N"+CStr(xN(
c9))+"O"+CStr(xO(c9))+"P"+CStr(xP(c9))+"S"+CStr(
xS(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((x_mz(c9)+(1+z)*H)/z))/((
x_mz(c9)+(1+z)*H)/z)*1000000),4))&Chr(13)
2065 percent(c12-c9)=1
2066 x_location(c12-c9)=1
2067 else
2068 end if
2069 if oligofragments(c10)<CDbl((y_mz(c9)+(1+z)*H)/z)+
CDbl((y_mz(c9)+(1+z)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((y_mz(c9)+(1+z)*H)/z)-CDbl((
```

```
y_mz(c9)+(1+z)*H)/z)*ppmrange2 and d1<=c12 and CDbl(
oligofragments_zs(c10))=z then
2070 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
2071 f6.Write "y"+Chr(44)+Chr(9)+CStr(c9)+Chr(44)+Chr(
9)+Chr(43)+CStr(z)+Chr(44)+Chr(9)+CStr((y_mz(c9
)+(1+z)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(yC(c9))+"H"
+CStr(yH(c9)+(z+1))+"F"+CStr(yF(c9))+"N"+CStr(yN(
c9))+"O"+CStr(yO(c9))+"P"+CStr(yP(c9))+"S"+CStr(
yS(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((y_mz(c9)+(1+z)*H)/z))/((
y_mz(c9)+(1+z)*H)/z)*1000000),4))&Chr(13)
2072 percent(c12-c9)=1
2073 y_location(c12-c9)=1
2074 else
2075 end if
2076 if oligofragments(c10)<CDbl((z_mz(c9)+(1+z)*H)/z)+
CDbl((z_mz(c9)+(1+z)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((z_mz(c9)+(1+z)*H)/z)-CDbl((
z_mz(c9)+(1+z)*H)/z)*ppmrange2 and d1<=c12 and CDbl(
oligofragments_zs(c10))=z then
2077 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
2078 f6.Write "z"+Chr(44)+Chr(9)+CStr(c9)+Chr(44)+Chr(
9)+Chr(43)+CStr(z)+Chr(44)+Chr(9)+CStr((z_mz(c9
)+(1+z)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(zC(c9))+"H"
+CStr(zH(c9)+(z+1))+"F"+CStr(zF(c9))+"N"+CStr(zN(
c9))+"O"+CStr(zO(c9))+"P"+CStr(zP(c9))+"S"+CStr(
zS(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((z_mz(c9)+(1+z)*H)/z))/((
z_mz(c9)+(1+z)*H)/z)*1000000),4))&Chr(13)
2079 percent(c12-c9)=1
2080 z_location(c12-c9)=1
2081 else
2082 end if
2083 Next
2084 'c=c+1
2085 c10=c10+1
2086 Next
2087 Elseif charge_comparison="x" then
2088 c10=0
2089 For Each signal In oligofragments
2090 'For z=1 to max_charge_fragment
2091 if oligofragments(c10)<CDbl((a1_mz(c9)+(1*H)))+CDbl((
a1_mz(c9)+(1*H)))*ppmrange2 and oligofragments(c10)>CDbl
((a1_mz(c9)+(1*H)))-CDbl((a1_mz(c9)+(1*H)))*ppmrange2 and
d1<c12 then
2092 'the second if statmetn below is to check if the
signal in the oligofragments has a signal which is
between 0.97 and 1.03 daltons lower mass than the
signal itself, if false this peak is not a
descernable monooisotopic peak in this workflow
2093 'else if the peak is a descernable monoisotopic peak
it may be assigned
2094 light_logic=true
```

