2017

Integrating glycomics, proteomics and glycoproteomics to understand the structural basis for influenza a virus evolution and glycan mediated immune interactions

BOSTON UNIVERSITY

SCHOOL OF MEDICINE

Dissertation

INTEGRATING GLYCOMICS, PROTEOMICS AND GLYCOPROTEOMICS TO UNDERSTAND THE STRUCTURAL BASIS FOR INFLUENZA A VIRUS EVOLUTION AND GLYCAN MEDIATED IMMUNE INTERACTIONS

by

KSHITIJ KHATRI

B.Tech., Amity University, 2008
M.S., Northeastern University, 2010

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2017

Approved by

First Reader _____

Joseph Zaia, Ph.D.
Professor of Biochemistry

Second Reader _____

Cheng Lin, Ph.D.
Research Associate Professor of Biochemistry

DEDICATION

This thesis is dedicated to my late grandfather, Shri Kush Dev Khatri.

ACKNOWLEDGMENTS

I am extremely grateful to the people who supported me during the course of my pre-doctoral training. I feel fortunate to have been a part of the Center for Biomedical Mass Spectrometry and the department of Biochemistry. Over the years, the lab has become less of a workplace and more of a home for me, and it is going to be hard to say goodbye.

Above all, I would like to thank my advisor Prof. Joseph Zaia for his unwavering support and guidance that helped me develop my skills as an independent researcher. Joe maintains an unspoken understanding of his group members' personalities and work-styles and provides them the opportunities that help them become successful. Thank you Joe, for allowing me to work in your group, providing me great opportunities, freedom and advice and for being a patient and caring mentor! I could not have found a better advisor to pursue my graduate studies!

I would also like to thank my committee chair, Prof. Catherine E. Costello, for providing a collaborative and fun work environment at CBMS. Thank you Cathy for being a source of inspiration for all of us working under your leadership. I never cease to be amazed by the level of enthusiasm and energy you bring both to the workplace and to all social events. Thank you for your support and invaluable advice on matters related to science, career and life.

I am extremely grateful to Prof. Cheng Lin and Prof. Mark E. McComb for their guidance on many projects during my thesis and for serving on my thesis committee.

Also, many thanks go to my committee members Prof. Matthew A. Nugent and Prof. Barbara A. Seaton for providing their valuable time and feedback on my research.

During my pre-doctoral training, I have had the opportunity to work with and learn from many incredibly talented people, both at BU and through outside collaborations. I am particularly thankful to Joshua Klein for his help with numerous projects that required informatics skills and in general for adding so much value to our lab as a colleague and a friend. I owe much gratitude to Dr. Nancy Leymarie for her help in getting me started in the lab and not only teaching me the basics of mass spectrometry and glycoproteomics but also being a friend and a mentor when I needed guidance. I am also thankful to Dr. Deborah R. Leon for being a great friend and colleague and for many helpful discussions over the years.

I had the privilege to work closely with Dr. Yi Pu and Dr. Rebecca Glaskin on ExD and ion-mobility spectrometry related projects and would like to thank them for their collaborative efforts and for sharing with me their technical knowledge of these areas. I also wish to thank my collaborators on various projects - Prof. Kevan L. Hartshorn, Dr. Gregory O. Staples, Dr. Oliver C. Grant, Dr. Rosa Viner, Dr. Michael J. Rynkiewicz, Mitchell R. White and Ken Shun Yip. I especially want to thank my wonderful labmates Dr. Le Meng, Dr. Xiaofeng Shi, Dr. Yang Mao, Dr. Yu Huang, Dr. Han Hu, Dr. Chun Shao, Dr. Bo Yan, Dr. Kevin Chandler, John Haserick, Christian Heckendorf and Rekha Raghunathan for making the lab such a fun place to work and for their day-to-day help and scientific advice.

I am extremely grateful to Prof. Barbara Schreiber for the continued support and encouragement I have received from her during my time at BU and to Prof. Matthew Layne for serving as my advisor during the first year of the program. A special thanks goes to Denise Neves, Vicki Glenn, Melissa Howard, Erika Casavant and Jane Mahoney for taking care of the administrative tasks and enabling us to focus on the science.

Being an international student comes with many challenges, when you are thousands of miles from home, family and friends. I consider myself lucky to have a support system that made me feel comfortable and secure. Many thanks go to the Dhingra and Arora families for welcoming me into their lives and providing me a home away from home. I have enjoyed spending many holidays and memorable family events with them, which I will cherish throughout my life.

The credit for me being able to keep whatever little sanity I had through graduate school goes to my friends Tushar Murthy, Sushant Dhiman, Edwin Motari, Aleks Szymaniak and Deborah Leon. Thank you for your friendship and sharing all the fun moments that helped me get through stressful periods of work and life.

I am unable to express in words my gratitude toward my parents, Sucheta and Ashwani Khatri, who are the reason behind all of my success. I am thankful to my father, mother, grandmother and sister for the unconditional love I have received throughout my life. None of this would have been possible without my family's love and support – I love you all very much!

INTEGRATING GLYCOMICS, PROTEOMICS AND GLYCOPROTEOMICS TO UNDERSTAND THE STRUCTURAL BASIS FOR INFLUENZA A VIRUS EVOLUTION AND GLYCAN MEDIATED IMMUNE INTERACTIONS

KSHITIJ KHATRI

Boston University School of Medicine, 2017

Major Professor: Joseph Zaia, Ph.D., Professor of Biochemistry

ABSTRACT

Glycosylation modulates the range and specificity of interactions among glycoproteins and their binding partners. This is important in influenza A virus (IAV) biology because binding of host immune molecules depends on glycosylation of viral surface proteins such as hemagglutinin (HA). Circulating viruses mutate rapidly in response to pressure from the host immune system. As proteins mutate, the virus glycosylation patterns change. The consequence is that viruses evolve to evade host immune responses, which renders vaccines ineffective. Glycan biosynthesis is a non-template driven process, governed by stoichiometric and steric relationships between the enzymatic machinery for glycosylation and the protein being glycosylated. Consequently, protein glycosylation is heterogeneous, thereby making structural analysis and elucidation of precise biological functions extremely challenging. The lack of structural information has been a limiting factor in understanding the exact mechanisms of glycan-mediated interactions of the IAV with host immune-lectins. Genetic sequencing methods allow prediction of glycosylation sites along the protein backbone but are unable to

provide exact phenotypic information regarding site occupancy. Crystallography methods are also unable to determine the glycan structures beyond the core residues due to the flexible nature of carbohydrates. This dissertation centers on the development of chromatography and mass spectrometry methods for characterization of site-specific glycosylation in complex glycoproteins and application of these methods to IAV glycomics and glycoproteomics. We combined the site-specific glycosylation information generated using mass spectrometry with information from biochemical assays and structural modeling studies to identify key glycosylation sites mediating interactions of HA with immune lectin surfactant protein-D (SP-D). We also identified the structural features that control glycan processing at these sites, particularly those involving glycan maturation from high-mannose to complex-type, which, in turn, regulate interactions with SP-D. The work presented in this dissertation contributes significantly to the improvement of analytical and bioinformatics methods in glycan and glycoprotein analysis using mass spectrometry and greatly advances the understanding of the structural features regulating glycan microheterogeneity on HA and its interactions with host immune lectins.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2-AA .....Anthranilic Acid

2-AB.....2-Aminobenzamide

2-AP 2-Aminopyridine

AGP.....Alpha-1-acid glycoprotein/α1-acid glycoprotein

AIF ..... All ions fragmentation

AIF .....All-ions fragmentation

AMAC..... 9-Aminoacridone

ANTS ..... Aminonapththalene Trisulfolane

APTS.....Aminopyrene Trisulfolane

BGE..... Background Electrolyte

BPC.....Base Peak Chromatogram

CAD .....Collisionally-Activated Dissociation

Cam .....Carbamidomethylation (at cysteine)

CCS.....Collision Cross Section

CE .....Capillary Electrophoresis

CE-MS .....Capillary electrophoresis – mass spectrometry

ConA.....Concanavalin A

CRD .....Carbohydrate recognition domain

C-trap .....Curved RF-only quadrupole ion trap

D+R..... NCRD with two amino acid mutations

DC..... Direct Current

DDA.....Data-dependent acquisition

dHex.....Deoxyhexose

DIA ..... Data-Independent Acquisition

DMSO.....Dimethyl sulfoxide

dp..... Degree of polymerization

DTIMS ..... Drift-tube Ion-Mobility Spectrometry

ECD..... Electron Capture Dissociation

ECnoD..... Electron Capture with no Dissociation

EDD .....Electron Detachment Dissociation

EED.....Electronic Excitation Dissociation

EIC .....Extracted Ion Chromatogram

EIE .....Extracted-ion Electropherogram

EOF ..... Electroosmotic Flow

ER .....Endoplasmic Reticulum

ERLIC .....Electrostatic repulsion hydrophilic interaction liquid chromatography

ERManI..... Endoplasmic reticulum class I α-mannosidase

ESI..... Electrospray Ionization

ETD..... Electron Transfer Dissociation

EThcD .....Electron Transfer Dissociation with supplemental collisional activation

ETnoD ..... Electron Transfer with no Dissociation

ExD ..... Electron Activated Dissociation

FAIMS .....High Field Asymmetric Waveform Ion Mobility Spectrometry

FTICR ..... Fourier Transform Ion Cyclotron Resonance

Fuc..... Fucose

FucT ..... Fucosyl-transferase

GAG/GAGs.....Glycosaminoglycan/Glycosaminoglycans

Gal.....Galactose

GalNAc .....*N*-Acetylgalactosamine

Glc..... Glucose

GlcNAc .....*N*-Acetylglucosamine

GlcNAcT.....GlcNAc-transferase

HA.....Hemagglutinin

HCD .....Higher-energy Collisional Dissociation

hECD..... Hot Electron Capture Dissociation

Hex.....Hexose

HexNAc .....*N*-Acetylhexosamine

HILIC..... Hydrophilic Interaction Liquid Chromatography

HILIC-C18-MS...HILIC enrichment prior to C18 analytical separation with MS detection

HPLC ..... High Performance Liquid Chromatography

HR/AM .....High-Resolution/Accurate Mass

huNCRD ..... Wild-type human Neck and Carbohydrate Recognition Domain of SP-D

IAV .....Influenza A Virus

ICR.....Ion Cyclotron Resonance

IM-MS.....Ion-mobility-Mass Spectrometry

IMS ..... Ion-Mobility Spectrometry

INLIGHTIndividuality Normalization when Labeling with Isotopic Glycan Hydrazide Tags

LAC..... Lectin-Affinity Chromatography

LacdiNAc..... GalNAcβ1-4GlcNAc

LacNAc..... Galβ1-3GlcNAc (type 1) or Galβ1-4GlcNAc (type 2)

LC ..... Liquid Chromatography

LC-MS ..... Liquid Chromatography-Mass Spectrometry

LC-MS/MS ..... Liquid Chromatography-Tandem Mass Spectrometry

LIF..... Laser-Induced Fluorescence

LMWH..... Low Molecular Weight Heparin

m/z..... Mass-to-charge ratio

MALDI ..... Matrix-Assisted Laser Desorption/Ionization

MALDI-MS ..... Matrix-assisted Laser Desorption/Ionization – Mass Spectrometry

Man ..... Mannose

MD ..... Molecular Dynamics

M-LAC..... Multi-Lectin Affinity Column

MS..... Mass Spectrometry

MSn..... Sequential tandem MS (Multi-stage tandem MS)

NA..... Neuraminidase

NCE..... Normalized Collision Energy

NETD..... Negative Electron Transfer Dissociation

NeuAc/NeuNAc..... *N*-acetyl neuraminic acid

NeuGc .....*N*-glycolyl neuraminic acid

NMR ..... Nuclear Magnetic Resonance

PGC.....Porous Graphitized Carbon

Phil-82..... A/Philippines/2/1982 (H3N2)

Phil-BS..... A/Philippines/2/1982/BS (H3N2) (Bovine serum escape mutant)

PNGase F ..... peptide-N4-(*N*-acetyl-β-glucosaminyl) asparagine amidase

pSP-D.....Porcine SP-D

PR-08 ..... A/Puerto Rico/8/1934 (H1N1)

PTM .....Post-translational modification

Q.....Quadrupole

R343V ..... NCRD with single amino acid mutation

RF.....Radio Frequency

rhSPDII ..... Recombinant full length human SP-D

RNase B ..... Ribonuclease B

RPLC.....Reversed-Phase Liquid Chromatography

RpNCRD.....Recombinant Neck and Carbohydrate Recognition Domain of porcine surfactant protein-D

SA .....Sialic Acid

SAIMS .....Selected Accumulation-Ion Mobility Spectrometry

SA-TIMS.....Selective Accumulation – Trapped Ion Mobility Spectrometry

SAX.....Strong Anion eXchange

SEC .....Size Exclusion Chromatography

SEC-IS-MS .....Size Exclusion Chromatography-Ion Suppressor-Mass Spectrometry

SEC-MS ..... Size Exclusion Chromatography-Mass Spectrometry

SP-D.....Surfactant Protein-D (Human SP-D, unless mentioned otherwise)

SPE.....Solid-Phase Extraction

SRM .....Selected Reaction Monitoring

SWATH ..... Sequential Window Acquisition of All THeoretical fragment-ion spectra

TIC .....Total Ion Chromatogram

TIE .....Total Ion Electropherogram

TIMS ..... Trapped Ion Mobility Spectrometry

TMT ..... Tandem Mass Tags

TOF ..... Time of Flight

TWIMS ..... Traveling-Wave Ion-Mobility Spectrometry

WAX ..... Weak Anion eXchange

WGA .....Wheat Germ Agglutinin

Chapter 1: Introduction

1.1 Glycobiology

Glycobiology, the study of processes mediated by carbohydrates in biological systems, encompasses glycan biosynthesis and the roles of glycosylated molecules in health and disease. Glycans are ubiquitous in living organisms as free carbohydrates, lipid- and protein-linked. In addition to being the primary source for energy storage in most organisms, carbohydrates play important roles in defining the structural and biochemical properties of glycoconjugates. While the total number of native gene products coded by an organism's genome is well-defined, the presence of glycosylation greatly expands the range of phenotypes and functions of gene products.

Glycans define biological processes in two different ways: first, by modulating the structural properties of the protein to which they are attached, and second, by modulating interactions with binding partners. Glycans alter structural stability and confer proteolytic resistance to the proteins they decorate. The large size of branched oligosaccharides also shields functionally important regions of proteins against immune molecule or pathogen binding. Thus, the partners to which a glycoprotein binds depend to a large degree on binding between glycans and lectin domain-containing proteins. All animal cells have a surface layer of glycans and glycoconjugates coating them, known as the "glycocalyx", which defines cell-cell interactions and also mediates highly specific interactions with pathogens [134,163,540,541,542,547,643]. In tissue, cells are

surrounded by an extracellular matrix consisting of carbohydrates, glycoproteins, proteoglycans and collagens.

Glycans are made up of monosaccharides, joined together into linear or branched oligosaccharides (2-20 units) or polysaccharides (>20 units). The glycosidic bonds that join these monosaccharides are formed between the hemiacetal group of the anomeric carbon of one monosaccharide and a hydroxyl group of another monosaccharide. Thus, the glycan biopolymers usually contain reducing and non-reducing ends, with the reducing sugar bearing a free anomeric center that is not involved in a glycosidic bond. The reducing end with an aldehyde functional group is often used as a handle for glycan tagging and enrichment, as described in the following sections.

In contrast to the biosynthesis of DNA, RNA and proteins, which are all template driven processes, synthesis of glycans is controlled by interactions of substrates with the glycosylation machinery consisting of a series of enzymes (glycosyltransferases and glycosidases) in the endoplasmic reticulum (ER) and the Golgi apparatus. The factors that affect protein glycosylation include conformational access to the enzymes, enzyme and substrate concentrations, the physiochemical characteristics of the glycoprotein being processed, and the tissue or cell-type-specific properties of the secretory pathway. As a result, glycosylation as a rule is heterogeneous in structure and therefore function. While this makes the study of glycosylation extremely challenging, it also makes this one of the most interesting processes in nature [494,543].

Glycan heterogeneity exists at two levels: macroheterogeneity refers to the presence or absence of glycans at putative glycosylation sites; microheterogeneity refers

to the different glycoforms that are present at a particular site in a population of glycoproteins. Both macro- and microheterogeneity are important in defining the interactions of the glycoconjugates. Together, the overall glycan composition and specific motifs present dictate the range of lectin domain-containing binding partners. Such interactions are especially important in tissue patterning, recruitment of hormones, signaling proteins, infection of a cell by pathogens and recognition of antigens for immune response.

Alterations in these glycan distributions can be correlated with deviations from the normal or healthy state to a disease state. The lack of template and proof-reading mechanisms renders glycan biosynthesis susceptible to environmental influences; glycosylation is therefore spatially and temporally regulated. Consequently, glycans serve as sensitive biomarkers for pathological conditions ranging from metabolic diseases to cancer. Due to this, the development of glycan and glycoprotein analysis methods has received much attention in the last decade. The following sections describe an overview of the *N*-linked glycan biosynthetic pathway and the roles of glycosylation in Influenza.

1.1.1 N-Glycan Biosynthesis

Glycans are added to proteins co-translationally, as nascent polypeptide chains are extruded into the ER lumen. Asparagine-linked *N*-glycans are attached to a sequon characterized by the presence of an Asn-Xxx-Ser/Thr (where Xxx ≠ Pro), also abbreviated as N-X-S/T. *N*-glycan biosynthesis begins by addition of a GlcNAc-P (Phosphate) from UDP-GlcNAc in the cytoplasm to dolichol-P, which generates Dol-P-P-GlcNAc. This structure is extended to Dol-P-P-GlcNAc2-Man5 and then flipped to the

luminal side of the ER by enzymes known as flippases. There, four more mannose and three glucose residues are added to the immature glycan to build up the final composition Glc3Man9GlcNAc2. The glucose residues are recognized by chaperones calnexin and calreticulin, and act as a handle during protein folding quality control in the ER. The terminal glucose residues and one to two mannose residues must be trimmed by α-glucosidases I and II and ER α-mannosidase I, respectively, before the folded glycoprotein can be transferred to the Golgi apparatus. As a result, most glycoproteins entering the Golgi have a glycan composition of Man7-8GlcNAc2 [5,60,96,119,134,494].

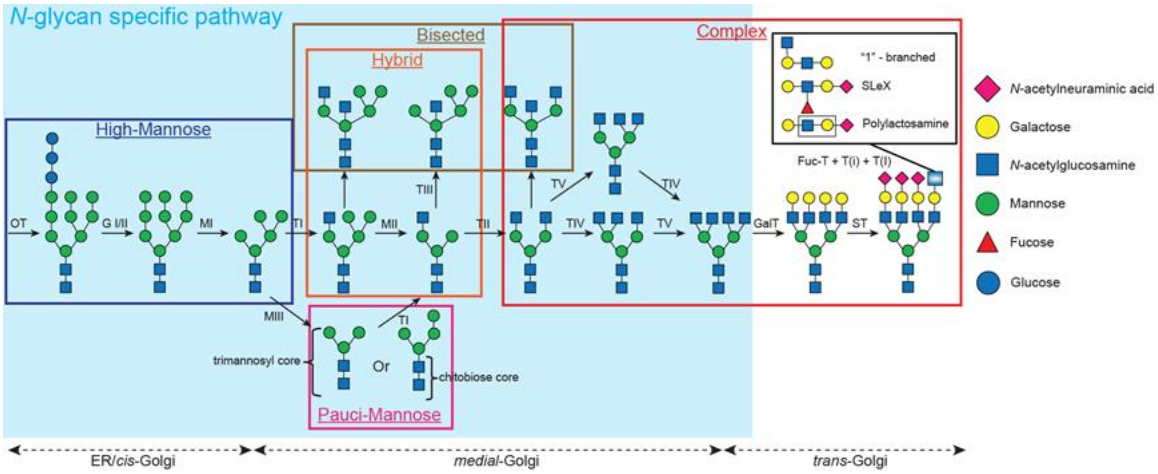


Figure 1: Schematic of N-linked glycan biosynthesis showing the Golgi compartments and the different classes of N-glycans: paucimannose-, high-mannose-, hybrid-, bisected-, and complex-type.

Abbreviations: oligosaccharyltransferase, OT; the α-glucosidases, GI, GII; the β-*N*-acetylglucosaminyltransferases, TI, TII, TIII, TIV, TV, T(i), T(I); the α1,2mannosidases, MI, α1,3/6mannosidases MII, MIII;β1,4-galactosyltransferases (Gal-T), α-fucosyltransferases (Fuc-T), α-sialyltransferases (ST). The terminal sequences are added to both *N*- and *O*-linked glycans by β1,4Gal-T; β1,3GlcNAc-T(i); β1,6GlcNAc-T(I); α2,3ST, α1,3Fuc-T. A number of substrate-product relationships govern *N*-glycan biosynthesis. Briefly, the GlcNAc-TI product is required for substrate recognition by α-mannosidase II and GlcNAc-TII, and necessary for biosynthesis of the complex-type glycans. GlcNAc-TIII substitutes the core β-mannose residue, re-directs the pathway into ‘‘bisected glycans,’’ and blocks any subsequent he action of α-mannosidase II and GlcNAc-TII. The lactosamine antenna initiated by GlcNAc-TV is preferentially substituted by β1,3GlcNAc-T(i), and elongated to produce polylactosamine (i.e.,Galβ1–4GlcNAcβ1–3 repeating units of 2 to .10 in length) and Lewis antigens. Polylactosamine synthesis is also affected by β1,3GlcNAc-T(i) activity, glycoprotein transit time in the trans-Golgi, and competition by chain-terminating enzymes including α1,2Fuc-Tand α2,6SA-T. Thus, control of polylactosamine and Lewis antigen expression appears to be, in large measure, a consequence of substrate preference, enzyme competition and tissue-specific regulation of glycosyltransferases. **Adapted from** [119]**; used with permission. Copyright © 1999 John Wiley & Sons, Inc.**

In the *cis*-Golgi apparatus, trimming of the mannose units produces a Man5GlcNAc2 structure for most proteins before they can be further processed to complex and hybrid-type *N*-glycans, as shown in Figure 1. Once in the *medial*-Golgi, the Man5GlcNAc2 structure may be further elaborated by GlcNAcT-I addition of a GlcNAc residue to the C-2 position of core α1-3-linked Man. This structure can be trimmed by α-mannosidase II to remove the terminal α1-3Man and α1-6Man, leaving behind a GlcNAcMan3GlcNAc2 structure. Upon removal of the two mannose units, another GlcNAc residue is added to the C-2 of the α1-6-linked Man in the core by GlcNAcT-II, resulting in the precursor structure for bi-antennary complex-type glycans. In the absence of α-mannosidase II activity, the α1-3Man and α1-6Man are left intact and are not further acted upon by the glycosylation machinery, leaving behind a Man5GlcNAc2 structure. However, in cases of incomplete α-mannosidase II activity, a Man4GlcNac2 structure is formed that acts as the precursor for hybrid type *N*-glycans. Paucimannose glycan structures (Man3-4GlcNAc2) can also be formed, usually as a result of GlcNAc residue removal from GlcNAcMan3-4GlcNAc2, by a Golgi hexosaminidase after α-mannosidase activity. Further branching of complex type *N*-glycans to tri- and tetra-antennary structures is carried out by GlcNAcT-IV and GlcNAcT-V, at the C-4 of the core mannose α1-3 and C-6 of core mannose α1-6, respectively. GlcNAcT-VI can initiate another branch at C-4 of the core-mannose α1-3 and a bisecting-GlcNAc residue can be attached to the β-mannose of the core by GlcNAcT-III.

After the branching is completed in the *medial*-Golgi, the elongation and addition of terminal glycan residues occurs in the *trans*-Golgi. The trimannosyl core can be

modified by addition of a fucose residue to the GlcNAc adjacent to the asparagine. FucT-VIII adds the fucose in α1-3 linkage (in plants), in α1-6 linkage (in vertebrates) and in α1-3 or α1-6 linkage (in invertebrates). Plants can have a β1-2 linked xylose present as a modification to the core.

Typical elongation of the branches involves addition of β-linked galactose to the GlcNAc, to form a Galβ1-3GlcNAc or Galβ1-4GlcNAc sequence, known as type-1 or type-2 *N*-acetyllactosamine sequences, respectively (commonly known as LacNAc). Sometimes, sequential additions of LacNAc repeats can form a polyLacNAc structure. A less common structure, resulting from the addition of a β-linked GalNAc to the terminal GlcNAc forms a LacdiNAc structure. *N*-Glycan branches typically terminate by addition of sialic acids by sialyltransferases, with the possibility of other modifications such as Fuc, Gal, GalNAc and sulfates to the branches. Sialic acids (NeuAc or NeuGc) are generally α-linked to the galactose units. While many other modifications and substituents have been reported, the above described features are most common in *N*-glycan biosynthesis and sufficient for the explanation of concepts spanning the scope of this thesis.

1.1.2 Influenza virus glycobiology

Influenza is responsible for an annual average of 36,000 deaths in the US alone and about ten times as many worldwide. The primary causative agent for influenza, the Influenza A virus (IAV), consists of eight single-stranded RNA segments that code for eleven proteins. Of these eleven, there are two coat surface glycoproteins; hemagglutinin (HA) and neuraminidase (NA), as shown in Figure 2. Hemagglutinin is a roughly bulb-

shaped trimeric protein with between 6 and 12 *N*-glycosylation sites that serves many important functions, including glycan mediated attachment to host cell-surface receptors, membrane fusion and internalization of the virus genetic payload for infection of the host cell. Most importantly, HA displays the immunodominant antigens on the virus. HA is almost 10 times more abundant than NA on the surface of IAV and is therefore the primary target for the host immune system. NA is important for cleaving sialic acid residues on the host cell surface to allow release of newly formed virions from the infected cell. HA and NA evolution is one of the mechanisms by which the virus responds to the selection pressures from the host immune system. The evolutionary process, also known as antigenic drift, involves changes in glycosylation at or near the antigenic sites on the surface of viral hemagglutinin. Because glycans can act as shields against the host antibodies, the IAV circulating in the human population evolves to gain *N*-linked glycosylation sequons around the antigenic sites on the HA head region. Genetic sequencing studies have shown that new strains entering a host population typically have a low sequon number on the HA head region; however, as they circulate seasonally, they acquire mutations in the protein backbone to change HA antigenicity by adding more glycosylation sequons [439,508]. On one hand, this glycosylation protects the HA protein from host antibodies [86,117,118,466,516]; on the other, increased high mannose *N*-glycosylation potentiates binding of lectins of the innate immune system, including surfactant Protein-D (SP-D) and mannose-Binding Lectin (MBL) that can neutralize IAV by binding HA glycans [81,109,545,564]. Thus, the virus strives to

maintain fitness by balancing addition or removal of glycosylation sequons, in the presence of selective pressures from the host adaptive and innate immune molecules.

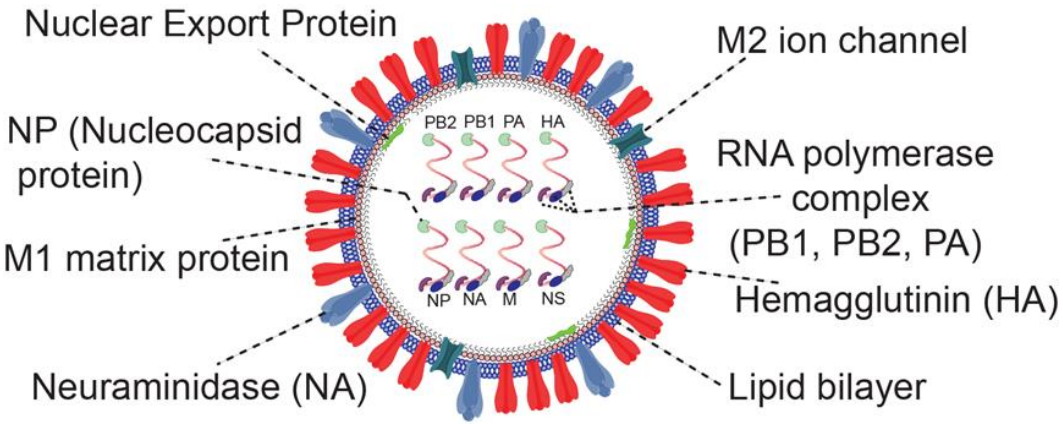


Figure 2: Model of the Influenza A virus

The innate immune system is the first line of defense against microbial pathogens. Collagen domain-containing C-type lectins (collectins) are soluble glycoproteins of the innate immune system that bind glycoconjugates on the surfaces of microbial pathogens. Collectin binding triggers microbial elimination through aggregation, complement activation, and/or activation of phagocytosis, the effects of which are to clear virus from the lungs [469,518,565]. The collectin surfactant protein-D is expressed constitutively by aveolar type II pneumocyte cells and non-ciliated bronchiolar cells (Clara cells) as a cruciform-shaped four-armed dodecameric glycoprotein, in which each arm represents a collagenous trimeric subunit that is covalently linked with three other subunits via the cysteine-rich N-terminal domain of the polypeptide [105]. Trimerization is initiated by a short 'neck' region of 33 amino acids that generates the formation of an α-helical coiled

coil domain and drives collagen trimerization further towards the N-terminus [243]. The C-terminus of each subunit clusters three calcium-dependent C-type lectin domains (Carbohydrate Recognition Domains, CRDs) that facilitate multivalent, high-affinity interactions with patterns of glycans expressed on the surface of pathogens including IAV. This categorizes SP-D as a β-inhibitor, a class of IAV inhibitors that share the ability to target viral carbohydrates in a calcium-dependent manner. SP-D molecules bind mannose residues of HA glycans, thereby acting as opsonins that modulate phagocytic neutrophils and macrophages against IAV [215]. IAV escaping from SP-D binding may be neutralized by antibodies in hosts previously exposed to a similar IAV strain, constituting an adaptive immune response.

It has been shown that porcine SP-D (pSP-D) possesses two unique structural features in its CRD that result in substantially stronger anti-IAV activity as compared to the antiviral activity of SP-Ds from other animal species, including human SP-D (hSP-D) [137]. The IAV HA has a sialic acid binding receptor that allows binding to sialylated host cell-surface receptor glycoproteins. This interaction is specific to sialic-acid linkage type and defines the host-specificity of an Influenza strain. Avian hosts predominantly display α-2,3 linked sialosides in their upper respiratory tract, while humans have α-2,6 linked sialosides in their upper respiratory tracts. This restricts the infection of these species by viruses that have a preference towards binding the respective sialylated glycans. This receptor specificity has been used by many research groups, as a tool for characterizing IAV strains and understanding virus-protein interactions [194,428,489,503,505]. Such studies are typically performed on glycan arrays where

fluorescently labeled virions or viral proteins are allowed to interact with a variety of immobilized glycans with defined structures. Porcine SP-D has a sialylated *N*-linked glycan in its CRD that facilitates a dual mechanism of interaction with IAV and that makes pSP-D unique amongst SP-Ds from different species in being a β-inhibitor as well as a γ-inhibitor against IAV infections. β-inhibitors are calcium-dependent lectins that bind carbohydrates on pathogen surface leading to aggregation and neutralization, while γ inhibitors (calcium-independent) have sialylated glycans that block the sialic acid binding receptor on HA, thereby preventing the interaction of the virus with host cell-surface receptors. Studies with pSP-D isolated from porcine bronchoalveolar lavage have suggested that this *N*-linked glycan is fully sialylated with α(2,6)-linked terminal SAs [538] and is responsible for the interactions between pSP-D and IA*Vs.* In addition, it was shown that the type of SA-linkage present on pSP-D is important for inhibition of IAV strains that have a corresponding receptor-binding specificity [538]. The adaptive changes that affect glycan binding specificity of a virus would therefore also affect its neutralization by a pSP-D.

Genetic sequencing studies can only predict the presence or absence of sequons, positions at which glycosylation could modulate the interactions of the virus with the host immune molecules. Such information has been used to model HA glycosylation by using a chitobiose core at the sequons [117,520,561,571]. However, crystallographic studies are unable to identify exact structural information pertaining to the carbohydrates linked to the protein backbone for both pSP-D and HA. This lack of glycan structural information prevents a systematic study of these interactions in a site-specific manner. Fortunately,

recent technological advancements in mass spectrometry have enabled site-specific glycosylation profiling. The information from mass spectrometry can be used for elucidating the specific interactions that mediate IAV neutralization by innate immune molecules like SP-D and how glycan macro- and micro-heterogeneity affects these interactions.

1.2 Glycan and Glycoconjugate analysis using mass spectrometry

Glycan and glycoprotein mass spectrometry includes a range of analytical methods that can be used to attain structural information from carbohydrates and glycoconjugates. There are many levels at which these biopolymers can be probed and each requires a specific set of methods to be used, ranging from the type of sample preparation and cleanup to the type of instrumentation required and the specific informatics tools that are necessary to deal with the data generated. While a complete review of the methods utilized for analysis of glycans and glycoconjugates would be impractical within the thesis, this section provides an overview of the most popular strategies in glycan and glycoprotein analysis. Figure 3 exhibits a generic flow of steps involved in mass spectrometry analysis and each step is elaborated upon, in the context of glycan and glycoconjugates samples, in the following sub-sections.

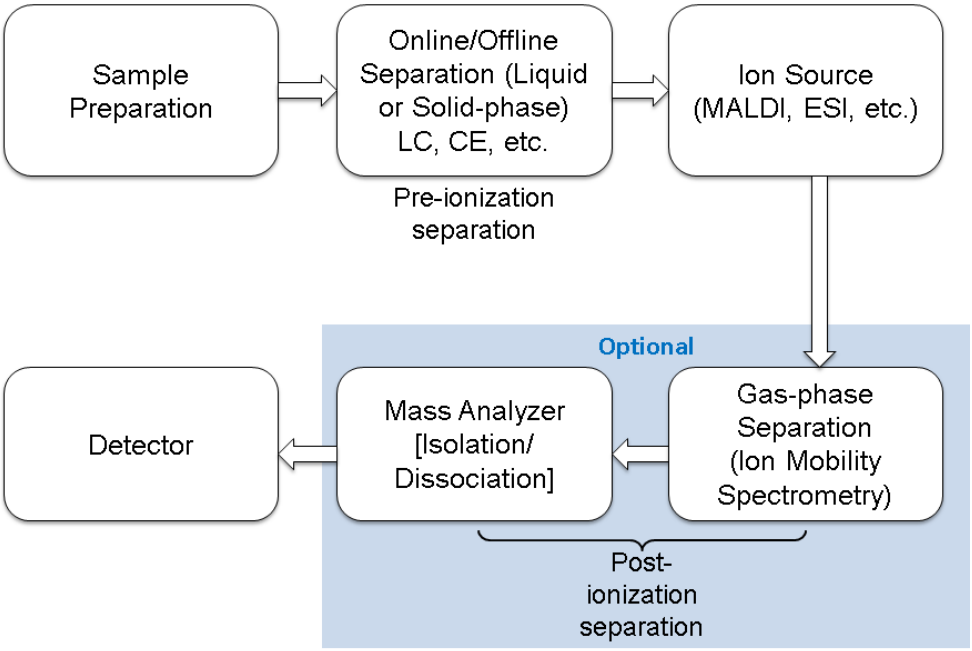


Figure 3: A workflow showing steps involved in mass spectral analysis combined with online and offline separation methods

(Courtesy of Prof. Catherine E. Costello)

1.2.1 Glycan sample preparation

Glycans attached to proteins are released either enzymatically or chemically, depending on the glycan type (linkage to protein) and the scale of desired experiment. *N*-Linked glycans can be released from asparagine residues on the protein backbones rather conveniently, using the enzyme peptidyl-*N*-glycosidase F (PNGase F), which typically releases all vertebrate *N*-glycans. Invertebrate and plant glycans may contain α-2,3 linked fucose on the core GlcNAc, a motif that prevents PNGase F release. PNGase*A*, on the other hand can be used to release all *N*-glycans, including those with a core fucose, provided that the protein is first proteolytically digested. There are several classes of *O*-glycosylation and no single glycoside acts on all; for many types, no glycosidase is

available. For mucin-type *O*-glycans, the *O*-glycanase enzyme releases Galβ1-3GalNAc, but not when this epitope is elaborated by monosaccharide substitution. As a result, mucin-type *O*-glycans are released by chemical means, most commonly by a reductive β-elimination reaction, first described by Carlson [70], which has been further modified and adapted by several groups [96,258,465]. Sodium borohydride is used as a reducing agent, to prevent the peeling reaction that can degrade the free reducing end generated by β-elimination, and converts the terminus to a glycan alditol. Glycosaminoglycans are also typically *O*-linked to proteins and are released by reductive β-elimination [95], after solubilization of tissue using guanidine-HCl or after digestion of the protein using proteases like pronase or papain. Another method for glycan release is hydrazinolysis, which causes complete release of both *N*- and *O*-linked glycans from proteins. This reaction can lead to loss of acetyl groups attached to the carbohydrate and therefore must be followed by a reacetylation step [310,311,416]. An advantage to the use of mild hydrazinolysis is the ability to completely regenerate the reducing end of the glycan by mild hydrolysis with mineral or Lewis acid, such as copper acetate [533]. The conditions used for release of the *O*-glycans degrade the protein/peptide backbone.

1.2.1.1 Glycan derivatization

Chemical derivatization of glycans is common for improving liquid chromatography and mass spectrometry analyses. Carbohydrates with a free reducing end present a reactive aldehyde functional group, which can be modified with a variety of linkers to improve analytical properties. Reducing-end tagging offers multiple advantages to glycan analysis, including increasing hydrophobicity for better reversed-phase

chromatographic retention, enabling optical detection, allowing quantification using tandem MS, ability to multiplex samples, etc. In addition, the available hydroxyl groups on the oligosaccharides can be modified by per-*O*-alkylation, such as permethylation, or per-*O*-esterification, e.g., peracetylation, to alter the liquid and gas-phase properties of glycans for enhanced sensitivity, enrichment, separation and detailed structural analysis by mass spectrometry [8,401]. Glycan labeling strategies have been discussed in detail in various reviews and research articles [279,383,384,450]. Figure 4, shows the most widely-used glycan derivatization schemes.

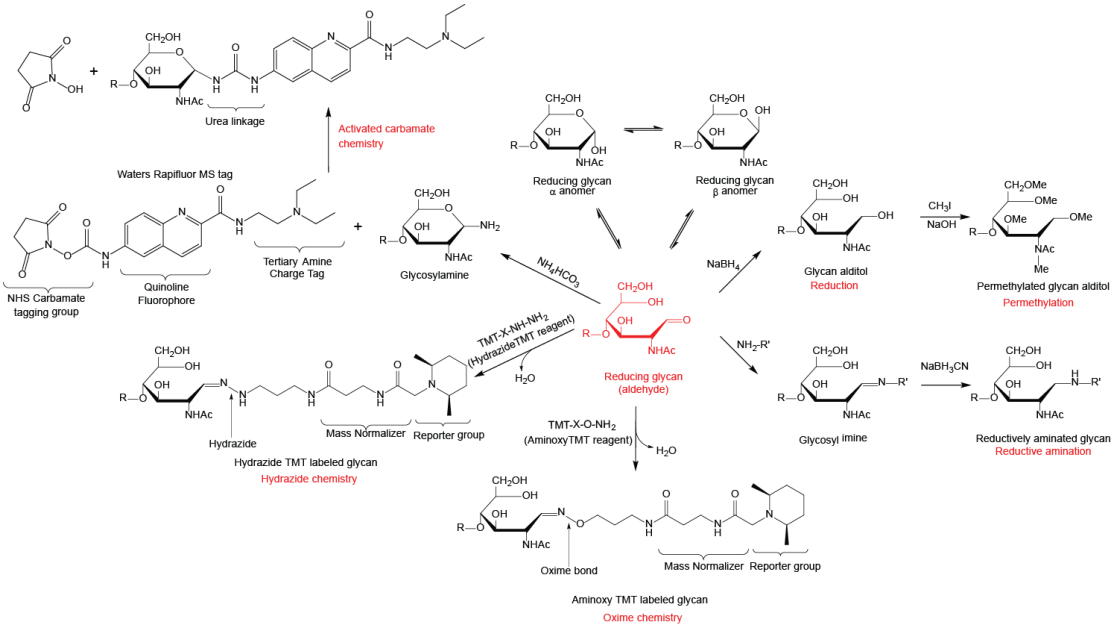


Figure 4: Commonly used glycan derivatization schemes

1.2.1.2 Reductive amination

Released glycans can be reductively aminated to add reducing end tags, where a primary amine group-containing tag reacts with the aldehyde on the free reducing end of the glycan to form a Schiff base or imine, which is then reduced in the presence of a

reducing agent such as sodium cyanoborohydride to a stable secondary amine. The most commonly used reagents for reductive amination are 2-AA (anthranilic acid), 2-AP (2-aminopyridine) and 2-AB (2-aminobenzamide), all of which are fluorophores that enable optical detection and quantification of the glycans using HPLC-based methods [25]. The labeled glycans can also be analyzed by mass spectrometry when coupled to online chromatographic separations [26,48,219,268]. The primary advantage of using these tags is the ability to get absolute quantification of glycan species by measuring optical response of the fluorescent tags against standards of known quantities. There are many variants of this approach that have been published by academic groups; some of these have been commercialized for glycan analysis. One of the first stable-isotope coded reductive amination tags for glycomics experiments was introduced by Bowman and Zaia [57,58]. These tags utilize mild reductive amination conditions and provide high labeling yields. Recently, isobaric aldehyde-reactive tags (iARTs) were used for reductive amination-based labeling, identification and relative quantification of *N*-glycans from the gp120 glycoprotein of human immunodeficiency virus [599].

1.2.1.3 Hydrazide labeling, oxazolone labeling, carbamate chemistry and aminoxy TMT labeling

Another tagging approach, described by Walker *et al.* [559], utilizes isotopically labeled hydrazide tags. These tags add a hydrophobic linker group to the glycan, which enables chromatographic separation of these glycans using reversed-phase chromatography. The tags also allow correction of any bias in quantification from isotope overlap by inclusion of internal standards labeled with light and heavy versions of the

tag. These tags are available commercially as the Individuality Normalization when Labeling with Isotopic Glycan Hydrazide Tags (INLIGHT) from Cambridge Isotope Laboratories, Tewksbury, MA. All of these tags add a hydrophobic group to the glycans, thereby also improving the ionization efficiency. Reaction of the carbonyl group of glycan reducing end aldehyde with an amine group results in formation of a hydrazone [313,558,600]. Conversion of glycans to phenylhydrazones has been shown to improve sensitivity in mass spectrometry-based analyses [315]. In addition to labeling the glycan for increasing hydrophobicity or with isobaric tags such as INLIGHT, hydrazide chemistry-based labels can also be used for glycan purification. One advantage of this chemistry is that it does not require addition of salts such as sodium borohydride, thus making it much cleaner and more adaptable to mass spectrometry based workflows. Lowering the pH allows hydrolysis of the hydrazine for facile release of the glycans. Hydrazide functionalized beads or biotin have been used in many studies for solid-phase glycan capture [597].

Cai *et al.* recently reported an oxazolone chemistry-based strategy for isotope-coded glycan tagging [64]. In this study, isotopically labeled amino acids were conjugated to glycans by reductive amination. The primary advantage here was the wide-availability of isotopically labeled amino acids, which makes this approach broadly accessible. Amine-reactive tandem mass tags (TMT) are widely used in proteomics analyses with great multiplexing and relative quantification capacity [94,364,447,456,526]. A carbonyl-reactive version of the TMT is available for tagging unreduced glycans. Hahne *et al.* compared aminoxy- and hydrazide-functionalized TMT

and concluded that the aminoxy-TMT outperformed the hydrazide version [206]. This reagent is also available commercially for glycan tagging and multiplexing for mass spectrometry analysis.

Another convenient glycan tag is the Rapifluor MS that has been developed by Waters Corporation. As shown in Figure 4, this tag combines an NHS-carbamate reactive group that reacts with *N*-glycosylamines with a quinoline fluorophore for optical detection and a tertiary amine group that improves MS response by enhancing ionization of the tagged glycans [316]. A similar derivatization chemistry has been used by Gong *et al.* for labeling glycans with isotopically coded TMT tags to enable multiplexing [184].

Shuang and coworkers have reported a quaternary amine-containing isobaric tag for quantitative glycomics (QUANTITY) [598]. While both the tertiary amine structure in aminoxy-TMT and the quaternary amine structure in QUANTITY fragment more easily during collisional dissociation compared to the glycosidic bonds, the quaternary amine is more labile than the tertiary amine structure and therefore QUANTITY generates a stronger reporter ion signal that can be used for relative quantification of multiplexed glycans.

1.2.1.4 Permethylation

Permethylation is the most widely used and arguably the most useful derivatization technique in glycan analysis. Native glycans are difficult to analyze by mass spectrometry owing to their hydrophilicity, which results in poor ionization efficiencies and MS sensitivity lower than the more hydrophobic compounds. The most widely available dissociation mode, CAD, typically yields glycosidic bond fragments for

native glycans and except in the negative mode analyses, where more extensive cross-ring fragmentation is observed and is also helpful in keeping sialic acids stable [98,129]. Permethylation converts the hydroxyl (-OH) groups on glycans to O-CH_3, thus making the glycans more hydrophobic and easier to desolvate. This reduces the bias from differences in ionization efficiencies for better quantification. Collisional dissociation of permethylated oligosaccharides has been shown to generate more cross-ring fragments compared to native glycans, thus improving the ability to assign linkages between different monosaccharides and also reduces gas-phase rearrangement of ions that is seen in tandem MS of native or reducing end labeled glycans [222,622]. Notably, gas-phase rearrangement may occur in high energy CAD of permethylated glycans and can be minimized in native glycans by metal-adduction and utilization of free radicals [120,323,586,587]. Permethylation also stabilizes sialic acid residues to enable positive mode analysis and tandem MS of sialylated glycans. Combined with isotopic labeling, permethylation can be used as an effective strategy to identify structural features like sulfation on glycans [320].

Isobaric labeling can be achieved by permethylation as shown by Atwood and colleagues [31]. The use of $^{13}CH_3I$ and $^{12}CH_2DI$, introduces the very small mass difference of 0.002922 Da, which generates derivatives that can be resolved when multiple sites are labeled. In addition to multiplexed analysis and quantification, permethylation also enables generation of useful fragments for structural analysis of the derivatized glycans. Kang *et al.* have described an approach for comparative glycomic profiling using quantitative permethylation and stable-isotope labeling [283]. The most

commonly used method for permethylation is a liquid-phase approach developed originally by Ciucanu and Kerek [90] and improved by Ciucanu and Costello [89], which involves incubation of glycans with sodium hydroxide dissolved in DMSO (dimethyl sulfoxide) and methyl iodide, followed by a liquid-liquid extraction of the derivatized oligosaccharides using chloroform and water. While this procedure is simple and well-established, solid-phase permethylation strategies have also been adopted in recent years for derivatization of glycans from small sample amounts . Solid-phase permethylation requires packaging of sodium hydroxide particles or beads into spin-columns or capillaries, followed by passing a solution of the glycan sample dissolved in DMSO and methyl iodide through the columns or capillaries. Residual salts from both liquid and solid-phase permethylation procedures can be removed easily, using reversed-phase spin columns or online with reversed-phase liquid chromatography by using a trapping column where the glycans can be washed and concentrated. Commonly, for glycans containing a free reducing end, the glycan sample is reduced prior to permethylation in order to open the reducing end monosaccharide to avoid chromatographic separation of the α and β anomers and to introduce a mass shift that facilitates discrimination of reducing and non-reducing end fragments. Glycans are reduced to oligosaccharide alditols using sodium cyanoborohydride or an ammonia-borane complex [7,162]. The symmetry between reducing and non-reducing end fragments and certain internal fragments can be further resolved by carrying out the reduction in the presence of isotopically labeled reagents like sodium borodeuteride [383]. A number of LC-MS/MS platforms have been developed for the analysis of permethylated glycan alditols [7,98].

1.2.1.5 Metabolic/enzymatic labeling of glycans to enable click-chemistry

Another approach for glycan labeling that has recently gained interest is the use of metabolic or enzymatic labeling followed by click-chemistry [448]. Both copper-catalyzed and copper-free azide-alkyne cycloadditions are relatively easy to perform and produce high yields. Monosaccharide analogs with azido functional groups can be used in cell culture for incorporation into glycans as they are being synthesized by the glycosylation machinery. Azido sugars are recognized by glycosylation enzymes and incorporated into glycans and glycoproteins. The azido sugars can then be probed using tags with alkyne groups. The efficiency of this method depends on the efficiency and specificity of the glycosylation machinery to accept the azido monosaccharides and then incorporate them into oligosaccharides in the ER and the Golgi apparatus.

Woo and colleagues recently developed a platform for targeted glycoproteomics using this metabolic labeling approach [578]. They used alkyne-functionalized isobaric tags with a cleavable biotin linker that enabled both the enrichment of the glycoproteins or glycopeptides and, after removal of the biotin linker, their subsequent analysis by LC-MS. Linkage-specific enzymes can also be used to probe specific motifs in glycans, by incorporating azido-labeled monosaccharides, as has been shown by Wu *et al.* [583]. When combined with powerful LC-MS workflows, these methods greatly improve the ability to identify features associated with glycans in normal and altered physiology and lead to understanding the biological pathways.

1.2.1.6 Sialic acid derivatization

Sialic acid residues are prone to dissociation during ionization, especially in MALDI-MS. In addition to glycan permethylation, sialic acids can be selectively modified to stabilize them in MALDI mass spectral analyses. Powell and Harvey have described a methyl esterification strategy for sialic acid stabilization [423]. This method is relatively straightforward and requires incubation of released glycans with DMSO and iodomethane. Another reagent, 3-Methyl-1-*p*-tolyltriazene, has been used for efficient solid-phase methyl-esterification of sialic acids [381]. More recently, the Wuhrer group have devised an approach where sialic acid esterification leads to stabilization of these labile monosaccharides and also enables discrimination of Gal-α-2,3 and Gal-α-2,6 linked neuraminic acids. This method utilizes 1-ethyl-3-(3-(dimethylamino)propyl)carbodiimide (EDC) and 1-hydroxybenzotriazole (HOBt) in presence of ethanol and ethyl-esterification or lactonization of α-2,6 or α-2,3 linked sialic acids, respectively. As shown in Figure 5, the mass shifts associated with these modifications allow easy identification and relative quantification of differently linked sialosides, in addition to stabilizing these glycans for positive mode MALDI-MS. Applications of this method for analysis of sialylated glycans will be discussed in subsequent chapters.

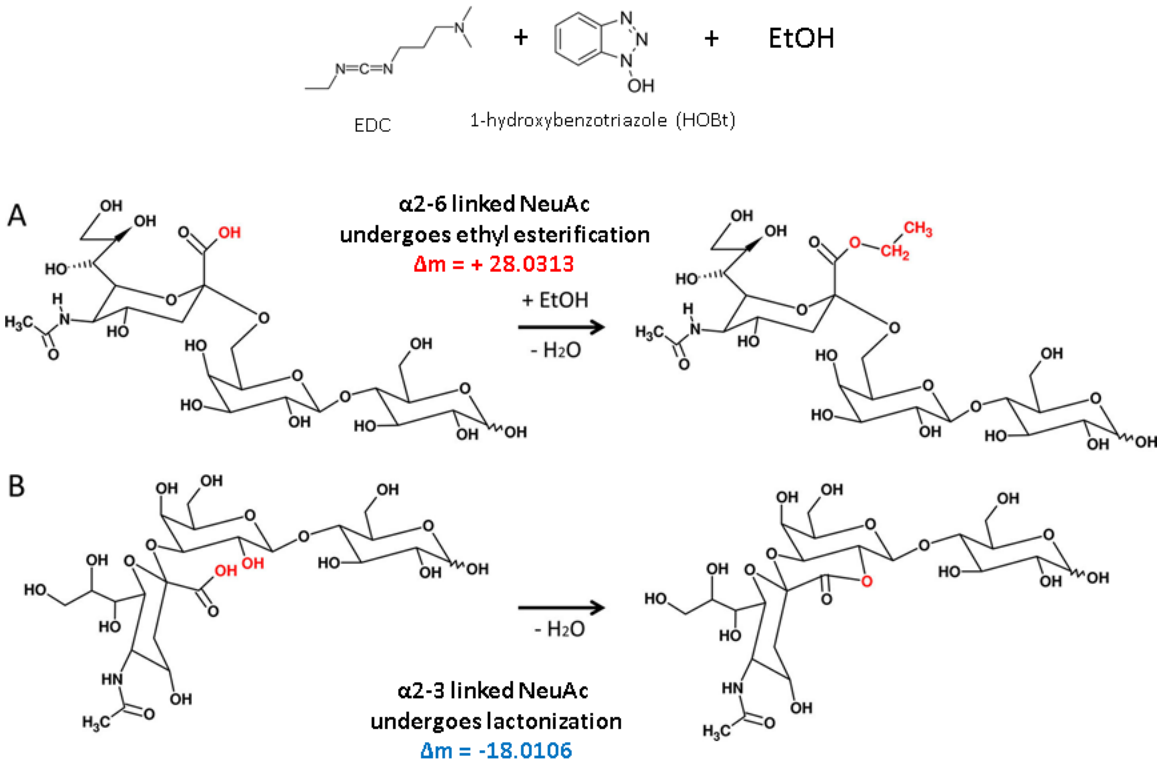


Figure 5: Chemical derivatization scheme for neutralization of carboxylic acid groups and differentiation of sialic acid linkages, as described by Reiding et al.[438]

Adapted from [438]**;used with permission. Copyright © 2014, American Chemical Society**

1.2.2 Glycopeptide sample preparation

Sample preparation for glycopeptide analysis or glycoproteomics is similar to that for proteomics. A bottom-up analysis entails reduction and alkylation of cysteine residues followed by proteolysis of proteins or glycoproteins and then analyzing the peptides and glycopeptides by LC-MS. Glycopeptides can be analyzed directly as part of a complex mixture of peptides and glycopeptides. The presence of a hydrophilic glycan hampers efficient desolvation and ionization of glycosylated peptides and results in lower response in mass spectrometry, compared to the non-glycosylated peptides. Additionally, glycan

heterogeneity divides the glycopeptide signal into many channels corresponding to multiple glycoforms. To prevent masking of the glycopeptide signal and obtain deeper coverage of glycan microheterogeneity, it is preferable to enrich or separate glycopeptides from non-glycosylated peptides prior to mass spectrometry. Glycopeptide enrichment can be performed using SPE or LC based methods that will be described in the following section.

1.2.3 Glycan and glycopeptide enrichment and pre-ionization separation

As the performance of MS instruments improves steadily both in speed and sensitivity, the depth and throughput in analyses is limited by front-end separations [481]. Advancement in separation methods is therefore imperative for bringing proteomics and glycomics to the same level of throughput as genomics analyses. Improved separations are necessitated by the heterogeneity in molecular phenotypes caused by glycosylation, where one gene product can have dozens of proteoforms. The range of molecular forms multiplies exponentially with the presence of glycosylation on a protein, because this modification brings more heterogeneity in terms of presence or absence at a site and the population of glycoforms that can occupy a site. Therefore to improve analysis by mass spectrometry, separation of differently modified or glycosylated forms, is necessary.

Liquid-chromatography is an indispensable separations method for analytical separation and fractionation of compounds in mixtures. A variety of stationary-phases enable orthogonal separation of mixtures. Depending on the solvent system required for mobile phases, the chromatography mode may or may not be compatible for separations

online with mass spectrometry. Owing to the hydrophilic nature, branching and the presence or absence of acidic substituents of glycans, their separations are most efficiently performed using HILIC (Hydrophilic Interaction Liquid Chromatography), ion-exchange or PGC (Porous Graphitized Carbon) modes [88,271,413,468,496,584]. Additionally, SEC (Size Exclusion Chromatography) can be used to resolve glycans based on their size [611,617]. While reversed-phase chromatography is not well-suited for the separation of native glycans, glycans labeled with hydrophobic linkers or permethylated glycans can be analyzed using RPLC (Reversed-phase Liquid Chromatography) modes, particularly with ion-pairing agents [7]. A comparison of different chromatography modes for glycan separation has been reported in literature [358,376].

1.2.3.1 HILIC

Hydrophilic interaction liquid chromatography takes advantage of polar stationary phases, including silica, amine- or hydroxy-based or zwitterionic resins and a mobile phase that contains a mixture of water and an organic solvent such as acetonitrile [11]. Ammonium salts are most commonly used to enhance resolution in MS-friendly mobile phases. The presence of polar groups on the surface of the column leads to creation of an aqueous layer, into which the hydrophilic analytes partition. A binary gradient starting with a higher concentration of the organic solvent is typically used for loading and adsorption of the hydrophilic analytes. An increasing aqueous composition in the mobile phase allows gradual elution of the glycans, in the order of increasing hydrophilicity – larger and highly charged glycans typically elute later than small and neutral

oligosaccharides [584]. HILIC works well for separation of hydrophilic compounds from hydrophobic compounds but generally offers lower resolution, broader peak shapes compared to other chromatography modes like PGC and reversed-phase, and is unable to resolve isomeric glycans, as shown in Figure 6. HILIC is therefore a preferred method for glycan and glycopeptide enrichment prior to RPLC-MS [16,68,293,321,390,412,468,523]. With online HILIC-MS of peptides and glycopeptides, non-glycosylated peptides elute earlier, while the glycosylated species elute later. Glycopeptide separation using HILIC is mostly glycan-centric and the resolution is much lower compared to reversed-phase methods [254,612]. Another disadvantage of using HILIC for glycopeptide LC-MS is that the presence of salts (typically ammonium) in the mobile phase causes formation of adducts during ionization, and these make the mass spectra more convoluted and decrease the sensitivity and speed of analysis, due to signal splitting and time wasted on data-dependent tandem MS of adducted species. Therefore, HILIC-enrichment using LC or solid-phase extraction (SPE) methods precedes glycopeptide analysis using reversed-phase LC-MS [293,468]. While HILIC-based methods still lag behind reversed-phase in terms of resolution, recently introduced zwitterionic HILIC stationary phases such as ZIC-HILIC offer much better resolution than Amide-based HILIC stationary-phases, so that the results are comparable to the resolution achieved on graphitized carbon columns [170,358]. With the advent of newer stationary phases, which contain both HILIC and weak anion exchange (WAX) properties such as the Thermo Scientific's GlycanPac™ AXH-1 columns, better resolution, mixed-mode separation of acidic oligosaccharides can be achieved. A similar

column (AXR-1) that combines reversed-phase and WAX characteristics can be used for separation for glycans with a hydrophobic tag or a linker or glycopeptides with acidic glycans. A reasonable concern with HILIC-based approaches is the solubility of the analyte in the high-organic starting conditions. Additionally, HILIC LC columns take longer to equilibrate, making the method less attractive for high-throughput analyses, whereas HILIC SPE is widely used in high-throughput workflows.

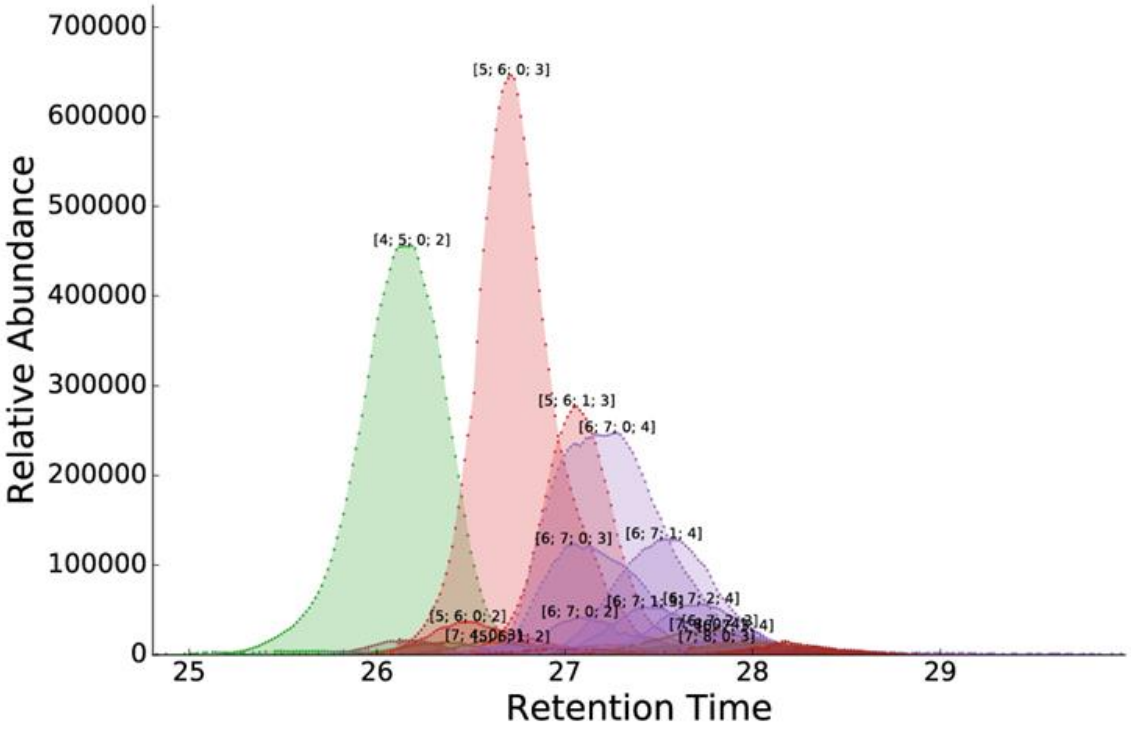


Figure 6: Overlaid extracted-ion chromatograms showing separation of released N-glycans from human α1-acid glycoprotein by HILIC-MS.

Glycan compositions are abbreviated as [HexNAc; Hex; dHex; NeuAc].

1.2.3.2 Porous Graphitized Carbon

A particular advantage of using PGC is its ability to resolve glycan isomers [286,287,408,409,609]. Because glycan retention on PGC stationary phases is a combination of hydrophobic, polar and ionic interactions [146,407], PGC has been

successfully used for separation of both native and permethylated oligosaccharides [98,130]. In addition to offline liquid chromatography separation and online LC-MS of glycans and glycopeptides, PGC is also commonly used in SPE cartridges for extraction, enrichment and desalting of oligosaccharides from complex samples [410]. Pabst *et al.* have performed a systematic study cataloging the retention times of isomeric glycan standards using PGC-MS and then populated a library for analysis of glycans in unknown mixtures [408]. The principles governing PGC separation of oligosaccharides and its applications have been reviewed in detail by Ruhaak and colleagues [451]. Hong *et al.* have described a PGC LC-MS workflow for glycopeptide quantification using a MRM (Multiple Reaction Monitoring) assay [242]. Alley *et al.* have compared the performance of RPLC and PGC for glycopeptide analysis and noted that while the performance of the two chromatography modes is equivalent, some hydrophilic glycopeptides are better retained on PGC compared to RPLC [10].

PGC LC methods can be conveniently coupled with UV detection of labeled glycans or online MS detection in the positive or negative mode. Molesworth and colleagues observed that it is important to keep the carbon column electrically isolated from the MS source to prevent polarization of the carbon material, which could prevent elution of acidic oligosaccharides [382]. After comparing multiple solvent systems for a range of oligosaccharides, an acidic eluent with pH 3.0 (30-65 mM ammonium formate) or lower was found to allow better retention and elution of all sialylated, sulfated or uronic acid containing oligosaccharides. However, for attaining a more sensitive analysis,

a higher pH and lower salt concentration, such as 10 mM ammonium carbonate, provided the highest sensitivity and elution power for a broad range of analytes.

PGC has been shown to have wide applicability for analysis of both *N*- and *O*-linked oligosaccharides, including sulfated and sialylated glycans [35,305]. Combining PGC with tandem MS allows simultaneous separation and characterization of glycan isomers, using diagnostic fragment ions [145]. This enables high-throughput characterization of glycan structures in complex mixtures of oligosaccharides, which can be extremely useful in identifying glycosylation changes associated with various biological processes and pathologies [6,143,452,499,527]. PGC has a few shortcomings in its susceptibility to salt buildup from samples that cause polarization of the stationary-phase and loss of reproducibility. The column must therefore be stripped periodically to remove the adsorbed material and recover performance. In addition, analyte recovery diminishes with increase in the number of acidic glycan substitutents.

1.2.3.3 Reversed-phase liquid chromatography (RPLC)

Reversed-phase chromatography is the most commonly used chromatography mode in biomolecule analysis. Reversed-phase stationary phases retain and resolve molecules based on their hydrophobicity by interaction or adsorption to alkyl chains of varying length for use with different compounds. The greatest advantage of using RPLC is the ease of coupling with mass spectrometers. Typical reversed-phase setups use MS friendly solvent systems for gradient elution of analytes. In addition to obviating the need for salts in the mobile-phase, RPLC also allows online desalting or washing of the samples thereby making resulting mass spectra cleaner and adduct-free. While RPLC is

the preferred mode for analysis of biopolymers like peptides, proteins and lipids, glycans are not retained well by reversed-phase columns due to their hydrophilic nature. However, addition of a hydrophobic tag or permethylation of glycans, as described in the previous sections, allows their retention and separation by RPLC [7,98,123,320,557,558,559]. Addition of ion-pairing agents such as trifluoroacetic acid improves the retention and separation of glycans and glycopeptides on reversed-phase columns but causes ion suppression [24,312]. Elevated temperatures have also been shown to improve chromatographic resolution of permethylated glycans on reversed-phase columns [628].

The above mentioned features make reversed-phase chromatography the most elegant method for glycopeptide analysis by LC-MS. Contrary to HILIC, where the separation is glycan-centric, glycopeptides are primarily retained and separated on RPLC columns by virtue of their peptide backbones. However, the glycan moieties also interact to an extent with either the un-bonded silica in the stationary-phase or in other cases where reversed-phase columns have polar modified stationary-phases, such as the Polaris™ stationary-phase from Agilent Technologies. This allows mixed-mode separation of glycopeptides, where glycopeptides with the same peptide backbone cluster together on the chromatogram and are separated to some extent based on their attached glycan moieties. The polar character allows longer retention of larger and sialylated glycoforms, as shown in Figure 7. Other mixed-mode stationary phases, such as Thermo Scientific's GlycanPac™ AXR-1, combine reversed-phase and WAX properties, thereby

allowing better glycoforms separation, particularly for sulfated and sialylated glycopeptides.

While RPLC offers unmatched resolution for glycopeptide analysis, enrichment of glycopeptides prior to LC-MS is important for improving depth of glycopeptide analysis. This is because non-glycosylated peptides desolvate more easily than glycosylated peptides upon ionization and produce a much stronger signal, which masks the signal from co-eluting glycopeptides. Additionally, the heterogeneity of glycoforms splits the signal from glycosylated peptides into many channels, effectively reducing the signal-to-noise of glycopeptide peaks. These factors prevent comprehensive glycopeptide analysis in the presence of unglycosylated peptides. HILIC and PGC are therefore typically used for offline glycopeptide enrichment prior to RPLC-MS analysis. We have also developed a workflow for online enrichment of glycosylated peptides, using HILIC, prior to reversed-phase separation coupled with tandem MS analysis [293].

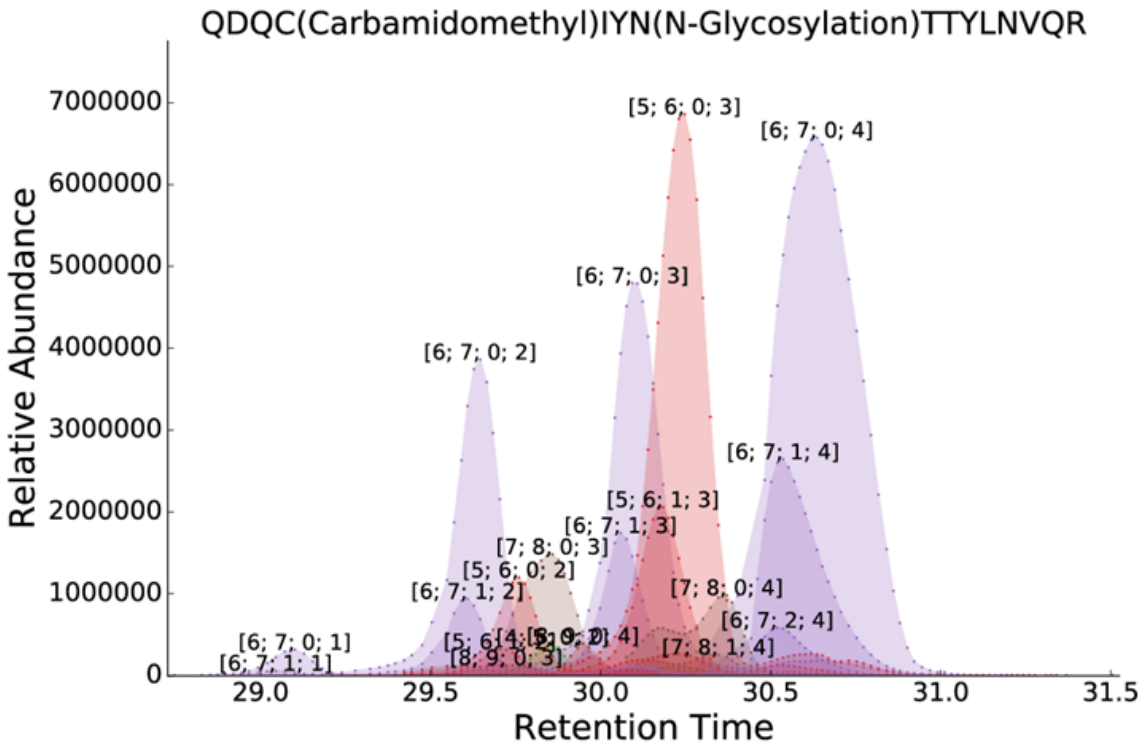


Figure 7: Overlaid extracted-ion chromatograms showing separation of the different glycoforms of a Human α1-acid glycopeptide QDQCIYNTTYLNVQR by C18 reversed-phase LC-MS

Glycan compositions are abbreviated as [HexNAc; Hex; dHex; NeuAc].

1.2.3.4 Size exclusion chromatography (SEC)

Size exclusion or gel filtration chromatography offers a mode of separation orthogonal to the above discussed modes, including PGC, HILIC and RPLC. Unlike HILIC, PGC or RPLC, in SEC analytes do not interact with the stationary phase based on their hydropathy. Instead molecules are resolved based on their size or molecular weight, typically under isocratic mobile phase delivery conditions. Larger analytes are unable to enter the pores or channels in the column packing material and are therefore excluded or elute earlier. The smaller analytes get trapped within the porous architecture of the SEC column and thus elute later. This allows facile separation of glycan polymers based on the degree of polymerization (dp). Isobaric saccharides with the same saccharide

composition can have different branching patterns, leading to multiple topologies, which cannot be efficiently resolved by SEC. However, linear polysaccharides such as GAGs (Glycosaminoglycans) typically consist of a heterogeneous mixture with high polydispersity. SEC becomes extremely useful for resolving the components in such mixtures because the components largely differ by size or the degree of polymerization. These separations can be performed online with mass spectrometry detection [228,608,609]. A major disadvantage of SEC is that it dilutes the sample, leading to lowered detection sensitivity. Therefore, SEC is sometimes performed offline using very long columns that offer higher resolution but dilute out the sample to levels lower than the detection limits of mass spectrometers. Development of UPLC SEC columns has now improved size-based separation both in terms of peak capacity and speed, such that very complex mixtures of linear polysaccharides can be resolved online with mass detection [611]. However, SEC uses a low salt content for minimizing interactions with stationary-phase and keeping analytes soluble by maintaining mobile-phase pH, these salts can cause ion-adduction, making the mass spectra difficult to analyze. We have demonstrated the application of an online ion-suppressor cell that exchanges cations in the mobile-phase with protons to prevent gas-phase adduction of analyte ions and generates cleaner, more interpretable spectra [611]. Sephadex (Dextran) based size exclusion columns, available from GE Healthcare™, are commonly used for desalting glycan preparations.

SEC can also be used for glycopeptide enrichment and separation from non-glycosylated peptides, because of the added size of the glycan on the peptide backbone [12]. While this is not the most efficient separation method and does not offer high

chromatographic resolution, it can resolve non-glycosylated, singly glycosylated and multiply glycosylated peptides. SEC can be interfaced with MS to ascertain the elution profiles of glycosylated peptides or glycans of different sizes and then offline fractionation can be performed, as shown in Figure 53, followed by RPLC-MS of the enriched fractions of interest.

1.2.3.5 Lectin Affinity chromatography

Lectin-affinity chromatography (LAC) is driven by carbohydrate-protein interactions [329]. Lectins bind oligosaccharides and thus glycoproteins/glycopeptides by varying degrees of specificity. This property can be used to enrich or separate glycans and glycoconjugates based on their saccharide compositions, structures or biologically relevant motifs [112]. The most commonly used lectins that offer broad specificities for glycan and glycopeptide enrichment are ConA (concanavalin A), which binds to alpha-mannose units and WGA (wheat germ agglutinin) that binds *N*-acetylglucosamine and to some extent sialic acid [112,160,280]. *O*-linked glycans and glycopeptides/glycoproteins can be selectively enriched using Jacalin, a lectin that binds the *O*-linked core disaccharide 1-β-galactopyranosyl-3-(α-2-acetamido-2-deoxygalactopyranoside) [245]. Lectins are most commonly immobilized on Sepharose for glycan/glycopeptide enrichment in SPE formats. While, lectins can be used individually for isolating a particular glycan or glycoconjugate population, for isolating all glycoforms from a complex biological sample, a multi-lectin affinity column (M-LAC) can be used where lectins with different glycan specificities are packed together [560,601,602]. Similar to other enrichment and analysis workflows, lectin affinity-enriched glycans and

glycopeptides are further desalted and analyzed using an appropriate LC-MS method [165]. Zhu *et al.* have compared the characteristics of various lectins for glycopeptide enrichment and fractionation combined with IMS-MS analyses [631,633]. Commercially available lectins are generally used for purification of common glycan compositions and motifs but in theory any biologically occurring lectin can be expressed or purified and then immobilized for mimicking biological interactions between biomolecules and capturing or resolving proteins of interest.

Other chromatography modes such as ion exchange that utilize high salt content cannot be directly interfaced with MS but can be used for orthogonal glycan and glycopeptide fractionation prior to LC-MS. A recent study by Totten *et al.* evaluates the performance of multiple chromatography methods for glycopeptide separation and enrichment [528]. The study found great value in performing electrostatic repulsion hydrophilic interaction liquid chromatography using strong anion exchange solid-phase extraction (SAX-ERLIC) over other methods for improving the depth in glycopeptide analysis. Capillary electrophoresis (CE) has recently gained popularity for glycan and glycopeptide analysis due to availability of improved MS interfaces and is discussed in Chapter 5.

1.2.4 Mass Spectrometry analysis of glycans and glycopeptides

Optical measurements are easy to combine with LC-based oligosaccharide and glycoconjugate separations but offer little insight into the composition and structure of the glycan or glycoconjugate and minimal capacity to differentiate co-eluting compounds. Accurate mass measurements and tandem mass spectrometry provide the

fields of glycomics and glycoproteomics with detailed compositional and structural analyses. While other analytical tools such as nuclear magnetic resonance (NMR) spectroscopy enable comprehensive structural analysis of biomolecules, the relatively low sensitivity of such methods and their inability to combine with online separation methods like nano-LC and CE, limits their use for analysis of complex biological samples. Mass spectrometry offers unparalleled sensitivity and specificity for unambiguous characterization of biomolecules. Tandem MS provides further resolution of isobaric analytes by dissociating them in the gas phase and then analyzing the fragments. Modern mass spectrometers are also able to combine gas-phase separation with mass detection, using Ion-Mobility Spectrometry (IMS), which takes advantage of differences in their cross-sections to resolve isomers and conformers that cannot be resolved by chromatography.

1.2.4.1 Sample introduction/ionization

Sample analysis by mass spectrometry requires ionization of the analytes. While many ionization methods exist for introduction of analytes into mass spectrometers, only electrospray Ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) are commonly used for analysis of oligosaccharides and glycopeptides [234,608,610]. A MALDI source typically employs a UV (e.g. nitrogen, 337 nm or Nd:YAG third harmonic, 355 nm) or an IR (Er:YAG, 2.94 μm) laser to liberate the analyte dried with a chemical matrix that absorbs at the laser wavelength. The energy from the laser pulse leads to photoexcitation and photoionization of the solid matrix and transfer of a proton to co-crystallized analyte molecules [221,230]. To prevent thermal disintegration of the

analyte, the laser is only pulsed for a short period of time which makes this ionization most suitable for use with a Time-of-Flight (TOF) mass analyzer. UV-MALDI ionization predominantly generates singly charged ions, with more highly charged ions observed in abundance only for large molecules such as intact proteins; multiply charged ions are common in IR-MALDI mass spectra. MALDI provides high spatial resolution in the plane of the sample stage and is relatively forgiving of contaminants in the sample. Thus, it is well suited for imaging mass spectrometry directly analysis of biological samples such as tissue sections. However, MALDI-Mass Spectrometry (MALDI-MS) cannot be combined with online separation methods such as liquid chromatography. Since analytes in a mixture are ionized together by MALDI, differences in ionization efficiencies and abundances of analytes favor sampling of molecules that ionize well and are abundant in a complex mixture. Pre-fractionation can help circumvent this limitation to a certain extent.

On-line LC-MS is the preferred method for analysis of complex mixtures of analytes, particularly those coming from biological matrices. Liquid chromatographs are interfaced with mass spectrometers using a continuous ionization source that allows introduction of analytes into the mass spectrometer as they elute from the column. ESI allows ionization of the sample from a continuous stream of analytes at atmospheric pressure. As analyte droplets emerge from the tip of a conductive sprayer or a capillary, an electric potential is applied between the capillary orifice and the mass spectrometer inlet, along with a drying gas that facilitates rapid evaporation of the solvent. As the solvent droplets evaporate, the charge concentration builds-up, leading to increased

Coulombic repulsions and division of the droplets carrying the analyte into smaller droplets and eventual release of the charged analyte into the gas phase [148]. While charged analyte molecules are drawn into the mass spectrometer under the influence of the electric field, the neutrals are not captured and solvent droplets dry off by application of a heated drying or curtain gas. In general, a combination of organic and aqueous solvents is used, which improves electrospray by decreasing surface tension, while maintaining the solvent conditions necessary for keeping the analyte soluble. For proteins and peptides, modifiers such as formic and acetic acid are also used in both static electrospray or as part of LC mobile phases to improve protonation.

Nano-ESI sources that operate at nanoliter flowrates provide much higher sensitivity and typically do not require application of a heated sheath gas to the source, owing to the low solvent volumes [570]. Unfortunately, these advantages to true nano-ESI are only observed for flow rates less than 50 nL/min. Because it remains impractical to interface on-line chromatography at such flowrates, sub-microliter per minute ESI sources operate as reduced scale conventional electrospray, not as true nano-ESI. Such devices have been integrated onto microfluidic separation devices such as the Agilent HPLC chip and the microfluidic-CE device (ZipChip) from 908 Devices [256,293,437,495]. Other convenient systems such as the Advion Triversa Nanomate are available for easy coupling of liquid chromatography to mass spectrometry using nano-ESI [291].

In contrast to MALDI, ESI generates abundant multiply charged ions, which allows analysis of biomolecules with higher neutral masses using a limited *m/z* analysis

range. The distribution of multiple charge states makes the mass spectra more complex than singly-charged mass spectra due to possible overlap of isotopic clusters and presence of multiple charge states. Both MALDI and ESI can be used in positive and negative modes for sample introduction but ESI is generally softer and therefore more appropriate for labile molecules such as oligosaccharides and glycopeptides. Even so, careful attention must be paid to the ESI source conditions to avoid loss of one or more labile sialic acid residues. Unless vibrational cooling is employed [265], MALDI can be reliably employed for sialylated glycans only after permethylation or use of other derivatization to stabilize the labile sialic acids [98,202,384,438],

1.2.4.2 MS analysis

Mass spectrometers consist of a series of ion-optical guides that direct beams or packets of ions towards a detector. Mass analyzers resolve analyte ions based on their *mass-to-charge* ratios (*m/z*). A variety of detectors are available on modern mass spectrometers that deliver varying detection or acquisition speed, mass resolution, mass accuracy, sensitivity and dynamic range. Irrespective of the mode of detection, all mass analyzers generate a spectrum with *mass-to-charge* (*m/z*) ratios of all ions detected.

The limits to acquisition speed or the rate at which mass spectra are acquired are a function of the physics of the mass analyzer. The realized mass resolution depends largely on the mode of operation of the mass spectrometer such as the choice of acquisition speed, acquisition *m/z* range, etc. and to a degree on the electronics involved in digitizing and converting the signal. The mass accuracy is measured as the deviation of

measured mass of an analyte from the theoretical mass of that analyte and is generally expressed in parts-per-million.

$$\textbf{Mass accuracy or mass error (ppm)} = \frac{(\textbf{measured } m/z - \textbf{ theoretical } m/z)}{\textbf{measured } m/z} \times 10^6 \qquad \textbf{Equation 1}$$

Mass resolving power is the ability of a mass analyzer to discriminate or separate masses that are very close to each other. Ion signals in mass spectra appear as normally distributed peaks. Mass resolving power is defined by IUPAC in terms of the overlap (or 'valley') between two peaks [641]. For two peaks of equal height, masses m1 and m2, when there is overlap between the two peaks to a stated percentage of either peak height (10% is recommended), then the resolving power is defined as:

$$\boldsymbol{R} = \frac{\boldsymbol{m1}}{\boldsymbol{m1 - m2}} = \frac{\boldsymbol{m}}{\Delta \boldsymbol{m}} = \frac{\boldsymbol{m}}{\Delta(\boldsymbol{m/z})} \qquad \textbf{Equation 2}$$

It should be noted that manufacturers' quotations of resolving power or resolution are often given with a 50% valley definition. For mass analyzers such as FT-ICR and Orbitraps, resolving power is a function of *m/z* value, so the *m/z* value must be included in the quotation of mass resolution.

The percentage overlap (or 'valley') concerned must always be stated. To calculate resolution from a single peak in a given spectrum, the peak width at half maximum is used as Δm.

Sensitivity and dynamic range of a mass analyzer define the lower limits of detection and the range of signal intensity that can be measured by the mass spectrometer. Sensitivity is the smallest amount of sample that can be used to generate a

signal discernible from the background noise and sometimes also referred to as the lowest limit of detection. This is different from the lowest limit of quantitation, which is the smallest amount of a sample needed to produce a robustly quantifiable signal. The sensitivity or limits of detection are typically defined in number of moles. Dynamic range on the other hand defines the difference between the highest and lowest analyte signal that can be measured together (usually in orders of magnitude of difference between the highest and lowest signals).

1.2.4.3 Types of mass analyzers

The most commonly used mass analyzers in biological mass spectrometry are described in brief below. Each mass analyzer has specific applicability and offers features for specific analyzer types as described in the following section.

Analyzer	**MS Acquisition speed**	**Resolving power (*m/z* 1000)**	**Mass accuracy**	**Upper limit of *m/z* range**	**Dynamic range**
Time-of-flight	50 Hz	20,000 (Linear) 40,000 (Reflectron)	< 5ppm	Unlimited	10^4
Quadrupole	10 Hz	2000	100 ppm	4,000	10^5
Orbitrap	18-22 Hz (for High-field Orbitraps)	50,000 – 10000 ***	1-5 ppm	3,000 (Up to *m/z* 8,000 on the Biopharma extended mass range instrument)	10^4
Ion-trap	50-60 Hz	4000	100 ppm	*m/z* 4000	10^3-10^4
FTICR-MS	10 Hz	500,000 *	< 1-2 ppm	3,000 (typical) Can be extended up to 5,000 **	10^4

Table 1: Comparison of different types of mass analyzers [127,235,349,354,462,562,577]

***varies with acquisition speed, transient length and magnetic field strength**
****While the analyzer can detect higher masses, the limit is imposed by transfer optics and the pressure in the ICR cell. Also, resolving power decreases significantly at higher *m/z*.**
***** Depends on the acquisition speed. Resolving power (R) decreases less significantly with increased *m/z* in Orbitrap compared to FTICR. For FTICR, R α 1/(*m/z*), while for Orbitrap R α 1/√ (*m/z*)**

1.2.4.3.1 Time-of-flight detector

As its name suggests, a time-of-flight or TOF mass analyzer measures the time taken by ions to fly through a "flight tube" of defined length and uses that information to measure the *m/z* value of the ions. The flight tube is a field free region into which the ions with the same charge state are injected with a constant kinetic energy. As a result, the

velocity of ions and their time-of-flight are a function of mass and charge, as described by the equations below [193].

$$\boldsymbol{Kinetic\ energy, KE = \frac{1}{2}mv^2 = qU = zeU}$$ **Equation 3**

$$v = \sqrt{\frac{2eU}{m/z}}$$ **Equation 4**

$$\boldsymbol{t = \frac{L}{v} = \frac{L}{\sqrt{\frac{2eU}{m/z}}}}$$ **Equation 5**

where, m = mass of the ion

v = velocity of the ion in the flight tube

z = number of charges on the ion

U= Ion acceleration potential

e = elementary charge (1.602×10^{-19} coulombs)

L = length of the flight tube

t = time of flight

A TOF instrument detects all ions pulsed into the analyzer at a given time. Thus, TOF analyzers are well suited for analysis of ions generated by the pulsed MALDI process. Electrospray generates a continuous beam of ions that must be compacted and pulsed into the TOF analyzer. TOF analyzers record all arriving signals and therefore sample a higher percent of the ions generated than quadrupole or ion-trap analyzers using

ESI. TOF instruments also have a duty-cycle advantage over trap instruments and can therefore reach MS acquisition speeds as high as 50 Hz on commercial instruments, which makes them the preferred analyzer for LC-MS and MALDI-MS. Resolution in TOF-MS is limited primarily by the length of the flight tube as well as by the temporal, spatial and kinetic energy spreads of the ion-packets, which can cause peak broadening. Reflectron TOF detectors can improve resolution by effectively increasing the length of the flight path, without having to increase the physical length of the flight tube, and more importantly by correcting the kinetic energy spread of ion packets. Additionally, delayed pulsed ion extraction can decrease the dispersion of ion packets at the detector, further improving the resolving power [62,350,544]. The detection system must be very fast (kHz) in high resolution TOF systems, in order to maintain the resolution provided by the analyzer.

1.2.4.3.2 Quadrupole mass analyzer or Quadrupole mass filter

Ideal quadrupole mass analyzers (Q) are made up of four parallel, hyperbolic rods assembled in a square configuration. In practice, cylindrical rods are often used. The opposite pairs of rods are connected electrically by a direct current (DC) potential with negative and positive polarities applied to the two different pairs. The two pairs of rods also have a radio frequency (RF) alternating current applied to them and are always 180 degrees out of phase. As ions travel along the center of the Q, parallel to the rods, the DC voltages destabilize the ions, while the application of RF maintains ions of certain *m/z* along the ion path [352,353]. The combination of RF and DC fields causes the ions to oscillate perpendicular to the length of the rods. The RF and DC current values can be

selected such that only ions of certain *m/z* can be stabilized and traverse the entire length of the quadrupole, while all other ions collide with the rods and remain undetected. The Q can therefore be used as a mass filter where specific potentials allow the transmission of ions of a selected *m/z* that can be detected by a downstream detector in hybrid instruments such as QQQ, Q-TOF, Q-Orbitrap and Q-FTICR. The range of *m/z* values that is allowed to pass through the Q can be very wide or very narrow. While the quadrupole mass analyzer suffers from low mass accuracy, low resolution and limited mass range, the combination of the quadrupole as a mass filter with higher resolution mass analyzers such as TOF, Orbitrap and FTICR enables high-resolution/accurate mass (HR/AM) measurements [142]. This also facilitates selective fragmentation of analyte molecules in the gas phase and analysis of its fragments by tandem MS, which is discussed in more detailed in later sections. Triple quadrupole mass analyzers (QQQ) are extremely useful in monitoring pairs of precursor and fragment ions with enhanced specificity and sensitivity, using multiple reaction monitoring (MRM) assays [20].

1.2.4.3.3 Ion trap mass analyzer

Ion traps are mass analyzers that use RF oscillating electric fields to store and analyze ions. Ion traps are categorized into 3-D quadrupole ion traps (QIT) or 2-D QIT or linear ion traps (LIT). The QIT operates with a 3-D quadrupole field by using a pair of opposing hyperbolic electrodes as end caps along with a ring electrode, as opposed to a quadrupole mass filter that uses four hyperbolic or cylindrical rods. QITs can have very high scan rates as the time to fill up the trap and acquire a spectrum is very short. However, the trapping efficiencies and dynamic range are limited due to the small

volume of the trapping region. This can also lead to poor spectral quality and mass accuracy from increased charge density [352,353]. The LIT achieves trapping by utilizing four sectioned quadrupole rods ending in lenses that trap ions radially by applying a RF potential between the rods and a DC potential at the end sections that confines the ions axially. Linear ion traps remedy the limitations of QIT by achieving higher ion storage capacity and minimizing space-charge effects. Mass measurements in ion-trap instruments are performed by ramping the RF field and measuring the amplitude that leads to ejection of ions based on their *m/z* [131,204]. While ion-trap instruments can achieve high scan rates, resolution depends heavily upon scan time and mass accuracy, even with increasing scan time, remains poor. The greatest advantage of using ion-trap instruments is that the ions can be stored and then excited inside the trap analyzers to achieve tandem MS and sequential MS/MS (MS^n). Furthermore, ion trap mass analyzers enable gas-phase ion-ion reactions that require analyte ions to be stored with reagent ions for varying periods of time. Such reactions and MS^n experiments cannot be performed in beam-type instruments. Ion traps can be used in combination with other mass analyzers such as Orbitrap or TOF to achieve high mass accuracy and resolution, following ion-ion reactions or in tandem MS [65,349].