```
light_upper_window=oligofragments(c10)-small_dalton
light_lower_window=oligofragments(c10)-big_dalton
for lightest_cluster_check_counter=1 to
oligofragmentcount
light_check=oligofragments(
lightest_cluster_check_counter)
if light_check<light_upper_window and light_check
>light_lower_window then
light_logic=false
else
End If
Next
if light_logic=true then
f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
f6.Write "a"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr(
9)+Chr(45)+CStr(0)+Chr(44)+Chr(9)+CStr((a1_mz(c9
)+(1*H)))+Chr(44)+Chr(9)+"C"+CStr(a1C(c9))+"H"+
CStr(a1H(c9)+1)+"F"+CStr(a1F(c9))+"N"+CStr(a1N(c9
))+"O"+CStr(a1O(c9))+"P"+CStr(a1P(c9))+"S"+CStr(
a1S(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-(a1_mz(c9)+(1*H)))/(a1_mz(c9
)+(1*H))*1000000),4))&Chr(13)
percent(d1)=1
a_location(d1)=1
Else
End if
else
End if
if oligofragments(c10)<CDbl((ab_mz(c9)+(1*H)))+CDbl((
ab_mz(c9)+(1*H)))*ppmrange2 and oligofragments(c10)>CDbl
((ab_mz(c9)+(1*H)))-CDbl((ab_mz(c9)+(1*H)))*ppmrange2 and
d1<c12 then
light_logic=true
light_upper_window=oligofragments(c10)-small_dalton
light_lower_window=oligofragments(c10)-big_dalton
for lightest_cluster_check_counter=1 to
oligofragmentcount
light_check=oligofragments(
lightest_cluster_check_counter)
if light_check<light_upper_window and light_check
>light_lower_window then
light_logic=false
else
End If
Next
if light_logic=true then
f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
f6.Write "ab"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr
(9)+Chr(45)+CStr(0)+Chr(44)+Chr(9)+CStr((ab_mz(c9
)+(1*H)))+Chr(44)+Chr(9)+"C"+CStr(abC(c9))+"H"+
CStr(abH(c9)+1)+"F"+CStr(abF(c9))+"N"+CStr(abN(c9
))+"O"+CStr(abO(c9))+"P"+CStr(abP(c9))+"S"+CStr(
ab_S(c9))+Chr(44)+Chr(9)+CStr(Round(((
```

```
oligofragments(c10)-(ab_mz(c9)+(1*H)))/(ab_mz(c9
)+(1*H))*1000000),4))&Chr(13)
2127 percent(d1)=1
2128 ab_location(d1)=1
2129 Else
2130 End if
2131 else
2132 end if
2133 if oligofragments(c10)<CDbl((b_mz(c9)+(1*H)))+CDbl((b_mz(
c9)+(1*H)))*ppmrange2 and oligofragments(c10)>CDbl((b_mz(
c9)+(1*H)))-CDbl((b_mz(c9)+(1*H)))*ppmrange2 and d1<c12
then
2134 light_logic=true
2135 light_upper_window=oligofragments(c10)-small_dalton
2136 light_lower_window=oligofragments(c10)-big_dalton
2137 for lightest_cluster_check_counter=1 to
oligofragmentcount
2138 light_check=oligofragments(
lightest_cluster_check_counter)
2139 if light_check<light_upper_window and light_check
>light_lower_window then
2140 light_logic=false
2141 else
2142 End If
2143 Next
2144 if light_logic=true then
2145 'annotate here
2146 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
2147 f6.Write "b"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr(
9)+Chr(45)+CStr(0)+Chr(44)+Chr(9)+CStr((b_mz(c9
)+(1*H)))+Chr(44)+Chr(9)+"C"+CStr(bC(c9))+"H"+
CStr(bH(c9)+1)+"F"+CStr(bF(c9))+"N"+CStr(bN(c9))+
"O"+CStr(bO(c9))+"P"+CStr(bP(c9))+"S"+CStr(bS(c9
))+Chr(44)+Chr(9)+CStr(Round(((oligofragments(c10
)-(b_mz(c9)+(1*H)))/(b_mz(c9)+(1*H))*1000000),4
))&Chr(13)
2148 percent(d1)=1
2149 b_location(d1)=1
2150 Else
2151 End If
2152 else
2153 end if
2154 if oligofragments(c10)<CDbl((c_mz(c9)+(1*H)))+CDbl((c_mz(
c9)+(1*H)))*ppmrange2 and oligofragments(c10)>CDbl((c_mz(
c9)+(1*H)))-CDbl((c_mz(c9)+(1*H)))*ppmrange2 and d1<c12
then
2155 light_logic=true
2156 light_upper_window=oligofragments(c10)-small_dalton
2157 light_lower_window=oligofragments(c10)-big_dalton
2158 for lightest_cluster_check_counter=1 to
oligofragmentcount
2159 'MsgBox
"m/z="&CStr(oligofragments(lightest_cluster_check
```

```
_counter))&Chr(13)&"light_upper_window="&CStr(lig
ht_upper_window)&Chr(13)&"c_mz="&CStr((c_mz(c9)+(
1*H)))&Chr(13)&"light_lower_window="&CStr(light_l
ower_window)
2160 light_check=oligofragments(
lightest_cluster_check_counter)
2161 if light_check<light_upper_window and light_check
>light_lower_window then
2162 light_logic=false
2163 else
2164 End if
2165 Next
2166 if light_logic=true then
2167 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
2168 f6.Write "c"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr(
9)+Chr(45)+CStr(0)+Chr(44)+Chr(9)+CStr((c_mz(c9
)+(1*H)))+Chr(44)+Chr(9)+"C"+CStr(cC(c9))+"H"+
CStr(cH(c9)+1)+"F"+CStr(cF(c9))+"N"+CStr(cN(c9))+
"O"+CStr(cO(c9))+"P"+CStr(cP(c9))+"S"+CStr(cS(c9
))+Chr(44)+Chr(9)+CStr(Round(((oligofragments(c10
)-(c_mz(c9)+(1*H))))/(c_mz(c9)+(1*H))*1000000),4
))&Chr(13)
2169 percent(d1)=1
2170 c_location(d1)=1
2171 Else
2172 End if
2173 else
2174 end if
2175 if oligofragments(c10)<CDbl((d_mz(c9)+(1*H)))+CDbl((d_mz(
c9)+(1*H)))*ppmrange2 and oligofragments(c10)>CDbl((d_mz(
c9)+(1*H)))-CDbl((d_mz(c9)+(1*H)))*ppmrange2 and d1<c12
then
2176 light_logic=true
2177 light_upper_window=oligofragments(c10)-small_dalton
2178 light_lower_window=oligofragments(c10)-big_dalton
2179 for lightest_cluster_check_counter=1 to
oligofragmentcount
2180 light_check=oligofragments(
lightest_cluster_check_counter)
2181 if light_check<light_upper_window and light_check
>light_lower_window then
2182 light_logic=false
2183 else
2184 End If
2185 Next
2186 if light_logic=true then
2187 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
2188 f6.Write "d"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr(
9)+Chr(45)+CStr(0)+Chr(44)+Chr(9)+CStr((d_mz(c9
)+(1*H)))+Chr(44)+Chr(9)+"C"+CStr(dC(c9))+"H"+
CStr(dH(c9)+1)+"F"+CStr(dF(c9))+"N"+CStr(dN(c9))+
"O"+CStr(d_O(c9))+"P"+CStr(dP(c9))+"S"+CStr(dS(c9
))+Chr(44)+Chr(9)+CStr(Round(((oligofragments(c10
```

```
)-(d_mz(c9)+(1*H)))/(d_mz(c9)+(1*H))*1000000),4
))&Chr(13)
2189 percent(d1)=1
2190 d_location(d1)=1
2191 Else
2192 end If
2193 else
2194 end if
2195 if oligofragments(c10)<CDbl((w_mz(c9)+(1*H)))+CDbl((w_mz(
c9)+(1*H)))*ppmrange2 and oligofragments(c10)>CDbl((w_mz(
c9)+(1*H)))-CDbl((w_mz(c9)+(1*H)))*ppmrange2 and d1<=c12
then
2196 light_logic=true
2197 light_upper_window=oligofragments(c10)-small_dalton
2198 light_lower_window=oligofragments(c10)-big_dalton
2199 for lightest_cluster_check_counter=1 to
oligofragmentcount
2200 light_check=oligofragments(
lightest_cluster_check_counter)
2201 if light_check<light_upper_window and light_check
>light_lower_window then
2202 light_logic=false
2203 else
2204 End If
2205 Next
2206 if light_logic=true then
2207 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
2208 f6.Write "w"+Chr(44)+Chr(9)+CStr(c9)+Chr(44)+Chr(
9)+Chr(45)+CStr(0)+Chr(44)+Chr(9)+CStr((w_mz(c9
)+(1*H)))+Chr(44)+Chr(9)+"C"+CStr(wC(c9))+"H"+
CStr(wH(c9)+1)+"F"+CStr(wF(c9))+"N"+CStr(wN(c9))+
"O"+CStr(wO(c9))+"P"+CStr(wP(c9))+"S"+CStr(wS(c9
))+Chr(44)+Chr(9)+CStr(Round(((oligofragments(c10
)-(w_mz(c9)+(1*H)))/(w_mz(c9)+(1*H))*1000000),4
))&Chr(13)
2209 percent(c12-c9)=1
2210 w_location(c12-c9)=1
2211 Else
2212 End if
2213 else
2214 end if
2215 if oligofragments(c10)<CDbl((x_mz(c9)+(1*H)))+CDbl((x_mz(
c9)+(1*H)))*ppmrange2 and oligofragments(c10)>CDbl((x_mz(
c9)+(1*H)))-CDbl((x_mz(c9)+(1*H)))*ppmrange2 and d1<=c12
then
2216 light_logic=true
2217 light_upper_window=oligofragments(c10)-small_dalton
2218 light_lower_window=oligofragments(c10)-big_dalton
2219 for lightest_cluster_check_counter=1 to
oligofragmentcount
2220 light_check=oligofragments(
lightest_cluster_check_counter)
2221 if light_check<light_upper_window and light_check
```