1.2.4.3.4 Fourier Transform Ion Cyclotron Resonance (FTICR) Mass Analyzer

FTICR mass spectrometers are the most versatile instruments and provide the highest resolution and mass accuracy among the mass analyzers available today. FTICR-MS is based on ion cyclotron resonance theory, which was first developed and evaluated for studying the fundamental properties of an atom by Lawrence and Edlefsen [317] and

Sommer *et al.* were the first to apply this technology in ICR mass spectrometry [232]. Marshall and Comisarow were the first to integrate Fourier Transform in ICR spectrometry [92,93]. Since then, instrumentation development and applications in FTICR-MS have increased exponentially. This section provides an overview of the fundamental principles of FTICR-MS, which have been covered in great detail in articles by Marshall *et al.* and by Amster [354,13]

There are four primary components in all FTICR-MS instruments: a high-field-strength magnet; an analyzer cell; an ultra-high vacuum system; a data system for data acquisition and processing. The first component is a magnet, most commonly a superconducting magnet. The newer FTICR-MS instruments all have solenoidal superconducting magnets that offer stronger magnetic fields and enable achievement of higher performance. All magnets have a cylindrical bore to allow incorporation of the analyzer cell.

A cubic analyzer was the first to be developed and is still used today. This cell comprises of six plates in a cubic arrangement, such that one opposing pair of plates (trapping plates) is orthogonal to the direction of the magnetic field and the remaining two opposing pairs lie parallel to the field. The trapping plates or the trapping electrodes enable axial trapping of ions and have small openings in them to allow introduction of analyte ions and electrons into the cell, whereas the other two pairs of plates are used for ion excitation and ion detection. Another cell design known as the open-ended cylindrical cell also has six electrodes that perform similar to those in a cubic analyzer cell but form a cylinder-like shape. The two ends of the cylinder are the trapping electrodes that are

perpendicular to the magnetic field, with the principal axis of the cylinder aligning with the magnetic field. The core of the cylinder is divided into four electrodes that allow excitation and detection of charged particles. A dynamically harmonized FTICR cell with specially shaped electrodes was recently introduced by Nikolaev and coworkers [55,309,396]. This new cell design utilizes specially shaped electrodes for compensation of inhomogeneity of the magnetic field, which makes the cell field hyperbolic after being averaged along the cyclotron motion trajectories. This results in long durations of synchronous ion motion and higher resolving power than other cell types.

The ultra-high vacuum system is present to maintain an extremely low pressure in the instrument (the order of 10^{-9} to 10^{-10} Torr), which is critical to achieve high mass resolution. This is because collisions between any gaseous particles and the analyte ions can lead to rapid decay of the transient signal, resulting in a lower resolution. A sophisticated data acquisition and analysis system is the fourth feature of an FTICR-MS instrument that includes a frequency synthesizer, delay pulse generator, broadband R.F. amplifier and preamplifier, a fast transient digitizer and an electronic system to coordinate all of the devices during the data acquisition, as well as to process and analyze the data.

The ion motion in the analyzer cell is governed by a combination of electric and magnetic fields as described by the equation below.

$$F = qE + q(v \otimes B) \qquad \textbf{Equation 6}$$

The electric field *E*, allows trapping of the ions between the trapping plates, while the ion motion in a direction perpendicular to the electric field is guided by a Lorentz force *F*, generated by the uniform magnetic field *B*. As a result, the cyclotron motion is defined by the following equation in the absence of the axial trapping field:

$$F = ma = mv\omega = \frac{mv^2}{r} = q(v \otimes B) \qquad \textbf{Equation 7}$$

$$r = \frac{mv}{qb} \qquad \textbf{Equation 8}$$

Where the Lorentz force (*F*) acts upon an ion of mass *m* and charge *q*, inside a uniform magnetic field *B* to move it along a circular path of radius *r*, with velocity *v*. Angular frequency is defined by, $\boldsymbol{\omega = v/r}$ and cyclotron frequency $\boldsymbol{f_c = \omega/2\pi}$. Therefore, the cyclotron frequency can be calculated as:

$$f_c = \frac{qB}{2\pi \mathrm{m}} \qquad \textbf{Equation 9}$$

Thus, measuring the cyclotron frequency, in a uniform magnetic field allows calculation of the mass-to-charge ratio, *m/q* or *m/z*.

The electric field required for trapping of the ions inside the cell leads to a magnetron motion, which doesn't provide any useful information regarding the mass-to-charge ratio of the ions. Magnetron motion, represented by the equation below, is

undesirable because it causes radial diffusion of the ions and adversely affects mass accuracy.

$$f_m = \frac{\alpha V}{\pi a^2 \mathrm{B}}$$ **Equation 10**

where, f_m is the magnetron frequency, α is the geometry factor for the analyzer cell, a is the distance between the trapping electrodes and V is the trapping potential or electric field necessary to store the ions during analysis.

Ion detection is achieved by applying a broadband excitation or a range of frequencies to excite all ions in that cyclotron frequency range. This is done by applying a rapid frequency sweep, which allows all ions in that range to be excited into large cyclotron orbits of the same radius and different cyclotron frequencies. When multiple ions are detected simultaneously with a broadband sweep, the resulting image current is a composite of sinusoids of different frequencies and amplitudes. The frequency components of the signal are isolated by applying a Fourier transform to the time domain transient. A calibration formula derived from the cyclotron equation allows conversion of the frequency spectrum to a mass spectrum. While, any portion of the transient provides information on the same set of ions, the mass resolving power is directly proportional to the length of transient recorded. This is described by the equation below, where R is the resolving power and T is the duration of the transient:

$$R = \frac{f_c T}{2}$$ **Equation 11**

The ultra-high resolution and capabilities for electron-based activation of ions in FTICR instruments have led to their frequent application to biopolymer structural studies and have been key in the development of top-down proteomics and detailed glycan characterization strategies [76,257,467,605,622,637,638]. The different dissociation modes used in protein, glycan and glycoconjugate analysis are discussed in later sections. Unfortunately, the high resolving power and versatility in dissociation techniques in FTICR-MS comes at the expense of data acquisition time. This prevents efficient coupling of these instruments to LC or CE separations. Further, the presence of many components in hybrid FTICR-MS instruments and the selectivity of the ion optics can result in low sensitivity compared to other mass spectrometers. Some of these limitations have been overcome in Orbitrap Mass Spectrometers that combine high-resolution with high acquisition speed and sensitivity, and this has led to their wide adoption for analysis of complex biological samples.

1.2.4.3.5 Orbitrap Mass Analyzer

Orbital trapping was demonstrated by Kingdon in the early 1900s but the Orbitrap mass analyzer was developed into a commercial instrument by Alexander Makarov more recently by solving the problem of efficiently injecting ions into it using a C-trap [295]. While Orbitrap mass spectra are also generated via Fourier transform of the transient signal, Orbitrap analyzers trap ions electrostatically [348], (*i.e.*, without the requirement for a high-field magnet) making them considerably less expensive than FTICR analyzers. The Orbitrap analyzer consists of an outer barrel-like electrode and a coaxial inner spindle-shaped electrode. The outer electrode is separated into two equal halves separated

by a very fine gap. Ions orbit the inner electrode in elliptical trajectories and also move along the axis of the inner electrode in a harmonic motion. The ions are injected tangentially into the electric field between the two electrodes, which is ramped to compress the ions towards the inner electrode, until they reach the desired orbit inside the trap [250]. The ions move at different axial frequencies (ω) based on their *m/z*. The ion axial motion in the outer electrodes induces an image current, which is acquired as a time-domain transient and Fourier Transformed to produce a frequency spectrum. Frequencies are converted to *m/z* by the equation below [417,639].

$$\omega = \left(\frac{kz}{m}\right)^{\frac{1}{2}}$$ **Equation 12**

where, ω is the axial oscillation frequency, *m* and *z* are the mass and charge of the ion, respectively. *k* is the field curvature, determined by the exact shape of the electrodes and the applied potential. Orbitrap instruments have quickly gained popularity due to their superior performance in terms of speed and sensitivity as well as the high resolution and high mass accuracy achievable.

The commercial Orbitrap instruments are fitted with a curved RF-only quadrupole ion trap (C-trap), which compresses and injects ion packets for analysis by rapidly ramping RF and DC voltages across the C-trap. Hybrid Orbitrap instruments use either an LIT upstream of the Orbitrap for mass isolation and dissociation of precursor ions by CID or ETD or a quadrupole, in conjunction with a Higher energy collisional dissociation (HCD) cell for collisional dissociation [349], or both [474].

1.2.5 Tandem Mass Spectrometry

Tandem mass spectrometry entails isolation of a designated *m/z* range followed by ion activation, dissociation, and analysis of the resulting product ions [367]. The analyte ions or precursor ions can be measured intact first, followed by ion activation by one of the methods described in the following subsections, and mass analysis of the resulting fragment ions.

Peptide fragments follow a nomenclature adapted from what was first proposed by Roepstorff and Fohlman [444]. Fragment ions containing the N-terminus are named a-, b- or c-type based on the position, as shown in Figure 8. Similarly, product ions containing the C-terminus are named x-, y- or z-type. The numbering denotes the bond cleaved counting from the N or the C-terminus.

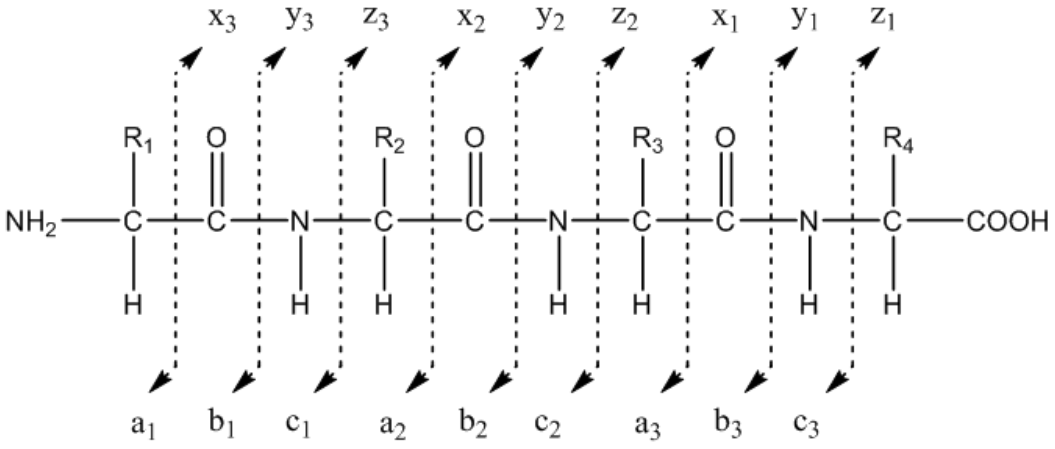


Figure 8: Nomenclature for peptide backbone fragments generated from gas phase ions

Similar to peptide fragmentation, a fragment nomenclature system for glycans was introduced by Domon and Costello [129]. Analogous to peptide fragmentation, fragments containing either end of the oligosaccharide (reducing or non-reducing end) are categorized as A-, B- and C-type (non-reducing end) or X-, Y- and Z-type (reducing

end) with a numbering system that indicates the position of the bond or residue at which cleavage occurred, as shown in Figure 9. While, B-, Y-, C- and Z-type ions are all generated by glycosidic bond fragmentation, A- and X-type ions are cross-ring fragments and therefore assigned with additional two numbers that denote the two bonds in a pyranose ring that were cleaved. The bonds are numbered from 0 to 5, starting with the O-C_1 bond. For branched oligosaccharides, Greek letters (α, β, γ) are used to annotate the branches or antennae. The largest branch is designated as α, the next largest as β and so on for subsequent branches in decreasing order of size. If primary branches are further branched, superscript primes are used with Greek letters to indicate the order of branching, such as α', β' and α'', β'' for primary and secondary branches, respectively.

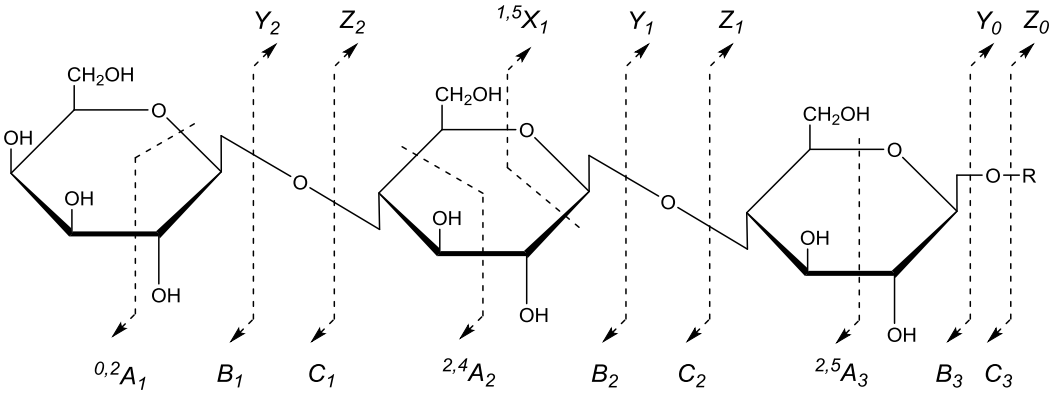


Figure 9: Domon and Costello nomenclature for glycan fragmentation

Reproduced from [129]**, with permission. Copyright © 1988, Glycoconjugate Journal**

1.2.5.1 Collisionally-activated dissociation or Collision-induced dissociation (CAD or CID)

Collisionally-activated dissociation activates precursor ions by exciting them in a dissociation cell that is filled with a neutral gas such as nitrogen, argon or helium.

Repeated collisions lead to vibrational excitation of the precursor ions and the bonds dissociate in the order of their decreasing fragility, leading to generation of fragment or product ions. CAD is the most commonly used dissociation mode on modern mass spectrometers. It allows quick and efficient fragmentation that enables fast spectral acquisition and deeper coverage in LC-MS/MS experiments compared to other dissociation modes. Collisional activation can be broadly categorized into resonant-excitation and beam-type. Resonant-excitation dissociation is achieved in ion trapping instruments, where product ions cool rapidly, thus limiting the extent of dissociation. Beam-type CAD product ions, on the other hand, undergo further collisions with gas, resulting in a higher extent of dissociation [261,365].

Collisional dissociation was the first dissociation method applied for the purposeof peptide sequencing and PTM identification by mass spectrometry; it is still widely used today for peptide and PTM identification (211–213). Collisional-dissociation of peptides primarily yields b- and y-type ions, which are useful for peptide sequencing. One disadvantage of using CAD is that any PTMs that are linked to the peptide backbone with a bond more fragile than the peptide amide bond (e.g. sulfation, phosphorylation, *O*-GlcNAc, etc.) may be lost in this process and only the unmodified peptide backbone can be determined from the tandem MS [415]. Another disadvantage is that, for large proteins, collisional activation only produces fragments near the N- and C-terminus while the central part of the protein remains poorly covered [76].

Glycan fragmentation by collisional-dissociation primarily yields glycosidic bond fragments, with cross-ring fragments in lower abundances [324]. More comprehensive

cross-ring fragmentation can be achieved by manipulating ion charge states, using metal ion adduction, or inducing high-energy (kV) dissociation, at the cost of higher analysis complexity.

1.2.5.2 Electron-activated dissociation (ExD)

Electron-activated dissociation is a broad term used to refer to a number of dissociation modes including electron capture dissociation (ECD), hot electron capture dissociation (hECD), electron transfer dissociation (ETD), negative electron transfer dissociation (NETD), electron detachment dissociation (EDD) and electronic excitation dissociation (EED) [33,97,378,488,589,636].

The ETD process combines ion-ion chemistry with tandem mass spectrometry [513]. In ETD, multiply charged analyte ions are reacted with radical anions, generated in a chemical ionization source, to transfer electrons to the precursor cations, and results in the formation of an odd-electron hypervalent species that dissociates by a non-ergodic process. ETD reactions are typically performed in analyzers that allow trapping of the precursor ions with reagent radical anions, such as those from anthracene and fluoranthene. Ion trap instruments only allow resonant-excitation of the precursors and suffer from an *m/z* range cutoff that limits acquisition in the low *m/z* range, depending on the RF trapping field magnitude. Newer instrument and dissociation cell designs now enable ETD and beam-type collisional dissociation to be performed in the same multipurpose dissociation cell [446]. This enables combination of collisional and electron-based dissociation for better sequence coverage on large proteins and glycans [161,331].

Negative electron transfer dissociation (NETD) is a variant of ETD used for negative mode analysis, where reagent cation radicals are reacted with precursor anions to induce dissociation. It is an extremely useful fragmentation mode for generating complementary information on analytes that do not ionize efficiently or fragment too extensively in the positive mode due to their acidic nature, including acidic glycans and ribonucleotides, as well as acidic peptides [257,366,442,453,576].

Electron capture dissociation (ECD) is achieved by directing a beam of low-energy electrons into a region containing multiply-charged analyte cations. While the exact mechanism of peptide fragmentation by ECD is still under debate, the overall reaction is driven by electron capture by the analyte polycations, followed by charge reduction and generation of an excited radical species that rapidly decays by bond cleavage to generate the fragment ions. Peptide fragmentation by ECD and ETD predominantly produces N-Cα bond cleavage. As a result, ECD and ETD primarily generate c- and z-type ions; some secondary fragmentation from side chain losses may be seen and are more common in ECD than ETD [553,637].

ETD and ECD generate more complete coverage for large proteins and informative cross-ring fragments for glycans [76,208,255,327,630,637]. Because these dissociation modes do not involve vibrational excitation, fragile PTMs often remain intact during ECD/ETD and their linkage site can be identified in the product ion spectra [370,380,567,637].

Increasing the electron energy during ECD fragmentation can lead to extensive secondary fragmentation, including amino acid side chain losses from z• ions and

formation of w ions. Additionally, b- and y-type ions may be seen. This type of fragmentation is termed hotECD (hECD) and has been shown to be useful in discriminating isomeric amino-acids such as norvaline from valine and leucine from isoleucine [298,299,568,606].

Further increasing the electron energy above what is needed for hECD can lead to a completely different dissociation process known as Electronic Excitation Dissociation (EED) [605]. For glycans, EED results from formation of a distonic ion due to electron detachment from an oxygen atom. This is followed by electron recapture to generate a diradical that rapidly undergoes fragmentation. The advantage of using EED is the generation of extensive fragmentation without charge-reduction, which minimizes the precursor charge state dependence that ETD and ECD suffer from. EED also produces significantly more structural information than ECD and ETD and is compatible with analysis of glycans ionized with a wide range of charge carriers.

Electron Detachment Dissociation (EDD) is the negative mode complement of ECD that utilizes higher-energy electrons (>10 eV). EDD is initiated by electron detachment from a multiply charged analyte anion, which produces an anionic radical that subsequently undergoes radical-induced fragmentation to generate a variety of product ions [63]. Similar to NETD, this fragmentation mode has been shown to be extremely useful in analysis of acidic analytes, including proteins, peptides and acidic glycans such as GAGs, without the loss of labile groups or side chains [3,492,574,575].

Until recently, most electron-based activation methods could only be implemented in FTICR mass spectrometers because a strong magnetic field is required to

contain the electrons for activation of precursor ions. The application and dissemination of these methods was therefore limited due to the expensive FTICR instrumentation whose relatively lower scan speed and lower sensitivity hindered their effective implementation during high-throughput LC-scale experiments; however, Voinov *et al.* have recently demonstrated electron activated dissociation of biomolecules on QQQ and Q-TOF instruments fitted with a radio-frequency-free electromagnetostatic cell [550,552,553].

1.2.5.3 Photodissociation

Photodissociation is an alternative method for ion activation and fragmentation, which offers higher selectivity than collisional and electron-based dissociation methods. This is because bond cleavages are specific to the wavelength of light used and the absorption spectrum of particular bonds. Two commonly used photodissociation modes are classified by the wavelength ranges they operate in: IRMPD (Infrared multiphoton dissociation) and UVPD (Ultraviolet photodissociation). While photodissociation is not widely available on commercial MS instruments, a number of academic research groups have used it for biopolymer characterization. IRMPD is often combined with CAD or ExD for supplemental activation, to achieve more extensive fragmentation. Photodissociation is typically implemented in instruments where ions can be trapped for some period of time, so that they can be exposed to photon irradiation. Without making major modifications to the mass spectrometer, a window is added to the instrument to allow introduction of a laser beam. A quartz window may be used for transmission of near UV light (e.g. 355 nm). UVPD (e.g., 193 nm) requires the use of a LiF or MgF_2

window, IRMPD by CO2 (10.6 μm) typically uses a ZnSe window. Photodissociation requires the analyte ion to have a suitable chromophore to allow photon absorption and ion excitation. Ly and Julian have summarized the use of UV lasers of different wavelengths in tandem MS [341]. In cases where the laser is not directly absorbed by the analyte, a chromophore may be attached covalently or non-covalently. IRMPD is typically used in conjunction with other dissociation modes such as ETD and facilitates better dissociation by pre-activating the ion. This involves unfolding of the precursor ion by irradiation leading to electron acceptor sites getting exposed for better ExD fragmentation. Many research groups have explored the use of UVPD and IRMPD in glycomics and proteomics, some of which are discussed in the following sections [124,125,303,318,344,442,446,480,492].

1.2.6 Ion-mobility-Mass Spectrometry

Ion-mobility spectrometry (IMS) is a method that provides an additional dimension of separation and structural information for gas-phase analytes. Ion-mobility techniques are broadly characterized into three different modes: time dispersive; space-dispersive and ion-confinement with selective release [361]. The different types of IMS techniques have been described in detail elsewhere [54,149,285,361,369,426,644]. IMS provides an extra dimension of separation for ions in complex biological mixtures and can be easily combined with liquid-chromatography and mass analysis. Combination of IMS with MS allows analysis of both molecular compositions and structural conformations in a single experiment, with tandem MS adding additional sequence and structure information to the already resolved conformations. This section reviews the

basic concepts and instrumentation used in IM-MS (Ion mobility-mass spectrometry) of biomolecules, while work focused on applications of IM-MS to biomolecular separations is discussed in later sections and in Chapter 5.

1.2.6.1 Drift-tube Ion Mobility Spectrometry (DTIMS)

A drift-tube ion mobility device is the most common IMS instrumentation, widely used in chemical detectors for security applications. In a DTIMS system, ions travel with constant velocities v_d under a uniform electric field E through a drift tube filled with a buffer gas.

$$v_d = KE \qquad \textbf{Equation 13}$$

$$K = \frac{v_d}{E} = \frac{L^2}{t_D V} \qquad \textbf{Equation 14}$$

Where K is the mobility of the ion. L is the length of the drift tube, t_D is the drift time of the ion, and V is the voltage drop across the drift tube. The mobility of an ion is often reported as its reduced mobility (K_0) at standard temperature and pressure, which can be calculated from K by:

$$K_0 = \frac{p}{760} \times \frac{273.15}{T} \times K \qquad \textbf{Equation 15}$$

Where p and T are the pressure and temperature of the buffer gas, respectively. Substituting for K,

$$K_0 = \frac{p}{760} \times \frac{273.15}{T} \times \frac{L^2}{t_D V}$$ **Equation 16**

The collision-cross-section (Ω), abbreviated as CCS, of an ion can be calculated from its measured drift time and mobility, as follows:

$$\Omega = \frac{(18\pi)^{\frac{1}{2}}}{16} \frac{ze}{(k_b T)^{\frac{1}{2}}} \left[\frac{1}{m_i} + \frac{1}{m_b}\right]^{\left(\frac{1}{2}\right)} \frac{t_D E}{L} \frac{760}{p} \frac{T}{273.2} \frac{1}{N}$$ **Equation 17**

Where z is the ion charge, N is the number density of the gas used, m_i is the mass of the ion, m_b is the mass of the buffer gas molecule and k_b is the Boltzmann constant.

The mobility resolving power R, for all temporally dispersive ion-mobility devices, is calculated as the drift time divided by the peak width at half-maxima (w):

$$R = \frac{t_D}{w}$$ **Equation 18**

Commercial DTIMS instruments deliver a resolving power of about 50-60, while certain custom-built instruments have been shown to have resolving power higher than 100 [172,176,285,515,535,591].

1.2.6.2 Traveling wave ion mobility spectrometry

A traveling wave ion mobility spectrometry (TWIMS) device comprises a series of electrodes that form a stacked ring ion guide. RF voltages with opposite phases are applied to the adjacent electrodes in the ion guide and this induces a wave pattern of voltages, known as a traveling wave (T-wave) [169,424,519]. The T-wave creates a

transient potential to move ions through the TWIMS device under static pressure of an inert buffer gas. While higher-mobility ions can keep up with the wave and therefore travel faster through the ion guide, lower-mobility ions fall over the apex of the traveling wave peak and are therefore slower to arrive at the end of the mobility device. In addition to the RF potential used to create the T-wave and push the ions through the ion guide, a second RF voltage component is applied to confine the ions radially. The TWIMS device is available on commercial mass spectrometers (Synapt) from Waters Corporation. While, the calibration of DTIMS is straightforward, the TWIMS device requires extensive calculations. Due to the absence of a physical model for TWIMS, the mobilities and CCS values of analytes must be extracted using empirical calibration performed using a set of analytes or calibrants, for which CCS values have been determined using DTIMS or another mobility device [484].

1.2.6.3 Trapped ion mobility spectrometry

Trapped ion mobility spectrometry (TIMS) is another technique for separating gas-phase ions based on their size-to-charge ratio[152,153,229,377,485]. This method utilizes an electric field to stabilize ions against the flow of a carrier gas that pushes ions towards the analyzer. A TIMS ion funnel consists of an entrance funnel, a mobility analyzer and an exit funnel. Radial confinement of ions is achieved by applying the same RF potential to the electrodes in all three regions of the TIMS device. The adjacent plates in the entrance and exit funnels have an RF potential applied to them, which is 180^0 out-of-phase. The RF phase between adjacent plates in the analyzer region is constantly alternating [175]. The entrance funnel traps and focuses ion packets that are periodically

pulsed into the analyzer region, where a laminar gas flow directs the ions from the entrance funnel towards the exit funnel. A scanning electric field is used to retard the motion of ions towards the exit funnel and, with the appropriate combination of electric field strength E and carrier gas velocity $v_{G,}$ the net axial velocity of an ion of mobility K becomes zero, thus the ion is trapped in the analyzer region:

$$v_{ion} = KE + v_G = 0 \qquad \textbf{Equation 19}$$

A gradual decrease in the electric field strength by reducing the voltage difference (ΔV) between the entrance and exit funnels allows sequential release of trapped ions of increasing mobilities [377,377,425]. Mass analyzers can be combined with TIMS devices to record spectra of ions as they sequentially elute from the mobility device.

A variation of the TIMS device, known as selected accumulation TIMS (SA-TIMS) has been recently introduced that allows selective storage or trapping of ions of desired mobility. In SA-TIMS, the axial electric field gradient is modified to include a steep ramp near the entrance that stops lower-mobility ions from entering the mobility analyzer, and a small barrier near the exit that does not retain higher-mobility ions; thus only ions with the mobility of interest are selectively accumulated in the TIMS analyzer. SA-TIMS allows combination of TIMS with mass analyzers and dissociation modes that require ion-trapping such as ExD on FTICR-MS [426].

1.2.6.4 High-field asymmetric waveform ion mobility spectrometry

DTIMS instruments use a low-field potential for mobility separations, where drift velocity is proportional to the electric field applied but the mobility is independent of the

electric field. When higher electric fields are applied, the drift velocity is no longer proportional to the electric field strength but the mobility becomes dependent on the electric field applied. High-field asymmetric waveform ion mobility spectrometry (FAIMS) utilizes the difference in mobility of an ion in the high-field and low-field limits as a mechanism for gas-phase separation of ions [191,421,441].

FAIMS uses an electric field applied between a pair of electrodes perpendicular to the flow of the carrier gas, such that it induces axial motion of the ions through the mobility device. One of the two electrodes is kept at ground potential and an asymmetric waveform is applied to the other. The asymmetric waveform consists of a shorter duration high-voltage component and a longer duration low-voltage component, of opposing polarities. The peak voltage of the asymmetric waveform is known as the dispersion voltage. For a majority of ions, the mobility in the high field is different from the mobility in the low field, resulting in a net transversal displacement from their original positions [191,192]. To eliminate this transversal displacement, a DC compensation voltage is applied. The ratio of high-field mobility to low-field mobility is unique for each ion in a mixture. As a result, a unique compensation voltage is required to transmit each ionic species in the FAIMS device. When the compensation voltage is set for transmission of one ion in a mixture, all other ions drift towards one of the electrodes and get neutralized, thus allowing the FAIMS device to act as a mobility filter. A compensation voltage scanning mode can be used to generate a FAIMS mobility spectrum, where ions of different mobilities get transmitted sequentially through the mobility device.

While other types of IMS instrumentation exists, the above described techniques are most commonly used for gas-phase separation of biomolecular ions. Our own work involving application of DTIMS to glycan and glycoconjugate separation is described in Chapter 5.

1.2.7 A review of mass spectrometry applications in glycan and glycopeptide analysis

Glycans and proteins are two different classes of biopolymers that have very different biochemical properties. While proteomics and glycomics present their own challenges in deep characterization, the level of complexity scales exponentially when these two biopolymers are combined to form glycoproteins with heterogeneous glycosylation. This section describes the state-of-the-art methodologies in achieving different levels of characterization of both released glycans and glycopeptides using mass spectrometry.

Comprehensive glycoproteomics analysis is many problems combined into one. This is because detailed characterization of the entire protein backbone and the decorating glycan(s) requires many different experiments. The methods to be used are dictated by the type and level of information that is desired to meet the needs in biological hypotheses - an overall glycomics profile may be sufficient in some cases, while detailed linkage information is necessary in others. Generally, the ability to characterize glycoproteins and glycans in a high-throughput manner decreases as deeper, more specific analyses become necessary. This decline in throughput is offset by

advancement in analytical workflows and improvement in instrumentation and informatics tools.

Biological hypotheses often focus on a specific set of biomolecules and their functions in normal and altered states. While omics level experiments measure the properties of a system at the level of all detectable proteins and glycans, the resulting information needs to be probed at varying levels of granularity for pinpointing the involvement of candidate biomolecules in health and disease. Naturally, more specific and detailed information can be obtained from a sample that has been enriched for a specific subset of proteins and glycoproteins based upon the hypothesis. Biochemical methods employed in fractionating, enriching and purifying glycoproteins from different matrices, cellular compartments or based on their physiochemical properties or affinity towards binding partners are generally well-established and previously reviewed [79,80,253,343,607,623].

Regardless of the method and stringency used in glycoprotein purification, contaminating proteins and glycoproteins are always present. It is therefore important to first establish the level of purity. With the assumption that proteomics workflows are more sensitive that glycoproteomics, a bottom-up proteomics experiment with or without deglycosylation helps account for all proteins and glycoproteins present in the sample, including targets and contaminants. Because glycosylated peptides are more difficult to analyze due to their poorer ionization efficiency and diverse glycoforms, deglycosylated proteomics enhances the ability to characterize the underlying peptide backbones. This is limited to *N*-linked glycopeptides, where the glycan can be removed using PNGase F,

without destroying the peptide backbone. The PNGase F release also induces a mass shift of 0.98 Da by converting the previously glycosylated asparagine to an aspartic acid. This mass shift can be used to account for the glycan site occupancy (macroheterogeneity) at that site [291,498,507,607]. The observed mass shift can be enhanced if the enzymatic deglycosylation is performed in the presence of $H_2{}^{18}O$, allowing incorporation of the heavier oxygen at the previous site of glycosylation [291,337,478,618]. Incorporation of an ^{18}O isotope also reduces false positives arising from *in-vivo* Asn deamidation. However, it is important to ensure that trypsin or other serine proteases have been inactivated prior to deglycosylation in presence of $H_2{}^{18}O$, to prevent artefactual labeling on peptide C-termini [23]. Also, spontaneous deamidation of asparagine residues in the presence of the heavier water can lead to labeling of the unmodified peptides, thereby confounding the data [291]. In addition to providing a measure of site-occupancy, proteomics of deglycosylated peptides also allows determination of all molecular forms of the peptide backbone present in the sample, along with any other PTMs present on the peptides apart from *N*-glycans [291,292]. For *O*-linked glycoproteins or glycopeptides, the level of information that a proteomics experiment yields may be limited to identification of all proteins present in the sample, since *O*-glycans cannot be removed without destroying the peptide backbone without using a series of enzymatic steps, and few relevant *O*-glycosidases are commercially available. Due to the multiplicity of molecular forms in a heterogeneous glycan population, it may not always be feasible to assign all peptide backbones and PTMs on glycosylated peptides.

Proteomic characterization is well-established and usually involves C18 reversed-phase LC-MS/MS with data-dependent acquisition (DDA). The mass spectrometer is programmed to select and fragment the most abundant analytes as they elute from the chromatograph [37,181,277,385]. The data are typically subjected to a proteomics database search for identification of proteins and peptidoforms present in the sample.

1.2.7.1 Glycan mass spectrometry

Once a biological sample has been characterized for proteins present and their respective proteoforms or peptidoforms, the total pool of glycans can be determined by analyzing released glycans. Glycans can be released from proteins as described in previous sections and further resolved by chromatography. Online LC-MS can be used to get an overall profile of glycan compositions using single stage MS. While native glycans may be analyzed in the positive mode, it is generally preferred that negative mode be used to prevent protonation and minimize loss of acidic moieties such as sialic acids and sulfate groups. LC-MS is the preferred method for glycan profiling. As described in earlier sections, reversed-phase chromatography can be used for glycan separation only when the glycans are peralkylated or are conjugated with a hydrophobic tag, frequently at the reducing end. Reversed-phase chromatography methods for LC-MS of derivatized glycans have been reviewed recently by Vreeker & Wuhrer [554]. The Wuhrer group has also developed numerous workflows for glycan profiling using MALDI-MS, which also involve linkage specific chemical derivatization and automation for high-throughput glycan profiling with sialic acid linkage discrimination [202,438,483]. They have also developed a data analysis pipeline for glycan assignment and relative quantitation from

MALDI-MS data and showed that relative quantification results from MALDI-MS and LC-MS are comparable [269].

The Mechref group has thoroughly explored the use of reversed phase LC-MS for glycomics and compared its performance with MALDI-MS for glycan analysis. They have demonstrated online trapping for desalting of permethylated glycans and use of higher temperatures for improving chromatographic resolution for permethylated glycan alditols [123,130,251,252,628]. They assert that while MALDI-MS is convenient and provides good coverage on glycoforms, it doesn't match the sensitivity of RPLC-ESI-MS for glycan profiling. Furthermore, they have shown that RPLC performed at elevated temperature produces better and narrower peak shapes and higher resolution, sufficient to separate glycan structural isomers, but their method could not separate linkage-specific isomers. Quantitative mass spectrometry-based methods for glycan profiling have been reviewed by Mechref [371]. These strategies involve labeling of released glycans using one of the methods described in the previous sections and then using either single-stage or tandem MS for label-free or label-based quantification.

While glycan profiling experiments are valuable in estimating the range of glycoforms and their relative abundances, the assignments often suffer from ambiguity in isobaric glycan compositions and structures. Further, false compositional assignments cannot be eliminated in the absence of glycan fragmentation. Tandem MS of oligosaccharides adds much value to glycan analysis both in terms of overall glycan profiling and confirming assignments, and detailed glycan structure analysis [15,323]. Permethylation of glycans enables LC-MS in the positive mode without any loss of acidic

residues. Tandem MS of permethylated glycans also yields more useful information for structural assignment compared to native glycans. Most recently, Zhou *et al.* have described the use of PGC LC-MS/MS for isomer characterization of permethylated *N*-glycan alditols [627]. They demonstrated the ability to resolve core and terminal fucosylation by chromatography and confirmed the assignments based on tandem MS.

PGC LC-MS is also used in many laboratories for analysis of native glycans. Everest-Dass *et al.* compiled a library of tandem mass spectra from glycan alditols analyzed using PGC-MS/MS on an ion-trap instrument using CAD [145]. They characterized over 200 *N*- and *O*-glycan structures generated by sequential treatment with exoglycosidases and used diagnostic ions for glycosidic and cross-ring fragments to discriminate isomers. Negative mode LC-MS is also commonly used in native glycan profiling experiments, particularly for acidic glycans including sialylated oligosaccharides and sulfated glycosaminoglycans [256,286,287,322,495,496]. While collisional activation in the positive mode generates primarily glycosidic bond cleavages, in the negative mode it can yield more cross-ring fragments that are informative towards structure elucidation. This is particularly the case when acidic groups are deprotonated or paired with a metal cation. Metal-adduction in the positive mode also enhances the ability to generate cross-ring fragments using collisional dissociation, compared to protonated oligosaccharides [220,238]. Ni and colleagues used ammonium fluoride in the mobile-phase for PGC LC-MS to form pseudomolecular oligosaccharide anions that yield useful product ions, including C-type glycosidic fragments, A-type cross-ring cleavages and gas-phase reaction products such as D- and E-type ions, which are useful for detailed

structural characterization of the glycan [395]. Additives to improve mass spectrometric ion responses and fragmentation patterns of oligosaccharides can also be added post-column continuously or pulsed into the electrospray to minimize salt build-up in the source [256,496]. Yu and colleagues have explored CAD on MALDI-Q-TOF and MALDI-TOF-TOF for permethylated glycan tandem MS and show that, while a Q-TOF instrument affords only glycosidic bond fragments generated by lower energy CID, TOF-TOF MS/MS provides high collisional energies to generate abundant cross-ring fragments for glycan sequencing [604]. Unfortunately, many of the more abundant cross-ring fragments are $^{1,5}X_n$ product ions which can be generated from any linkage isomer, and are therefore provide little structural information.

Another method for analyzing detailed glycan structure is sequential tandem MS (MS^n). The Reinhold group has thoroughly investigated the use of MS^n for glycan structure elucidation [28,29,30,273,273]. Sequential tandem MS can only be performed in ion-trap instruments where multi-stage dissociation can be performed efficiently on precursors and resulting fragments. Ion trap instruments are also extremely valuable for combining CID and ETD fragmentation in MS^n experiments to generate complementary information on the same precursors. Han and Costello analyzed the structure of milk oligosaccharides using CID, ETD and sequential MS/MS on an ion-trap instrument, even including sequential ETD and CID steps, in either order [208]. Their results showed an abundance of cross-ring fragments in ETD spectra and the fragment ion assignments were verified using MS^3 and stable isotope labelling at the reducing end, with or without reduction.

Although sequential tandem MS has its advantages in producing detailed structural information, the process is extremely low throughput and requires relatively large quantities of pure starting material for long periods of continuous infusion. This also hinders the coupling of the MS^n approach to chromatography and limits its use in omics-scale experiments. In addition, manual involvement and expert knowledge are necessary for selecting fragments appropriate for sequential tandem MS and the data acquisition and analysis are difficult to automate. In addition, there is concern in the community that product ion identification is ambiguous in the absence of high mass resolution and mass accuracy, although the use of stable isotopic labels and verification with parallel analyses performed on FTICR instruments can mitigate potential ambiguity.

Electron activated dissociation of glycans is greatly advantageous over single stage collisional activation for tandem MS. While ECD of large branched oligosaccharides can require supplemental activation such as IRMPD to yield efficient fragmentation, higher energy electrons in hECD can generate extensive glycan fragmentation without supplemental activation [2]. Zhao *et al.* have compared CAD and hECD of branched permethylated oligosaccharides on an ESI-FTICR mass spectrometer [622]. The study showed that the two fragmentation modes generate complementary structural information, in terms of the types of ions generated. While CAD generated abundant B- and Y-type glyosidic fragments, hECD generated predominantly C and Z ions. CAD also generated A-type cross ring fragments and hECD produced paired A and X ions that helped assign linkages positions. Devakumar and colleagues have also used ultraviolet photodissociation (UVPD) using a 157-nm laser to identify structural isomers

in permethylated glycans [124,125]. The authors showed that photofragmentation generates more A-type cross ring fragments compared to CID. Ko and Brodbelt have used a 193-nm UV laser for photoactivation of sialylated glycans in the negative mode to attain extensive A- and X-type cleavages for sialylated glycans [303]. They note that operation in the negative mode also alleviates the need to permethylate glycans [61].

Another method for producing rich glycan structural information is Electronic Excitation Dissociation (EED), which can yield extensive cross-ring fragmentation without any charge reduction [605]. Electron Detachment Dissociation (EDD) allows negative mode fragmentation of oligosaccharides, which has been shown to be more extensive than both CAD and IRMPD [3]. This has enabled negative mode fragmentation of acidic oligosaccharides such as GAGs, to generate detailed structural information [575]. In addition to EDD, NETD (Negative Electron Transfer Dissociation) has also shown promising results in attaining structurally informative fragments for highly sulfated glycans in the negative ion mode [257,576]. Overall, electron activated dissociation modes can provide comprehensive structural information that can be combined with chromatographic separations for better coverage. As ExD becomes available on faster, more sensitive instruments such as the systems being developed by Voinov *et al.* [550,551,552,553], online LC-MS methods with ExD will become more accessible to researchers in glycoanalytics.

Even with the best chromatographic resolution and most efficient tandem MS methodologies, it is difficult to characterize glycans that are present in biological samples. Glycans in nature are always heterogeneous in terms of compositions and fine

structure, including branching, positions of substituents and linkages. Structural isomers can be separated to a limited extent by chromatographic and offline purification methods and these co-eluting isobaric molecules generate chimeric spectra where the exact structural features remain unresolved. Ion-mobility spectrometry is becoming increasingly powerful in resolving isobaric molecules with structural and conformational differences. With the addition of ion-mobility devices to mass spectrometers, it has become possible to resolve isomeric glycans prior to MS and tandem MS measurements. The first commercial IMS-MS instrument, the Waters Synapt TWIMS-Q-IM-TOF, became available in 2006. Since then, newer generations of the Synapt instrument have been released and the latest addition to the family of commercial IM-Q-TOF instruments is the Agilent 6560 drift tube IMS (DTIMS) IM-QTOF.

In 2013, Pagel and Harvey measured the CCS (Collision Cross Section) values of many sodiated *N*-glycans using both nitrogen and helium on a TWIMS system [411]. They also performed in-source fragmentation and measured CCS values of both the native glycans and fragments thereof, to correlate trends in varying compositions/structures with CCS values. They were able to resolve structural isomers, of the same glycan composition (bi-antennary disialylated *N*-glycan), from fetuin, thyroglobulin and ovalbumin. Plasencia *et al.* have analyzed permethylated *N*-linked glycan structural isomers from Ovalbumin using a home-built modular DTIMS-TOF system [419]. Such data from glycan IMS studies are being populated into an online database, GlycoMob, cataloging CCS values for different glycans [506]. As the size of this database grows, these CCS values can be used for confirming glycan identities in

discovery glycomics experiments [237]. Ion-mobility separation has also been used for resolving linear glycosaminoglycan oligosaccharides that have different conformations due to differences in sulfation patterns [463,475]. Zheng *et al.* have studied the separation of glycan isomers adducted with metal ions in both positive and negative modes, using DTIMS [625]. Their results show that the presence of metal ions has a significant impact on glycan conformations and this helps improve gas-phase separations. Hofmann and colleagues recently performed relative quantification of configurational trisaccharide isomers in a mixture, using a TWIMS-Q-TOF system [236]. The Clemmer group has characterized isomers of the Man7 glycan from bovine ribonuclease B after permethylation, using a modified DT-IMS-LTQ instrument with a 1 meter long drift tube, where they were able to select mobility separated glycan isomers using a selection gate at the end of the drift tube [632]. Mobility selected ions were dissociated in the linear ion trap using collisional activation. Glaskin et al. have reported the incorporation of an ECD cell into the Agilent 6560 IM-QTOF MS and its use for oligosaccharide analysis [99,178]. Such combination of ion-mobility and tandem MS can be extremely useful for separation and identification of glycan structural isomers. Pu and co-workers have demonstrated the power of combining ExD with ion-mobility separation on a hybrid Qh-FTICR mass spectrometer, modified with a front-end selective accumulation-trapped ion mobility spectrometry (SA-TIMS) device [426]. SA-TIMS provides higher resolution than the drift-tube and traveling wave devices. Combined with high resolution mass spectrometry and extensive structural information from ExD, this method provides unambiguous assignment of glycan structures from a mixture of isomers. Together with

liquid and gas-phase separations, tandem MS of glycans has made a great impact in revealing structural complexities of glycans in biological systems and enabled large scale glycan profiling and comprehensive characterization studies.

1.2.7.2 Glycopeptide mass spectrometry

Complete structural elucidation of glycoproteins requires analysis of site-specific glycosylation to model the functions of glycosylated proteins and to understand their biological functions. Released glycan analysis by MS-based approaches yields useful information for compositional and structural assignments but does not provide information on the linkage of glycans to proteins. Addition of glycosylation sites to proteins enhances the heterogeneity of protein molecular forms exponentially, due to the added heterogeneity of glycan molecular forms. This adds considerable challenges in both acquisition and interpretation of site-specific glycosylation data.

The glycan micro- and macro-heterogeneity are primarily regulated by the glycosylation enzymatic machinery and its accessibility to the site(s) of glycosylation [4,210,291]. Since glycan synthesis is a non-template driven process, it is difficult to predict the exact compositions or structures present at each glycosylation site. Because each glycosylation site can have a variety of glycoforms decorating it, addition of glycosylation sites quickly balloons the range of possible glycoproteoforms from the combined heterogeneity at each site. This complexity is further exacerbated by the presence of other PTMs on the protein and the difficulties in studying glycosylation as a PTM, as described in the previous sections.

Typical proteomics experiments entail enrichment, sample cleanup, proteolysis and LC-MS analysis, followed by database searching for identification and then quantification, when necessary. Glycoproteomics differs from typical proteomics experiments in many ways, due to the inherent differences in the physiochemical properties of glycosylated peptides. The first step in all protein mass spectrometry studies is sample acquisition and preparation. This includes collection of biological samples such as tissue, biofluids or cultured cells and then appropriate treatment for analytical workflows. For a discovery or targeted proteomics experiment, a simple cleanup followed by proteolysis and desalting may suffice; a glycoproteomics experiment can be performed the same way but it is generally better to add an enrichment step for glycoconjugates to prevent masking of signal from non-glycosylated species. The enrichment methods used are typically selected based on the goals of the experiment. It is important to select an enrichment method that would minimize any bias in enrichment of different glycoforms, particularly if quantitative analyses are desirable. Typically, SPE methodologies are more convenient, robust and high-throughput for sample enrichment or fractionation compared to LC based fractionation, unless fine resolution of biomolecules over a long gradient is needed. Methods like HILIC provide the best overall enrichment of glycosylated proteins and peptides, without any major bias toward glycoforms. For more selective enrichment of particular glycoforms or glycan substructures/motifs, a lectin-based approach is preferable. Ion-exchange fractionation methods are very useful for isolating acidic glycans and glycopeptides (sulfated and sialylated species) from neutral glycoconjugates. In addition to offline glycopeptide

enrichment approaches, online enrichment of glycopeptides has also been demonstrated. We have designed a method for online HILIC enrichment of glycopeptides prior to reversed-phase separation and nanoESI on a microfluidics HPLC chip device from Agilent Technologies [293]. Reversed-phase separation remains the most efficient chromatography mode for glycopeptide LC-MS. Details of chromatographic separation methods have been discussed in the previous sections.

Confident glycopeptide identification using mass spectrometry requires a tandem MS step [122]. This is because the combinations of theoretical glycan and peptide molecular forms against which data are searched typically contain multiple ambiguous matches even when high resolution/accurate mass measurements are used. Furthermore, in the absence of tandem MS, it is not feasible to assign the exact locations of PTMs or to *de novo* sequence peptides that are not present in a theoretical search space. Before the development of hybrid mass spectrometers capable of delivering higher energy beam-type CAD, such as Q-TOF and Orbitrap instruments, collisional activation primarily generated glycan fragments and little or no peptide backbone coverage was present in the tandem mass spectra [579,585]. While this approach provides some information on the glycan composition and topology, in most cases, the peptide backbone has to be inferred from the residual mass after the glycan composition has been identified. This was still an error-prone process, since peptide backbone intact mass could have many more possible combinations given the number of amino acids and PTMs that can be combined together.

One method that allows sequencing of the peptide backbone using low-energy resonant CAD uses MS^3 of the glycopeptide Y_1 ion (341–343). Such experiments can

only be performed on ion trap and ion trap-Orbitrap instruments. The only other way to confirm a peptide match was to perform ETD to get peptide backbone fragmentation with the glycan still attached to the peptide backbone fragments [121,634]. ETD fragmentation is slower than CAD due to the reaction time required to initiate radical-mediated dissociation. While ETD performance has improved dramatically, particularly on newer generations of the Orbitrap instruments, it still lags behind CAD in terms of speed and sensitivity [248]. Furthermore, efficient ETD experiments can only be performed on the more expensive high-end hybrid Orbitrap instrumentation, thus limiting access among the research community.

Fortunately, in the last few years CAD has gained popularity for glycopeptide analysis. We and others have shown that peptide backbone sequencing is possible using higher-energy collisional-activation (CAD or HCD) of glycopeptides [18,293]. This has enabled confident glycopeptide analysis by single-stage CAD tandem MS, using low-cost Q-TOF and Q-Orbitrap instruments. Such methods can be easily disseminated for data-dependent acquisition of high-confidence glycoproteomics data [291]. While CAD works well for identification of singly-glycosylated glycopeptides, it has serious limitations in the ability to characterize multiply glycosylated peptides. Because glycosidic bonds are more labile than peptide bonds, vibrational energy provided by collisional activation results in complete obliteration of the glycans before the peptide backbone starts to fragment significantly. As a result, in higher energy CAD spectra where the peptide backbone can be sequenced, the glycan composition must be inferred from the residual mass of the precursor. This becomes challenging when analyzing multiply glycosylated

peptides for the following reasons. First, the abundances of peptide backbone dissociation ion for multiply glycosylated peptides are low with CAD. Second, the composition of individual glycan modifications remains unknown. Third, the exact location(s) of glycan attachment(s) remains ambiguous. Unless the glycosylation sites can be separated by the use of proteolytic enzymes, such glycopeptides can only be assigned confidently using ExD, most commonly ETD. The use of ExD for glycans and glycopeptides is discussed in Chapter 5. The Brodbelt group is also exploring the use of UVPD for glycopeptide analysis. They had earlier compared the use of CAD with UVPD for glycopeptide fragmentation, using a 193-nm laser in a trap instrument [303]. They showed that negative mode UVPD generated more extensive fragmentation for both the peptide and the glycan compared to positive mode UVPD or CAD in positive or negative modes. More recently, they have presented the negative mode UVPD results on acidic glycopeptides from kappa Casein glycoprotein and an *Acinetobacter baumannii* Ompa/MotB glycoprotein [344]. In both cases, UVPD produced abundant a- and x-type peptide backbone fragments that retained the labile glycan modifications. In addition, useful B/Y- and C/Z-type glycan fragments were observed, along with some cross-ring fragments that helped sequence the glycan attached.

Because glycopeptide glycoforms cannot be fully resolved using online LC-MS, ion mobility is increasingly being used in glycopeptide analysis. IMS adds the much needed peak capacity for resolving glycoforms and achieving a deeper coverage of the glycoproteome. Li and co-workers studied the correlation in FAIMS trendlines with different charge states of glycopeptides from α1-acid glycoprotein (AGP) and

Antithrombin III and compared them with non-glycosylated peptides in both positive and negative modes. They were able to detect isomeric mobility peaks for specific glycopeptides [325]. The glycosylated peptides were well separated from any non-glycosylated peptides. Such separation allows better identification of the glycoconjugates without interference from the more abundant peptides. Ion-mobility separation has also been applied to analysis of intact monoclonal antibodies and tryptic glycopeptides from the antibody. Olivova *et al.* analyzed the glycan heterogeneity in a monoclonal antibody by LC-ESI-IM-MS where the light and heavy chains were well-resolved in the gas phase. They also performed LC-ESI-IM-MS with two step-collisional dissociation of the glycopeptides to generate both glycan and peptide information [404]. Creese and Cooper have used FAIMS to separate and identify isomeric glycopeptides [104]. They analyzed two *O*-GalNAc linked glycopeptides with the same peptide backbones, only differing in the location of the *O*-GalNAc-linked using LC-FAIMS-MS. The glycopeptides co-eluted from the chromatograph but were partially separated in the gas-phase. The the locations of the GalNAc residues in the mobility-separated isomers were identified by tandem MS using ETD, which preserved the glycan modification during fragmentation. Glycoscience researchers are now starting to apply ion-mobility MS to yield information on biologically-relevant motifs. A recent study by Hinneburg and colleagues demonstrated the ability to resolve isomeric glycopeptides based on differences in *N*-acetylneuraminic acid linkages [231]. Our own work has focused on cataloging the collision-cross sections of a variety of glycans and glycopeptides to populate a database that can be used for

inferring the glycan compositions in large-scale glycoproteomics and glycomics studies employing IMS. The details of this work are presented in Chapter 5.

Improvement in glycoproteomics technologies has propelled biological studies in the last few years to the point where deep coverage of glycoform heterogeneity is now feasible. As the identification coverage increases, the focus is now shifting to being able to perform quantitative glycoproteomics. Recent studies have demonstrated relative quantitation of site-specific glycopepetide glycoforms [116,291,319,592,593,603]. Thaysen-Andersen *et al.* have summarized the recent glycoproteomics studies with the level of quantification achieved in a recent review [524]. We and others have described the available strategies and computational tools in both glycopeptide identification and quantification [113,248,249]. Data-dependent acquisition is the most commonly used acquisition method in LC-MS based proteomics and glycoproteomics, where precursors are isolated and fragmented based on their ion abundances. This method is limited by the duty cycle of the mass spectrometer in use and only samples the most abundant species. Additionally, the amount of time spent fragmenting each precursor is inconsistent because of variations in the ion abundances, chromatographic-peak shapes and changes in background. With increasing instrument speed, it is now possible to sample the entire mass range in small increments, without isolating specific precursor ions, by repeatedly cycling through sequential isolation windows – these methods are collectively referred to as Data Independent Acquisition (DIA) methods [164,333,414]. This has enabled better coverage of all molecular forms present in the proteome/glycoproteome. Most importantly, every peak gets sampled consistently, so there is no bias in tandem MS-

based quantitation. Liu *et al.* have compared the quantitative power of SWATH-MS (sequential window acquisition of all theoretical fragment-ion spectra) with SRM (selected reaction monitoring), which is regarded as the gold standard in quantitative mass spectrometry, for *N*-linked glycoprotein analysis [336]. SWATH measurements had slightly lower sensitivity for quantification but its reproducibility was comparable to SRM. However, a much larger number of peptides and glycopeptides could be quantified using SWATH-MS. SWATH and other DIA methods also obviate the need to create a targeted assay and yield information that can be used for both identification and quantification. The same group also applied the SWATH-MS acquisition method for glycoproteomics analysis of prostate cancer tissues and identified potential biomarkers for tumor aggressiveness [335]. Xu and colleagues have used SWATH-MS as a novel approach for *N*-glycosylation site occupancy measurement [595]. Sanda and Goldman have presented a workflow for DIA of IgG glycoforms from unfractionated human plasma samples [459]. They optimized CAD conditions for quantitative measurement of Y-ions resulting from saccharide unit losses and showed that this approach enhances sensitivity of label-free glycoform measurement in a complex sample.

Chapter 2: Development of LC-MS/MS methods for direct analysis of site-specific glycosylation in complex glycopeptide and peptide mixtures

(Part of this work is published in [293])

2.1 Introduction

Occurring through a series of regulated biosynthetic events in the endoplasmic reticulum and Golgi apparatus, the extent of glycosylation of mature proteins depends on numerous factors. Thus, the glycan structures assembled onto a given protein core are regulated according to cell type, developmental stage and spatial considerations. It is therefore not surprising that distinct glycoprotein glycosylation patterns characterize normal *versus* disease states [53,449,532]. As has been described in a recent report by the National Research Council [391], methods for study of glycoprotein structures are essential to progress in biomedicine. Many glycoproteins have several sites of glycosylation, each with a range of glycoforms. Glycoproteins pose analytical challenges based both on their chemical properties and the nature of the information produced. Thus, analytical methods used for proteomics and glycomics do not suffice for analysis of glycoproteins [323].

Mass spectrometry workflows for discovery and targeted proteomics are applicable to post-translational modifications (PTM) with defined molecular weight values, including phosphorylation, acetylation, methylation, ubiquitination and β-*O*-GlcNAcylation [87,264]. Such workflows have been applied to analysis of protein *N*-glycosylation site occupancy only following enzymatic release of glycans [296]. Thus,

because complex glycosylation is beyond the scope of traditional proteomics workflows, there is a paucity of information regarding the structures of *N*-glycans that occupy each protein site [522]. In order to address this problem, algorithms that allow users to define an unlimited number of peptide modifications have been developed [43].

In order to determine the site-specific glycosylation structure of mature glycoproteins, it is necessary to acquire accurate proteolytic peptide maps that include the range of glycan structures present at each glycosylated amino acid. In practice, accurate determination of the site specific glycan structures present on complex glycoproteins (those containing more than one glycosylation site) requires purified starting material. This is demonstrated by a recent interlaboratory study [321] in which prostate specific antigen, a glycoprotein with a single site *of N*-glycosylation was analyzed. Although there was a consensus in the interlaboratory results regarding the distribution of the most abundant *N*-glycan compositions present, it was clear that a more complex glycoprotein would pose serious challenges to investigators using analytical methods in current practice. The reason for this is that glycosylation multiplies the number of molecular forms of each glycosylated peptide by more than 10-fold. In addition, the most-often used reversed phase chromatography methods in proteomics workflows lack an enrichment step and therefore do not perform well for mapping glycopeptides from glycoproteins that contain several glycosylation sites.

Glycosylated proteins and peptides ionize poorly relative to non-glycosylated molecules. In order to produce confident identification of peptide backbone and glycan composition, it is desirable to analyze intact glycopeptides. Intact analysis eliminates the

possibility of false positive identifications that result from detection of deglycosylated peptides as surrogates for glycopeptides. Specifically, peptides that become deamidated, either through chemical or enzymatic processes, may be confused with formerly glycosylated peptides [413]. On the other hand, direct MS analysis of glycopeptides is made challenging by their aforementioned heterogeneity and poor ionization properties. Fortunately, hydrophilic interaction liquid chromatography (HILIC) has proven effective for enrichment of glycopeptides based on the hydrophilic glycan moiety [390,405,585]. However, these enrichment strategies, often in the form of solid-phase extraction (SPE) cartridges, add a number of steps to the sample preparation workflow. We conceived that it would be possible to integrate HILIC enrichment of glycopeptides directly to the reversed phase separation.

A glycoprotein may contain several sites of glycosylation, each of which is heterogeneous. As a consequence of glycoform diversity and signal suppression from nonglycosylated peptides that ionize more efficiently, typical reversed-phase LC-MS and bottom-up proteomics database searching workflows do not perform well for identification of site-specific glycosylation for complex glycoproteins. We present an LC-MS system for enrichment, separation, and analysis of glycopeptides from complex glycoproteins (>4 *N*-glycosylation sequons) in a single step. This system uses an online HILIC enrichment trap prior to reversed-phase C18-MS analysis. We demonstrated the effectiveness of the system using a set of glycoproteins including human transferrin (2 sequons), human α1-acid glycoprotein (5 sequons), and influenza A virus hemagglutinin (9 sequons). The online enrichment renders glycopeptides the most abundant ions

detected, thereby facilitating the generation of high-quality data-dependent tandem mass spectra. The tandem mass spectra exhibited product ions from both glycan and peptide backbone dissociation for a majority of the glycopeptides tested using collisionally activated dissociation that served to confidently assign site-specific glycosylation. We demonstrated the value of our system to define site-specific glycosylation using a hemagglutinin containing 9 *N*-glycosylation sequons from a single HILIC-C18-MS acquisition. HILIC has been directly interfaced to revered phase separation in previous reports for 2D separations [52,572]. In these cases, significant dilution of the fractions is necessary in order to ensure proper sample binding on the second dimension phase. This complication arises from the fact that HILIC separations typically begin in high percent organic mobile phase and reversed phased separations in high percent aqueous mobile phase. The design of a platform to automate HILIC enrichment followed by reversed phase separation is considerably simpler in architecture than a 2D HILIC-reversed phase separation.

We found that HILIC enrichment could be performed online with reversed phase separation so long as the enrichment column volume was kept sufficiently small compared to the separation column volume. Here, we present the use of on-line HILIC enrichment prior to C18 analytical separation with MS detection (HILIC-C18-MS). While we used a microfluidic chromatography chip to perform the experiments, such online enrichment and separation arrangement can be achieved using any HPLC system with appropriate valves for flow control and plumbing for low dead volume. Using this system, the most abundant ions detected corresponded to glycopeptides.

As has been pointed out by Desaire and Hua [122], it is difficult to assign glycopeptides based solely on MS data, even with the use of high resolution Fourier transform-MS instruments with high mass accuracy. As the number of glycosylation sites, variable modifications and peptide miscleavages increase, the usefulness of high-resolution MS data becomes limited in analyzing complex glycoproteins. A tandem mass spectrometry step therefore becomes necessary, in order to map glycopeptides with confidence. A recent review of software tools for glycopeptide analysis also highlights the importance of tandem fragmentation data for high-confidence glycopeptide assignments [113]. Thus, the need to improve quality of glycopeptide tandem mass spectra acquired during LC-MS (liquid chromatography-mass spectrometry) experiments also drove the present work. Beam-type collision-activated dissociation (CAD) of glycopeptides produces abundant oxonium ions, the *m/z* values of which serve as useful indicators of glycosylation [9,71]. Under typical CAD conditions, ions produced from dissociation of the peptide backbone are not detected [362]. As a result, separate electron activated dissociation (ExD) methods are often used to produce abundant peptide backbone dissociation while leaving the glycan intact [75]. The on-line HILIC-C18-MS system we describe increases the abundances of glycopeptide precursor ions and is in principle applicable to any tandem MS dissociation method.

Using elevated collision energies, we observed peptide backbone fragments from glycopeptides, in addition to oxonium ions and intact peptide ions with varying numbers of saccharide units attached (stub glycopeptides). An and colleagues [18], who used elevated collision energies to fragment all ions entering the mass spectrometer without

ion isolation/selection (known as MS^E) have also reported glycopeptide backbone fragmentation and others have reported similar glycopeptide backbone fragmentation using higher energy collisional (C-trap) dissociation (HCD) [468,471]. We demonstrate that HILIC-C18-MS enriches proteolytic glycopeptides from complex glycoproteins efficiently and thereby improves the ability to acquire data-dependent MS/MS on these molecules. As a result, the use of HILIC-C18-MS/MS generated significantly greater informational value than did C18-MS/MS. We demonstrated this method by analyzing glycoproteins including human transferrin and human α1 acid glycoprotein (AGP) and influenza A virus (IAV) hemagglutinin (HA). HA is a viral coat protein that is heavily glycosylated with the number of glycosylation sites varying between 5 and 12, depending on the strain. It binds sialylated glycoproteins present in host airway epithelia to facilitate infection and presents as a target for both innate and adaptive host response. HA glycosylation therefore represents an attractive target for researchers studying the pathogenicity of IAV [1,18,301,392,517].

2.2 Experimental Section

2.2.1 Materials and Methods

Human transferrin and human α1-acid glycoprotein were purchased from Sigma-Aldrich, St. Louis, MO. Recombinant HA1 from *A/USSR/90/1977 (H1N1)* (amino acid 18-345) (Genebank #: ABD60933), was purchased from Immune Technology Corp., New York. Sequencing grade trypsin was purchased from Promega Corp., Madison, WI.

LC-MS grade solvents were purchased from Fisher Scientific, Pittsburgh, PA. All other chemicals and reagents were purchased from Sigma-Aldrich, unless otherwise stated.

Glycoproteins were subjected to reduction (dithiothreitol), alkylation (iodoacetamide) and overnight tryptic digestion in the presence of 2,2,2-trifluoroethanol (5%) as denaturant, and ammonium bicarbonate (100 mM) as pH buffer. Reactions were stopped by heating at 95^0 C for 5 min. The masses of resulting peptides were analyzed using a Bruker Ultraflextreme MALDI MS (Bruker Daltonics, Billerica, MA) using 2,5-dihydroxybenzoic acid (DHB) as matrix. The samples were then dried down in a centrifugal evaporator.

Dry tryptic digests were reconstituted in a 1:1 mixture of water and acetonitrile with 0.1% formic acid to a 1 mg/mL concentration. The samples were then diluted in the appropriate loading conditions for C18 or HILIC-C18 analysis to approximately 100 fmol/µL and injected into the LC/MS system using the Agilent 1200 series autosampler. All analyses were performed on an Agilent 1200 series LC system capable of delivering nanoliter flow rates coupled to an Agilent 6550 ion-funnel quadrupole time-of-flight (Q-TOF) mass spectrometer, equipped with a chip-cube ionization source (Agilent Technologies, Santa Clara, CA), in the positive ion mode.

Approximately, 1 µg and 0.1 µg of the total proteolytic digests were injected without any cleanup for analyses on HILIC-C18 or C18 chip-LC/MS, respectively. Lower amounts of total material were loaded onto the C18 chip because in the absence of enrichment, the non-glycosylated peptides led to signal saturation. The Agilent HPLC chip consisted of laser ablated channels that were packed with appropriate stationary

phases. The C18 chip used was a commercially available Polaris-HR-Chip-3C18 chip (Agilent Technologies, Santa Clara, CA) with a 75µm X 150mm separation/analytical column and a 360nL enrichment column, both packed with 3µm particle size and 180Å pore size, Agilent Polaris C18-A stationary phase. The HILIC-C18 hybrid chip consisted of a 75µm X 150mm analytical column packed with 5µm particle size and Polaris C18-A stationary phase and a 40nL enrichment column, packed with Amide-80 stationary phase (Tosoh Biosciences, Montgomeryville, PA), 5µm particle size and 80Å pore size. This chip is available from the vendor as a custom order. A binary pump with microliter flow rates was used as the "loading pump" to deliver sample and solvent to the trapping/enrichment column and a nanoscale pump was used as the "gradient pump" to deliver the gradient once the trapping and analytical columns were inline, as shown in Figure 10.

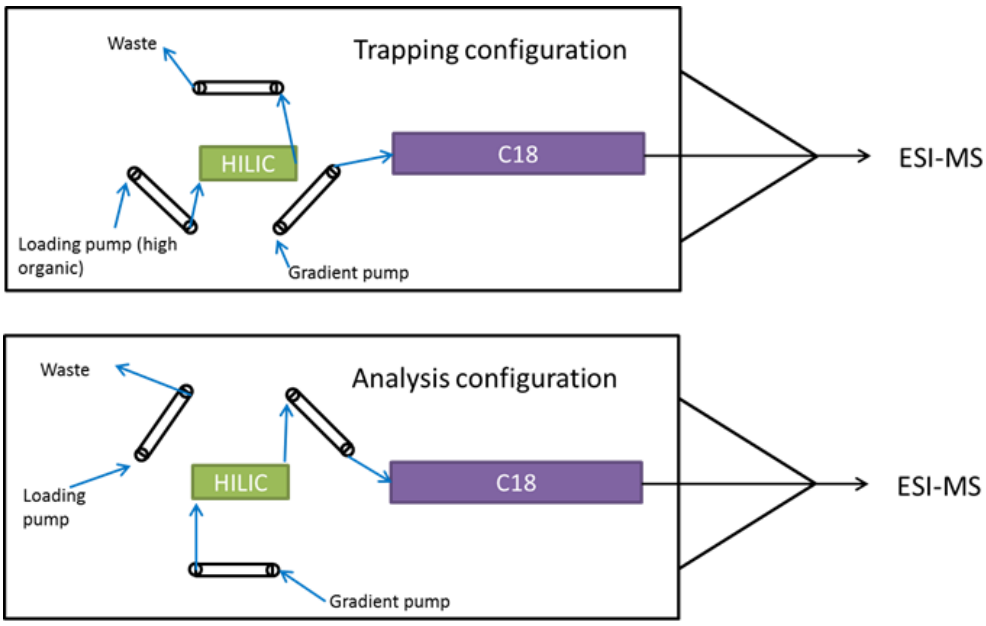


Figure 10: Schematic showing trapping and analysis configurations of a HILIC-C18 HPLC-chip for glycopeptide analysis

The HPLC mobile phases for the C18 chip were as follows: Solvent A was 99% water, 1% acetonitrile modified with 0.1% formic acid. Solvent B was 99% acetonitrile, 1% water modified with 0.1% formic acid. The enrichment column was equilibrated using the loading pump which was constantly pumping 98% solvent A and 2% solvent B as loading conditions. The sample was washed on the enrichment column with 4µL of the loading condition solvent and then the 6 port valve was switched to bring the enrichment column in-line with the analytical column. The analytical pump was used to deliver the solvent after the valve switched. 98% solvent A was pumped for the first 10 minutes after a gradient was started from 98% solvent A to 15% solvent A over 30 min. The solvent conditions were then changed to 5% solvent A in the next 2 min and kept constant for 5 min thereafter to wash the columns. The columns were brought back to the starting condition, 98% solvent A, in 3 min and then equilibrated for 5 min.

For the HILIC-C18 chip, the HPLC conditions were as follows. Solvent A was 99% water, 1% acetonitrile with 0.1% formic acid. Solvent B was 99% acetonitrile, 1% water with 0.1% formic acid. A premixed special loading solvent, 80% acetonitrile, 20% water, modified with 0.1% trifluoroacetic acid, was delivered using the loading pump in this case as the loading condition. The samples were loaded and washed on the trapping column using the loading pump with 4µL of the special loading solvent. These high organic conditions were used to allow binding of the glycosylated peptides to the HILIC trapping column which the non-glycosylated peptides were washed away to waste. The analytical column was equilibrated with the regular mobile phases at 98% solvent A.

After the washing step was complete, the trapping column was brought in-line with the analytical column and the setup received flow from the nanoscale analytical pump, which delivered 98% solvent A for the first 10 min. These high-aqueous conditions allowed elution of the glycosylated peptides from the HILIC trapping column and binding to the C18 reversed phase analytical column. After this a gradient was delivered by the analytical pump to resolve and elute the glycopeptides from the C18 analytical column. The gradient was similar to that used in the C18 chip setup. Mobile phase was changed from 98% solvent A to 15% solvent A over 30 min. The solvent conditions were then changed to 5% solvent A in the next 2 min and kept constant for 5 min thereafter to wash the columns. The columns were brought back to the starting condition, 98% solvent A, in 3 min and then equilibrated for 5 min.

Both HPLC chips (C18 and HILIC-C18) consisted of an integrated nano-ESI source, to allow sample introduction into the mass spectrometer. The MS and tandem MS conditions used for both C18 and HILIC-C18 runs were the same. Data dependent tandem mass spectra were acquired for the first 40 minutes of the run. Top 5 most abundant, multiply charged precursor ions were selected for fragmentation with dynamic exclusion enabled after 10 spectra for each precursor. MS spectra were acquired in the *m/z* 200-2000 range and MS/MS spectra in the *m/z* 50-3000 range. MS and MS/MS spectra were acquired at rates of 4 and 1 spectra/s, respectively. An isolation width of 4 *m/z* was used. Nitrogen was used as the collision gas and collision energy varied with precursor *m/z*. The collision energy was determined dynamically as a linear function of precursor *m/z*, with a slope of 5 and offset of -4.8 to yield peptide backbone fragments.

For only glycan dissociation, a lower collision energy can be used, also calculated using the precursor *m/z*, with a slope of 0.9 and offset of 2. Data were acquired in the positive ion mode with source capillary voltage set at 2150 V for the first 40 minutes of the run. Drying gas flow was 13 l/min for the first 25 minutes and 11 l/min, thereafter with gas temperature set at 225 ^{0}C throughout the run. Data files were acquired using Agilent MassHunter Workstation Data Acquisition software and analyzed using Agilent MassHunter Qualitative Analysis software package.

2.2.2 LC-MS data analysis

For automated data analysis, entire LC-MS data files were deconvoluted using DeconTools [266]. Deconvoluted data were matched and scored against theoretical peptide and glycopeptide neutral mass lists, using GlycReSoft [359], with a 10 ppm mass tolerance. For quantitative evaluation of the ability to perform tandem MS on glycopeptides, LC-MS data files were also searched for precursor ions, which produced oxonium ions. The search was performed using software written in-house that generates a list with the total number of precursor ions in an LC-MS/MS data file and precursors that produce oxonium ions. Data files were searched for the following oxonium ion masses [624] with a 0.01 unit mass tolerance: *m/z* 204.08635, 186.07539, 168.065, 138.05423, 366.13942, 274.0920 and 292.1026. Only the precursor ions that produced 3 or more of the listed oxonium ions were identified as glycopeptide precursor ions. The number of unique glycopeptide precursors selected for data dependent tandem MS were determined in this way and compared among HILIC-C18 and C18 data.

Tandem mass spectra were analyzed both manually and using a prototype software tool, written in-house, to confirm peptide backbone fragment ions and glycan compositions for manually assigned glycopeptides. Theoretical peptide fragment ion masses were calculated and matched, using a 20 ppm mass error tolerance. For manual matching of tandem MS peaks a minimum signal threshold of 5 counts was used.

2.3 Results and Discussion

2.3.1 Glycopeptide identification based on accurate intact mass

We compared the usefulness of LC-MS data acquired using HILIC-C18-MS *versus* C18-MS for profiling glycoprotein proteolytic peptides. For this purpose, the neutral mass values extracted from the data sets were searched against a set of theoretical glycopeptide masses, with a 10 ppm mass-error tolerance. The theoretical masses consisted of the protein proteolytic peptides with up to two missed cleavage sites. Masses for those peptides containing an *N*-glycosylation sequon were calculated as a set of glycosylation variants using *N*-glycosylation compositions ranging from core *N*-glycan structures to penta-antennary complex-type *N*-glycans containing N-acetylneuraminic acid and high-mannose *N*-glycans. The GlycReSoft program [359] was used to match and score deconvoluted masses from an LC-MS experiment against this list of theoretical compound accurate masses. Because the complexity in glycopeptide data increases with the number of glycosylation sites, intact mass assignments suffer from ambiguous matches and false positives that cannot be verified in the absence of tandem MS.

We acquired LC-MS profiling data for three glycoproteins with a range of complexities. Transferrin is known to have two *N*-linked glycosylation sites with complex *N*-glycans and the glycan heterogeneity for this protein is fairly limited with [5,4,0,2,0] [Hex, HexNAc, dHex, NeuAc, NeuGc] contributing to over 90% of the glycan distribution [596]. Thus, this glycoprotein presents low complexity, which makes the search space small and minimizes the chances for false positives and ambiguous assignments. AGP, by contrast, has five known *N*-linked glycosylation sites [530] with multiple genetic variants and a more diverse set of complex glycans that make this a relatively more complex glycoprotein. Recombinant hemagglutinin from *A/USSR/90/1977 (H1N1)*, referred to herein as HA, was the most complex of the three glycoproteins analyzed, with 9 consensus glycosylation sites. In addition, theoretical tryptic cleavage of HA produced 2 peptides that presented more than one putative glycosylation site on a single peptide. Thus, expected ambiguity and false positives scaled with increasing number of glycosylation sites in transferrin, AGP and HA. The number of theoretical glycopeptide compositions used for each glycoprotein is given in Table 2.

We observed a significantly higher number of glycopeptide matches using HILIC-C18 LC-MS data over C18 LC-MS data, for all glycoproteins tested. Table 2 shows the number of glycopeptides and peptides identified based on intact glycopeptide masses. Even with the limited the possibility of false matches when searching solely based on theoretical masses, it was evident that HILIC-C18 enriched glycopeptides efficiently.

Data were acquired as analytical triplicates and peptide/glycopeptide compositions found in all three replicates were considered matches.

2.3.2 Oxonium ion distributions

In order to confirm the ion compositions from the MS data, we used data-dependent LC-tandem mass spectrometry. Collisional dissociation of glycopeptides leads to formation of oxonium ions that are useful as features for identifying glycopeptides in a LC-MS/MS chromatograms [71]. We therefore used the absence of oxonium ions to disqualify a precursor ion as glycopeptide. Based on this rationale, we compared the abundances of oxonium ions in extracted ion chromatograms (EICs) of glycopeptides observed using C18-MS/MS *versus* HILIC-C18-MS/MS data. Figure 11 compares the HexNAc oxonium ion (*m/z* 204) EIC with the BPC (base peak chromatogram) for hemagglutinin. When using HILIC-C18, we observed that the abundances of the HexNAc (*m/z* 204) oxonium ion profile were similar to those of the BPC, consistent with the fact that most of the detected ions corresponded to glycopeptides. For the C18 LC-MS experiments, the abundances of HexNAc oxonium ion were 10-fold lower than the BPC, indicating that most of the detected ions corresponded to unglycosylated peptides.

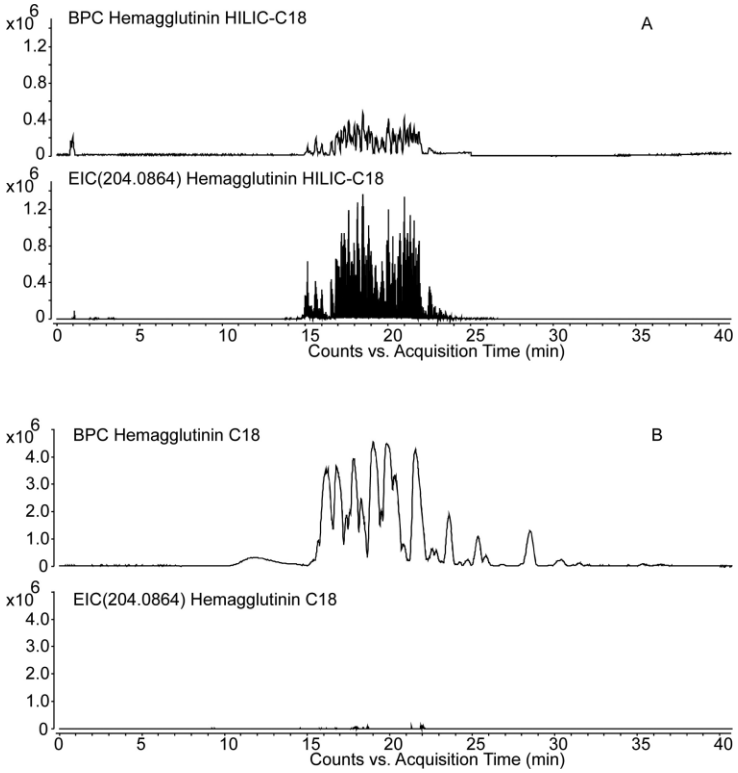


Figure 11: Comparison of tryptic glycopeptide abundances using (A) HILIC-C18-MS and (B) C18-MS

Each panel shows the base peak chromatogram (BPC) and the extracted ion chromatogram (EIC) for the HexNAc oxonium ion (*m/z* 204.08). BPC and EIC are shown using the same Y-scale.

Using C18-MS, unmodified peptides were the most abundant ions selected by the acquisition software for tandem MS based on abundance. By contrast, for HILIC-C18 data, glycopeptides were the most abundant ions present and were selected automatically for tandem MS. The oxonium ion distributions were consistent with the results obtained from the LC-MS data, shown in Table 2, indicating the contribution of glycopeptides to total ion abundances in C18 and HILIC-C18 runs.

Glycoprotein		Transferrin	AGP	Hemagglutinin
Number of theoretical peptides + glycopeptides		1576	4009	57358
Total number of non-glycosylated peptides identified (MS1 only)	**HILIC-C18**	2	2	4
	C18	77	19	39
Number of theoretical glycopeptides (search space)		1230	3936	56886
Number of ambiguous matches (more than one match)	**HILIC-C18**	6	16	725
	C18	5	23	278
Total number of glycopeptides identified (MS1 only)	**HILIC-C18**	54	155	1220
	C18	38	141	530
Total % relative abundance of glycopeptides	**HILIC-C18**	97.67%	99.93%	99.97%
	C18	0.86%	3.40%	2.52%

Table 2: Total number of identified peptides and glycopeptides from LC-MS data sets, using GlycReSoft (MS only)

Only the glycopeptides/peptides found in all three replicate data were included. Relative abundances were calculated as a percentage of total abundances of all identified compounds. Note: Deamidation was considered as a variable modification only in the case of HA.

Oxonium ions detected early in the chromatogram using HILIC-C18-MS resulted from a small fraction of the enriched glycopeptides from the HILIC trapping column getting washed away with the dead volume containing high organic mobile phase. This was a fixed and unbiased loss that accounted for less than 2% of the total sample abundance and did not affect the relative abundances of the compounds being retained on the C18 analytical columns.

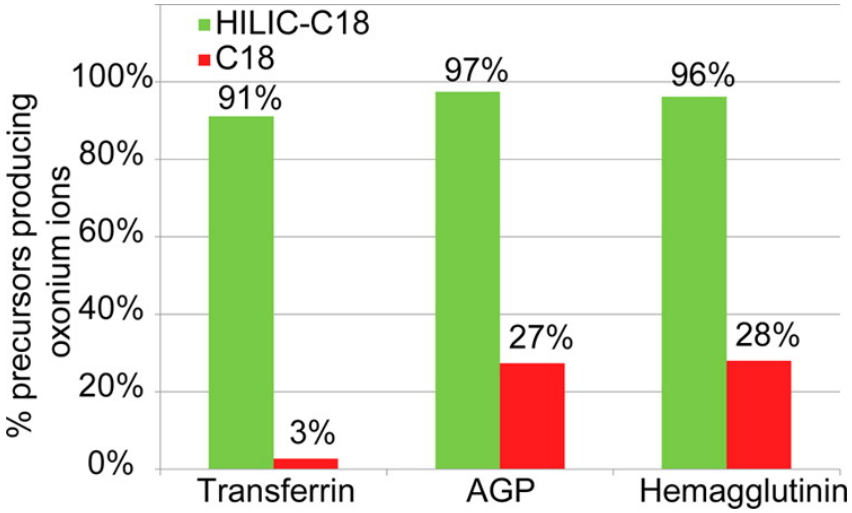


Figure 12: Comparison of percent of precursor ions that generate diagnostic oxonium ions between HILIC-C18-MS (green) and C18-MS (red)

Results are shown for the glycoproteins labeled on the X axis.

Figure 12 shows the percent of precursor ions identified as glycopeptides based on formation of oxonium ions in data dependent LC-MS/MS runs using HILIC-C18-MS *versus* C18-MS. We concluded that use of HILIC-C18-MS resulted in significant increase in the ability to analyze glycopeptides using data-dependent LC-MS/MS. These data demonstrate that the HILIC-C18-MS increased abundances of glycopeptides relative to unglycosylated peptides and thus the quality of data dependent LC-tandem MS data. The improved data quality increased confidence in assignments and decreased false identifications.

Two AGP glycopeptides (ENGTISR/ENGTVSR and NEEYNK) were detected only using HILIC-C18-MS. This was due to the fact that tryptic cleavage of AGP produces glycopeptides that are only 5 to 7 amino acids long and are not retained during the trapping step when using C18-MS. These glycopeptides were trapped on the HILIC

enrichment column when using HILIC-C18-MS and eluted soon after the solvent-front on the analytical chromatogram, which allowed them to undergo MS and MS/MS analysis. The C18 trapping step can be eliminated to retain any shorter glycopeptides in a sample. However, in our experience an on-line trapping step is useful means of elimination of salts and other contaminants while minimizing manual manipulation.

2.3.3 Tandem MS of glycopeptides

While the presence of oxonium ions from dissociation of glycosidic bonds is a useful feature, the formation of peptide backbone product ions is necessary for unambiguous assignment of the peptide sequence and glycan composition. We therefore investigated the extent to which the increased glycopeptide precursor ion abundances obtained using HILIC-C18-MS improved the quality of the resulting glycopeptide tandem mass spectra. Collisional dissociation using typical conditions for peptide tandem MS results in formation of abundant ions from dissociation of the glycan with those from peptide backbone dissociation often not detected. In the interest of maximizing the information produced from a CAD LC-MS/MS experiment, we investigated use of higher collision energies for glycopeptides. Figure 13 shows an example of glycopeptide tandem mass spectra from a hemagglutinin glycopeptide; collision energies were calculated as per the equations described in the methods section. As expected, glycopeptide dissociation produced abundant oxonium ions in the low *m/z* range that confirmed the presence of monosaccharides in the precursor ion composition. In addition, peptide backbone product ions were detected that enabled direct identification of the peptide. This information increased confidence in true matches and decreased those in incorrect

matches. This was important in cases where more than one theoretical glycopeptide matched an observed mass value. Complete peptide backbone or parts of the peptide backbone could be sequenced for glycopeptide precursors as shown in the annotated spectra. Intact peptide and glycopeptides+saccharide (referred to as stub glycopeptide) ions were also detected in the higher *m/z* range of the spectra that matched the masses of peptide with *N*-glycan core structures. The presence of intact peptide or stub glycopeptide ions significantly increased confidence of assignments over those made from MS-only data. In the majority of LC-tandem mass spectra, ions corresponding to peptide backbone product ions plus a HexNAc residue were detected, which confirmed the site of glycosylation.

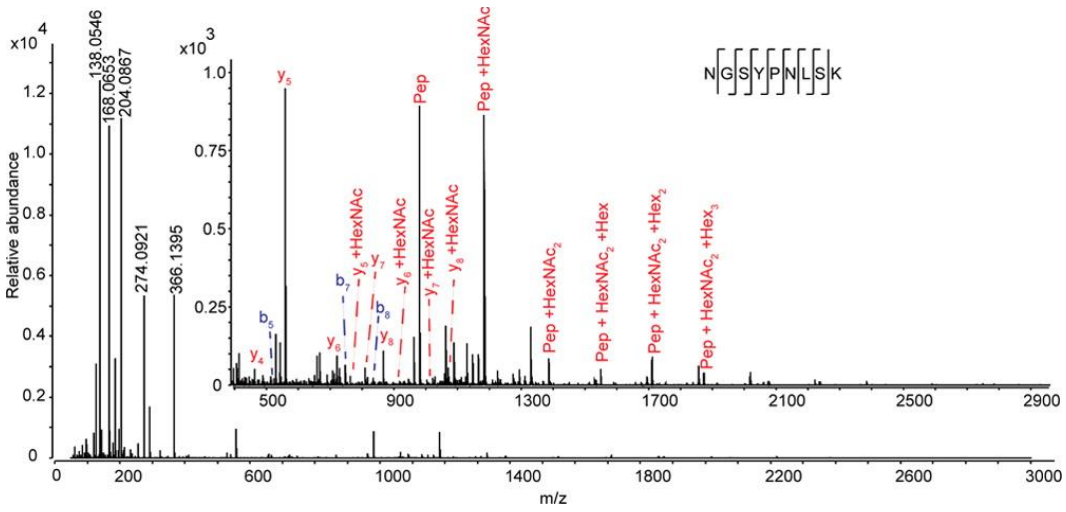


Figure 13: Tandem mass spectrum of Hemagglutinin glycopeptide precursor ion 1013.7474 [M+3H]3+, identified as NGSYPNLSK-[5,4,1,1,0]

Glycopeptide identifier is listed as Peptide-[a,b,c,d,e], where: a = number of Hexoses; b = number of N-acetylhexosamines; c = number of Deoxyhexoses; d = number of N-acetylneuraminic acids; e = number of N-glycolylneuraminic acids.

Although, the relative abundance of a glycopeptide precursor ion had an effect on the quality of tandem MS by affecting the ability to select the ion in a data-dependent

LC-MS/MS experiment, the absolute abundance did not appear to play a role in generation of useful fragment ions for confident assignment of the glycopeptides. In order to demonstrate that the detection of peptide backbone product ions was a universal phenomenon, we analyzed glycopeptides with different peptide backbones and varying absolute abundances, from each of the three glycoproteins studied. Table 3 summarizes the types of product ions observed using HILC-C18 *vs.* C18 LC-tandem MS.

Glycoprotein	Glycresoft assignments based on intact neutral mass				Precursor ion m/z and charge	Manual interpretation of tandem MS															
						HILIC-C18								C18							
	HILIC-C18 Compositions	Avg. % relative volume (out of matched compounds)	C18 Compositions	Avg. % relative volume (out of matched compounds)	Precursor ion m/z and charge	Total precursor ion abundance (area)	Selected for Tandem MS	Oxonium ions	Intact peptide	Stub glycopeptides	Peptide sequence	% y-ion coverage	% b-ion coverage	Glycosylation site (b/y-ion with HexNAc)	Total precursor ion abundance (area)	Selected for Tandem MS	Oxonium ions	Intact peptide	Stub glycopeptides	Peptide sequence	% y-ion coverage	% b-ion coverage	Glycosylation site (b/y-ion with HexNAc)
Transferrin	QQQHLFGSNVTDCSGNFCLFR-[5,4,0,2,0]	30.04%	QQQHLFGSNVTDCSGNFCLFR-[5,4,0,2,0]	0.27%	1180.7328 (4+)	1.02E+07	√	√	√	√	√	95%	70%	√	7.77E+06	√	√	√	√	√	85%	70%	√
	CGLVPVLAENYNK-[5,4,0,2,0]	18.57%	CGLVPVLAENYNK-[5,4,0,2,0]	0.25%	921.1389 (4+)	5.76E+06	√	√	√	√	√	67%	67%	√	6.66E+06	√	√	√	√	√	67%	67%	√
	QQQHLFGSNVTDCSGNFCLFR-[5,4,1,2,0]	0.50%	QQQHLFGSNVTDCSGNFCLFR-[5,4,1,2,0]	0.02%	1217.2468 (4+)	1.10E+06	√	√	√	√	√	85%	55%	√	8.87E+05	x	x	x	x	x	0%	0%	x
	QQQHLFGSNVTDCSGNFCLFR-[6,5,1,3,0]	0.91%	QQQHLFGSNVTDCSGNFCLFR-[6,5,1,3,0]	0.00%	1381.3025 (4+)	3.21E+05	√	√	√	√	√	80%	50%	√	1.11E+05	x	x	x	x	x	0%	0%	x
AGP	ITGKWFYIASAFRNEEYNK-[4,6,1,1,0]	10.32%	ITGKWFYIASAFRNEEYNK-[4,6,1,1,0]	0.76%	1161.0091 (4+)	1.56E+07	√	√	x	x	x	0%	0%	x	1.58E+07	√	√	x	x	x	0%	0%	x
	SVQEIQATFFYFTPNK-[7,6,0,4,0] or QNQCFYNSSYLNVQRENGTVSR-[6,6,0,2,0]	10.80%	SVQEIQATFFYFTPNK-[7,6,0,4,0] or QNQCFYNSSYLNVQRENGTVSR-[6,6,0,2,0]	0.02%	1088.2441 (5+)	1.49E+07	√	√	√	√	√	93%	80%	√	0.00E+00	x	x	x	x	x	0%	0%	x
	QNQCFYNSSYLNVQRENGTVSR-[4,4,0,1,0]	4.46%	QNQCFYNSSYLNVQRENGTVSR-[4,4,0,1,0]	0.27%	1104.7291 (4+)	1.22E+07	√	√	x	x	x	0%	0%	x	5.88E+06	√	√	x	x	x	0%	0%	x
	QDQCIYNTTYLNVQR-[6,5,0,3,0]	1.94%	QDQCIYNTTYLNVQR-[6,5,0,3,0]	0.01%	1194.9868 (4+)	6.04E+06	√	√	√	√	√	100%	57%	√	1.73E+06	√	√	√	√	√	43%	14%	√
	NEEYNK-[5,4,0,2,0]	2.77%	Not found	0.00%	751.0409 (4+)	4.74E+06	√	√	√	√	√	100%	80%	√	0.00E+00	x	x	x	x	x	0%	0%	x
	ENGTISR-[7,6,0,4,0]	1.18%	ENGTISR-[7,6,0,4,0]	0.00%	1074.1627 (4+)	3.97E+06	√	√	√	√	√	100%	83%	√	0.00E+00	x	x	x	x	x	0%	0%	x
	SVQEIQATFFYFTPNKTEDTIFLR-[7,6,0,4,0]	1.91%	Not found	0.00%	1283.3416 (5+)	3.80E+06	√	√	√	√	√	74%	70%	√	0.00E+00	x	x	x	x	x	0%	0%	x
	ENGTVSR-[7,6,0,4,0]	0.24%	Not found	0.00%	1070.6586 (4+)	6.95E+05	√	√	√	√	√	100%	33%	√	0.00E+00	x	x	x	x	x	0%	0%	x
	ENGTISR-[6,5,1,3,0]	0.19%	Not found	0.00%	1261.8232 (3+)	2.90E+05	√	√	√	√	x	17%	0%	x	0.00E+00	x	x	x	x	x	0%	0%	x
HA USSR	NGSYPNLSK-[5,4,1,1,0]	0.02%	NGSYPNLSK-[5,4,1,1,0]	0.01%	1013.7474 (3+)	1.04E+06	√	√	√	√	√	88%	63%	√	4.80E+05	√	√	√	√	√	100%	63%	x
	NVTVTHSVNLLEDSHNGK(1-Deamidated)-[7,6,1,3,0] NGSYPNLSKSYVNNK(2Deamidated)-[8,8,3,1,0]	0.02%	NVTVTHSVNLLEDSHNGK(1-Deamidated)-[7,6,1,3,0] NGSYPNLSKSYVNNK(2Deamidated)-[8,8,3,1,0]	0.03%	1068.2338 (5+)	6.38E+05	√	√	√	√	√	88%	88%	√	8.60E+04	x	x	x	x	x	0%	0%	x
	GFGSGIITSNASMDECDTK-[7,6,1,3,0]	0.03%	GFGSGIITSNASMDECDTK-[7,6,1,3,0]	0.01%	1341.2684 (4+)	6.35E+05	√	√	√	√	√	83%	61%	√	8.00E+04	x	x	x	x	x	0%	0%	x
	SWSYIAETPNSENGTCYPGYFADYEELR-[7,6,1,3,0] or NGSYPNLSKSYVNNKEK1Deamidated-14,8,0,3,0] or NGSYPNLSKSYVNNK2Deamidated-[12,13,3,0,0]	0.22%	SWSYIAETPNSENGTCYPGYFADYEELR-[7,6,1,3,0] or NGSYPNLSKSYVNNKEK1Deamidated-14,8,0,3,0] or NGSYPNLSKSYVNNK2Deamidated-[12,13,3,0,0]	0.01%	1339.1270 (5+)	3.09E+05	√	√	√	√	√	81%	59%	√	3.23E+04	x	x	x	x	x	0%	0%	x
	GFGSGIITSNASMDECDTK-[7,6,1,4,0]	0.06%	GFGSGIITSNASMDECDTK-[7,6,1,4,0]	0.01%	1131.4390 (5+)	1.68E+05	√	√	√	√	√	78%	67%	√	2.27E+04	x	x	x	x	x	0%	0%	x

Table 3: Summary of glycopeptide tandem mass spectra for intact-mass matched compounds

The features used in identification are listed in the table as yes/present (✓) or no/absent (x). Abundances refer to the composite values of all charge states identified and matched for a given glycopeptide, presented as a percent of the total volume of all matched compounds. The absolute precursor ion abundances were determined from extracted ion chromatograms. Peptide backbone coverage was reported as percent product ions detected of total possible peptide product ions. Glycopeptide compositions in red indicate false/incorrect match. Glycopeptide nomenclature has been described in the text. Detailed mass tables confirming glycopeptide assignments have been presented with the supplemental information in [293]. Glycopeptide identifiers are listed as Peptide-[a,b,c,d,e], where: a = number of Hexoses; b = number of N-acetylhexosamines; c = number of Deoxyhexoses; d = number of N-acetylneuraminic acids; e = number of N-glycolylneuraminic acids

For all glycopeptides assignments that were confirmed by tandem MS, the following criteria were used (1) intact glycopeptide mass match; (2) presence of oxonium ions; (3) presence of either an abundant protonated peptide ion and/or a peptide + HexNAc ion; (4) presence of significant peptide backbone product ion coverage. It is significant that some of the glycopeptide compositions that were assigned to high abundance precursor ions based on intact mass that also produced oxonium ions, were rejected (shown in red font) because the tandem MS stub glycopeptide or peptide y-ions were not consistent with the assignment. This emphasizes the need for tandem MS on glycopeptides for confident assignments. Although the ions shown in Table 3 were quite abundant in the mixtures analyzed, many were not selected for data dependent tandem MS when using C18-MS due to the presence of more abundant non-glycosylated peptides. In addition, GlycReSoft failed to find and match some of these compositions in C18 data, which was probably due to very low abundances or presence of overlapping isotopic peaks. HILIC-C18-MS improved the abundances of glycopeptides relative to non-glycosylated peptides, which allowed these ions to be matched by the GlycReSoft software. Data-dependent acquisition of tandem mass spectra led to systematic identification of these glycopeptides and allowed confident assignments due to the presence of peptide backbone product ions. The abundant product ions containing stub glycopeptides and oxonium ions were useful for confirming peptide identities.

Results for transferrin. Transferrin presents two glycosylation sites, with the most abundant *N*-linked glycan of composition [5,4,0,2,0] at each of those two sequons. In agreement with this, glycopeptides QQQHLFGSNVTDCSGNFCLFR-[5,4,0,2,0] and

CGLVPVLAENYNK-[5,4,0,2,0] could be identified using tandem MS in both HILIC-C18 and C18 analyses. Other transferrin glycopeptides, although identified by GlycReSoft in the C18 LC-MS data, did not get selected for data dependent tandem MS due to presence of abundant non-glycosylated peptides. However, these glycopeptides underwent tandem MS and could be confidently assigned in case of HILIC-C18.

Results for α1-acid glycoprotein. Compared to transferrin which has only two glycosylation sites and limited glycan heterogeneity, AGP is more complex with five different glycosylation sites and differences in peptide backbone arising from genetic variants. The HILIC-C18 method was significantly more effective at acquiring data dependent tandem mass spectra of sufficient quality to enable identification of the peptide backbone. We observed formation of useful peptide backbone ions, even for precursor ions with moderate abundances (~1-2x10^5 counts). Tandem mass spectra were useful in discriminating single amino acid genetic variants , for example, in Table 3, two AGP glycopeptides have similar peptide sequences, ENGTVSR and ENGTISR, for which we observed complete peptide sequences in the glycopeptide tandem mass spectra of precursors 1070.6586 (4+) and 1074.1627 (4+). These peptides were not retained using the C18-MS system.

Results for influenza A virus hemagglutinin. HA presents a considerably greater analytical challenge than AGP. Due to the presence of 9 putative glycosylation sites with a wide distribution of possible glycan compositions at each site, the number of theoretical glycopeptides (the search space) was more than 10-fold greater than that of AGP, as shown in Table 2. In addition, the number of glycopeptide glycoforms detected was

greater, and as a result of higher glycoform heterogeneity the corresponding relative abundances of each glycopeptide precursor ion were generally lower, than those observed for transferrin and AGP. In Table 3, we compared 5 different glycopeptide compositions that were identified by GlycReSoft in both HILIC-C18 and C18 data from HA analyses. Out of the five glycopeptides assigned by intact mass, only one, NGSYPNLSK-[5,4,1,1,0], underwent tandem MS with C18 analysis. However, in case of HILIC-C18, useful tandem MS data were acquired on all 5 glycopeptides, leading to formation of not only peptide backbone ions but also backbone ions with an attached HexNAc, facilitating identification of the glycosylation site. This was important in case of NGSYPNLSK-[5,4,1,1,0], where two glycosylation sequons NGS and NLS were present on the same glycopeptide; HILIC-C18 tandem MS data allowed us to identify NLS as the occupied sequon as shown in Figure 13. Tandem MS was also useful in identifying other modifications. For example, the exact site of deamidation could be identified in case of HA glycopeptide NVTVTHSVNLLEDSH**N**GK(1-Deamidation)-[7,6,1,3,0].

As shown in Table 2, measurement of glycopeptide mass did not suffice to identify the peptide and glycan composition because of the size of the search space. Thus, observed masses consistent with more than one glycopeptide were not uncommon. We showed that tandem MS could be used to resolve ambiguities. As shown in Table 3, HA precursor ion *m/z* 1339.1270 (5+) matched three different compositions SWSYIAETPNSENGTCYPGYFADYEELR-[7,6,1,3,0], NGSYPNLSKSYVNNKEK (1 deamidation)-[14,8,0,3,0] and NGSYPNLSKSYVNNK (2 deamidation)-[12,13,3,0,0], within a 10 ppm mass error tolerance. Peptide backbone and stub glycopeptide ions

proved useful in assigning SWSYIAETPNSENGTCYPGYFADYEELR-[7,6,1,3,0] as the correct composition. In case of AGP, SVQEIQATFFYFTPNK-[7,6,0,4,0] or QNQCFYNSSYLNVQRENGTVSR-[6,6,0,2,0] also corresponded to the same precursor mass 1088.2441 (5+) and SVQEIQATFFYFTPNK-[7,6,0,4,0] was assigned as the correct composition based on tandem MS data. The fact that HILIC-C18-MS resulted in relative precursor ion abundances sufficient to allow selection for tandem MS and consequently, make these assignments, demonstrates the value of this approach.

2.3.4 Confident site-specific glycan profiling:

We used the glycopeptides assignments to construct a site specific glycosylation map for HA and AGP (Figure 14). Glycan compositions were classified based on the number of HexNAc units, as high-mannose (2 HexNAc), bi-antennary (3 or 4 HexNAc), tri-antennary (5 HexNAc), tetra-antennary (6 HexNAc) and penta-antennary (7 HexNAc). Only glycopeptides identified with at least 40% peptide backbone coverage from tandem MS data were included in the glycan profile maps. In some cases, particularly for HA, two putative glycosylation sites appeared on a single tryptic peptide. The peptide backbone fragment ions with an attached HexNAc helped resolve the ambiguity in such cases by helping to identify the occupied site. Such site-specific glycan profile assignments were possible only when using HILIC-C18-MS/MS whereby at least 7 times more glycopeptides underwent tandem-MS, from evaluation of precursors generating oxonium ions. It is evident from the data that the usefulness of HILIC-C18-MS scales with the number of glycosylation sites and glycopeptides in the theoretical search space. For transferrin, where the glycopeptide and glycoform diversity is very

limited, the C18 and HILIC-C18 platforms yield similar results. However, with AGP and hemagglutinin there is an obvious improvement in data quality with the use of HILIC-C18 over C18. We made these assignments using assumptions regarding the protein sequences, glycosylation sites, glycoform distributions and post-translational modifications. It is therefore possible that modified peptides not included in our assumptions exist. Since the goal of this study was to compare performances of the HILIC-C18 and C18 systems, our glycopeptide analyses sufficed to reach clear conclusions despite the absence of exhaustive discovery-mode proteomics data interpretation.

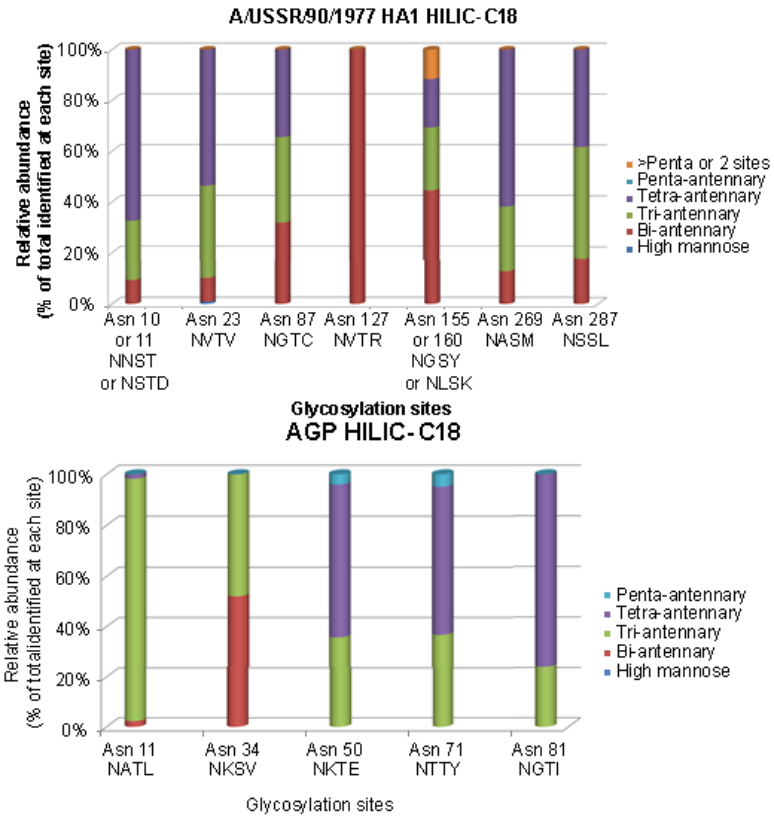


Figure 14 Site-specific glycan profile of hemagglutinin from influenza A virus (top) And Human AGP (bottom)

2.4 Conclusions:

We assert that even for purified glycoproteins, a tandem MS step is necessary in order to confidently assign glycopeptide compositions with respect to peptide and glycan. Because the traditional data-dependent mass spectrometry algorithms used in proteomics select the most abundant precursor ions, most glycopeptides are not selected for tandem MS when using C18 chromatography. We therefore developed a method in which glycopeptides are enriched using a HILIC trapping column on-line prior to a C18 analytical column. Using this method, we observed that glycopeptides were the most abundant ions in the data sets and therefore preferentially selected for data-dependent LC-MS/MS. The HILIC trapping column provides efficient enrichment of glycopeptides. The C18 analytical column provides high chromatographic resolution. Thus, we demonstrated that online enrichment and separation can be conveniently performed without manual sample manipulation. This combination enables facile glycopeptide enrichment, separation and analysis with online LC-MS/MS using a commercially available instrument system.

CAD-based tandem MS of glycopeptides is useful for determining peptide sequence and glycan composition; it does not determine glycan structure in detail. We found that by programming the MS system to acquire CAD tandem MS using elevated collision energy, we observed peptide backbone product ions for the majority of the glycopeptides selected for tandem MS. Even for the lowest abundance glycopeptide precursor ions so selected, we observed glycan modified peptide stubs that were useful for confirming the mass of the peptide moiety. These tandem mass spectra improve the

informational value of an LC-MS/MS experiment, from the point-of-view of glycopeptide analysis. In principle, however, use of on-line enrichment of glycopeptides with HILIC-C18-MS will improve tandem MS results for any dissociation method.

2.5 Acknowledgements

This work was supported by NIH grant P41GM104603.

Chapter 3: Development of data-analysis workflows for glycoproteomics and glycomics

(Part of this work is published in [292])

3.1 Introduction: the need for tailored bioinformatics workflows in glycoproteomics and glycomics

Proteomics methods have focused on the analysis of PTMs with single, defined mass shifts, including phosphorylation, acetylation, methylation and ubiquitination [32,87,122,621]. Glycosylation does not fit into the traditional proteomics paradigm because, owing to its microheterogeneity, the peptide mass cannot be accurately and unambiguously predicted from the molecular mass [122]. In addition, the molecular weight heterogeneity is multiplied by the presence of other modifications, including mutations, oxidations, deamidations, dehydrations, incomplete and/or unexpected proteolytic cleavages. Therefore, tailored bioinformatics algorithms are needed to assign glycopeptide mass spectral data with confidence.

As an example, consider the tryptic peptide sequence QNQC(Carbamidomethyl)FYNSSYLNVQR from α1-acid glycoprotein 2 http://www.uniprot.org/uniprot/P19652 [346,537], which contains a single *N*-glycosylation site. We do not have information on the actual frequency of Asparagine (N) deamidation for this protein. Therefore, we have to rely on the assumption that deamidation can occur on N not occupied by a glycan, and there are two Ns in this sequence excluding the one in the glycosylation sequon. Deamidation can occur

spontaneously, so we cannot *a priori* designate where it may occur, so if we are constructing a theoretical database, we must include all combinations. The number of peptidoforms from variable deamidation is $\sum_{i=0}^{2}\binom{2}{i}$. There are 4 theoretical peptidoforms from deamidation alone. Next we combine these peptidoforms with some library of glycan compositions. Assuming a list of 200 *N*-glycan compositions, which may be considered conservatively small, we have to consider 4 * 200 = 800 glycopeptides for a single tryptic peptide with one missed cleavage and one glycosite. If we use the target-decoy method [139,635], that means we must search at least 800 decoy glycopeptides as well. We don't even expect many of those cases to be real in any case, but we must dedicate compute time and statistical power to analyze them. As the above example shows, it is possible to estimate the search space for a glycopeptide by making assumptions regarding the multiplicity of molecular forms present. This method of describing peptidoforms using only subjective information we refer to as *naïve*. Such assumptions can lead to erroneous data interpretations as shown by Knudsen *et al.* [302].

In order to minimize the margin for error, we want to make as few assumptions about the contents of our sample as possible. The set of theoretical glycopeptides, known as a search space, consists of the peptide sequences, multiplied by number of possible glycan structures, multiplied by any other structural/chemical variants (such as oxidations, deamidations, etc.). Ideally we would like to use the mass spectral data to calculate confidence values for the search space so as to identify the site-specific glycosylation present. In order to do this, the search space must be defined accurately. The best way to define the proteins present in the sample is to measure the proteome.

Likewise, analytical measurement of the released glycome is the best way to determine the glycans present. Methods for measuring proteomes and glycomes of glycoprotein samples are mature and available for use by the same mass spectrometry instruments used for glycoprotein analysis. To what extent do these data domains improve the confidence whereby site specific glycosylation can be assigned? This chapter describes our bioinformatics workflow and examines the effectiveness of glycoproteomics under scenarios where the search space is defined based on naïve assumptions *versus* using glycomics and proteomics data and the effects of increasing matrix complexity on data acquisition and interpretation.

Bottom-up proteomics experiments generally contain peptide precursor ions that have low charge states (+2 or +3) that during tandem MS tend to produce singly charged product ions. These ions can be matched against a theoretical database without the need to determine charge states. As a result, assignment of charge states in proteomics data is not critical. Most proteomics database search programs attempt to identify isotopic clusters and then match the theoretical fragment ions to the monoisotopic peak (A ion) in the cluster. Occasionally, incorrect assignment of the A+1 peak as the monoisotopic peak can also be confused as presence of deamidation or another modification that adds about 1 unit to the mass of the peptide. While it is not clear how commercial proteomics software packages deal with the issue of charge state deconvolution, many popular open-source software tools identify charge states by measuring the spacing between isotopic peaks but do not take into account the isotopic pattern fit [140,157,294,563].

Glycopeptides generally have higher charge states than unglycosylated peptides. As a result, they generate abundant multiply charged fragment ions, as shown in Figure 15. Furthermore, overlapping peak clusters can often lead to misassignment of the monoisotopic peak or the charge state. As shown in Section 5.3, polydisperse mixtures of oligosaccharides can generate overlapping ion clusters. Also, presence of sulfation on glycans can complicate spectra and makes even manual assignment of monoisotopic peaks challenging. It therefore becomes important to deconvolute and deisotope glycan and glycopeptide mass spectra before automated peak matching can be performed. Our glycoproteomics workflow described in this chapter utilizes an averagine model for deconvolution that models the isotopic patterns based on an average elemental composition for that class of analytes, as described by Senko and implemented by Horn [244,473].

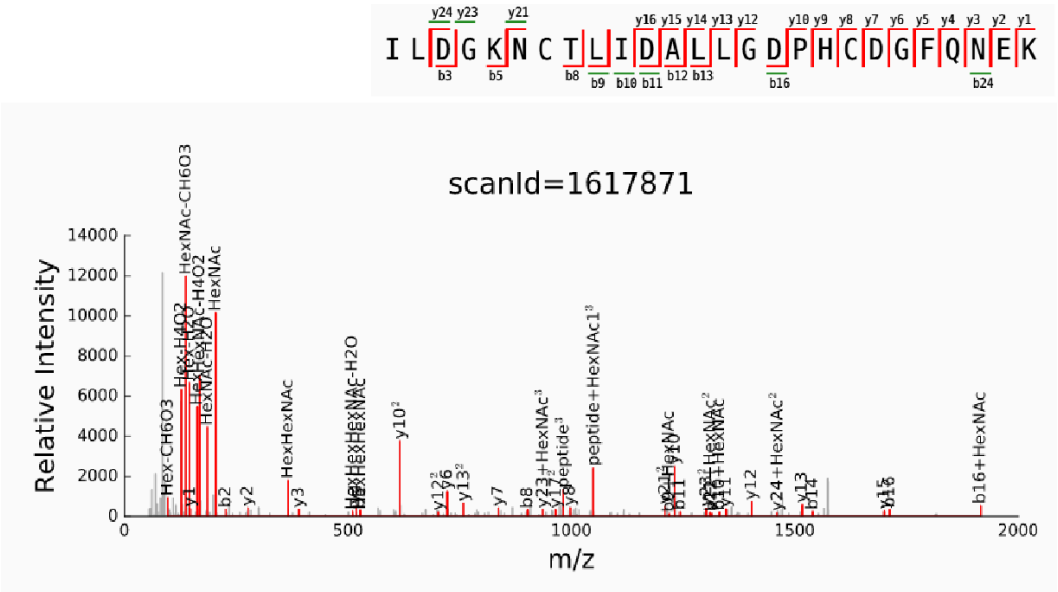


Figure 15: Tandem MS of IAV glycopeptide ILDGKN(*N*-Glycosylation)CTLIDALLGDPHCDGFQNEK Hex5 HexNAc2 assigned by GlycReSoft, showing multiply charged fragment ions

For mixtures of sulfated oligosaccharides, an averagine model is not sufficient, since the varying number of sulfates skews the isotopic pattern beyond what can be modeled using an average composition. For sulfated oligosaccharides, we therefore used a targeted pattern-fitting approach, where theoretical isotopic patterns were calculated using an implementation of the BRAIN algorithm for all theoretical compounds. Both mass and the isotopic pattern were used in matching the mass spectra, as described in Section 5.3 and elsewhere[128,611].

3.2 Methods

All the work described in this section was carried out under the guidance of Prof. Joseph Zaia. Mass spectrometry data were acquired by Kshitij Khatri. Joshua A. Klein developed the software tools including the GlycReSoft pipeline used in glycoproteomics data analysis. Data analysis and interpretation were performed by Joshua A. Klein and Kshitij Khatri.

Standard human serum glycoproteins were purchased from Sigma-Aldrich (St. Louis, MO). Glycoproteins were reduced with dithithreitol and alkylated using iodoacetamide. The samples were then subjected to tryptic digestion and individual glycoprotein tryptic digests were mixed to generate samples of scaling complexity levels as below:

α1-acid glycoprotein (AGP) (Complexity level 1)

Glycoprotein mix 3: AGP + transferrin + fetuin

Glycoprotein mix 4: AGP+ transferrin + fetuin + haptoglobin

Glycoprotein mix 5: AGP + transferrin + fetuin + haptoglobin + α2-macroglobulin

The highest complexity level analyzed was AGP in pooled human serum (Sigma-Aldrich, St. Louis, MO). Since, serum contains very high abundances of albumin and immunoglobulin G (IgG), it was necessary to deplete these resident proteins to observe coverage on the glycoproteins of interest [21,135,613]. Albumin and IgG were depleted using ProteoExtract® Albumin/IgG Removal Kit (EMD Millipore, Billerica MA). Depleted serum and glycoprotein mixes were subjected to sample preparation and analysis using a multi-pronged approach, as described elsewhere [291].

3.2.1 Proteomics analyses

Tryptic samples were deglycosylated using peptide *N*-glycosidase F (PNGase F, New England Biolabs, Ipswich MA) and subjected to triplicate C18 LC-MS/MS analyses using a Q-Exactive Plus mass spectrometer (Thermo Scientific, San Jose, CA) equipped with an Advion NanoMate (Advion Inc., Ithaca, NY) nanoESI source, coupled to a Waters NanoAcquity nanoLC system. A Waters Xbridge™ reversed-phase column (150 μm × 100 mm) with 1.7 μm BEH C18 resin and a Waters trap column (180 μm × 20 mm) packed with 5 μm Symmetry™ C18 stationary phase, were used for online desalting and separation of the proteomics samples. The mass spectrometer was programmed to acquire data-dependent tandem MS, using instrument parameters described previously [291]. Glycoproteins used in complex mixtures were analyzed both individually and as mixtures of increasing complexity.

3.2.2 Glycomics analyses

N-Linked glycans were released from tryptic digests using PNGase F. Released glycans were separated from deglycosylated peptides using C18 reversed-phase spin columns (Thermo Pierce, Rockford IL). Glycans were desalted using PD MiniTrap G-10 columns (GE Healthcare, Piscataway NJ), as per the manufacturer's instructions. Glycan samples were analyzed, in triplicate, using HILIC-MS using an Agilent 6520 LC-MS system with a chip-cube nanoESI source, as described previously [496].

3.2.3 Glycoproteomics analyses

Tryptic digests of glycoprotein mixes and depleted serum were analyzed using a previously described chip LC-MS platform [293]. Glycopeptide enrichment was performed online using a HILIC trapping column followed by C18 reversed-phase separation and data-dependent tandem MS analysis using collisional-dissociation.

3.2.4 Proteomics data analysis

Proteomics data were analyzed using Peaks Studio 7.5 [615] (Bioinformatics Solutions Inc., Waterloo ON). Triplicate data files for individual samples were combined in Peaks Spider Searches, which combine database, error-tolerant PTM searching, and *de novo* sequencing. Data were searched against a Uniprot protein database, using a precursor mass tolerance of 10 ppm and a product ion mass tolerance of 0.1 Da. A maximum of 2 missed cleavages per peptide were specified; carbamidomethylation at cysteine was used as a fixed modification and deamidation (Asn, Gln), oxidation (Met), acetylation (Lys and N-terminus), Na^+ adduction (Asp, Glu, C-term), and

pyroglutamination (Gln at N-terminus) were used as variable modifications in the database searches. The PTM search examined the data for presence of 484 variable modifications from the Unimod database [103] and possible mutations leading to amino acid substitutions. The results from combination of database, PTM and *de novo* searches were accepted at a false discovery rate (FDR) threshold of 0.1%. Peptides with spectrum matches of -10 log p values 15 or higher were used for downstream processing and building a glycoproteomics database.

3.2.5 LC-MS/MS Data Preprocessing

3.2.5.1 LC-MS Deconvolution and Relative Quantitation with Decon2LS

LC-MS data are pre-processed in the MS^1 dimension using Decon2LS/DeconTools [266], determining charge state, monoisotopic mass, isotopic cluster continuity, and total abundance integration. This data is used for gross composition matching, and in downstream analysis for relative quantitation.

3.2.5.2 MS/MS Deconvolution with BUPID/ MasSPIKE

MS/MS data taken from the LC-MS run were preprocessed using an implementation of MasSPIKE [288] implemented for BUPID [525]. MasSPIKE provided substantially more sensitive deconvolution of tandem MS, allowing detection of low abundance fragment ion peaks indicative of residually glycosylated peptide backbone fragments [293].

3.2.6 Glycomics data analysis

Glycomics data were deconvoluted and deisotoped using DeconTools (version 1.0.5501) [244,266]. Scan-specific monoisotopic peaklists from LC-MS analyses were matched against a combinatorically generated glycan composition database, using the following parameters:

Monosaccharide lower and upper bounds:

Hexose	3 - 10
HexNAc	2 - 8
Fucose (deoxyhexose)	0 - 5
NeuAc	0 - 4

Rules:

Fucose (dHex) <= HexNAc – 1
NeuAc <= HexNAc – 2

Database generation and matching were performed using a 10 ppm mass error tolerance using GlycReSoft [359].

LC-MS data preprocessed by Decon2LS were clustered using a method similar to that reported previously [359], described in Method 1. The sorted order invariant ensures that the most abundant peak will be the seed mass for a group, which is expected to have the most accurate mass. These peak clusters were merged following a new algorithm to capture scan density, signal to noise ratio, charge state multiplicity, monoisotopic mass, and isotopic cluster consistency.

```
SequentialNeutralMassClusteringByDescendingIntensity(MSPeaks, GroupMassErrorTolerance):
Clusters := ∅
SortedMSPeaks := SortByIntensityDescending(MSPeaks)
For each Peak in SortedMSPeaks:
      PeakGroup := FindNearestMass(Peak, Clusters, GroupMassErrorTolerance)
      If PeakGroup != null:
            AddToGroup(Peak, PeakGroup)
      Else:
            CreateNewGroup(Peak, Clusters)
Return Clusters
```

Method 1: Sequential Clustering

3.2.6.1 MS1 Database Search

After clustering MS peaks and computing cluster features, clusters were compared against a hypothesis database as constructed above. Each entry in the hypothesis database was searched against the peak group clusters, selecting matches which are within some ppm error tolerance threshold t, for our tests, 10 ppm. Adduction and incomplete derivatization can cause a glycan precursor mass to shift by fixed amounts, which are taken into account

3.2.6.2 Peak Group Evaluation

We converted peak groups into feature vectors comprised of charge state count, scan density, scan count, total volume, A to A+2 peak intensity estimate error, centroid scan estimate error, and average signal to noise. We then constructed a label vector to reflect the match label, and fit a logistic regression model to find peak group parameters which distinguish groups that match the hypothesis database, and those that do not. We then score all peak groups with this model, independent of whether or not they matched a

database entry. Because MS^1 information alone is insufficient for identification, this model application does not predict the quality of a match, but a peak group's likeness to peak groups which matched database entries. This metric evaluates the completeness of the database, intended to identify peak groups which look like real compositions but that were not found in the database.

3.2.7 Glycoproteomics data analysis

A naïve glycopeptide hypothesis is a database of glycopeptides from an *in-silico* enzymatic digest with up to k missed cleavages of one or more reference sequences of putative glycoproteins with all combinations of selected post-translational modifications, combined with one or more *N*-Glycan compositions. We used pyteomics [182] to provide our dictionary of proteolytic enzymes. Post-translational modifications may either be fixed, guaranteed to be present, or variable, occurring or not at each viable site as defined by a subset of rules selected from database such as Unimod [103]or ProteinProspector's MS-Digest [32]. If a glycopeptide has more than one glycosylation site, combinations of glycans are considered, up to a limit of g glycosylations on a single peptide. A hypothesis is only constructed for MS^1 compositions, so exact placement of variable modifications and glycosylations need not be determined exactly. Only unique glycopeptides are considered for each reference protein.

After construction, the hypothesis is searched against clustered LC-MS data as for glycomics. The results are used to construct an MS^2 hypothesis.

3.2.7.1 MS^1 Informed Hypothesis Building

An informed hypothesis begins with an mzIdentML file [278] and a library of glycan compositions, like a hypothesis or database search results set from a glycomics experiment. Each protein and associated peptidoforms are extracted from the mzIdentML file, filtered for just those which contain glycosylation sites following biosynthetic-rules, user specifications or other experimental information and then all combinations of peptidoforms and glycans are generated as in the naïve case. Because post-translational modifications and their locations have already been identified, no position information is removed. This may cause the informed hypothesis to be larger than the naïve hypothesis at the MS1 level, but each match at this level will translate to exactly one sequence searched at MS2 instead one or more. As in the naïve case, the informed MS1 hypothesis is evaluated in the same way.

3.2.7.2 MS^2 Naïve Hypothesis Building

After having performed MS^1 database searches, an MS2 hypothesis can be constructed. Each hit glycopeptide composition with present but un-localized PTMs is expanded into a set of glycopeptide sequences where PTM locations are exactly specified. This results in a combinatorial expansion of glycopeptide sequences, many of which may be indistinguishable without high coverage. This process included positioning the glycosylation attachment site(s) for each isoform. We generated theoretical peptide b and y ion series for each sequence, as well as precursor stub ions and oxonium ions. A precursor stub ion is composed of the mass of the intact peptide backbone with portions of the *N*-glycan core attached. Oxonium ions are low mass diagnostic saccharide ions

generated by fragmentation of the glycan units on glycoconjugates during collisional dissociation, analogous to immonium ions generated by peptide dissociation.

3.2.7.3 MS^2 Informed Hypothesis Building

Each glycopeptide matched at the MS1 level from an informed hypothesis is translated to exactly one glycopeptide sequence. Informed databases for integrated omics were generated as described previously [291], except that the only glycans which were included from the glycomics analysis had an aggregated abundance greater than twice the mean abundance of all glycan matches.. Tandem mass spectra were identified by first recalculating the precursor ion monoisotopic mass and charge, followed by a database search procedure with a precursor mass tolerance of 10 ppm. For each glycopeptide which fell within the acceptable mass range, theoretical product ions for b, y, b plus HexNAc, y plus HexNAc, and intact peptide plus incremental losses of saccharide units (known as "stub ions", were assigned. The tandem spectra were deconvolved and peaks were matched for each theoretical product ion with an error tolerance of 10 ppm, constructing a glycopeptide-spectrum match (GSM).

3.2.7.4 MS^2 Database Search

For each sequence in the hypothesis, for each tandem spectrum whose precursor neutral mass is within t_1 ppm mass error window, we compute fragment matches. Theoretical fragment masses are compared against the neutral masses for each tandem spectrum peak with an error tolerance of t_2 ppm. For our datasets, $t_1 = 10$, $t_2 = 20$. We first search for oxonium ions, evidence that the scan included a glycopeptide, and if none

are found, no further matching is performed. Multiple glycopeptides regularly match the same spectrum, as in proteomics, so spectral evidence is assigned to the best matching cases. Under the simple scoring model, this is just the set of glycopeptides which match the most peaks in the spectrum.

3.2.7.5 MS^2 Scoring

Under a simple scoring regime, which does not assume any instrument-specific information, we built the score for each glycopeptide match from the set of all spectrum matches assigned to it. We computed three scores and combined them into a single summary $MS2\ Score$ following user parameterized weights depending upon the amount of backbone information that was expected, as shown in Method 2.

For each GSM, a score based upon peptide backbone coverage and presence of stub ions was computed and scaled to be between 0.0 and 1.0. For each spectrum where multiple glycopeptides could be assigned, all GSMs tied for the highest score were added to a results set.

For each experiment, a forward-sequence (target) and a reverse-sequence-with-valid-sequon (decoy) database was searched, and a q-value without PIT as described by Käll *et al.* [281], was computed for the paired result sets. Target spectra which had a q-value < 0.05 were selected for inclusion in the final reported results.

All of the glycomics and glycoproteomics data analysis, including database generation, matching and FDR estimation was performed using components of a prototype GlycReSoft data analysis pipeline, developed by Joshua A. Klein.

```
MeanCoverage(Glycopeptide, MatchedFragments):
bIonSeries := {0 for i in [1, Length(Glycopeptide)]}
yIonSeries := {0 for i in [1, Length(Glycopeptide)]}

For each Fragment in MatchedFragments:
      If IonSeries(Fragment) == b:
            bIonSeries[SequenceIndex(Fragment)] = 1
      Else If IonSeries(Fragment) == y:
            yIonSeries[SequenceIndex(Fragment)] = 1
Coverage := bIonSeries + Reversed(yIonSeries)
MeanCoverage := Sum({log2 1+Coverage [i] / log2 1.5*Length(Glycopeptide ) for i in [1, Length(Glycopeptide)]})
Return MeanCoverage


MeanHexNAcCoverage(Glycopeptide, MatchedFragments):
bSeriesObserved := 0
ySeriesObserved := 0

For each Fragment in MatchedFragments:
      If IonSeries(Fragment) == b && Glycosylated(Fragment):
            bSeriesObserved += 1
      Else If IonSeries(Fragment) == y && Glycosylated(Fragment):
            ySeriesObserved += 1

bSeriesEnumerated := Length(TheoreticalGlycopeptideSeries(Glycopeptide, b))
ySeriesEnumerated := Length(TheoreticalGlycopeptideSeries(Glycopeptide, y))

MeanHexNAcCoverage := bSeriesObserved + ySeriesObserved / bSeriesEnumerated + ySeriesEnumerated
Return MeanHexNAcCoverage


MS2Score(Glycopeptide, MatchedFragments):
MeanCoverage := MeanCoverage(Glycopeptide, MatchedFragments)
MeanHexNAcCoverage := MeanHexNAcCoverage(Glycopeptide, MatchedFragments)

BackboneScore := (MeanCoverage * BackBoneWeight) + (MeanHexNAcCoverage *
GlycosylatedWeight)

BackboneFactor := 1 − StubIonWeight
StubIonScore := Min(CountStubIonsObserved(MatchedFragments) / 3.0, 3)

MS2Score := (BackboneScore * BackboneFactor) + (StubIonScore * StubIonWeight)
Return MS2Score
```

Method 2: Glycopeptide MS2 scoring scheme

We implemented a false discovery rate (FDR) estimate using the target decoy method [139,635] using a reversed target sequence decoy database, comparing $MS2Score$ between target and decoy. In order to preserve as much biological context as possible, the *N*-glycan sequon in each decoy sequence retained its original orientation.

For example, a target sequence YPVLN(HexNAc)VTMPNNGK becomes GNNPMN(HexNAc)VTLVPYK. This produces deterministic, if reduced entropy decoys, which is desirable for reproducibility [197]. To accommodate for the assumption that the majority of the sequences in even a small combinatorial database are unlikely to be present, we used a "percent incorrect target" adjustment to the Target Decoy ratio calculation and also reported a *q*-value for each glycopeptide match as described by Käll and colleagues [281].

3.3 Results

The complexity of a sample mixture has implications for both the analytics and informatics methods used in data acquisition and data analysis. Assumptions about sample complexity or purity can have serious consequences on the quality of data and information generated. Because population heterogeneity is inherent in glycoconjugates, the number of glycosylated molecular forms propagates quickly with added heterogeneity from bottom-up sample preparation methods and presence of other PTMs. Additional proteins or glycoproteins in the sample matrix will impact the analytical and informatics performance. Therefore, we assessed the effects of sample matrix complexity on analytical methods and informatics. We also demonstrated how integration of different data domains improves estimation of complexity and keeps false discovery rates in check.

3.3.1 Effects of assumptions about glycoprotein search space

Typical assumptions in glycoprotein characterization are related to sample purity. Most search spaces do not account for presence of contaminating proteins/glycoproteins in the sample analyzed. For biotherapeutic formulations, these contaminants are often components of the culture system used for production of the biomolecule. For biological samples derived from animals or plants, co-purified molecules from the sample matrix are commonly present. Commercially available standard glycoproteins are commonly used for method development and validation and are usually assumed to be pure; however, while the target glycoprotein may be the most abundant molecule present in the preparations, it is not safe to assume that it is the only (glyco)protein present. We assessed purity of commercially available AGP and transferrin glycoprotein standards using proteomics. Figure 16A shows the total number of proteins identified in a proteomics analysis of each purified glycoprotein. While all standard glycoproteins originated from serum or plasma, each has its own unique subset of contaminants, which are co-purified despite the enrichment and purification steps.

3.3.2 Evaluation of search space construction methods: naïve versus informed

In addition to contaminating proteins present in the sample, presence of PTMs, non-specific cleavages or missed-cleavages in bottom-up sample preparation methods can also confound the data analysis. In the absence of a proteomics experiment, it is almost impossible to gauge the number and type of protein variants in the sample. Traditional proteomics search methods require the user to specify a list of modifications

and proteolytic cleavage types to be considered, which also limits the search space to assumptions [141].

The degree to which the naïve approach estimates accurately the false discovery rate (FDR) depends on the assumptions; unfortunately, in the absence of data, there is no way of evaluating the correctness of these assumptions. As a result, the naïve method may over- or under-estimate search space size. Underestimation arises when a glycoprotein is incorrectly assumed to be pure. This situation leads to false confidence in the results when FDR is calculated for glycopeptides derived from the pure glycoprotein. The presence of contaminating glycoproteins in the sample means that the search space is in reality much larger than assumed and that the FDR calculated from the assumed search space underestimates the rate of false identifications.

Overestimation of search space size is a result of including too many glycoproteins and glycoforms in the search space and may lead to inability to differentiate true from false identifications; in such cases many high quality tandem mass spectra will have spurious matches against components in the search space. In summary, the validity of an FDR calculation depends on the assumptions of search space size. To the extent that the search space accurately estimates sample complexity, the FDR correctly expresses the confidence of glycopeptide assignments.

Error-tolerant searches and *de novo* sequencing tools have gained popularity in proteomics [102,351,509,615]. These allow the user to analyze data without bias from pre-defined search parameters to maximize coverage of the proteome with respect to the different molecular forms present. Glycosylation adds a large molecular weight and

considerable macro and micro-heterogeneity, in addition to other non-glycan modifications present on proteins or peptides. Glycosylation also introduces glycosidic product ions in tandem MS, not found for unglycosylated peptides, complicating the identification process. As a result proteomics search engines that perform error-tolerant searches are unable to handle complex glycosylation as a modification. For similar reasons, emerging software tools that are capable of assigning glycopeptide compositions are unable to perform large error-tolerant searches for other PTMs and therefore rely on user specified search spaces. We assert that the most efficient circumvention to this problem is to use the identifications from proteomics data to inform the search space for glycoproteomics.

To prove this, we used mzIdentML proteomics search results from a combination of error-tolerant and *de novo* searches and the confidently identified peptide molecular forms to build a search space for all possible glycopeptidoforms present in our samples. Similarly, to define our glycoform search space, we performed glycomics searches on our released glycan pools and iteratively combined the results with the proteomics identifications that contained glycosylation sites. Figure 16 compares the sizes for a naïve search space (B), built for glycoprotein mixture 3 based on assumptions about glycoproteins and PTMs present *versus* an informed search space (C), built using the proteomics search results to define the range of glycoproteins and PTMs present in the sample and a further constrained informed search space (D) that included glycomics data. Strikingly, the proteomics and glycomics informed search space in panel D has less than half the number of total theoretical glycopeptides compared to the naïve search space in

panel B. While proteomics information helps identify all contaminating glycoproteins in the sample, it can lead to an explosion of the search space size, which can render the glycopeptide search process inefficient; glycomics helps restrict the number of glycoforms to those that are actually detected in the glycomics analyses. It is therefore important to use both proteomics and glycomics in glycopeptide search space construction.

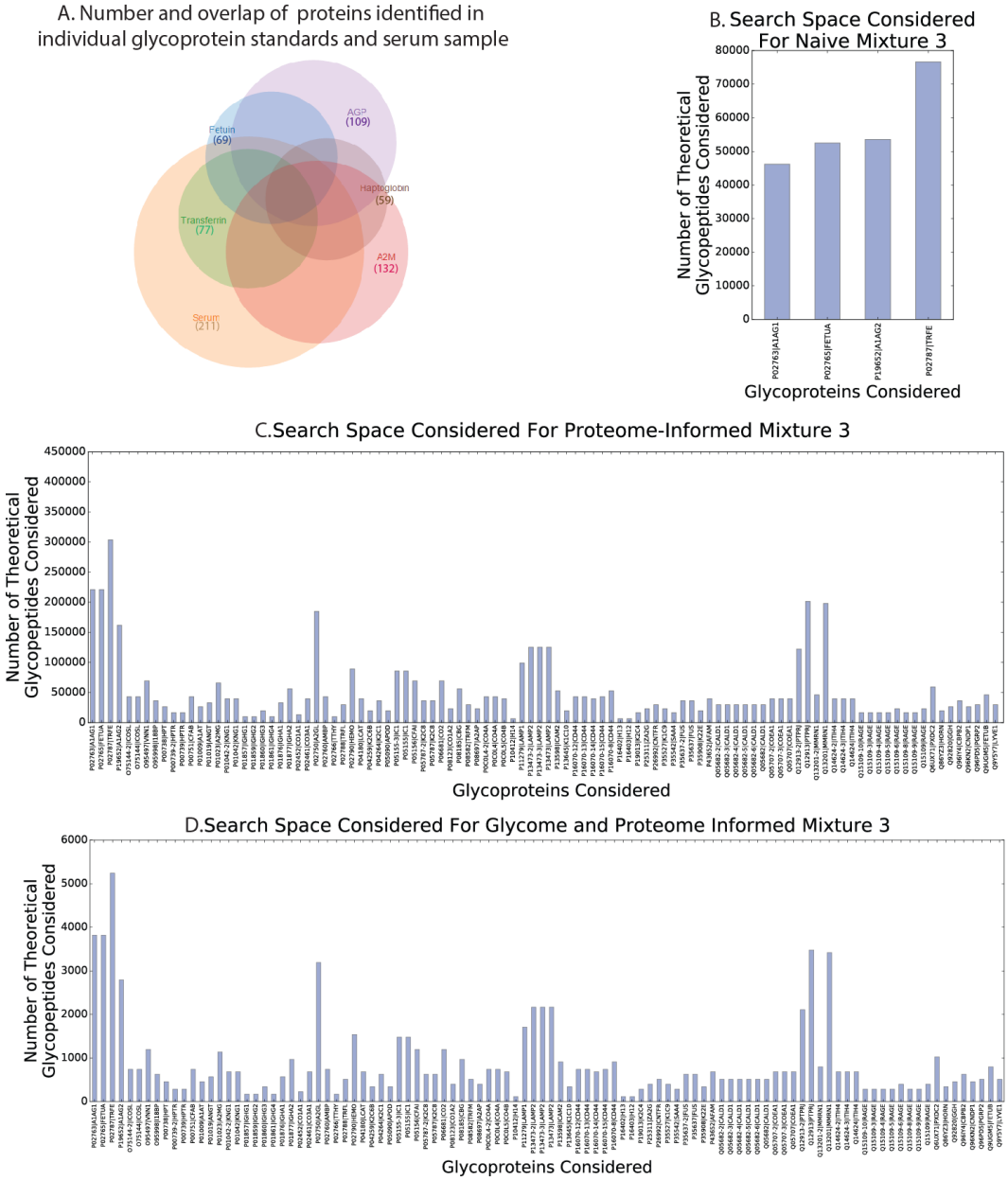


Figure 16: Glycoproteomics search space

(A) Venn diagram showing number and overlap of proteins identified from proteomics analyses of individual serum glycoproteins and albumin and IgG depleted serum. Only proteins identified at a 1% FDR with 2 or more unique peptides identified were included. Comparison of a naïve search space (B) with proteomics informed (C) and proteomics + glycomics informed (D) search spaces for glycoprotein mixture 3. X-axes show the Swissprot identifiers for glycoproteins included in the hypotheses and Y-axis shows number of glycopeptides per glycoprotein considered. Glycoprotein standards AGP (A1AG), fetuin (FETUA) and transferrin (TRFE) are shown on the left-most side of the X-axes.

3.3.3 The site-specific glycoproteomes of transferrin and α1-acid glycoprotein

Figure 17 shows the base-peak chromatogram (A) for size-enriched glycopeptides from commercially sourced human transferrin. Chromatographic peaks corresponding to glycopeptides were identified based on the extracted ion chromatogram for the HexNAc and Hex-HexNAc oxonium ions (B and C). We interpreted the glycopeptide tandem mass spectra using a search space containing putative transferrin glycopeptides; this analysis identified three of the EIC peaks as transferrin glycopeptides. Next, we acquired proteomics data on the PNGase F deglycosylated sample and found that hemopexin, a protein with five *N*-glycosylation sequons, was in the sample, in addition to transferrin. When we added hemopexin to the search space, we identified abundant chromatographic peaks corresponding to hemopexin glycopeptides.

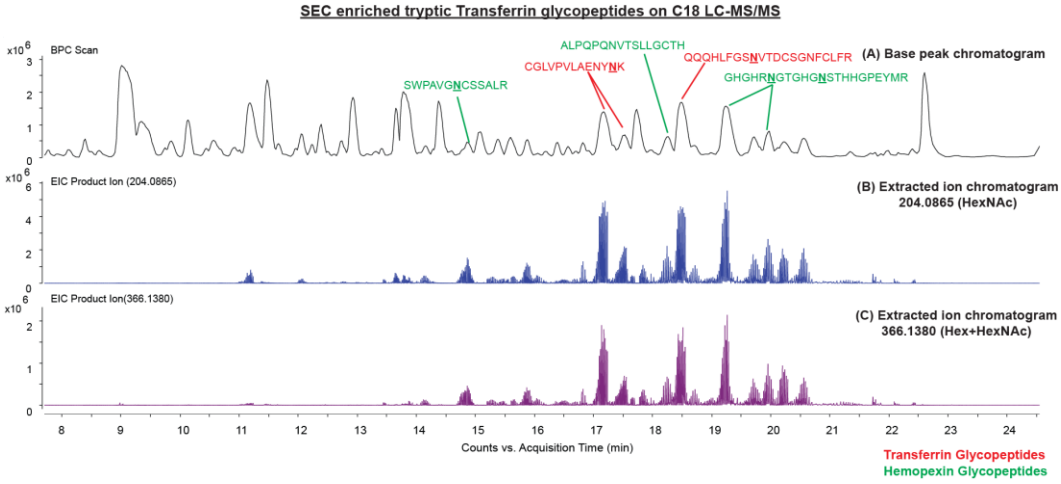


Figure 17: Base peak chromatogram of enriched transferrin glycopeptides stacked over extracted ion chromatograms for saccharide oxonium ions, generated by MS/MS, indicating the presence of glycopeptides in the eluting peaks

α1-Acid glycoprotein contains two protein isoforms with five sequons, each [293,398,530:1-]. To construct a naïve glycopeptide search space, we allowed 0-2 missed tryptic cleavage sites, methionine oxidation, and 0-1 sites of deamidation for asparagine-

containing glycopeptides. Table 4 shows the naïve search space size calculated assuming pure AGP, AGP plus haptoglobin, and AGP plus haptoglobin plus transferrin. We then compared the results from our informed *versus* naïve search spaces.

The search space size ballooned 10-fold for the three protein mixture relative to the hypothetically pure AGP protein. We acquired proteomics data for human α1-acid glycoprotein (AGP) and found that in addition to the two expected isoforms, the sample contained haptoglobin, a protein that contains four *N*-glycosylation sequons, among the top scoring identified glycoproteins. For both AGP and haptoglobin, there are a large number of hypothetical glycopeptides. Tryptic haptoglobin has glycopeptides containing two sequons, thus the high mass and large number of hypothetical haptoglobin glycopeptides.

Sample	**Naïve Search Space Size**	**Informed Search Space Size**
Naïve Pure AGP	14,577	2,000
Naïve Contaminated AGP (AGP + Haptoglobin)	45,371	2,085
Naïve AGP + Transferrin (AGP + Haptoglobin + Transferrin)	115,910	25,850

Table 4: Search space sizes for AGP and AGP mixed with transferrin when calculated with assumptions about proteoforms and glycoforms (naïve) and using data from proteomics and glycomics experiments

Because of the sample complexity, we compared the ability to assign glycopeptide tandem mass spectra for a naïve search space *versus* a search space

informed by proteomics and released glycan glycomics (Figure 18A). The glycomics results for AGP are presented in Figure 19.

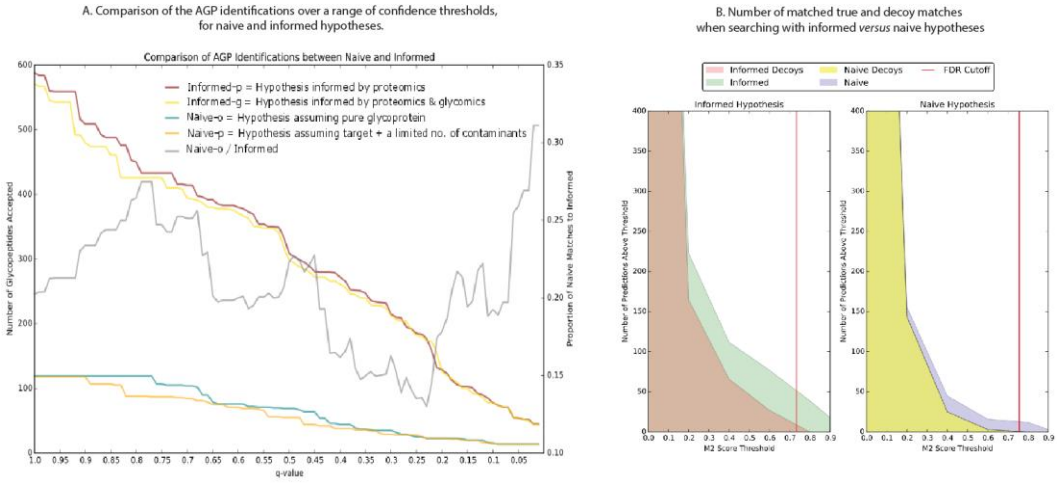


Figure 18: Comparison of the performance of naïve *versus* informed search spaces for an AGP glycoproteomics result.

(A) Number of glycopeptide matches over a range of confidence thresholds (q-value) based on glycoproteomics searches followed by FDR calculation by decoy database searches. Colored lines represent different hypotheses as indicated in the legend. (B) A comparison of number of glycopeptides and decoys matches for naïve and informed search spaces at different MS2 score thresholds.

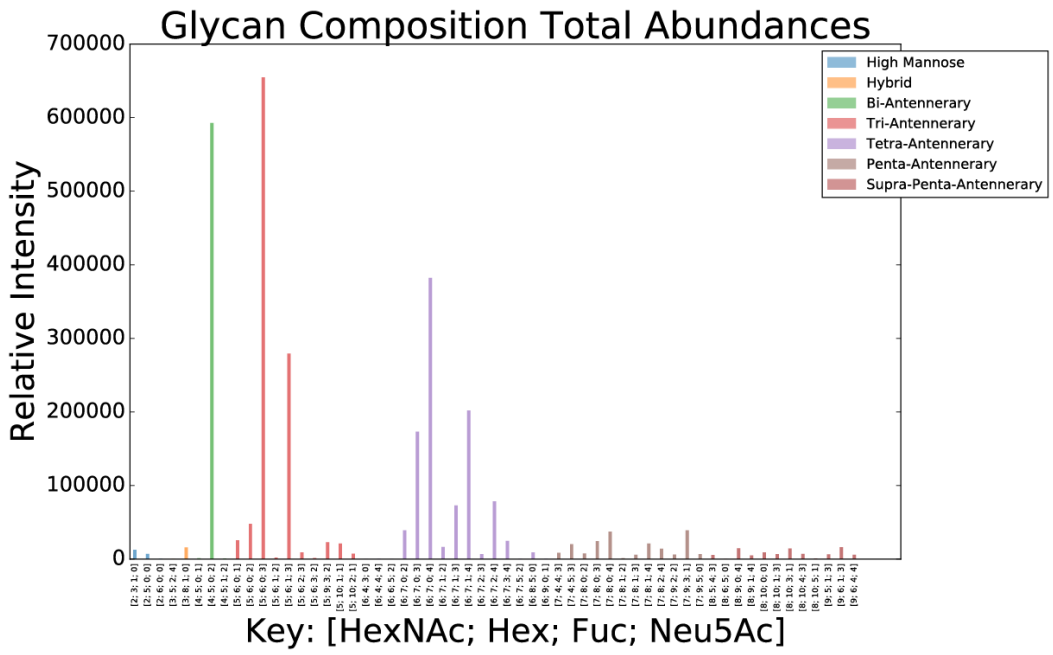


Figure 19: Glycomics results for α1-acid glycoprotein

Figure 18B, compares glycoproteomics search results using naïve *versus* informed hypotheses. The height of the curve at the red bar tells us how many predictions were accepted at our confidence threshold. The informed hypothesis retains substantially more predictions above the confidence threshold, compared to the naive hypothesis. Therefore, use of a search space informed by proteomics and glycomics data boosts our ability to assign glycopeptides from tandem mass spectral data with confidence. The profile of site specific AGP glycosylation informed by proteomics and glycomics is shown in Figure 21.

3.3.4 Confidence in assignment of site-specific glycosylation of an analyte of interest in an increasingly complex matrix.

Most glycoproteomics experiments rely on data dependent acquisition that is directly affected by the complexity of the sample matrix. With increase in matrix complexity the ability of the analytical methods to interrogate the analyte of interest decreases. The mass spectrometer may not be able to detect glycopeptides from the glycoprotein of interest depending on the relative abundance of the impurities and the dynamic range of the instrument. Moreover, even when the glycopeptides of interest are within the dynamic range, data-dependent acquisition methods spend less time on collecting tandem MS on these analytes in presence of contaminating glycopeptides/peptides.

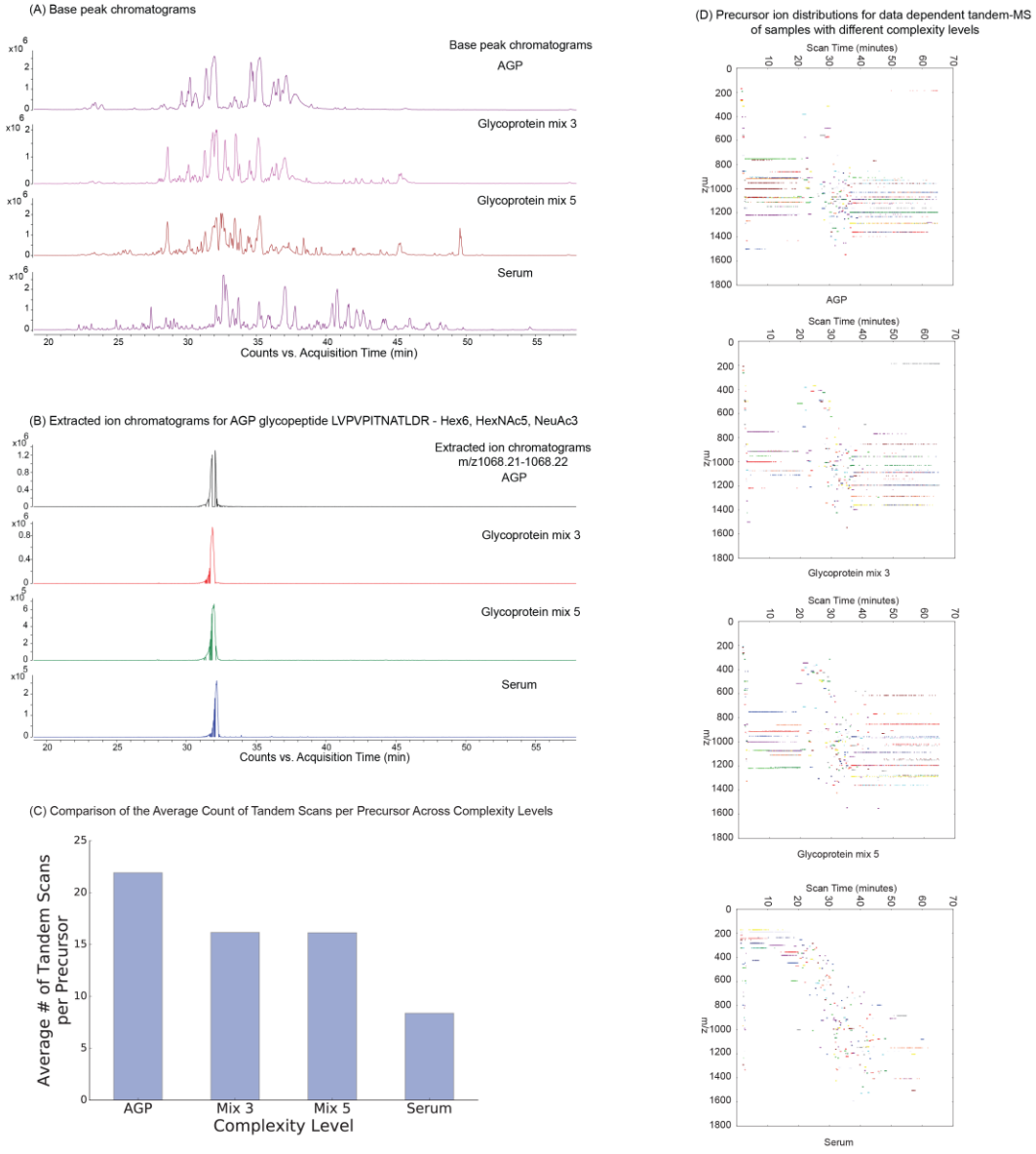


Figure 20: Effects of increasing sample complexity on analytical methods

(A) Base peak chromatograms (MS only) for samples with different levels of complexity. (B) Extracted ion chromatograms showing the abundance of an AGP glycopeptide (LVPVPITNATLDR-Hex6, HexNAc5, NeuAc3) in the different complexity levels. (C) The average number of tandem mass spectra acquired per precursor across samples of different complexity. (D) Distributions of precursors selected for tandem MS across a LC-tandem MS experiment for samples with scaling complexity levels.

We constructed mixtures with scaling complexity by mixing together fixed quantities of different glycoprotein samples. Figure 20 shows how data quality for the

AGP glycopeptides of interest deteriorates when complexity increases. Figure 20A shows overlaid base peak chromatograms for AGP alone, glycoprotein mix 3, glycoprotein mix 5 and serum (see the Experimental section for mixture definitions). With each complexity level, the chromatogram is visibly more crowded. The non-AGP glycopeptides in the sample can easily lead to saturation of the column binding capacity leading to a decrease in intensity of the glycoprotein/peptides of interest, as seen in Figure 20B. This could be a problem when trying to get quantitative LC-MS data, since sample matrix complexity could affect the dynamic range of the assay at both the column binding capacity and mass spectrometer detection levels. Figure 20D shows how the precursors were selected for tandem MS as sample complexity grows. The colored dots trailing along the X-axis represent tandem MS for different precursors. The long trails of colored dots in the 10-20 and 40-60 represent tandem MS on low abundance contaminants that elute over long periods of the LC gradient. In lower complexity samples, these contaminants are repeatedly selected for tandem MS due to the absence of other precursors eluting in these retention time ranges. In the 20-40 min retention-time range where most glycopeptides are found to be eluting in the AGP chromatogram, multiple tandem MS are acquired for each precursor. This also allows proper triggering of tandem MS for the precursor to be selected at the chromatographic peak apex and results in good quality tandem mass spectra. On the other hand, the distribution of precursors in glycoprotein mix 5 and serum is clearly very sparse and in most cases only a single tandem mass spectrum is being collected per precursor. For a better comparison of the complexity of these samples, in Figure 20C, we have compared the average number of tandem MS per precursor for each

complexity level. This shows that as the total number of unique precursors increases with increasing complexity, the attention or time-spent acquiring MS/MS on each precursor diminishes. This becomes a serious limiting factor for low abundance glycopeptide glycoforms, which are either not selected for tandem MS or for which poor quality tandem MS are acquired. Thus, as the complexity increases, in glycoprotein mixes and serum, the mass spectrometer is unable to handle the increase in co-eluting peptides/glycopeptides and acquires fewer tandem mass spectra per precursor, resulting in poor data quality and lower identification rates.

The effect of increasing complexity is reflected in the depth of glycoform coverage for AGP glycosylation sites as shown in Figure 21. The most comprehensive assignment of glycopeptides comes from the AGP only sample. As mixture complexity increases, the overall scores diminish, resulting in fewer assignments falling above the scoring threshold. This is due to the decreasing number of tandem mass spectra per precursor ion resulting from increased mixture complexity and the finite scanning speed of the instrument. The AGP precursor ion abundances decreased with mixture complexity. This is likely due to a combination of ionization effects and the limited capacity of the glycopeptide enrichment step, which limits the amount of total protein/peptides that are loaded on to the analytical column. Since the bar plots represent the sum of fragment ion abundances for all scans matching a glycopeptide, the precursor that have the most high-scoring spectral matches also show up as the most abundant matches on the bar plots.

It was important to integrate data from proteomics and glycomics domains, for accurate estimation of the FDR and for setting confidence thresholds for accepting matched glycopeptide spectra for all complexity levels, as already shown in Figure 18. While a tandem MS may be acquired for all eluting glycopeptides, good spectral quality is essential for confident matching above the FDR thresholds. For some of the AGP glycosylation sites shown in Figure 21, glycoprotein mix 5 appears to have more glycopeptide matches with glycoprotein mix 3. This is generally due to differences in the profile of co-eluting contaminating peptides that affect tandem MS acquisition.

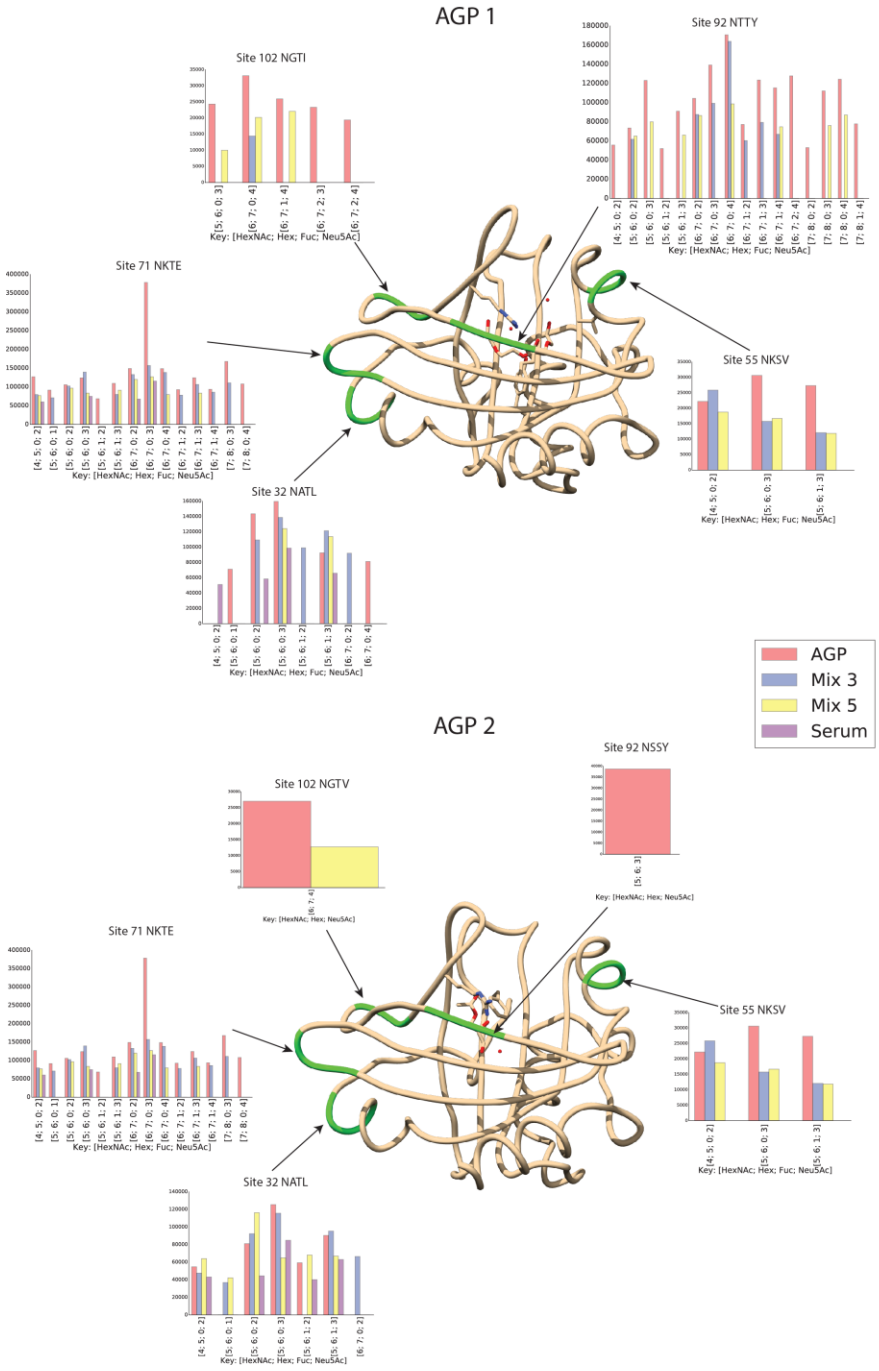


Figure 21: Grouped bar plots showing identified site-specific AGP glycoforms in samples of different complexity levels and in depleted serum

The different glycoforms for each glycosylation site are listed on the X-axes while the Y-axes indicate summed abundances of all fragment ions in the matched glycopeptide spectra. Different colored bars are specific for pure AGP, glycoprotein mixtures or depleted serum as indicated in the legend.

3.4 Discussion

For most proteomics and glycoproteomics experiments, multiple sample clean-up and fractionation steps are used for reducing complexity before online LC-MS. While sample fractionation increases glycopeptide coverage, overall matrix complexity impacts the analysis of our molecule of interest. Since, heterogeneity of glycoforms is a major challenge in the analysis of glycoconjugates, the problem of defining the correct boundaries for data analysis is amplified many fold over non-glycosylated molecules. It is therefore important to account for the matrix contribution during data analysis so that assumptions about sample purity do not affect confidence in results. A better understanding of the matrix comes from integrating different data dimensions and prevents erroneous assignment of contaminant data to the analyte of interest, thereby lowering the false discovery rates. Since this step helps determine the nature of contaminants present in the sample matrix, it also guides choice of the appropriate methods required for further sample purification. This can greatly impact successful method development in an industry setting for analysis of biologicals [434]. Collectively, these results demonstrate the effect of assumptions about sample complexity on the false discovery rates and therefore confidence in analyses. Analytical method development benefits greatly from proper data analysis and results visualization to identify any issues that limit performance.

While method performance is heavily dependent on instrument hardware and software, it is essential to establish standard methods and datasets to use as benchmarks for tracking any changes in instrument and method suitability both within an experiment

and over long-term use. The human α1-acid glycoprotein is reasonably complex, widely available and can be handled by most LC-MS analytical platforms used for glycoproteomics. AGP therefore serves as a good standard for method development, validation and suitability tracking.

We demonstrated that predetermination of sample complexity is of utmost importance, when dealing with complex biological matrices. In the absence of empirically derived information, the multiplicity of molecular forms in glycoproteomics can easily overwhelm analytical methods that are not setup keeping sample and matrix complexity in mind. The ramifications are even more serious on the informatics end. Contaminating proteins, sample processing artefacts, unusual PTMs, mutations and above all glycoforms can confound glycoproteomics data and severely inflate false discovery rates. Measurement of the proteome of glycoprotein sample constrains the search space size by defining the peptide variants actually present. This allows inclusion of contaminating glycoproteins detected in the proteomics data that were not included in the naïve search space, while preventing addition of unnecessary contaminants based on assumptions about matrix that may not be present in the sample, thus fixing the proteomics search space. In addition, profiling of the released glycans reduces the list of glycans present in the sample and efficiently constrains the search space size. Thus, proteomics and glycomics data help develop an efficient search space that prevents over- or underestimation of the glycosylated molecular forms present in a sample. We present a new glycoproteomics data analysis pipeline GlycReSoft that maximizes informatics power by combining glycomics and proteomics data in mining the glycoproteomes.

Naïve searches fail to eliminate unreasonable hypotheses because of the complexity of real life glycoprotein samples. As a result, the level of false identifications cannot be determined accurately. Informed searches allow unbiased definition of search spaces, thus helping determine confidence levels in assignments, accurately.

We conclude that sample complexity at which acceptable rates of false identifications occur need to be defined using data. We suggest use of AGP as a glycoprotein standard of complexity suitable to test the glycoproteomics workflows. This will help harmonize efforts in the glycoproteomics community and enable a better comparison of data quality and controls. Further, we suggest controls in which AGP is doped into the sample matrix to demonstrate that acceptable rate of false identification results. This will help the experimentalists identify and eliminate any flaws in their analytical strategies.

3.5 Acknowledgements

Funding was provided from NIH grants P41 GM105603 and R21 177476. Thermo-Fisher Scientific provided access to the Q-Exactive Plus mass spectrometer used in this work.

Chapter 4: Influenza virus glycobiology and structural studies using mass spectrometry

(Part of this work is published in [291]).

4.1 Introduction

Influenza A virus (IAV) remains a major cause of animal and human mortality and morbidity, with no effective broadly neutralizing vaccines available. The IAV envelope-glycoprotein hemagglutinin (HA) binds sialic acids on respiratory airway surface glycoconjugates and facilitates viral entry into host cells, thereby leading to infection. The envelope-glycoprotein neuraminidase (NA) cleaves sialic acids to allow newly formed virions to escape the host cell membrane. Because the surface proteins undergo rapid mutation, it has not been possible to develop broadly neutralizing vaccines against IAV. In particular, IAV strains undergo amino acid mutations to accumulate new *N*-glycosylation sequons during seasonal circulation in the human population [49,86,546,561]. Thus, based primarily on genetic sequencing evidence [72,225,345], glycans appear to shield HA antigenic sites from host antibody recognition, as part of the evolutionary antigenic drift [111,118,340,616]. While this putative shielding effect of glycans apparently improves the ability of the virus to escape antibody neutralization, the virus must compensate for loss in receptor avidity from steric interference caused by the glycans proximal to the sialic acid binding sites on the HA head [117,541].

T-cells of the adaptive immune system reach the lung approximately five days after IAV infection [259]. During this time, viral replication and lung damage is

mitigated by the innate immune system. Among the soluble innate immune inhibitors, the calcium-dependent lectin inhibitors (known as β-inhibitors) bind to carbohydrates on pathogen surfaces. For seasonal IAV strains, surfactant protein-D (SP-D) is the most important soluble innate immune factor present in bronchial lavage fluid [215,565]. The activity of this lectin depends on the presence of high mannose *N*-glycans on viral HA; IAV strains lacking glycosylation on the HA head region resist neutralization by SP-D. These include the pandemic strains of H1N1 1918 and 2009, the H3N2 of 1968 and the H2N2 of 1957 [218,274,427]. Thus, the pathogenicity of pandemic IAV correlates with ability to escape neutralization by SP-D [215,224,427]. Generally speaking, as seasonal strains acquire additional *N*-glycosylation sequons they show increased susceptibility to SP-D, as they circulate in humans. While IAV mutates to avoid antibody neutralization by accumulating sequons, the mechanisms whereby viruses maintain fitness despite increased binding to SP-D and other lectins of the innate immune system remain poorly understood [19,213,214,218,436,516,565].

The H3N2 subtype has circulated in humans since 1968 and remains the major seasonal IAV health threat. The H1N1 subtype was introduced into the human population with the disastrous 1918 pandemic and was supplanted by the H2N2 subtype in 1957. The H1N1 subtype was re-introduced into the human population in 1977 and displayed several additional *N*-glycosylation sequons on HA not present on earlier strains [508]. In 2009, it underwent antigenic shift by incorporation of genetic elements from the swine and avian sources to emerge as a new pandemic subtype. The pattern of *N*-glycosylation

sequons of pandemic H1N1 2009 more closely resembled the 1918 pandemic H1N1 (minimal glycosylation) than that circulating seasonally prior to 2009 [561].

Despite the central role of glycosylation in viral evolution, there remains a paucity of structural information on HA glycosylation. A recent study of engineered variants of the Hong Kong 1968 H3N2 subtype demonstrated that exclusively high-mannose *N*-glycans occupy HA sites 165 and 246 [17]. Although genetic sequence analyses identify the creation or disruption of *N*-glycosylation sequons, they provide no information on the glycan site occupancy, structure and micro-heterogeneity. In addition, crystal structures typically show only the chitobiose core and do not define glycan antennae [487,504,571]. We therefore built a predictive model for seasonal IAV fitness and antigenicity that incorporates (i) data on changes in site-specific HA glycosylation that occur as the virus evolves to escape SP-D neutralization and (ii) the HA structural features that regulate glycan microheterogeneity at each site that influence binding to SP-D. We demonstrate the value of complete information on the mature glycosylated HA structures to inform better understanding of the evolution of IAV in response to immune system pressure.

In addition to defining changes in site-specific glycosylation on HA in response to immune-selection pressures, a parallel study was performed to understand the role of sialylated glycans present on porcine SP-D (pSP-D), in potentiating the neutralization of Influenza in pigs. It has been shown that pSP-D possesses unique structural features in its CRD that result in substantially stronger anti-IAV activity as compared to SP-Ds from other animal species that have been characterized to date, including human SP-D (hSP-D) [137]. pSP-D is known to have a highly sialylated *N*-linked glycan on its CRD that

generates a dual mechanism of interaction with IAV and that makes pSP-D unique amongst all other SP-Ds in being a β-inhibitor as well as a γ-inhibitor against IAV infections. In addition, it was shown that the type of sialic acid (SA)-linkage present on pSP-D is important for inhibition of IAV strains that have a corresponding receptor-binding specificity [538]. The presence of this sialoside on pSP-D makes it a potent inhibitor of IAV strains, such as PR-08, that lack glycans on the head of HA and are therefore not inhibited efficiently by SP-D from other species, but are inhibited by pSP-D since the HA recognizes and binds sialylated glycans on the pSP-D [538]. To better understand the mechanisms involved in inhibition of IAV by pSP-D, it is important to elucidate the structural features of pSP-D, with a particular focus on the sialylated glycan present in the CRD. We performed mass spectral characterization of recombinant porcine NCRD (RpNCRD). These data will support a study comparing the structural features and binding and neutralization activities of sialylated and asialo pSP-D that is being performed at Utrecht University by Dr. Martin van Eijk and coworkers.

4.2 Experimental procedures

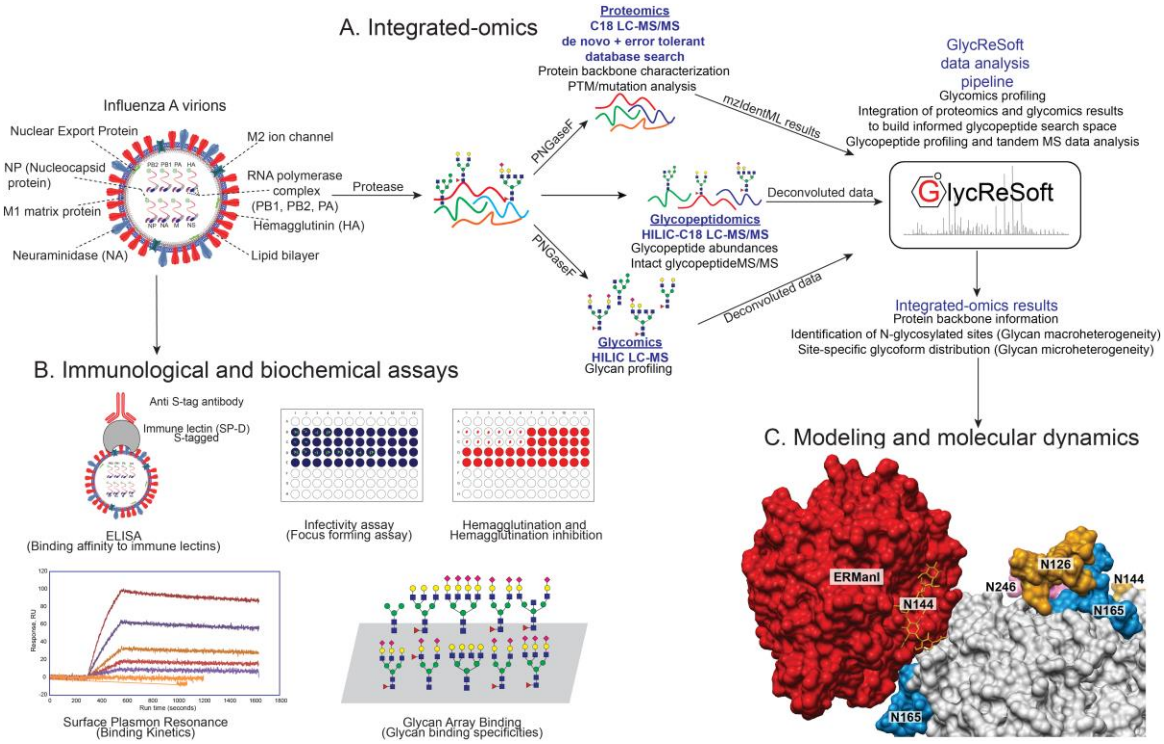


Figure 22: Experimental Workflow for IAV analysis

A: Integrated-omics: Workflow for acquiring and combining proteomics, glycomics and glycopeptidomics information to enable confident assignment of site-specific glycoforms; B: Immunological and biochemical assays used for correlating changes in virus glycosylation with bioactivity; C: Modeling and molecular dynamics simulations driven by structural information from integrated-omics analyses to understand structural basis for host-virus interactions and glycan processing at important sites. Graphic shows ERManI interacting with the glycan at Asn 144 on the head region of a Phil-82 hemagglutinin trimer.

All mass spectrometry data were acquired by Kshitij Khatri, and the data acquisition and analysis were directed by Prof. Joseph Zaia. Mass spectrometry/Omics data were analyzed by Kshitij Khatri and Joshua A. Klein. The IAV preparations used in this work were grown by Mitchell R. White. Bioassays were performed by Kshitij Khatri and Mitchell R. White, under the guidance of Prof. Kevan L. Hartshorn and Prof. Joseph Zaia. Dr. Oliver C. Grant performed the computational glycobiology and molecular modeling work under the guidance of Prof. Robert J. Woods. Dr. Erika C. Crouch

(Washington University School of Medicine) provided the Surface Plasmon Resonance data, shown in Figure 28**.** Dr. Martin van Eijk at Utrecht University generated the RpNCRD used in this work. Prof. D.F. Smith carried out the CFG glycan array analyses at the National Center for Functional Glycomics (Emory University School of Medicine) funded by the NIH grant P41GM103694. Dr. Michael J. Rynkiewicz and Prof. Barbara A. Seaton participated in discussions pertinent to this project and provided helpful suggestions in designing IAV and porcine SP-D experiments.

4.2.1 Virus preparation

The following IAV strains were used in the study:

Phil-82: A high growth reassortant of A/Philippines/2/82(H3N2) and A/Puerto Rico/8/34(H1N1). The seed stock was a gift from Dr. E. Margot Anders, University of Melbourne. This strain consisted of an H3 hemagglutinin and N2 neuraminidase from A/Philippines/2/82, while the remaining virus components were from A/Puerto Rico/8/34.

Phil-BS: A mutant version of Phil-82 that was selected by growing Phil-82 in the presence of bovine serum, found to be resistant to neutralization by bovine conglutinin, as described previously [19,213,214,435]. The seed stock was from Dr. E. Margot Anders, University of Melbourne.

PR-08: The mouse-adapted A/Puerto Rico/8/34(H1N1) strain. The seed stock was from Dr. Jon S. Abramson, Wake Forest University.

All viruses were grown in the chorioallantoic cavity of 10 day old embryonated chicken eggs. Allantoic fluid was harvested after incubating for 48 hours and insoluble debris were pelleted by centrifugation at 1000x *g*. The virions were then pelleted at 135,000x *g* and the pellets were purified on a sucrose density gradient followed by dialysis against phosphate buffered saline (PBS).

RpNCRD was expressed in and purified from HEK293 cells by Dr. Martin van Eijk, as described previously[136].

4.2.2 Experimental Design and Statistical Rationale for IAV analyses

As shown in Figure 22, virus samples were subjected to proteolytic digestion. Proteolytic samples were divided into three fractions for proteomics, glycomics and glycoproteomics. Proteomics and glycomics fractions were subjected to PNGase *F* glycan release. All three fractions for each sample were subjected to triplicate LC-MS analyses, as described in sections below. Three technical replicates of LC-MS acquisitions were used to calculate mean analyte abundances and standard deviations in measurements. Human α1-acid glycoprotein (Sigma-Aldrich, St. Louis, MO), a widely studied standard glycoprotein, was used as a control for all sample preparation and verifying LC-MS instrument performance.

4.2.3 LC-MS analyses of IAV

4.2.3.1 Sample preparation

Virus preparations were subjected to membrane disruption, reduction, alkylation and proteolysis. Virus preparations containing approximately 60 µg total protein were

dried down in a centrifugal evaporator and re-suspended in LC-MS grade methanol (Fisher Scientific), followed by sonication for 10 minutes in a water bath to disrupt viral membranes. Samples were dried and re-suspended in 50% 100 mM ammonium bicarbonate and 50% 2, 2, 2-trifluoro ethanol (TFE) (Sigma-Aldrich, St. Louis, MO) (60 µl). 200mM dithiothreitol (DTT) (2 µl) was added to the samples and heated at 95^0C for 30 mins to reduce disulfide bonds. Iodoacetamide (10 µl, 200 mM in water) was then added to the samples and incubated for 1 hour in dark, at room temperature. To quench excess iodoacetamide, another (2 µl) of 200 mM DTT was added to the samples and incubated for one hour at room temperature. LC-MS water (300 µl) was added to the samples to dilute TFE to less than 5% of the total volume and 100 mM ammonium bicarbonate (200 µl) was added to raise pH to 7.5. Sequencing grade Trypsin or Chymotrypsin (Promega Corp., Madison, WI) were added to the samples and incubated overnight at 37^0C, for proteolysis. Proteolytic peptides and glycopeptides were split into aliquots for performing ^{16}O-proteomics, ^{18}O-proteomics (for site-occupancy analysis) and glycoproteomics. ^{16}O-proteomics and ^{18}O-proteomics samples were deglycosylated using PNGase F (New England Biolabs, Ipswitch MA) in $H_2{}^{16}O$ or $H_2{}^{18}O$ respectively, while the glycoproteomics samples were incubated in the deglycosylation conditions without PNGase F. Released glycans from deglycosylated samples were isolated using C18 spin columns (Pierce Biotechnology, Rockford, IL) and pooled (^{16}O and ^{18}O) for glycomics analyses, and deglycosylated peptides were eluted and collected separately for proteomics analyses. As a control for spontaneous deamidation at inherently non-glycosylated Asparagine residues, the glycosylated samples (no PNGase *F* treatment)

were also analyzed, using the proteomics workflow. Mass spectrometry data were acquired on each sample as three technical replicates, unless otherwise stated.

4.2.3.2 Proteomics analyses

Deglycosylated peptides were subjected to LC-MS/MS using a Waters™ NanoAcquity™ nano-flow chromatograph (Waters Corp., Milford, MA) mounted with a Waters™ Xbridge™ reversed-phase column (150 µm x 100 mm) packed with 1.7µm BEH C18 resin and a Waters™ trap column (180 µm x 20 mm) packed with 5µm Symmetry™ C18 stationary phase. The chromatography was performed online with a Q-Exactive Plus™ mass spectrometer (ThermoFisher Scientific™, San Jose, CA), operated in positive ion mode. The mass spectrometer was mounted with an Advion NanoMate (Advion Inc., Ithaca, NY) source for introduction of the LC eluate by nano-ESI. The source was operated at 1.7 kV with transfer-capillary temperature maintained at 250^0C and S-Lens RF level set at 55. Mass spectra were acquired in the Orbitrap mass analyzer with 1 microscan per spectrum for both MS and MS/MS. Resolving power for MS and MS/MS were set at 70,000 and 17,500, respectively. Tandem MS data were acquired in parallel with MS, on the top 20 most abundant multiply charged precursors, with higher energy collisional dissociation (HCD) at normalized collision energy of 27V. Precursors were isolated using a 1.4 Th window and dynamic exclusion of 10 seconds was enabled during precursor selection.

4.2.3.3 Glycomics analyses

Released *N*-glycans were desalted using a Superdex™ Peptide™ 3.2/30 size exclusion column on a Beckman™ Gold chromatograph. Desalted *N*-glycans were analyzed in triplicate using negative mode hydrophilic interaction liquid chromatography (HILIC) –MS with a constant acetonitrile post-column make-up-flow on an Agilent™ 6520 LC-MS system with chip-cube™ nanoESI source, as described previously [496].

4.2.3.4 Glycoproteomics analyses

Entire tryptic digests were subjected to triplicate analyses using chip-LC-MS/MS with online HILIC enrichment and C18 separation on an Agilent™ 6550 LC-MS system with chip-cube™ ionization source as described previously [293] and by C18 LC-MS/MS without any enrichment on the Q-Exactive Plus™ instrument as described for the proteomics samples. Chymotryptic samples were analyzed only using C18 LC-MS/MS on the Q-Exactive Plus™.