```
>light_lower_window then
2222 light_logic=false
2223 else
2224 End If
2225 Next
2226 if light_logic=true then
2227 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
2228 f6.Write "x"+Chr(44)+Chr(9)+CStr(c9)+Chr(44)+Chr(
9)+Chr(45)+CStr(0)+Chr(44)+Chr(9)+CStr((x_mz(c9
)+(1*H)))+Chr(44)+Chr(9)+"C"+CStr(xC(c9))+"H"+
CStr(xH(c9)+1)+"F"+CStr(xF(c9))+"N"+CStr(xN(c9))+
"O"+CStr(xO(c9))+"P"+CStr(xP(c9))+"S"+CStr(xS(c9
))+Chr(44)+Chr(9)+CStr(Round(((oligofragments(c10
)-(x_mz(c9)+(1*H)))/(x_mz(c9)+(1*H))*1000000),4
))&Chr(13)
2229 percent(c12-c9)=1
2230 x_location(c12-c9)=1
2231 Else
2232 end if
2233 else
2234 end if
2235 if oligofragments(c10)<CDbl((y_mz(c9)+(1*H)))+CDbl((y_mz(
c9)+(1*H)))*ppmrange2 and oligofragments(c10)>CDbl((y_mz(
c9)+(1*H)))-CDbl((y_mz(c9)+(1*H)))*ppmrange2 and d1<=c12
then
2236 light_logic=true
2237 light_upper_window=oligofragments(c10)-small_dalton
2238 light_lower_window=oligofragments(c10)-big_dalton
2239 for lightest_cluster_check_counter=1 to
oligofragmentcount
2240 light_check=oligofragments(
lightest_cluster_check_counter)
2241 if light_check<light_upper_window and light_check
>light_lower_window then
2242 light_logic=false
2243 else
2244 End If
2245 Next
2246 if light_logic=true then
2247 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
2248 f6.Write "y"+Chr(44)+Chr(9)+CStr(c9)+Chr(44)+Chr(
9)+Chr(45)+CStr(0)+Chr(44)+Chr(9)+CStr((y_mz(c9
)+(1*H)))+Chr(44)+Chr(9)+"C"+CStr(yC(c9))+"H"+
CStr(yH(c9)+1)+"F"+CStr(yF(c9))+"N"+CStr(yN(c9))+
"O"+CStr(yO(c9))+"P"+CStr(yP(c9))+"S"+CStr(yS(c9
))+Chr(44)+Chr(9)+CStr(Round(((oligofragments(c10
)-(y_mz(c9)+(1*H)))/(y_mz(c9)+(1*H))*1000000),4
))&Chr(13)
2249 percent(c12-c9)=1
2250 y_location(c12-c9)=1
2251 Else
2252 End if
2253 else
```

```
2254 end if
2255 if oligofragments(c10)<CDbl((z_mz(c9)+(1*H)))+CDbl((z_mz(
c9)+(1*H)))*ppmrange2 and oligofragments(c10)>CDbl((z_mz(
c9)+(1*H)))-CDbl((z_mz(c9)+(1*H)))*ppmrange2 and d1<=c12
then
2256 light_logic=true
2257 light_upper_window=oligofragments(c10)-small_dalton
2258 light_lower_window=oligofragments(c10)-big_dalton
2259 for lightest_cluster_check_counter=1 to
oligofragmentcount
2260 light_check=oligofragments(
lightest_cluster_check_counter)
2261 if light_check<light_upper_window and light_check
>light_lower_window then
2262 light_logic=false
2263 else
2264 End If
2265 Next
2266 if light_logic=true then
2267 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
2268 f6.Write "z"+Chr(44)+Chr(9)+CStr(c9)+Chr(44)+Chr(
9)+Chr(45)+CStr(0)+Chr(44)+Chr(9)+CStr((z_mz(c9
)+(1*H)))+Chr(44)+Chr(9)+"C"+CStr(zC(c9))+"H"+
CStr(zH(c9)+1)+"F"+CStr(zF(c9))+"N"+CStr(zN(c9))+
"O"+CStr(zO(c9))+"P"+CStr(zP(c9))+"S"+CStr(zS(c9
))+Chr(44)+Chr(9)+CStr(Round(((oligofragments(c10
)-(z_mz(c9)+(1*H)))/(z_mz(c9)+(1*H))*1000000),4
))&Chr(13)
2269 percent(c12-c9)=1
2270 z_location(c12-c9)=1
2271 else
2272 End If
2273 else
2274 end if
2275 'Next
2276 'c=c+1
2277 c10=c10+1
2278 Next
2279 Else
2280 c10=0
2281 For Each signal In oligofragments
2282
2283 For z=1 to max_charge_fragment
2284 if oligofragments(c10)<CDbl((a1_mz(c9)+(1+z)*H)/z)+
CDbl((a1_mz(c9)+(1+z)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((a1_mz(c9)+(1+z)*H)/z)-CDbl
((a1_mz(c9)+(1+z)*H)/z)*ppmrange2 and d1<c12 then
2285 'annotate here
2286 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
2287 f6.Write "a"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr(
9)+Chr(43)+CStr(z)+Chr(44)+Chr(9)+CStr((a1_mz(c9
)+(1+z)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(a1C(c9))+
"H"+CStr(a1H(c9)+(z+1))+"F"+CStr(a1F(c9))+"N"+
```

```
CStr(a1N(c9))+"O"+CStr(a1O(c9))+"P"+CStr(a1P(c9
))+"S"+CStr(a1S(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((a1_mz(c9)+(1+z)*H)/z))/((
a1_mz(c9)+(1+z)*H)/z)*1000000),4))&Chr(13)
percent(d1)=1
a_location(d1)=1
else
End if
if oligofragments(c10)<CDbl((ab_mz(c9)+(1+z)*H)/z)+
CDbl((ab_mz(c9)+(1+z)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((ab_mz(c9)+(1+z)*H)/z)-CDbl
((ab_mz(c9)+(1+z)*H)/z)*ppmrange2 and d1<c12 then
'Cmpd(2).Annotations.AddAnnotation ab_mz_L,
prime_int*1.2, ab_mz_H, prime_int*1.2,
prime_int*1.6, "a-b"+CStr(d), True
'annotate here
f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
f6.Write "ab"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr
(9)+Chr(43)+CStr(z)+Chr(44)+Chr(9)+CStr((ab_mz(c9
)+(1+z)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(abC(c9))+
"H"+CStr(abH(c9)+(z+1))+"F"+CStr(abF(c9))+"N"+
CStr(abN(c9))+"O"+CStr(abO(c9))+"P"+CStr(abP(c9
))+"S"+CStr(ab_S(c9))+Chr(44)+Chr(9)+CStr(Round
(((oligofragments(c10)-((ab_mz(c9)+(1+z)*H)/z
))/((ab_mz(c9)+(1+z)*H)/z)*1000000),4))&Chr(13)
percent(d1)=1
ab_location(d1)=1
else
end if
if oligofragments(c10)<CDbl((b_mz(c9)+(1+z)*H)/z)+
CDbl((b_mz(c9)+(1+z)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((b_mz(c9)+(1+z)*H)/z)-CDbl((
b_mz(c9)+(1+z)*H)/z)*ppmrange2 and d1<c12 then
'annotate here
f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
f6.Write "b"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr(
9)+Chr(43)+CStr(z)+Chr(44)+Chr(9)+CStr((b_mz(c9
)+(1+z)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(bC(c9))+"H"
+CStr(bH(c9)+(z+1))+"F"+CStr(bF(c9))+"N"+CStr(bN(
c9))+"O"+CStr(bO(c9))+"P"+CStr(bP(c9))+"S"+CStr(
bS(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((b_mz(c9)+(1+z)*H)/z))/((
b_mz(c9)+(1+z)*H)/z)*1000000),4))&Chr(13)
percent(d1)=1
b_location(d1)=1
else
end if
if oligofragments(c10)<CDbl((c_mz(c9)+(1+z)*H)/z)+
CDbl((c_mz(c9)+(1+z)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((c_mz(c9)+(1+z)*H)/z)-CDbl((
c_mz(c9)+(1+z)*H)/z)*ppmrange2 and d1<c12 then
f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
f6.Write "c"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr(
```