4.2.4 Data analysis

4.2.4.1 Proteomics

Proteomics data were subjected to PEAKS SPIDER (*de novo* + database + error tolerant PTM) searches in PEAKS [615] Studio 7.5 (Bioinformatics Solutions Inc., Waterloo, ON) against a combined Uniprot [346] and Influenza A virus protein database [493]. The FASTA sequence database used for searches has been included in the data repository. A 50 ppm error tolerance for the precursor (MS^1) and 0.1 Da mass error tolerance for fragment ions (MS^2) were specified. A maximum of 2 missed cleavages per

peptide were allowed for the database search. Trypsin specificity was defined as cleavage after Arg and Lys, when not followed by a Pro. Chymotrypsin cleavage specificity was after Phe, Leu, Met, Trp and Tyr residues, when not followed by a Pro. Cysteine carbamidomethylation was specified as a fixed modification; deamidation (Asn) and oxidation (Met) were specified as variable modifications. ^{18}O labeling of deglycosylated asparagine was also used as a variable modification in case of ^{18}O-proteomics samples. After a regular database search, an error-tolerant PTM search was also performed searching for a larger subset of modifications from Unimod [103] and any amino-acid substitutions. The final results were a combination of database search, *de novo* search and the error-tolerant search. False discovery rates (FDR) were calculated using a decoy-fusion approach in PEAKS 7.5, as described previously [615]. Identified peptide-spectrum-matches with -10logP value of 15 or higher were kept, at a FDR threshold of 0.1%.

4.2.4.2 Site-occupancy analysis

Site-occupancy analysis was performed on Phil-82 for glycan modeling and molecular dynamics. While ^{18}O-proteomics was performed to define glycan site-occupancy, in cases where multiple asparagine and glutamine residues existed on peptides containing the *N*-glycosylation sequons, spontaneous deamidation of these residues in presence of $H_2{}^{18}O$, confounded the data. Therefore, ^{16}O-proteomics results were instead used to define site-occupancy information.

From the proteomics search results, peptides spanning glycosylation sites were selected and extracted ion chromatograms were integrated to generate chromatographic peak areas for paired analyses of unoccupied peptides bearing glycosylation sequon in glycosylated samples and the same peptides in PNGase F deglycosylated sample. Average peak areas of three replicate analyses were compared after normalization using a high-confidence peptide from a different protein identified across runs. Site-occupancy for sites 8 and 22 were calculated from proteomics results of chymotryptic digests, while for the remaining sites, proteomics results from tryptic samples were used.

Percent site occupancy was calculated as:

$$Site\ occupancy\ = \frac{Abundance\ (deglycosylated) - Abundance\ (glycosylated)}{Abundance\ (deglycosylated)} \quad \textbf{Equation 20}$$

4.2.4.3 Glycomics

Negative mode HILIC-MS data were deconvoluted and deisotoped using the THRASH algorithm in DeconTools/Decon2LS (version 1.0.5501) [244,266]. The deconvoluted/deisotoped peaklists were matched against a theoretical composition hypothesis containing human and avian *N*-linked glycans from GlycomeDB [432,433], using GlycReSoft [359]. Up to 2 formate adducts were allowed while searching for glycan compositions, with a match error tolerance of 20 ppm in the theoretical masses. Abundances of different adducted forms were grouped together and reported as mean of three technical replicates.

4.2.4.4 Glycoproteomics/Integrated-omics

Glycopeptide LC-MS/MS data were deconvoluted and deisotoped using DeconTools [266] for MS^1 spectra. Data were converted to mzML format [355] using the MSConvert tool in Proteowizard (version 3.0.7692) [290] and the MS^2 spectra were then deconvoluted using MasSpike [288]. Deconvoluted data were analyzed using an in-house version of GlycReSoft and validated manually. Source code for GlycReSoft is publically hosted on GitHub (https://github.com/GlycReSoft2). Naïve and informed glycopeptide hypotheses were created for mining glycopeptide data. For naïve hypothesis generation, theoretical digests of glycoproteins were generated for combining *N*-glycosylation sequon containing peptides with *N*-glycan compositions from GlycomeDB [432] glycan databases.

For informed hypothesis generation, proteomics search results exported in mzIdentML 1.1 format, were used. Hemagglutinin peptides that contained Asn from the *N*-linked glycosylation sequon, with -10logp values of 20 or higher were included in the hypotheses. Peptide variants were combined with glycoforms identified in glycomics data, to build the glycoproteomics hypotheses. MS^1 deconvoluted data were searched against glycopeptide hypotheses, using a match error-tolerance of 20ppm. Theoretical fragment ions were generated for the MS^1 matches and then searched against the MS^2 deconvoluted peak lists, using a 20 ppm error tolerance. Features searched for in MS^2 data of glycopeptides included oxonium ions, peptide backbone ions, peptide backbone ions with an attached HexNAc and stub-glycopeptide ions [293]. A glycopeptide spectrum match was required to have mono or disaccharide oxonium ions. Automated

glycopeptide analysis results have been included in the data repository. MS^1 scores were calculated as described previously [359]. MS^2 scores are representative of number of fragment ions found (peptide backbone and stub ion coverage) [293]. q-values depict minimal FDR threshold at which the identification is accepted, as described by Käll *et al.* [281]. For cases, where more than one glycopeptide composition matched a given precursor and fragment ion spectrum, only the highest scoring match was retained for reporting in the results. Glycopeptide analysis results represent site-specific glycoform abundances for glycopeptides assigned confidently based on MS^2 spectra by GlycReSoft and validated manually. Chymotryptic glycopeptide data were manually analyzed for resolving glycoforms on Asn8 and Asn22, on Phil-82 and Phil-BS and integrated with site-specific glycosylation data from tryptic samples processed using GlycReSoft.

4.2.5 Bioassays

Biochemical and immunological assays were performed to correlate changes in site-specific glycan distributions in IAV strains with virus fitness and susceptibility towards host innate immune system. SP-D is a collectin produced and secreted by lung epithelial cells in humans and other mammals. SP-D is also known to inhibit IAV by binding its surface glycoproteins HA and NA [215,435,224,531]. For this study, a recombinant full length human SP-D (rhSPDII) that exists as dodecamers was prepared, as described previously by Nikolaidis *et al.* [106,397]. In addition to the full-length SP-D, recombinant neck and carbohydrate recognition domain (NCRD) from SP-D were also used. These NCRDs lack the *N*-terminal domain and the extended collagen domain

present in the full-length SP-D and as a result form homotrimers instead of higher order multimers. The wild-type recombinant human NCRD (huNCRD) has been shown to bind glycans but lacks antiviral activity due to absence of cooperative binding effect from the multiple heads [108,217,397]. Gain-of-function mutants R343V and double mutant D325A+R343V (D+R), which have been shown to possess increased virus binding and neutralizing activity [106,397,107,109,179], were also used in bioassays. Mutant names are indicative of amino-acid substitutions flanking the lectin site in the huNCRD protein.

4.2.5.1 Hemagglutination assays

Hemagglutination and Hemagglutination inhibition (HI) assays were performed in round bottom 96 well plates, using PBS containing calcium and magnesium (PBS++) as a diluent. Viral titers were first measured using hemagglutination assays, performed as described previously [216], using type O- red blood cells from human donors. HI assays were performed with recombinant SP-D preparations described above. Inhibition or disruption of agglutination was measured using serial dilutions of SP-D preparations.

4.2.5.2 Enzyme-linked immunosorbent assay (ELISA)

96-well plates were coated overnight at 4^0C with 0.1µg/well of Phil-82, Phil-BS and PR-08. Serial dilutions of NCRDs were added to control wells. Wild-type huNCRD and mutant version D+R were used. Plates were washed 3x with 100µL of PBS++ (PBS containing 0.9mM CaCl2, 0.52mM MgCl2 and 0.16mM $MgSO_4$) and blocked with blocking buffer (Superblock, Pierce). To the wells coated with IAV, different doses of NCRD, matching those in control wells, were added and incubated for 30 min at 37^0C.

Plates were then washed 3x with wash buffer. To each well, horseradish peroxidase conjugated anti-S-protein antibody was added and incubated at 37^0C. Following incubation, plates were washed 3x with PBS++ and 50µL of 1-step ELISA substrate (Pierce) was added. After 20 min incubation at room temperature with the ELISA substrate, the reaction was stopped by addition of 50ul 1N sulfuric acid. Plates were read for absorption at 450nm wavelength with wavelength correction at 540nm.

4.2.5.3 Fluorescent Focus assay (IAV infectivity assays)

Infectivity assays were performed using human bronchial/tracheal epithelial (BTE) cells (Lifeline cell technology, Walkersville, MD). BTE cell monolayers were grown to about 70% confluency in 96 well plates. IAV preparations were pre-incubated with SP-D preparations for 30 min at 37^0C, to allow SP-D binding and inhibition of HA. The BTE cell monolayers were infected with these IAV preparations for 45 min at 37^0C, in PBS++. Cells were tested for presence of IAV after 18 hours of virus addition using an anti IAV nucleoprotein monoclonal antibody (Millipore) and fluorescence detection, as described previously [566].

4.2.5.4 Surface Plasmon Resonance

Surface Plasmon Resonance (SPR) experiments were performed as described previously by Crouch *et al.* [109]. Recombinant human SP-D NCRD (Neck and carbohydrate recognition domain) preparations were immobilized using N-terminal His tags on NiNTA, with the binding domain oriented upward. Binding of Phil/82 and

Phil/82/BS B-HA were measured against recombinant wild type huNCRD and double mutant D+R.

4.2.5.5 Glycan array binding

Intact virions were labeled with Alexa 488 succinimidyl ester conjugated dye (Life Technologies/Thermo Fisher Scientific, Carlsbad, CA) using the manufacturer's protocol. Post-labeling, virions were dialyzed against 1x PBS. HA titers were measured and the virus preparations were appropriately diluted to adjust for differences in HA titers before binding experiments. Glycan array binding was performed using standard protocols at the Consortium for Functional Glycomics [51,429,490] Protein-Glycan Interaction Core at Emory University, using array version 5.2 (https://glycopattern.emory.edu/structures/view?version=5.2). Binding was measured as fluorescent units of bound fluorescently labeled influenza virions.

4.2.6 Computational glycobiology

4.2.6.1 3D structure generation

A homology model for the Phil/2/1982 HA trimer was generated in Modeller [457,154,144] using PDBIDs 2YP3, 2YP7, 1HA0, 4KVN, 4WE5, 4O58 and 4WA1 as templates. In order to decrease the time required to perform a molecular dynamics (MD) simulation, a smaller structure was generated from the HA0 trimer that contained only residues F83 to D86 of the stalk and D57 to S270 of the head group. Glycans were attached to the head group using a customized version of the glycoprotein builder available on GLYCAM-Web [297], with the following modifications. When attaching the

N-glycan to an Asn sidechain, the χ1, χ2, ψ and φ torsion angles were adjusted to match observed rotamers reported for *N*-linked glycoproteins in the PDB [418]. Any vdW overlaps were relieved by adjusting the glycosidic torsion angles within normal bounds [400]. In all cases, the Asn sidechain conformation that was observed to be most common in the PDB was selected, provided that it could be attached without irreconcilable steric overlaps with the protein surface. $Man_9GlcNAc_2$ glycans were added to each of the five *N*-glycosylation sites on the HA headgroup; at sites Asn63, Asn126, Asn144, Asn165 and Asn246.

4.2.6.2 System preparation for simulation

Hydrogen atoms were added to the head group structure, with the protonation states of the ionizable side chains being assigned by tleap [73]. The structures were placed in a periodic box of approximately 35,000 TIP5P waters [347] to provide an 8 Å buffer between the glycan and the edge of the periodic box.

4.2.6.3 Energy minimization and solvent equilibration

Energy minimization of all atoms was performed for 20,000 steps (10,000 steepest decent, followed by 10,000 conjugate gradient). The energy-minimized structures were equilibrated at 300 K under nPT conditions for 400 ps, with 5 kcal/mol-$Å^2$ Cartesian restraints on the solute heavy atoms.

4.2.6.4 System equilibration and production MD

After solvent equilibration, the heavy atom restraints were removed, and 1 ns of equilibration MD followed by a 500 ns production simulation were performed. The Cα atoms of the terminus residues were restrained (5 kcal/mol-$Å^2$).

4.2.6.5 Energy minimization and molecular dynamics (MD) simulations

All simulations were performed with the CUDA implementation of PMEMD [186,458] in the Amber14 software suite [73]. The GLYCAM06-j [297] and Amber14SB parameters [73] were employed for the carbohydrate and protein portions, respectively. A Berendsen barostat with a time constant of 1 ps was employed for pressure regulation, while a Langevin thermostat with a collision frequency of 2 ps^{-1} was employed for temperature regulation. A nonbonded interaction cutoff of 8 Å was employed. Long-range electrostatics were treated with the particle-mesh Ewald (PME) method [115]. Covalent bonds involving hydrogen were constrained with the SHAKE algorithm allowing a time step of 2 fs [455].

4.2.6.6 Post processing

An ensemble of 1,000 conformations for the glycosylated Phil-2-1982 HA glycoprotein were taken at regular 500 ps intervals of the MD simulation using cpptraj [443].

4.2.6.7 Assessment of glycan accessibility to ERManI

A new crystal structure of ERManI, kindly provided by Dr. Kelley Moremen's laboratory, includes a full high mannose glycan in the binding site. This allowed

assessment of the accessibility of the glycan to ERManI by superimposing the ERManI-bound Manβ1-4GlcNAcβ residues onto each *N*-glycan on the H3 head group. This was performed for each *N*-glycan over 1,000 snapshots taken at regular intervals from the MD simulation. If surface overlaps greater than ~36 $Å^2$ (equivalent to a buried Carbon atom) were observed, the glycan was deemed to be inaccessible for that particular timepoint of the simulation.

4.2.6.8 Assessment of glycan accessibility to SP-D

A 3D structure has been determined previously for unliganded human SP-D (PDBID 1B08), as well as porcine SP-D in complex with Manα (PDBID 4DN8) [136]. Despite structural differences in the carbohydrate binding sites of the porcine and human proteins, the relative spacing of the binding sites is identical, with a linear distance of approximately 45 Å between each Manα binding site. By monitoring the distances between the non-reducing terminal Manα residues in each *N*-glycan over the course of the simulation, it was determined which sets of three *N*-glycans were positioned for trimeric interaction with SP-D (sets that were all 43-47 Å apart for any of the 1,000 snapshots).

4.2.7 LC-MS analysis of RpNCRD

4.2.7.1 Identification of glycosylation site and assessment of glycosylation site-occupancy on RpNCRD

Theoretical recombinant porcine SP-D protein sequence:

GITALRQQVETLQGQVQRLQKAFSQYKKVELFPNGRGVGEKIFKTGGFEKTFQDAQQVCTQA GGQMASPRSETENEALSQLVTAQNKAAFLSMTDIKTEGNFTYPTGEPLVYANWAPGEPNNNGGSSGAE NCVEIFPNGKWNDKACGELRLVICEF

Glycosite identification and site occupancy determination were performed using ^{18}O labeling of the glycopeptide and analysis of the intact glycoprotein. Glycosylation site was also confirmed by analysis of intact glycopeptides by LC-tandem-MS.

Chymotrypsin was purchased from Promega Corporation, Madison WI. PNgase F was purchased from New England Biolabs, Ipswitch, MA. ^{18}O-Water was purchased from Cambridge Isotope Laboratories, Andover, MA. All other chemicals were obtained from Sigma-Aldrich, St. Louis, MO, unless otherwise stated.

RpNCRD samples were subjected to reduction (dithiothreitol, DTT), alkylation (iodoacetamide) and chymotryptic digestion in the presence of 2,2,2-trifluoroethanol as denaturant. Reactions were stopped by heating at 95^0 C for 5 min and addition of trifluoroacetic acid to lower the pH. The masses of the resultant peptides were immediately analyzed using a Bruker Ultraflextreme matrix-assisted laser desorption/ionization (MALDI) MS (Bruker Daltonics, Billerica, MA) using 2,5-dihydroxybenzoic acid (DHB) as matrix. The samples were then dried down in a centrifugal evaporator. Chymotryptic RpNCRD peptides were deglycosylated using PNGase F in the presence of $H_2{}^{18}O$ or $H_2{}^{16}O$. Aprotinin was then added to the reactions to inhibit any residual chymotrypsin activity. Samples were desalted using Pierce C18 spin columns and analyzed using MALDI MS with DHB as matrix.

The labeled (deglycosylated) RpNCRD peptide samples were subjected to LC-MS/MS using a Nano-Acquity™ chromatograph and a Symmetry® 5µm C18 trapping

column (180µm x 20mm) and BEH™ 30 C18 analytical column (100µm x 100mm, Waters Corporation, Milford, MA), coupled to a Thermo® LTQ-Orbitrap™ via an Advion® Nanomate™, nano-ESI robot. Samples were concentrated on the trapping column at a flow rate of 4µL/min for 4 min, with 98% Buffer A, 2% Buffer B (Buffer A: 99% Water, 1% Acetonitrile, 0.1% Formic acid v/v; buffer B: 99% Acetonitrile, 1% Water, 0.1% Formic acid). A linear gradient of buffers A and B was used to elute the peptides from the analytical column, starting at 5% buffer B to 40% buffer B over 40 min, at a flow rate of 0.5µL/min. ESI was performed using the NanoMate, at 1.7kV capillary voltage and 150^0C capillary temperature, for introduction of samples into the mass spectrometer. Spectra were acquired in the positive ion mode in the Orbitrap using *m/z* range 300-2000 at a resolution of 60,000. AutoMS/MS was simultaneously performed on the top 3 most abundant precursors from the MS1 scan. Multiply charged precursors were fragmented in the LTQ, using CID (Collision-induced dissociation) as the activation type. Normalized collision energy was set at 30%, with Helium as the collision gas and the minimum signal cutoff set at 50000, 30000 and 15000 for the top3 precursors, respectively. Lock mass was active during all spectral acquisitions. Dynamic precursor exclusion was enabled for MS/MS runs to allow only two MS/MS spectra for each precursor to be acquired. Three additional analytical replicates were prepared and analyzed for both O-18 and O-16 labeled samples, with the same LC-MS setup, where tandem MS spectra were acquired in the LTQ instead of the Orbitrap.

LC-MS/MS data was analyzed using Thermo Xcalibur 2.2 SP 1.48 and Thermo Proteome Discoverer 1.3.0.339 (Thermo Fisher Scientific). Protein and peptide searches

were performed using an in-house licensed MASCOT server (Matrix Science, Boston, MA) against the mammalian subset of Swiss-Prot database (SwissProt_2013_01.fasta, version 2.4), allowing a maximum of 5 miscleavages using Trypsin/Chymotrypsin as the enzyme, with precursor and fragment mass tolerance set at 10 ppm and 0.8 Da, respectively. Carbamidomethylation was set as fixed modification at Cysteine residues. Deamidation (NQ) and Dioxidation (M) were set as dynamic modifications. Custom modifications were added to the MASCOT server for identification of the labeled deglycosylated peptide after PNGase F *N*-glycan release for both O-16 and O-18 modifications. All searches included decoy database search with target False Discovery Rate set at 0.01 (strict) and 0.05 (relaxed).

Precursor ion abundances for the peptides verified with MS/MS were used for quantification. Peak areas for extracted ion chromatograms of labeled and unlabeled peptides were integrated in Xcalibur and compared to give site occupancy information. Considering the possibility for spontaneous deamidation, chromatograms were extracted for masses corresponding to both deamidated and non-deamidated forms of the peptides (unlabeled) containing the sequon. It was noticed that most of the unlabeled/unglycosylated sequon containing peptides had undergone deamidation. To confirm this, a deamidation control was analyzed using LC-MS under the same conditions, except the mass spectra were acquired in the LTQ. This control sample was a chymotryptic digest of RpNCRD, without any PNGase F *N*-glycan release. The sample showed that almost all of the unglycosylated sequon containing peptides were present in the deamidated form.

A second approach was used to quantify the site occupancy of RpNCRD. The RpNCRD was reduced and alkylated using DTT and iodoacetamide and then desalted using C18 ZipTips (Millipore). The protein samples were spotted on a ground steel MALDI target and spectra were acquired in linear TOF mode at 35% laser power using DHB as the MALDI matrix. A total of 4 analytical replicates were analyzed. The unglycosylated and the glycosylated protein populations were resolved in the MALDI mass spectra as shown in Figure 32. Site occupancy was determined by comparing peak areas for spectral peaks corresponding to glycosylated and unglycosylated protein species. Peak areas were calculated by pixel counting using image processing software.

4.2.7.2 Glycan profiling of RpNCRD and determination of the degree of sialylation

Glycan profiling/composition analysis was performed using LC-MS on released *N*-glycans from RpNCRD. RpNCRD samples were subjected to reduction and alkylation followed by tryptic digestion. PNGase F release was performed as described above in the presence of aprotinin. Released *N*-glycans were fractionated using C18 spin columns (Pierce) and then subjected to HILIC (hydrophilic interaction liquid chromatography)-MS in the negative ion mode using an Agilent® 6520 accurate mass Q-TOF mass spectrometer (Agilent Technologies, Santa Clara, CA) coupled with an Agilent 1200 series HPLC stack capable of delivering micro-flow with a flow-splitter and a chip-cube for online LC separation and Nano-ESI sample introduction into the mass-spectrometer. A custom-made HILIC-chip was used for chromatography, packed with 5µm TSK-gel

Amide 80 HILIC stationary phase (Tosoh Bioscience). The outer valve of the chip was used to deliver a make-up-flow as described by Staples, et al [495,496].

Data were analyzed manually using Agilent Mass Hunter Qualitative Analysis software (Version B.05.00) and glycan compositions were assigned to *m/z* values using an in-house excel based tool, using a 20ppm mass error tolerance. The extracted ion chromatograms for monoisotopic *m/z* values were integrated to get peak areas and relative ion abundances for each glycan composition. Three analytical replicates were analyzed and average relative ion abundances were plotted only for the glycans that were present in all three replicates. The adducted ion forms and charge states were combined for each glycan composition.

4.2.7.3 Determination of sialic acid linkage on RpNCRD

A chemical derivatization method reported for determination of NeuAc linkage by Reiding *et al.* was used [438]. The chemical reaction leads to formation of different products, as shown in Figure 5, depending on the sialic acid linkage and causes specific mass shifts in glycan mass, which can be easily detected using LC-MS.

The chemical derivatization was performed by incubating released *N*-glycans from RpNCRD in 20µl of 0.25M1-ethyl-3-(3-(dimethylamino)propyl)-carbodiimide (EDC) and 0.25M 1-hydroxybenzotriazole (HOBt) in ethanol, at 37°C for 1 hour. Subsequently, 20µL acetonitrile was added to the samples and incubated in -20°C for 15 minutes. Glycan standards with known linkage for NeuAc were purchased from V-Labs (Covington, LA) and used as controls for the reaction. EDC and HOBt were purchased

from Sigma-Aldrich(St. Louis, MO). The samples were desalted using PD MiniTrap G-10 columns (GE Healthcare, Piscataway NJ), as per the manufacturer's instructions and dried in a centrifugal evaporator. Data acquisition and analysis were performed as described for native glycans.

4.3 Results and Discussion

We modeled the changes in HA mature structure that occur as an H3N2 subtype evolves during repeated growth in eggs in the presence of bovine serum. This process rendered the virus resistant to conglutinin and mannose binding lectin in serum, and to SP-D as well [215]. Owing to the error prone nature of virus replication and the inclusion of host proteins in the virion structure [262], we acquired deep proteomes of H3N2 before and after development of resistance to SP-D. We then profiled the range of glycans present on the virions and used this information to drive determination of confident site-specific glycosylation from glycoproteomics data. For HA, the combination of peptide proteolytic variants, other PTMs including oxidation and deamidation, and glycosylation yield an extremely large number of theoretical structures. To address this, we combined proteomics, glycomics and glycoproteomics analyses to produce site-specific glycosylation assignments.

We and others have developed liquid-chromatography-mass spectrometry methods with sufficient power for analysis of glycopeptides from complex glycoproteins [14,121,241,293,523]. In addition, improvements to software tools further enhance the ability to analyze complex glycoproteins [113,580]. We built an exhaustive glycopeptide

bioinformatics search-space that made no assumptions regarding peptide backbone, proteolytic variants, PTMs, or glycome present in the IAV samples. We then defined the mature structures of H3N2 HA and the changes associated with escape from SP-D neutralization. We used viral infectivity assays, glycan array binding, lectin binding and other biochemical methods to demonstrate the changes in biological interactions and activities associated with evolution of glycosylation structure among IAV subtypes and strains. In order to understand how particular glycoforms are generated, we modeled the interactions between mature glycosylated HA and ER class I α-mannosidase (ERManI), a key enzyme in the glycan biosynthetic pathway. We then defined the key glycans mediating interactions of SP-D with HA using structure modeling and molecular simulation studies.

4.3.1 Proteomics

Influenza A shows host species-specificity in infection governed by HA binding to sialylated glycans present on host cell-surface receptors. Avian-strains prefer binding to NeuAc-α-2,3-Gal, while human-strains prefer binding to NeuAc-α-2,6-Gal [100,511,503,502]. Minor changes to the protein sequence correlate with a switch in sialic-acid binding preferences leading to changes in receptor specificities and altered susceptibility towards antibodies and lectins [174,274,445,521].

We performed deep proteomics on deglycosylated IAV digests. Despite the fact that IAV contains only 8 gene segments, corresponding to roughly 11 proteins, the virions include host proteins as integral parts of its architecture [262]. In addition, host

proteins may co-purify with the virions. We therefore accounted for all proteins and glycoproteins present in the samples in order to maximize the confidence of site-specific glycosylation assignments. The error-prone manner of IAV replication allows incorporation of mutations in viral proteins that may affect glycosylation sites and neighboring residues. Our deep proteomics analyses of virus preparations covered all glycosylation sites and identified glycopeptide variants that included other PTMs and incomplete proteolytic cleavages. We therefore used an unbiased bioinformatics search space for glycopeptide assignment.

In addition to identifying peptide backbone variants, we identified a revertant population in the Phil-BS sample. The *N*-glycosylation sequon at Asn165 apparently became disrupted due to an acquired mutation leading to conversion of NVTM to NVAM in the majority of Phil-BS virion population; however, a small fraction of the population contained the wild-type sequon NVTM. Based on extracted ion chromatogram peak areas, the revertant corresponded to approximately 2% of the population. We observed no reversion at the second disrupted sequon (Asn246 NSTG) mutated in Phil-BS. To check for presence of glycosylation at the revertant sequon, we included both Phil-82 and Phil-BS protein sequences in the Phil-BS search hypothesis.

A majority of proteins identified were common among the three IAV samples (Figure 23D). Gene ontology analysis performed using STRAP [44], showed similar distributions of proteins in terms of molecular functions and cellular localization. Functional annotation of the proteins showed close similarity to the results described by Hutchison et al [262], showing a wide-variety of host-proteins involved in various

structural and molecular roles present in the virions. We observed deamidation (NQ) as the most commonly encountered PTM in the proteomics results, followed by non-specific carbamidomethylation, oxidation (M), sodium adduction and dehydration.

We identified 10 and 8 sequons for Phil-82 and Phil-BS respectively, as shown in Figure 23. All the peptide variants identified and used to construct our informed glycoproteomics bioinformatics hypotheses are shown as part of supplemental information in [291].

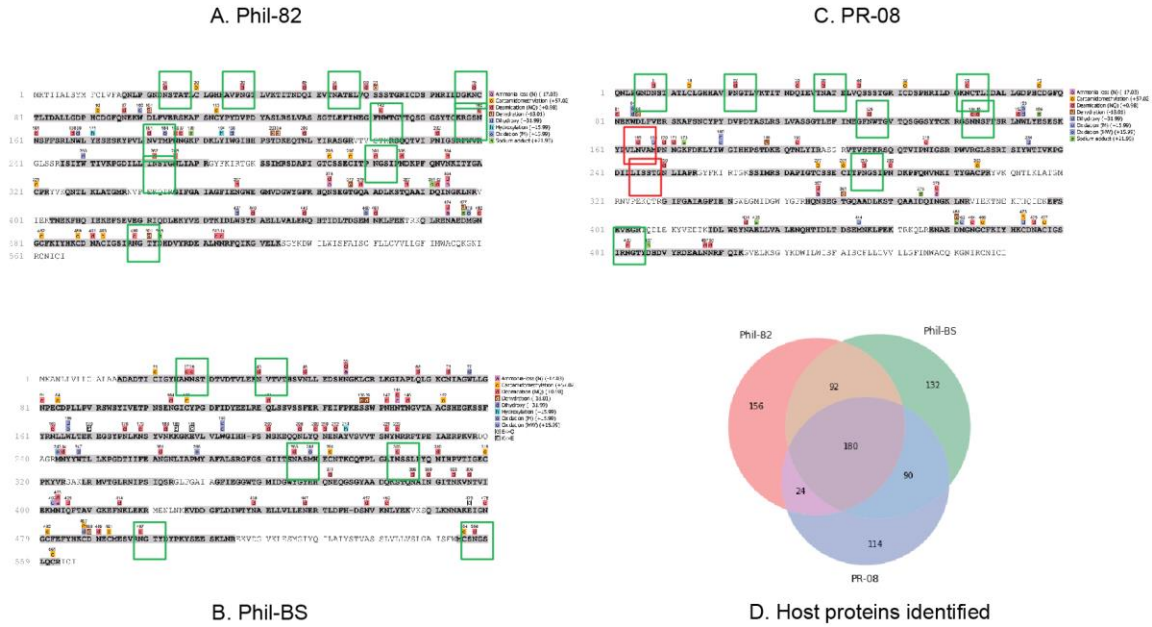


Figure 23: IAV Proteomics results

Proteomics coverage of Phil-82 (A), Phil-BS (B) and PR-08 (C) hemagglutinins. Portions of sequence covered are highlighted in grey, with colored symbols above the sequence rows representing modifications and mutations. Green boxes show N-glycosylation sequons and red boxes indicate mutated/disrupted sequons in Phil-BS; D: Total number of proteins including host-proteins identified in proteomics of the three IAV samples. Only IAV and chicken (Gallus gallus) proteins identified with 2 or more unique peptides were counted.

4.3.2 Glycan site-occupancy analysis

For modeling and molecular-dynamics studies, we determined glycosylation site occupancy of Phil-82 HA. We evaluated the NetNGlyc server [198] (http://www.cbs.dtu.dk/services/NetNGlyc/), used in published studies [40,549,275,183,17] to provide a qualitative measure of the likelihood for *N*-linked glycosylation, given the protein sequence; however, we found that NetNGlyc predictions inconsistent with our results. Table 5 compares NetNGlyc predictions and site-occupancy results from proteomics data for Phil-82. While NetNGlyc predicted no glycosylation at sites 38 and 144, we identified glycopeptides containing those two sites with high

confidence from glycopeptide tandem MS data. The data showed low site occupancy at Asn144 and high occupancy for Asn 38. These results re-emphasize the importance of analyzing both deglycosylated peptides and intact glycopeptides to minimize assumptions about samples and shifting reliance from prediction tools to empirical evidence.

Site		NetNGlyc Potential	NetNGlyc Jury Agreement	NetNGlyc Results	Site-occupancy calculated from proteomics
8	NSTA	0.7987	9/9	+++	1.0
22	NGTL	0.7153	9/9	++	1.0
38	NATE	0.4996	3/9	-	0.99
63	NCTL	0.6366	9/9	++	0.99
126	NWTG	0.5393	6/9	+	0.73
144	NNSF	0.4293	6/9	-	0.17
165	NVTM	0.7626	9/9	+++	0.99
246	NSTG	0.5880	7/9	+	0.99
285	NGSI	0.6716	9/9	++	0.95
483	NGTY	0.5207	8/9	+	0.95

Table 5: NetNGlyc predictions for glycosylation site occupancy and site-occupancy from proteomics data for Phil-82

4.3.3 Glycomics

We used glycomics profiling to define the range of glycans present on virions precisely. The glycan profiling results for Phil-BS (38% high-mannose) showed an approximately 10% decrease in relative abundances of high-mannose type *N*-glycans compared to Phil-82 (48%) (Figure 24A). PR-08 displayed a very small proportion of *N*-

glycans (ca. 6%) occurring as the high mannose type. Detailed comparison of abundances for individual glycan compositions (Figure 24B, C, D) showed a marked decrease in the relative abundances of immature $Hex_9HexNAc_2$ and $Hex_{10}HexNAc_2$ glycan compositions in Phil-BS. The glycan topologies shown in Figure 24 are speculative. Glycomics profiling using LC-MS provides saccharide compositions only – topological/structural information cannot be derived from these experiments. Detailed results from glycomics analyses have been included in the data repository.

From the proteomics results, the primary difference between Phil-82 and Phil-BS was the disruption of two *N*-linked glycosylation sequons at positions 165 and 246. The glycomics results implied that a large proportion of the glycans occupying these two sequons were high-mannose type. To confirm this theory, we carried out intact glycopeptide analysis.

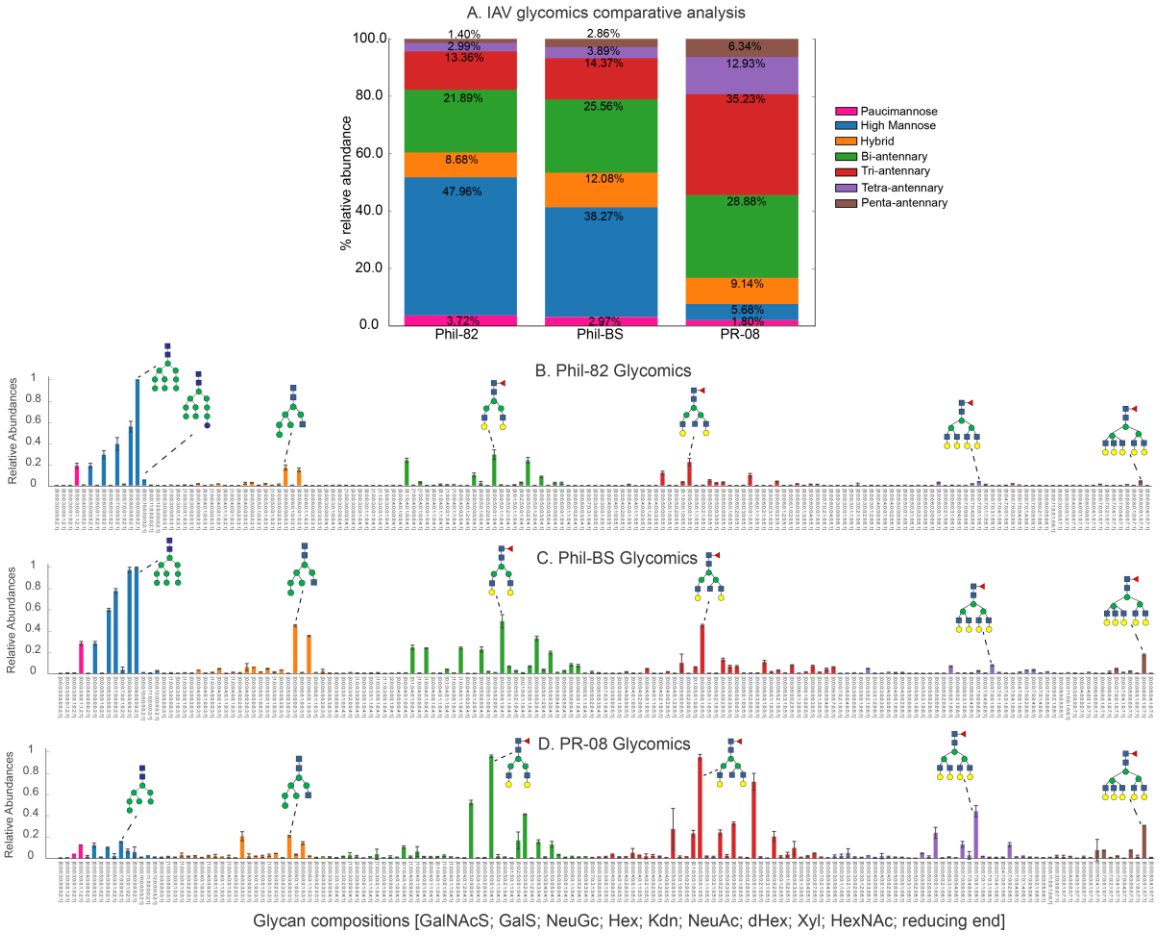


Figure 24 : Glycomics results from negative mode LC-MS profiling of released IAV N-glycans

A: Comparison of the three strains. Stacked bars represent mean composite relative abundances of different glycan classes, categorized by HexNAc and Hexose units in the identified glycan compositions (Paucimannose: HexNAc = 2, Hexose <= 4; High Mannose: HexNAc = 2, Hexose >= 5; Hybrid: HexNAc = 3; Bi-antennary: HexNAc = 4; Tri-antennary: HexNAc = 5; Tetra-antennary: HexNAc = 6 and Penta-antennary: HexNAc = 7). Bar plots show individual glycan relative abundances for B: Phil-82; C: Phil-BS; D: PR-08. Colored bars represent mean abundances relative to the most abundant composition detected for each sample. Error bars represent standard deviation. The most abundant glycoforms in each category are labeled with putative topologies using the Consortium for Functional Glycomics glycan representation scheme.

4.3.4 Glycoproteomics

For confident assignment of site-specific glycoforms, we analyzed intact glycopeptides using LC-tandem-MS and mined the data using both naïve and informed bioinformatics search space hypotheses, as described in the experimental section.

Collisional dissociation of glycopeptides in the gas-phase leads to abundant ions from glycosidic bond dissociation; however, provided application of sufficient collision energy, peptide bond dissociation also occurs [293]. Figure 25 compares the features in collisional dissociation of glycopeptides using low *versus* high collision energies.

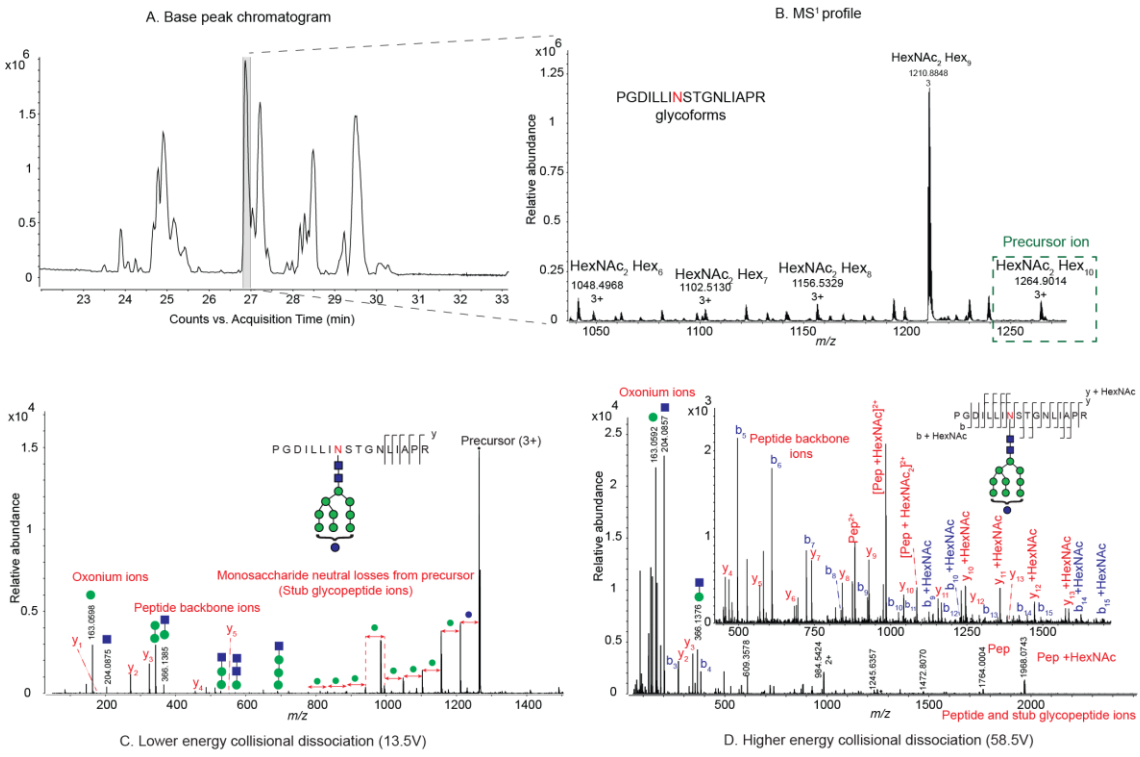


Figure 25: Glycoproteomics analysis data dimensions

A: Chromatographic separation and associated peak areas of enriched Phil-82 IAV glycopeptides; B. MS1 profiling and intact mass assignments of Phil-82 hemagglutinin glycopeptide PGDILLINSTGNLIAPR glycoforms; Tandem MS of glycopeptide PGDILLINSTGNLIAPR - Hex10HexNAc2 showing product ions resulting from lower-energy collisional dissociation (C) and higher-energy collisional dissociation (D). Glycan topologies shown are speculative based on inferred glycan compositions.

As seen in Figure 25C, lower-energy collisional dissociation yields abundant ions from glycosidic bond dissociation. Higher collision energies (Figure 25D) yield tandem

mass spectra that contain glycan oxonium ions, peptide backbone dissociation ions and intact-peptide with glycan additions (known as stub-glycopeptides). We used these features in our automated data analysis workflow for glycopeptide assignment.

4.3.5 Integrated-omics

Analysis of intact glycopeptides is often challenging due to the complexity and size of the bioinformatics search space [122,293]. Typically, hypotheses for glycopeptide mass spectrometry data analysis are limited by assumptions about the peptide variants and glycoforms present in a sample. This limitation has been overcome in proteomics analyses with the use of error-tolerant database searches [102,330,509,393], which check for modifications [103] and mutations that best explain the data, in addition to those specified by the user. This is difficult to perform in case of glycopeptides due to the already large number of glycoforms associated with each glycosylation site. A number of software tools have emerged recently that mine only glycopeptide tandem MS [180,276,422,440] or perform both proteomics and glycoproteomics searches [43]. However, the analyses are limited to the user defined boundaries of proteoforms and glycoforms. Use of assumptions while defining these boundaries or search-spaces can lead to under or over-estimation of the proteoforms, thus leading to loss of information or high false discovery rates, respectively.

For a glycoprotein as complex as HA, tandem MS can be used to resolve ambiguous assignments; however, glycopeptide tandem MS alone does not allow adequate depth and confidence in assignment for glycoproteoforms. Additionally,

glycoproteomics alone does not account for possible mutations incorporated into viral genome as a result of evolutionary pressures.

To produce the most confident information on influenza glycoproteome, we adhered to a *lex parsimoniae* approach, described here as "integrated-omics". Proteomics and glycomics workflows are well-established and easy to implement. As a first step in the workflow, we performed bottom-up proteomics on PNGase F deglycosylated influenza virions. The data were mined using hybrid *de novo* and error-tolerant proteomics database searches, which accounted for all possible PTMs, mutations and peptide variants from sample handling. The released IAV *N*-glycans were analyzed in a separate experiment and searched against glycan databases to define the glycoform search space. Thus, in an effort to minimize assumptions and efficiently mine glycoproteomics data, the glycomics and proteomics information were integrated to generate a glycopeptide database, against which the glycopeptide MS and tandem MS data were searched. Tandem MS matches were scored based on fragment ion coverage. To calculate a confidence score, a decoy database containing reversed target sequences was used, where the three residue sequon was maintained in the correct order. For example, glycopeptide SVQEIQEIQTFFYFTPN(HexNAc)KTEDTIFLR had a decoy LFITDEN(HexNAc)KTPTFYFFTQIEQIEQVSR. Confidence scores (q-values) were calculated as described by Käll *et al.* [281,635].

The tandem mass spectrometric assignment scores for some glycopeptides, such as those spanning site 126 (NWTG) were lower than the accepted score threshold, due to incomplete coverage and low precursor and product ion abundances. Manual validation

helped confirm the assignments and inclusion in the final results. All automated glycopeptide assignment results have been included as comma separated text files in the data repository.

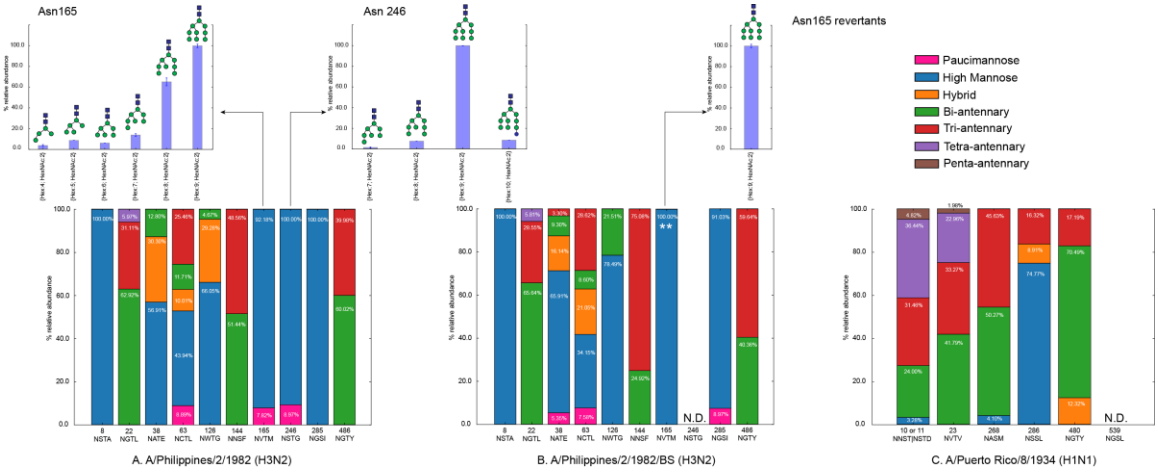
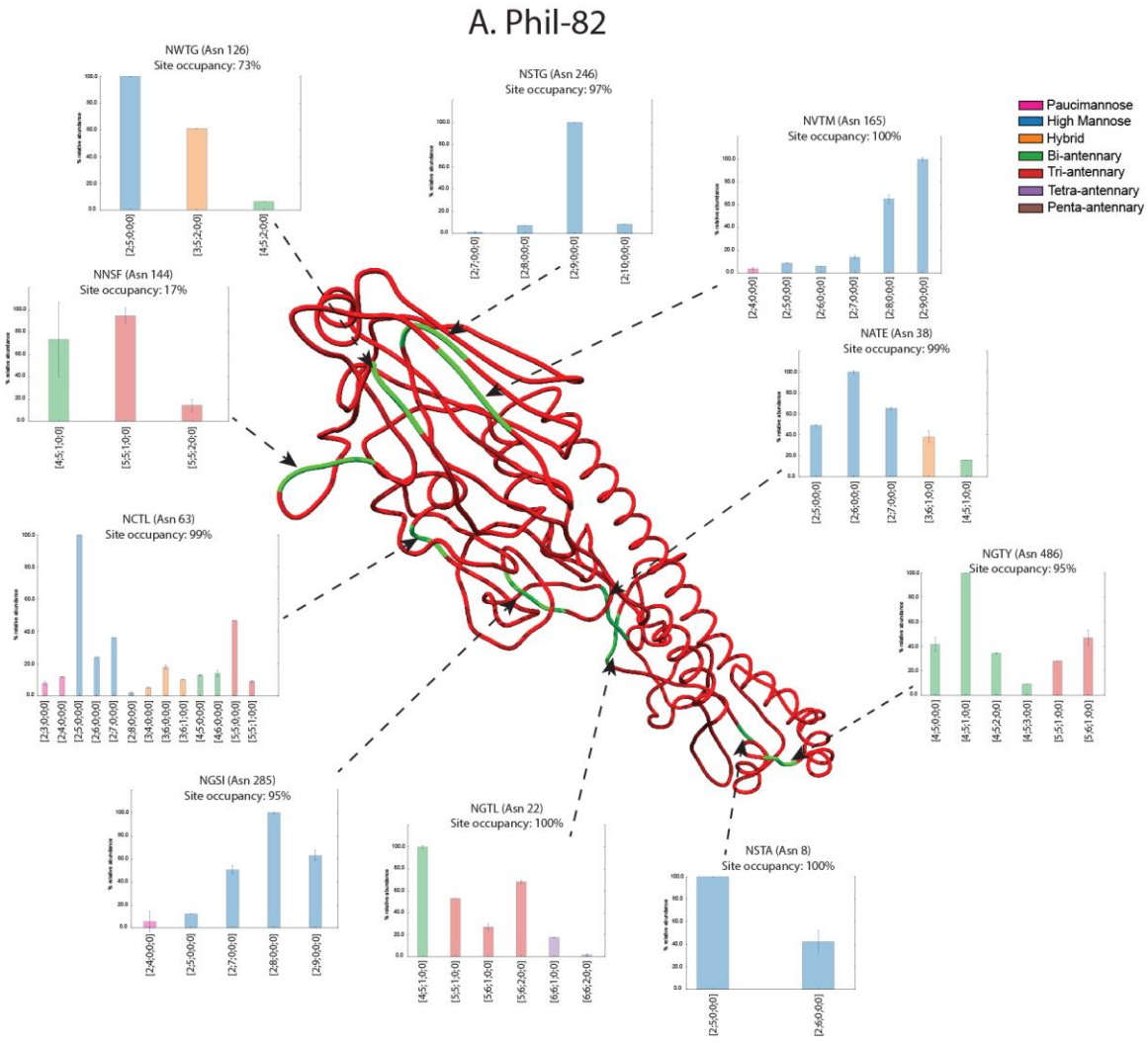


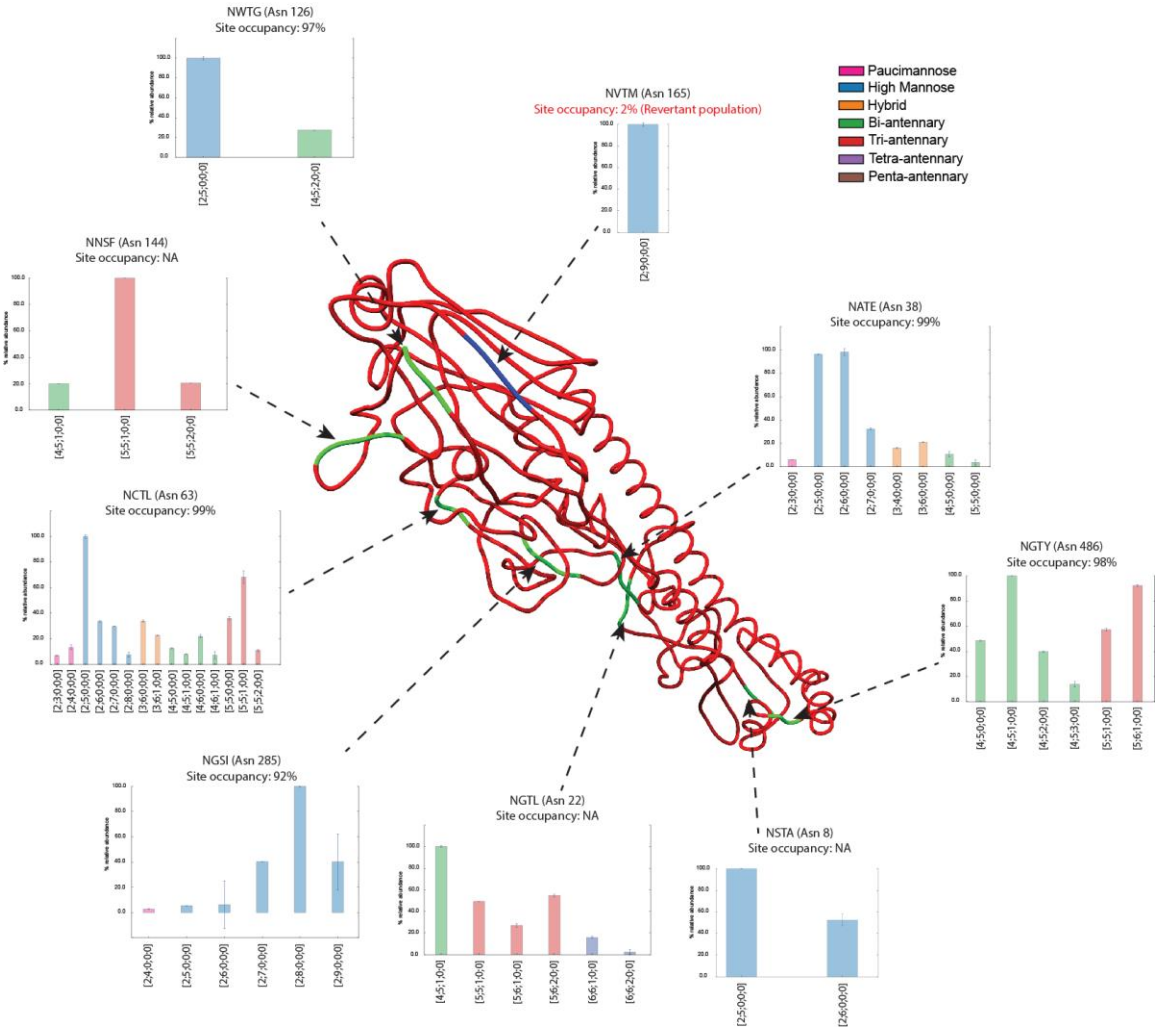
Figure 26: Integrated-omics results: Site-specific glycan distributions for hemagglutinins

A. Phil-82; B. Phil-BS; C. PR-08. Stacked bars represent composite mean relative abundances for N-glycan compositions categorized by number of HexNAc units, as described above. All stacked bars have been scaled to 100 percent to represent percentages of individual glycan classes among all glycans identified. Insets show individual glycoform relative abundances for sites identified in the interaction with immune-lectins. Error bars represent standard deviation in three measurements of glycoform abundances. N.D.: Not Detected. **Revertant population at site 165 in Phil-BS.

The site-specific glycosylation plots, shown in Figure 26, represent normalized percent relative abundances of different glycoform classes at each glycosylation site for the hemagglutinins from three IAV strains. While the plots present results from glycopeptide LC-tandem MS analyses, the databases against which glycopeptide LC-tandem MS data were searched were generated by combining the proteomics and glycomics results for respective virions. These results therefore reflect the integration of different data domains. Individual glycan compositions contributing to the abundances at sites 165 and 246 have been shown in Figure 26 (inset). It is clear from the integrated-

omics results that glycosylation at Asn165 and Asn246 is lost in the escape mutant Phil-BS from the selection pressures inflicted by growth in presence of bovine serum. All site-specific glycan distributions are shown aligned to individual HA monomer models for better visualization in Figure 27.




B. Phil-BS
NWTG (Asn 126)
Site occupancy: 97%
NVTM (Asn 165)
Site occupancy: 2% (Revertant population)
Paucimannose
High Mannose
Hybrid
Bi-antennary
Tri-antennary
Tetra-antennary
Penta-antennary
NNSF (Asn 144)
Site occupancy: NA
NATE (Asn 38)
Site occupancy: 99%
NCTL (Asn 63)
Site occupancy: 99%
NGTY (Asn 486)
Site occupancy: 98%
NGSI (Asn 285)
Site occupancy: 92%
NGTL (Asn 22)
Site occupancy: NA
NSTA (Asn 8)
Site occupancy: NA

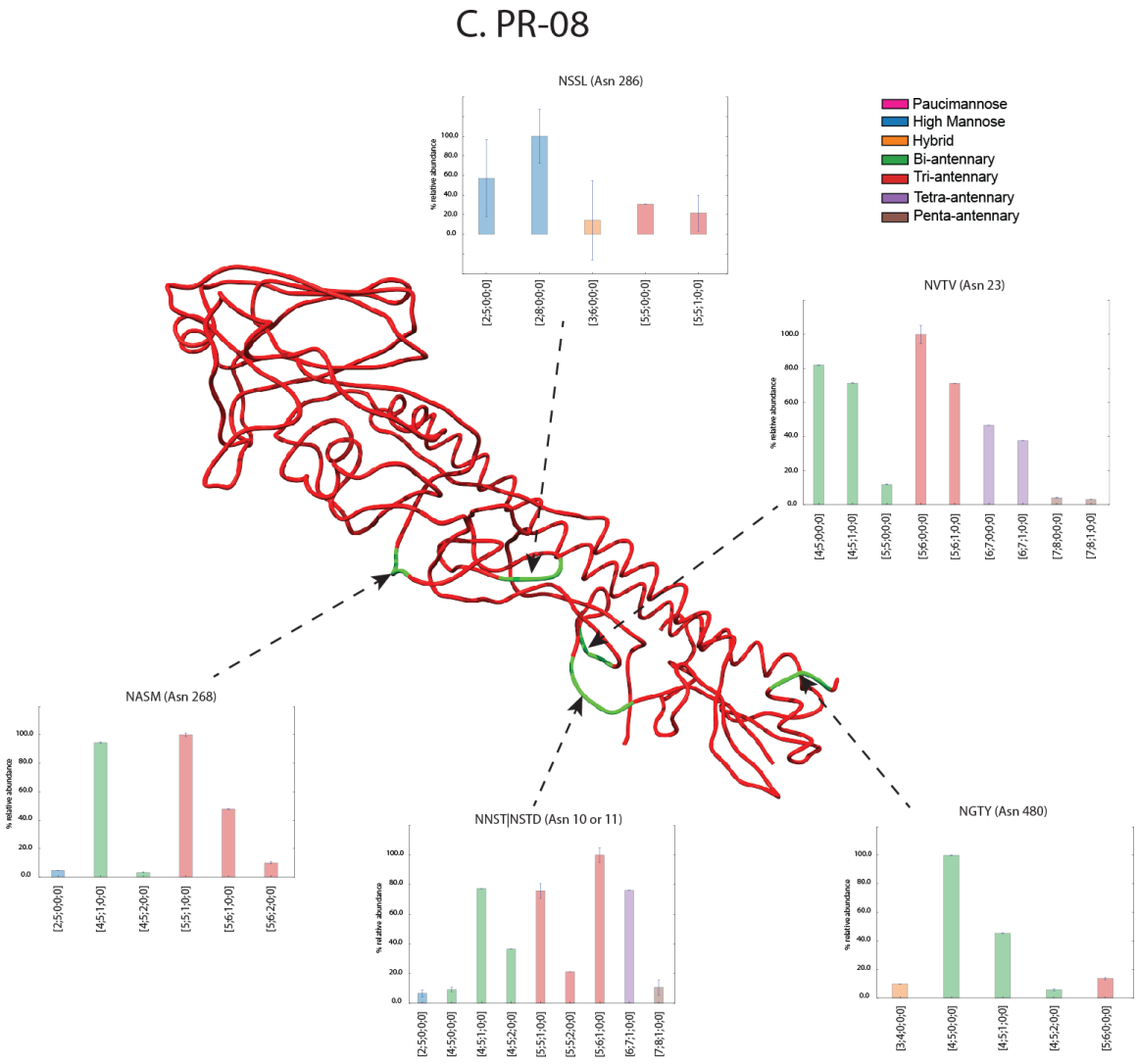


Figure 27: Integrated omics results: Site-specific glycan distributions for hemagglutinins aligned to their respective 3-D models. A. Phil-82; B. Phil-BS; C.PR-08.

Glycan compositions are listed as [HexNAc, Hex, dHex, NeuAc, NeuGc].

Hartley *et al.* had concluded that Asn246 on Phil-82 and Phil-BS is not occupied by a glycan [214]. These conclusions were derived using results from genetic sequence analyses and mobility shifts in electrophoresis, showing loss of glycosylation in escape mutants. By contrast, we observed that Asn246 is in fact glycosylated in Phil-82, while

Phil-BS loses glycosylation at both Asn165 and Asn246, due to acquired mutations. This exemplifies the need for site-specific structural information on glycoproteins, which genetic sequencing and classical biochemical methods fail to provide.

Consistent with the proteomics results, we identified glycopeptides from a revertant population only for Asn165 (NVTM); we observed no reversion at Asn246 (NSTG). To check what percentage of the total virus population reverted to the intact sequon at Asn165, we compared AUC (arbitrary unit counts) from integrated extracted ion chromatograms in the proteomics data. We also calculated site-occupancy for this revertant population as described above. In the proteomics analyses, 2% of the total Phil-BS population contained the intact revertant sequon (NVTM) at Asn165, with a glycan site-occupancy of 92%.

Glycosylation changes were limited to loss of glycosylation at Asn165 and Asn246, while that at other sequons remained largely unchanged. Both Asn165 and 246, present on the HA head, are accessible to host immune lectins. Any differences observed in bioassays could therefore be attributed to differences in glycosylation at these two sites. Although, high-mannose glycans dominated at all sites on H3 HA head, the level of glycan processing varied. In particular, for Phil-82, we noticed a striking difference in extent of high mannose *N*-glycan processing between Asn165 and 246. In the early stages of biosynthesis, glycoproteins are occupied by glucosylated high mannose ($Glc_3Man_9GlcNAc_2$) *N*-glycans. The glucose residues are recognized by calnexin and calreticulin as a quality-control mechanism for protein folding. The three glucose units are removed subsequently in the endoplasmic reticulum by glucosidases. A significant

number of glycoproteins are also acted upon by ER mannosidase-I, which removes a terminal mannose unit, after which the glycoproteins are transported to the Golgi apparatus with a $Man_9GlcNAc_2$ or $Man_8GlcNAc_2$ glycan. In the Golgi apparatus, the mannose residues are trimmed further before becoming extended into complex type structures [4,210,226,267,357,494,588]. Both Asn165 and 246 displayed higher abundances of relatively immature glycans containing a higher number of hexose units in their compositions, as seen in the site-specific glycan distributions (Figure 26 inset). The composition $Hex_9HexNAc_2$ was the most abundant at both Asn165 and 246 and the presence of $Hex_{10}HexNAc_2$ was observed at Asn246 by tandem MS, as shown in Figure 25(C and D).

The glycosylation profile on PR-08 contrasted with those of the two H3N2 strains in that complex type *N*-glycans dominated at all sequons. Given that all three viruses were grown using the same egg culture conditions, this ruled out the involvement of external factors including cell type, growth media and culture system in presence of under-processed *N*-glycans on H3N2 virions. Another recent study has shown similar predominance of high-mannose type glycans on the head region of a different H3N2 strain [17]. While the homologous H3 HAs of the same subtype but different strains show similar glycosylation profiles, the H1 HA from PR-08 contains much more highly processed *N*-glycans. This implicates protein backbone and structural features of the viral subtypes and strains as the drivers in glycan processing and resulting microheterogeneity.

4.3.6 Bioassays

We used biochemical and virological assays to survey virion binding preferences to host glycans and correlate differences in glycosylation with susceptibility of HAs towards the immune lectin SP-D (Figure 28).

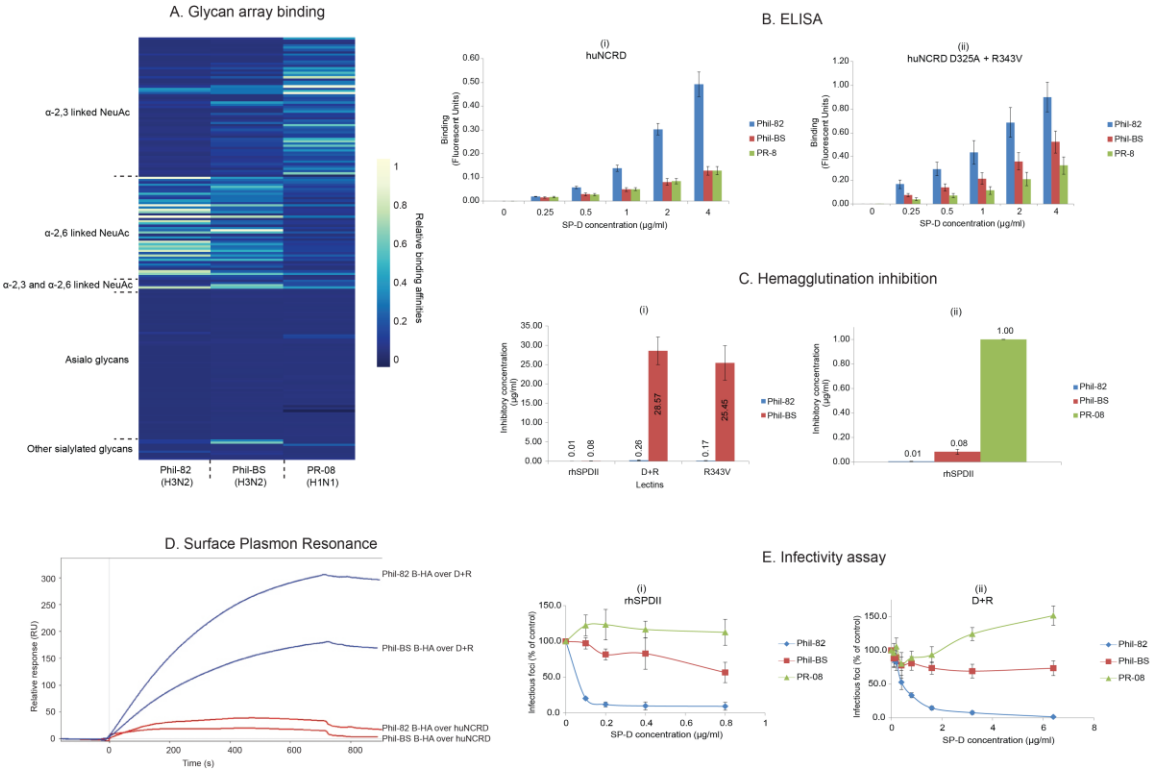


Figure 28: Results from biochemical, virological and immunological assays

A: CFG glycan array analysis of three IAV strains. Only glycans where relative binding was greater than 0.1% for any one of the three strains have been included in the heat-map. Array glycans are sorted into different categories (Y-axis) and the three virus strains are listed on X-axis; B: ELISA results. Bar-charts show dose-dependent binding to human surfactant protein-D NCRD wild-type (i) and double mutant D+R (ii). X-axis shows the concentration of SP-D used and Y-axis shows binding response in fluorescent units measured. Error bars represent standard error; C: Hemagglutination inhibition of IAV strains. (i) Comparison of inhibitory effect of different lectins on Phil-82 and Phil-BS. (ii) HA inhibitory concentrations of rhSPDII for the three IAV strains studied. Error bars represent standard error; D: Surface Plasmon Resonance results. Bromelain cleaved hemagglutinin (B-HA) from Phil-82 and Phil-BS, were introduced in the mobile phase and binding towards immobilized wild type or mutant NCRDs was measured; E: Infectivity assay results. Infectivity of the three IAV strains was compared after pre-incubation with increasing concentrations of (i) rhSPDII and (ii) D+R mutant SP-D. Data points show infectivity measured as infectious foci expressing influenza nucleoprotein, relative to control (no SP-D pre-incubation).

4.3.6.1 Glycan array binding

In addition to the genetic reassortments that cause antigenic-shift and zoonosis, differences in protein backbone and in glycosylation states affect receptor binding specificities of influenza A virus [521,529,555,270,520,501,594]. Glycan array results, in Figure 28A, show binding affinities of the three fluorescently labeled IAV strains to glycans normalized to the strongest binder for each strain. Only glycans where the relative binding was found greater than 0.1% of the strongest bound glycan for any IAV strain have been included. A complete list of glycans and fluorescence readouts from bound IAV are included in supplemental information for [291].

Both Phil-82 and Phil-BS strains showed preferential binding to α-2,6 linked sialic acid containing glycans, *versus* α-2,3 linked for the PR-08 H1N1 subtype (Figure 28A). This showed that the reassorted H3N2 virions, generated with A/Puerto Rico/8/34 backbone, retain the binding specificities on their HAs as expected. Furthermore, the mutations induced by selection pressure leading to loss of two *N*-linked glycans on Phil-BS, did not alter its binding preferences for sialylated glycans and therefore host specificity.

4.3.6.2 ELISA

ELISA results performed at different concentrations of the wild-type huNCRD show stronger binding for Phil-82, compared to Phil-BS and PR-08, which show similar binding affinities, as seen in Figure 28B(i). The D+R mutant huNCRD has mutations in the amino acids surrounding the lectin site that improve hemagglutination activity [106,110]. Consistent with previous reports, we observed higher binding for all three

strains with D+R mutant huNCRD Figure 28B (ii), the extent of which was substantially higher in the case of Phil-82 and Phil-BS than for PR-08. These results correlated well with our observed abundances for high-mannose type glycans present on the three strains from integrated omics. While Phil-BS had lower abundance of high-mannose type glycans compared to Phil-82 due to loss of glycosylation at sites 165 and 246, the abundances of immature glycans remained higher than on PR-08, for which mature complex type *N*-glycans predominated.

4.3.6.3 Hemagglutination inhibition

Surfactant protein-D plays an important role in inhibition of IAV as part of the host innate immune response. SP-D has a carbohydrate recognition domain that binds to mannose containing *N*-glycans on viral hemagglutinin. This presumably inhibits HA binding to host cell surface receptors, preventing docking and entry in to the host cell. Typical hemagglutination inhibition assays are performed using serum to detect the presence and levels of antibodies against HA. Because SP-D prevents IAV interaction with host cells, it can be used to inhibit hemagglutination activity, in a dose-dependent manner[215,224,179].

We first determined the sensitivity of Phil-82 and Phil-BS strains towards SP-D lectins. As seen in Figure 28C(i), rhSPDII (full-length SP-D) was the most effective in inhibiting the hemagglutination activity between the three SP-D constructs compared. While rhSPDII inhibited both Phil-82 and Phil-BS, the D+R and R343V constructs were effective only against Phil-82. Figure 28C(ii) shows the measured inhibitory

concentrations of rhSPDII for all three IAV strains. Phil-BS required 8-fold higher rhSPDII concentration than Phil-82, while PR-08 could not be inhibited even at the highest lectin concentration (1 µg/ml), used in the experiment. PR-08 was also found to be insensitive to D+R and R343V (data not shown). We concluded from these ELISA and hemagglutination inhibition assay results that the presence of high-mannose type *N*-glycans on the virus surface correlates with extent of neutralization of hemagglutination by SP-D.

4.3.6.4 Surface plasmon resonance

Although HA is almost 10 times more abundant than NA [262], it was important to identify the primary contributor to SP-D interactions and changes therein. We therefore performed SPR experiments using bromelain cleaved hemagglutinins (B-HA) from the wild-type and mutant H3N2 IAV strains. As shown in Figure 28D, both wild-type and mutant SP-Ds display markedly stronger binding to the Phil-82 B-HA compared to Phil-BS B-HA. Thus, differing SP-D binding amongst the tested IAV strains can be attributed to changes in the HA. Therefore, ELISA, hemagglutination inhibition and SPR assays established a clear difference in SP-D binding and activity of HA from different IAV strains.

4.3.6.5 Infectivity assay

We next determined the ability of SP-D to neutralize virus infectivity using fluorescent focus assays. For these experiments we used primary human bronchial/tracheal epithelial cells to ensure that the results reflected behavior of the

viruses in non-malignant human respiratory epithelia. Infectivity was measured via production of viral nucleoprotein by infected host cells, using an anti-nucleoprotein antibody. Cells were infected using virions pre-incubated without SP-D or with different concentrations of full-length rhSPDII and the strong binding NCRD mutant D+R. Figure 28E shows infectivity of the three virus preparations relative to cells infected without SP-D pre-incubation. With increasing concentration of rhSPDII during pre-incubation, the infectivity of Phil-82 declined sharply. By contrast, Phil-BS showed only a slight decrease in infectivity at higher lectin concentrations. With trimeric NCRD D+R, Phil-82 required pre-incubation with higher concentrations of lectin for loss in infectivity while no significant loss in infectivity was observed for Phil-BS. PR-08 infectivity was insensitive to either lectin preparation.

These results substantiate the involvement of Asn165 and Asn246 in HA interactions with SP-D, a collectin that attenuates viral interactions with host cell surface receptors and causes viral aggregation and recognition by immune cells. We concluded that the presence of high-mannose type glycans at Asn165 and Asn246 is crucial towards recognition by collectins.

4.3.7 Modeling and molecular dynamics

The integrated-omics results and bioassays showed reduced glycan processing at the two sites involved in immune recognition for Phil-82. The predominance of the $Hex_9HexNAc_2$ composition at these sites, suggested decreased processing by ERManI. The presence of $Hex_{10}HexNAc_2$ at Asn246 also suggested a decreased degree of *N*-

glycan processing by other enzymes, including glucosidases. We therefore performed three-dimensional (3D) structural modeling and molecular dynamics (MD) simulations to understand the structural basis for the low level of glycan processing at these sites and the interactions of these immature glycans with SP-D.

4.3.7.1 Molecular dynamics simulation

A 500 ns simulation was performed on the head domain of the H3 trimer, with $Man_9GlcNAc_2$ attached to each of the 15 glycosites (namely N65, N144, N126, N165 and N246 on each protomer). Despite the initial threefold symmetry of the trimer, unique glycan-glycan and glycan-protein interactions formed on different protomers, a consequence of both the algorithm employed to attach the glycans and of the randomly assigned initial atomic velocities. Over the timescale of the simulation the unique interactions observed on individual protomers were stable and did not interconvert, indicating that interconversion occurs on larger timescales.

4.3.7.2 Assessment of glycan accessibility to ERManI

The accessibility of each *N*-glycan to ERManI was assessed over the course of the simulation. The co-complexed *N*-glycan in the ERManI 3D structure was superimposed onto each *N*-glycan on the H3 head domain. Large surface overlaps between ERManI and the H3 glycoprotein indicated that the *N*-glycan was inaccessible at that particular timepoint of the simulation. The accessibility varied depending on the unique structures formed by the glycans on each protomer, but the relative percentage of time an *N*-glycan

was accessible to ERManI correlated well with the amount of processing observed experimentally (Figure 29B).

Both Asn165 and Asn246 remained completely inaccessible to ERManI throughout the simulation, which agrees with the experimental observation that these sites contain only high mannose glycans. Glycans at Asn63 and Asn126, which showed some processing to mature glycoforms, become accessible to ERManI over the course of the simulation. The most processed site, Asn144, adopts shapes accessible to ERManI for the longest amount of time during the simulation.

These results demonstrate a clear correlation between the percentage of time an *N*-glycan is accessible to ERManI and the degree of processing in the glycoforms observed at a site; however, the results indicate that longer simulation times, not feasible using present state-of-the art processors, would be required to observe interconversion between the conformations adopted at each site. Further, it should be noted that the actual value of percentage time available is based on an arbitrary overlap cutoff, and that assessment of accessibility does not account for induced fit in ERManI or any of the overlapping glycans and protein. Thus it is appropriate to consider only the relative values among sites.

4.3.7.3 SP-D binding to glycosylated HA

SP-D binds to Manα, which is present in high-mannose *N*-glycans that coat viral surface proteins. As SP-D adopts a trimeric structure with approximately 45 Å between each binding site, we hypothesized that sets of *N*-glycans on the H3 head domain may be

positioned such that SP-D could form a multimeric interaction with three *N*-glycans simultaneously, enhancing the interaction strength (see Figure 29B). It is interesting to note that HAs also have threefold symmetry and an approximately 45 Å linear distance between binding sites.

We took an ensemble of conformations adopted by the $Man_9GlcNAc_2$ glycosylated H3 from the MD simulation. By monitoring the distances between the non-reducing terminal Manα residues in each *N*-glycan over the course of the simulation, we determined which sets of three *N*-glycans were positioned for trimeric interaction with SP-D (sets that were 43-47 Å apart). Of the 20 total sets, 18 contained either Asn165 or Asn246 (Table 6). The data imply that removal of Asn165 and Asn246 would greatly reduce the potential for trimeric interactions to form between SP-D and the H3 head domain. This finding agrees with the experimental assays (Figure 28), where we observed that removal of these two glycosites reduced binding of SP-D.

Sets of N-glycan sites (protomer number)		
Site 1	**Site 2**	**Site 3**
N165(1)	N126(1)	N63(3)
	N246(1)	N126(3)
N165(2)	N165(3)	N63(1)
	N63(2)	N126(3)
N165(3)	N63(1)	N126(2)
N246(1)	N126(2)	N126(3)
	N165(2)	N126(3)
		N165(3)
		N246(3)
	N165(3)	N126(2)
		N126(3)
	N246(2)	N126(1)
	N63(1)	N126(2)
N246(2)	N126(2)	N165(3)
		N246(3)
		N63(1)
N246(3)	N126(2)	N63(1)
		N63(2)
N126(2)	N63(1)	N63(2)
	N63(2)	N126(3)

Table 6: Sets of *N*-glycosylation sites[a] whose spacing and glycan dynamics allow trimeric interaction with a trimeric SP-D

[a]Site N144 was excluded from the analysis as it was found to have both low site occupancy and a mature glycoform distribution.

We observed that the binding sites of trimeric architecture of SP-D are positioned perfectly to interact with glycans on the exposed influenza H3 head domain. Our molecular modeling predicts that the high mannose *N*-glycan at sites N165 and N246 play a pivotal role in forming trimeric interactions with SP-D, in agreement with the experimental evidence.

Enzyme accessibility is only one of the factors that determine the degree of glycan processing at a particular site. Residues outside of the traditional N-X-S/T (where X ≠ P) sequon can profoundly affect both the rate of glycosylation by oligosaccharyltransferase and the predominant glycoform observed [389]. Further, mechanisms that govern intracellular proteostasis have been observed to impact the *N*-glycosylation pattern of secreted glycoproteins [126]. We note, however, that a clear correlation between the relative accessibility of an *N*-glycan computed from an MD simulation and the degree of processing has been observed both here and in previous work [210].

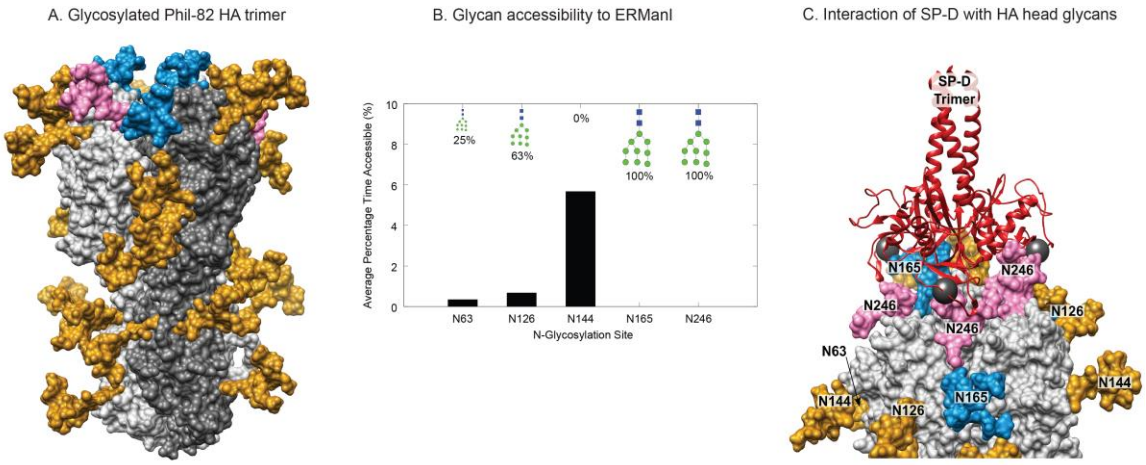


Figure 29: Structure modeling and molecular dynamics results

A: 3D model of glycosylated Phil-82 hemagglutinin trimer (HA0). One HA monomer is shown in dark grey. Glycans at Asn165 and Asn246 are shown in blue and pink, respectively. B: Percentage of simulation time that an N-glycan is accessible to ERManI. The values are averaged over each protomer of the H3 trimer. The amount of Man9GlcNAc2 at each site is depicted on the upper axis, demonstrating that increased accessibility to ERManI correlates with less Man9GlcNAc2 at each site. C: A model of trimeric SP-D (red ribbons) bound to three N-glycans on the H3 head domain (light grey surface). The 3D structure of the glycosylated H3 was taken from the MD simulation. In this snapshot, Manα residues at sites Asn 246, Asn 165, and Asn 246 on different protomers were approximately 45 Å apart, such that SP-D could bind each residue simultaneously. The calcium atoms found in each SP-D Manα binding site are shown (dark grey spheres).

4.3.8 RpNCRD analysis

4.3.8.1 Glycosylation site identification and site-occupancy analysis

Glycosylation site was identified by a mass shift upon release of *N*-glycans using PNGase F. The glycan release was performed in the presence of $H_2{}^{18}O$, allowing incorporation of the ^{18}O at the site of glycosylation leading to a mass shift approximating 3 Da, as shown in Figure 30. A representative MALDI-MS showing an ^{18}O labeled *vs.* unlabeled chymotryptic peptide is shown in the bottom panel of Figure 30. The formerly glycosylated and unglycosylated peptides were identified and differentially quantitated using LC-MS/MS. A tandem MS of the intact glycopeptide carrying the sialylated glycan is shown in Figure 31.

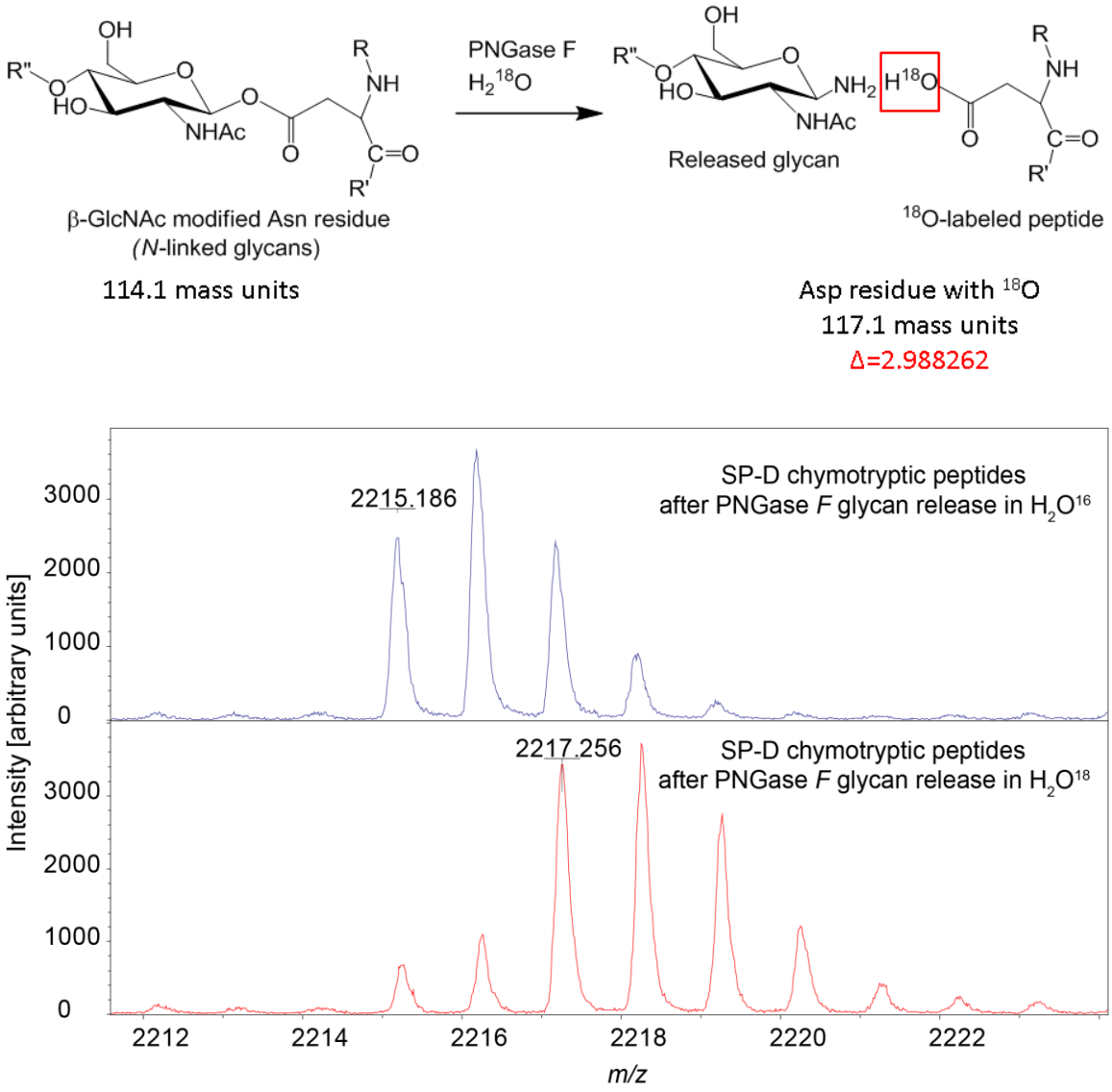


Figure 30: ^{18}O labeling of *N*-glycosylation sites

Top: Isotopic labeling of the N-glycosylation site for glycosylation site identification and site occupancy analysis. Bottom: Representative MALDI-MS of deglycosylated SP-D peptide LSMTDIKTEGNFTYPTGEPL showing incorporation of ^{18}O

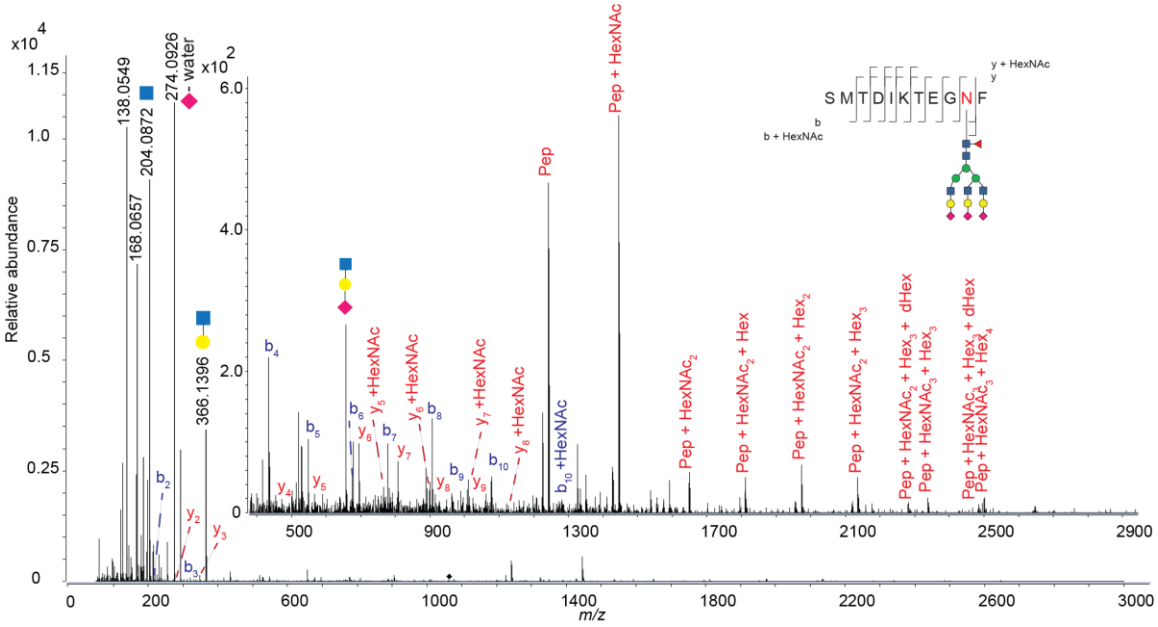


Figure 31: Tandem mass spectrum for chymotryptic pSP-D precursor *m/z* 1063.1666(3+) corresponding to RpNCRD glycopeptide SMTDIKTEGNF-Hex6HexNAc5dHex1

Symbolic representation of glycans as specified by Consortium for Functional Glycomics has been used. Glycan topology shown is speculative.

Quantification of the EICs (Extracted ion chromatograms) was performed for the identified labeled and unlabeled peptides with 2 and 3 miscleavages with peptide sequences SMTDIKTEGNFTYPTGEPL and LSMTDIKTEGNFTYPTGEPL, respectively. The site occupancy values were corrected for the effective $H_2{}^{18}O$ concentration in the PNGase F reaction. Effective $H_2{}^{18}O$ was only 85% due to presence of buffer and enzyme in the reaction. The average site occupancy was determined as 98.75 ± 0.55% from this data. Data are summarized in Table 7 below:

Filename	% O18 labeled peptide (precursor 1052) SMTDIKTEGNFTYPTGEPL	% O18 labeled peptide (precursor 1109) LSMTDIKTEGNFTYPTGEPL	Average Site occupancy (labeled/ glycosylated peptide)	Average Site occupancy (labeled/ glycosylated peptide) Corrected for effective $H_2{}^{18}O$ in the reaction	Std. deviation
SPD_O18_CID_Orbi_Top3	83.44 %	84.10 %	83.94%	98.75%	0.55%
SPD_O18_CID_Trap_Top5	83.45 %	84.30 %			
SPD_O18_Quanti_replicate1	83.28 %	84.77 %			
SPD_O18_Quanti_replicate2_a	83.68 %	84.49 %			

Table 7: RpNCRD site-occupancy determination from ^{18}O labeling

The average site occupancy from intact protein MALDI-MS was determined as 90.81 ± 0.73%. Data are summarized in Table 8 below and an example spectrum is shown in Figure 32.