```
9)+Chr(43)+CStr(z)+Chr(44)+Chr(9)+CStr((c_mz(c9
)+(1+z)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(cC(c9))+"H"
+CStr(cH(c9)+(z+1))+"F"+CStr(cF(c9))+"N"+CStr(cN(
c9))+"O"+CStr(cO(c9))+"P"+CStr(cP(c9))+"S"+CStr(
cS(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((c_mz(c9)+(1+z)*H)/z))/((
c_mz(c9)+(1+z)*H)/z)*1000000),4))&Chr(13)
2312 percent(d1)=1
2313 c_location(d1)=1
2314 else
2315 end if
2316 if oligofragments(c10)<CDbl((d_mz(c9)+(1+z)*H)/z)+
CDbl((d_mz(c9)+(1+z)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((d_mz(c9)+(1+z)*H)/z)-CDbl((
d_mz(c9)+(1+z)*H)/z)*ppmrange2 and d1<c12 then
2317 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
2318 f6.Write "d"+Chr(44)+Chr(9)+CStr(d1)+Chr(44)+Chr(
9)+Chr(43)+CStr(z)+Chr(44)+Chr(9)+CStr((d_mz(c9
)+(1+z)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(dC(c9))+"H"
+CStr(dH(c9)+2)+"F"+CStr(dF(c9))+"N"+CStr(dN(c9
))+"O"+CStr(d_O(c9))+"P"+CStr(dP(c9))+"S"+CStr(dS
(c9))+Chr(44)+Chr(9)+CStr(Round(((oligofragments(
c10)-((d_mz(c9)+(1+z)*H)/z))/((d_mz(c9)+(1+z)*H)/
z)*1000000),4))&Chr(13)
2319 percent(d1)=1
2320 d_location(d1)=1
2321 else
2322 end if
2323 if oligofragments(c10)<CDbl((w_mz(c9)+(1+z)*H)/z)+
CDbl((w_mz(c9)+(1+z)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((w_mz(c9)+(1+z)*H)/z)-CDbl((
w_mz(c9)+(1+z)*H)/z)*ppmrange2 and d1<=c12 then
2324 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
2325 f6.Write "w"+Chr(44)+Chr(9)+CStr(c9)+Chr(44)+Chr(
9)+Chr(43)+CStr(z)+Chr(44)+Chr(9)+CStr((w_mz(c9
)+(1+z)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(wC(c9))+"H"
+CStr(wH(c9)+(z+1))+"F"+CStr(wF(c9))+"N"+CStr(wN(
c9))+"O"+CStr(wO(c9))+"P"+CStr(wP(c9))+"S"+CStr(
wS(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((w_mz(c9)+(1+z)*H)/z))/((
w_mz(c9)+(1+z)*H)/z)*1000000),4))&Chr(13)
2326 'MsgBox "c12="&c12
2327 'MsgBox "c9="&c9
2328 percent(c12-c9)=1
2329 w_location(c12-c9)=1
2330 else
2331 end if
2332 if oligofragments(c10)<CDbl((x_mz(c9)+(1+z)*H)/z)+
CDbl((x_mz(c9)+(1+z)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((x_mz(c9)+(1+z)*H)/z)-CDbl((
x_mz(c9)+(1+z)*H)/z)*ppmrange2 and d1<=c12 then
2333 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
2334 f6.Write "x"+Chr(44)+Chr(9)+CStr(c9)+Chr(44)+Chr(
```