Sample	Site occupancy (pixels in glycosylated/total pixels in two peaks)	Average Site occupancy	Std. deviation
20130412 No enz 35%	90.63%	90.81%	0.73 %
20130318 No enz 35%	91.57%		
20130409 No enz 1 35%	89.87%		
20130409 No enz 2 35%	91.14%		

Table 8: RpNCRD site-occupancy determination from intact protein MALDI-MS

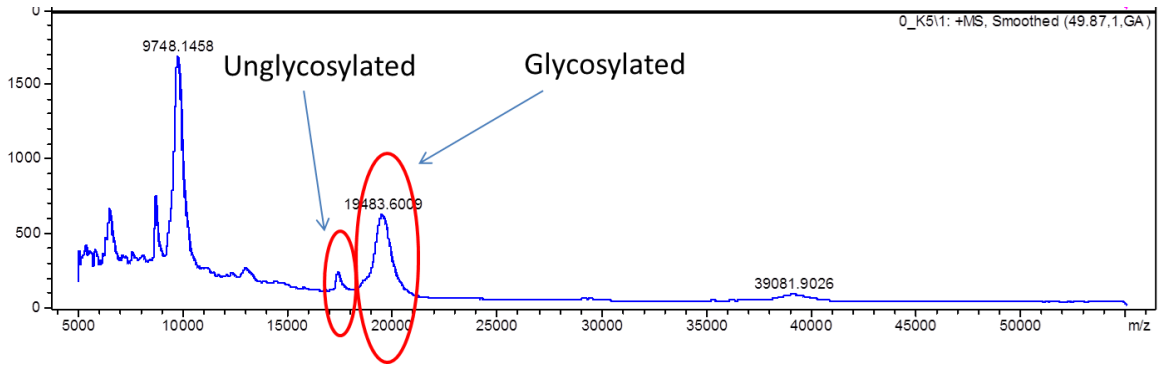


Figure 32: Site occupancy determination from intact protein MALDI-MS

From the combined results of LC-MS/MS of ^{18}O-labeled samples and intact protein MALDI, the RpNCRD was found to be more than 90% glycosylated.

4.3.8.2 Glycan profiling and analysis of degree of sialylation

Figure 33 shows relative abundances only for glycans that had a relative abundance greater than 1% of total abundance from all identified glycans.and the error bars represent standard deviation between three analytical replicates. The numeric code refers to the monosaccharides in the glycans as [Hex, HexNAc, dHex, NeuAc, NeuGc].

The average percent relative abundance by the number of sialic acids is reported in Figure 34. Approximately 50% *N*-glycans from RpNCRD were found to contain *N*-acetyl Neuraminic acid. The presence of NeuAc was confirmed by profiling the release *N*-glycans after neuraminidase treatment. Results are summarized in Table 9.

Sample	% sialylation before sialidase treatment	% sialylation after sialidase treatment
Surfactant protein-D	49.96% +/- 2.84%	1.52% +/- 0.28%
Transferrin (Control; data not shown)	95.37%	0.90%

Table 9: Degree of sialylation of N-glycans before and after neuraminidase treatment

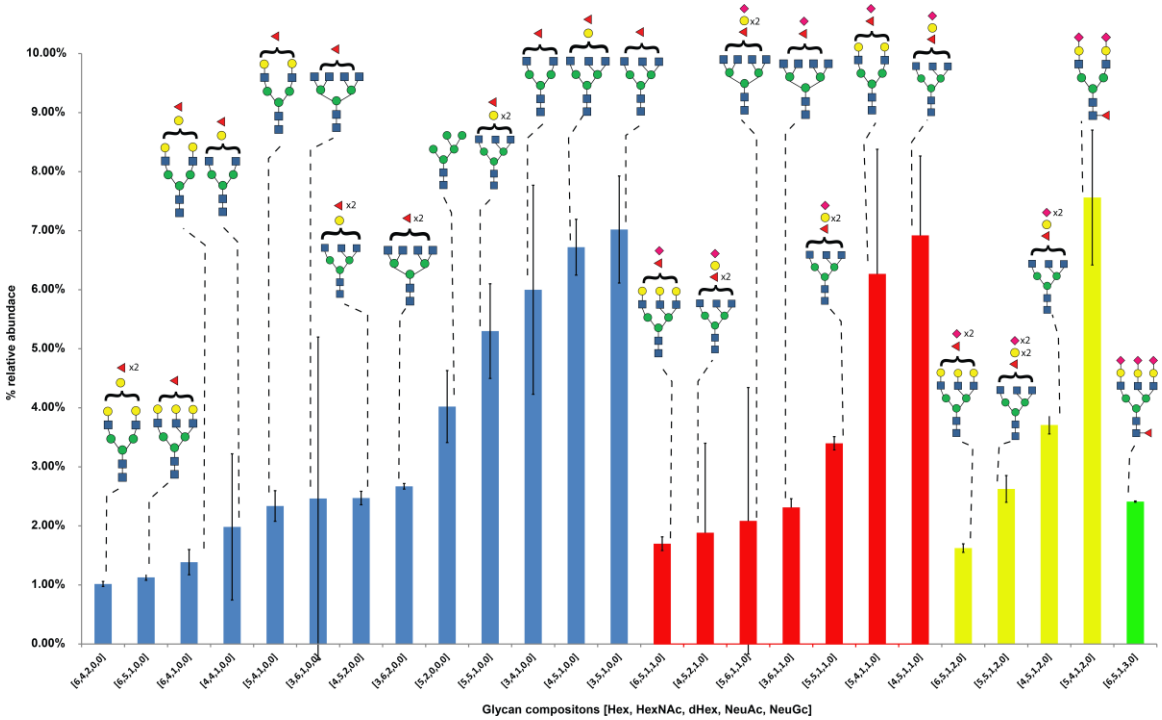


Figure 33: *N*-glycan profile of recombinant SP-D showing glycans with relative abundances greater than 1% of total identified

Percent relative abundances are average of three replicates; error bars represent standard deviation. Glycan symbolic representation as defined by Consortium for Functional Glycomics is used. Glycan topologies shown are speculative. The numeric code refers to the monosaccharides in the glycans as [Hex, HexNAc, dHex, NeuAc, NeuGc].

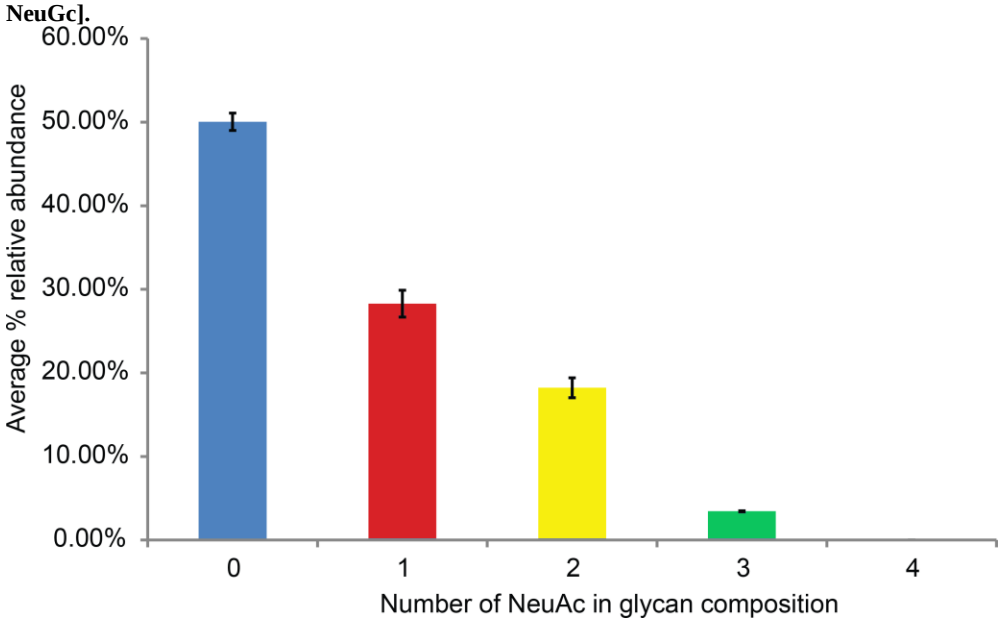


Figure 34: *N*-glycan summarized by number of sialic acids (NeuAc)

4.3.8.3 Sialic acid linkage analysis

The results for NeuAc linkage analysis have been summarized in Figure 35. Ions were assigned to the different groups shown in the figure based on lowest ppm error matches against derivatized glycan compositions. Ions accounting for approximately 12% of the total relative abundance from matched ions could not be matched unambiguously (had more than one match). Data bars indicate mean % relative abundances and error bars represent % standard deviation.

A majority of the abundance from sialylated *N*-glycans released from RpNCRD was contributed by glycans containing α2-3 linked NeuAc, while a small fraction was found to be contributed by those containing α2-6 linked NeuAc or both α2-3 and α2-6 linked NeuAc.

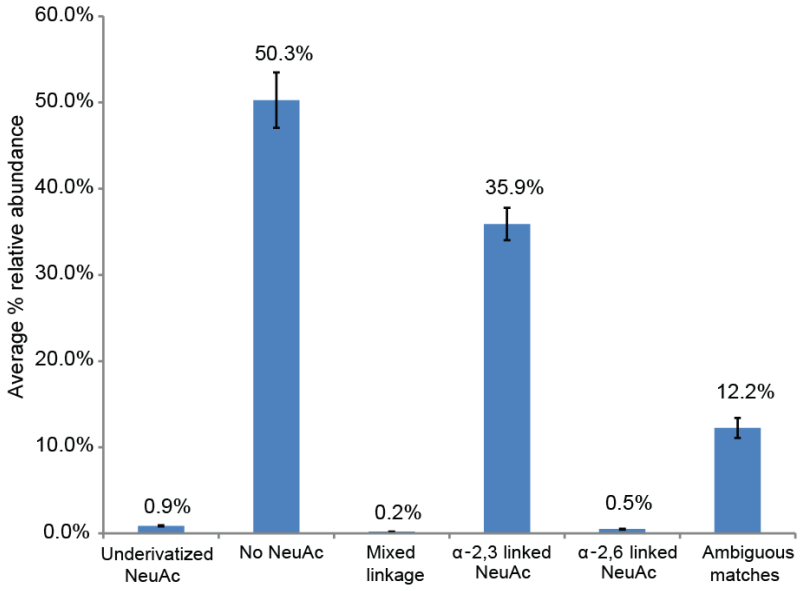


Figure 35: Composite relative abundances of SP-D N-glycans identified after chemical derivatization of NeuAc

4.4 Concluding remarks

Several recent studies have underlined the importance of HA glycosylation in influenza virus pathology [69,263,270,345,374,620]. In addition, a number of groups have attempted to trigger a broadly neutralizing antibody response towards influenza hemagglutinins, by artificially modulating HA glycosylation [84,339]. Therefore, HA glycosylation is a clear indicator of virus fitness and evolution [117,118,374,516]. Naturally, the ability to predict the fitness of an IAV strain based on its glycosylation profile will help designate strains for vaccine development.

We identified trends in viral evolution with a focus on changes in glycosylation patterns. Our results detailed the site-specific structural changes in HA glycosylation that occur as IAV adapts to immune pressure. We showed that not only the presence or absence of glycosylation, but also the level of glycan processing need to be considered with regard to prediction of interaction with molecules of the host immune system.

Most importantly, we demonstrated how comprehensive analysis of site-specific glycosylation fills in the gaps between modeling and structural studies. We demonstrated that with detailed site-specific glycosylation information, and computational glycosylation modeling, we can identify structural determinants of site-specific glycoforms. Our study describes a new strategy to elucidate IAV-host interactions and controlled evolution studies for understanding the effects of selection pressures in virus evolution.

4.5 Acknowledgements

This work was supported by NIH grants P41GM104603 (JZ, KK, JAK), NIH R01 GM100058, and P41 GM103390 (OCG, RJW) and R01 HL069031 (KLH and MRW). We thank Thermo-Fisher Scientific for providing access to the Q-Exactive Plus mass spectrometer used in this work.

Chapter 5 Improving glycopeptide and glycan separations and analysis using advanced analytical methods

The advancement of sample preparation methods, online separation, fragmentation techniques and high resolution mass spectrometry workflows has been central to improving biomolecular analyses. There is a constant push for improving speed, sensitivity and accuracy in analytical methods. Any progress made in the different steps involved in analytical workflows has a direct impact on the ability to further biological assays. This section describes the efforts directed at improvement and integration of analytical workflows, for application in biologically relevant studies.

5.1 Gas-phase separation of glycans and glycopeptides using ion-mobility spectrometry and building a collision-cross-section database

(This work has been published in [177]).

5.1.1 Introduction

Interest in ion mobility spectrometry (IMS) has grown because of its capability to separate conformations of gas-phase ions, including biomolecules, explosives, drugs, and chemical-warfare agents [101,285]. In IMS, ions are periodically introduced into a drift tube that consists of a chamber filled with a buffer gas that contains a series of electrostatic lenses to which a weak uniform electric field is applied. Ions separate in the drift tube based on their cross-section-to-charge ratios, as a result of the applied electric field and interactions with the buffer gas [54,91,101,285,441].

A major benefit of IMS is the timescale of the mobility separation, which enables the coupling of a drift tube to a time-of-flight mass spectrometer. With the IMS separation occurring on a millisecond timescale and flight times on a microsecond timescale, thousands of mass spectra can be collected during each ion mobility separation, generating a two-dimensional data array wherein the mobilities and mass-to-charge (*m/z*) values of ions are recorded [233]. In addition to the extra dimension of separation produced by having IMS prior to mass spectrometry, structural information can be obtained with the calculation of collision cross section (CCS, Ω) values, which provide information on the orientationally averaged surface area of the ions [356].

IMS has demonstrated its capability for separations of lipids, carbohydrates, and peptides, as well as proteins and protein assemblies [150,151,171,306,368,369,454,534,573,581,590]. This capability results from differences in the CCS values that occur as the mass and/or extension are increased, for each of these different classes of biomolecules. Previous work has shown that each class tends to fold into particular patterns, dependent on the charge state, degrees of freedom, and intramolecular folding forces. IMS also separates charge states into distinct trend lines, with ions falling along these lines exhibiting a strong dependence between mobility (CCS) and mass. When CCS is plotted as a function of *m/z*, more compact ions will fall below the trend line. This has been displayed previously for the series of biomolecules described above [360]. Although GlycoMob, a database for CCS of free glycans, was announced recently [66], to date there have been few reports that can provide a database that includes glycopeptides, the corresponding released glycans and deglycosylated peptides. Such data can be used to determine whether trends in the CCS values of glycopeptides follow those determined for peptides or glycans, and whether their variation shows some dependence on peptide length and/or glycan composition and branching.

Most eukaryotic proteins are modified by *N*- and/or *O*-linked glycosylation [27,41]. Cell surface and extracellular matrix proteins may include one or more *N*-linked sites where the glycan is covalently bound to asparagine [185,207,307,338], almost always in a consensus sequon Asn-Xxx-Thr/Ser (Xxx ≠ Pro) [38,39,494]. In addition, *O*-linked glycans can modify the hydroxyl groups of Ser/Thr and, less frequently,

hydroxyPro [212,543]. Glycosylation impacts the structure, folding, stability, function, and activity of proteins [185,207,307,338]. In particular, glycosylation enables binding to modular lectin domains present on many extracellular proteins [132,479]. Aberrant glycosylation associated with cancer, neurodegenerative disorders, and genetic abnormalities reflects disease physiology [114,133,158] and many infectious agents show specificity for particular cell surface glycans [22]. Thus, essential physiological functions including cell-cell and cell-matrix recognition [209], cellular adhesion, and inter- and intracellular interactions depend on protein-carbohydrate interactions [614].

Glycoconjugate glycan expression reflects enzyme-dependent remodeling and biosynthetic steps in the endoplasmic reticulum and Golgi apparatus that vary according to cell type, development, disease state and species-specific expression of glycosyltransferases [159,323]. In addition, the population of glycans present at each site reflects the efficiencies of the individual biosynthetic steps as that particular glycoprotein is generated and processed. Thus, a given site exists as a set of glycoform variants that differ with regard to degree of occupancy, composition, linkage positions and branching points [622]. Each glycoprotein is expressed as a set of glycoforms that share a common amino acid sequence as the backbone but differ in the localizations, structures and/or sequences of carbohydrate units [209].

Owing to the large number of possible glycan structures and localizations, glycoprotein and even glycopeptide compositions cannot be assigned solely from high resolution mass measurements [122]. Tandem MS experiments increase the structural detail for glycopeptides; however, statistically rigorous scoring methods for glycopeptide

MS/MS data are not yet available. Thus, the confidence level for many glycoproteomics experiments remains poorly defined. IMS offers the advantages of providing an additional, orthogonal mode for resolving mixture components and CCS data that may be used as a feature to improve the depth to which glycopeptides may be reliably assigned, provided that there is an understanding of the influence of the glycan structure on mobility [122].

Here we have measured the mobilities of glycopeptides produced by protease digestions of well-characterized glycoproteins and have utilized exoglycosidases to generate additional glycoforms, in order to observe how the volume increments that are due to individual monosaccharide units affect the CCS values. We are compiling the measured values into a database that we are developing for *N*-linked glycopeptides, reduced and permethylated *N*-glycans, and deglycosylated/deamidated peptides. In particular, we observed that the cross section values measured for the $[M+3H]^{3+}$ ions of a given glycopeptide or glycan changed linearly as the number of residues of a particular monosaccharide at the termini of its antennae was varied, and that the incremental differences were dependent on the identity of the monosaccharide. Furthermore, the relationships showed little dependence on the peptide structures. This work establishes the feasibility of building a database of glycopeptide and glycan mobilities as a function of glycan composition that will be useful for mobility-based glycopeptide and glycan profiling, which will likely extend to other types of glycoconjugates as well.

5.1.2 Experimental

Sample preparation was carried out by Kshitij Khatri, data acquisition and data interpretation were performed by Dr. Rebecca S. Glaskin with help from Kshitij Khatri under the guidance of Prof. Joseph Zaia and Prof. Catherine E. Costello. Final data integration and figure generation were carried out by Dr. Rebecca S. Glaskin and Prof. Catherine E. Costello.

5.1.3 Glycoproteins, enzymes, solvents and reagents

Bovine ribonuclease B (RNase B), human transferrin, bovine fetuin and human α_1-acid glycoprotein (AGP) were obtained from Sigma-Aldrich (St. Louis, MO). Sequencing grade trypsin and chymotrypsin were purchased from Promega Corp. (Madison, WI). Endoglycosidase F and exoglycosidases were prepared by New England Biolabs (Beverly, MA). LC-MS solvents were acquired from Fisher Scientific (Pittsburgh, PA). Unless specifically stated, all other reagents were obtained from Sigma-Aldrich.

5.1.4 Methods

5.1.4.1 Glycoprotein reduction, alkylation and protease digestion

Each glycoprotein was reduced, alkylated, and subjected to tryptic or chymotryptic digestion, as described below. The glycopeptide fraction was enriched with a ProteoExtract Glycopeptide Enrichment Kit (EMD Millipore, Billerica, MA), following the manufacturer's protocol.

5.1.4.2 Glycopeptide sample preparation, stepwise exoglycosidase digestion, and endoglycosidase release of N-linked glycans.

Tryptic digestion. Each standard glycoprotein was reduced (dithiothreitol), alkylated (iodoacetamide), and subjected to overnight tryptic digestion with the use of 2,2,2-trifluoroethanol (5%) as a denaturant, and ammonium bicarbonate (100 mM) as a pH buffer. Trypsin was dissolved in 50 mM acetic acid (1 μg/μL) and added in the ratio of 1:20 protease/protein. After digestion, the samples were dried with a centrifugal evaporator (Speedvac, Thermo Fisher Savant, Waltham, MA) and glycopeptides that resulted from the digestion were individually enriched with a ProteoExtract Glycopeptide Enrichment Kit (EMD Millipore, Billerica, MA) following the manufacturer's protocol.

Chymotryptic digestion. For digestion with chymotrypsin, 500 μg of transferrin, fetuin or AGP were dissolved in 400 μL of water and 100 μL of 100 mM ammonium bicarbonate and reduced and alkylated as above. After reduction and alkylation, 100 μL of 100 mM Tris-HCl, 10 mM $CaCl_2$ (pH 8) were added to each sample for digestion with chymotrypsin. Chymotrypsin was dissolved in 1 mM HCl (1 μg/μL) and added in the ratio of 1:20 protease/protein. The chymotryptic digestion products were dried with the Speedvac centrifugal evaporator and desalted with Pierce C18 spin columns (Thermo Scientific, Rockford, IL).

Exoglycosidase treatments. Exoglycosidases were used to sequentially remove individual saccharide units from the glycan non-reducing termini. The stepwise reactions were conducted at 37 °C. First, an aliquot of the zwitterionic hydrophilic interaction chromatography (ZIC-HILIC) enriched glycopeptides of a proteolytic digestion of each protein was analyzed by LC-IM-MS to determine CCSs for the intact glycopeptides. The

remainder of the sample was treated with α2-3,6,8 neuraminidase to remove sialic acid from the termini of the bi- or multi-antennary structures. A portion of each product mixture was desalted using a C18 spin column prior to analysis by LC-IM-MS, while the remainder of the sample was treated with β1-4 galactosidase to remove galactose from new non-reducing termini. A portion of the product mixture was desalted and then analyzed by LC-IM-MS, while the remainder of the sample was further digested using the exoglycosidase β–*N*-acetylglucosaminidase to remove the newly exposed *N*-Acetylglucosamine residues (GlcNAc). Each product mixture was desalted and analyzed by LC-IM-MS.

A 20-μg aliquot of each glycopeptide-enriched fraction was reconstituted with 80:20:0.1 water/acetonitrile/formic acid (H_2O/ACN/FA) to a concentration of 2 μg/μL; these solutions were diluted (with 99:1:0.1 H_2O/ACN/FA) to 200 ng/μL prior to C18-reversed phase separation.

Another portion of each glycopeptide-enriched fraction (20-40 μg) was sequentially treated with three exoglycosidases: α2-3,6,8-neuraminidase, β1-4-galactosidase, and β–*N*-acetylglucosaminidase. The products from each step were desalted using Pierce C18 spin columns (Thermo Scientific, Rockford, IL) and prepared for analysis by LC-IM-MS in 80:20:0.1 H_2O/ACN/FA.

N-glycans were released from glycopeptides using PNGase F. From each of the nonenriched tryptic and chymotryptic proteolysis product solutions, 50 μL (2 μg/μL in water) was added to 7.5 μL of PNGase F (500,000 units/mL), 5 μL 10x GlycoBuffer, 1 μL aprotinin (2 mg/mL in water) and the reaction mixtures were incubated overnight at

37 °C. The products were desalted using Pierce C18 spin columns and dried with the Speedvac. Flow-through and wash fractions contained salts and released glycans. The peptides resulting from the PNGase F release were resuspended in 25 µL of 90:10:0.1 H_2O/ACN/FA. For LC-MS, 0.5 µL of each peptide solution was diluted into 4.5 µL of a mixture of 99:1:0.1 H_2O/ACN/FA; 2.5 µL of the resulting solution was injected by the autosampler.

5.1.4.3 Glycan reduction, permethylation and analysis

The glycan pools resulting from endoglycosidase F treatments of RNase B, transferrin, fetuin, and AGP, with or without subsequent stepwise exoglycosidase treatment, were each reduced with $NaBD_4$ and then permethylated with a method introduced by Ciucanu and Kerek [90] and later modified by Ciucanu and Costello [89]. An approximately 50-µg portion of the glycans released from each sample by treatment with Endo F, with or without exoglycosidase treatment similar to that used for the glycopeptides, was reduced with NaBD4 and the product mixture was suspended in 200 µL of DMSO/NaOH solution and vortexed for one hour at room temperature. Methyl iodide (100 µL) and 100 µL of DMSO/NaOH were added and the reaction mixture was vortexed for another hour at room temperature. An additional 100 µL of methyl iodide and 200 µL of DMSO/NaOH were added and the reaction mixture was vortexed for another hour at room temperature; the procedure was repeated four times. To stop the permethylation reaction and extract the permethylated glycans, 1 mL of chloroform was added to the reaction mixture. Water (1 mL) was added to dissolve salt from the samples.

After the sample tubes were vortexed and centrifuged, the aqueous layer was removed, and the organic phase that contained the permethylated glycans was retained. This washing cycle was repeated 18-20 times and, finally, the products were dried with the use of a centrifugal evaporator.

The derivatized glycan products were suspended in 50 µL of 50:50:0.1 H_2O/ACN/FA. For HPLC-MS analysis 1 µL of the RNase B glycans solution was diluted 1:8 with 90:10:0.1 H_2O/ACN/FA; for transferrin, fetuin and AGP, the glycan solution was diluted 1:4 with 96:4:0.1 H_2O/ACN/FA. For each experiment, an 8-µL aliquot was injected.

5.1.4.4 Instrumentation

The Agilent 1200 series HPLC-Chip Cube interface equipped with a Polaris-HR-Chip-3C18 LC chip was coupled to a 6560 Ion Mobility Quadrupole Time-of-Flight (IM-QTOF) mass spectrometer (Agilent Technologies, Santa Clara, CA). The LC system included an autosampler, a binary capillary-flow pump used for loading samples, and a nano-flow gradient pump. Tryptic and chymotryptic digests of each reduced and alkylated protein were separated using the Polaris-HR-Chip-3C18 (3 µm, 180 Å pore size) that had a 360 nL enrichment column followed by a 75 µm × 150 mm separation/analytical column.

For the analysis of glycopeptides and peptides, the gradient was 99% A (99% water, 1% acetonitrile, and 0.1% formic acid) and 1% B (1% water, 99% acetonitrile, 0.1% formic acid) to 60% A for minutes 0 to 25, from 60% to 40% A in minutes 25 to

35, from 40% to 5% A in minutes 35 to 36, 5% A from minutes 36 to 41, 5% to 99% A in minutes 41 to 42, and 99% A from minutes 42 to 55. The flow rates for the capillary-flow and nano-flow pumps were 1.50 μL/min and 0.25 μL/min, respectively. For the analysis of deutero-reduced, permethylated glycans, the gradient was 90% A and 10% B for minutes 0 to 5, 90% to 35% A for minutes 5 to 40, 35% to 5% A for minutes 40 to 50, 5% A for minutes 50 to 55, 5% to 90% A for minutes 55 to 60, and 90% A for minutes 60 to 65. For the nanoflow and capillary-flow pumps, the flow rates were 0.25 μL/min and 2.00 μL/min, respectively.

The HPLC chip was held at a DC bias of 1950-2250 V above the ion source entrance aperture, which is followed by a dual ion funnel interface into the 6560 IM-QTOF analyzer. IMS separation took place in the ~80-cm long drift tube, operated at 18.5 V/cm (temperature, 30.5-33.5 °C; pressure, 3.91 to 4.04 Torr nitrogen). Ions exiting the drift tube are refocused in a rear ion funnel prior to entering the mass analyzer. The dual ion funnel interface and rear ion funnels were operated at 100 and 120 V peak-to-peak, respectively. For the Agilent tune mix, the typical values were: drift resolution, 50; mass resolution, 30,000; mass error, 2 ppm. The tune mix was infused during LC-MS runs and mass calibration was corrected based on its signal at *m/z* 1221. As discussed below, likely due to linkage and conformational heterogeneity, the values observed during sample measurements were lower: drift resolution, 15-60; mass resolution, 24,000 – 32,000; mass error, 1-10 ppm.

LC-IMS-MS data files were extracted and deconvoluted using DeconTools [266]. Our software package GlycReSoft [359] was used to match the deconvoluted peak lists

against theoretical peptide and glycopeptide neutral mass lists, with mass tolerance 20 ppm.

All measurements were made in triplicate. Measured CCS values were plotted *vs. m/z*. Linear fits were applied, unless noted otherwise. The results are presented in the figures that follow.

5.1.5 Results and Discussion

5.1.5.1 Analysis of RNase B Glycopeptides

Previous reports on IMS-MS of the high mannose (Man) glycans of RNase B have analyzed the glycans released after tryptic digestion of the glycoprotein [411,569,632]. The Clemmer group, utilizing a one-meter drift tube coupled to a linear ion trap, detected four isomeric structures for doubly sodiated Man_7 [632]. The Bowers and Harvey groups used a home-built drift tube and a Synapt G2-S instrument (Waters Corporation, Manchester, U.K.) [569]. Pagel and Harvey later used both unmodified and modified Synapt HDMS instruments (Waters Corporation, Manchester, U.K.) to analyze these glycans [411]. Their experiments distinguished the high Mannose *N*-glycan isomers based on drift time and determined their absolute CCS values. IM-MS studies of the proteolytic glycopeptides obtained from the proteins analyzed here (including RNase B) have not been reported, nor have the IM-MS properties been determined for the shortened glycoforms of the peptides or for the released, reduced and permethylated glycans.

In our experiments, we first studied the RNase B tryptic glycopeptide ^{34}N*LTK^{37} (where N* = occupied glycosylation site). Figure 36 displays a 2-D plot of *m/z vs.* drift

time over the drift time range during which the Man_{5-9} $[M+2H]^{2+}$ ions were observed. These results are comparable to those obtained previously by Bowers and Harvey for the released glycans.

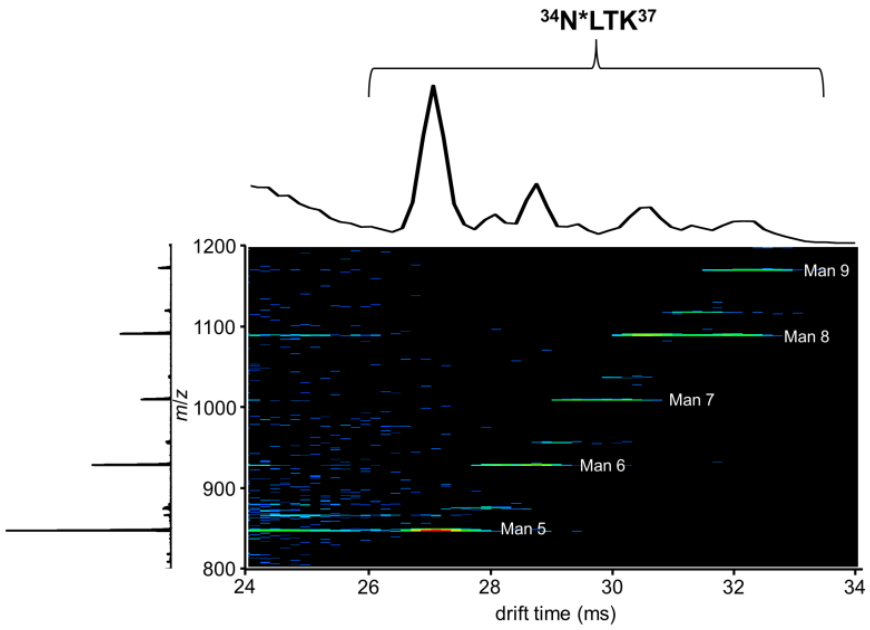


Figure 36: A two-dimensional plot displaying *m/z* as a function of drift time represented with a false color scale (least intense features in blue; most intense in red) for the RNase B tryptic glycopeptide ^{34}N*LTK^{37}.

To the left and above the 2-D plot are the mass and drift time spectra, respectively, for Man_{5-9} $[M+2H]^{2+}$ glycopeptide ions.

For the Man_{5-9} glycoforms observed in Figure 36, the extracted drift time distributions are plotted in Figure 37A. The drift time distributions for the $[M+2H]^{2+}$ ions display multiple peaks for Man_{6-8}, consistent with separation of the known multiple isomeric structures.

Here and elsewhere, the observations of lower drift time resolution (with respect to the data obtained for polyfluoroalkoxyphosphazine standards) and multiple peaks at

given *m/z* values are likely due to the presence of multiple conformers of individual isomers as well as branching isomers, some of which overlap in their drift times.

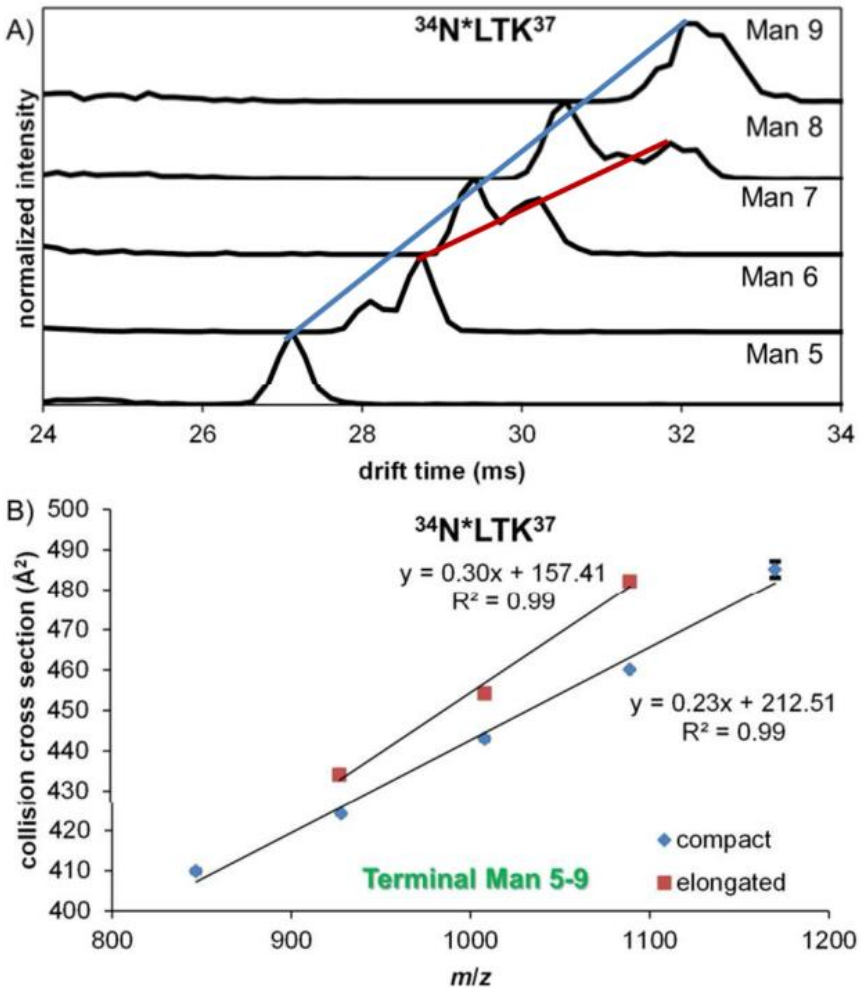


Figure 37: RNAse B glycopeptide IMS-MS

A) Extracted drift time distributions for the Man_{5-9} $[M+2H]^{2+}$ ions observed for the tryptic glycopeptide $^{34}N*LTK^{37}$ from RNase B. B) CCS values were plotted as a function of *m/z* for the high mannose glycoforms. Error bars: standard deviations obtained from triplicate measurements.

CCS values calculated from the drift times obtained for the $[M+2H]^{2+}$ ions of the Man_{5-9} glycoforms ranged from 400 to 500 $Å^2$ (Figure 37A). The measured CCS values

were plotted as a function of *m/z* (Figure 37B). Two distinct linear trends can be observed as the number of Man units on the glycopeptide varies, one corresponding to the compact conformations (peaks falling along the blue line in Figure 37A) observed for Man_{5-9}, and the other, to the elongated conformations (peaks falling along the red line in Figure 37A) of less highly branched glycoforms of Man_{6-8}.

5.1.5.2 Analysis of Transferrin Glycopeptides

Transferrin has two *N*-linked glycosylation sites, each carrying a complex biantennary structure [211,289,293]. The transferrin glycopeptides were analyzed using procedures similar to those described above. However, unlike RNase B, abundant shorter glycoforms of transferrin do not occur naturally (except in the case of disease or metabolic disorders). We therefore used exoglycosidases to sequentially remove individual saccharide units from the termini of the biantennary structures, in order to produce convenient quantities of the shorter glycoforms.

Figure 38 displays the results obtained for tryptic glycopeptides corresponding to the two known glycosylation sites of transferrin, after modification by carbamidomethylation (Cam) at cysteine, ^{421}C(Cam)GLVPVLAENYN*K^{433} and ^{622}QQQHLFGSN*VTDC(Cam)SGNFC(Cam)LFR^{642}. Figure 38A and D, B and E, and C and F display the CCS plotted *vs. m/z* for the glycoforms that varied in their content of sialic acid (NeuAc), galactose (Gal), and *N*-acetylglucosamine (GlcNAc), in that order, after these residues were stepwise removed from the non-reducing termini.

In Figure 38A, the CCS values obtained for the $[M+4H]^{4+}$ and $[M+3H]^{3+}$ ions of the biantennary glycopeptide ^{421}C(Cam)GLVPVLAENYN*K^{433} are plotted *vs. m/z* for the species containing $NeuAc_{0\text{-}2}$. Unique linear trendlines can clearly be distinguished for the two observed charge states, as a result of the increased force experienced by ions in the higher charge state during travel through the gas-filled drift tube. Linear trends were also observed for the biantennary glycoforms with $Gal_{0\text{-}2}$ (Figure 38B) and then $GlcNAc_{0\text{-}2}$ (Figure 38C) at the non-reducing termini.

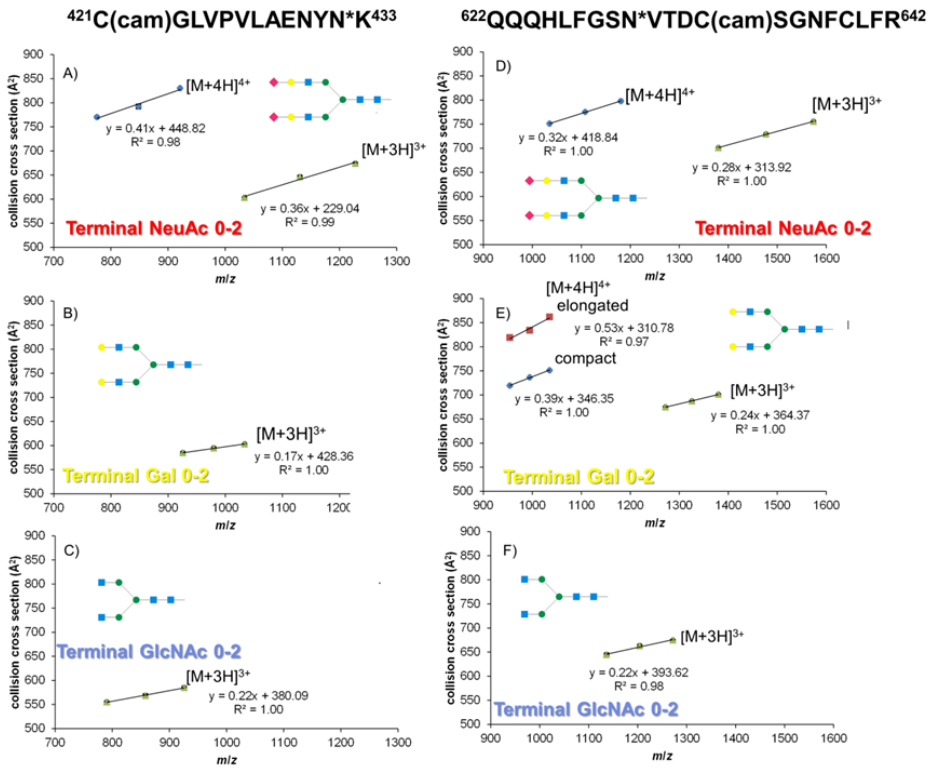


Figure 38: Plots displaying CCS *vs. m/z* for the two biantennary complex *N*-linked tryptic glycopeptides from transferrin.

The left and right columns correspond to variations in the number of each type of terminal saccharide unit on ^{421}C(Cam)GLVPVLAENYN*K^{433} and ^{622}QQQHLFGSN*VTDC(Cam)SGNFC(Cam)LFR^{642}, respectively, after each round of treatment with exoglycosidase. A) and D) display CCS measurements for products with incremental changes in the number of NeuAc residues; B) and E) show changes in Gal; and C) and F) vary the number of GalNAc. Blue diamonds: compact $[M+4H]^{4+}$ ions; red squares: elongated $[M+4H]^{4+}$ ions; green triangles: $[M+3H]^{3+}$ ions. Error bars: standard deviations from triplicate measurements.

Figure 38D shows plots of CCS *vs. m/z* for the $[M+4H]^{4+}$ and $[M+3H]^{3+}$ ions observed for the species with $NeuAc_{0-2}$ on the biantennary structure of the glycopeptide at the second glycosylation site, ^{622}QQQHLFGSN*VTDC(Cam)SGNFC(Cam)LFR^{642}. Figure 38E presents the results obtained for the series Gal_{0-2} resulting from subsequent β-galactosidase treatment of the desialylated glycoforms. Both compact and elongated conformations were observed for the $[M+4H]^{4+}$ and $[M+3H]^{3+}$ ions. Figure 38F shows that, after further treatment with β–*N*-acetylglucosaminidase, the $[M+3H]^{3+}$ ions of the $GlcNAc_{0-2}$ glycoforms also exhibited a linear trend.

A parallel set of IM-MS analyses was carried out after the digestion of transferrin with chymotrypsin, to determine if the changes in the observed CCS were primarily dependent on the glycan composition or were influenced by the length of the peptide (Data not shown).

5.1.5.3 Analysis of Fetuin Glycopeptides

Complex *N*-glycans of bovine fetuin have been previously studied by MS [205]. Here, we analyzed peptides corresponding to one of the three occupied glycosylation sites. The procedures for sequential exoglycosidase treatments followed those above.

In Figure 39, CCS was plotted *vs. m/z* for full-length tryptic glycopeptides from bovine fetuin and their products after two stepwise glycosidase treatments, ^{145}LC(Cam)PDC(Cam)PLLAPLN*DSR^{159} and its analog that contains one missed cleavage site, ^{144}KLC(Cam)PDC(Cam)PLLAPLN*DSR^{159}. These exhibited similar trends. For each set, we observed $[M+5H]^{5+}$, compact $[M+4H]^{4+}$, elongated $[M+4H]^{4+}$,

and $[M+3H]^{3+}$ ions. Although the CCS *vs. m/z* plots showed similar slopes for the $[M+3H]^{3+}$ ions as the number of monosaccharide units varied, the $[M+4H]^{4+}$ ions did not exhibit a linear trend.

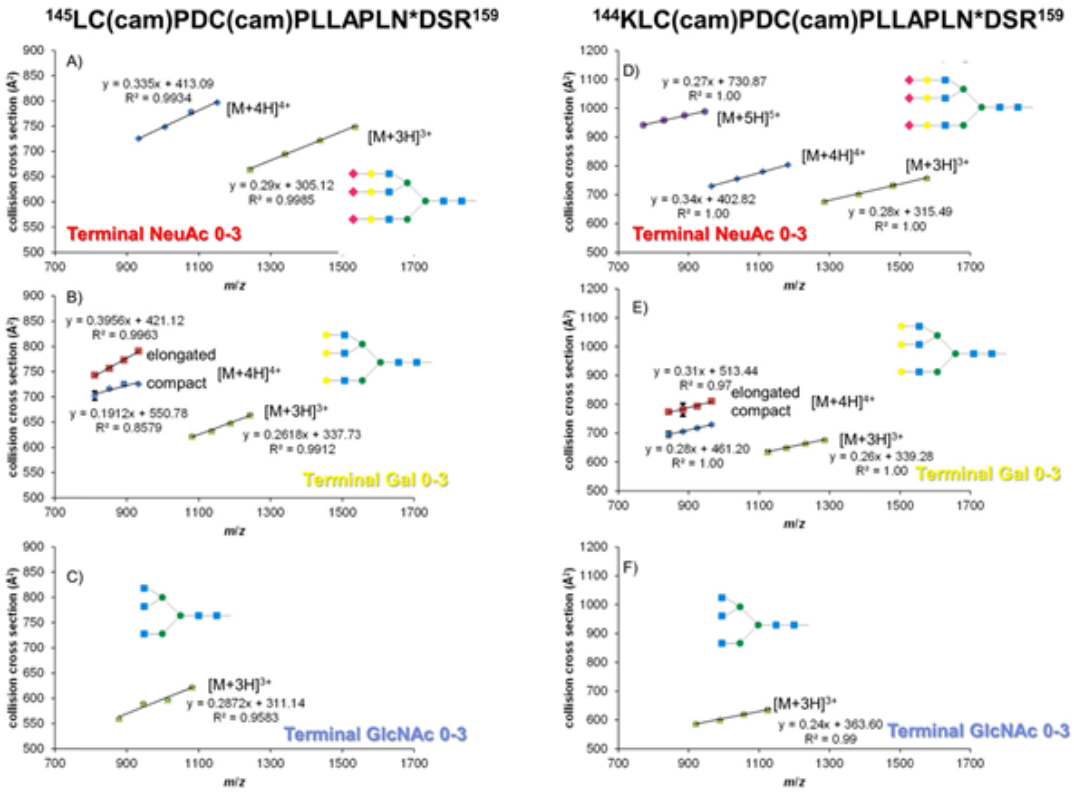


Figure 39: Plots of the CCS as a function of *m/z* for the triantennary complex *N*-linked glycopeptides resulting from the tryptic digestion of Fetuin and further exoglycosidase treatments.

Similar experiments were performed on the fetuin triantennary glycopeptides. These trends followed those described above for the biantennary transferrin glycopeptides. Figure 39 shows plots of CCS as a function of *m/z* for the triantennary complex *N*-linked glycopeptides resulting from the tryptic digestion of fetuin after sequential treatments with α2-3,6,8 neuraminidase, β-1-4-galactosidase and β–*N*-acetylglucosaminidase.

Fetuin was also digested with chymotrypsin. The results obtained as the numbers of A) NeuAc and B) Gal residues varied for the triantennary chymotryptic glycopeptide ^{156}N*DSRVVHAVEVAL^{168} are shown in Figure 40. These results exhibited no significant difference in the variance of CCS upon change in the number of NeuAc residues present on two glycopeptides containing the same sequon and the same inner glycan structure but composed of different numbers of amino acids, the tryptic peptide ^{145}LC(cam)PDC(cam)PLLAPLN*DSR^{159} and the chymotryptic peptide ^{156}N*DSRVVHAVEVAL^{168}. As the number of NeuAc residues changed, the two sets of glycoforms exhibited the same slope for the plot of CCS *vs. m/z* for the triply charged species. For the $[M+4H]^{4+}$ ions, the slope of the line corresponding to variation in the number of NeuAc residues differed significantly for this pair of glycopeptides (0.33 *vs.* 0.22) (Figure 40A). It is likely that the excess charge on the $[M+4H]^{4+}$ ions increases the extension of the overall structure as compared to $[M+3H]^{3+}$ ions; the increased flexibility of the $[M+4H]^{4+}$ ions which have a more elongated structure due to increased Coulombic repulsion is also reflected by the difference in the slopes for the $[M+4H]^{4+}$ ions, whereas the slopes for the $[M+3H]^{3+}$ ions are almost identical. Figure 39B shows the incremental changes due to variation in the terminal Gal residues on the asialo glycoforms observed for $[M+3H]^{3+}$ ions of glycopeptides resulting from tryptic and chymotryptic digestions. Here again, the situation was reversed: the slopes observed for CCS *vs. m/z* were: tryptic glycopeptides, 0.26; chymotryptic, 0.31.

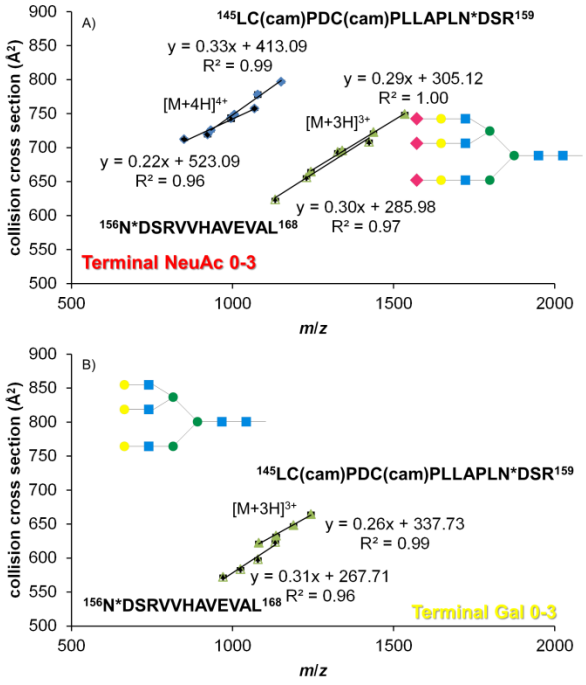


Figure 40: Plots displaying the CCS *vs. m/z* for variations in the number of terminal A) NeuAc and B) Gal residues on the triantennary structure of the glycopeptides ^{156}N*DSRVVHAVEVAL^{168} and ^{145}LC(Cam)PDC(Cam)PLLAPLN*DSR^{159}, resulting from proteolytic digestion of fetuin with trypsin (solid symbols) or chymotrypsin (open symbols)

Blue diamonds: $[M+4H]^{4+}$ ions; green triangles: $[M+3H]^{3+}$ ions. Error bars: standard deviations obtained from triplicate measurements.

Values measured for the $NeuAc_{0-3}$ series of the $[M+4H]^{4+}$ and $[M+3H]^{3+}$ ions of the triantennary glycopeptide ^{98}AN*C(Cam)SVRQQTQHAVE^{111} and similar plots for the triantennary chymotryptic glycopeptide ^{98}AN*C(Cam)SVRQQTQHAVEGDC(Cam)DIHVL^{119}, analyzed for sequential variances in the numbers of A) NeuAc, B) Gal, and C) GlcNAc residues. For $NeuAc_{0-3}$, $[M+4H]^{4+}$ and $[M+3H]^{3+}$ ions both displayed linear trends. The slope for the line for CCS *vs. m/z* observed for the $[M+3H]^{3+}$ ions with $NeuAc_{0-3}$ was 0.26, the same value observed for the $[M+3H]^{3+}$ ions of ^{98}AN*C(Cam)SVRQQTQHAVE^{111} with $NeuAc_{0-3}$. A linear trend was

determined for the $[M+4H]^{4+}$ and $[M+3H]^{3+}$ ions for Gal_{0-3} and for $[M+3H]^{3+}$ ions with $GlcNAc_{0-3}$. (Data not shown)

5.1.5.4 Analysis of α1-Acid Glycoprotein Glycopeptides

Parallel analyses were carried out for human AGP, a glycoprotein that contains five *N*-linked glycosylation sites [292,293,530]. For AGP, tryptic glycopeptides exhibited zero, one, or two missed cleavage sites. Both tetra-antennary and triantennary structures were observed for most of the AGP glycopeptides discussed here.

For the glycoforms of the AGP tryptic glycopeptide ^{87}QNQC(Cam)FYN*SSYLNVQR^{101} differing in the numbers of A) NeuAc, B) Gal, and C) GlcNAc residues, CCS was plotted *vs. m/z*, for (Figure 41). Figure 41A displays the glycoforms containing $NeuAc_{0-4}$ observed as $[M+5H]^{5+}$ ions and clearly distinguished from the $[M+4H]^{4+}$ ions.

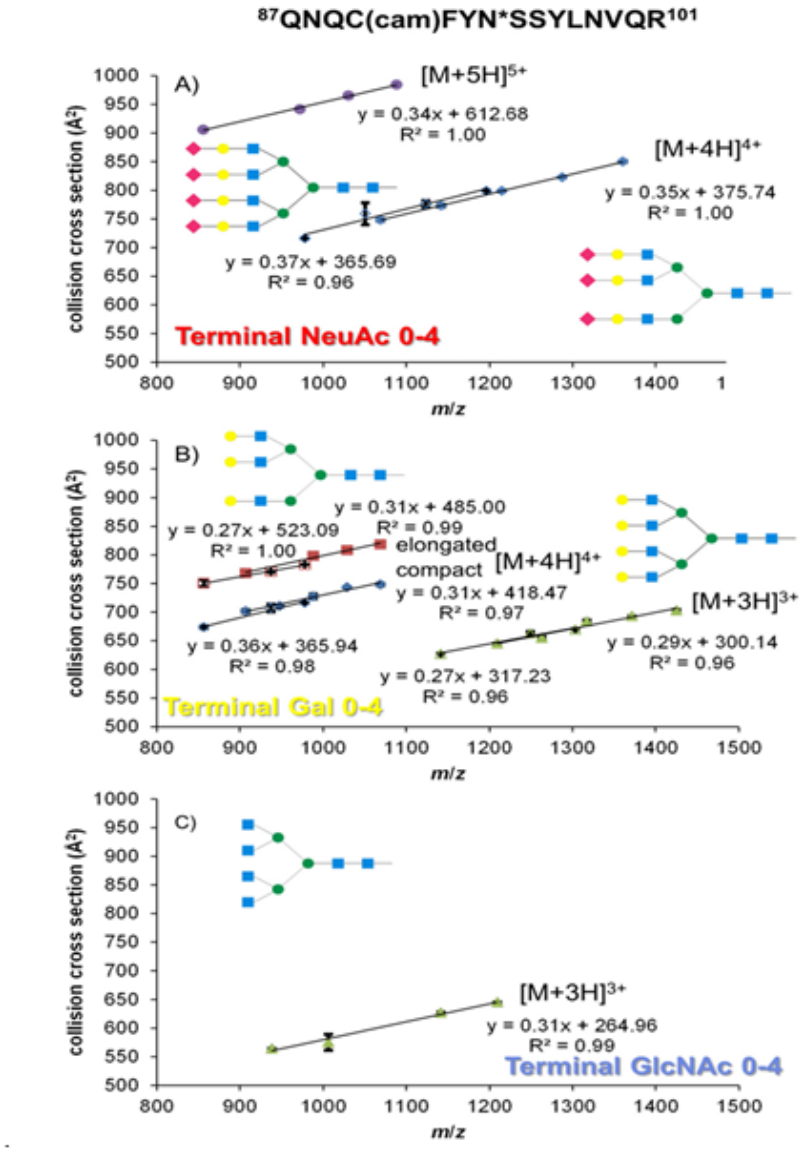


Figure 41: CCS plotted *vs. m/z* for glycoforms that vary in the numbers of non-reducing terminal NeuAc, Gal and GlcNAc residues on the multiantennary structures of the listed AGP glycopeptide.

Both tetra-antennary and triantennary structures for the $[M+4H]^{4+}$ ions fall along the same diagonal with similar slopes, 0.35 and 0.37, respectively. For the $[M+4H]^{4+}$ ions of tetra-antennary and triantennary glycoforms with Gal_{0-4}, compact and elongated structures were observed, with comparable trendlines. (Figure 41B) The values determined for the $[M+3H]^{3+}$ ions for the structures shown in Figure 41B overlap along

the same diagonal and display linear trends for the plots of CCS *vs. m/z*, with slopes of 0.29 (tetra-antennary structures) and 0.27 (triantennary). Figure 41C displays a plot of CCS *vs. m/z* with a linear trendline applied for the $[M+3H]^{3+}$ ion series corresponding to tetra-antennary structures with $GlcNAc_{0-4}$.

Both tetra- and triantennary structures were observed for glycoforms of the AGP peptide ^{98}NVQREN*GTISRY^{109}. Linear trends were found for CCS plotted *vs. m/z* (Figure 42A); the observed slopes were tetra-antennary $[M+4H]^{4+}$ ions, 0.37; triantennary $[M+4H]^{4+}$, 0.40; tetra-antennary $[M+3H]^{3+}$, 0.26; triantennary $[M+3H]^{3+}$, 0.30.

In Figure 42B, CCS was plotted *vs. m/z* for the series $NeuAc_{0-3}$ on the triantennary structures of $[M+3H]^{3+}$ ions of the tryptic and chymotryptic AGP glycopeptides ^{102}EN*GTISR^{108} and ^{98}NVQREN*GTISRY^{109}. The slopes for CCS *vs. m/z* were tryptic glycopeptides, 0.29; chymotryptic, 0.30. As discussed above for fetuin, the observed changes in CCS were similar for the pair of glycopeptides with the same sugar attached, independent of the difference in peptide lengths. Here again, it appeared that the slope in the correlation of CCS with incremental changes in glycan depended on sugar composition and was independent of peptide length.

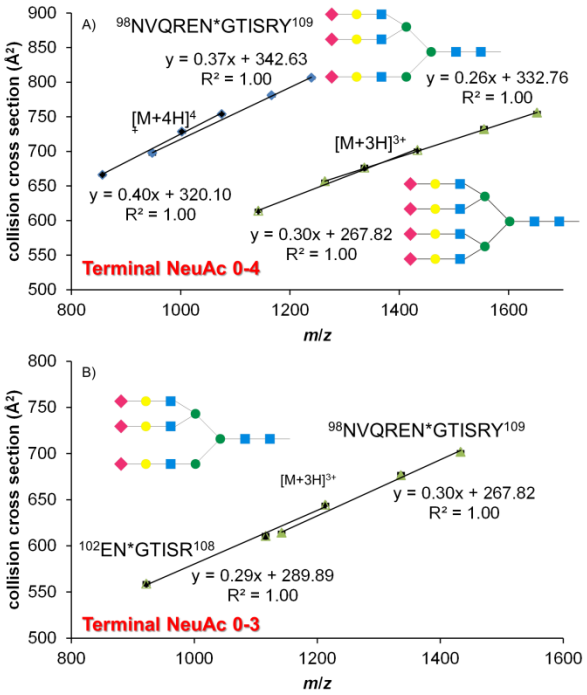


Figure 42: A) Display of the change in CCS with incremental changes in the number of NeuAc residues present on the $[M+4H]^{4+}$ and $[M+3H]^{3+}$ ions of tetra-antennary and triantennary glycoforms of the AGP chymotryptic glycopeptide ^{98}NVQREN*GTISRY^{109}, with CCS plotted *vs. m/z*.

Solid and open blue diamonds correspond to $[M+4H]^{4+}$ ions of the tetra-antennary and triantennary structures, respectively. For the $[M+3H]^{3+}$ ions, solid and open green triangles correspond to the tetra-antennary and triantennary structures, respectively. B) A plot displaying CCS *vs. m/z* for $[M+3H]^{3+}$ ions in the series of NeuAc$_{0-3}$ glycoforms of the triantennary glycopeptides from AGP ^{102}EN*GTISR^{108} (tryptic peptide, open green triangles) and ^{98}NVQREN*GTISRY^{109} (chymotryptic peptide, solid green triangles). Error bars: standard deviations resulting from triplicate measurements.

Analyses similar to those described above were conducted on additional glycopeptides resulting from tryptic and chymotryptic digestion of AGP and sequential treatment with the exoglycosidases (results not shown).

For the tetra- and triantennary glycoforms of ^{87}QDQC(Cam)IYN*TTYLNVQR^{101} with varying number of NeuAc (Figure 43A), the slopes resulting from the linear fits applied to the $[M+5H]^{5+}$ ions deviated more than those observed for the compact and extended forms of the $[M+4H]^{4+}$ and $[M+3H]^{3+}$ ions, suggesting that increased

Coulombic repulsion allows the antennae to adopt more elongated and flexible structures. The values obtained for the $[M+3H]^{3+}$ ions overlapped along a comparable diagonal; the slopes in the plots of the CCS values for $NeuAc_{0-4}$ were tetra-antennary, 0.26; triantennary, 0.29.

For glycoforms with terminal Gal_{0-4} on tetra- and triantennary structures of this glycopeptide, the compact and elongated conformations observed for $[M+4H]^{4+}$ ions of both sets of structures displayed linear trends, with deviations in their slopes (Figure 43B). The $[M+3H]^{3+}$ ions of these structures exhibited comparable diagonals (tetra-antennary slope 0.20; triantennary, 0.17).

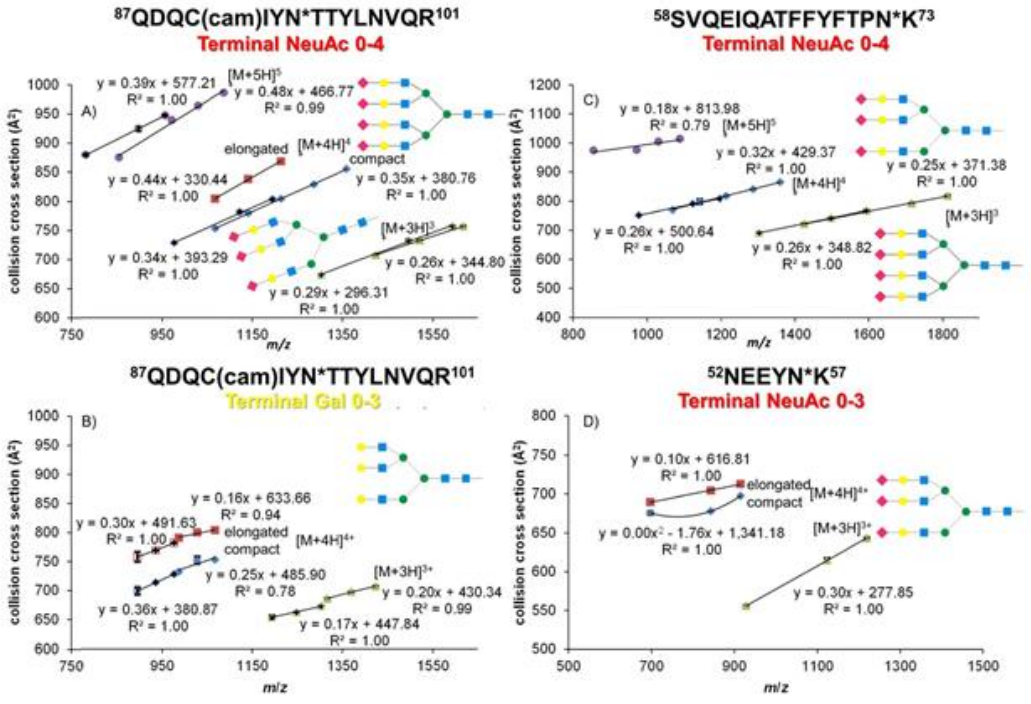


Figure 43: Plots of CCS values *vs. m/z* for glycoforms that differ in the number of NeuAc and Gal residues, present at the termini of multiantennary structures for AGP glycopeptides listed.

Results for the AGP tryptic glycopeptide ^{58}SVQEIQATFFYFTPN*K^{73} were similarly analyzed (Figure 43C); CCS values were plotted *vs. m/z* for $[M+3H]^{3+}$ to

$[M+5H]^{5+}$ ions. $[M+5H]^{5+}$ ions were observed for the tetra-antennary structure. CCS *vs. m/z* plots for the $[M+4H]^{4+}$ ions fall along similar diagonals (tetra-antennary slope 0.32; triantennary, 0.26) with variation in NeuAc. The $[M+3H]^{3+}$ ions exhibited similar behavior (tetra-antennary slope, 0.25; triantennary, 0.26).

For the triantennary glycoforms of the AGP tryptic glycopeptide ^{52}NEEYN*K^{57} ($NeuAc_{0-3}$), both compact and elongated structures were observed for the $[M+4H]^{4+}$ ions (Figure 43D). Compact ions showed a quadratic trend for CCS *vs. m/z*, whereas elongated ions followed a linear trend. Compact ions for this particular charge state may exhibit tighter folding for glycoforms having fewer NeuAc residues. CCS of the $[M+3H]^{3+}$ ions displayed a linear trend.

5.1.5.5 Analysis of Deutero-Reduced Permethylated N-glycans and Deglycosylated/Deamidated Peptides

The same type of analysis was carried out for the deutero-reduced permethylated *N*-glycans and the corresponding deglycosylated/deamidated peptides. Linear trends were observed when CCS values of the derivatized glycans were plotted *vs. m/z* (Figure 44). For the derivatized high mannose *N*-glycans (Figure 44A), the CCS values were 395 - 538 $Å^2$, with a slope of 0.27 in the line for CCS *vs. m/z*. The CCS values of the derivatized complex *N*-glycans and the shortened glycans obtained via glycosidase digestions were also measured. $[M+3H]^{3+}$ ions were observed for both tetra- and triantennary structures. The slopes in the plots for CCS *vs. m/z* with variation in NeuAc residues were similar to slopes observed for the glycopeptides (tetra-antennary, 0.34; triantennary, 0.39) (Figure 44B). For the $[M+2H]^{2+}$ ions of the tri- and biantennary

structures, the slopes for the lines reporting CCS *vs. m/z* were 0.20 and 0.25, respectively. As the number of terminal Gal residues varied, the CCS values changed at rates similar to those observed for the glycopeptides. For ions corresponding to structures that differed in the number of Gal residues (Figure 44C), the observed slopes were: tetra-antennary $[M+3H]^{3+}$ ions, 0.40; triantennary 0.39; and, for $[M+2H]^{2+}$ ions, were tetra-antennary, 0.24; triantennary, 0.27; biantennary, 0.23. The deglycosylated/deamidated peptides did not exhibit significant conformational heterogeneity and yielded well-defined, narrow CCS distributions.

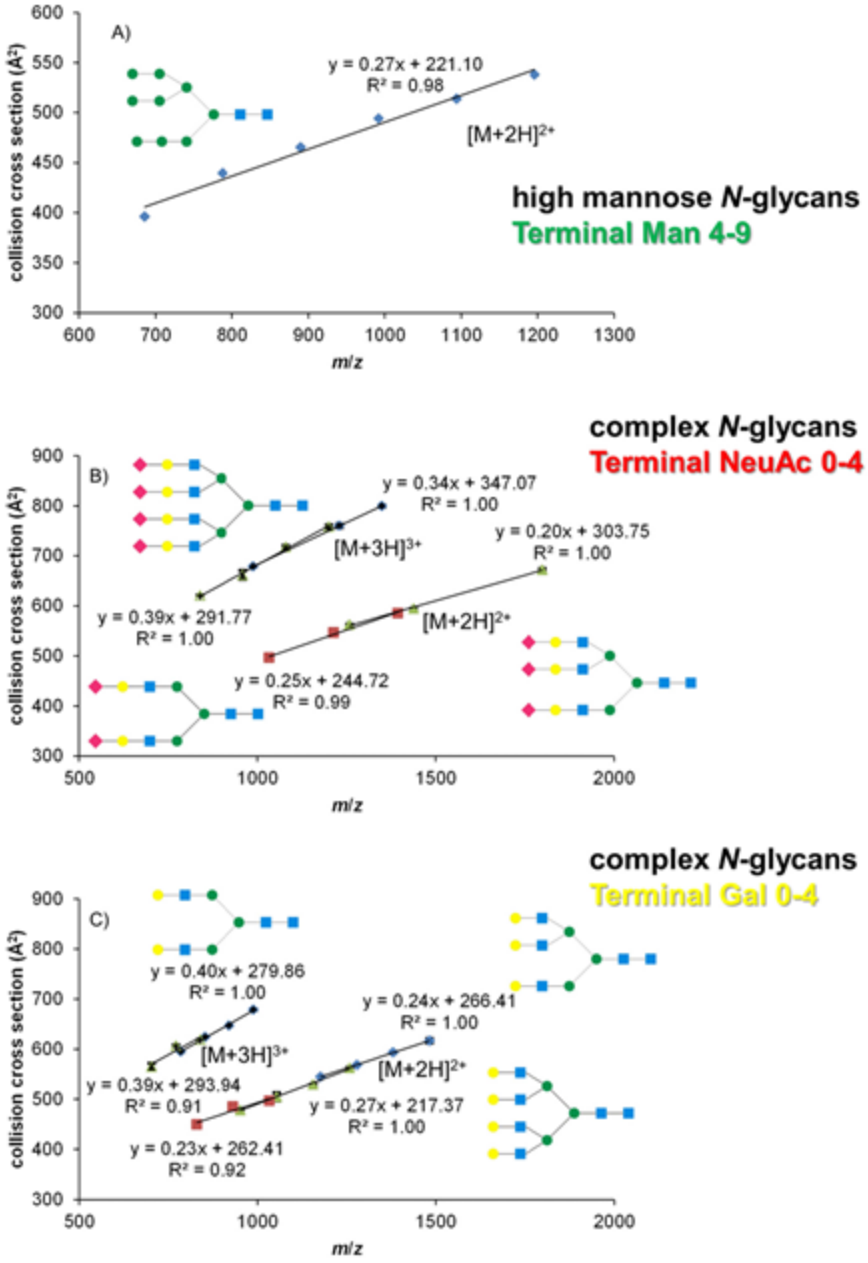


Figure 44: Plots of CCS values *vs. m/z* for deuteron-reduced and permethylated high-mannose (A) and complex glycans (B & C).

5.1.5.6 Combination of Glycopeptides, Deutero-Reduced Permethylated N-Glycans, and Deglycosylated/Deamidated Peptides

To enable visualization of the characteristic separations of the three classes of biomolecules, we assembled these data and plotted the CCS values *vs. m/z*, as shown in Figure 45. For charge states $[M+2H]^{2+}$ and $[M+3H]^{3+}$, it appears that the average densities differ in the order: glycopeptides > peptides > permethylated glycans. The differences in CCS values between the classes of biomolecules are due to class-specific structural folding (dependent on the particular intramolecular folding forces). Although both the deutero-reduced permethylated glycans and the glycopeptides may possess highly branched structures, the deutero-reduced permethylated glycans have larger CCS values than the glycopeptides with similar *m/z* values, probably because the peptide portions of the latter serve as the major reservoirs of charge localization, and thus the glycan structure does not need to expand its cross section quite so much to accommodate the higher charge states. It is also possible that the attachment of the glycan to the peptide causes compaction of the glycopeptide, and therefore the resulting CCS values for the glycopeptide and corresponding glycan moiety were found to be similar.

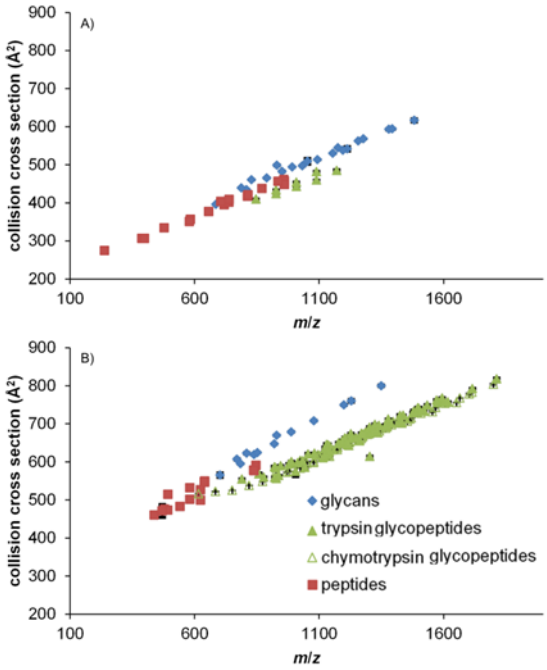


Figure 45: CCS plotted as a function of *m/z* for three classes of biomolecules: deutero-reduced, permethylated glycans, deglycosylated/deamidated peptides, and glycopeptides

A) $[M+2H]^{2+}$ ions and B) $[M+3H]^{3+}$ ions. Solid green triangles: tryptic glycopeptides; open green triangles: chymotryptic glycopeptides. Error bars: standard deviations obtained from triplicate measurements.

5.1.6 Conclusions

With coupled IM-MS, we determined CCS values for glycopeptides resulting from tryptic or chymotryptic digestion of a set of well-characterized glycoprotein standards (bovine RNase B, human transferrin, bovine fetuin and human α_1-acid glycoprotein), the deutero-reduced permethylated derivatives of the *N*-linked glycans released from these sources, and the corresponding deglycosylated/deamidated peptides. Multiple charge states and conformations were observed for the glycopeptides and derivatized glycans. We calculated CCS values for the glycopeptides generated by proteolysis, the deglycosylated/deamidated peptides and deutero-reduced permethylated glycans, including shorter glycoforms that occur naturally or that resulted from stepwise

application of exoglycosidases. We determined a linear dependence for all $[M+3H]^{3+}$ ions and for a majority of $[M+4H]^{4+}$ ions, when the CCS values were plotted *vs. m/z* for glycoforms that differed from one another by known increments. The most highly consistent result was the overlapping linear fits observed for the complete $[M+3H]^{3+}$ series for glycoform sets constructed from a given glycopeptide or deutero-reduced permethylated *N*-glycan. For some sets of glycoforms in $[M+4H]^{4+}$ and $[M+5H]^{5+}$ charge states, quadratic curves were observed. Our results from analysis of the $[M+3H]^{3+}$ ions indicate that the rate of change in CCS upon increments of individual saccharide units is independent of peptide length. The monosaccharide incremental changes in CCS values for both $[M+3H]^{3+}$ and $[M+4H]^{4+}$ glycopeptide ions were 22 to 30 (±5) $Å^2$ per NeuAc residue, 11 to 19 (±5) $Å^2$ for Gal, and 13 to 15 (±5) $Å^2$ for GlcNAc. For both $[M+2H]^{2+}$ and $[M+3H]^{3+}$ deutero-reduced permethylated glycans, the monosaccharide incremental change in CCS value was 32 to 49 (±7) $Å^2$ per NeuAc residue, 23 to 28 (±2) $Å^2$ for Gal, and 19 to 44 (±9) $Å^2$ for Man. Future work will utilize reference glycans with specific linkages and additional proteases, and will examine the behavior of *O*-linked glycopeptides and released *O*-glycans, as well as negative ions, and will refine the results by incorporating tandem MS data and, when possible, molecular dynamics calculations.

5.1.7 Acknowledgements

This research was supported by NIH grant P41 GM104603.

5.2 Electron-activated dissociation (ExD) of glycans and glycoconjugates

5.2.1 Introduction

Electron-based activation of glycans and glycoconjugates yields informative product ions that provide information critical to confident assignment of detailed structures and site-specific post-translational modifications. The application of these methods to biological samples has primarily been limited by instrument costs, sensitivity and speed. Instrument development and workflow development need to go hand in hand for efficient application and dissemination of these methods. We developed and tested approaches for standardization of sample preparation and ExD analysis of glycans and glycopeptides that are not characterized efficiently using collisional-dissociation methods. These efforts have helped identify the key features in ExD of these biopolymers and also generated data that are being used for development of bioinformatics software.

A particular problem with analyzing glycopeptides using collisional dissociation is the loss of information on the glycan moiety. For singly glycosylated peptides, this does not pose a problem because the glycan composition can be inferred from the total mass of the precursor ions, once the peptide backbone has been identified. However, when analyzing multiply glycosylated peptides, the loss of saccharide from peptide backbone fragments prevents assignment of site-specific glycosylation. In addition, CAD does not produce good peptide backbone fragmentation for multiply *N*-glycosylated peptides. This limits the applicability of this method to analysis of heavily glycosylated peptides such as *O*-glycosylated peptides, which contain stretches of serine and threonine

residues with multiple glycans attached, or long peptides with more than one *N*-glycosylation sequon. In some cases, it may be possible to separate the multiple sites of glycosylation by using a combination of proteolytic enzymes; however, this limitation of CAD prevents its application to intact protein and glycoprotein analysis.

When analyzing in a bottom-up fashion, we lose the information concerning how different pieces of a molecule exist together in the intact molecular structure and are thus unable to differentiate among the many forms that are present as part of the population that we analyze. For example, there may be a range of glycan structures present at each site on a multiply glycosylated protein. When we analyze using a bottom-up method, we are able to determine the range of glycan forms that are present at each site but we are unable to establish what combinations of glycans appear together on multiple sites of a complete protein. Thus, we are always building back to a molecular form that is an average of all molecules in a measured population. This hampers the ability to fully understand the interactions between specific molecular forms and to identify the forms that are relevant to a particular biological response or disease.

To overcome this barrier, methods are being developed to study proteins using a "top-down" approach, which means the intact molecule is analyzed instead of studying smaller portions and building back. This requires ExD fragmentation combined with high-resolution and high-sensitivity mass spectrometers. These methods are primarily being applied to non-glycosylated proteins at present. The problem is that the addition of glycans to proteins multiplies the total number of molecular forms many-fold due to the added heterogeneity of this second biopolymer. The number of protein-glycan

combinations increases exponentially with an increase in the number of glycosylation sites. As a result, for large glycoproteins with multiple glycosylation sites only a "middle-down" approach is conceivable, where moderately-sized pieces of protein that contain 2-3 sites of glycosylation can be studied [67,156,314,420,500]. We evaluated the performance of different fragmentation modes for analysis of bottom-up and middle-down glycopeptides from standard glycoproteins including Human Transferrin and Human α1-acid glycoprotein (AGP).

For glycan dissociation, CAD is known to produce abundant glycosidic bond fragments with little to no cross-ring fragmentation. It is therefore important to utilize electron-based activation for detailed structural characterization [3,208,257,575,605,634]. To date there is a paucity of high-quality oligosaccharide ExD data. ExD of glycans generates very complex tandem mass spectra that are very tedious to interpret manually. Because the availability of ExD data is limited, automated spectral interpretation tools that can handle such data are not well-developed. Our work included generation of CAD and ExD data on chemically defined standards to compare the characteristics of these fragmentation modes and to make available high-quality data on oligosaccharide standards for development of bioinformatics pipelines.

5.2.2 Materials and methods

5.2.2.1 Sample preparation

For bottom-up glycopeptide analysis, human transferrin and human AGP (Sigma-Aldrich, St. Louis, MO) were denatured by heating in presence of 2,2,2-trifluoroethanol.

Samples were then reduced with dithitreitol (DTT), alkylated using iodoacetamide (IAM) and digested with Trypsin (Promega Corp., Madison, WI) in presence of 100mM ammonium bicarbonate (Sigma- Aldrich, St. Louis, MO) as buffer. Detailed digestion protocol has been described previously [291,292,293]. Glycopeptides were enriched using ZIC-HILIC glycopeptide enrichment kit (EMD Millipore, Billerica, MA), as per the manufacturer's protocol. Samples were desalted, where necessary, using Pierce Pepclean C18 spin columns (Thermo Fisher Scientific, Pittsburgh, PA).

For middle-down glycopeptide analysis, AGP was denatured, reduced, alkylated and digested using AspN endproteinase (Promega Corp., Madison, WI). Middle-down glycopeptides were enriched by fractionation on the AspN digest using a Superdex 75 (3.2/300) (GE Healthcare, Pittsburgh, PA) on a Beckman Gold HPLC system (Beckman Coulter, Inc, Indianapolis, IN). 25mM ammonium formate (pH 4.5) with 10% acetonitrile was used as mobile phase for the separation at an isocratic flow rate of 50µL/min. Fractions were collected manually based on UV absorbance at 230nm and further fractionated using a Vydac C18 reversed-phase HPLC column (W.R. Grace & Co., Columbia, MD) on an Agilent 1200 series chromatograph (Agilent, Inc., Santa Clara, CA), fitted with an automated fraction collector.

Where indicated, for both bottom-up and middle-down analyses, samples were also desialylated using α-2,3,6,8 Neuraminidase (New England Biolabs, Ipswitch, MA), prior to LC-MS or nanoESI-MS to reduce glycoform heterogeneity.

Chemo-enzymatically synthesized glycan standards were used for glycan ExD and CAD experiments. The standards were kindly donated by Chemily, LLC (Atlanta,

GA). Glycan samples were first reduced using sodium cyanoborodeuteride (Sigma-Aldrich, St. Louis, MO); excess borate salts were removed by repeated resuspension and drying in methanol. Deutero-reduced glycan alditols were permethylated using the protocol described by Cicuanu and Costello [89,90]. Following liquid-liquid extraction of permethylated glycan alditols, desalting was performed using C18 spin columns.

5.2.2.2 Data acquisition

Bottom-up glycopeptide samples were either directly infused for analysis using the Advion Nanomate alone or analyzed by online LC-MS using LC-MS at nanoliter flow-rates on a Bruker Solarix 12T hybrid Q-FTICR mass spectrometer mounted with an Advion Nanomate nanoESI source (Advion Inc., Ithaca, NY). A Waters™ NanoAcquity™ nano-flow chromatograph (Waters Corp., Milford, MA) mounted with a Waters™ Xbridge™ reversed-phase column (150 μm × 100 mm) packed with 1.7 μm BEH C18 resin and a Waters™ trap column (180 μm × 20 mm) packed with 5 μm Symmetry™ C18 stationary phase, was used for online LC-MS. Glycopeptides were analyzed by collisional-dissociation on a Thermo-Fisher Scientific Q-Exactive Plus mass spectrometer (Thermo Scientific, San Jose, CA), using a similar LC-MS setup, as described previously [291]. Bottom-up Transferrin glycopeptides were also analyzed using CAD on an Agilent 6550 Q-TOF mass spectrometer using online HILIC enrichment combined with reversed-phase separation, as described previously [293].

Middle-down glycopeptide fractions from C18 LC separation were analyzed by nanoESI-MS using a Bruker Solarix 12T FTICR-MS mounted with an Advion Nanomate

source. ETD and EThcD (Electron Transfer Dissociation with supplemental collisional activation) experiments were performed on ZIC-HILIC enriched and unfractionated bottom-up and middle-down glycopeptide samples in collaboration with Dr. Rosa Viner at Thermo-Fisher Scientific (San Jose, CA) by LC-MS using an EASY-nLC chromatograph with an EASY-Spray C18 LC column on a Thermo Orbitrap Fusion instrument. The instrument was set to perform HCD-triggered EThcD, which allowed EThcD to be performed only on precursor ions that generated saccharide oxonium ions.

Permethylated glycan alditols were resusupended in 1:1 methanol:water with 0.1% formic acid and infused via nano-ESI (Advion Nanomate) into a Bruker Solarix 12T FTICR-MS for ExD experiments or into a Thermo Fisher Scientific Q-Exactive Plus mass spectrometer for CAD (HCD) experiments. Sodium metal adduction was achieved by addition of sodium hydroxide to the samples and glycan EED was performed as described previously [605].

Experiments on the Bruker Solarix FTICR-MS were performed in collaboration with Dr. Yi Pu. For data acquisition on the Solarix by LC-MS/MS, precursor ions of interest were isolated using a front-end quadrupole mass-filter and accumulated in the collision-cell for 100 to 1000 ms prior to MS/MS analyses. hECD was performed by irradiating trapped ions in the ICR-cell with 12-14 eV electrons, for up to 1 s. Cathode current was set at 1.5 A and the cathode bias was set between -12 and -14 V. Summed spectra were used to improve signal-to-noise ratio. LC-MS/MS experiments were performed by data-dependent precursor selection of the most-abundant parent ion or using a targeted inclusion list.

5.2.2.3 Data analysis

Glycopeptide data analysis was performed manually or in a semi-automated manner using python scripts developed in-house, using the GlyPy library [300] (https://pypi.python.org/pypi/glypy/0.0.5rc2) developed by Joshua A. Klein and with help from Prof. Cheng Lin and Dr. Yi Pu, at Center for Biomedical Mass Spectrometry, Boston University. Glycan tandem MS data were deconvoluted using the SNAP algorithm version 2.0 in the Bruker DataAnalysis software, using a quality factor threshold of 0.01, signal-to-noise ratio of 2, relative intensity threshold of 0.01% and a peptide averagine model for isotope fitting [244,308]. Deconvoluted peaklists were annotated using GlyPy and a prototype version of GlycoDeNovo (Prof. Pengyu Hong, Brandeis University). The data are being used for further development of these software pipelines.

5.2.3 Results and Discussion

5.2.3.1 Glycopeptide analysis by different dissociation methods

As described previously, glycopeptide fragmentation by collisional-dissociation causes dissociation of the labile glycosidic bonds, leading to loss of the carbohydrate before the peptide backbone can be fragmented. This complicates assignment of the glycosylation sites for multiply glycosylated peptides. For bottom-up *N*-glycopeptide analysis, peptide backbone fragments often retain a HexNAc since the amide bond between the asparagine residue and HexNAc is stronger than the glycosidic bonds and survives CAD. Further, for multiply glycosylated peptides, CAD is generally unable to

resolve site-specific glycosylation. Therefore, ExD becomes necessary in cases where multiple glycans are attached to the peptide to keep the glycan moieties intact, while fragmenting the peptide backbone.

5.2.3.2 Comparison of dissociation modes for bottom-up glycopeptide tandem-MS

To compare the dissociation characteristics of different dissociation modes, we acquired tandem MS of the same transferrin glycopeptide (QQQHLFGSNVTDCSGNFCLFR – Hex5 HexNAc4 NeuAc2) using CAD, ETD and ECD. Figure 46 shows the tandem MS results using CAD. As expected [293], extensive dissociation of the glycan is seen leading to generation of abundant oxonium ions. The intact peptide ion without any glycans attached and a glycan Y_1 ion are also seen, which confirm the intact peptide and glycan mass. A series of peptide backbone ions are also seen that have relatively low abundances. In addition to the bare peptide backbone ions, peptide fragment ions with an attached HexNAc (n_x/Y_1) are also observed. Together these features allow unambiguous assignment of the peptide sequence and the glycan composition can be inferred from the residual precursor mass. The same precursor was also fragmented using CAD on the FTICR MS. While similar CAD dissociation patterns were observed, fewer peptide backbone fragments were observed, owing to the lower sensitivity of the FTICR MS compared to the Q-TOF.

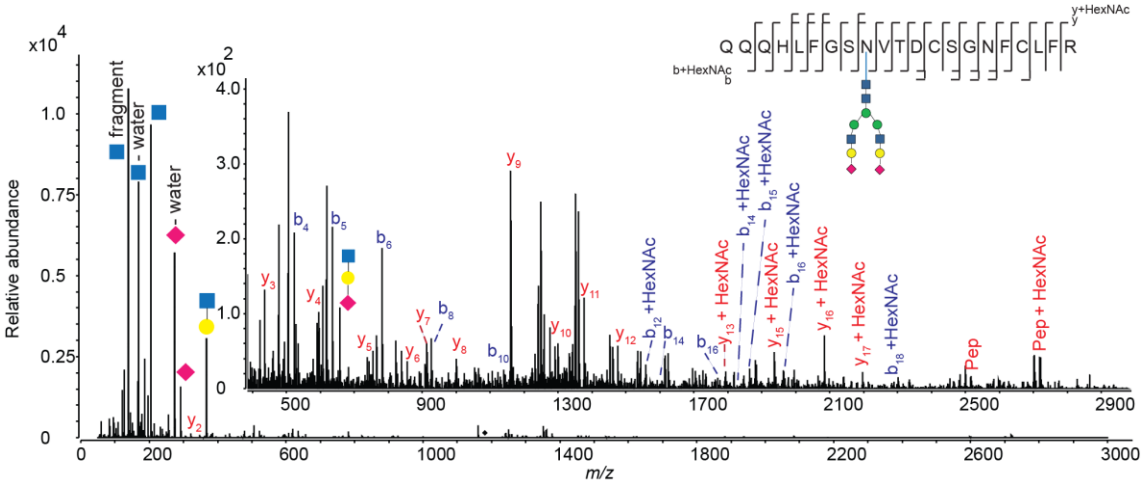


Figure 46: CAD Tandem MS of a Transferrin glycopeptide ($[M+4H]^{4+}$, *m/z* 1180.7328) on a Q-TOF instrument

An ETD spectrum for the same precursor, generated on the FTICR instrument, is shown in Figure 47. It was interesting to see that ETD generated few, if any, peptide backbone fragmentation for this precursor, while some glycan fragmentation was observed from vibrational excitation, probably induced by ion transfer in the optics. The intact precursor and abundant charge-reduced species were observed in the ETD spectrum, indicating an ETnoD (Electron Transfer with no Dissociation) phenomenon, where the electron is transferred to the precursor leading to charge reduction but no fragmentation takes place [195,272].

Contrastingly, ECD of the same precursor generated extensive glycopeptide fragmentation on the FTICR MS (Figure 48). We observed abundant c and z-type peptide backbone ions with the intact glycan attached in the range *m/z* 1600-2400. In the lower *m/z* range, peptide backbone fragments not containing the glycosylation site were observed. A few b/y-type peptide ions and some peaks resulting from glycan

fragmentation (oxonium ions and saccharide losses from the precursor) were also observed, indicating slight vibrational excitation.

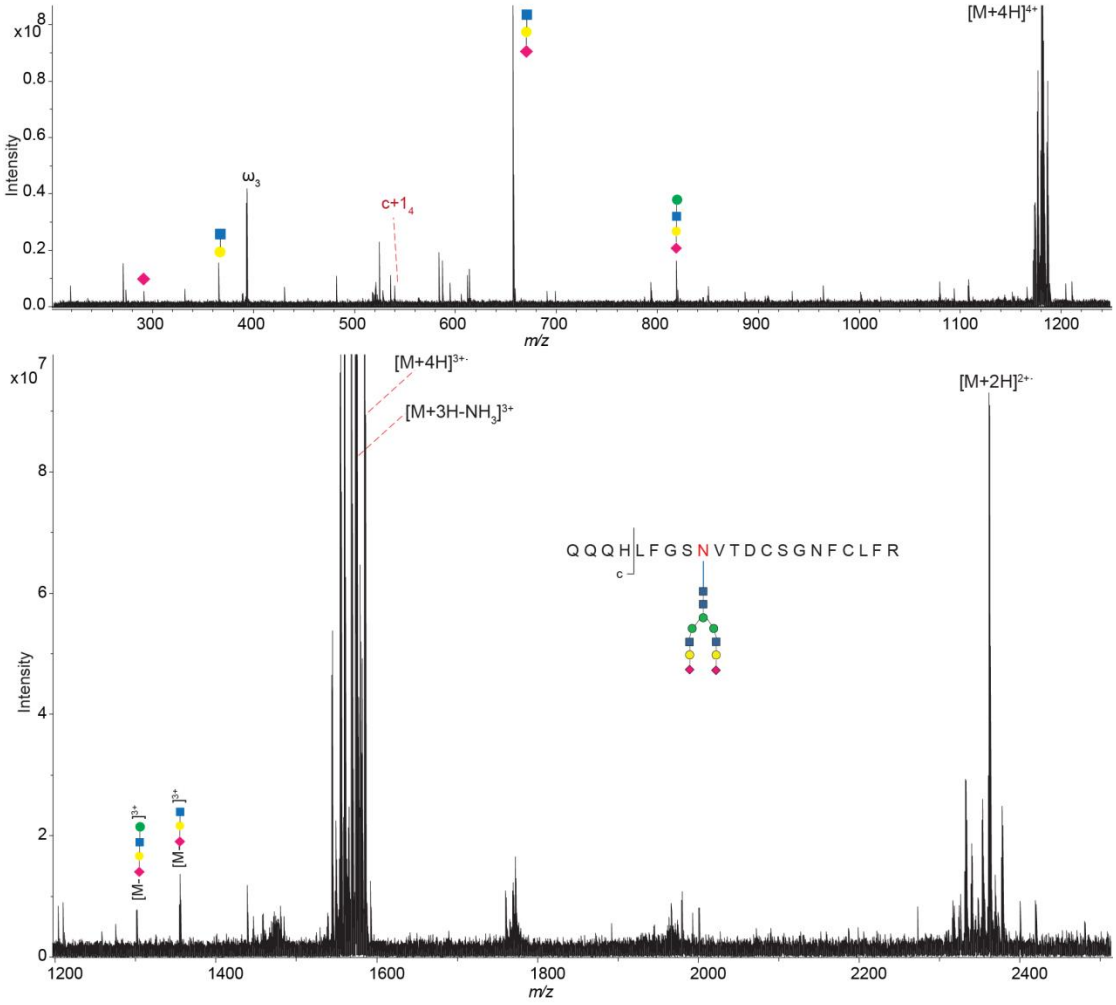


Figure 47: ETD Tandem MS of a Transferrin glycopeptide ($[M+4H]^{4+}$, *m/z* 1180.7328) on an FTICR-MS instrument

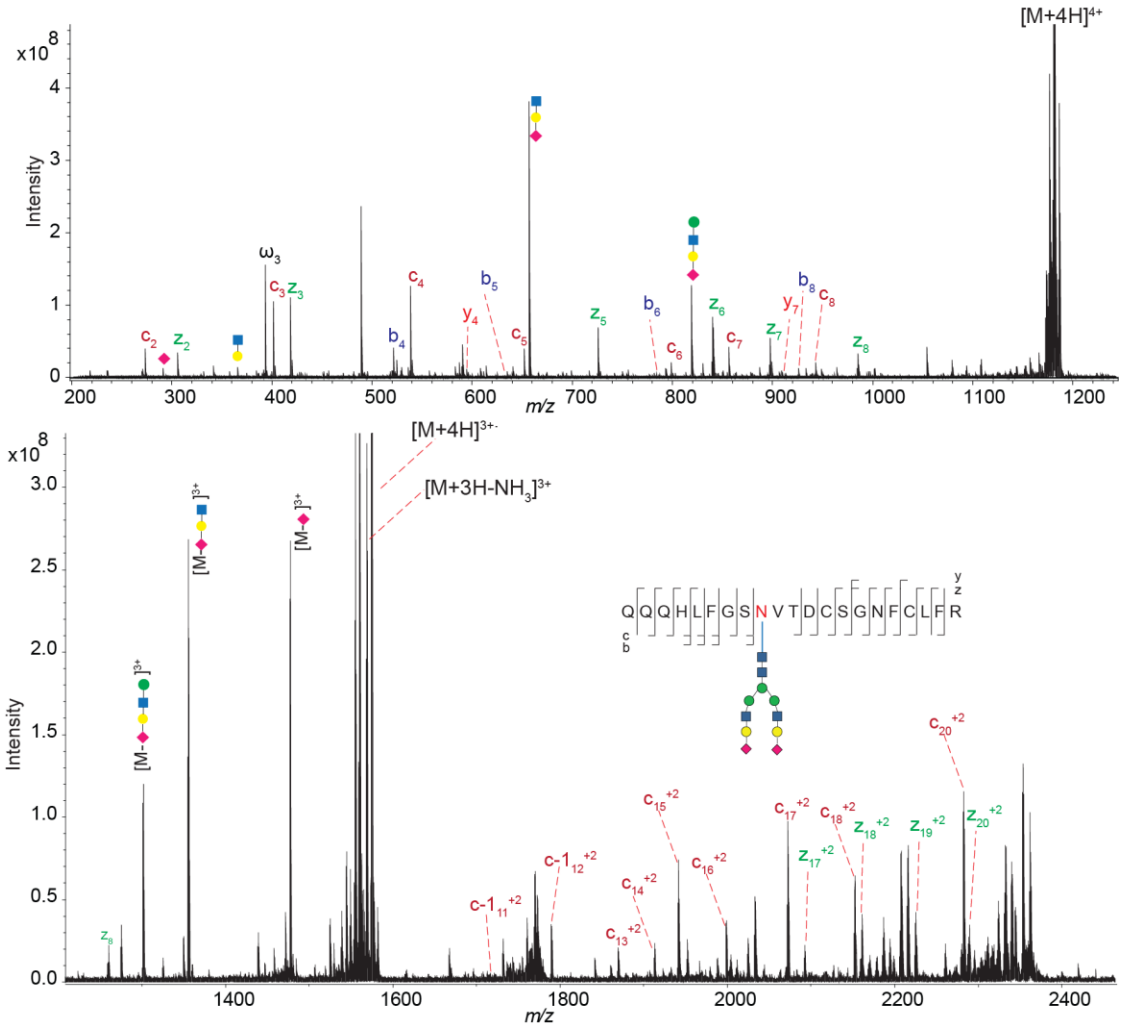


Figure 48: ECD Tandem MS of a Transferrin glycopeptide ($[M+4H]^{4+}$, *m/z* 1180.7328) on an FTICR-MS instrument

While ECD of the 4+ precursor of the transferrin glycopeptide generated abundant peptide backbone product ions, a triply charged precursor of the same glycopeptide led to Electron Capture with no Dissociation (ECnoD) and therefore did not yield useful tandem MS (data not shown but presented in [425]). Further experiments

were performed in an attempt to understand and alleviate this problem of charge-state dependence.

To improve glycopeptide fragmentation for lower-charge precursors by ECD, we first desialylated the transferrin glycopeptide samples using α-2,3,6,8 Neuraminidase. Desialylation was performed to collapse the overall glycan microheterogeneity for the same, thereby increasing precursor ion abundances and to eliminate any intramolecular interactions of the sialic acids preventing efficient dissociation of the odd-electron species. However, increased precursor abundance and removal of sialic acids did not seem to improve ECD performance for lower-charge precursors (Figure 49). Only charge-reduced species were observed in addition to low-abundance oxonium ions, confirming presence of ECnoD. To circumvent the problem of charge-state dependence, we tested hECD for the same precursor. hECD yielded abundant glycopeptide fragments for the same triply charged precursor that could not be fragmented by ECD (infusion-hECD data not shown).

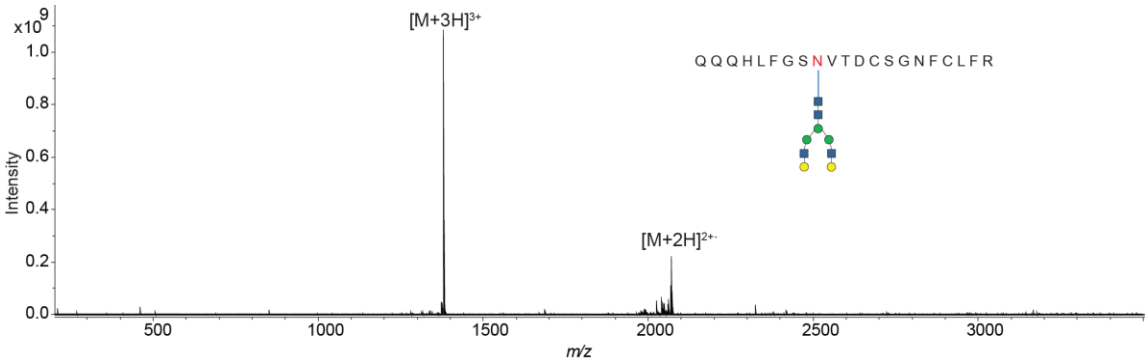


Figure 49: ECD Tandem MS of an asialo Transferrin glycopeptide ($[M+3H]^{3+}$, *m/z* 1379.9068) on a FTICR-MS instrument

The ExD fragmentation modes tested so far all used static nanoelectrospray on the FTICR instrument. Practical application of these analytical methods requires their implementation for LC-scale experiments to allow efficient characterization of the glycoproteome. Owing to the high fragmentation efficiency of hECD, we decided to implement this method with online LC-separation for fragmentation of Transferrin glycopeptides. Figure 50 shows the LC-hECD results for enriched glycopeptides from a Transferrin tryptic digest. Comprehensive glycopeptide structural information was generated by hECD in a single tandem-MS scan, where the peptide backbone was covered by both c/z- and b/y-type ions. In addition, many secondary cleavages were observed in the tandem-MS. While the increased electron energy caused some glycan dissociation, many peptide backbone fragments with intact glycan attached, were observed. Overall, hECD generated the most complete glycopeptide fragmentation among all ExD modes evaluated. While CAD is still more efficient for bottom-up glycopeptide fragmentation, its applicability to peptides with more than one glycosylation site is limited due to the loss of glycan linkage information in vibrational excitation, which prevents assignment of site-specific glycan compositions.

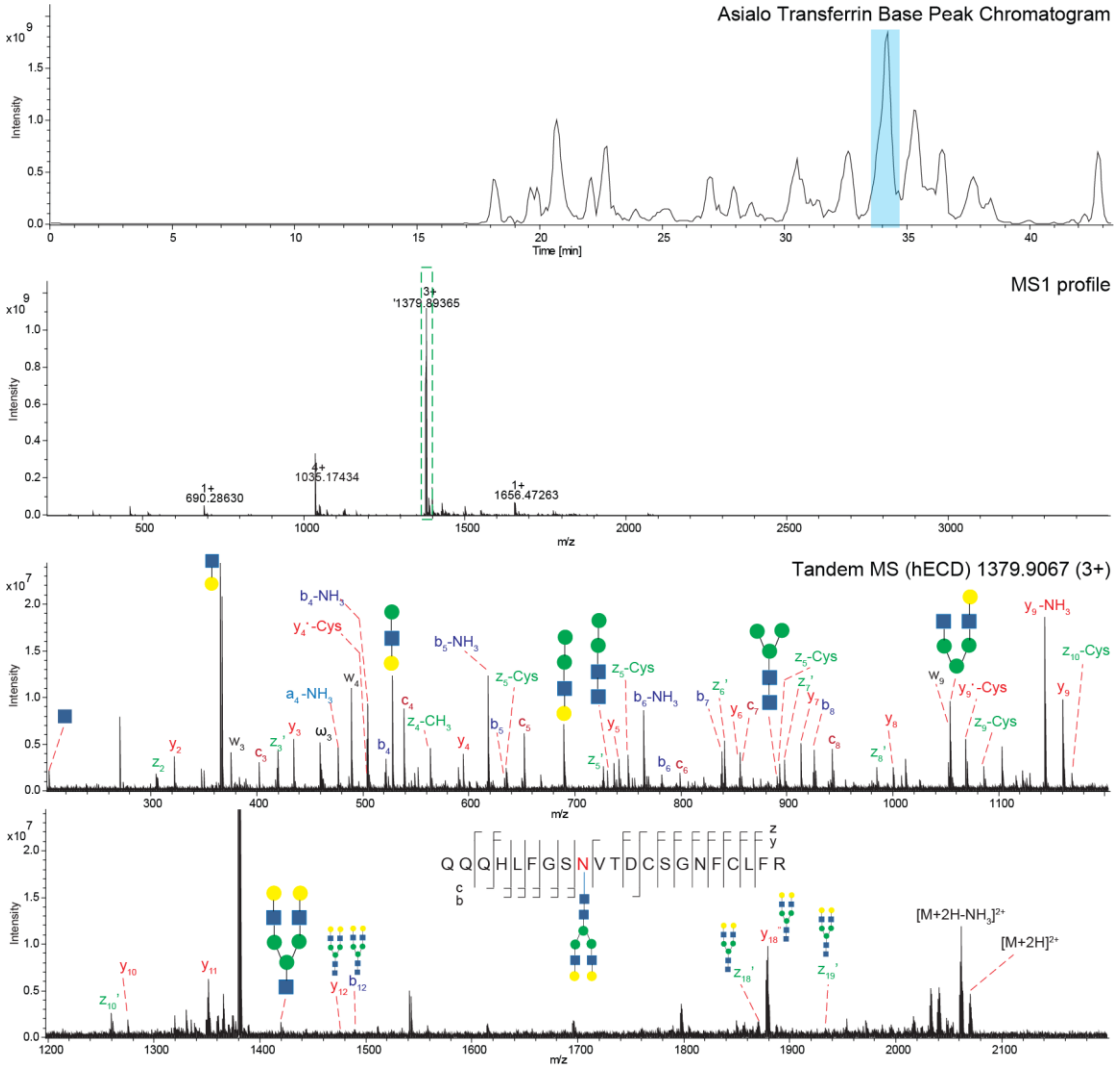


Figure 50: LC-hECD of enriched tryptic glycopeptides from human Transferrin

However, hECD can only be performed on FTICR-MS instruments to improve fragmentation of glycopeptides, when ECD or ETD fragmentation is not efficient. FTICR-MS instruments suffer from the lack of sensitivity and speed prohibiting their use in large-scale glycoproteomics studies. While FITCR MS is well suited for targeted analyses, the most successful implementation of ExD for large scale studies has been

possible on hybrid Orbitrap instruments capable of performing ETD [370,486,548,567,619]. Many groups have demonstrated further improvement in the efficiency of ETD fragmentation on Orbitrap instruments by use of supplemental collisional activation [78,161,331,512].

We evaluated collisional dissociation (HCD) (Figure 51) and EThcD (ETD with supplemental collisional activation) (Figure 52) of an AGP glycopeptide on a Thermo Orbitrap Fusion instrument. As expected, collisional dissociation of the precursor by HCD primarily generated peptide backbone fragments while the glycan was reduced to mono-, di- and tri-saccharide oxonium ions. In addition to bare peptide backbone ions, y-type ions with an attached HexNAc can be observed. Also, an intact peptide ion with a HexNAc (Glycan Y_1 ion) is seen in the spectrum. EThcD is performed by first allowing the precursor ions to react with the electron donating radical anion, which leads to formation of the odd-electron species. In many cases, this reaction leads to formation of a charge-reduced species. The ETD reaction takes place in the LTQ, after which these ions get transferred to the HCD cell, where they are subjected to gentle collisional activation that facilitates efficient dissociation of the peptide backbone by disrupting non-covalent interactions, while keeping the glycan largely intact. As seen in Figure 52, EThcD of an AGP glycopeptide generates extensive coverage with c-type ions and a few b-ions. The z_7 ion is observed with the intact glycan attached. Additionally, Y_1 and Y_2 ions, saccharides losses from the precursor and oxonium ions are seen due to collisional-activation but their abundances are much lower relative to backbone ions as compared to HCD spectra.

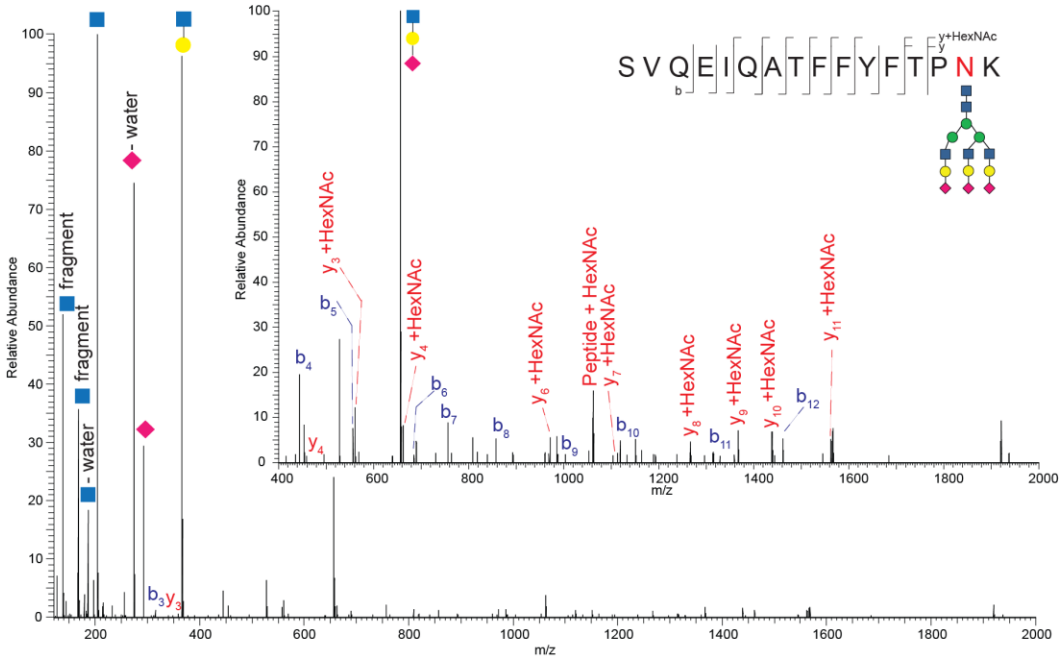


Figure 51: HCD of a Human AGP glycopeptide on Orbitrap-MS (*m/z*: 1195.9939, $[M+4H]^{4+}$)

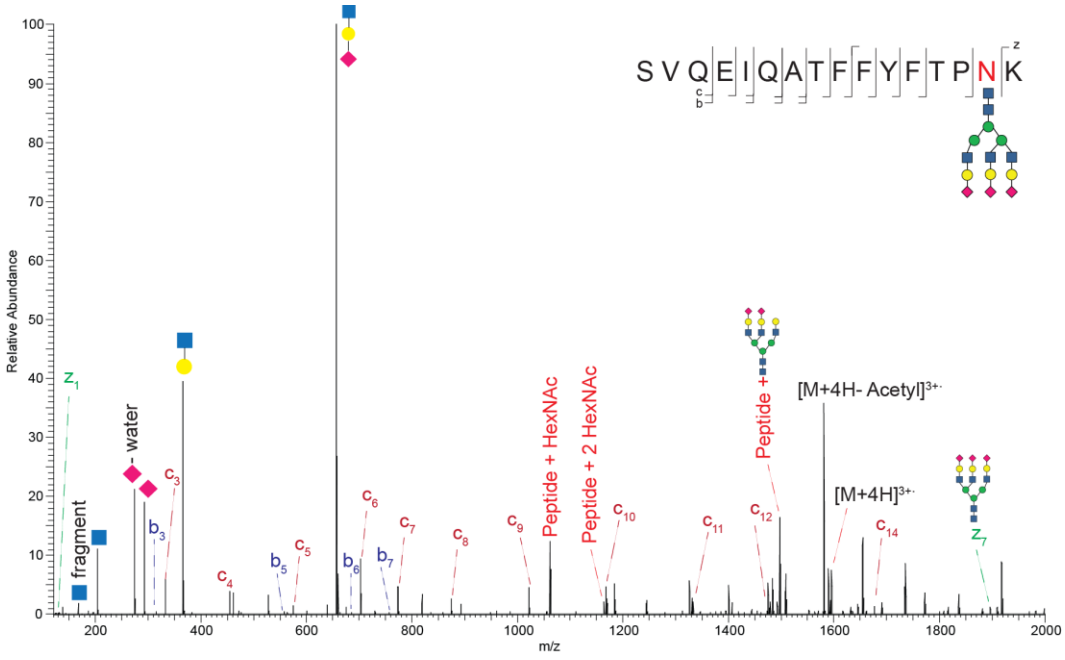


Figure 52: EThcD of a Human AGP glycopeptide on Orbitrap-MS (*m/z*: 1195.9939, $[M+4H]^{4+}$)

Supplemental activation can also be performed prior to electron-based activation by photodissociation methods such as IRMPD, as shown by the Coon group [446,366,442]. Supplemental activation using collisional and photodissociation methods generally decreases the charge-state dependency of ETD and ECD and has been shown to improve fragmentation for both bottom-up and top-down analyses [56,78,480].

From our evaluation of the different dissociation modes for bottom-up glycopeptide fragmentation, CAD based fragmentation was found to be the most efficient in terms of speed and sensitivity. However, when considering the level and quality of information generated, CAD suffered from poor abundance of peptide backbone fragment ions and inability to resolve multiple glycans attached to the same peptide. EThcD and hECD generated abundant peptide fragments, while keeping the glycan moiety intact during the fragmentation process, thus indicating potential for application in the analysis of middle-down glycopeptides with multiple glycosylation sites.

5.2.3.3 Method development for middle-down glycopeptide analysis

Bottom-up glycoproteomics has matured to the point where we can measure site-specific glycosylation in very complex biological matrices. For our purposes, bottom-up glycopeptide analysis refers to tandem mass spectral analysis of glycopeptides with one glycosylation site. Bottom-up analyses determine the distributions of glycosylation at specific sites but not relational information among the glycans that exist on different sequons on a mature glycoprotein simultaneously. As a result, we are able to model complete glycoprotein structures with glycan distributions at each site [291] but are

unable to generate structures of fully-glycosylated proteoforms present in a heterogeneous mixture. This hampers the complete biological characterization of glycoproteins in terms of identifying the most functionally relevant glycoforms. Middle-down glycoproteomics is a relatively new technique involving analysis of large glycopeptides with two or more glycosylation sites. By analyzing multiple glycosylation sites connected together, we can assign relationships between glycosylation sites and assess the interplay of glycosylation and structural features on the overall glycosylated proteoform.

The greatest challenge in glycopeptide analysis is the heterogeneity of glycoforms. As the number of glycosylation sites occurring together on a glycopeptide increases, the glycan heterogeneity multiplies thereby increasing the number of possible glycoform combinations exponentially. This extreme heterogeneity has prevented the development of efficient top- and middle-down glycoproteomics methodologies. Our efforts described in this section have focused on the development of offline and online separation methods and mass-spectral dissociation methods for middle-down glycopeptidomics.

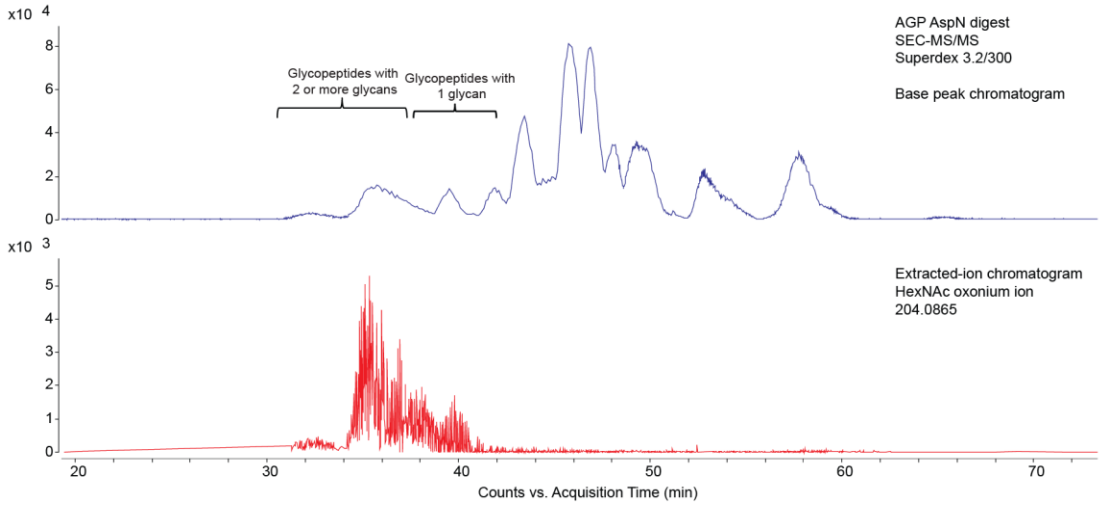


Figure 53: SEC-MS/MS (CAD) of Human AGP digested with AspN, showing separation of glycopeptides and peptides based on size and number of glycans attached

We selected the human AGP for middle-down methods development, which we and others have characterized in detail using bottom-up mass spectral methods [292,293,319,530]. AGP has 5 *N*-glycosylation sites, each with a variety of complex-type *N*-glycans. Furthermore, AGP occurs as a mixture of two protein isoforms (AGP1 and AGP2) [292,406]. Digestion of AGP using endoproteinase AspN generates multiple peptides with two *N*-linked glycosylation sequons. Therefore, AGP is an ideal glycoprotein for development and evaluation of methods for middle-down glycopeptide analysis. Since, glycan microheterogeneity poses the greatest challenge in analysis of peptides with multiple glycosylation sites; we decided to reduce glycan heterogeneity by trimming them from the non-reducing end. We used α-2,3,6,8 Neuraminidase to remove NeuAc residues from the AGP AspN digests. The resulting mixtures after proteolysis contained unglycosylated peptides and glycopeptides with 1, 2 or 3 glycosylation sites. To resolve glycopeptides with 2 glycosylation sites from those with fewer or greater

number of sites, we used a size exclusion chromatography fractionation step, as shown in Figure 53 and described above. SEC provided more specificity in enrichment of glycopeptides with 2 glycosylation sites over HILIC SPE techniques, such as those used for bottom-up sample preparation.

After further C18 fractionation of the SEC fractions containing two glycosylation sites, the glycopeptides were analyzed by FTICR-MS. We compared the different ExD modes available on the FTICR-MS for middle-down glycopeptide analysis. As expected, ETD of middle-down glycopeptides only produced charge-reduced species, similar to bottom-up glycopeptides and while ECD produced a few fragments covering the ends of the peptide, no fragmentation between the glycosylation sites was observed (data not shown). Middle-down glycopeptides were then irradiated with higher energy electrons to achieve hECD. We analyzed a few examples of hECD spectra to assess the ability of this fragmentation mode to resolve glycan compositions on the two *N*-glycosylation sites.

As shown in Figure 54, comprehensive fragmentation of a middle-down glycopeptide from AGP could be achieved using hECD. Predominantly c-type ions were seen in the tandem-MS. Overall, better coverage was observed along the peptide portions not carrying glycosylation, while peptide backbone coverage between the two glycosylation sites was sparse. Nevertheless, a few c-type fragment ions carrying the intact glycan could be identified. Vibrational excitation also led to generation of three peptide b-ions that still had intact glycans attached to them. A few z-ions were also observed in the tandem-MS complementing the information from the c-ion series. This example had the same glycan composition present at the two glycosylation sites. Another

middle-down glycopeptide tandem-mass spectrum is shown in Figure 55, where two different glycan compositions were identified at the two glycosylation sites. In this case, abundant c- and z-type ion series were observed again covering the N- and C-termini of the peptide, along with presence of y-type ions from vibrational excitation. Coverage between the two glycosylation sites was again poor with only one fragment ion each of c- and z-type identified.

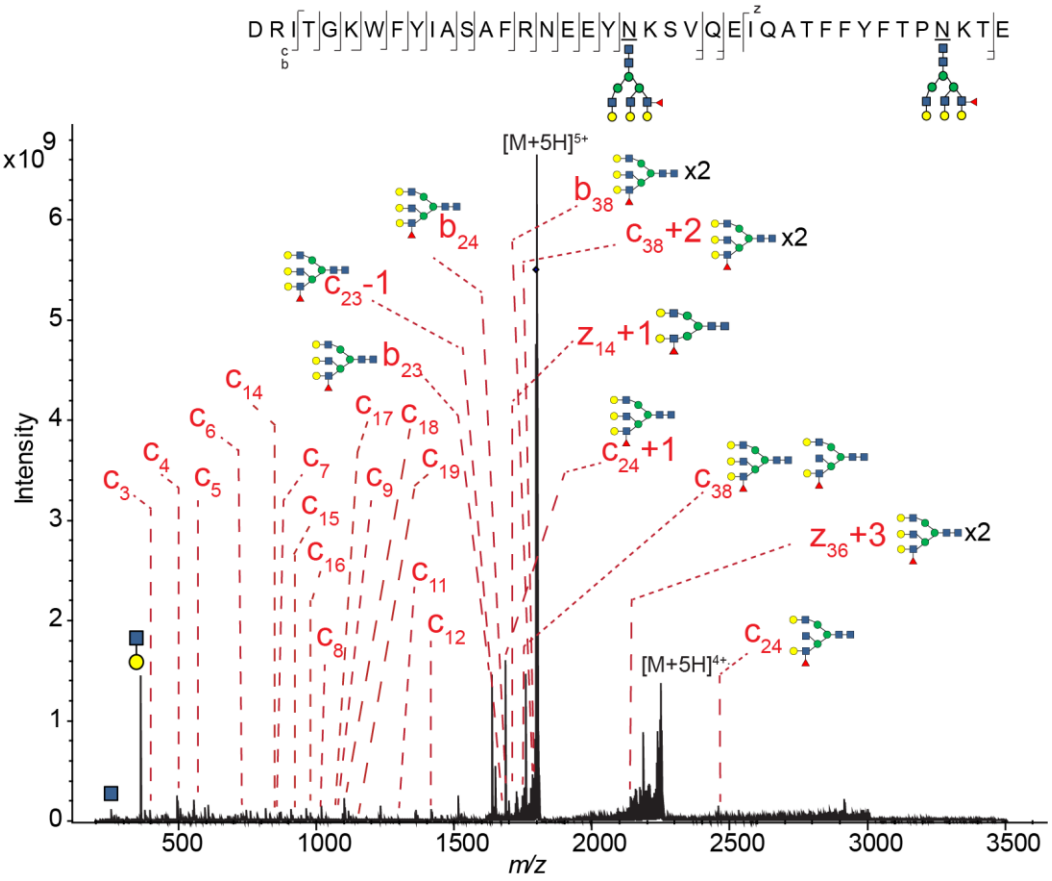


Figure 54: Tandem MS (hECD) of an AGP AspN glycopeptide with two glycosylation sites on FTICR-MS (*m/z*: 1802.1786, $[M+5H]^{5+}$)

Our results for hECD fragmentation of middle-down glycopeptides with two glycosylation sites indicated that although overall the peptide backbone was reasonably

well-covered, coverage between the glycosylation sites needed improvement. Despite the poor coverage between glycosylation sites, we show that presence of as few as one or two fragment ions between the glycosylation sequons is sufficient to discriminate the exact glycan compositions simultaneously occupying the glycopeptide. Data acquisition on the FTICR-MS for middle-down glycopeptide tandem-MS was performed using static nano-ESI and could not be coupled to online LC-separations.

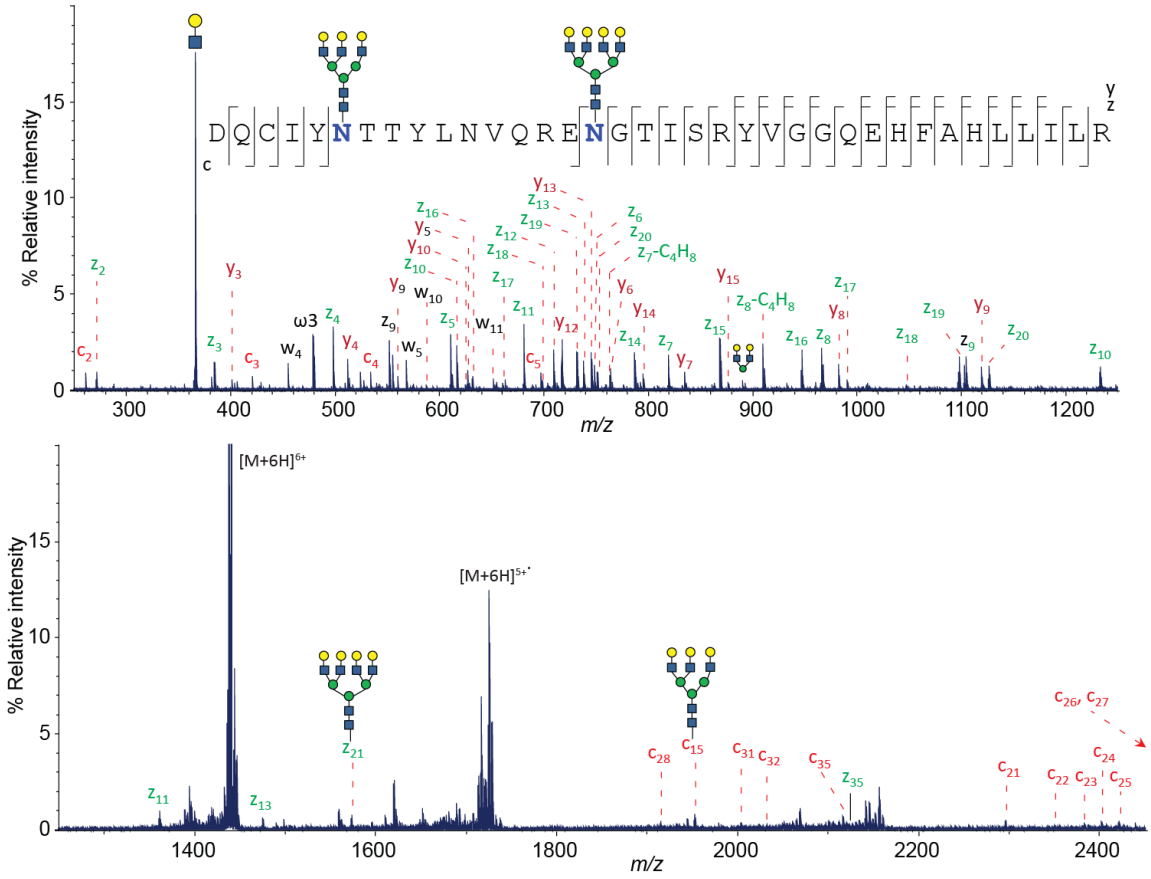


Figure 55: Tandem MS (hECD) of an AGP AspN glycopeptide with two glycosylation sites on FTICR-MS (*m/z*: 1437.4571, $[M+6H]^{6+}$)

Since, the hybrid Orbitrap instruments had shown promising results for online LC-MS/MS of bottom-up glycopeptides using EThcD, we evaluated the performance of

this method for middle-down glycopeptides. HILIC SPE-enriched glycopeptides were subjected to online C18 LC-MS on an Orbitrap Fusion instrument. Similar to bottom-up glycopeptide separation, middle-down glycopeptide resolution on the reversed-phase column was peptide-centric leading to clustering of the different glycoforms for the same glycopeptide in the LC-elution profile. An example MS1 spectrum showing a distribution of co-eluting middle-down glycopeptide glycoforms is shown in Figure 56. The higher speed and sensitivity of Orbitrap-MS made data-dependent LC-tandem-MS more feasible on this instrument, compared to FTICR-MS.

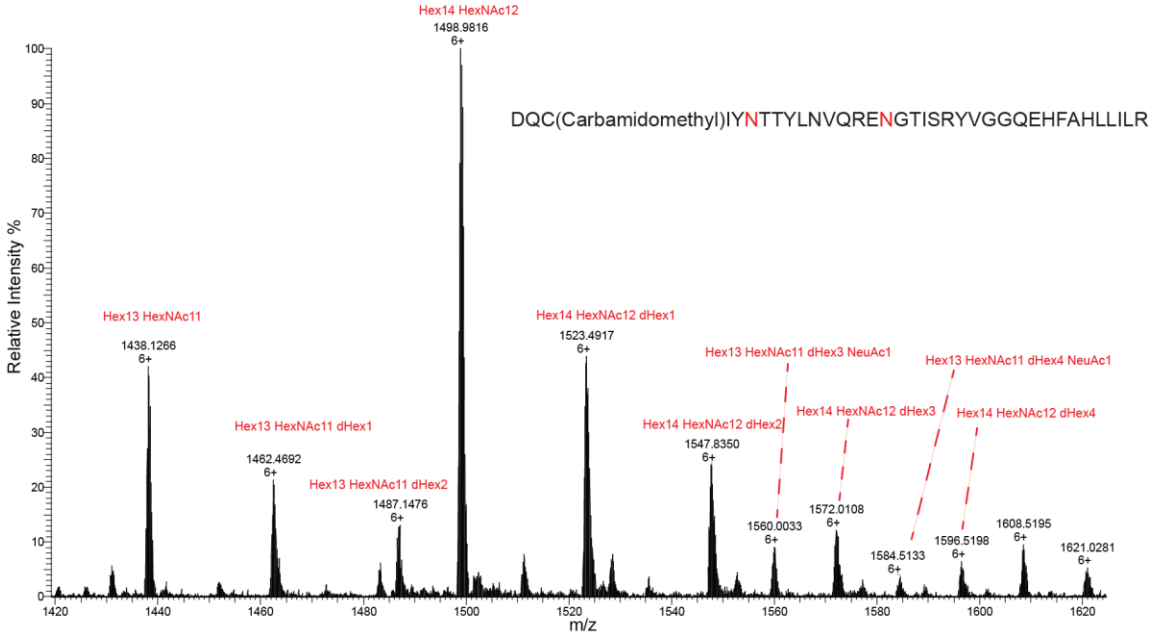


Figure 56: A MS1 spectrum showing glycan heterogeneity for an AGP middle-down glycopeptide with two glycosylation sites

While Orbitrap-MS was able to sample many more precursors for tandem-MS during the data-dependent acquisition experiment, only a few middle-down glycopeptides generated fragments suitable for resolving glycosylation site microheterogeneity. As shown in Figure 57, the terminal regions of the peptide backbone are well covered in the

tandem-MS but coverage for fragment ions carrying the glycan moieties was again sparse, similar to EThcD results. A z-ion series was predominant for this glycopeptide, with only two ions covering the N-terminal region of the peptide. Supplemental collisional activation also generated y-type fragment ions and glycan oxonium ions. The two z-type ions with intact glycan attached, covering region between two glycosylation sites provided adequate confidence in assigning the glycan compositions at the two sites.

From our evaluation of methods for middle-down glycopeptide analysis, it is clear that front-end separation plays a crucial role in managing the extreme heterogeneity of glycoforms. Even with a combination of different enrichment, fractionation and online separation methods, the glycopeptide complexity extended far beyond what is typically seen for bottom-up glycopeptides. Another level of complexity comes from the presence of glycoform combinations that are isobaric due to presence of glycosylation site positional isomers. Such isobaric compositions can't be resolved using liquid-phase separation methods and generate chimeric tandem-MS due to co-isolation in the mass spectrometer. Gas-phase separation methods such as ion-mobility might be useful in separating such molecules prior to fragmentation. Thus, successful middle-down glycopeptide analysis would require a combination of analytical tools.

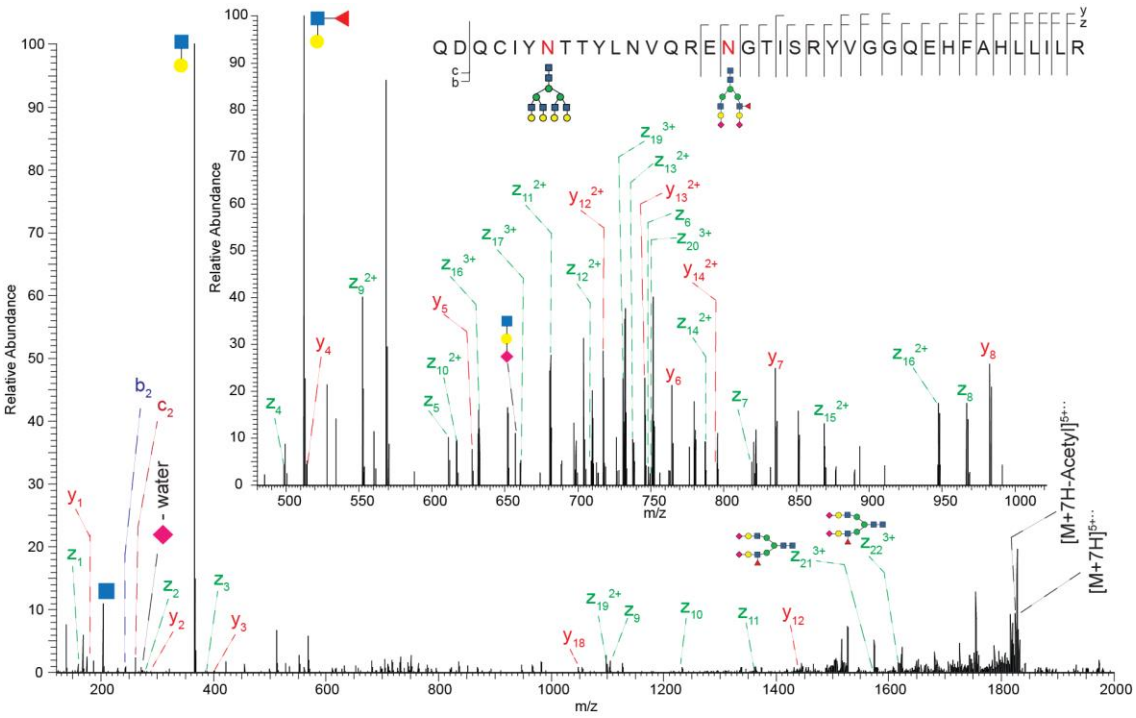


Figure 57: Tandem MS (EThcD) of an AGP middle-down glycopeptide on an Orbitrap Fusion instrument (precursor *m/z*: 1308.2788, $[M+7H]^{7+}$)

Our pilot study focusing on method development for middle-down glycopeptide analysis has helped identify features in different types of analytical methods and their suitability for separation and analysis. We have also generated a dataset, which would enable development of bioinformatics strategies. The data discussed here were analyzed using a combination of manual interpretation and programmatic data analysis tools prototyped written using Pyteomics [182] and GlyPy [300] (https://pypi.python.org/pypi/glypy/0.0.5rc2) Python libraries. It is clear from these analyses that middle-down glycopeptide analyses alone are not sufficient for efficient bioinformatics and defining glycosylation site relationships would require a combination of bottom-up and middle-down information sets. Our proposed workflow would require

integration of the two data domains for efficient and comprehensive data analysis, as shown in Figure 58.

Further development of analytical methods is necessary to enable acquisition of high-quality middle-down glycopeptide tandem MS. While hECD and EThcD have shown promising results in the analysis of multiply glycosylated peptides, the sensitivity and speed of these methods remains mediocre. Several recent studies have underlined the value of combining collisional and photo-activation modes with electron-based activation methods [61,239,344,366,442,512]. At the same time, electron-based activation modes are now being implemented on faster, more sensitive instruments [551,552,553]. Technology development in these areas will play a key role in advancing top- and middle-down workflows and dissemination of these methods in the glycol-analytics community.

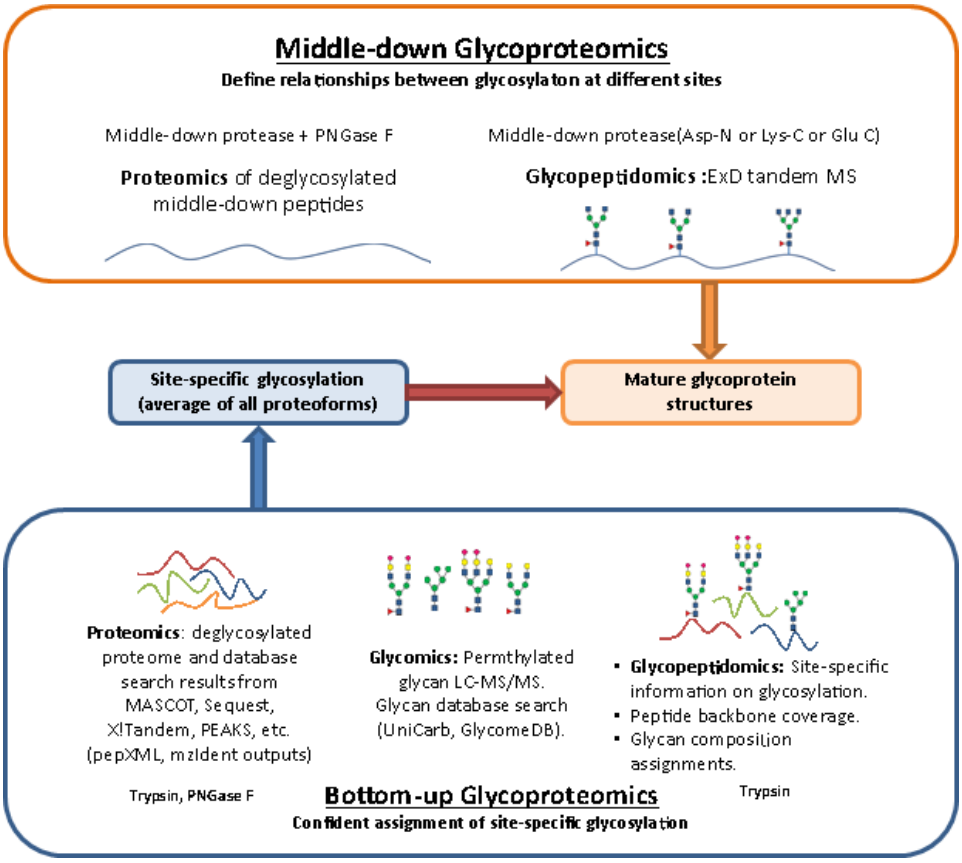


Figure 58: Proposed workflow for combining bottom-up and middle-down glycoproteomics

5.2.3.4 N-glycan ExD and HCD data generation for bioinformatics development

A major factor affecting glycan data interpretation is the lack of high-quality training data. Glycans from biological sources generally occur as mixtures of isomeric forms and are generally limited in quantities, which makes it difficult to obtain pure glycan structures for ExD experiments. While our group has developed and applied electron-based dissociation methods, such as EED, for detailed characterization of glycans, the extent of these applications has remained limited to relatively simple glycans [605]. We recently obtained chemo-enzymatically synthesized glycan standards from

Chemily, LLC, containing larger, branched oligosaccharides with *N*-glycan like structures. We generated collisional (HCD) and electron-activated (EED) tandem-MS data on nine (9) different glycan structures shown in Figure 59, with their identifiers for correlating with data files. All glycans were deutero-reduced and permethylated before analyses and precursors with different charge states for sodiated ions were analyzed by MS/MS. HCD data were acquired using a range of different collision energies. These data are being used to support the development of GlycoDeNovo, a *de novo* glycan data analysis program being developed in collaboration with Prof. Pengyu Hong at Brandeis University (Waltham, MA) and for mining features and development of in-house glycan data interpretation tools like GlyPy (https://pypi.python.org/pypi/glypy/0.0.5rc2).

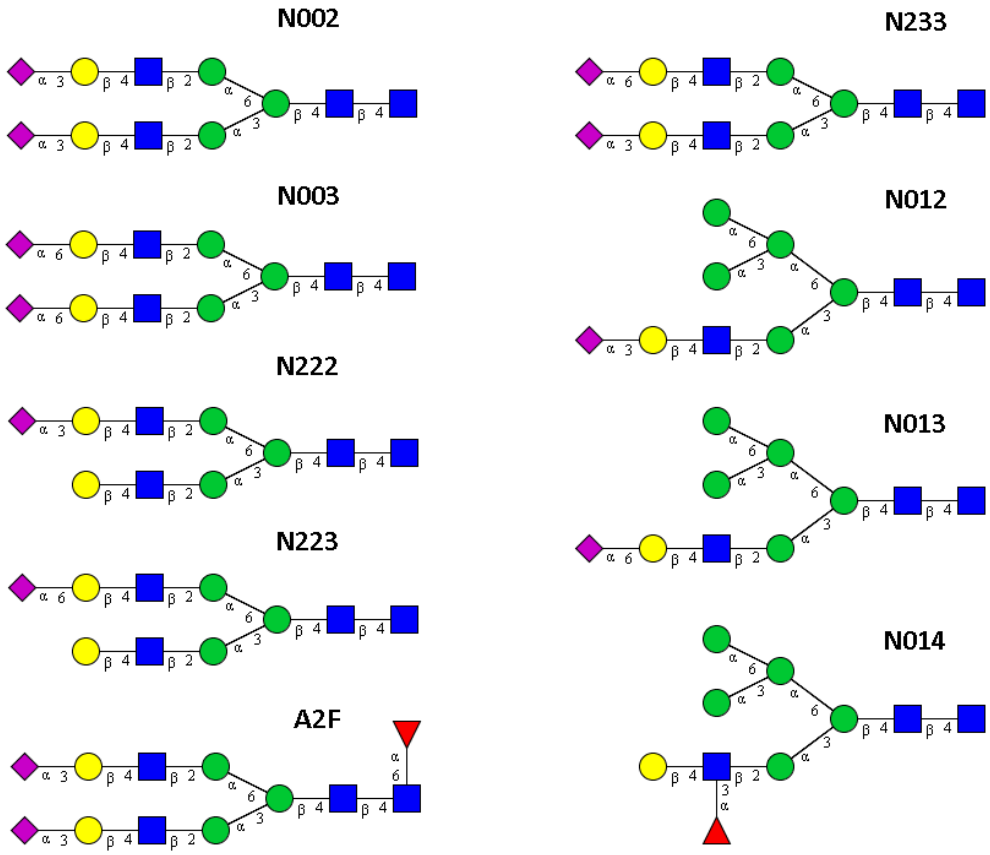


Figure 59: Pure glycan standards analyzed by EED and HCD for bioinformatics development

Exploratory data analysis was performed on HCD and EED data for N002 and N014, using the GlyPy machinery. Theoretical fragment ions were generated for both structures and matched programmatically against deconvoluted and deisotoped peaklists. Glycans subjected to CAD and ExD yield different fragment types. Initial tests for fragmentation using ETD and ECD of sodium adducted permethylated glycan alditols primarily resulted in charge reduction with poor fragmentation. EED resulted in extensive fragmentation of the permethylated glycan alditols. CAD (HCD) generated abundant glycosidic bond fragments at lower collision energies but with increase in collision

energies, cross-ring cleavages were apparent in the spectra. However, fragments from both dissociation modes had multiple ambiguous matches in a theoretical fragment database necessitating the use of more sophisticated bioinformatics approaches along with complementary analytical techniques for confident assignment of glycan fine structure.

5.2.4 Summary

In summary, we made significant advancement in analytical methodologies for middle-down glycopeptide analysis and compared the characteristics of different dissociation modes for both bottom-up and middle-down glycopeptide analysis. This work will contribute towards development of a comprehensive workflow for middle-down glycopeptide analysis and the training data generated will be used toward development of informatics strategies for both glycan and glycopeptide mass spectral data.

5.2.5 Acknowledgments

This work was supported by NIH grants P41 GM104603 and S10 RR025082. Thermo Fisher Scientific provided access to the Orbitrap Fusion instrument.

5.3 Improved analysis of sulfated oligosaccharides using size-exclusion chromatography-ion-suppressor-high resolution mass spectrometry

(This work has been published in [611])

5.3.1 Introduction

Pharmaceutical heparin is a sulfated, linear polysaccharide prepared from porcine intestinal mucosa. Consisting of repeating uronic acid-glucosamine disaccharide units modified with an average of 2.3 sulfate groups, unfractionated heparin has an average molecular weight of 17,800 Da with a high degree of polydispersity [388]. Pharmaceutical heparin, used as an anticoagulant in surgical suites, is neutralizable with protamine and requires constant patient monitoring. Heparin-induced thrombocytopenia, a serious risk of heparin treatment, results from an immunological reaction against a complex between heparin and platelet factor 4 [42,379]. Low molecular weight heparins (LMWH) were developed in order to minimize such risks to allow administration as an outpatient anticoagulant. LMWH are produced from unfractionated heparin by limited chemical or enzymatic cleavage [328] to produce a biopolymer with average molecular weight of 4000-6000 Da, depending on the preparation [188]. The chemical structure and pharmacological activities of LMWH depend on the method used for partial depolymerization [328] (see Figure 60).

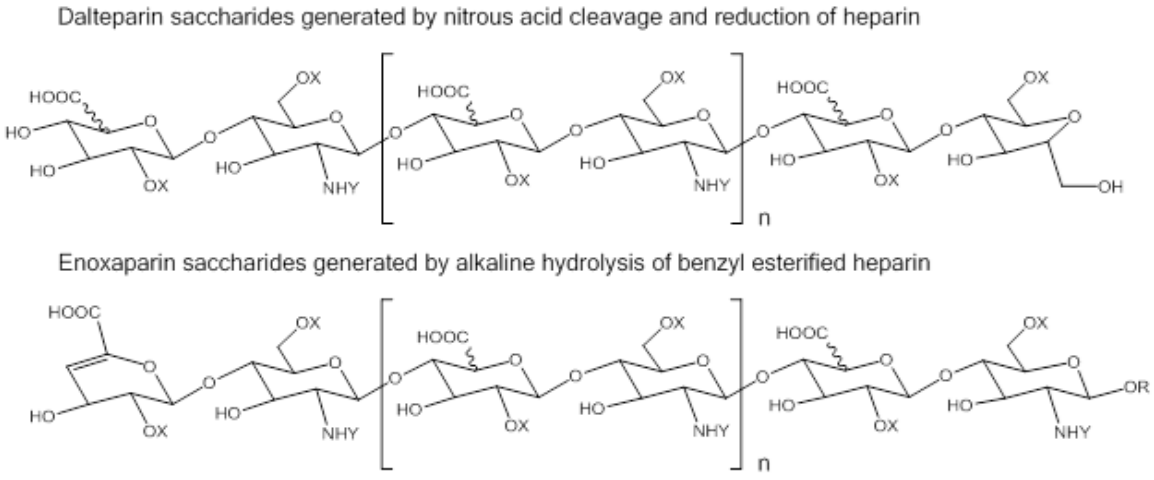


Figure 60: General structures for dalteparin and enoxaparin saccharides. X = H or SO3H, Y = Ac or SO3H

Since their introduction, LMWH have captured a large share of the US heparin market. This has driven effort to develop biosimilar heparins, the first of which for the enoxaparin LMWH was approved by the Food and Drug Administration in 2010 [642]. Because the complexity is vast, it is not possible to characterize every molecular species in LMWH preparations; thus, the FDA approval was based on the ability to reverse engineer LMWH using analytical and pharmacologic assay data representative of the preparation. Such biosimilars development depends on the demonstration of acceptably low risk for use of the biosimilar heparin. The ability to demonstrate acceptable risk improves with increases in analytical capability.

Molecular weight distributions for unfractionated heparin and LMWH have been measured traditionally using size exclusion chromatography with refractive index detection [386,387,388]. In order to increase the detail of the chromatographic LMWH separations we and others have analyzed LMWH fractions using hydrophilic interaction chromatography (HILIC)-MS [326,497]. While the chromatographic resolution of

HILIC is higher than that of SEC, the ability to interpret the mass spectral data is limited by the presence of abundant ammonium adducted forms of the LMWH oligosaccharide ions. This adduction results from pairing of ammonium cations with deprotonated oligosaccharide anions, resulting in proton transfer reactions that decrease the magnitude of ion charge. Thus, abundant ions with low degree of charge and high degree of ammonia adduction are observed. As a result it is not possible to extract neutral masses from LMWH saccharides larger than degree-of-polymerization (dp) 20 using LC-MS. Similar problems limit the use of MS to detect LMWH eluting from a SEC column since ammonium ions must be included in the mobile phase.

Tinzaparin LMWH was analyzed using 6 μm particle size SEC columns [228]. These investigators used a self-regenerating cation exchange column to remove ammonium salts prior to the MS source. Chromatographic resolution was modest but they reported ability to identify ions corresponding to Tinzaparin oligomers up to dp18. The availability of ultraperformance SEC columns with 1.7 μm particles greatly improved the chromatographic resolution for LC-MS separation of LMWH [617]. Again, the LMWH extracted mass spectra suffered from ammonium adduction, limiting the ability to interpret data for >dp 18.

In this work, we solved this ammonium adduction problem for SEC-MS of LMWH by using an on-line ion suppressor to exchange remove ammonium ions for protons, dramatically reducing the abundances of ammonium adducted LMWH ions. This allowed us to measure mass profiles for LMWH up to dp30, essentially the complete

LMWH biopolymeric range. We developed a specialized bioinformatics approach to interpret the data.

5.3.2 Experimental Section

5.3.2.1 Materials

The United States Pharmacopeia standard for enoxaparin sodium was purchased from Sigma Aldrich (St. Louis, MO). Dalteparin sodium as Fragmin was obtained from Pharmacia Corporation (Kalamazoo, MI).

Stock solutions of enoxaparin sodium and dalteparin sodium were prepared at 10 nmol/μL in water. They were then desalted by dialysis against 50 mM ammonium formate (pH 6.8) in methanol:water (20:80) using 500 Da molecular weight membranes (The Nest Group, Southborough, MA).

5.3.2.2 SEC LC-MS Analysis with Ion Suppressor

Size exclusion chromatography-ion suppressor-mass spectrometry (SEC-IS-MS) analyses were performed using Tandem Acquity UPLC BEH columns (4.6 mm × 150 mm and 4.6 mm × 300 mm, Waters Corp., Milford, MA) were used to separate the LMWH preparations using conditions similar to a published method[617]. The mobile phase contained 50 mM ammonium formate (pH 6.8) in methanol:water (80:20). The constituents were eluted in 90 min at a flow rate of 75 μL/min. The analyses were initiated and performed in part by Dr. Yuewei Sheng.

The LC column was connected to a chemically regenerated ion suppressor (ACRS 500, 2 mm, Thermo Fisher Scientific/Dionex, San Jose, CA) for online desalting. The ion

suppressor exchanged ammonium salts to their corresponding organic acids, thereby eliminating the problem of ammonium adduction of the saccharide ions. A solution of 100 mM sulfuric acid was used to regenerate the suppressor. The suppressor effluent was connected to the standard ESI source of a Q-Executive Plus mass spectrometer (Thermo Fisher Scientific, San Jose, CA). The mass spectra were acquired using the negative ionization mode at resolution 75,000 and *m/z* range 200-2000 Th with three microscans. The source parameters for the analyses were optimized to minimize sulfate loss and included a spray voltage of 3.8 kV, a capillary temperature of 280 ºC, a sheath gas flow of 10, and an S-lens RF level of 40.

5.3.2.3 Data Deconvolution

The acquired spectra were manually examined and averaged over retention time ranges using the Xcalibur software (version 3.0.63, ThermoFisher Scientific, San Jose, CA). The combined spectra were auto-deconvoluted by an in-house software, developed by Joshua A. Klein and Dr. Yuewei Sheng, GAGdecon (https://github.com/BostonUniversityCBMS/gagdecon). GAGdecon uses a database search algorithm for structural assignment at the MS level. It maps the isotopic distribution of theoretical species using a python implementation of the BRAIN (Baffling Recursive Algorithm for Isotopic distribution calculations) [128,246] available at (https://github.com/BostonUniversityCBMS/brainpy) to the experimental data. The data processing and validation was performed together with Joshua A. Klein and Prof. Joseph Zaia.

GAGdecon works by constructing a combinatorial database of GAG compositions, computing their theoretical monoisotopic mass at a range of user specified charge states. The composition includes the number of monosaccharide species: HexN, HexA, HexNAc, dHexA, ManN, as well as substituents and adducts: Ac, Na, K, Ca, and Li. The disaccharide pattern was extrapolated from the user input GAG species, and the counts of each monosaccharide scale with chain length (degree of polymerization, dp), which may either be a range, or a fixed value. GAGdecon can process either centroid or profile data, using a Lorentzian peak shape fit to calculate the centroids for profile peaks. Once the input spectra have been preprocessed, database search is carried out by searching each theoretical GAG composition and charge state, generating a theoretical isotopic cluster based on the elemental composition of the theoretical molecular composition, including the first 95% of the isotopic pattern. Each theoretical isotopic pattern was fitted to the peak list, producing a score of the form:

$$G = 2\sum_{i}\frac{O_i}{\sum_j O_j}\log\left(\frac{T_i/\sum_j T_j}{O_i/\sum_j O_j}\right) \qquad \textbf{Equation 21}$$

O_i is the abundance of the *i*th observed peak and T_i is the abundance of the *i*th theoretical peak. Optimal fits minimize$G$. Each fit with a score below a parameterized threshold is reported to the user.

5.3.3 Results and Discussion

We analyzed the two most widely used LMWH in the US, dalteparin and enoxaparin, the general structures for which are shown in Figure 60.

5.3.3.1 SEC-MS of dalteparin

Dalteparin is produced by limited nitrous acid depolymerization of unfractionated heparin, resulting in chains that range from 5600-6400 Da (European Pharmacopeia range)[188] that terminate with an anhydromannitol residue at the reducing end [482]. Notably, a side reaction occurs in that corresponds to GlcN ring contraction without chain cleavage (Figure 61) [196]. SEC-MS of a synthetic heparin saccharide (Arixtra) in the absence of an ion suppressor results in abundant ammonium adducted ions that complicate the assignment of exact masses and saccharide compositions. Use of a post-column ion suppressor results in observation of deprotonated ions with an absence of ammonium adducts (Figure 62).

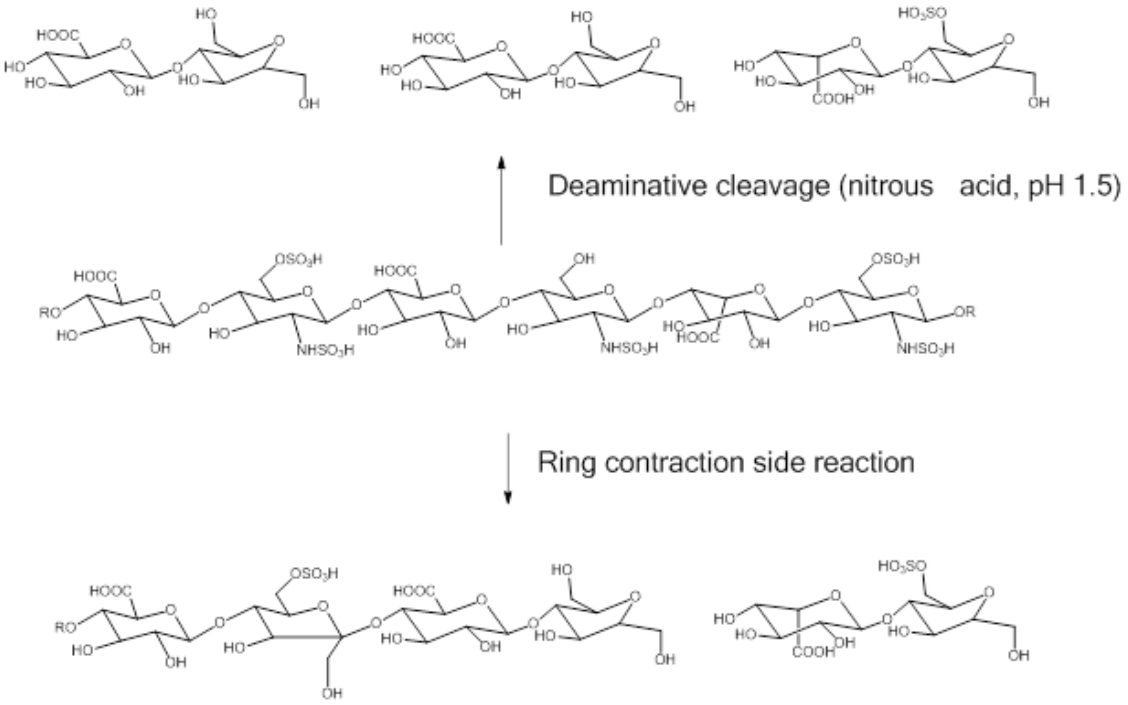


Figure 61: Top: deaminative cleavage of heparin using nitrous acid, pH 1.5. Bottom: ring contraction side reaction that occurs with deaminative cleavage, equivalent to loss of NH[196]

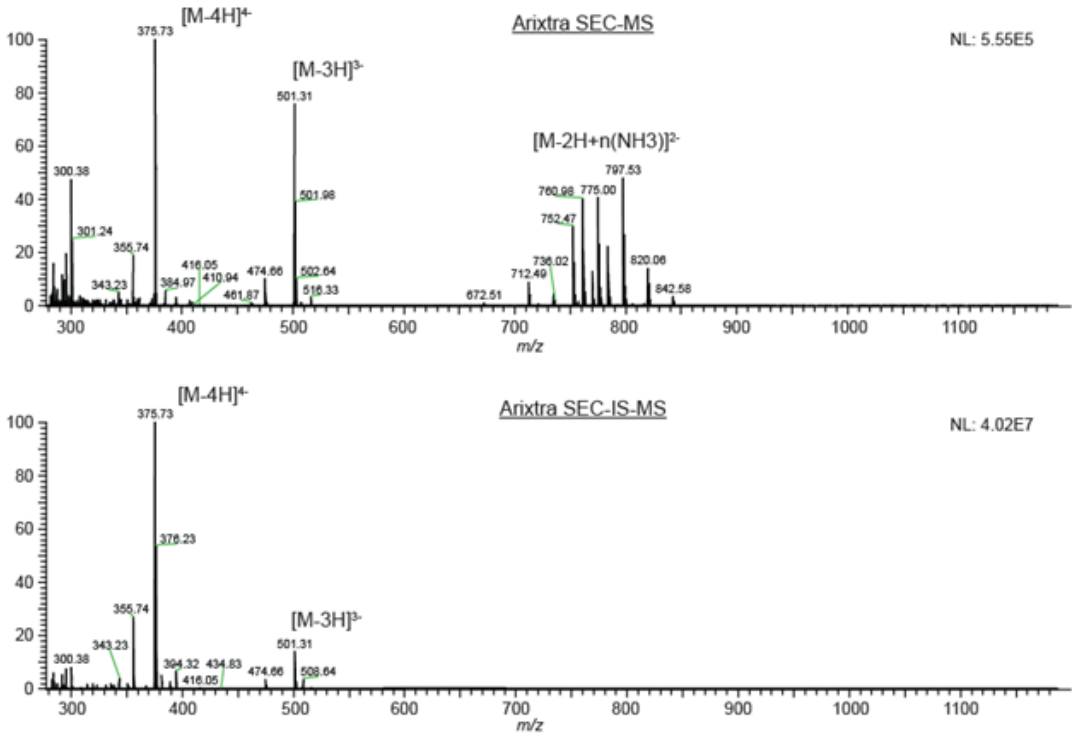


Figure 62: SEC-MS extracted mass spectra for Arixtra

Top: mass spectrum with no ion suppression, showing abundant ammonium adducts for the 2– charge state. Bottom: use of an online ion suppressor results in deprotonated ions with no ammonium adducts.

Figure 63a shows the total ion chromatogram (TIC) for SEC-ion suppressor (IS)-MS of dalteparin. Distinct chromatographic peaks are observed corresponding to saccharides ranging from dp6-30 as labeled in the figure. The extracted mass spectrum for dp16 saccharides obtained using the ion suppressor is shown in Figure 63b in which deprotonated ions are observed without ammonium adducts. This absence of adduction greatly facilitates interpretation of the mass spectral data; ammonium adducts increase the incidence of overlapping isotopic clusters and limit the size of heparin saccharides that yield interpretable mass spectra. We noticed several pairs of isotope clusters separated by 15.0 Da in the data. We reasoned that this resulted from deamination of

heparin chains without chain cleavage, resulting from the nitrous acid depolymerization reaction used in the formulation of dalteparin [196]. This corresponds to the substitution of an amino group with a proton for a net loss of NH (15.01 Da) (Figure 61). This pattern of paired isotopic clusters differing by NH units was observed throughout the SEC-IS-MS dataset on dalteparin. The extracted mass spectrum of dp26 saccharides (Figure 63c) show absence of ammonium adducts that allow unambiguous assignment of compositions. Saccharides resulting from NH loss were too low in abundances to be assigned.

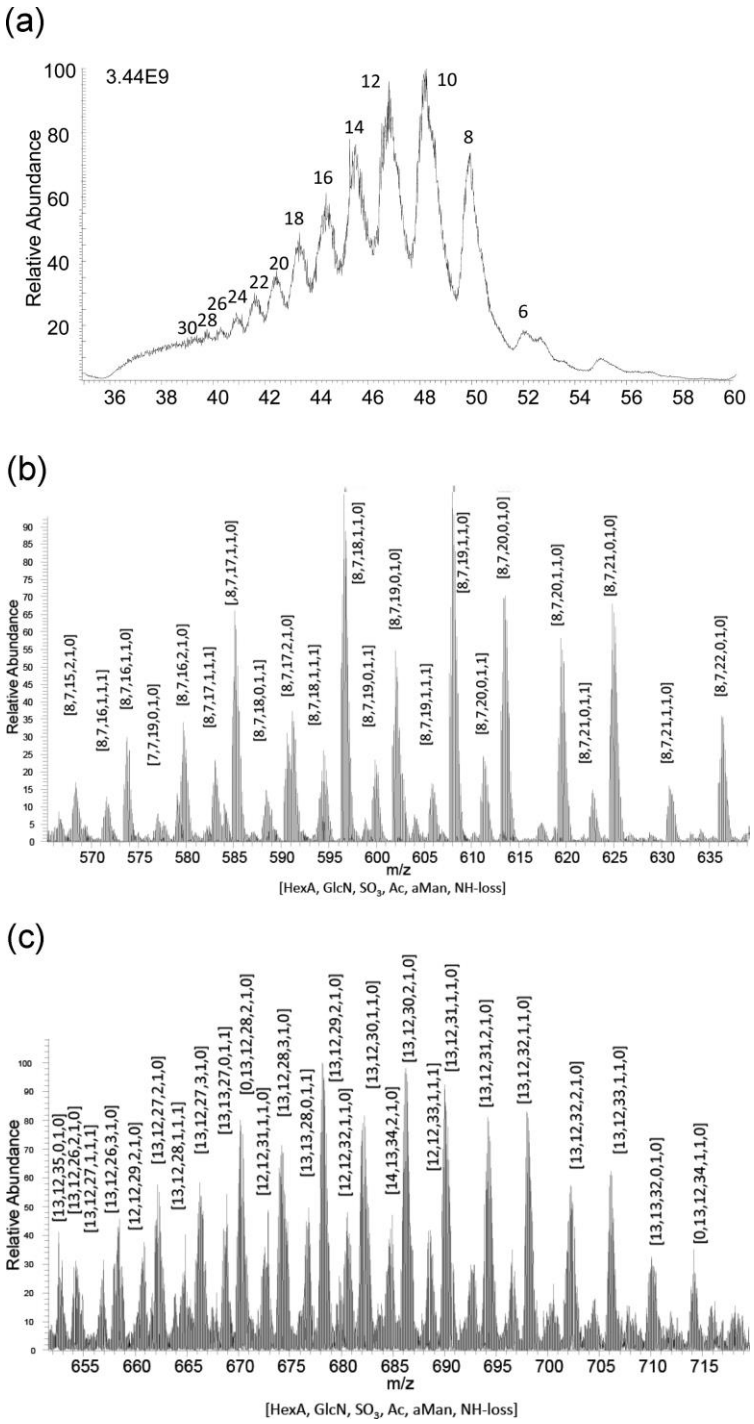


Figure 63: SEC-IS-MS of Dalteparin

(a) SEC-IS-MS total ion chromatogram for Dalteparin. The oligomer size (degree-of-polymerization) is shown for each chromatographic peak, (b) extracted mass spectrum for 43.6–44.7 min, and (c) extracted mass spectrum for 39.9–40.4 min.

In previous work, we used hydrophilic interaction chromatography (HILIC) MS for analysis of heparins [256,326,497]. In these studies we used Horn deconvolution [244,266,359] to convert mass spectra data to neutral masses. As shown in Figure 64, the decrease in abundances of the monoisotopic ions for LMWH saccharides that occur as molecular weight increases poses a serious problem for identification of monoisotopic ions for dp >18. For the dp14 and dp18 saccharides, the monoisotopic peak is readily identifiable. This identification becomes more difficult for larger saccharides as shown by the dp26 and dp30 examples because of the lower abundance of the monoisotopic peak. By overlaying the calculated isotopic patterns (shown in orange), we found that it was possible to assign dalteparin saccharides accurately through dp30.

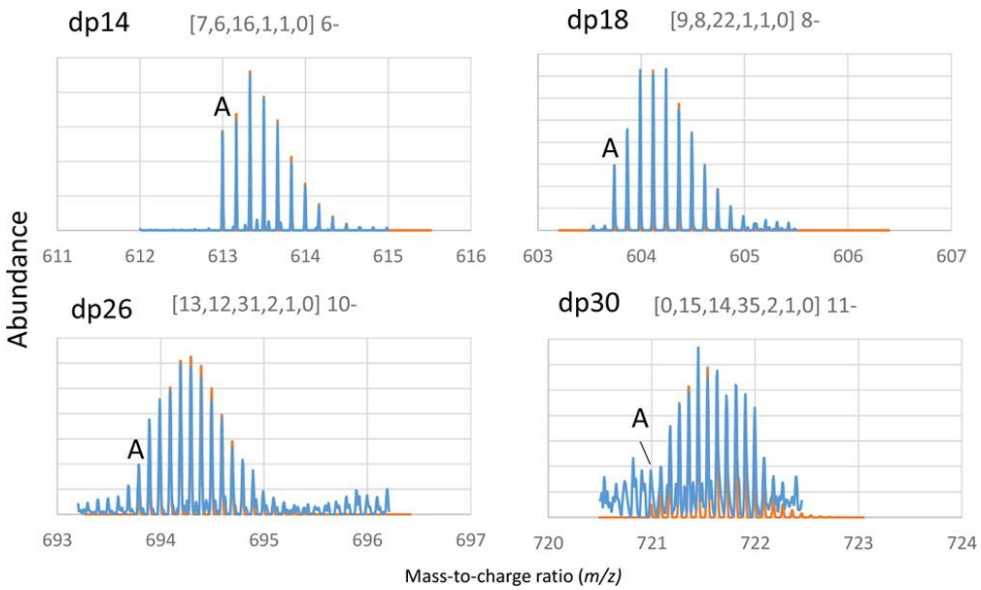


Figure 64: Dalteparin mass spectra – theoretical *vs.* observed

Expanded *m/z* ranges for (a) dp 8 [7,6,16,1,1,0] 6–, (b) dp 18 [9,8,22,1,1,0] 8–, (c) dp 26 [13,12, 31, 2, 1, 0] 10–, and (d) dp30 [15,14,35,2,1,0] 11–. Compositions are given as [HexA, HexN, SO3, Ac, aMan, NH loss]. Experimental data (blue) overlay theoretical isotope patterns (orange). Theoretical isotopic distributions were calculated using IsoPro 3.1 at 40 000 resolution.

In order to determine the LMWH compositions, it is essential for the algorithm to assign charge states correctly. Horn deconvolution seeks to identify the monoisotopic peak using a molecular averagine. This task becomes increasingly difficult as the LMWH oligomer size increases, to the point that it cannot easily be determined even by manual inspection (see examples for dp26 and dp30 in Figure 64). We therefore developed a bioinformatics algorithm (GAGdecon) to accomplish the task of isotopic pattern matching and scoring for LMWH datasets. For this, the algorithm generated theoretical dalteparin compositions and exact isotope patterns, provided with supplemental data of the published article[611].

As described in the methods section, exact masses for LMWH saccharides were assigned by fitting calculated isotope patterns. The compositional profile aggregated for all observed charge states for dalteparin dp16 saccharides is shown in Figure 65a. Scores were higher (poorer) for saccharides with NH losses, owing to the lower abundances and correspondingly greater difficulty in matching isotope clusters. For dp30 saccharides (Figure 65b) it was possible to assign only the compositions without NH losses with acceptable confidence. The complete set of compositional profiles are shown in the supplemental information for the published article[611]. These compositional profiles show a small degree of incorrect compositions for low abundance saccharides due to incorrect charge state assignments at half the correct value. This is due to multiple well-scoring interpretations of the signal which share peaks. These cases are uncommon, requiring that the isotopic pattern of one composition be similar on a relative

scale to the overlapped components of another composition. Often, a difference in scores can be used to discriminate the correct fit.

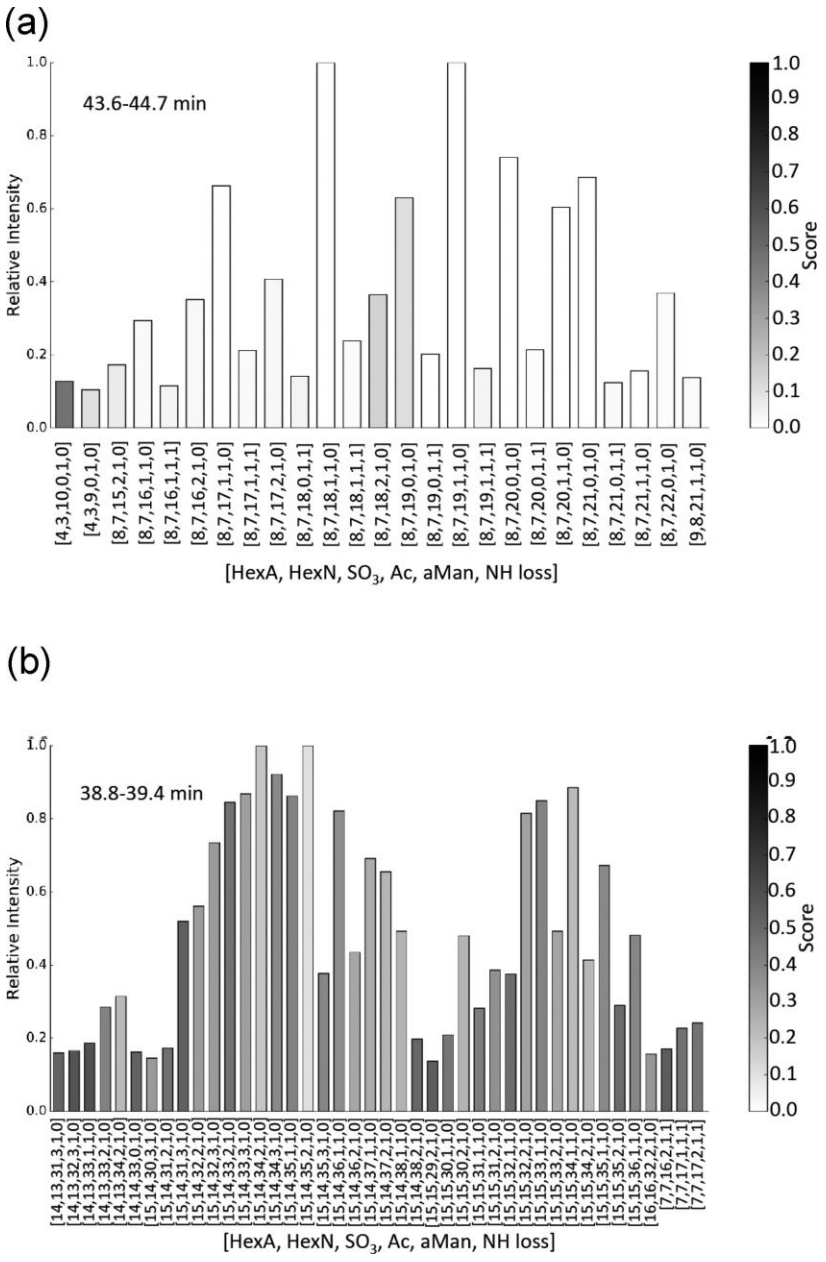


Figure 65: Compositions, scores, and abundances of dalteparin saccharides

(a) retention time 43.6–44.6 min, corresponding to dp16 and (b) retention time 38.8–39.4 min, corresponding to dp30. Ring contraction is given as NH loss in the compositions. The lowest score values reflect the greatest confidence.

In order to validate that the assignment of saccharide compositions was correct, we plotted the average saccharide chain length *versus* elution time. The resulting plot (Figure 66a) shows that average chain length diminishes with elution time as expected. This is consistent with overall correctness of assignment of saccharide compositions. Instances of incorrect assignments would skew this trend. This plot also demonstrates that charge state assignments were generally correct. We next calculated the number of sulfation per saccharide chain length across the SEC elution profile (Figure 66b). This showed that the number of sulfate groups ranged from 1.10 to 1.14, indicating a relatively narrow distribution of sulfation per saccharide across the elution profile. This corresponds to an average of 2.3 sulfate groups per disaccharide unit for dalteparin, in close agreement with the expected value [388]}. We also plotted the number of acetate groups per dalteparin disaccharide unit as shown in Figure 66c. This plot showed that the number acetate groups per disaccharide unit diminished as the dalteparin chain length diminished. This is expected since nitrous acid cleavage occurs at *N*-sulfated glucosamine residues to produce anhydromannitol, requiring the absence of acetate from the reducing terminal residue.

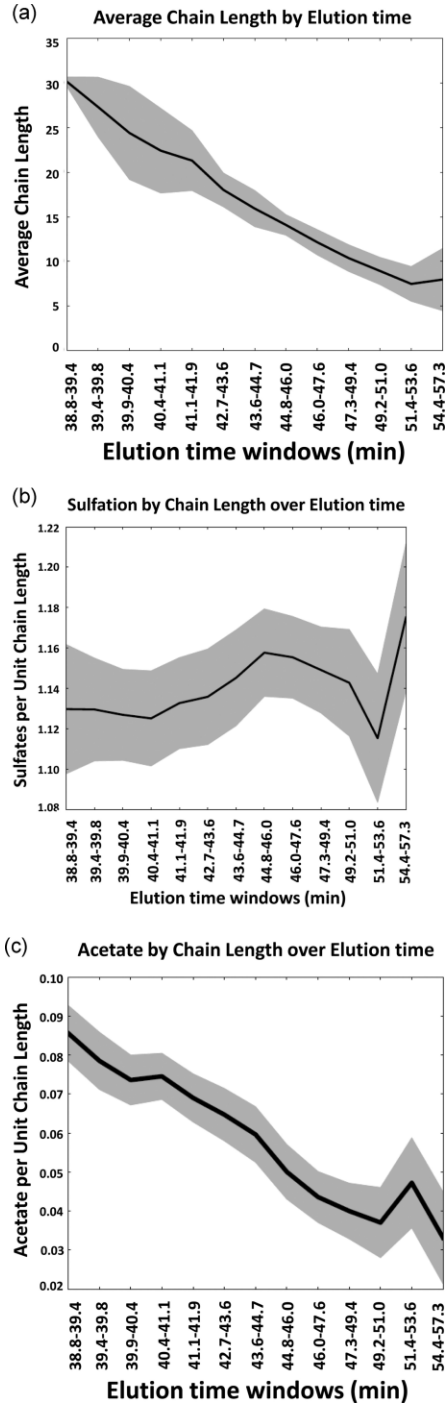


Figure 66: (a) Plot of dalteparin chain length as a function of SEC retention time. The shaded regions surrounding the trend lines indicate standard deviation of the mean at each point. (b) Plot of number of sulfate groups per unit chain length. The shaded region for the relationships indicate the confidence interval of the coefficient for the sample. (c) Plot of number of acetate groups per unit chain length. The shaded regions surrounding the trend lines indicate standard deviation of the mean at each point.

5.3.3.2 SEC-IS-MS of enoxaparin

Enoxaparin is produced by alkaline hydrolysis of heparin benzyl esters [328]. The package insert for the commercially available Lovenox® formulation states that 20% of the molecular weight distribution is below 2000 Da, 68% between 2000 and 8000 Da and 15% above 8000 Da[640]. The European Pharmakopeia range is listed as 3800-5000 Da[188]}. In comparison with that of dalteparin (Figure 63a) the SEC-IS-MS total ion chromatogram (TIC) of enoxaparin (Figure 67a) shows more abundant dp4 saccharides. The 43.7-44.7 min retention time window (Figure 67b) shows a range of enoxaparin dp16 saccharides in three compositional forms. The fully saturated saccharides correspond to enoxaparin containing the original saturated non-reducing HexA residue of the parent heparin chains. The saccharides showing one water loss correspond to chains cleaved by alkaline hydrolysis in which a $\Delta^{4,5}$-unsaturated uronic acid (ΔHexA) is present on the non-reducing chain ends. Saccharides of the same mass would be produced by cyclic ester formation at the reducing end. The saccharides showing two water losses correspond to saccharides with Δ-unsaturated non-reducing ends and a cyclic ester formed from the reducing end glucosamine (also known as a cyclic acetal). These compositions are given as [HexA, GlcN, SO_3, Ac, H_2O-loss] in the figure. Fully saturated enoxaparin dp16 saccharides (0 water losses) were low in abundances (Figure 67b). The assignments for a summed mass spectrum for the 39.9-40.4 min retention time range, corresponding to dp30 are shown in Figure 67c. As shown in Figure 68a for dp16, abundances for enoxaparin saccharides fully saturated, with one water loss and with two water losses, respectively, fall in the order 1>2>0. The same order is observed for dp30

saccharides (Figure 68b). As shown by the shading (darker shade indicating lower confidence), the confidence of assignments for dp30 saccharides of low relative abundances is low due to the difficulty of assigning isotopic patterns for isotope clusters with low signal-to-noise ratios.

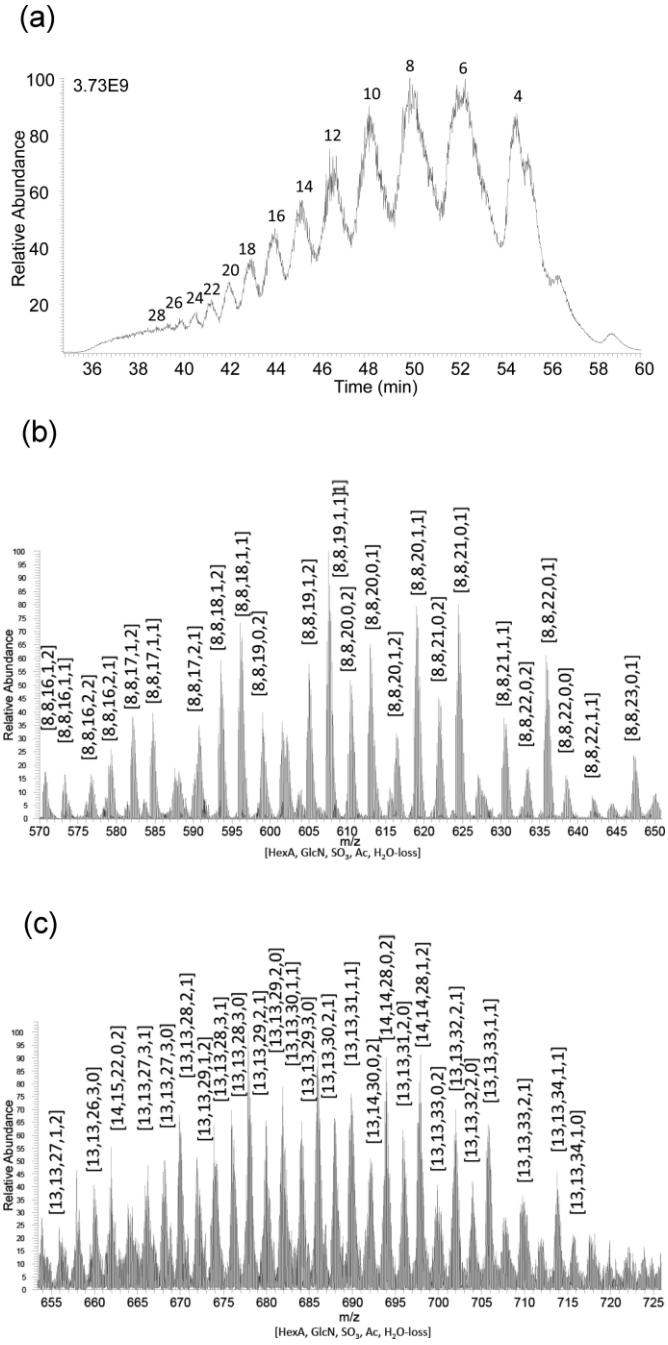


Figure 67: SEC-IS-MS of enoxaparin

(a) total ion chromatogram, (b) summed mass spectrum for the 43.7–44.7 min retention time window, corresponding to dp16, (c) summed mass spectrum for the 39.9–40.4 min retention time window, corresponding to dp26–28. Compositions are given as [ΔHexA, HexA, GlcN, SO3, Ac, H2O-loss].

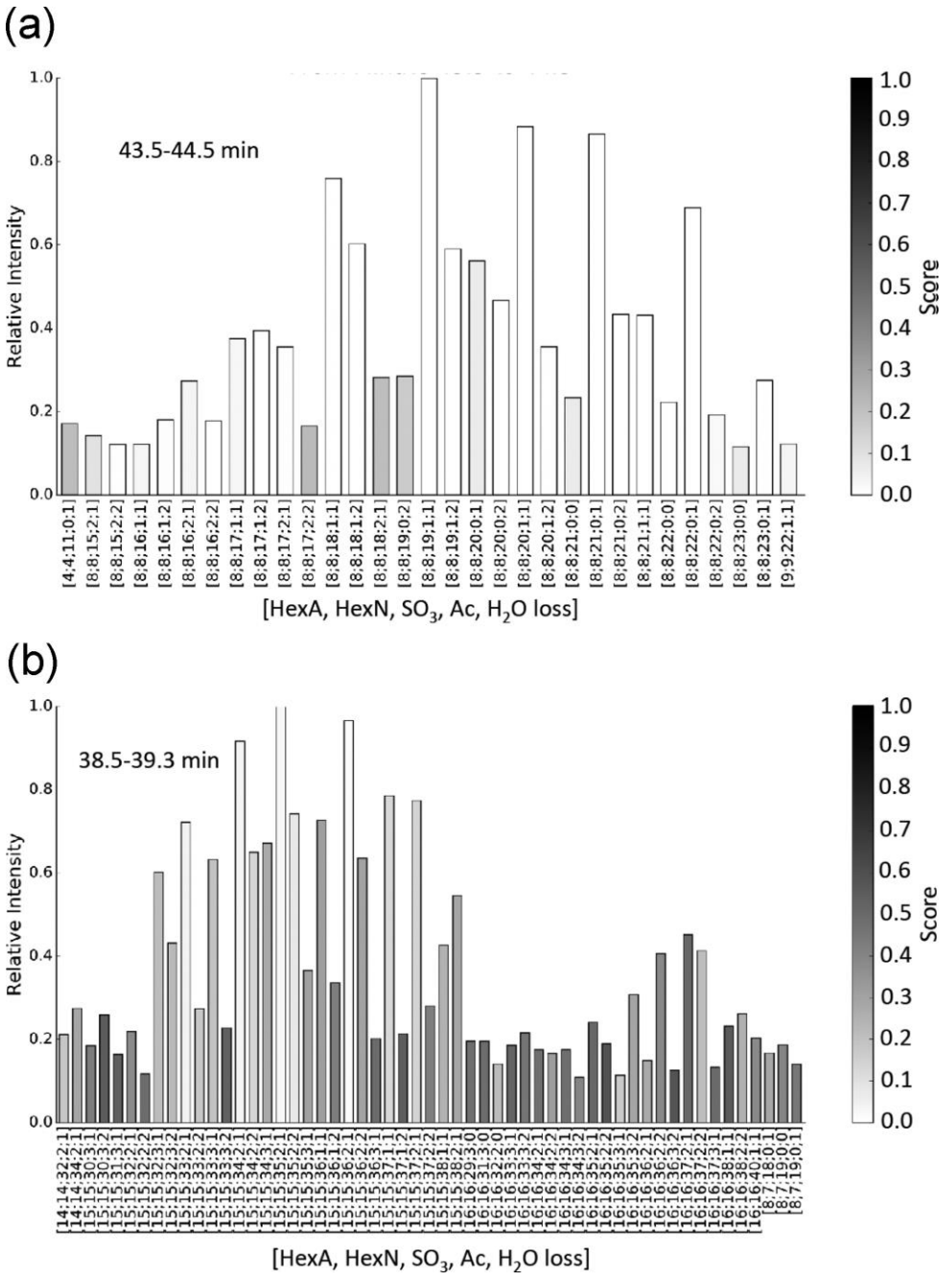


Figure 68: Compositions, scores and abundances of enoxaparin saccharides

(a) Retention time 43.5–44.5 min, corresponding to dp16. (b) Retention time 38.5–39.3 min, corresponding to dp30. The lowest score values reflect the greatest confidence.