```
9)+Chr(43)+CStr(z)+Chr(44)+Chr(9)+CStr((x_mz(c9
)+(1+z)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(xC(c9))+"H"
+CStr(xH(c9)+(z+1))+"F"+CStr(xF(c9))+"N"+CStr(xN(
c9))+"O"+CStr(xO(c9))+"P"+CStr(xP(c9))+"S"+CStr(
xS(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((x_mz(c9)+(1+z)*H)/z))/((
x_mz(c9)+(1+z)*H)/z)*1000000),4))&Chr(13)
2335 percent(c12-c9)=1
2336 x_location(c12-c9)=1
2337 else
2338 end if
2339 if oligofragments(c10)<CDbl((y_mz(c9)+(1+z)*H)/z)+
CDbl((y_mz(c9)+(1+z)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((y_mz(c9)+(1+z)*H)/z)-CDbl((
y_mz(c9)+(1+z)*H)/z)*ppmrange2 and d1<=c12 then
2340 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
2341 f6.Write "y"+Chr(44)+Chr(9)+CStr(c9)+Chr(44)+Chr(
9)+Chr(43)+CStr(z)+Chr(44)+Chr(9)+CStr((y_mz(c9
)+(1+z)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(yC(c9))+"H"
+CStr(yH(c9)+(z+1))+"F"+CStr(yF(c9))+"N"+CStr(yN(
c9))+"O"+CStr(yO(c9))+"P"+CStr(yP(c9))+"S"+CStr(
yS(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((y_mz(c9)+(1+z)*H)/z))/((
y_mz(c9)+(1+z)*H)/z)*1000000),4))&Chr(13)
2342 percent(c12-c9)=1
2343 y_location(c12-c9)=1
2344 else
2345 end if
2346 if oligofragments(c10)<CDbl((z_mz(c9)+(1+z)*H)/z)+
CDbl((z_mz(c9)+(1+z)*H)/z)*ppmrange2 and
oligofragments(c10)>CDbl((z_mz(c9)+(1+z)*H)/z)-CDbl((
z_mz(c9)+(1+z)*H)/z)*ppmrange2 and d1<=c12 then
2347 f6.Write CStr(oligofragments(c10))+Chr(44)+Chr(9)
2348 f6.Write "z"+Chr(44)+Chr(9)+CStr(c9)+Chr(44)+Chr(
9)+Chr(43)+CStr(z)+Chr(44)+Chr(9)+CStr((z_mz(c9
)+(1+z)*H)/z)+Chr(44)+Chr(9)+"C"+CStr(zC(c9))+"H"
+CStr(zH(c9)+(z+1))+"F"+CStr(zF(c9))+"N"+CStr(zN(
c9))+"O"+CStr(zO(c9))+"P"+CStr(zP(c9))+"S"+CStr(
zS(c9))+Chr(44)+Chr(9)+CStr(Round(((
oligofragments(c10)-((z_mz(c9)+(1+z)*H)/z))/((
z_mz(c9)+(1+z)*H)/z)*1000000),4))&Chr(13)
2349 percent(c12-c9)=1
2350 z_location(c12-c9)=1
2351 else
2352 end if
2353 Next
2354 'c=c+1
2355 c10=c10+1
2356 Next
2357 End if
2358 c9=c9+1
2359 Next
2360 c11=0
```

```
2361 sum=0
2362 'coverage=0
2363 For Each location in percent
2364 sum=sum+percent(c11)
2365 c11=c11+1
2366 Next
2367
2368 'sequence coverage table writing
2369 '
2370
2371 ''
2372 f6.Write "ab1"+Chr(9)+"-"
2373 c13=1
2374 For location=1 to (c11-1)
2375 f6.Write CStr(ab_location(c13))+Chr(44)+Chr(32)
2376 c13=c13+1
2377 Next
2378 f6.Write Chr(13)
2379 ''
2380 ''
2381 f6.Write "a1"+Chr(9)+"-"
2382 c13=1
2383 For location=1 to (c11-1)
2384 f6.Write CStr(a_location(c13))+Chr(44)+Chr(32)
2385 c13=c13+1
2386 Next
2387 f6.Write Chr(13)
2388 ''
2389 ''
2390 f6.Write "b1"+Chr(9)+"-"
2391 c13=1
2392 For location=1 to (c11-1)
2393 f6.Write CStr(b_location(c13))+Chr(44)+Chr(32)
2394 c13=c13+1
2395 Next
2396 f6.Write Chr(13)
2397 ''
2398 ''
2399 f6.Write "c1"+Chr(9)+"-"
2400 c13=1
2401 For location=1 to (c11-1)
2402 f6.Write CStr(c_location(c13))+Chr(44)+Chr(32)
2403 c13=c13+1
2404 Next
2405 f6.Write Chr(13)
2406 ''
2407 ''
2408 f6.Write "d1"+Chr(9)+"-"
2409 c13=1
2410 For location=1 to (c11-1)
2411 f6.Write CStr(d_location(c13))+Chr(44)+Chr(32)
2412 c13=c13+1
2413 Next
```