As with dalteparin, we plotted the average chain length *versus* SEC elution time (Figure 69a). This plot showed that as expected the average chain length for enoxaparin diminished with time indicating that GAGdecon identified enoxaparin compositions correctly. We plotted the sulfation divided by saccharide chain length to determine the density of sulfation of enoxaparin per unit chain length (Figure 69b). We found that sulfation of enoxaparin saccharides diminished with decreasing chain length. This trend is consistent with less sulfated regions of the precursor chains being of increased likelihood to undergo alkaline hydrolysis. We next plotted the number of acetate groups per enoxaparin saccharide chain length (Figure 69c). This showed that the number of sulfate groups per unit chain length varied to a small degree for the range of saccharides detected in the SEC-IS-MS runs. This is consistent with acetylated glucosamine playing little influence on the alkaline hydrolysis reaction.

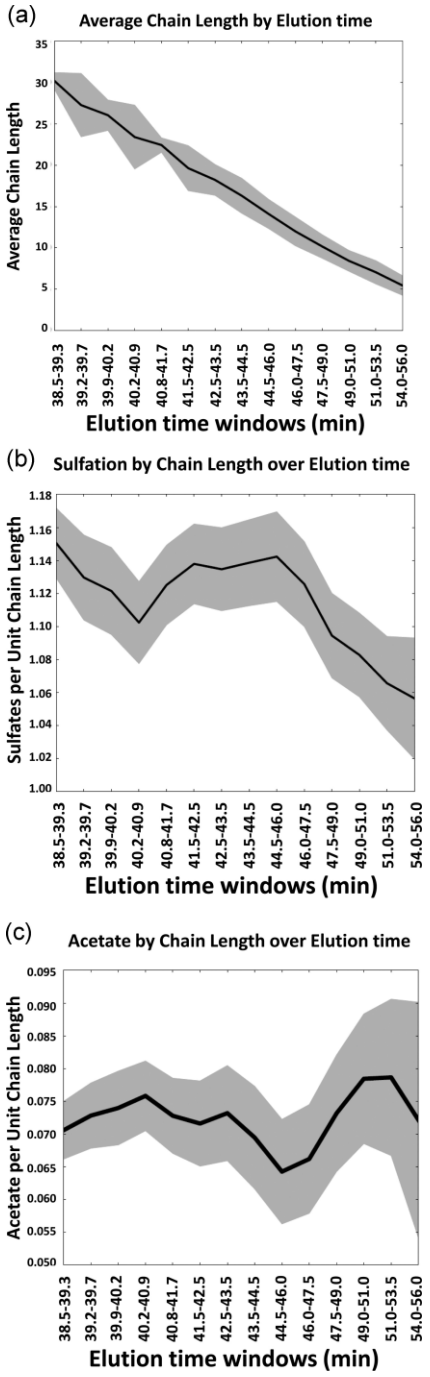


Figure 69: (a) Plot of enoxaparin chain length as a function of SEC retention time. The shaded regions surrounding the trend lines indicate standard deviation of the mean at each point. (b) Plot of number of sulfate groups per unit chain length. The shaded region for the relationships indicate the confidence interval of the coefficient for the sample. (c) Plot of number of acetate groups per unit chain length. The shaded regions surrounding the trend lines indicate standard deviation of the mean at each point.

5.3.4 Conclusions

The United States Food and Drug Administration has defined five criteria for demonstrating sameness for LMWH abbreviated new drug applications [642]. Among these, improvements in physical and chemical methods that occur over time were recognized to increase the ability to characterize sameness of LMWH preparations. While the establishment of sameness impacts both the development of abbreviated new drugs for LMWH, and LMWH batch release criteria, the identification of contaminants remains a critical need. This was shown by the 2007-2008 heparin contamination crisis in which criminal adulteration of the heparin severely impacted the global heparin supply[189,190,332,461].

We demonstrate the ability to produce exact mass values on the entire polymeric distribution of LMWH preparations using SEC-IS-MS. This capability adds a dimension to traditional SEC analysis of LMWH distributions by enabling calculation of elemental compositions for each detected LMWH species. We assert that this new capability will impact the ability to characterize the compositions of LMWH preparations in the global market favorably.

The components used in this first time SEC-IS-MS method for LMWH profiling are commercially available. It was necessary, however, to develop specialized bioinformatics to assign the compositions of the LMWH preparation. This involved assigning elemental compositions from theoretically calculated values. Using this approach, we defined the distribution of saccharides present in the LMWH preparations.

We suggest that this approach will be useful for analyzing LMWH as part of drug development and quality control efforts.

5.3.5 Acknowledgements

This work was supported by National Institutes for Health grants P41GM104603 and R21HL131554. Thermo-Fisher Scientific provided access to the Q-Exactive plus mass spectrometer and the ACRS 500 ion suppressor used in this work.

5.4 Microfluidic Capillary Electrophoresis-Mass Spectrometry of Glycans and Glycopeptides

This work has been submitted as a manuscript to the journal Analytical Chemistry and is currently under peer review.

5.4.1 Introduction

With improvements in mass spectrometer speed, sensitivity, analysis time, and depth, effectiveness of omics characterizations depends largely on front-end separation methods [481]. Liquid chromatography (LC) remains the preferred separation mode for online interface with mass spectrometry (MS), due to mature interfaces, solvent compatibility and a variety of available chromatography modes. Glycosylation increases the number of molecular forms of proteins (proteoforms) and thus the complexity of a biological sample. CE offers high peak capacity for glycoform resolution because the separation is based on charge, size and shape [34,83,304,373]. In contrast to LC and nano-LC, CE systems have lacked convenient interfaces for mass spectrometers, and this situation has limited their utilization for applications in large-scale glycomics and glycoproteomics studies. Due to recent developments in technology, microfluidic CE systems with integrated nano-electrospray ionization interfaces are commercially available for easy coupling to mass spectrometers [77,375,437]. The compact scale of these CE devices improves sensitivity and facilitates analysis of small sample quantities.

As stated above, a key factor affecting the implementation of CE-MS to biomolecule analysis has been the development of interfaces for mass spectrometry

coupling. One challenge has been the low liquid flow-rates in CE that are insufficient for ESI. Introduction of a coaxial sheath liquid by Smith and coworkers provided for mixing of a sheath liquid with CE effluent to increase overall flow rates, thereby enabling online coupling of CE to a mass spectrometer [491]. The sheath flow interface is easy to implement but suffers from low sensitivity due to sample dilution and chemical noise from the sheath liquid. Sheathless-flow interfaces were also developed around the same time [403] and modified versions of this interface were later introduced [147,363]; however they suffered from problems in spray-stability and reproducibility. A liquid-junction interface, implemented by many groups, enables partial disconnection of the CE and ESI components. This makes it possible to optimize CE and ESI individually because the flows from the two components are combined at a junction to produce a stable electrospray [155,476,477,556].

The use of bare fused-silica capillaries compromises separation efficiencies due to interactions with the analyte [430]. It also deteriorates migration time reproducibility, which is critical to proteomics analyses. To prevent such interactions, the use of surface coatings is common [260]. A number of groups have evaluated a polyamine-based coating, PolyE-323, that adheres to glass through ionic interactions with silanol groups and produces a net positive charge on the coated surfaces under neutral to acidic conditions [375,536]. The authors reported efficient separation for tryptic peptides using capillaries with this coating. The coating also provided a stable anodal electroosmotic flow (EOF), thus enabling stable ESI on the microfluidic device. Therefore, a variety of neutral or positively-charged coatings have been explored to counteract unfavorable

surface interactions for proteomics applications [74,138,470]. These coatings include a monoquaternarized piperazine compound, polybrene-poly(vinyl sulfonate) and an *N*-methylpolyvinylpyridinium cationic polymer.

CE was widely adopted for glycan analysis after its coupling to a laser-induced fluorescence detector (LIF). This method requires derivatization of the glycans with a fluorophore, with aminiopyrene trisulfolane (APTS) being the most commonly used tag [199,200,510]. Optical detection allows identification only of molecules for which standard migration times have been determined. Gennaro and colleagues evaluated aminonaphthalene trisulfolane (ANTS) as a derivatization reagent for CE-MS of glycans, along with a low molarity (20 mM) 6-aminocaproic acid CE separation buffer [166]. APTS and ANTS both impart negative charge to the labeled glycans and make negative ion mode analyses more efficient by improving ionization efficiencies [166,200,460]. Gennaro and Salas–Solano demonstrated the use of online CE-LIF-MS for APTS-labeled glycans on a CE system with PVA-coated capillaries and a running buffer of 40 mM aminocaproic acid and 0.02% hydroxypropylmethylcellulose [167].

Monosaccharide analysis is commonly performed using gas chromatography-mass spectrometry (GC-MS). This method requires derivatization (generally peralkylation) of monosaccharides to generate hydrophobic derivatives that can be resolved by gas chromatography [384]. Partially methylated alditol acetates can be used for linkage analysis by GC-MS [227]. Isobaric monosaccharides are differentiated based on the retention time differences measured for purified standards. A method for monosaccharide analysis using CE was developed by Guttman, who has described the

separation of APTS-labeled monosaccharides and AMAC-labeled sialic acids coupled with LIF detection [201]. This setup used 25 mM lithium tetraborate buffer in the separation channels. This method has been adopted by many laboratories but employs buffers incompatible with mass spectrometry.

Guttman also described sample preparation and separation methods using CE with optical detection for analysis of oligosaccharides [199,200,402,510,514]. These methods are incompatible with mass spectral analysis due to their running buffers. More recently, a study described the characterization of *N*- and *O*-linked glycopeptides from human erythropoietin using CE-MS [173]. This study utilized bare fused-silica capillaries with an electrolyte solution containing a mixture of acetic and formic acid, with 50% isopropanol used for sheath flow. Váradi et al. used isotopically coded 2-aminobenzamide for glycan labeling and analysis by CE-MS. This method enabled duplex analysis of glycan samples [539]. The aminoxy TMT (tandem mass tag) reagents have been used for efficient labeling and multiplexing of glycans in CE-MS/MS experiments [626,629]. This reagent facilitates fast and efficient labeling of glycans with minimal post-labeling cleanup required and enables straightforward tandem-MS based quantification and sample multiplexing.

The CE-MS setup used in this work integrates a microfluidic separation system with a nanoESI source on a single piece of hardware [375] that works as a removable interface for mass spectrometers. As shown in the schematic in Figure 70, sample injection, background electrolyte and ESI solution are controlled independently in a manner similar to the liquid-junction interface described above. The microfluidic setup

minimizes gaps, dead volume, sample dilution, and band broadening, to maintain speed and sensitivity in analyses. The separation capillary surfaces are coated with an aminopropyl silane reagent to minimize EOF, as described previously [36].

We developed a CE-MS method for analysis of monosaccharides, oligosaccharides and glycopeptides to enable comprehensive glycoprotein analysis. The method uses the same electrophoresis solutions for all three compound classes, greatly simplifying the task of glycoprotein characterization. We used the aminoxy TMT regent to improve electrophoretic migration of neutral saccharides and to enable multiplexing in tandem-MS experiments. We also used a sialic acid derivatization method to prevent interaction of the negatively charged carboxyl groups with the positively charged capillary surface coating and discriminate sialic acid linkages. In summary, we present a CE-MS method using a convenient mass spectrometry interface for glycan, monosaccharide and glycopeptide analysis.

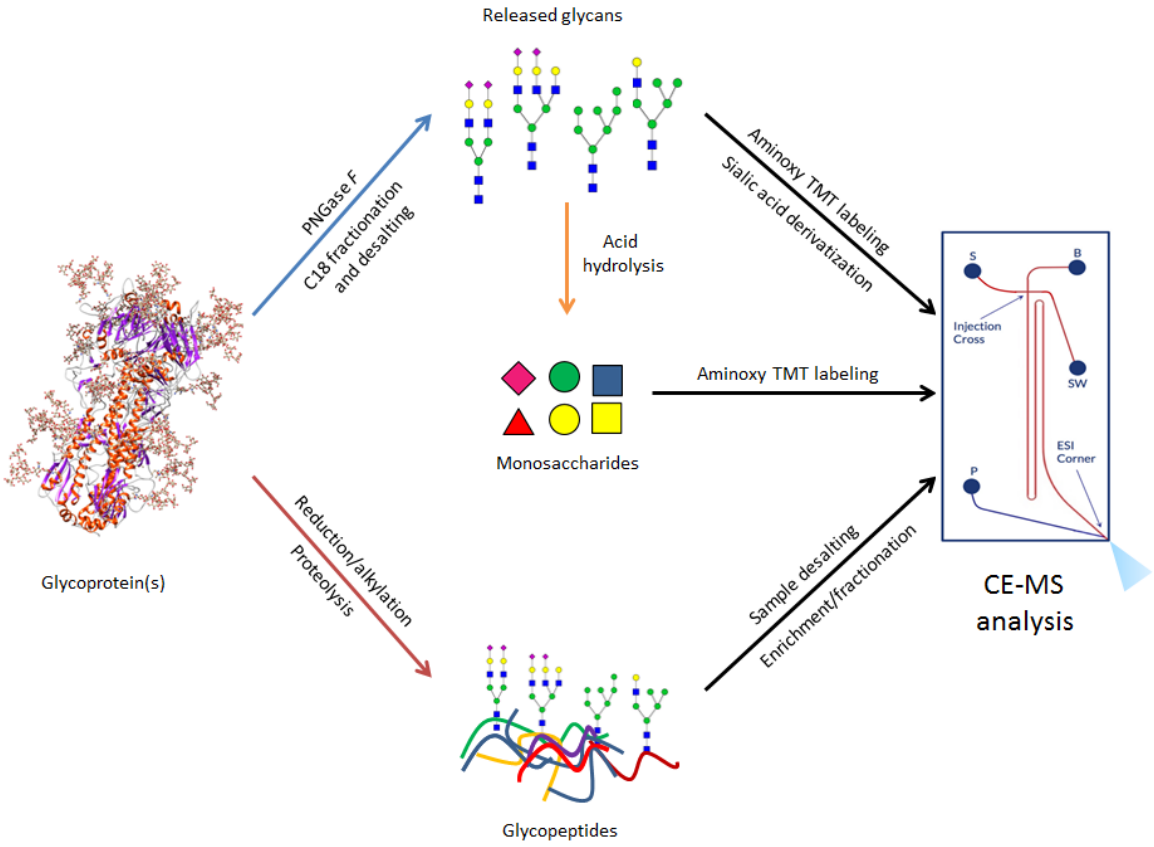


Figure 70: Overview of workflow for analysis of monosaccharides, oligosaccharides and glycopeptides using microfluidic CE-ESI-MS

Chip schematic on the right shows sample reservoir "S", background reservoir "B", ESI pump "P" and sample waste "SW".

5.4.2 Materials and methods

All experiments were performed on a ZipChip capillary electrophoresis-electrospray ionization interface (CE-ESI) (908 Devices, Boston, MA), using a high-resolution chip with capillary length 22 cm, coupled to a Q-Exactive Plus mass spectrometer (Thermo Scientific, San Jose, CA). A CE voltage of 20 kV was applied with a shield potential set at 500 V. ESI voltage varied between 2200-2700 V for the different samples and chips. A background electrolyte (BGE) and ESI solution containing

50% methanol, 48% water and 2% formic acid was used for all experiments. Pressure-assisted CE, a feature of the CE-MS interface, was used for oligosaccharide sample analysis to speed analysis times. Pressure-driven injection allows transient isotachophoresis for the analyte but dilutes the sample solution in cases where low sample volumes are loaded in the sample well. All MS data were acquired in the positive-ion mode. Released glycans and monosaccharides were derivatized with aminoxy-TMT prior to analysis. Sialic acid derivatization was performed using the method described by Reiding *et al.*[438].

Glycopeptide and oligosaccharide sample preparation and data acquisition were carried out by Kshitij Khatri. Monosaccharide samples were prepared by John R. Haserick and the data were acquired by John R. Haserick and Kshitij Khatri. Oligosaccharide and glycopeptide data analysis was carried out by Kshitij Khatri and Joshua A. Klein and monosaccharide data were analyzed by John R. Haserick. Prof. Mark E. McComb supervised all CE-MS method development and data acquisition. Data interpretation and figure generation was carried out with guidance from Prof. Joseph Zaia, Prof. Catherine E. Costello and Prof. Mark E. McComb.

5.4.2.1 Glycopeptide sample preparation and data acquisition

The glycoprotein standards, bovine ribonuclease B (RNase B), human transferrin (transferrin) and human alpha-1-acid glycoprotein (AGP), were purchased from Sigma-Aldrich (St. Louis, MO). Approximately 250µg of each glycoprotein was subjected to reduction and alkylation, using dithiothreitol and iodoacetamide, followed by proteolysis

using trypsin (Promega Corp., Madison, WI), as described previously [293]. Tryptic digests were desalted using C18 spin columns (Thermo Pierce) and then dried down using a centrifugal evaporator. Dry peptide/glycopeptide samples were resuspended in 50µl of the background electrolyte (BGE) containing 100 mM ammonium acetate, to improve conductivity and sample stacking via isotachophoresis. MS1 spectra were acquired over the range *m/z* 665 - 2000, with 1 microscan per spectrum at a resolution of 70,000. A high starting *m/z* was used to avoid interference from abundant siloxane peaks that were coming from the CE-ESI source. Siloxane peaks corresponding to *m/z* 667.1764 and 741.1952 were consistently present in the spectra and were used as lock masses for internal mass calibration. The automatic-gain-control (AGC) target set at 1e6 and a maximum ion injection time of 110 msec. Data-dependent tandem MS were acquired on the top 10 precursors with charge states ranging between 3 and 6 using (higher energy collisional induced dissociation) collisional-dissociation at a normalized collision energy (NCE) of 27 eV. Tandem mass spectra were acquired with 2 microscans per spectrum at a resolution of 35,000 and an AGC target set at 5e5. Peaks for migrating glycopeptides were between 15-30 seconds wide. A minimum intensity threshold of 1.4 e5 was used for precursor selection and following a tandem MS precursors were dynamically excluded for 1 sec from being selected again. Source capillary temperature and RF level were set at 320^0C and 55, respectively.

5.4.2.2 N-glycan sample preparation and data acquisition

N-linked glycans were released from 100µg of standard glycoprotein tryptic digests using PNGase-F (New England Biolabs, Ipswich, MA). The released glycans were separated from the peptides using a C18 spin column, as the unbound fraction. Glycan samples were dried and then subjected to reducing-end labeling using Aminoxy-TMT^{131} (Thermo Scientific) reagent as per the manufacturer's protocol. Labeled glycans were desalted using PD Minitrap G10 columns (GE Healthcare, Marlborough, MA). Desalted glycan samples were dried *in vacuo* and resuspended in 20µL of background electrolyte containing 100mM ammonium acetate.

Glycan MS1 spectra were acquired over the range *m/z* 600 – 2000, with 1 microscan per spectrum at a resolution of 70,000, with the AGC target set at 1e6 and a maximum ion injection time of 110 msec. Siloxane peaks with *m/z* 610.18416, 667.1764 and 741.1952 were used as lock-masses. Data-dependent tandem MS were acquired on the top 5 precursors with charge states of 2 or 3, using collisional-dissociation at a stepped NCE of 25 and 30 eV. While glycan compositions were assigned based on MS1 only, a low minimum intensity threshold of 5.5 e3 was used for precursor selection in data dependent tandem MS to confirm that the migrating peaks were glycans and following a tandem MS precursors were dynamically excluded for 1 sec from being selected again. Tandem mass spectra were acquired with 2 microscans per spectrum at a resolution of 35,000 and an AGC target set at 2e4 to enable tandem MS of low abundance glycoforms. Source capillary temperature and RF level were set at 320^{0}C and 55, respectively.

5.4.2.3 Monosaccharide sample preparation

5.4.2.3.1 Preparation of TMT-labeled Monosaccharide Standards

Monosaccharide Monosaccharide standard pentoses (D-xylose, D-ribose, D-arabinose), deoxyhexoses (L-fucose, L-rhamnose), hexoses (D-glucose, D-galactose, D-mannose), *N*-Acetyl-hexoses (*N*-acetyl-D-glucosamine, *N*-acetyl-D-galactosamine, *N*-acetyl-D-mannosamine), and a sialic acid, *N*-acetylneuraminic acid, were purchased from Sigma Aldrich. Aqueous solutions of each monosaccharide were made at a concentration of 20 mg/mL in ddH_2O. One microliter of each was added to three separate tubes, in groups, to be labeled with the aminoxyTMT^6-127, aminoxyTMT^6-128, and aminoxyTMT^6-130 tags. The monosaccharides were grouped in the following manner: aminoxyTMT^6-127 (Xyl, Fuc, Gal, GlcNAc), aminoxyTMT^6-128 (Rib, Rha, Man, GalNAc), and aminoxyTMT^6-130 (Ara, Glc, ManNAc, Neu5Ac). Prior to labeling, the three mixes were dried under vacuum and then reactions were performed according to the manufacturer's protocol, excluding quenching with acetone. Briefly, the entire contents of a packet of aminoxy-TMT reagent (0.8 mg) was re-suspended in 200 µL of 95% methanol/ 0.1% acetic acid, and added to one of the tubes containing dry monosaccharide samples. The tubes were then vortexed vigorously, and allowed to sit for 10 minutes at ambient temperature. The solvents were removed with a centrifugal evaporator. Then, 200 µL of 95% methanol was added; the samples were vortexed and allowed to sit for 10 min. at room temperature, and dried using the centrifugal evaporator. After drying, the three mixes were stored at -20°C until analyzed.

Just prior to analysis, the three standard monosaccharide mixes were reconstituted with 50:50 methanol:water, and each solution was mixed with an equal volume of 50:50 methanol:water containing 200 mM ammonium acetate to yield a final concentration of 100 mM ammonium acetate for CE-MS analysis.

5.4.2.3.2 AGP monosaccharide preparation

A 1 mg quantity of AGP was treated with PNGase-F overnight at 37 ^{0}C. The released *N*-glycans were separated from the protein using C-18 Sep-Pak cartridges (Waters Corporation, Milford, MA), as described previously [2,3]. The released, purified *N*-glycans were placed into a conical vial and subjected to acid hydrolysis using 2 M trifluoroacetic acid at 110 ^{0}C, for 30 min. The resulting monosaccharide mixtures were dried down in a centrifugal evaporator and labeled with aminoxyTMT^6-127 as described above. The labeled sample was dried and re-suspended for analysis in 50 µL of 50:50 methanol:water. A 1:200 dilution was made in 50:50 MeOH/H_2O containing 100 mM ammonium acetate and this solution was used for CE-MS analysis.

5.4.2.3.3 CE-MS of Monosaccharides

MS1 alternating with all-ions fragmentation (AIF) was performed (MS/AIF). For MS, the chosen parameters were resolution 17,500, AGC target 2e5, maximum injection Time 40 ms, scan range *m/z* 440 - 640 For AIF, the chosen parameters were NCE 35 eV, resolution 17,500, AGC target 2e5, maximum injection time 40 ms, scan range *m/z* 440-640. Lock masses used were, *m/z* 445.12003, 519.13888 and 610.18416. Source capillary temperature and RF level were set at 320^{0}C and 55, respectively.

5.4.2.4 Data analysis

Data were analyzed manually using Thermo XCalibur data analysis software and using in-house data analysis tools for generation of extracted ion electropherograms (EIEs), as described below.

5.4.2.4.1 Database Construction

The database of theoretical glycan compositions was generated using a combinatorial algorithm considering every combination of Hex between 3 and 9, HexNAc between 2 and 8, Fuc between 0 and 4, and NeuAc between 0 and 5. The number of NeuAc residues was constrained to be ≤ (# HexNAc – 1) and the number of fucose were constrained to be ≤ (# HexNAc) residues. Each composition was considered with and without a tandem mass tag (Mass Shift Formula: C10 C[13]5 H28 N4 O2). For each composition containing NeuAc, each NeuAc was considered in three forms, underivatized, dimethylaminated, or lactonized.

The glycopeptide database was constructed by combining this same list of glycan compositions without derivatization or tandem mass tags with a list of common tryptic peptide sequences around sequons of interest.

5.4.2.4.2 Migration Profile Reconstruction

Each MS1 scan for each sample was processed by a deisotoping and charge state deconvolution algorithm using an *N*-glycan-specific averagine[473] formula C7 H11.83 N0.5 O5.17, using the BRAIN [128] algorithm to generate isotopic patterns for the interpolated experimental composition. Extrapolated neutral mass monoisotopic peaks

were aggregated into migration profiles by mass using a 10 parts-per-million error tolerance to collect peaks across scans. Profiles which contained gaps greater 250 milliseconds wide were split into separate migration profiles, corresponding to separate events. Migration profiles were expected to conform to a bi-Gaussian shape after Gaussian smoothing ($\sigma = 1$). Un-gapped multimodal profiles were partitioned at valleys and evaluated independently. The bi-Gaussian fit for each smoothed profile was computed using non-linear least squares and the residual-sum-of-squares (RSS) for this fit was recorded, called the "shape fit RSS". Each migration profile was also compared against a null model which assumed that the distribution of signal over time was random, using a horizontal line at the average intensity across the migration profile to calculate the "null model RSS". Profiles for which $1 - \frac{\text{shape fit RSS}}{\text{null model RSS}} > 0.4$ were considered well resolved.

For glycomics data, reconstructed electropherograms were searched against a database of theoretical glycan compositions by neutral mass, using a 10 parts-per-million error tolerance to assign putative compositions. Under the conditions used, ammonium adducts were expected, and a second round of searches was performed using a mass shift corresponding to one or two ammonium adducts. The results which matched overlapping migration profiles of different adduct types were summed. These summed profiles were also evaluated following the above shape rules. Glycopeptide electropherograms were reconstructed in the same fashion; however, adducts were not considered.

5.4.2.4.3 Migration Profile Visualization

All migration profile figures were produced by applying the same Gaussian smoothing described above to the matched raw profile data.

5.4.3 Results and Discussion

5.4.3.1 N-glycan analysis

Zhong et al. have shown the utility of aminoxy-TMT tags in quantitative analysis of released glycans by CE-MS [626]. The TMT tags also improve ionization efficiency and thus the sensitivity of sialylated glycans in the positive-ion mode. The aminoxy TMT group imparts a positive charge to labeled neutral glycans in acidic buffers, thereby enabling electrophoretic migration. Further, TMT-based multiplexing has been shown to provide robust quantitative results for paired glycomics samples in both LC-MS and CE-MS [626,629]. We therefore used aminoxy-TMT labeling for analyses of released glycans by CE-MS.

The migration and separation of neutral glycans appeared to be dependent on the size and number of monosaccharide units, as shown in EIEs in Figure 71 for pauci- and high-mannose *N*-glycans (left panel) observed from a sample consisting of a mixture of released glycans from ribonuclease B, human transferrin and AGP. The EIEs for asialo complex-type glycans in the mixture are shown in the right panel. Glycan compositions are reported as [HexNAc; Hex; dHex; Total NeuAc; Esterified NeuAc; lactonized NeuAc] in the text and figures. Migration for high-mannose-type *N*-glycans shows a clear correlation with increasing size; however, multiple peaks are observed for paucimannose

glycans ($HexNAc_2Hex_{3-4}$) [2;3;0;0;0;0] – [2;4;0;0;0;0]. The earliest migrating peaks indicate the existence of these two glycans in solution prior to the CE-MS analysis. In addition, the data show that [2;3;0;0;0;0] and [2;4;0;0;0;0] EIEs contain signal from MS-induced breakdown products of larger high-mannose type glycans. MS-induced breakdown products in the EIEs for larger glycans also exist but are difficult to visualize due to the scale of the Y-axis and the presence of data-dependent tandem MS, which decreases the number of data points per electropherogram.

The extracted mass spectra showed protonated analyte ions with varying degrees of ammoniated, sodiated and potassiated ions forming due to presence of ammonium acetate in the sample loading solution and other salts that may be present in the sample matrix or introduced from sample preparation. The abundances of adducted ions ranged from 5-20% abundance relative to protonated ions and generally decreased with the size of the glycan. The EIEs were plotted for protonated ions only. Complex-type asialo glycans showed similar increasing migration times with added monosaccharide units and branching. The right panel in Figure 71 shows EIEs for bi-, tri- and tetra-antennary complex-type *N*-glycan compositions, with or without a fucose residue.

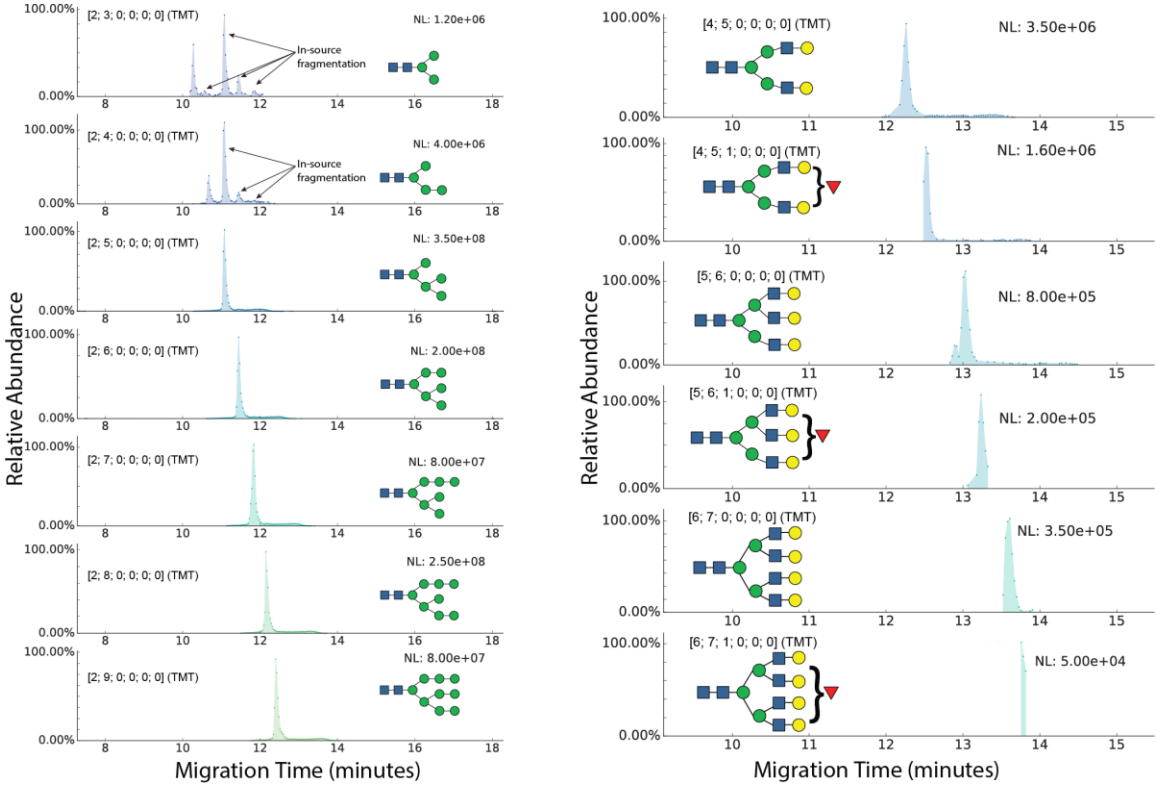


Figure 71: Neutral glycans – stacked EIEs showing high-mannose *N*-glycans from bovine RNAseB (left panel) mixed with complex glycans from human transferrin and AGP

Glycan compositions are represented as [HexNAc; Hex; dHex; Total NeuAc; Esterified NeuAc; Lactonized NeuAc].

We next analyzed the behavior of *N*-glycans containing neuraminic acids (NeuAc) in their compositions. Figure 72 shows stacked EIEs for sialylated glycan compositions, with a TMT label on the reducing end. While neutral glycans migrated through the separation capillary within the first 15 minutes of the CE-MS run, their sialylated counterparts migrated more slowly. The retardation of sialylated glycoforms was proportional to the number of sialic acids present in the glycan compositions. Additionally, we detected split peaks or multiple peaks for multiply sialylated glycoforms. For compositions including [5;6;0;3;0;0], migration time differences

between peaks were as high as 14 minutes. While CE has been shown to resolve isomeric glycans, such large differences in migration times for the same glycan composition were not expected [626]. We hypothesized that the increased migration times and elution in multiple peaks was related to a combination of increased glycan net negative charge, increased hydrodynamic radius and non-specific interactions of acidic glycans with the positively charged capillary coating.

Methyl esterification stabilizes sialic acids for MALDI- and LC-MS analyses [40,41]. Recently, Wuhrer and colleagues described a MALDI-MS method for analysis of sialylated glycoforms by sialic acid stabilization that allows facile discrimination of sialic acid linkages by ethyl esterification [438]. Using this approach, α2,6-linked NeuAc residues are esterified and α2,3-linked NeuAc form cyclic lactones. To eliminate any non-specific interactions of the sialylated glycoforms, we employed their ethyl esterification strategy to derivatize the NeuAc residues..

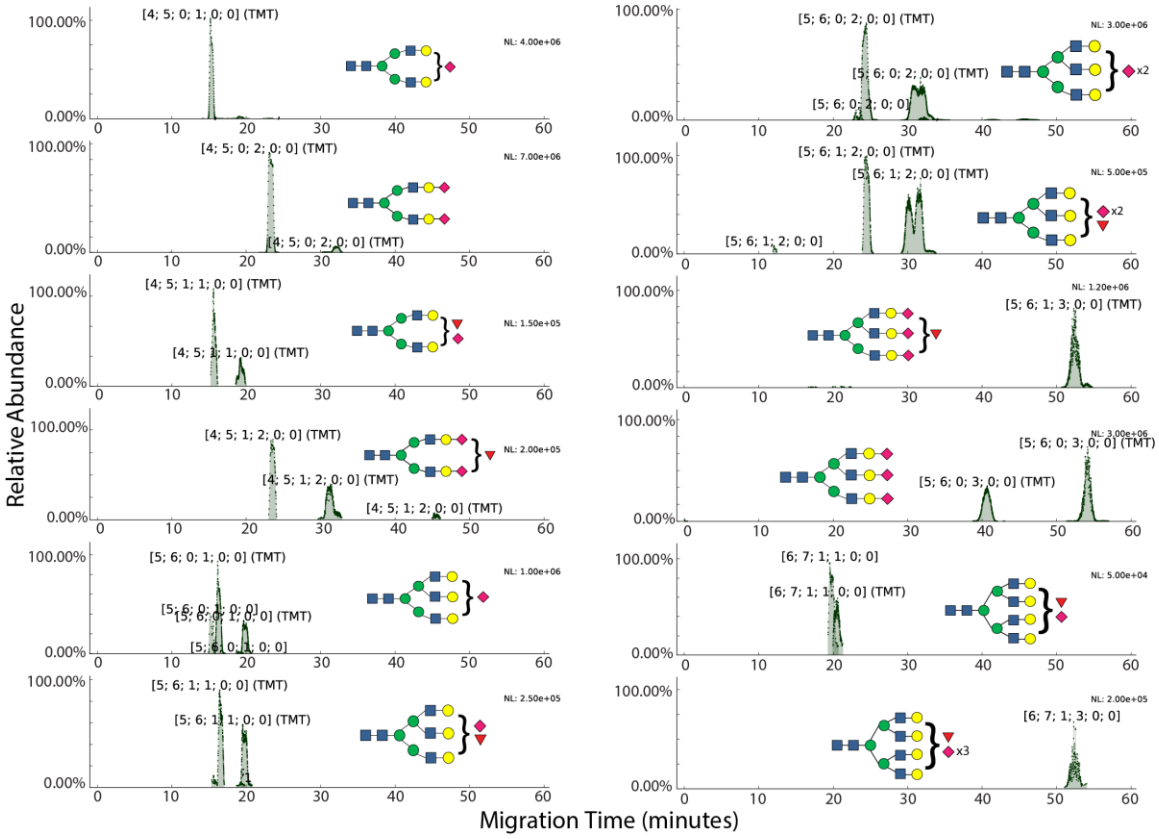


Figure 72: Sialylated glycans – EIEs showing sialylated glycan with underivatized sialic acids, from human transferrin and AGP

Glycan compositions are represented as [HexNAc; Hex; dHex; Total NeuAc; Esterified NeuAc; Lactonized NeuAc].

The ethyl-esterification step allowed us to track the migration characteristics of TMT-labeled glycans without the influence of the sialic acid carboxyl groups. We identified *N*-glycans with completely and partially derivatized sialic acids, as shown in Figure 73. As expected, fully derivatized *N*-glycans shifted to migration times of 20 minutes or less, while populations of partially derivatized glycans with one or more sialic acids left underivatized eluted in multiple peaks with longer migration times. Additionally, the mass shifts induced by ethyl esterification *versus* lactonization allowed characterization of sialic-acid linkages.

In most cases, the differently linked sialosides displayed the same electrophoretic mobility. In some cases, migration times of different compositions overlapped. For example, [4;5;1;2;0;2], overlapped with [4;5;1;2;1;1] and [4;5;1;2;2;0]. A relatively wide, bimodal peak was also observed for [5;6;1;1;1;0], indicating isomer separation. Peak shoulders and tailing were also observed for sialylated glycans with unoccupied non-reducing end Gal residues. It thus appears that NeuAc positional isomers exist on these glycans. The presence of abundant underivatized glycans necessitates the improvement of derivatization method to achieve higher efficiency. Several methods for sialic acid derivatization have been published, indicating that this is an ongoing and challenging problem [50,82,85,168,202,203,240,334,394,394,399,423,431,438,472].

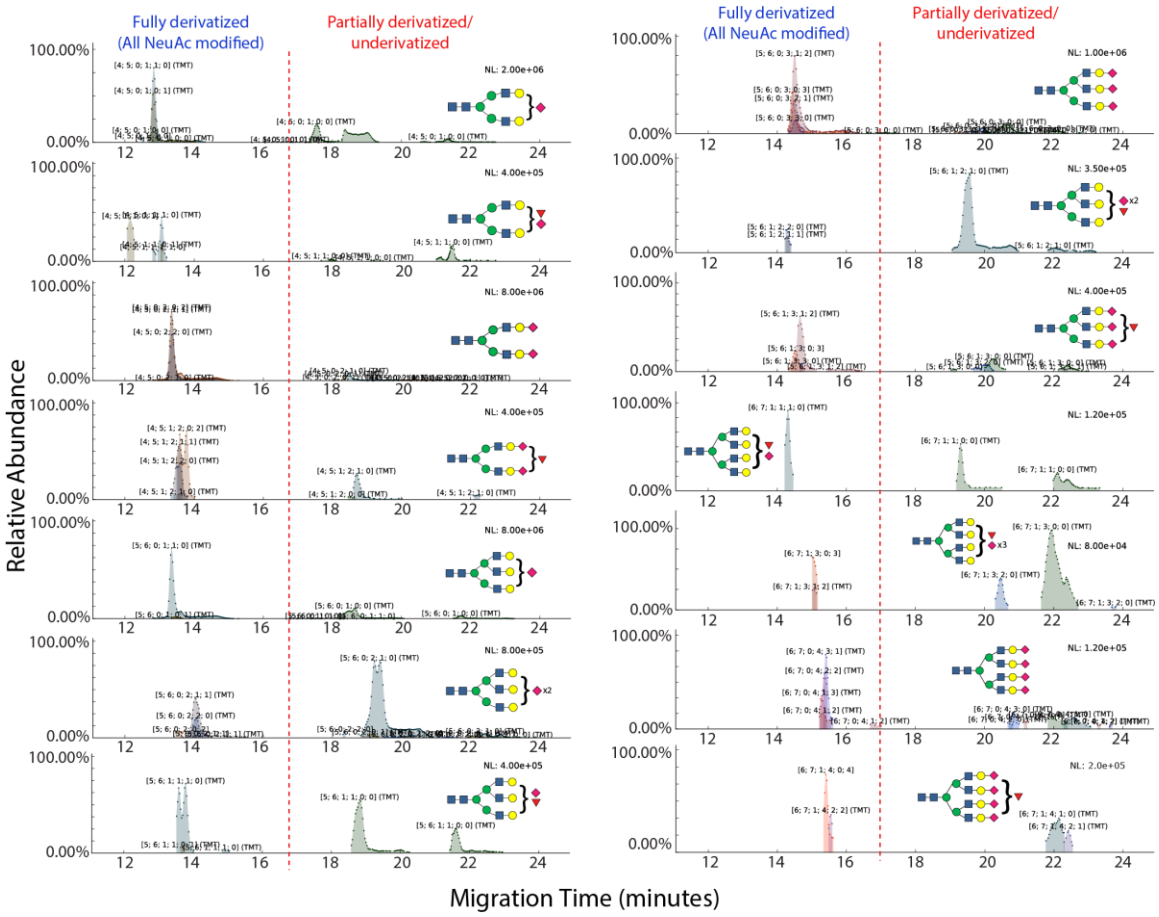


Figure 73: Derivatized sialylated glycans – EIEs showing sialylated glycan with derivatized sialic acids, from human transferrin and AGP

Glycan compositions are represented as [HexNAc; Hex; dHex; Total NeuAc; Esterified NeuAc; Lactonized NeuAc].

5.4.3.2 Glycopeptide analysis

Capillary electrophoresis separates analytes based on both charge and size. The size of glycans on peptide backbones has a significant influence on their electrophoretic mobilities. Thus, glycosylated peptides migrate more slowly through the separation capillary compared to their non-glycosylated counterparts [173]. Alpha-1-acid glycoprotein contains five glycosylation sites with a wide-ranging distribution of complex-type *N*-glycans [530,464,293,292]. In addition, the protein has two isoforms that

add variability in protein sequence in the region spanning the glycosylation sites. This allowed us to generate and study the migration characteristics of a variety of peptide backbone sequences with combinations of glycoforms using a tryptic digest of human AGP. Figure 74 shows a total-ion electropherogram (TIE) for an AGP tryptic digest stacked over the EIEs for oxonium ions indicating presence of HexNAc (middle) and NeuAc (bottom) on the precursor ions selected for tandem MS. As indicated by the overlay of oxonium ions, the non-glycosylated peptides in an AGP tryptic digest elute in the first 10 minutes, while the majority of glycosylated peptides elute after 8 minutes. All the peaks selected using data-dependent acquisition of tandem mass spectra after 8 min migration time were glycopeptides. The migration positions for two glycoforms of the peptide LVPVPITNATLDR are indicated in the TIE. Figure 75 and Figure 76 show HCD tandem mass spectra confirming the assignments. Note that addition of a single sialic acid resulted in a migration time difference of more than 5 minutes between the two glycoforms. Overall, most glycosylated peptides with neutral (asialo) glycans appeared prior to 14 minutes, while the sialylated glycoforms were slower migrating.

.

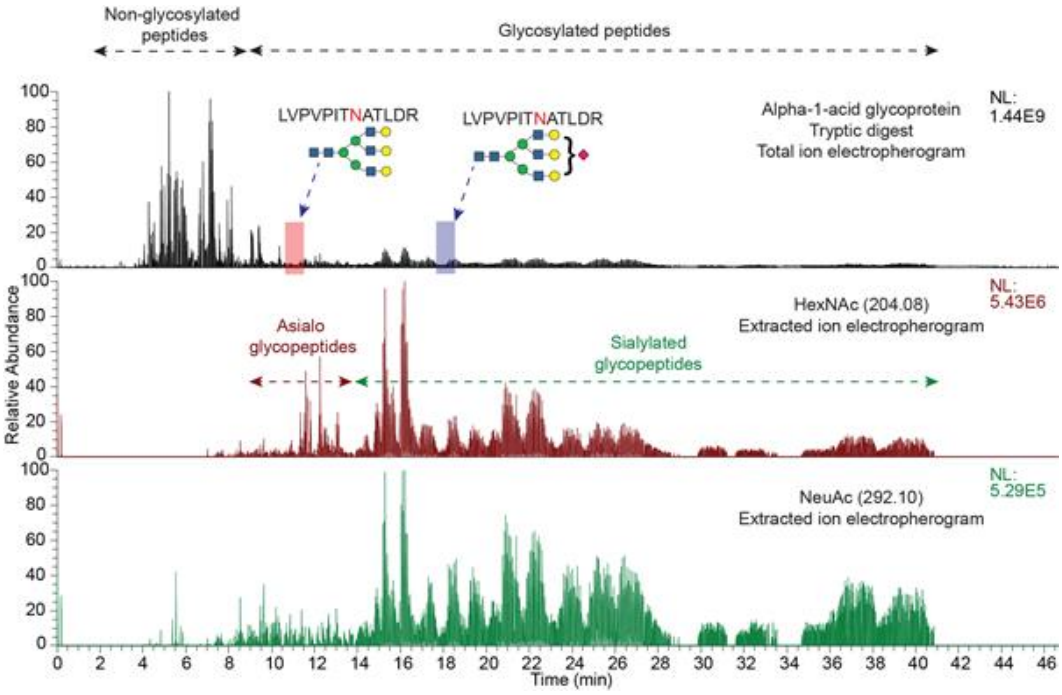


Figure 74: Total ion electropherogram of a tryptic digest of AGP, analyzed by CE-MS, overlaid with extracted ion electropherograms for HexNAc and NeuAc oxonium ions, showing migration characteristics of glycosylated and non-glycosylated peptides

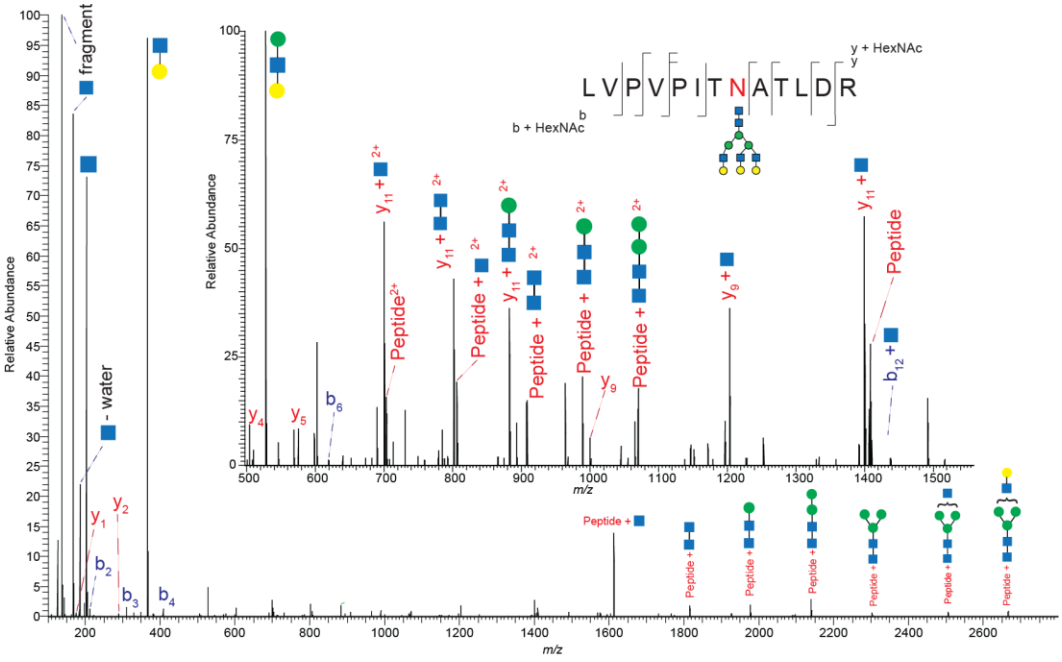


Figure 75: Tandem MS of an asialo AGP glycopeptide from data dependent CE-MS/MS

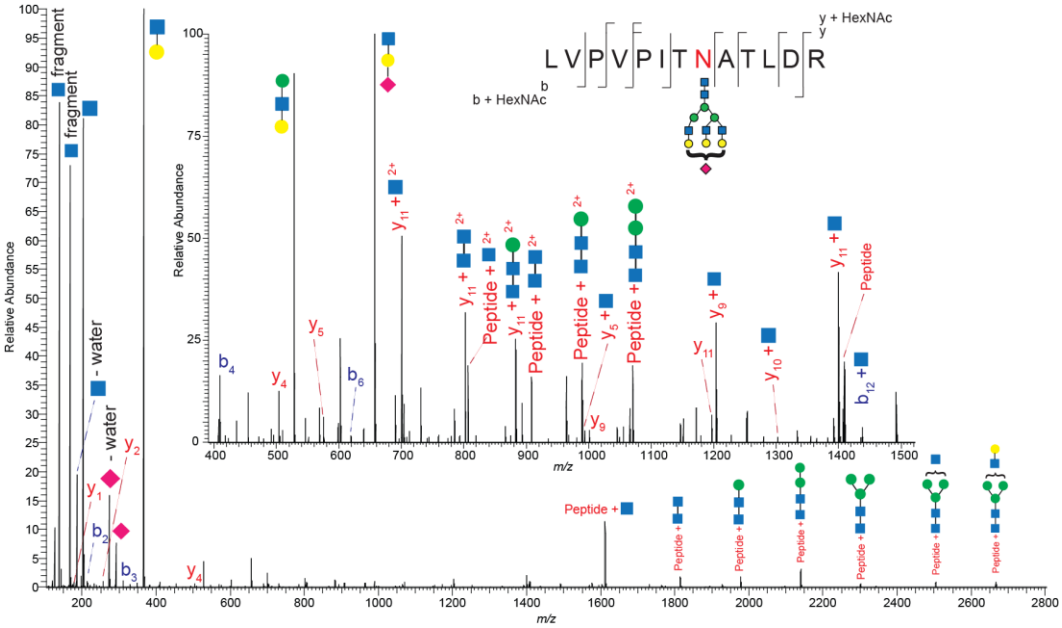


Figure 76: Tandem MS of an asialo AGP glycopeptide from data dependent CE-MS/MS

In a manner similar to the released glycans, glycopeptide migration through the separation capillary depended on size. Figure 77 compares the migration profiles of glycopeptides grouped according to peptide backbone by overlaying different peptide backbones with the same glycan composition on the same plot. Figure 78 compares the overlaid migration profiles of glycopeptides with the same peptide backbone but different glycoforms. Migrations times increased with both the peptide length and the glycan size. Further, the trend in migration of neutral *vs.* sialylated glycoforms, as seen in released glycans, persisted in glycopeptides. All asialo glycopeptides had migration times of 15 minutes or less. With increase in the number of sialic acids, broader and multiple peaks were observed. While these factors increase the density of the data, the overall separation of different glycoforms is useful for deeper characterization of site-specific glycan microheterogeneity. The value of this separation will become more apparent for complex

biological samples where multitudes of glycosylated peptides and their numerous glycoforms are present. Although CE separates non-glycosylated peptides from glycosylated peptides, employment of a prior enrichment step such as HILIC would greatly enhance the loading capacity and allow a better use of the instrument speed and dynamic range.

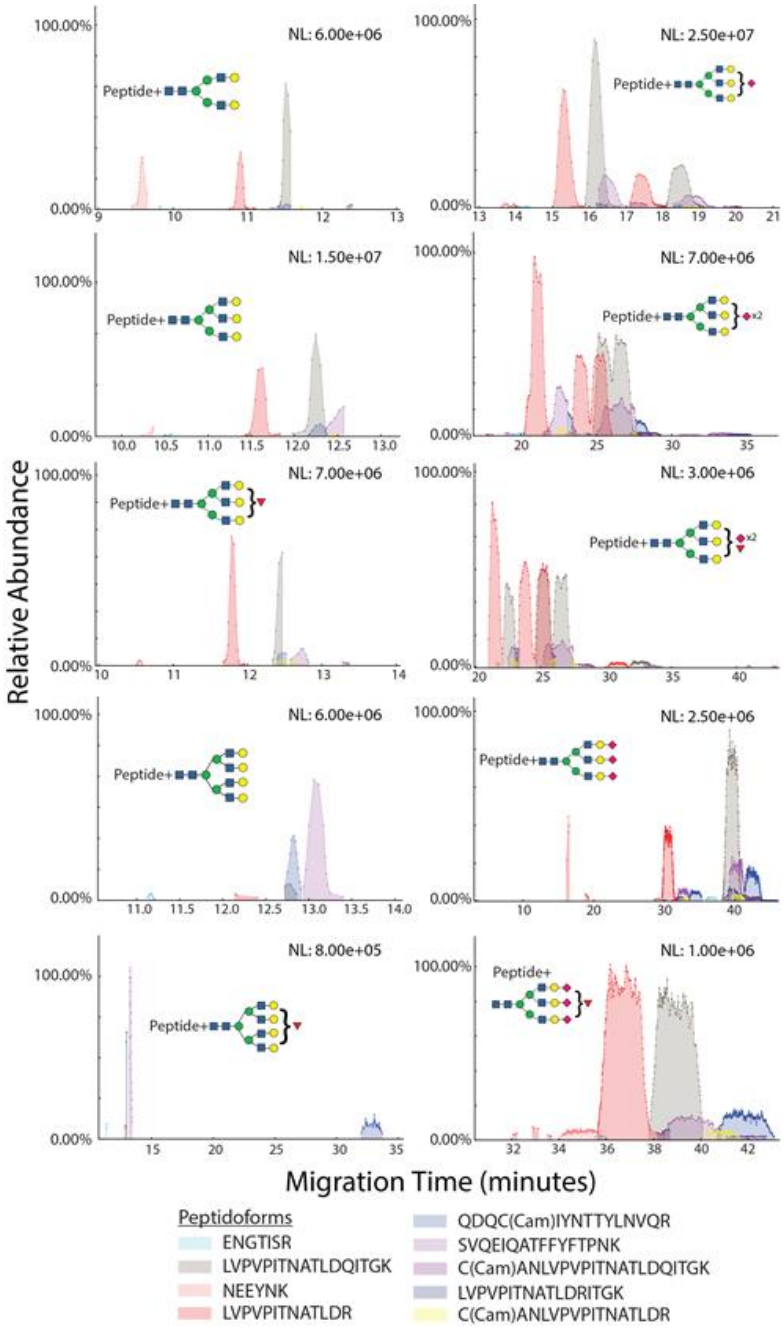


Figure 77: Glycopeptide migration by peptide backbone

Each electropherogram overlays the extracted ion traces for different peptide backbones of the same glycoform.

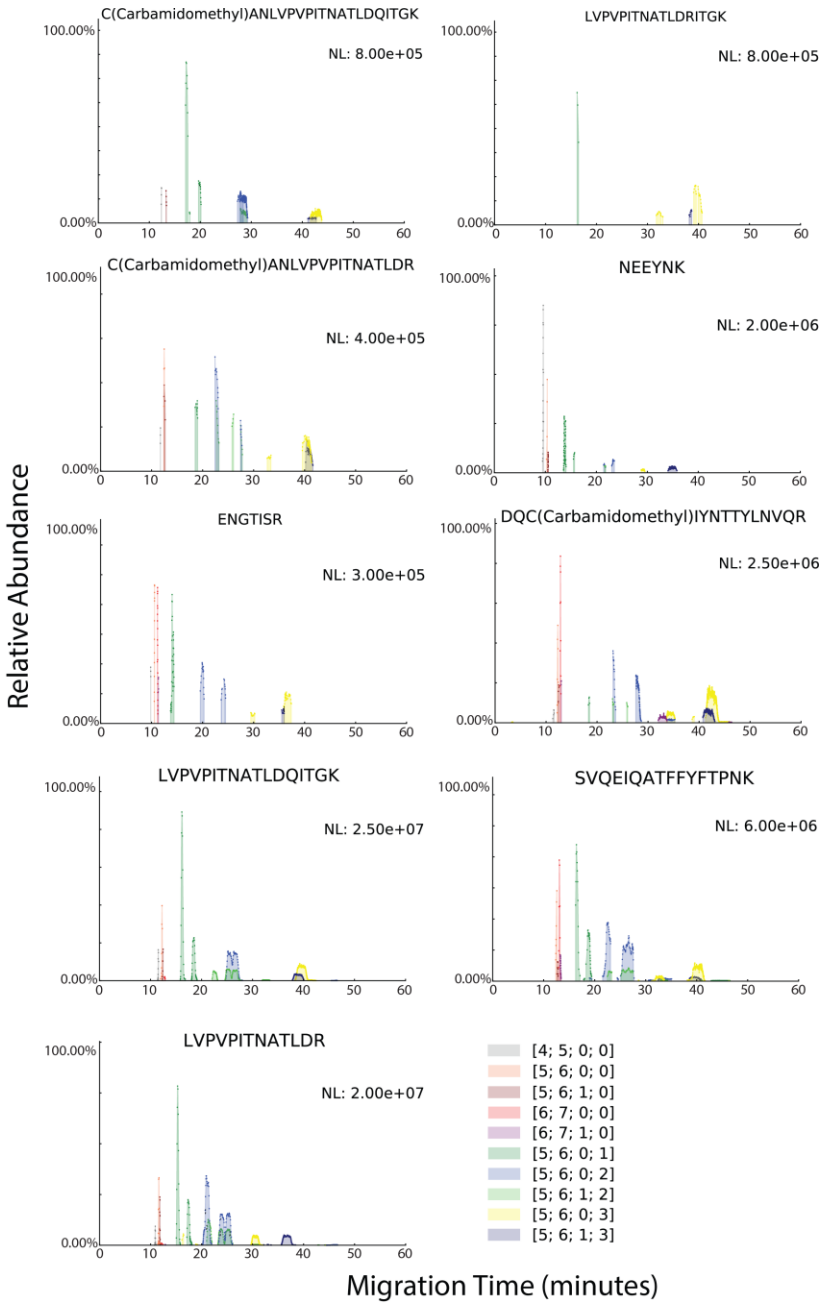


Figure 78: Glycopeptide migration by glycan composition

Each electropherogram overlays the extracted ion traces for different glycan compositions decorating the same peptide backbone.

To eliminate the effects of sialylated glycoforms on migration times, we also explored derivatization schemes similar to those used for released glycans. The Wuhrer group has reported a derivatization method for sialic acid stabilization and linkage discrimination in glycopeptide MALDI-MS, similar to their ethyl esterification procedure which we used for glycan derivatization [202]. This reaction converts α-2,6 linked sialic acid to dimethylamide, and α-2,3 linked sialic acid to a cyclic lactone with the adjacent galactose. We explored this derivatization method for CE-MS of sialylated glycopeptides from AGP. Although it did stabilize the sialic acids, the reaction introduced multiple modifications along the peptide backbone. This, combined with ammonium adduction from the ammonium acetate present in the stacking buffer, divided the signal from each glycopeptide into many channels, lowering the overall signal-to-noise ratio and making it extremely difficult to identify the glycopeptide compositions. This derivatization scheme necessitated the use of an additional HILIC-enrichment or cleanup step prior to CE-MS analysis because the reagents used for the derivatization were not soluble in aqueous solvents, and thus could not be removed via reversed-phase cleanup. To simplify the analysis, data were acquired using transferrin tryptic glycopeptides, which have only two glycosylation sites and a narrow glycoform distribution [187,293]. Methyl esterification of the glycopeptides using the method of Powell and Harvey was also attempted [423]. A few methyl esterified glycopeptides could be identified by *de novo* sequencing of the peptide backbone, but the problem of multiple backbone modifications persisted.

.

5.4.3.3 Monosaccharide analysis

Monosaccharide analysis provides insight into the overall variety of saccharide constituents in a sample. The monosaccharide compositions of unknown samples are identified and quantitated based on the retention/migration times of standards. Following analysis of the standard monosaccharide pools that had been derivatized with aminoxy TMT-127, TMT-128 or TMT-130, for which the results are shown in Figure 79, a mixture of monosaccharides, obtained from AGP by acid hydrolysis and derivatized with TMT-127, were analyzed by CE-MS. The monosaccharide identities were assigned using a table of migration times established for standard monosaccharides, as shown in Figure 80. As expected, AGP monosaccharides included fucose, mannose, galactose, *N*-acetyl glucosamine and sialic acid. The peak apex and peak width corresponding to hexoses indicated presence of mannose and galactose but no glucose (see Figure 81).

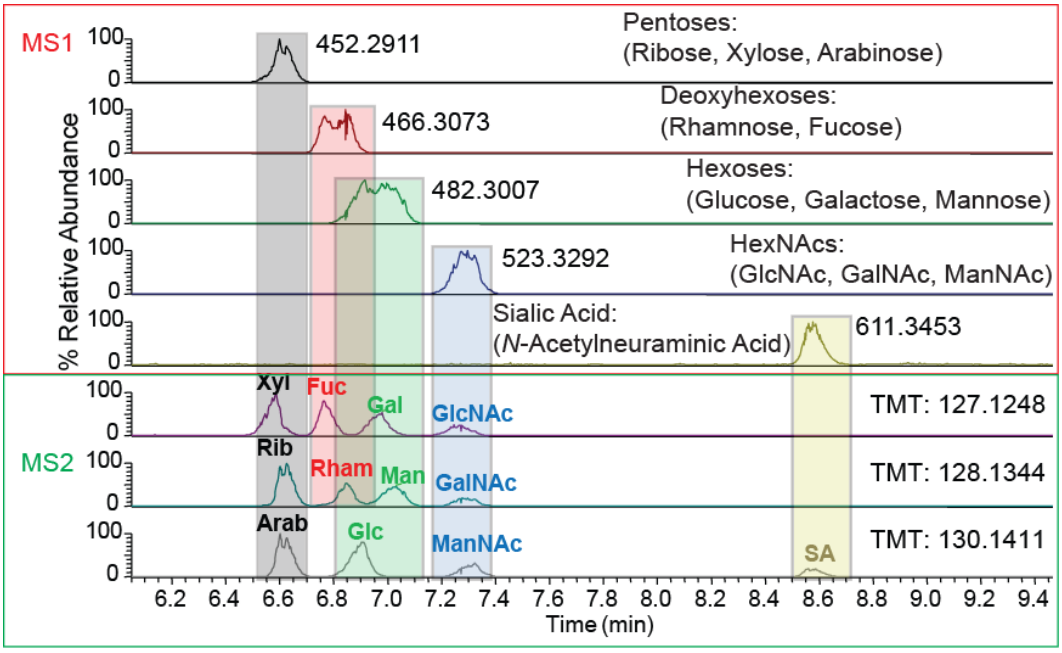


Figure 79: Monosaccharide standards multiplexed using the TMT reagent to enable identification of isomers by tandem MS

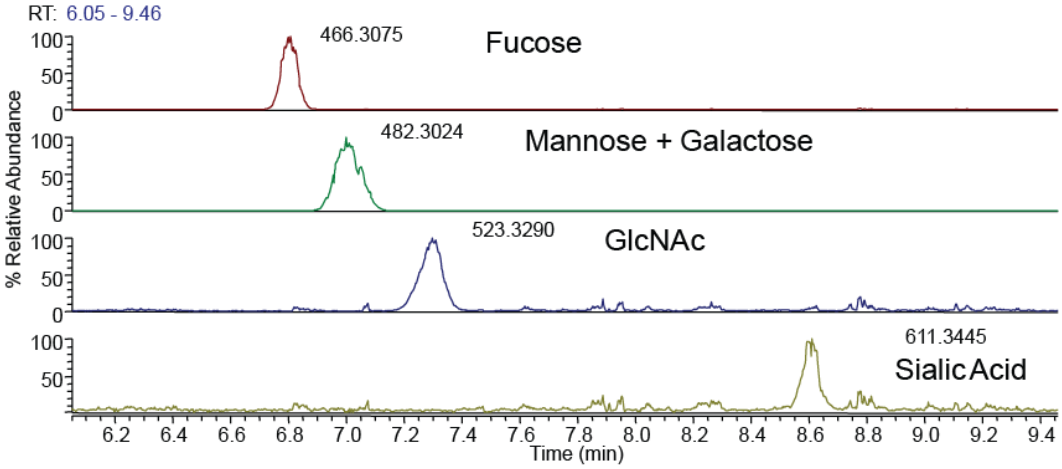


Figure 80: Extracted ion electropherograms for monosaccharides derived from AGP.

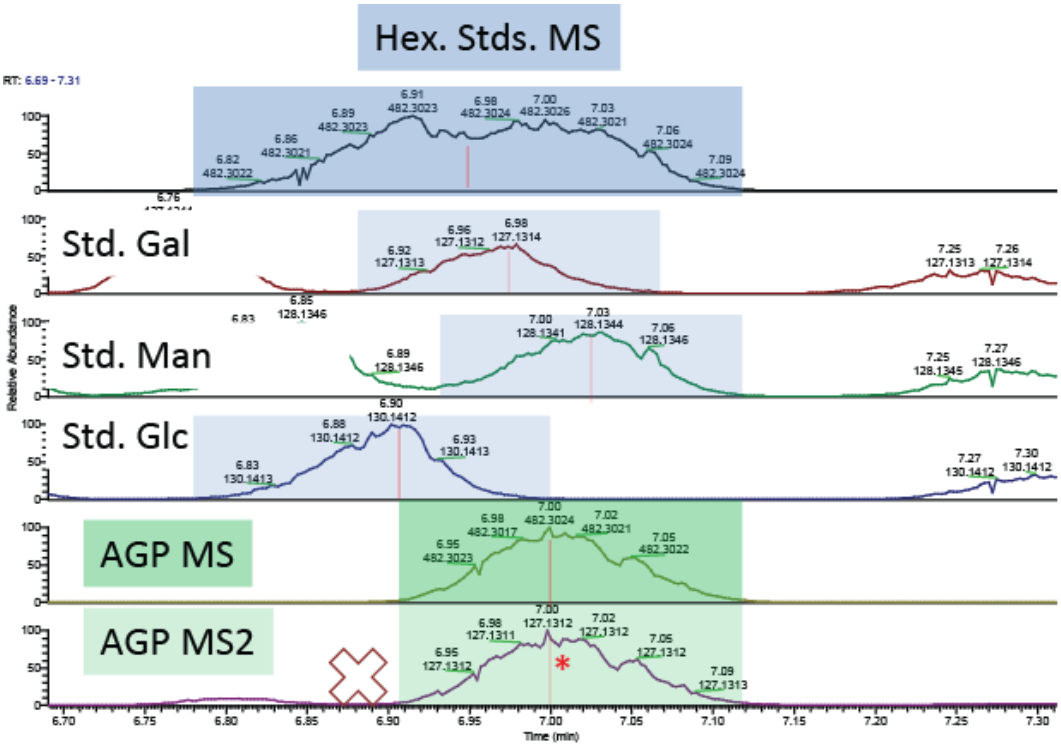


Figure 81: Overlaid traces for hexose standards and monosaccharides derived from AGP showing that AGP contains a mixture of mannose and galactose but no glucose

This workflow for monosaccharide analysis allows easy estimation of the monosaccharide composition of a sample. Even though some the HexNAc monosaccharide isomers are incompletely resolved by CE, using our testing conditions, baseline separation was achievable between the pentoses, deoxyhexoses, hexoses, HexNAcs, and the sialic acid. With the exception of the HexNAcs, separation was broad enough, and the retention time reproducible, such that each monosaccharide can be identified when compared to a standard mix. The use of TMT multiplexing offers the flexibility to develop this method to provide a better monosaccharide identity. For example, internal standards could be included in the analyte, which are labeled with a different TMT-tag, to not only provide internal RT calibrants, but enable the possibility

of quantification. Similarly, this method could be developed further to compare *N*-glycan monosaccharide compositions, and *O*-glycan compositions within the same protein sample, by differential release of *O*- and *N*-glycans, labeling with different TMT-tags, and analyzing simultaneously to have a quantifiable comparison. The ease of TMT-labeling and high-throughput capability of CE-MS makes monosaccharide analysis accessible to most with basic laboratory equipment.

5.4.4 Conclusions

The ZipChip CE-ESI device allowed easy interfacing of the CE separation capabilities with mass spectral detection. The integrated sprayer obviated the need for complex ESI source setups on CE-MS systems and delivered stable nanoESI for all experiments. Our methods utilized the same mass spectrometer friendly buffer system for glycopeptides, released glycans and monosaccharides. We showed that the same CE conditions allow analysis of three different analyte classes, thus eliminating the need for switching among different CE methods and capillary chemistries. The CE device delivered rapid separation and sensitive detection of oligosaccharides, with the sialic acid derivatization procedure allowing identification of sialic acid linkages.

Aminoxy TMT reagent was used to fix a positive charge to improve electrophoretic migration of small, neutral glycans. This reagent has been shown to be useful for relative quantitation and multiplexing of biological samples [626,629]. Sialylated glycoforms in oligosaccharides and glycopeptides that retained a negatively charged functional group under the experimental conditions appeared to interact with the

coated capillary surface, thereby causing band broadening. We used chemical derivatization to neutralize the acidic monosaccharides and eliminate this interaction for both oligosaccharides and glycopeptides. This derivatization induced linkage-specific mass-shifts for sialylated glycans, which were used for linkage-discrimination in complex mixtures. All sialic acid derivatized oligosaccharides eluted in less than 20 minutes, considerably reducing analysis times over commonly used LC modes. Glycopeptides in tryptic digests of glycoproteins were well-separated from non-glycosylated peptides, thereby reducing the need for enrichment prior to mass spectrometry analysis. Glycopeptide glycoforms were resolved based on glycan size and composition showing glycoform resolution much higher than observed using liquid chromatography separation.

Monosaccharide standards were derivatized using TMT tags and analyzed using CE-MS to establish migration times, followed by analysis of monosaccharides derived from a standard glycoprotein. Isomeric monosaccharide standards were multiplexed using TMT to generate different reporter ions allowing discrimination between co-migrating isobars. Different types of monosaccharides found on mammalian *N*-glycans were easily separated.

Glycopeptide separation by CE-MS opens avenues for improved resolution and analysis of these analytes in complex biological samples. While separation of glycopeptides from peptides obviates the need for pre-enrichment, an added enrichment step would increase loading capacity of glycopeptides and enable better tandem MS of low-abundance glycoforms.

The development of newer capillary coatings will increase compatibility with acidic analytes and improvements in hardware and availability of reversed polarity will enable analysis of acidic oligosaccharides and other polyanionic compounds. Because the separation mechanism is orthogonal to commonly used chromatography modes, our CE-MS method is attractive separation mode to combine with offline chromatographic fractionation.

5.4.5 Acknowledgements

This work was funded by NIH grant P41 GM104603. Thermo-Fisher Corporation provided access to the Q-Exactive Plus mass spectrometer. We thank 908devices Inc. for providing a loan of the ZipChip integrated capillary electrophoresis-electrospray chip. We thank Dr. Erin Redman and Dr. J. Scott Mellors for helpful advice.

Chapter 6: Conclusions and future perspectives

The work presented in this dissertation describes the development of analytical methods for analysis of glycans and glycoconjugates, with a focus on mass spectrometry and hyphenated methods such as LC and CE for separation of analytes online with mass spectral analysis. The majority of these methods were developed to tackle the challenges in analysis of complex glycoproteins in Influenza virions but all methods have wide applicability in the field of glycoprotein and glycan analysis. In addition, a bioinformatics workflow has been described that was developed to enable high-throughput and detailed analysis of the large glycoproteomics and glycomics datasets.

The first chapter summarizes the state-of-the art in glycan and glycoprotein analysis using mass spectrometry and related methods. This chapter reviews the most widely-used methods for chemical derivatization, front-end liquid-phase separation and mass spectrometry, including an overview of MS hardware and dissociation modes.

Our study of Influenza glycoproteins allowed us to elucidate site-specific glycosylation on hemagglutinins from three different IAV strains. This enabled direct comparison of changes in glycosylation on a wild-type and a mutant strain, which had very different sensitivities to inhibition by immune lectins as shown by our biochemical and binding assays. Using this information, we identified the sites involved in interactions of the virions with the immune molecule SP-D. We also utilized our glycan macro- and micro-heterogeneity information to perform modeling and simulations studies to pinpoint the structural determinants for interactions of HA with SP-D and for glycan processing at the sites involved in these interactions. This is the first IAV study that

combines site-specific glycosylation data with binding assays and *in silico* structural studies, in a systematic manner. The resolution of our multidimensional analyses is clear from our results. Our analytical and informatics workflows are now ready for analysis of a series of IAV strains where glycosylation sequons have been systematically mutated. The phenotypic information combined with bioassays and structural studies, as detailed in Chapter 4, will help identify the structural features that impart virulence and fitness to IAV and help build a model for prediction of IAV fitness based on its site-specific glycosylation profile. Together, these analyses can be used for rapid classification of emerging virus strains that can help select the best prophylactic and therapeutic interventions. These methods in general are applicable to study of glycosylation in any complex biological system.

Our informatics pipeline, GlycReSoft, takes advantage of proteomics and glycomics data to maximize the depth in analysis of glycopeptides without wasting resources on unnecessary assumptions about sample search space. Using empirically defined search spaces makes data analyses both more efficient and more confident. This pipeline was used in the analysis of data from IAV samples and also a series of increasing complexity samples containing serum glycoproteins to demonstrate the value of this approach.

Chapter 5 describes efforts in method development for glycan and glycoconjugate analysis. The first part describes development and application of a novel IM-MS workflow for adding a gas-phase separation step to biopolymers. Using this approach, we have cataloged the CCS values for scores of glycans and glycopeptides. This information

will be integrated into our data analysis pipeline using machine learning to add even more confidence to the information already delivered by our LC-MS/MS analyses. Furthermore, this approach can help resolve structural isomers and conformers to help identify structural features in biomolecules. The value of combining IMS with ExD modes has already been demonstrated by our group on an FTICR-MS system [425,426]. Recent developments in instrumentation have paved way for such applications on instruments that are capable of delivering higher scan speed that can be combined with LC-scale omics experiments [550,551,552]. While it may take some time to combine IMS and ExD on instruments that deliver high acquisition speed and sensitivity, such combinations on home-built and prototype instruments can now be realized, as demonstrated recently [99,178]. Further, we have compared the features in different collisional and electron-based dissociation modes for analysis of glycans and glycopeptides, including middle-down glycopeptides. The data from these experiments will enable development of bioinformatics strategies that are not currently available due to paucity of training data.

Also described in Chapter 5 is the application of a microfluidic CE-ESI interface for MS analysis of monosaccharides, oligosaccharides and glycopeptides. This is an important development in glycan and glycopeptide front-end separations, which often suffer from peak capacity limitations in chromatography modes that are unable to resolve glycoforms. While CE-MS methods have existed for a long time, the development of this interface has made it extremely easy to couple this orthogonal separations method with high-resolution mass spectrometry. Future work would include testing alternative

capillary coatings to minimize EOF for sialylated and sulfated glycans and different buffer conditions to improve electrophoretic separation. Further, combining CE separations with IMS-MS and with ExD would enable separation of isomeric glycans and identification of the isomers.

In addition to analysis of *N*-glycans, we have also developed methods for separation of extremely complex mixtures of linear sulfated polysaccharides of heparan sulfate. This method utilizes UPLC SEC separations and eliminates a major limitation of salt-based buffer systems in chromatography modes by online removal of salt-adducts by using an ion-suppressor system. The method was applied for profiling the complete molecular weight range of pharmaceutical preparations of low-molecular-weight heparins. The work also describes a novel data analysis workflow that takes into account the isotopic patterns for the elemental compositions of individual matches. This system is now ready for application to complex mixtures of polysaccharides from biological samples. The ion-suppressor system can be used with any chromatography mode such as HILIC or PGC, thereby extending its applicability to analysis of other glycan classes. Further, this system can be combined with ExD methods for detailed GAG structure analysis that have been developed by our group [247,257].

When analyzing proteins in a bottom-up fashion, we lose the information regarding how different parts of a molecule exist together in the intact molecular structure and are thus unable to differentiate between the many forms that are present as part of the population that we analyze. For example, there may be a range of glycan structures present at each site on a multiply glycosylated protein. When we analyze using

a bottom-up method, we are able to determine the range of glycan forms that are present at each site but we are unable to establish what combination(s) of glycans appear together on multiple sites of a complete protein. Thus, we are always building back to a molecular form that is an average of all molecules in a measured population. This hampers the ability to fully understand the interactions between specific molecular forms and to identify the forms that are relevant to a particular biological response or disease.

Together, the new methodologies that we have developed both in liquid- and gas-phase separations in mass spectral analyses and our unique informatics approach will enable realization of middle-down glycoprotein analyses. A middle-down methodology would enable us to pinpoint individual glycoproteoforms of HA and NA that interact differently with immune molecules and we can correlate these information sets to understand the relationship and effects of glycosylation that occurs simultaneously at different protein sites. Such information will allow production of better recombinant vaccines that will generate a potent and broadly neutralizing antibody response, by replicating the exact interaction of mature glycoprotein structures present on the most infectious virions with host molecules. Middle-down analyses and definition of detailed structural features in mature glycoproteoforms will require all our analytical tools and informatics strategies in both bottom-up and middle-down domains to work in concert, as shown in Figure 82.

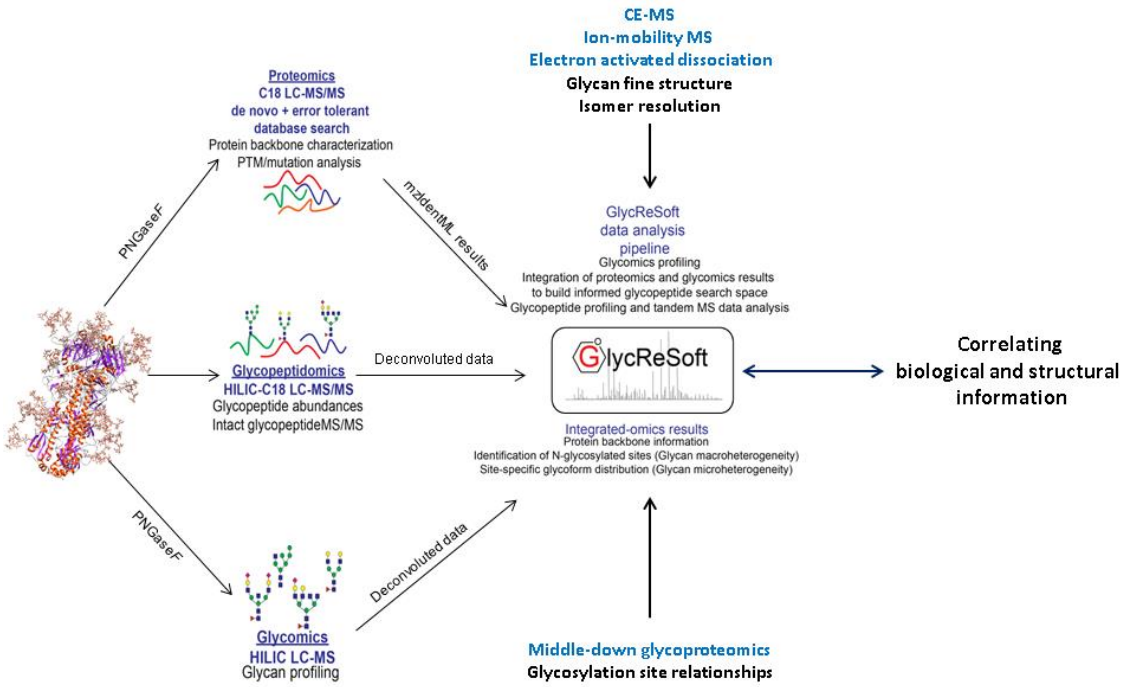


Figure 82: Extending the bottom-up analytical workflow by adding information from CE, IMS and ExD

In summary, this dissertation describes significant advancements made to methods in mass spectrometry for application to glycan and glycoprotein analysis in biological systems. While method development is pivotal to analytical technology, without real application the value of these methods remains limited to conjecture. The resolution of our methods is evident from our studies of influenza virus glycobiology and the wide-applicability of these methods will allow easy adaptation to other biological projects. It is incontrovertible that a single method or technique is not sufficient for detailed analysis of glycans and glycoproteins; for complete structural analysis, a combination of methods is necessary. We therefore conclude that bioinformatics is

crucial in logical integration of information from different analytical methods and informatics development must therefore be a concerted effort with analytics.

APPENDIX

In addition to the projects that were completed as part of this dissertation, some exploratory work was done on projects that remain incomplete at the time of writing this dissertation. This section provides an overview of the progress made on these projects and provides an outlook for the future of these projects.

Using Click-chemistry to probe specific glycan motifs in complex biological samples

Introduction

Glycan structures contain specific motifs that play important roles in biological processes. The presence and absence of certain motifs and linkages is often indicative of altered biological state. It is therefore of interest to probe biological systems for such glycans and stratify the biomolecular environment using glycans as sensitive biomarkers. Since glycan biosynthesis is regulated by enzymes that only act on substrates that contain the correct structural features and intermediate states, it is conceivable to use the enzymatic machinery to identify glycans that contain the structural features of interest and use these features as handles to isolate the glycans and glycoconjugates of interest from complex mixtures for detailed analysis.

Wu and colleagues have demonstrated the ability to probe sialosides using sialyltransferases to add monosaccharide analogs containing an azido group [583]. The

azido group can be used for azido-alkyne cycloaddition. Azido-labeled monosaccharide analogs are generally incorporated into glycans using metabolic labeling, followed by subsequent enrichment and analysis of the modified glycans or glycopeptides, as previously shown by the Bertozzi group [59,578]. Labeling using metabolic methods cannot be controlled except by modifying the genetic expression of certain glycosylation enzymes and can be performed only in samples derived from cell culture systems. Thus, these labeling approaches are not representative of the actual biological systems that one might want to study. The enzymatic labeling as described by Wu *et al.*, allows labeling of any glycoprotein derived from any biological sample.

Methods

Our initial efforts in enzymatic labeling of specific glycan structures entailed desialylalation of bovine fetuin glycoprotein using a non-specific sialidase (New England Biolabs, Ipswich MA). Asialo fetuin was then probed with recombinant sialyltrasferase (R&D systems, Minneapolis, MN) specific to adding Galα1-6NeuAc in presence of azido-CMP-sialic acid. Following the reaction, labeled fetuin was digested using trypsin as previously described [291,292,293]. Labeled glycopeptides were then reacted with an alkyne-functionalized TMT reagent, shown in Figure 83. Labeled samples were cleaned up using C18 spin columns and analyzed by data-dependent LC-MS/MS on a Q-Exactive Plus mass spectrometer using methods described previously [291].

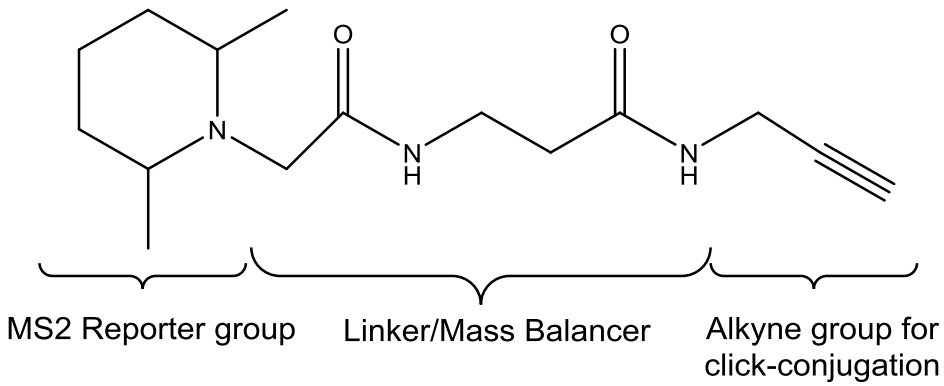


Figure 83: Alkyne-functionlized TMT reagent

Results

The TMT reagent and the modified sialic acid generated specific reporter ions in tandem MS that could be used to identify the precursors that were labeled by the sialyltransferase and TMT, as shown in

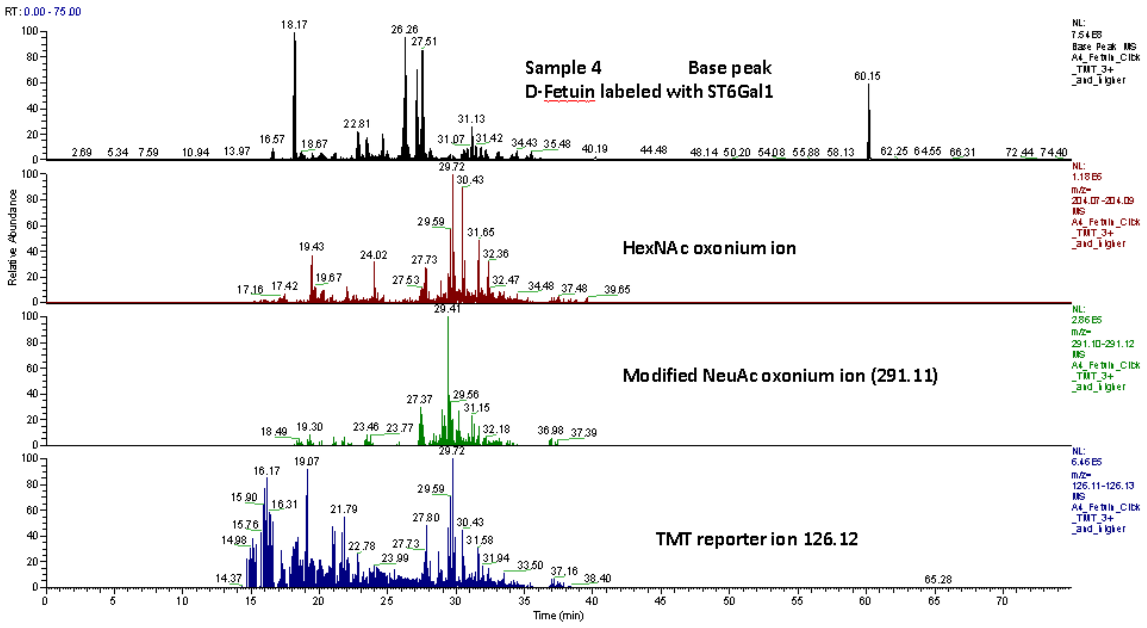


Figure 84: Overlaid chromatograms showing presence of azido-sialic acid from enzymatic labeling and click-conjugated TMT on fetuin glycopeptides

Spectra containing reporter ions were analyzed manually to identify the labeled glycopeptides, as shown in Figure 85. As shown in the figure, low mass reporter ions were present and confirmed the presence of azido-NeuAc and TMT-labeling. Other oxonium ions, stub-glycopeptide ions and peptide backbone ions were used to confirm the identity of the glycopeptide.

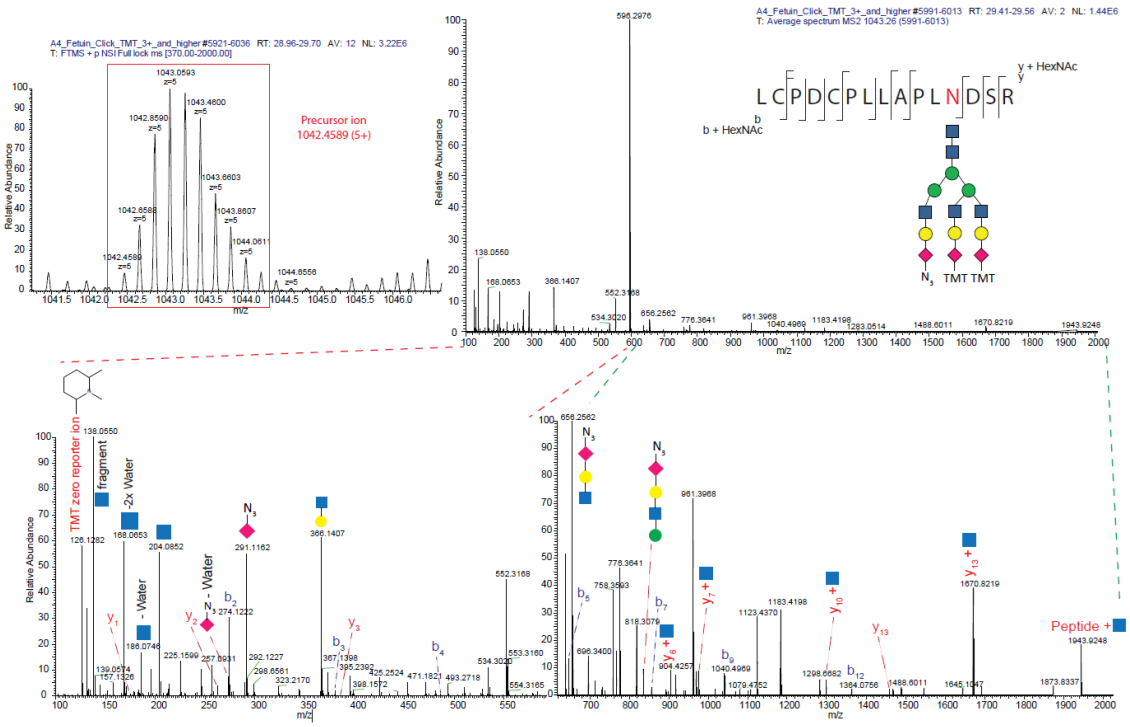


Figure 85: MS and tandem MS of a fetuin glycopeptide labeled with azido sialic acids and probed using alkyne functionalized TMT

Conclusions and future perspectives

The click-chemistry approach provides an unbiased method for specifically labeling and identifying glycan structural motifs in a site-specific manner from complex biological matrices. The TMT labeled glycopeptides can be enriched using an anti-TMT antibody that is available from Thermo Scientific. Alternatively, alkyne functionalized biotin can

be used to label, followed by affinity purification using streptavidin. The TMT labeling provides an advantage in the ability to analyze paired biological samples using multiplexing. In this pilot study, we used a sialyltransferase for labeling, but a variety of glycosyltransferases are available from R&D systems that can be used for labeling specific linkages and motifs, as needed.