```
f6.Write Chr(13)
''
'''
''
f6.Write "w"+CStr(c11-1)+Chr(9)+"-"
c13=1
For location=1 to (c11-1)
f6.Write CStr(w_location(c13))+Chr(44)+Chr(32)
c13=c13+1
Next
f6.Write Chr(13)
''
''
f6.Write "x"+CStr(c11-1)+Chr(9)+"-"
c13=1
For location=1 to (c11-1)
f6.Write CStr(x_location(c13))+Chr(44)+Chr(32)
c13=c13+1
Next
f6.Write Chr(13)
''
''
f6.Write "y"+CStr(c11-1)+Chr(9)+"-"
c13=1
For location=1 to (c11-1)
f6.Write CStr(y_location(c13))+Chr(44)+Chr(32)
c13=c13+1
Next
f6.Write Chr(13)
''
''
f6.Write "z"+CStr(c11-1)+Chr(9)+"-"
c13=1
For location=1 to (c11-1)
f6.Write CStr(z_location(c13))+Chr(44)+Chr(32)
c13=c13+1
Next
f6.Write Chr(13)
''
'

f6.Write "percent coverage(%)="+CStr(sum/(c11-1)*100)&Chr(13)
else
'f6.Write oligoprecursor_H&Chr(13)
End If
c5=c5+1
Next

f6.Close
Next
'if c2=100 then
' '----'
' Dim WshShell
```

```
2467 ' Set WshShell = WScript.CreateObject("WScript.Shell")
2468 ' popup = WshShell.Popup("MS/MS spectra read="&d&"/"&begin_ion_count,
0.5)
2469 ' '----'
2470 ' c2=0
2471 'else
2472 'End If
2473 c2=c2+1
2474 MyFile.Close
2475 Set MyFile=Nothing
2476 c6=c6+1
2477 d=d+1
2478 c4=c4+1
2479 Next
2480 '-----------------------------------------------------------------
2481 'the interface code whichsets the user options
2482 '-----------------------------------------------
2483 ccc=0
2484 c8=c8+1
2485 'MsgBox c8
2486 Next
2487 Set f6 = fso.OpenTextFile(fragfilepath+"\"+"report.txt", 8, True)
2488 f6.Write CStr(WScript.ScriptName)&Chr(13)
2489 f6.Write "user inputs:"&Chr(13)
2490 f6.Write "precoursor mz tolerance="+CStr(ppmrange_precursor)&Chr(13)
2491 f6.Write "fragment mz tolerance="+CStr(ppmrange)&Chr(13)
2492 f6.Write "max charge of precoursor searched="+CStr(max_charge_per_base)&Chr(13)
2493 f6.Write "max charge of fragment searched="+CStr(max_charge_fragment)&Chr(13)
2494 f6.Write "require charge state comparison="+CStr(charge_comparison)&Chr(13)
2495 c15=0
2496 f6.Write "sequences searched:"&Chr(13)
2497 For Each oligo in oligos
2498 f6.Write CStr(oligos(c15))&Chr(13)
2499 'oligos(c)=MyoligoFile.ReadLine
2500 'oligo_2=Split(oligos(c),"=")
2501 'oligo_1(c)=oligo_2(0)
2502 'oligoname(c)=oligo_2(1)
2503 'MsgBox oligoname(c)
2504 'fragsfilepaths(c)=fragfilepath
2505 c15=c15+1
2506 Next
2507
2508 f6.Close
2509
2510 MyTime2=Time
2511 'MsgBox Mytime1
2512 'MsgBox MyTime2
2513
2514 MsgBox CStr(Mytime1)+Chr(13)+CStr(Mytime2)+Chr(13)+"Done"
2515 Set objshell = Nothing
2516 WScript.Quit()
2517 WSH.Quit()
2518
```