A major limitation of this approach is the yield of the labeling reaction using the glycosytransferases. Application to biological samples would require consistent and high-labeling efficiencies both for the enzymatic addition of azido-monosaccharides and for subsequent click-conjugation. Further experiments will evaluate labeling efficiencies and approaches to improve yields in these reactions.

BIBLIOGRAPHY

1. Y. Abe, E. Takashita, K. Sugawara, Y. Matsuzaki, Y. Muraki, and S. Hongo. 2004. Effect of the addition of oligosaccharides on the biological activities and antigenicity of influenza A/H3N2 virus hemagglutinin. *Journal of Virology* 78: 9605–9611.

2. Julie T. Adamson and Kristina Håkansson. 2007. Electron Capture Dissociation of Oligosaccharides Ionized with Alkali, Alkaline Earth, and Transition Metals. *Anal. Chem.* 79, 7: 2901–2910.

3. Julie T. Adamson and Kristina Håkansson. 2007. Electron Detachment Dissociation of Neutral and Sialylated Oligosaccharides. *Journal of the American Society for Mass Spectrometry* 18, 12: 2162–2172.

4. Markus Aebi. 2013. N-linked protein glycosylation in the ER. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1833, 11: 2430–2437.

5. Markus Aebi, Riccardo Bernasconi, Simone Clerc, and Maurizio Molinari. 2010. N-glycan structures: recognition and processing in the ER. *Trends in Biochemical Sciences* 35, 2: 74–82.

6. Danielle Aldredge, Hyun Joo An, Ning Tang, Keith Waddell, and Carlito B. Lebrilla. 2012. Annotation of a Serum N-Glycan Library for Rapid Identification of Structures. *Journal of Proteome Research* 11, 3: 1958–1968.

7. William R. Alley, Milan Madera, Yehia Mechref, and Milos V. Novotny. 2010. Chip-based Reversed-phase Liquid Chromatography-Mass Spectrometry of Permethylated N-linked Glycans: a Potential Methodology for Cancer-biomarker Discovery. *Analytical chemistry* 82, 12: 5095–5106.

8. William R. Alley, Benjamin F. Mann, and Milos V. Novotny. 2013. High-sensitivity analytical approaches for the structural characterization of glycoproteins. *Chemical Reviews* 113, 4: 2668–2732.

9. William R. Alley, Yehia Mechref, and Milos V. Novotny. 2009. Characterization of glycopeptides by combining collision-induced dissociation and electron-transfer dissociation mass spectrometry data. *Rapid Communications in Mass Spectrometry* 23, 1: 161–170.

10. William R. Alley, Yehia Mechref, and Milos V. Novotny. 2009. Use of activated graphitized carbon chips for liquid chromatography/mass spectrometric and tandem mass spectrometric analysis of tryptic glycopeptides. *Rapid communications in mass spectrometry: RCM* 23, 4: 495–505.

11. A. J. Alpert. 1990. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *Journal of Chromatography* 499: 177–196.

12. Gerardo Alvarez-Manilla, Atwood, Yan Guo, Nicole Lynn Warren, Ron Orlando, and Michael Pierce. 2006. Tools for Glycoproteomic Analysis: Size Exclusion Chromatography Facilitates Identification of Tryptic Glycopeptides with N-linked Glycosylation Sites. *Journal of Proteome Research* 5, 3: 701–708.

13. I. Jonathan Amster. 1996. Fourier Transform Mass Spectrometry. *Journal of Mass Spectrometry* 31, 12: 1325–1337.

14. Hyun Joo An, John W. Froehlich, and Carlito B. Lebrilla. 2009. Determination of glycosylation sites and site-specific heterogeneity in glycoproteins. *Current Opinion in Chemical Biology* 13, 4: 421–426.

15. Hyun Joo An and Carlito B. Lebrilla. 2011. Structure elucidation of native N- and O-linked glycans by tandem mass spectrometry (tutorial). *Mass Spectrometry Reviews* 30, 4: 560–578.

16. Yanming An and John F. Cipollo. 2011. An unbiased approach for analysis of protein glycosylation and application to influenza vaccine hemagglutinin. *Analytical Biochemistry* 415, 1: 67–80.

17. Yanming An, Jonathan A. McCullers, Irina Alymova, Lisa M. Parsons, and John F. Cipollo. 2015. Glycosylation Analysis of Engineered H3N2 Influenza A Virus Hemagglutinins with Sequentially Added Historically Relevant Glycosylation Sites. *Journal of Proteome Research* 14, 9: 3957–3969.

18. Yanming An, Joseph A. Rininger, Donald L. Jarvis, et al. 2013. Comparative Glycomics Analysis of Influenza Hemagglutinin (H5N1) Produced in Vaccine Relevant Cell Platforms. *Journal of Proteome Research* 12, 8: 3707–3720.

19. E M Anders, C A Hartley, and D C Jackson. 1990. Bovine and mouse serum beta inhibitors of influenza A viruses are mannose-binding lectins. *Proceedings of the National Academy of Sciences of the United States of America* 87, 12: 4485–4489.

20. Leigh Anderson and Christie L. Hunter. 2006. Quantitative Mass Spectrometric Multiple Reaction Monitoring Assays for Major Plasma Proteins. *Molecular & Cellular Proteomics* 5, 4: 573–588.

21. N. Leigh Anderson and Norman G. Anderson. 2002. The Human Plasma Proteome History, Character, and Diagnostic Prospects. *Molecular & Cellular Proteomics* 1, 11: 845–867.

22. T. Angata and A. Varki. 2002. Chemical diversity in the sialic acids and related alpha-keto acids: An evolutionary perspective. *Chemical Reviews* 102: 439–469.

23. Peggi M. Angel, Jae-Min Lim, Lance Wells, Carl Bergmann, and Ron Orlando. 2007. A potential pitfall in 18O-based N-linked glycosylation site mapping. *Rapid Communications in Mass Spectrometry* 21, 5: 674–682.

24. Thomas M. Annesley. 2003. Ion Suppression in Mass Spectrometry. *Clinical Chemistry* 49, 7: 1041–1044.

25. Kalyan R. Anumula. 2006. Advances in fluorescence derivatization methods for high-performance liquid chromatographic analysis of glycoprotein carbohydrates. *Analytical Biochemistry* 350, 1: 1–23.

26. Kalyan Rao Anumula and Shirish T. Dhume. 1998. High resolution and high sensitivity methods for oligosaccharide mapping and characterization by normal phase high performance liquid chromatography following derivatization with highly fluorescent anthranilic acid. *Glycobiology* 8, 7: 685–694.

27. R. Apweiler, H. Hermjakob, and N. Sharon. 1999. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica Et Biophysica Acta-General Subjects* 1473: 4–8.

28. David J. Ashline, Anthony J. Lapadula, Yan-Hui Liu, et al. 2007. Carbohydrate Structural Isomers Analyzed by Sequential Mass Spectrometry. *Analytical chemistry* 79, 10: 3830–3842.

29. David J. Ashline, Ying Yu, Yi Lasanajak, et al. 2014. Structural characterization by MSn of human milk glycans recognized by human rotaviruses. *Molecular & Cellular Proteomics*: mcp.M114.039925.

30. David Ashline, Suddham Singh, Andy Hanneman, and Vernon Reinhold. 2005. Congruent strategies for carbohydrate sequencing. 1. Mining structural details by MSn. *Analytical chemistry* 77, 19: 6250–6262.

31. James A. Atwood, Lei Cheng, Gerardo Alvarez-Manilla, Nicole L. Warren, William S. York, and Ron Orlando. 2008. Quantitation by isobaric labeling: applications to glycomics. *Journal of Proteome Research* 7, 1: 367–374.

32. Peter R. Baker, Jonathan C. Trinidad, and Robert J. Chalkley. 2011. Modification site localization scoring integrated into a search engine. *Molecular & Cellular Proteomics: MCP* 10, 7: M111.008078.

33. Ray Bakhtiar and Ziqiang Guan. 2006. Electron capture dissociation mass spectrometry in characterization of peptides and proteins. *Biotechnology Letters* 28, 14: 1047–1059.

34. Elvira Balaguer and Christian Neusüss. 2006. Glycoprotein Characterization Combining Intact Protein and Glycan Analysis by Capillary Electrophoresis-Electrospray Ionization-Mass Spectrometry. *Anal. Chem.* 78, 15: 5384–5393.

35. Begona Barroso, Mihaela Didraga, and Rainer Bischoff. 2005. Analysis of proteoglycans derived sulphated disaccharides by liquid chromatography/mass spectrometry. *Journal of Chromatography A* 1080, 1: 43–48.

36. Nicholas G. Batz, J. Scott Mellors, Jean Pierre Alarie, and J. Michael Ramsey. 2014. Chemical Vapor Deposition of Aminopropyl Silanes in Microfluidic Channels for Highly Efficient Microchip Capillary Electrophoresis-Electrospray Ionization-Mass Spectrometry. *Analytical Chemistry* 86, 7: 3493–3500.

37. Manuel Bauer, Erik Ahrné, Anna P. Baron, et al. 2014. Evaluation of Data-Dependent and -Independent Mass Spectrometric Workflows for Sensitive Quantification of Proteins and Phosphorylation Sites. *Journal of Proteome Research* 13, 12: 5973–5988.

38. E. Bause and H. Hettkamp. 1979. Primary Structural Requirements for N-Glycosylation of Peptides in Rat-Liver. *Febs Letters* 108: 341–344.

39. E. Bause and G. Legler. 1981. The Role of the Hydroxy Amino-Acid in the Triplet Sequence Asn-Xaa-Thr(Ser) for the N-Glycosylation Step during Glycoprotein-Biosynthesis. *Biochemical Journal* 195: 639–644.

40. Anne P. Beigneux, Peter Gin, Brandon S. J. Davies, et al. 2008. Glycosylation of Asn-76 in mouse GPIHBP1 is critical for its appearance on the cell surface and the binding of chylomicrons and lipoprotein lipase. *Journal of Lipid Research* 49, 6: 1312–1321.

41. S. Ben-Dor, N. Esterman, E. Rubin, and N. Sharon. 2004. Biases and complex patterns in the residues flanking protein N-glycosylation sites. *Glycobiology* 14: 95–101.

42. R. E. Benton and B. J. Gersh. 1998. Cardiology for the primary care provider: heparin-induced thrombocytopenia. *South Med J* 91: 133–7.

43. Marshall Bern, Yong J. Kil, and Christopher Becker. 2012. Byonic: advanced peptide and protein identification software. *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]* Chapter 13: Unit13.20.

44. Vivek N. Bhatia, David H. Perlman, Catherine E. Costello, and Mark E. McComb. 2009. Software Tool for Researching Annotations of Proteins: Open-Source Protein Annotation Software with Data Visualization. *Analytical Chemistry* 81, 23: 9819–9823.

45. K. Biemann and H. A. Scoble. 1987. Characterization by tandem mass spectrometry of structural modifications in proteins. *Science* 237, 4818: 992–998.

46. Klaus Biemann. 1990. [25] Sequencing of peptides by tandem mass spectrometry and high-energy collision-induced dissociation. In B.-M. in Enzymology, ed., Academic Press, 455–479.

47. Klaus Biemann and Stephen A. Martin. 1987. Mass spectrometric determination of the amino acid sequence of peptides and proteins. *Mass Spectrometry Reviews* 6, 1: 1–75.

48. J. C. Bigge, T. P. Patel, J. A. Bruce, P. N. Goulding, S. M. Charles, and R. B. Parekh. 1995. Nonselective and efficient fluorescent labeling of glycans using 2-amino benzamide and anthranilic acid. *Analytical Biochemistry* 230, 2: 229–238.

49. Benjamin P. Blackburne, Alan J. Hay, and Richard A. Goldstein. 2008. Changing Selective Pressure during Antigenic Changes in Human Influenza H3. *PLoS Pathog* 4, 5: e1000058.

50. Marco R. Bladergroen, Karli R. Reiding, Agnes L. Hipgrave Ederveen, et al. 2015. Automation of High-Throughput Mass Spectrometry-Based Plasma N-Glycome Analysis with Linkage-Specific Sialic Acid Esterification. *Journal of Proteome Research* 14, 9: 4080–4086.

51. Ola Blixt, Steve Head, Tony Mondala, et al. 2004. Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proceedings of the National Academy of Sciences of the United States of America* 101, 49: 17033–17038.

52. Paul J. Boersema, Nullin Divecha, Albert J. R. Heck, and Shabaz Mohammed. 2007. Evaluation and optimization of ZIC-HILIC-RP as an alternative MudPIT strategy. *Journal of Proteome Research* 6: 937–946.

53. Paul J. Boersema, Tamar Geiger, Jacek R. Wisniewski, and Matthias Mann. 2013. Quantification of the N-glycosylated Secretome by Super-SILAC During Breast Cancer Progression and in Human Blood Samples. *Molecular & Cellular Proteomics* 12: 158–171.

54. Brian C. Bohrer, Samuel I. Merenbloom, Stormy L. Koeniger, Amy E. Hilderbrand, and David E. Clemmer. 2008. Biomolecule analysis by ion mobility spectrometry. *Annual Review of Analytical Chemistry (Palo Alto, Calif.)* 1: 293–327.

55. Ivan A. Boldin and Eugene N. Nikolaev. 2011. Fourier transform ion cyclotron resonance cell with dynamic harmonization of the electric field in the whole volume by shaping of the excitation and detection electrode assembly. *Rapid Communications in Mass Spectrometry* 25, 1: 122–126.

56. Sandrine Bourgoin-Voillard, Nancy Leymarie, and Catherine E. Costello. 2014. Top-down tandem mass spectrometry on RNase A and B using a Qh/FT-ICR hybrid mass spectrometer. *Proteomics* 14, 10: 1174–1184.

57. Michael J. Bowman and Joseph Zaia. 2007. Tags for the stable isotopic labeling of carbohydrates and quantitative analysis by mass spectrometry. *Analytical Chemistry* 79, 15: 5777–5784.

58. Michael J. Bowman and Joseph Zaia. 2010. Comparative Glycomics using A Tetraplex Stable-Isotope Coded Tag. *Analytical chemistry* 82, 7: 3023–3031.

59. Mark A. Breidenbach, Jennifer E. G. Gallagher, David S. King, Brian P. Smart, Peng Wu, and Carolyn R. Bertozzi. 2010. Targeted Metabolic Labeling of Yeast N-Glycans with Unnatural Sugars. *Proceedings of the National Academy of Sciences*.

60. Inka Brockhausen and Harry Schachter. 1996. Glycosyltransferases Involved in N– and O–Glycan Biosynthesis. In H.-J. Gabius and S. Gabius, eds., *Glycosciences*. Wiley-VCH Verlag GmbH, 79–113.

61. Jennifer S. Brodbelt. 2014. Photodissociation mass spectrometry: New tools for characterization of biological molecules. *Chemical Society reviews* 43, 8: 2757–2783.

62. Robert S. Brown and John J. Lennon. 1995. Mass Resolution Improvement by Incorporation of Pulsed Ion Extraction in a Matrix-Assisted Laser Desorption/Ionization Linear Time-of-Flight Mass Spectrometer. *Analytical Chemistry* 67, 13: 1998–2003.

63. Bogdan A. Budnik, Kim F. Haselmann, and Roman A. Zubarev. 2001. Electron detachment dissociation of peptide di-anions: an electron–hole recombination phenomenon. *Chemical Physics Letters* 342, 3–4: 299–302.

64. Yan Cai, Jing Jiao, Zhichao Bin, Ying Zhang, Pengyuan Yang, and Haojie Lu. 2015. Glycan reductive isotope-coded amino acid labeling (GRIAL) for mass

spectrometry-based quantitative N-glycomics. *Chemical Communications (Cambridge, England)* 51, 4: 772–775.

65. J. M. Campbell, B. A. Collings, and D. J. Douglas. 1998. A new linear ion trap time-of-flight system with tandem mass spectrometry capabilities. *Rapid Communications in Mass Spectrometry* 12, 20: 1463–1474.

66. M. P. Campbell, W. Struwe, K. Pagel, J. L. P. Benesch, and D. Harvey. 2015. GlycoMob - an Ion Mobility Mass Spectrometry Database. *Glycobiology* 25: 1275–1276.

67. Joe Cannon, Karen Lohnes, Colin Wynne, Yan Wang, Nathan Edwards, and Catherine Fenselau. 2010. High-Throughput Middle-Down Analysis Using an Orbitrap. *Journal of Proteome Research* 9, 8: 3886–3890.

68. Weiqian Cao, Jiangming Huang, Biyun Jiang, Xing Gao, and Pengyuan Yang. 2016. Highly Selective Enrichment of Glycopeptides Based on Zwitterionically Functionalized Soluble Nanopolymers. *Scientific Reports* 6: 29776.

69. Vincenzo Carbone, Elena K. Schneider, Steve Rockman, et al. 2015. Molecular Characterisation of the Haemagglutinin Glycan-Binding Specificity of Egg-Adapted Vaccine Strains of the Pandemic 2009 H1N1 Swine Influenza A Virus. *Molecules* 20, 6: 10415–10434.

70. Don M. Carlson. 1966. Oligosaccharides Isolated from Pig Submaxillary Mucin. *Journal of Biological Chemistry* 241, 12: 2984–2986.

71. Steven A. Carr, Michael J. Huddleston, and Mark F. Bean. 1993. Selective identification and differentiation of N-and O-linked oligosaccharides in glycoproteins by liquid chromatography-mass spectrometry. *Protein Science* 2, 2: 183–196.

72. F. Carrat and A. Flahault. 2007. Influenza vaccine: The challenge of antigenic drift. *Vaccine* 25, 39–40: 6852–6862.

73. DA Case, JT Berryman, RM Betz, et al. 2015. AMBER 2015. *University of California, San Francisco*.

74. Jonatan R. Catai, Javier Sastre Toraño, Gerhardus J. de Jong, and Govert W. Somsen. 2006. Efficient and highly reproducible capillary electrophoresis-mass spectrometry of peptides using Polybrene-poly(vinyl sulfonate)-coated capillaries. *Electrophoresis* 27, 11: 2091–2099.

75. M. Isabel Catalina, Carolien A. M. Koeleman, André M. Deelder, and Manfred Wuhrer. 2007. Electron transfer dissociation of N-glycopeptides: loss of the entire

N-glycosylated asparagine side chain. *Rapid communications in mass spectrometry: RCM* 21, 6: 1053–1061.

76. Adam D. Catherman, Owen S. Skinner, and Neil L. Kelleher. 2014. Top Down proteomics: Facts and perspectives. *Biochemical and Biophysical Research Communications* 445, 4: 683–693.

77. Andrew G. Chambers, J. Scott Mellors, W. Hampton Henley, and J. Michael Ramsey. 2011. Monolithic Integration of Two-Dimensional Liquid Chromatography−Capillary Electrophoresis and Electrospray Ionization on a Microfluidic Device. *Analytical Chemistry* 83, 3: 842–849.

78. Wai Yi Kelly Chan, T. W. Dominic Chan, and Peter B. O'Connor. 2010. Electron Transfer Dissociation with Supplemental Activation to Differentiate Aspartic and Isoaspartic Residues in Doubly Charged Peptide Cations. *Journal of the American Society for Mass Spectrometry* 21, 6: 1012–1015.

79. Kevin Brown Chandler and Catherine E. Costello. 2016. Glycomics and glycoproteomics of membrane proteins and cell-surface receptors: Present trends and future opportunities. *Electrophoresis* 37, 11: 1407–1419.

80. Kevin Chandler and Radoslav Goldman. 2013. Glycoprotein Disease Markers and Single Protein-omics. *Molecular & Cellular Proteomics* 12, 4: 836–845.

81. Wei-Chuan Chang, Mitchell R White, Patience Moyo, et al. 2010. Lack of the pattern recognition molecule mannose-binding lectin increases susceptibility to influenza A virus infection. *BMC Immunology* 11: 64.

82. F. Y. Che, X. X. Shao, K. Y. Wang, and Q. C. Xia. 1999. Characterization of derivatization of sialic acid with 2-aminoacridone and determination of sialic acid content in glycoproteins by capillary electrophoresis and high performance liquid chromatography--ion trap mass spectrometry. *Electrophoresis* 20, 14: 2930–2937.

83. Fu-Tai A. Chen, Thomas S. Dobashi, and Ramon A. Evangelista. 1998. Quantitative analysis of sugar constituents of glycoproteins by capillary electrophoresis. *Glycobiology* 8, 11: 1045–1052.

84. Juine-Ruey Chen, Yueh-Hsiang Yu, Yung-Chieh Tseng, et al. 2014. Vaccination of monoglycosylated hemagglutinin induces cross-strain protection against influenza virus infections. *Proceedings of the National Academy of Sciences* 111, 7: 2476–2481.

85. P. Chen, U. Werner-Zwanziger, D. Wiesler, M. Pagel, and M. V. Novotny. 1999. Mass spectrometric analysis of benzoylated sialooligosaccharides and

differentiation of terminal alpha 2-->3 and alpha 2-->6 sialogalactosylated linkages at subpicomole levels. *Analytical Chemistry* 71, 21: 4969–4973.

86. Joshua L. Cherry, David J. Lipman, Anastasia Nikolskaya, and Yuri I Wolf. 2009. Evolutionary Dynamics of N-Glycosylation Sites of Influenza Virus Hemagglutinin. *PLoS Currents* 1.

87. Chunaram Choudhary and Matthias Mann. 2010. Decoding signalling networks by mass spectrometry-based proteomics. *Nature Reviews Molecular Cell Biology* 11: 427–439.

88. Caroline S. Chu, Milady R. Niñonuevo, Brian H. Clowers, et al. 2009. Profile of Native N-linked Glycan Structures from Human Serum Using High Performance Liquid Chromatography on a Microfluidic Chip and Time-of-Flight Mass Spectrometry. *Proteomics* 9, 7: 1939–1951.

89. Ionel Ciucanu and Catherine E. Costello. 2003. Elimination of Oxidative Degradation during the per-O-Methylation of Carbohydrates. *Journal of the American Chemical Society* 125, 52: 16213–16219.

90. Ionel Ciucanu and Francisc Kerek. 1984. A simple and rapid method for the permethylation of carbohydrates. *Carbohydrate Research* 131, 2: 209–217.

91. D. E. Clemmer and M. F. Jarrold. 1997. Ion mobility measurements and their applications to clusters and biomolecules. *Journal of Mass Spectrometry* 32: 577–592.

92. Melvin B. Comisarow and Alan G. Marshall. 1974. Fourier transform ion cyclotron resonance spectroscopy. *Chemical Physics Letters* 25, 2: 282–283.

93. Melvin B. Comisarow and Alan G. Marshall. 1974. Frequency-sweep fourier transform ion cyclotron resonance spectroscopy. *Chemical Physics Letters* 26, 4: 489–490.

94. Yu-Sheng Cong, Ellen Fan, and Eugenia Wang. 2006. Simultaneous proteomic profiling of four different growth states of human fibroblasts, using amine-reactive isobaric tagging reagents and tandem mass spectrometry. *Mechanisms of Ageing and Development* 127, 4: 332–343.

95. H. Edward Conrad. 2001. β-Elimination for Release of O-Linked Glycosaminoglycans from Proteoglycans. In *Current Protocols in Molecular Biology*. John Wiley & Sons, Inc.

96. C. A. Cooper, N. H. Packer, and J. W. Redmond. 1994. The elimination of O-linked glycans from glycoproteins under non-reducing conditions. *Glycoconjugate Journal* 11, 2: 163–167.

97. Helen J. Cooper, Kristina Håkansson, and Alan G. Marshall. 2005. The role of electron capture dissociation in biomolecular analysis. *Mass Spectrometry Reviews* 24, 2: 201–222.

98. Catherine E. Costello, Joy May Contado-Miller, and John F. Cipollo. 2007. A glycomics platform for the analysis of permethylated oligosaccharide alditols. *Journal of the American Society for Mass Spectrometry* 18, 10: 1799–1812.

99. Catherine E Costello, Rebecca S. Glaskin, Valery G. Voinov, et al. 2016. Applications of Ion Mobility Separation with Electron-based Dissociation Tandem Mass Spectrometry. *Proceedings of the 21st International Mass Spectrometry Conference*.

100. J. N. Couceiro, J. C. Paulson, and L. G. Baum. 1993. Influenza virus strains selectively recognize sialyloligosaccharides on human respiratory epithelium; the role of the host cell in selection of hemagglutinin receptor specificity. *Virus Research* 29, 2: 155–165.

101. C. S. Creaser, J. R. Griffiths, C. J. Bramwell, S. Noreen, C. A. Hill, and C. L. P. Thomas. 2004. Ion mobility spectrometry: a review. Part 1. Structural analysis by mobility measurement. *Analyst* 129: 984–994.

102. David M. Creasy and John S. Cottrell. 2002. Error tolerant searching of uninterpreted tandem mass spectrometry data. *PROTEOMICS* 2, 10: 1426–1434.

103. David M. Creasy and John S. Cottrell. 2004. Unimod: Protein modifications for mass spectrometry. *Proteomics* 4, 6: 1534–1536.

104. Andrew J. Creese and Helen J. Cooper. 2012. Separation and Identification of Isomeric Glycopeptides by High Field Asymmetric Waveform Ion Mobility Spectrometry. *Analytical Chemistry* 84, 5: 2597–2601.

105. E. Crouch, A. Persson, D. Chang, and J. Heuser. 1994. Molecular structure of pulmonary surfactant protein D (SP-D). *The Journal of Biological Chemistry* 269, 25: 17311–17319.

106. Erika Crouch, Kevan Hartshorn, Tim Horlacher, et al. 2009. Recognition of Mannosylated Ligands and Influenza A Virus by Human Surfactant Protein D: Contributions of an Extended Site and Residue 343. *Biochemistry* 48, 15: 3335–3345.

107. Erika Crouch, Barbara McDonald, Kelly Smith, et al. 2007. Critical role of Arg/Lys343 in the species-dependent recognition of phosphatidylinositol by pulmonary surfactant protein D. *Biochemistry* 46, 17: 5160–5169.

108. Erika Crouch, Barbara McDonald, Kelly Smith, Tanya Cafarella, Barbara Seaton, and James Head. 2006. Contributions of Phenylalanine 335 to Ligand Recognition by Human Surfactant Protein D Ring Interactions with SP-D Ligands. *Journal of Biological Chemistry* 281, 26: 18008–18014.

109. Erika Crouch, Nikolaos Nikolaidis, Francis X. McCormack, et al. 2011. Mutagenesis of surfactant protein D informed by evolution and x-ray crystallography enhances defenses against influenza A virus in vivo. *The Journal of Biological Chemistry* 286, 47: 40681–40692.

110. Erika Crouch, Yizheng Tu, David Briner, et al. 2005. Ligand Specificity of Human Surfactant Protein D EXPRESSION OF A MUTANT TRIMERIC COLLECTIN THAT SHOWS ENHANCED INTERACTIONS WITH INFLUENZA A VIRUS. *Journal of Biological Chemistry* 280, 17: 17046–17056.

111. Jike Cui, Temple Smith, Phillips W. Robbins, and John Samuelson. 2009. Darwinian selection for sites of Asn-linked glycosylation in phylogenetically disparate eukaryotes and viruses. *Proceedings of the National Academy of Sciences* 106, 32: 13421–13426.

112. R. D. Cummings and S. Kornfeld. 1982. Fractionation of asparagine-linked oligosaccharides by serial lectin-Agarose affinity chromatography. A rapid, sensitive, and specific technique. *Journal of Biological Chemistry* 257, 19: 11235–11240.

113. David C Dallas, William F Martin, Serenus Hua, and J Bruce German. 2013. Automated glycopeptide analysis--review of current state and future directions. *Briefings in bioinformatics* 14, 3: 361–374.

114. M. A. Daniels, K. A. Hogquist, and S. C. Jameson. 2002. Sweet “n” sour: the impact of differential glycosylation on T cell responses. *Nature Immunology* 3: 903–910.

115. Tom Darden, Darrin York, and Lee Pedersen. 1993. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* 98, 12: 10089–10092.

116. Petra Darebna, Petr Novak, Radek Kucera, et al. 2017. Changes in the expression of N- and O-glycopeptides in patients with colorectal cancer and hepatocellular

carcinoma quantified by full-MS scan FT-ICR and multiple reaction monitoring. *Journal of Proteomics* 153: 44–52.

117. Suman R. Das, Scott E. Hensley, Alexandre David, et al. 2011. Fitness costs limit influenza A virus hemagglutinin glycosylation as an immune evasion strategy. *Proceedings of the National Academy of Sciences* 108, 51: E1417–E1422.

118. Suman R. Das, Pere Puigbò, Scott E. Hensley, Darrell E. Hurt, Jack R. Bennink, and Jonathan W. Yewdell. 2010. Glycosylation Focuses Sequence Variation in the Influenza A Virus H1 Hemagglutinin Globular Domain. *PLoS Pathog* 6, 11: e1001211.

119. James W. Dennis, Maria Granovsky, and Charles E. Warren. 1999. Protein glycosylation in development and disease. *BioEssays* 21, 5: 412–421.

120. Nikunj Desai, Daniel A. Thomas, Jungeun Lee, Jinshan Gao, and J. L. Beauchamp. 2016. Eradicating mass spectrometric glycan rearrangement by utilizing free radicals. *Chemical Science* 7, 8: 5390–5397.

121. Heather Desaire. 2013. Glycopeptide Analysis, Recent Developments and Applications. *Molecular & Cellular Proteomics* 12, 4: 893–901.

122. Heather Desaire and David Hua. 2009. When can glycopeptides be assigned based solely on high-resolution mass spectrometry data? *International Journal of Mass Spectrometry* 287, 1–3: 21–26.

123. Janie L. Desantos-Garcia, Sarah I. Khalil, Ahmed Hussein, Yunli Hu, and Yehia Mechref. 2011. Enhanced sensitivity of LC-MS analysis of permethylated N-glycans through online purification. *Electrophoresis* 32, 24: 3516–3525.

124. Arugadoss Devakumar, Yehia Mechref, Pilsoo Kang, Milos V. Novotny, and James P. Reilly. 2007. Laser-induced photofragmentation of neutral and acidic glycans inside an ion-trap mass spectrometer. *Rapid communications in mass spectrometry: RCM* 21, 8: 1452–1460.

125. Arugadoss Devakumar, Yehia Mechref, Pilsoo Kang, Milos V. Novotny, and James P. Reilly. 2008. Identification of Isomeric N-Glycan Structures by Mass Spectrometry with 157 nm Laser-Induced Photofragmentation. *Journal of the American Society for Mass Spectrometry* 19, 7: 1027–1040.

126. Mahender B. Dewal, Andrew S. DiChiara, Aristotelis Antonopoulos, et al. 2015. XBP1s Links the Unfolded Protein Response to the Molecular Architecture of Mature N-Glycans. *Chemistry & Biology* 22, 10: 1301–1312.

127. Francesco Di Girolamo, Isabella Lante, Maurizio Muraca, and Lorenza Putignani. 2013. The Role of Mass Spectrometry in the "Omics" Era. *Current Organic Chemistry* 17, 23: 2891–2905.

128. P. Dittwald, J. Claesen, T. Burzykowski, D. Valkenborg, and A. Gambin. 2013. BRAIN: a universal tool for high-throughput calculations of the isotopic distribution for mass spectrometry. *Anal Chem* 85: 1991–4.

129. Bruno Domon and Catherine E Costello. 1988. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconjugate Journal* 5, 4: 397–409.

130. Xue Dong, Shiyue Zhou, and Yehia Mechref. 2016. LC-MS/MS analysis of permethylated free oligosaccharides and N-glycans derived from human, bovine, and goat milk samples. *Electrophoresis* 37, 11: 1532–1548.

131. Donald J. Douglas, Aaron J. Frank, and Dunmin Mao. 2005. Linear ion traps in mass spectrometry. *Mass Spectrometry Reviews* 24, 1: 1–29.

132. K. Drickamer. 1988. 2 Distinct Classes of Carbohydrate-Recognition Domains in Animal Lectins. *Journal of Biological Chemistry* 263: 9557–9560.

133. M. V. Dwek, H. A. Ross, and A. J. C. Leathem. 2001. Proteome and glycosylation mapping identifies post-translational modifications associated with aggressive breast cancer. *Proteomics* 1: 756–762.

134. Raymond A. Dwek. 1996. Glycobiology: Toward Understanding the Function of Sugars. *Chemical Reviews* 96, 2: 683–720.

135. Lynn A. Echan, Hsin-Yao Tang, Nadeem Ali-Khan, KiBeom Lee, and David W. Speicher. 2005. Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and plasma. *Proteomics* 5, 13: 3292–3303.

136. Martin van Eijk, Michael J. Rynkiewicz, Mitchell R. White, et al. 2012. A Unique Sugar-binding Site Mediates the Distinct Anti-influenza Activity of Pig Surfactant Protein D. *Journal of Biological Chemistry* 287, 32: 26666–26677.

137. Martin van Eijk, Mitchell R. White, Erika C. Crouch, et al. 2003. Porcine pulmonary collectins show distinct interactions with influenza A viruses: role of the N-linked oligosaccharides in the carbohydrate recognition domain. *Journal of Immunology (Baltimore, Md.: 1950)* 171, 3: 1431–1440.

138. Anisa Elhamili, Magnus Wetterhall, Björn Arvidsson, Roberto Sebastiano, Pier Giorgio Righetti, and Jonas Bergquist. 2008. Rapid capillary electrophoresis time-of-flight mass spectrometry separations of peptides and proteins using a

monoquaternarized piperazine compound (M7C4I) for capillary coatings. *Electrophoresis* 29, 8: 1619–1625.

139. Joshua E. Elias and Steven P. Gygi. 2010. Target-decoy search strategy for mass spectrometry-based proteomics. *Methods in Molecular Biology (Clifton, N.J.)* 604: 55–71.

140. Jimmy K. Eng, Tahmina A. Jahan, and Michael R. Hoopmann. 2013. Comet: an open-source MS/MS sequence database search tool. *Proteomics* 13, 1: 22–24.

141. Jimmy K. Eng, Ashley L. McCormack, and John R. Yates. 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry* 5, 11: 976–989.

142. Werner Ens and Kenneth G. Standing. 2005. Hybrid quadrupole/time-of-flight mass spectrometers for analysis of biomolecules. *Methods in Enzymology* 402: 49–78.

143. Ruby P. Estrella, John M. Whitelock, Nicolle H. Packer, and Niclas G. Karlsson. 2007. Graphitized Carbon LC−MS Characterization of the Chondroitin Sulfate Oligosaccharides of Aggrecan. *Analytical Chemistry* 79, 10: 3597–3606.

144. Narayanan Eswar, Ben Webb, Marc A. Marti-Renom, et al. 2007. Comparative protein structure modeling using MODELLER. *Current Protocols in Protein Science* Chapter 2: Unit 2.9.

145. Arun V. Everest-Dass, Jodie L. Abrahams, Daniel Kolarich, Nicolle H. Packer, and Matthew P. Campbell. 2013. Structural Feature Ions for Distinguishing N- and O-Linked Glycan Isomers by LC-ESI-IT MS/MS. *Journal of The American Society for Mass Spectrometry* 24, 6: 895–906.

146. J. Q. Fan, A. Kondo, I. Kato, and Y. C. Lee. 1994. High-Performance Liquid Chromatography of Glycopeptides and Oligosaccharides on Graphitized Carbon Columns. *Analytical Biochemistry* 219, 2: 224–229.

147. Liling Fang, Rong Zhang, Evan R. Williams, and Richard N. Zare. 1994. Online Time-of-Flight Mass Spectrometric Analysis of Peptides Separated by Capillary Electrophoresis. *Analytical Chemistry* 66, 21: 3696–3701.

148. J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, and C. M. Whitehouse. 1989. Electrospray ionization for mass spectrometry of large biomolecules. *Science (New York, N.Y.)* 246, 4926: 64–71.

149. Larissa S. Fenn, Michal Kliman, Ablatt Mahsut, Sophie R. Zhao, and John A. McLean. 2009. Characterizing ion mobility-mass spectrometry conformation space

for the analysis of complex biological samples. *Analytical and Bioanalytical Chemistry* 394, 1: 235–244.

150. Larissa S. Fenn and John A. McLean. 2008. Biomolecular structural separations by ion mobility-mass spectrometry. *Analytical and Bioanalytical Chemistry* 391, 3: 905–909.

151. Larissa S. Fenn and John A. McLean. 2009. Simultaneous glycoproteomics on the basis of structure using ion mobility-mass spectrometry. *Molecular BioSystems* 5, 11: 1298.

152. F. A. Fernandez-Lima, D. A. Kaplan, and M. A. Park. 2011. Note: Integration of trapped ion mobility spectrometry with mass spectrometry. *The Review of Scientific Instruments* 82, 12: 126106.

153. Francisco Fernandez-Lima, Desmond A. Kaplan, J. Suetering, and Melvin A. Park. 2011. Gas-phase separation using a trapped ion mobility spectrometer. *International Journal for Ion Mobility Spectrometry* 14, 2–3: 93–98.

154. András Fiser and Andrej Šali. 2003. Modeller: Generation and Refinement of Homology-Based Protein Structure Models. In B.-M. in Enzymology, ed., Academic Press, 461–491.

155. Frantisek Foret, Haihong Zhou, Eric Gangl, and Barry L. Karger. 2000. Subatmospheric electrospray interface for coupling of microcolumn separations with mass spectrometry. *Electrophoresis* 21, 7: 1363–1371.

156. Luca Fornelli, Daniel Ayoub, Konstantin Aizikov, Alain Beck, and Yury O. Tsybin. 2014. Middle-Down Analysis of Monoclonal Antibodies with Electron Transfer Dissociation Orbitrap Fourier Transform Mass Spectrometry. *Analytical Chemistry* 86, 6: 3005–3012.

157. Ari Frank and Pavel Pevzner. 2005. PepNovo: De Novo Peptide Sequencing via Probabilistic Network Modeling. *Analytical Chemistry* 77, 4: 964–973.

158. H. H. Freeze. 2001. Update and perspectives on congenital disorders of glycosylation. *Glycobiology* 11: 129r–143r.

159. H. H. Freeze. 2006. Genetic defects in the human glycome. *Nature Reviews Genetics* 7: 537–551.

160. Hudson H. Freeze. 2001. Lectin Affinity Chromatography. In *Current Protocols in Protein Science*. John Wiley & Sons, Inc.

161. Christian K. Frese, Houjiang Zhou, Thomas Taus, et al. 2013. Unambiguous Phosphosite Localization using Electron-Transfer/Higher-Energy Collision Dissociation (EThcD). *Journal of Proteome Research* 12, 3: 1520–1525.

162. Minoru Fukuda. 2001. Chemical Labeling of Carbohydrates by Oxidation and Sodium Borohydride Reduction. In *Current Protocols in Molecular Biology*. John Wiley & Sons, Inc.

163. Hans-J. Gabius, Klaus Kayser, Sabine André, and Sigrun Gabius. 1996. Glycobiology of Host Defense Mechanisms. In H.-J. Gabius and S. Gabius, eds., *Glycosciences*. Wiley-VCH Verlag GmbH, 497–505.

164. Tamar Geiger, Juergen Cox, and Matthias Mann. 2010. Proteomics on an Orbitrap benchtop mass spectrometer using all-ion fragmentation. *Molecular & cellular proteomics: MCP* 9, 10: 2252–2261.

165. Ming Geng, Xiang Zhang, Minou Bina, and Fred Regnier. 2001. Proteomics of glycoproteins based on affinity selection of glycopeptides from tryptic digests. *Journal of Chromatography B: Biomedical Sciences and Applications* 752, 2: 293–306.

166. Lynn A. Gennaro, Jeannine Delaney, Paul Vouros, David J. Harvey, and Bruno Domon. 2002. Capillary electrophoresis/electrospray ion trap mass spectrometry for the analysis of negatively charged derivatized and underivatized glycans. *Rapid communications in mass spectrometry: RCM* 16, 3: 192–200.

167. Lynn A. Gennaro and Oscar Salas-Solano. 2008. On-Line CE−LIF−MS Technology for the Direct Characterization of N-Linked Glycans from Therapeutic Antibodies. *Analytical Chemistry* 80, 10: 3838–3845.

168. Geun-Cheol Gil, Bryce Iliff, Ron Cerny, William H. Velander, and Kevin E. Van Cott. 2010. High Throughput Quantification of N-Glycans Using One-Pot Sialic Acid Modification and Matrix Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry. *Analytical Chemistry* 82, 15: 6613–6620.

169. Kevin Giles, Steven D. Pringle, Kenneth R. Worthington, David Little, Jason L. Wildgoose, and Robert H. Bateman. 2004. Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Communications in Mass Spectrometry* 18, 20: 2401–2414.

170. Vanessa Leah Gill, Udayanath Aich, Srinivasa Rao, Chris Pohl, and Joseph Zaia. 2013. Disaccharide Analysis of Glycosaminoglycans Using Hydrophilic Interaction Chromatography and Mass Spectrometry. *Analytical Chemistry* 85, 2: 1138–1145.

171. K. J. Gillig, B. Ruotolo, E. G. Stone, et al. 2000. Coupling high-pressure MALDI with ion mobility/orthogonal time-of flight mass spectrometry. *Analytical Chemistry* 72: 3965–3971.

172. Kent J. Gillig, Brandon T. Ruotolo, Earle G. Stone, and David H. Russell. 2004. An electrostatic focusing ion guide for ion mobility-mass spectrometry. *International Journal of Mass Spectrometry* 239, 1: 43–49.

173. Estela Giménez, Raquel Ramos-Hernan, Fernando Benavente, José Barbosa, and Victoria Sanz-Nebot. 2011. Capillary electrophoresis time-of-flight mass spectrometry for a confident elucidation of a glycopeptide map of recombinant human erythropoietin. *Rapid Communications in Mass Spectrometry* 25, 16: 2307–2316.

174. Laurel Glaser, James Stevens, Dmitriy Zamarin, et al. 2005. A Single Amino Acid Substitution in 1918 Influenza Virus Hemagglutinin Changes Receptor Binding Specificity. *Journal of Virology* 79, 17: 11533–11536.

175. Rebecca S. Glaskin. 2013. New techniques for high mobility resolution and ion dynamics in a circular drift tube. Retrieved February 12, 2017 from http://gradworks.umi.com/36/01/3601799.html.

176. Rebecca S. Glaskin, Michael A. Ewing, and David E. Clemmer. 2013. Ion Trapping for Ion Mobility Spectrometry Measurements in a Cyclical Drift Tube. *Analytical Chemistry* 85, 15: 7003–7008.

177. Rebecca S. Glaskin, Kshitij Khatri, Qi Wang, Joseph Zaia, and Catherine E Costello. 2017. Construction of a Database of Collision Cross Section Values for Glycopeptides, Glycans, and Peptides Determined by IM-MS. *Analytical Chemistry*.

178. Rebecca S. Glaskin, Kenneth Newton, Ruwan T. Kurulugama, et al. 2016. Ion Mobility Quadrupole Time-Of-Flight Mass Spectrometer Modified For Electron Capture Dissociation Of Glycans, Glycoconjugates, Peptides, And Proteins. *Proceedings of the 64th ASMS Conference on Mass Spectrometry and Allied Topics*.

179. Boon Chong Goh, Michael J. Rynkiewicz, Tanya R. Cafarella, et al. 2013. Molecular Mechanisms of Inhibition of Influenza by Surfactant Protein D Revealed by Large-Scale Molecular Dynamics Simulation. *Biochemistry* 52, 47: 8527–8538.

180. David Goldberg, Marshall Bern, Simon Parry, et al. 2007. Automated N-Glycopeptide Identification Using a Combination of Single- and Tandem-MS. *Journal of Proteome Research* 6, 10: 3995–4005.

181. Dennis Goldfarb, Wei Wang, and Michael B. Major. 2016. MSAcquisitionSimulator: data-dependent acquisition simulator for LC-MS shotgun proteomics. *Bioinformatics (Oxford, England)* 32, 8: 1269–1271.

182. Anton A. Goloborodko, Lev I. Levitsky, Mark V. Ivanov, and Mikhail V. Gorshkov. 2013. Pyteomics--a Python framework for exploratory data analysis and rapid software prototyping in proteomics. *Journal of the American Society for Mass Spectrometry* 24, 2: 301–304.

183. Alvin Gomez, Jana Nekvindova, Sandra Travica, et al. 2010. Colorectal Cancer-Specific Cytochrome P450 2W1: Intracellular Localization, Glycosylation, and Catalytic Activity. *Molecular Pharmacology* 78, 6: 1004–1011.

184. Bing Gong, Erik Hoyt, Heather Lynaugh, et al. 2013. N-Glycosylamine-mediated isotope labeling for mass spectrometry-based quantitative analysis of N-linked glycans. *Analytical and Bioanalytical Chemistry* 405, 17: 5825–5831.

185. E. Gorelik, U. Galili, and A. Raz. 2001. On the role of cell surface carbohydrates and their binding proteins (lectins) in tumor metastasis. *Cancer and Metastasis Reviews* 20: 245–277.

186. Andreas W. Götz, Mark J. Williamson, Dong Xu, Duncan Poole, Scott Le Grand, and Ross C. Walker. 2012. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *Journal of Chemical Theory and Computation* 8, 5: 1542–1555.

187. I Graham and J Williams. 1975. A comparsion of glycopeptides from the transferrins of several species. *Biochemical Journal* 145, 2: 263–279.

188. E. Gray and B. Mulloy. 2009. Biosimilar low molecular weight heparin products. *Journal of thrombosis and haemostasis : JTH* 7: 1218–21.

189. M. Guerrini, D. Beccati, Z. Shriver, et al. 2008. Oversulfated chondroitin sulfate is a contaminant in heparin associated with adverse clinical events. *Nat Biotechnol* 26: 669–75.

190. M. Guerrini, Z. Shriver, A. Bisio, et al. 2009. The tainted heparin story: an update. *Thrombosis and haemostasis* 102: 907–11.

191. Roger Guevremont. 2004. High-field asymmetric waveform ion mobility spectrometry: A new tool for mass spectrometry. *Journal of Chromatography A* 1058, 1–2: 3–19.

192. Roger Guevremont and Randy W Purves. 1999. High field asymmetric waveform ion mobility spectrometry–mass spectrometry: an investigation of leucine

enkephalin ions produced by electrospray ionization. *Journal of the American Society for Mass Spectrometry* 10, 6: 492–501.

193. Michael Guilhaus. 1995. Special feature: Tutorial. Principles and instrumentation in time-of-flight mass spectrometry. Physical and instrumental concepts. *Journal of Mass Spectrometry* 30, 11: 1519–1532.

194. Shelly Gulati, Yi Lasanajak, David F Smith, Richard D Cummings, and Gillian M Air. 2014. Glycan array analysis of influenza H1N1 binding and release. *Cancer biomarkers: section A of Disease markers* 14, 1: 43–53.

195. Harsha P. Gunawardena, Min He, Paul A. Chrisman, et al. 2005. Electron transfer versus proton transfer in gas-phase ion/ion reactions of polyprotonated peptides. *Journal of the American Chemical Society* 127, 36: 12627–12639.

196. Y. C. Guo and H. E. Conrad. 1989. The disaccharide composition of heparins and heparan sulfates. *Anal Biochem* 176: 96–104.

197. Nitin Gupta, Nuno Bandeira, Uri Keich, and Pavel A. Pevzner. 2011. Target-decoy approach and false discovery rate: when things may go wrong. *Journal of the American Society for Mass Spectrometry* 22, 7: 1111–1120.

198. R. Gupta, E. Jung, and S. Brunak. Prediction of N-glycosylation sites in human proteins. *In preparation, 2004.*

199. A. Guttman, F. T. Chen, R. A. Evangelista, and N. Cooke. 1996. High-resolution capillary gel electrophoresis of reducing oligosaccharides labeled with 1-aminopyrene-3,6,8-trisulfonate. *Analytical Biochemistry* 233, 2: 234–242.

200. A. Guttman and T. Pritchett. 1995. Capillary gel electrophoresis separation of high-mannose type oligosaccharides derivatized by 1-aminopyrene-3,6,8-trisulfonic acid. *Electrophoresis* 16, 10: 1906–1911.

201. András Guttman. 1997. Analysis of monosaccharide composition by capillary electrophoresis. *Journal of Chromatography A* 763, 1–2: 271–277.

202. Noortje de Haan, Karli R. Reiding, Markus Haberger, Dietmar Reusch, David Falck, and Manfred Wuhrer. 2015. Linkage-Specific Sialic Acid Derivatization for MALDI-TOF-MS Profiling of IgG Glycopeptides. *Analytical Chemistry* 87, 16: 8284–8291.

203. Noortje de Haan, KarliR. Reiding, and Manfred Wuhrer. 2017. Sialic Acid Derivatization for the Rapid Subclass- and Sialic Acid Linkage-Specific MALDI-TOF-MS Analysis of IgG Fc-Glycopeptides. In G. Lauc and M. Wuhrer, eds., *High-Throughput Glycomics and Glycoproteomics*. Springer New York, 49–62.

204. James W. Hager. 2002. A new linear ion trap mass spectrometer. *Rapid Communications in Mass Spectrometry* 16, 6: 512–526.

205. Per Hägglund, Jakob Bunkenborg, Felix Elortza, Ole Nørregaard Jensen, and Peter Roepstorff. 2004. A New Strategy for Identification of N-Glycosylated Proteins and Unambiguous Assignment of Their Glycosylation Sites Using HILIC Enrichment and Partial Deglycosylation. *Journal of Proteome Research* 3, 3: 556–566.

206. Hannes Hahne, Patrick Neubert, Karsten Kuhn, et al. 2012. Carbonyl-reactive tandem mass tags for the proteome-wide quantification of N-linked glycans. *Analytical Chemistry* 84, 8: 3716–3724.

207. R. S. Haltiwanger and J. B. Lowe. 2004. Role of glycosylation indevelopment. *Annual Review of Biochemistry* 73: 491–537.

208. Liang Han and Catherine Costello. 2011. Electron Transfer Dissociation of Milk Oligosaccharides. *Journal of The American Society for Mass Spectrometry* 22, 6: 997–1013.

209. Liang Han and Catherine E. Costello. 2013. Mass spectrometry of glycans. *Biochemistry (Moscow)* 78, 7: 710–720.

210. Ivan Hang, Chia-wei Lin, Oliver C. Grant, et al. 2015. Analysis of site-specific N-glycan remodeling in the endoplasmic reticulum and the Golgi. *Glycobiology* 25, 12: 1335–1349.

211. P. L. Hao, Y. Ren, and Y. M. Xie. 2010. An Improved Protocol for N-Glycosylation Analysis of Gel-Separated Sialylated Glycoproteins by MALDI-TOF/TOF. *Plos One* 5.

212. G. W. Hart, R. S. Haltiwanger, G. D. Holt, and W. G. Kelly. 1989. Glycosylation in the Nucleus and Cytoplasm. *Annual Review of Biochemistry* 58: 841–874.

213. C A Hartley, D C Jackson, and E M Anders. 1992. Two distinct serum mannose-binding lectins function as beta inhibitors of influenza virus: identification of bovine serum beta inhibitor as conglutinin. *Journal of Virology* 66, 7: 4358–4363.

214. C. A. Hartley, P. C. Reading, A. C. Ward, and E. M. Anders. 1997. Changes in the hemagglutinin molecule of influenza type A (H3N2) virus associated with increased virulence for mice. *Archives of Virology* 142, 1: 75–88.

215. K L Hartshorn, E C Crouch, M R White, et al. 1994. Evidence for a protective role of pulmonary surfactant protein D (SP-D) against influenza A viruses. *Journal of Clinical Investigation* 94, 1: 311–319.

216. K. L. Hartshorn, M. R. White, V. Shepherd, K. Reid, J. C. Jensenius, and E. C. Crouch. 1997. Mechanisms of anti-influenza activity of surfactant proteins A and D: comparison with serum collectins. *The American Journal of Physiology* 273, 6 Pt 1: L1156-1166.

217. K. L. Hartshorn, M. R. White, K. Smith, et al. 2010. Increasing antiviral activity of surfactant protein d trimers by introducing residues from bovine serum collectins: dissociation of mannan-binding and antiviral activity. *Scandinavian Journal of Immunology* 72, 1: 22–30.

218. Kevan L Hartshorn, Richard Webby, Mitchell R White, et al. 2008. Role of viral hemagglutinin glycosylation in anti-influenza activities of recombinant surfactant protein D. *Respiratory Research* 9, 1: 65.

219. D. J. Harvey. 2000. Electrospray mass spectrometry and fragmentation of N-linked carbohydrates derivatized at the reducing terminus. *Journal of the American Society for Mass Spectrometry* 11, 10: 900–915.

220. David J Harvey. 2001. Ionization and collision-induced fragmentation of N-linked and related carbohydrates using divalent cations1. *Journal of the American Society for Mass Spectrometry* 12, 8: 926–937.

221. David J Harvey. 2003. Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates and glycoconjugates. *International Journal of Mass Spectrometry* 226, 1: 1–35.

222. David J. Harvey, Taj S. Mattu, Mark R. Wormald, Louise Royle, Raymond A. Dwek, and Pauline M. Rudd. 2002. "Internal Residue Loss": Rearrangements Occurring during the Fragmentation of Carbohydrates Derivatized at the Reducing Terminus. *Analytical Chemistry* 74, 4: 734–740.

223. John R. Haserick, Deborah R. Leon, John Samuelson, and Catherine E. Costello. 2017. Asparagine-linked Glycans of Cryptosporidium parvum Contain a Single Long Arm, Are Barely Processed in the ER or Golgi, and Show a Strong Bias for Sites with Threonine. *Molecular & cellular proteomics: MCP*.

224. Samuel Hawgood, Cynthia Brown, Jess Edmondson, et al. 2004. Pulmonary collectins modulate strain-specific influenza a virus infection and host responses. *Journal of Virology* 78, 16: 8565–8572.

225. Jie-Long He, Yi-Chung Chiu, Shih-Chung Chang, Ching-Ho Wang, and Rong-Huay Juang. 2015. Glycosylation at hemagglutinin Asn-167 protects the H6N1 avian influenza virus from tryptic cleavage at Arg-201 and maintains the viral infectivity. *Virus Research* 197: 101–107.

226. Ari Helenius and Markus Aebi. 2004. Roles of N-linked glycans in the endoplasmic reticulum. *Annual Review of Biochemistry* 73: 1019–1049.

227. C. G. Hellerqvist. 1990. Linkage analysis using Lindberg method. *Methods in Enzymology* 193: 554–573.

228. Jens Henriksen, Lene Hoffmeyer Ringborg, and Peter Roepstorrf. 2004. On-line size-exclusion chromatography/mass spectrometry of low molecular mass heparin. *Journal of Mass Spectrometry* 39, 11: 1305–1312.

229. Diana Rosa Hernandez, John Daniel Debord, Mark E. Ridgeway, Desmond A. Kaplan, Melvin A. Park, and Francisco Fernandez-Lima. 2014. Ion dynamics in a trapped ion mobility spectrometer. *The Analyst* 139, 8: 1913–1921.

230. Franz Hillenkamp, Michael Karas, Ronald C. Beavis, and Brian T. Chait. 1991. Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Biopolymers. *Analytical Chemistry* 63, 24: 1193A–1203A.

231. H. Hinneburg, J. Hofmann, W. B. Struwe, et al. 2016. Distinguishing N - acetylneuraminic acid linkage isomers on glycopeptides by ion mobility-mass spectrometry. *Chemical Communications* 52, 23: 4381–4384.

232. J. A. Hipple, H. Sommer, and H. A. Thomas. 1949. A Precise Method of Determining the Faraday by Magnetic Resonance. *Physical Review* 76, 12: 1877–1878.

233. C. S. Hoaglund, S. J. Valentine, C. R. Sporleder, J. P. Reilly, and D. E. Clemmer. 1998. Three-dimensional ion mobility TOFMS analysis of electrosprayed biomolecules. *Analytical Chemistry* 70: 2236–2242.

234. Edmond de Hoffmann. 2000. Mass Spectrometry. In *Kirk-Othmer Encyclopedia of Chemical Technology*. John Wiley & Sons, Inc.

235. Edmond de Hoffmann and Vincent Stroobant. *Mass Spectrometry: Principles and Applications, 3rd Edition*. Wiley.

236. J. Hofmann, H. S. Hahm, P. H. Seeberger, and K. Pagel. 2015. Identification of carbohydrate anomers using ion mobility-mass spectrometry. *Nature* 526, 7572: 241–244.

237. Johanna Hofmann, Weston B. Struwe, Charlotte A. Scarff, James H. Scrivens, David J. Harvey, and Kevin Pagel. 2014. Estimating Collision Cross Sections of Negatively Charged N-Glycans using Traveling Wave Ion Mobility-Mass Spectrometry. *Analytical Chemistry* 86, 21: 10789–10795.

238. Gretchen E. Hofmeister, Zhongrui Zhou, and Julie A. Leary. 1991. Linkage position determination in lithium-cationized disaccharides: tandem mass spectrometry and semiempirical calculations. *Journal of the American Chemical Society* 113, 16: 5964–5970.

239. Dustin D. Holden, William M. McGee, and Jennifer S. Brodbelt. 2016. Integration of Ultraviolet Photodissociation with Proton Transfer Reactions and Ion Parking for Analysis of Intact Proteins. *Analytical Chemistry* 88, 1: 1008–1016.

240. Stephanie Holst, Bram Heijs, Noortje de Haan, et al. 2016. Linkage-Specific in Situ Sialic Acid Derivatization for N-Glycan Mass Spectrometry Imaging of Formalin-Fixed Paraffin-Embedded Tissues. *Analytical Chemistry* 88, 11: 5904–5913.

241. Qiuting Hong, Carlito B. Lebrilla, Suzanne Miyamoto, and L. Renee Ruhaak. 2013. Absolute Quantitation of Immunoglobulin G and Its Glycoforms Using Multiple Reaction Monitoring. *Analytical Chemistry* 85, 18: 8585–8593.

242. Qiuting Hong, L. Renee Ruhaak, Carol Stroble, et al. 2015. A Method for Comprehensive Glycosite-Mapping and Direct Quantitation of Serum Glycoproteins. *Journal of Proteome Research* 14, 12: 5179–5192.

243. H. J. Hoppe, P. N. Barlow, and K. B. Reid. 1994. A parallel three stranded alpha-helical bundle at the nucleation site of collagen triple-helix formation. *FEBS letters* 344, 2–3: 191–195.

244. David M. Horn, Roman A. Zubarev, and Fred W. McLafferty. 2000. Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *Journal of the American Society for Mass Spectrometry* 11, 4: 320–332.

245. Glen L. Hortin and Beverly L. Trimpe. 1990. Lectin affinity chromatography of proteins bearing O-linked oligosaccharides: Application of jacalin-agarose. *Analytical Biochemistry* 188, 2: 271–277.

246. Han Hu, Piotr Dittwald, Joseph Zaia, and Dirk Valkenborg. 2013. Comment on "Computation of Isotopic Peak Center-Mass Distribution by Fourier Transform." *Analytical Chemistry* 85, 24: 12189–12192.

247. Han Hu, Yu Huang, Yang Mao, et al. 2014. A Computational Framework for Heparan Sulfate Sequencing Using High-resolution Tandem Mass Spectra. *Molecular & Cellular Proteomics* 13, 9: 2490–2502.

248. Han Hu, Kshitij Khatri, Joshua Klein, Nancy Leymarie, and Joseph Zaia. 2015. A review of methods for interpretation of glycopeptide tandem mass spectral data. *Glycoconjugate Journal*: 1–12.

249. Han Hu, Kshitij Khatri, and Joseph Zaia. 2016. Algorithms and design strategies towards automated glycoproteomics analysis. *Mass Spectrometry Reviews*: n/a-n/a.

250. Qizhi Hu, Robert J. Noll, Hongyan Li, Alexander Makarov, Mark Hardman, and R. Graham Cooks. 2005. The Orbitrap: a new mass spectrometer. *Journal of mass spectrometry: JMS* 40, 4: 430–443.

251. Yunli Hu and Yehia Mechref. 2012. Comparing MALDI-MS, RP-LC-MALDI-MS and RP-LC-ESI-MS glycomic profiles of permethylated N-glycans derived from model glycoproteins and human blood serum. *Electrophoresis* 33, 12: 1768–1777.

252. Yunli Hu, Tarek Shihab, Shiyue Zhou, Kerry Wooding, and Yehia Mechref. 2016. LC–MS/MS of permethylated N-glycans derived from model and human blood serum glycoproteins. *Electrophoresis* 37, 11: 1498–1505.

253. Lei Huang, Gulia Harvie, Jerald S. Feitelson, et al. 2005. Immunoaffinity separation of plasma proteins by IgY microbeads: Meeting the needs of proteomic sample preparation and analysis. *Proteomics* 5, 13: 3314–3328.

254. Yining Huang, Yongxin Nie, Barry Boyes, and Ron Orlando. 2016. Resolving Isomeric Glycopeptide Glycoforms with Hydrophilic Interaction Chromatography (HILIC). *Journal of Biomolecular Techniques : JBT* 27, 3: 98–104.

255. Yiqun Huang, Yi Pu, Xiang Yu, Catherine E. Costello, and Cheng Lin. 2014. Mechanistic study on electron capture dissociation of the oligosaccharide-Mg^{2+} complex. *Journal of the American Society for Mass Spectrometry* 25, 8: 1451–1460.

256. Yu Huang, Xiaofeng Shi, Xiang Yu, et al. 2011. Improved Liquid Chromatography-MS/MS of Heparan Sulfate Oligosaccharides via Chip-Based Pulsed Makeup Flow. *Analytical Chemistry* 83, 21: 8222–8229.

257. Yu Huang, Xiang Yu, Yang Mao, Catherine E. Costello, Joseph Zaia, and Cheng Lin. 2013. De Novo Sequencing of Heparan Sulfate Oligosaccharides by Electron-Activated Dissociation. *Analytical Chemistry* 85, 24: 11979–11986.

258. Yunping Huang, Yehia Mechref, and Milos V. Novotny. 2001. Microscale Nonreductive Release of O-Linked Glycans for Subsequent Analysis through MALDI Mass Spectrometry and Capillary Electrophoresis. *Analytical Chemistry* 73, 24: 6063–6069.

259. Matthew M. Hufford, Graham Richardson, Haixia Zhou, et al. 2012. Influenza-Infected Neutrophils within the Infected Lungs Act as Antigen Presenting Cells for Anti-Viral CD8+ T Cells. *PLoS ONE* 7, 10: e46581.

260. C. Huhn, R. Ramautar, M. Wuhrer, and G. W. Somsen. 2010. Relevance and use of capillary coatings in capillary electrophoresis-mass spectrometry. *Analytical and Bioanalytical Chemistry* 396, 1: 297–314.

261. Donald F. Hunt, Alexander M. Buko, John M. Ballard, Jeffrey Shabanowitz, and Anne B. Giordani. 1981. Sequence analysis of polypeptides by collision activated dissociation on a triple quadrupole mass spectrometer. *Biological Mass Spectrometry* 8, 9: 397–408.

262. Edward C. Hutchinson, Philip D. Charles, Svenja S. Hester, et al. 2014. Conserved and host-specific features of influenza virion architecture. *Nature Communications* 5: 4816.

263. Julia Hütter, Jana V. Rödig, Dirk Höper, et al. 2013. Toward Animal Cell Culture–Based Influenza Vaccine Design: Viral Hemagglutinin N-Glycosylation Markedly Impacts Immunogenicity. *The Journal of Immunology* 190, 1: 220–230.

264. A. C. Ivens, C. S. Peacock, E. A. Worthey, et al. 2005. The genome of the kinetoplastid parasite, Leishmania major. *Science* 309: 436–42.

265. Vera B. Ivleva, Yuri N. Elkin, Bogdan A. Budnik, Susanne C. Moyer, Peter B. O'Connor, and Catherine E. Costello. 2004. Coupling Thin-Layer Chromatography with Vibrational Cooling Matrix-Assisted Laser Desorption/Ionization Fourier Transform Mass Spectrometry for the Analysis of Ganglioside Mixtures. *Analytical Chemistry* 76, 21: 6484–6491.

266. Navdeep Jaitly, Anoop Mayampurath, Kyle Littlefield, Joshua N Adkins, Gordon A Anderson, and Richard D Smith. 2009. Decon2LS: An open-source software package for automated processing and visualization of high resolution mass spectrometry data. *BMC Bioinformatics* 10, 1: 87.

267. Claude A. Jakob, Patricie Burda, Jürgen Roth, and Markus Aebi. 1998. Degradation of Misfolded Endoplasmic Reticulum Glycoproteins in Saccharomyces cerevisiae Is Determined by a Specific Oligosaccharide Structure. *The Journal of Cell Biology* 142, 5: 1223–1233.

268. Ewa Jankowska and John Cipollo. 2011. Platform for analysis of anthranilic acid *N* -glycan derivatives utilizing multipolarity mode LC–MS with hydrophilic interaction chromatography separation and ion trap MS/MS. *Bioanalysis* 3, 21: 2401–2417.

269. Bas C. Jansen, Karli R. Reiding, Albert Bondt, et al. 2015. MassyTools: A High-Throughput Targeted Data Processing Tool for Relative Quantitation and Quality

Control Developed for Glycomic and Glycoproteomic MALDI-MS. *Journal of Proteome Research* 14, 12: 5088–5098.

270. Akila Jayaraman, Xiaoying Koh, Jing Li, et al. 2012. Glycosylation at Asn ^{91} of H1N1 haemagglutinin affects binding to glycan receptors. *Biochemical Journal* 444, 3: 429–435.

271. Pia H. Jensen, Niclas G. Karlsson, Daniel Kolarich, and Nicolle H. Packer. 2012. Structural analysis of N- and O-glycans released from glycoproteins. *Nature Protocols* 7, 7: 1299–1310.

272. Jacquelyn R. Jhingree, Rebecca Beveridge, Eleanor R. Dickinson, et al. Electron transfer with no dissociation ion mobility–mass spectrometry (ETnoD IM-MS). The effect of charge reduction on protein conformation. *International Journal of Mass Spectrometry*.

273. Jenny Jiao, Hailong Zhang, and Vernon N. Reinhold. 2011. High Performance IT-MS Sequencing of Glycans (Spatial Resolution of Ovalbumin Isomers). *International Journal of Mass Spectrometry* 303, 2–3: 109–117.

274. Emma R. Job, Barbara Bottazzi, Kirsty R. Short, et al. 2014. A Single Amino Acid Substitution in the Hemagglutinin of H3N2 Subtype Influenza A Viruses Is Associated with Resistance to the Long Pentraxin PTX3 and Enhanced Virulence in Mice. *The Journal of Immunology* 192, 1: 271–281.

275. Emma R. Job, Yi-Mo Deng, Michelle D. Tate, et al. 2010. Pandemic H1N1 Influenza A Viruses Are Resistant to the Antiviral Activities of Innate Immune Proteins of the Collectin and Pentraxin Superfamilies. *The Journal of Immunology* 185, 7: 4284–4291.

276. Sakari Joenväärä, Ilja Ritamo, Hannu Peltoniemi, and Risto Renkonen. 2008. N-Glycoproteomics – An automated workflow approach. *Glycobiology* 18, 4: 339–349.

277. Darryl Johnson, Barry Boyes, Taylor Fields, Rachel Kopkin, and Ron Orlando. 2013. Optimization of Data-Dependent Acquisition Parameters for Coupling High-Speed Separations with LC-MS/MS for Protein Identifications. *Journal of Biomolecular Techniques : JBT* 24, 2: 62–72.

278. Andrew R. Jones, Martin Eisenacher, Gerhard Mayer, et al. 2012. The mzIdentML Data Standard for Mass Spectrometry-Based Proteomics Results. *Molecular & Cellular Proteomics : MCP* 11, 7.

279. Muchena J. Kailemia, L. Renee Ruhaak, Carlito B. Lebrilla, and I. Jonathan Amster. 2014. Oligosaccharide Analysis by Mass Spectrometry: A Review of Recent Developments. *Analytical Chemistry* 86, 1: 196–212.

280. Hiroyuki Kaji, Haruna Saito, Yoshio Yamauchi, et al. 2003. Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nature Biotechnology* 21, 6: 667–672.

281. Lukas Käll, John D. Storey, Michael J. MacCoss, and William Stafford Noble. 2008. Assigning Significance to Peptides Identified by Tandem Mass Spectrometry Using Decoy Databases. *Journal of Proteome Research* 7, 1: 29–34.

282. Pilsoo Kang, Yehia Mechref, Iveta Klouckova, and Milos V. Novotny. 2005. Solid-phase permethylation of glycans for mass spectrometric analysis. *Rapid communications in mass spectrometry : RCM* 19, 23: 3421–3428.

283. Pilsoo Kang, Yehia Mechref, Zuzana Kyselova, John A. Goetz, and Milos V. Novotny. 2007. Comparative Glycomic Mapping through Quantitative Permethylation and Stable-Isotope Labeling. *Analytical Chemistry* 79, 16: 6064–6073.

284. Pilsoo Kang, Yehia Mechref, and Milos V. Novotny. 2008. High-throughput solid-phase permethylation of glycans prior to mass spectrometry. *Rapid communications in mass spectrometry: RCM* 22, 5: 721–734.

285. Abu B. Kanu, Prabha Dwivedi, Maggie Tam, Laura Matz, and Herbert H. Hill. 2008. Ion mobility–mass spectrometry. *Journal of Mass Spectrometry* 43, 1: 1–22.

286. Niclas G. Karlsson, Benjamin L. Schulz, and Nicolle H. Packer. 2004. Structural determination of neutral O-linked oligosaccharide alditols by negative ion LC-electrospray-MSn. *Journal of the American Society for Mass Spectrometry* 15, 5: 659–672.

287. Niclas G. Karlsson, Nicole L. Wilson, Hans-Jürgen Wirth, Peter Dawes, Hiren Joshi, and Nicolle H. Packer. 2004. Negative ion graphitised carbon nano-liquid chromatography/mass spectrometry increases sensitivity for glycoprotein oligosaccharide analysis. *Rapid communications in mass spectrometry: RCM* 18, 19: 2282–2292.

288. Parminder Kaur and Peter B. O'Connor. 2006. Algorithms for Automatic Interpretation of High Resolution Mass Spectra. *Journal of the American Society for Mass Spectrometry* 17, 3: 459–468.

289. J. P. Kerckaert and B. Bayard. 1982. Glycan Uniformity within Molecular Variants of Transferrin with Distinct Affinity for Concanavalin-A. *Biochemical and Biophysical Research Communications* 105: 1023–1030.

290. Darren Kessner, Matt Chambers, Robert Burke, David Agus, and Parag Mallick. 2008. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* 24, 21: 2534–2536.

291. Kshitij Khatri, Joshua A. Klein, Mitchell R. White, et al. 2016. Integrated omics and computational glycobiology reveal structural basis for Influenza A virus glycan microheterogeneity and host interactions. *Molecular & Cellular Proteomics*: mcp.M116.058016.

292. Kshitij Khatri, Joshua A. Klein, and Joseph Zaia. 2017. Use of an informed search space maximizes confidence of site-specific assignment of glycoprotein glycosylation. *Analytical and Bioanalytical Chemistry* 409, 2: 607–618.

293. Kshitij Khatri, Gregory O. Staples, Nancy Leymarie, et al. 2014. Confident Assignment of Site-Specific Glycosylation in Complex Glycoproteins in a Single Step. *Journal of Proteome Research* 13, 10: 4347–4355.

294. Sangtae Kim and Pavel A. Pevzner. 2014. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nature Communications* 5: 5277.

295. K. H. Kingdon. 1923. A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures. *Physical Review* 21, 4: 408–418.

296. S. Kirsch, M. Zarei, M. Cindric, J. Muthing, L. Bindila, and J. Peter-Katalinic. 2008. On-line nano-HPLC/ESI QTOF MS and tandem MS for separation, detection, and structural elucidation of human erythrocytes neutral glycosphingolipid mixture. *Analytical Chemistry* 80: 4711–22.

297. Karl N. Kirschner, Austin B. Yongye, Sarah M. Tschampel, et al. 2008. GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *Journal of Computational Chemistry* 29, 4: 622–655.

298. Frank Kjeldsen, Kim F. Haselmann, Bogdan A. Budnik, Frank Jensen, and Roman A. Zubarev. 2002. Dissociative capture of hot (3–13 eV) electrons by polypeptide polycations: an efficient process accompanied by secondary fragmentation. *Chemical Physics Letters* 356, 3–4: 201–206.

299. Frank Kjeldsen and Roman Zubarev. 2003. Secondary losses via gamma-lactam formation in hot electron capture dissociation: a missing link to complete de novo

sequencing of proteins? *Journal of the American Chemical Society* 125, 22: 6628–6629.

300. Joshua A. Klein, Kshitij Khatri, Luis Carvalho, and Joseph Zaia. 2016. A Toolkit for Interactive and Batch Analysis of Glycomics and Glycoproteomics Mass Spectrometry Data. *Glycobiology*, 1458–1458.

301. Hans-Dieter Klenk, Ralf Wagner, Dagmar Heuer, and Thorsten Wolff. 2002. Importance of hemagglutinin glycosylation for the biological functions of influenza virus. *Virus research* 82, 1–2: 73–75.

302. Giselle M. Knudsen and Robert J. Chalkley. 2011. The Effect of Using an Inappropriate Protein Database for Proteomic Data Analysis. *PLOS ONE* 6, 6: e20873.

303. Byoung Joon Ko and Jennifer S. Brodbelt. 2011. 193 nm ultraviolet photodissociation of deprotonated sialylated oligosaccharides. *Analytical Chemistry* 83, 21: 8192–8200.

304. Felix J. Kohl, Laura Sánchez-Hernández, and Christian Neusüß. 2015. Capillary electrophoresis in two-dimensional separation systems: Techniques and applications. *Electrophoresis* 36, 1: 144–158.

305. Daniel Kolarich, Markus Windwarder, Kathirvel Alagesan, and Friedrich Altmann. 2015. Isomer-Specific Analysis of Released N-Glycans by LC-ESI MS/MS with Porous Graphitized Carbon. *Methods in Molecular Biology (Clifton, N.J.)* 1321: 427–435.

306. J. M. Koomen, B. T. Ruotolo, K. J. Gillig, et al. 2002. Oligonucleotide analysis with MALDI-ion-mobility-TOFMS. *Analytical and Bioanalytical Chemistry* 373: 612–617.

307. S. Kornfeld. 1998. Diseases of abnormal protein glycosylation: An emerging area. *Journal of Clinical Investigation* 101: 1293–1295.

308. C. Koster and A. Holle. 1999. A new intelligent annotation procedure: SNAP. *Proceedings of the 47th ASMS Conference on Mass Spectrometry and Allied Topics*.

309. Yury I. Kostyukevich, Gleb N. Vladimirov, and Eugene N. Nikolaev. 2012. Dynamically Harmonized FT-ICR Cell with Specially Shaped Electrodes for Compensation of Inhomogeneity of the Magnetic Field. Computer Simulations of the Electric Field and Ion Motion Dynamics. *Journal of The American Society for Mass Spectrometry* 23, 12: 2198–2207.

310. Radoslaw P. Kozak, Louise Royle, Richard A. Gardner, Albert Bondt, Daryl L. Fernandes, and Manfred Wuhrer. 2014. Improved nonreductive O-glycan release by hydrazinolysis with ethylenediaminetetraacetic acid addition. *Analytical Biochemistry* 453: 29–37.

311. Radoslaw P. Kozak, Louise Royle, Richard A. Gardner, Daryl L. Fernandes, and Manfred Wuhrer. 2012. Suppression of peeling during the release of O-glycans by hydrazinolysis. *Analytical Biochemistry* 423, 1: 119–128.

312. Balagurunathan Kuberan, Miroslaw Lech, Lijuan Zhang, Zhengliang L. Wu, David L. Beeler, and Robert D. Rosenberg. 2002. Analysis of heparan sulfate oligosaccharides with ion pair-reverse phase capillary high performance liquid chromatography-microelectrospray ionization time-of-flight mass spectrometry. *Journal of the American Chemical Society* 124, 29: 8707–8718.

313. Fotini N Lamari, Reinhard Kuhn, and Nikos K Karamanos. 2003. Derivatization of carbohydrates for chromatographic, electrophoretic and mass spectrometric structure analysis. *Journal of Chromatography B* 793, 1: 15–36.

314. Ünige A. Laskay, Anna A. Lobas, Kristina Srzentić, Mikhail V. Gorshkov, and Yury O. Tsybin. 2013. Proteome Digestion Specificity Analysis for Rational Design of Extended Bottom-up and Middle-down Proteomics Experiments. *Journal of Proteome Research* 12, 12: 5558–5569.

315. Erika Lattova and Hélène Perreault. 2003. Labelling saccharides with phenylhydrazine for electrospray and matrix-assisted laser desorption-ionization mass spectrometry. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences* 793, 1: 167–179.

316. Matthew A. Lauber, Ying-Qing Yu, Darryl W. Brousmiche, et al. 2015. Rapid Preparation of Released N-Glycans for HILIC Analysis Using a Labeling Reagent that Facilitates Sensitive Fluorescence and ESI-MS Detection. *Analytical Chemistry* 87, 10: 5401–5409.

317. Ernest Orlando Lawrence, Niels E. Edlefsen, and Gilbert Newton Lewis. 1930. *On the production of high speed protons*. publisher not identified.

318. Aaron R. Ledvina, Christopher M. Rose, Graeme C. McAlister, et al. 2013. Activated ion ETD performed in a modified collision cell on a hybrid QLT-Oribtrap mass spectrometer. *Journal of the American Society for Mass Spectrometry* 24, 11: 1623–1633.

319. Ju Yeon Lee, Hyun Kyoung Lee, Gun Wook Park, et al. 2016. Characterization of Site-Specific N-Glycopeptide Isoforms of α-1-Acid Glycoprotein from an

Interlaboratory Study Using LC–MS/MS. *Journal of Proteome Research* 15, 12: 4146–4164.

320. Ming Lei, Yehia Mechref, and Milos V. Novotny. 2009. Structural Analysis of Sulfated Glycans by Sequential Double-Permethylation Using Methyl Iodide and Deuteromethyl Iodide. *Journal of the American Society for Mass Spectrometry* 20, 9: 1660–1671.

321. Nancy Leymarie, Paula J. Griffin, Karen Jonscher, et al. 2013. Interlaboratory Study on Differential Analysis of Protein Glycosylation by Mass Spectrometry: the ABRF Glycoprotein Research Multi-Institutional Study 2012. *Molecular & Cellular Proteomics*.

322. Nancy Leymarie, Mark E. McComb, Hicham Naimy, Gregory O. Staples, and Joseph Zaia. 2012. Differential Characterization and Classification of Tissue Specific Glycosaminoglycans by Tandem Mass Spectrometry and Statistical Methods. *International journal of mass spectrometry* 312: 144–154.

323. Nancy Leymarie and Joseph Zaia. 2012. Effective Use of Mass Spectrometry for Glycan and Glycopeptide Structural Analysis. *Analytical Chemistry* 84, 7: 3040–3048.

324. Bensheng Li, Hyun Joo An, Jerry L. Hedrick, and Carlito B. Lebrilla. 2009. Collision-induced dissociation tandem mass spectrometry for structural elucidation of glycans. *Methods in Molecular Biology (Clifton, N.J.)* 534: 133–145.

325. Hongli Li, Brad Bendiak, William F. Siems, David R. Gang, and Herbert H. Hill Jr. 2013. Ion mobility-mass correlation trend line separation of glycoprotein digests without deglycosylation. *International Journal for Ion Mobility Spectrometry* 16, 2: 105–115.

326. Lingyun Li, Fuming Zhang, Joseph Zaia, and Robert J. Linhardt. 2012. Top-Down Approach for the Direct Characterization of Low Molecular Weight Heparins Using LC-FT-MS. *Analytical Chemistry* 84: 8822–8829.

327. Xiaojuan Li, Cheng Lin, Liang Han, Catherine E. Costello, and Peter B. O'Connor. 2010. Charge remote fragmentation in electron capture and electron transfer dissociation. *Journal of the American Society for Mass Spectrometry* 21, 4: 646–656.

328. R. J. Linhardt and N. S. Gunay. 1999. Production and chemical processing of low molecular weight heparins. *Semin Thromb Hemost* 25 Suppl 3: 5–16.

329. H. Lis, N. Sharon, and E. Katchalski. 1969. Identification of the carbohydrate-protein linking group in soybean hemagglutinin. *Biochimica Et Biophysica Acta* 192, 2: 364–366.

330. Adam J. Liska and Andrej Shevchenko. 2003. Expanding the organismal scope of proteomics: Cross-species protein identification by mass spectrometry and its implications. *Proteomics* 3, 1: 19–28.

331. Fan Liu, Bas van Breukelen, and Albert J. R. Heck. 2014. Facilitating Protein Disulfide Mapping by a Combination of Pepsin Digestion, Electron Transfer Higher Energy Dissociation (EThcD), and a Dedicated Search Algorithm SlinkS. *Molecular & Cellular Proteomics* 13, 10: 2776–2786.

332. Haiying Liu, Zhenqing Zhang, and Robert J. Linhardt. 2009. Lessons learned from the contamination of heparin. *Natural Product Reports* 26: 313–321.

333. Hongbin Liu, Rovshan G. Sadygov, and John R. Yates. 2004. A Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics. *Analytical Chemistry* 76, 14: 4193–4201.

334. Xin Liu, Xianyu Li, Kenneth Chan, et al. 2007. "One-Pot" Methylation in Glycomics Application: Esterification of Sialic Acids and Permanent Charge Construction. *Analytical Chemistry* 79, 10: 3894–3900.

335. Yansheng Liu, Jing Chen, Atul Sethi, et al. 2014. Glycoproteomic Analysis of Prostate Cancer Tissues by SWATH Mass Spectrometry Discovers N-acylethanolamine Acid Amidase and Protein Tyrosine Kinase 7 as Signatures for Tumor Aggressiveness. *Molecular & Cellular Proteomics* 13, 7: 1753–1768.

336. Yansheng Liu, Ruth Hüttenhain, Silvia Surinova, et al. 2013. Quantitative measurements of N-linked glycoproteins in human plasma by SWATH-MS. *Proteomics* 13, 8: 1247–1256.

337. Ze Liu, Jing Cao, Yifeng He, et al. 2010. Tandem 18O Stable Isotope Labeling for Quantification of N-Glycoproteome. *Journal of Proteome Research* 9, 1: 227–236.

338. J. B. Lowe. 2003. Glycan-dependent leukocyte adhesion and recruitment in inflammation. *Current Opinion in Cell Biology* 15: 531–538.

339. Yuan Lu, John P. Welsh, and James R. Swartz. 2014. Production and stabilization of the trimeric influenza hemagglutinin stem domain for potentially broadly protective influenza vaccines. *Proceedings of the National Academy of Sciences* 111, 1: 125–130.

340. Marta Łuksza and Michael Lässig. 2014. A predictive fitness model for influenza. *Nature* 507, 7490: 57–61.

341. Tony Ly and Ryan R. Julian. 2009. Ultraviolet Photodissociation: Developments towards Applications for Mass-Spectrometry-Based Proteomics. *Angewandte Chemie International Edition* 48, 39: 7130–7137.

342. Ke-Shiuan Lynn, Chen-Chun Chen, T. Mamie Lih, et al. 2015. MAGIC: An Automated N-Linked Glycoprotein Identification Tool Using a Y1-Ion Pattern Matching Algorithm and in Silico MS2 Approach. *Analytical Chemistry* 87, 4: 2466–2473.

343. Milan Madera, Benjamin Mann, Yehia Mechref, and Milos V. Novotny. 2008. Efficacy of glycoprotein enrichment by microscale lectin affinity chromatography. *Journal of Separation Science* 31, 14: 2722–2732.

344. James A. Madsen, Byoung Joon Ko, Hua Xu, et al. 2013. Concurrent automated sequencing of the glycan and peptide portions of O-linked glycopeptide anions by ultraviolet photodissociation mass spectrometry. *Analytical Chemistry* 85, 19: 9253–9261.

345. Javier G. Magadán, Meghan O. Altman, William L. Ince, et al. 2014. Biogenesis of Influenza A Virus Hemagglutinin Cross-Protective Stem Epitopes. *PLoS Pathog* 10, 6: e1004204.

346. Michele Magrane and UniProt Consortium. 2011. UniProt Knowledgebase: a hub of integrated protein data. *Database: The Journal of Biological Databases and Curation* 2011.

347. Michael W. Mahoney and William L. Jorgensen. 2000. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *The Journal of Chemical Physics* 112, 20: 8910–8922.

348. Alexander Makarov. 2000. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Analytical Chemistry* 72, 6: 1156–1162.

349. Alexander Makarov, Eduard Denisov, Alexander Kholomeev, et al. 2006. Performance Evaluation of a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer. *Analytical Chemistry* 78, 7: 2113–2120.

350. B. A. Mamyrin, V. I. Karataev, D. V. Shmikk, and V. A. Zagulin. 1973. The mass-reflectron, a new nonmagnetic time-of-flight mass spectrometer with high resolution. *Soviet Journal of Experimental and Theoretical Physics* 37: 45.

351. M. Mann and M. Wilm. 1994. Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Analytical Chemistry* 66, 24: 4390–4399.

352. Raymond E. March. 1997. An Introduction to Quadrupole Ion Trap Mass Spectrometry. *Journal of Mass Spectrometry* 32, 4: 351–369.

353. Raymond E. March and John F. Todd. 2005. *Quadrupole ion trap mass spectrometry*. John Wiley & Sons.

354. Alan G. Marshall, Christopher L. Hendrickson, and George S. Jackson. 1998. Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrometry Reviews* 17, 1: 1–35.

355. Lennart Martens, Matthew Chambers, Marc Sturm, et al. 2011. mzML—a Community Standard for Mass Spectrometry Data. *Molecular & Cellular Proteomics* 10, 1: R110.000133.

356. Edward A. Mason and Earl Wadsworth McDaniel. 1988. *Transport properties of ions in gases*. Wiley, New York.

357. Steven W. Mast and Kelley W. Moremen. 2006. Family 47 alpha-mannosidases in N-glycan processing. *Methods in Enzymology* 415: 31–46.

358. Lea Mauko, Nathan A. Lacher, Matthias Pelzing, Anna Nordborg, Paul R. Haddad, and Emily F. Hilder. 2012. Comparison of ZIC-HILIC and graphitized carbon-based analytical approaches combined with exoglycosidase digestions for analysis of glycans from monoclonal antibodies. *Journal of Chromatography B* 911: 93–104.

359. Evan Maxwell, Yan Tan, Yuxiang Tan, et al. 2012. GlycReSoft: A Software Package for Automated Recognition of Glycans from LC/MS Data. *PLoS ONE* 7, 9: e45474.

360. J. C. May, C. R. Goodwin, N. M. Lareau, et al. 2014. Conformational Ordering of Biomolecules in the Gas Phase: Nitrogen Collision Cross Sections Measured on a Prototype High Resolution Drift Tube Ion Mobility-Mass Spectrometer. *Analytical Chemistry* 86: 2107–2116.

361. Jody C. May and John A. McLean. 2015. Ion Mobility-Mass Spectrometry: Time-Dispersive Instrumentation. *Analytical Chemistry* 87, 3: 1422–1436.

362. Anoop Mayampurath, Chuan-Yih Yu, Ehwang Song, Jagadheshwar Balan, Yehia Mechref, and Haixu Tang. 2013. Computational Framework for Identification of Intact Glycopeptides in Complex Samples. *Analytical Chemistry* 86, 1: 453–463.

363. M. Mazereeuw, A.j.p. Hofte, U.r. Tjaden, and J. van der Greef. 1997. A novel sheathless and electrodeless microelectrospray interface for the on-line coupling of capillary zone electrophoresis to mass spectrometry. *Rapid Communications in Mass Spectrometry* 11, 9: 981–986.

364. Graeme C. McAlister, Edward L. Huttlin, Wilhelm Haas, et al. 2012. Increasing the Multiplexing Capacity of TMTs Using Reporter Ion Isotopologues with Isobaric Masses. *Analytical Chemistry* 84, 17: 7469–7478.

365. Graeme C. McAlister, Douglas H. Phanstiel, Justin Brumbaugh, Michael S. Westphall, and Joshua J. Coon. 2011. Higher-energy Collision-activated Dissociation Without a Dedicated Collision Cell. *Molecular & Cellular Proteomics : MCP* 10, 5.

366. Graeme C. McAlister, Jason D. Russell, Neil G. Rumachik, et al. 2012. Analysis of the Acidic Proteome with Negative Electron-Transfer Dissociation Mass Spectrometry. *Analytical Chemistry* 84, 6: 2875–2882.

367. F. W. McLafferty. 1981. Tandem mass spectrometry. *Science* 214, 4518: 280–287.

368. J. A. McLean. 2009. The Mass-Mobility Correlation Redux: The Conformational Landscape of Anhydrous Biomolecules. *Journal of the American Society for Mass Spectrometry* 20: 1775–1781.

369. John A. McLean, Brandon T. Ruotolo, Kent J. Gillig, and David H. Russell. 2005. Ion mobility–mass spectrometry: a new paradigm for proteomics. *International Journal of Mass Spectrometry* 240, 3: 301–315.

370. Yehia Mechref. 2012. Use of CID/ETD Mass Spectrometry to Analyze Glycopeptides. *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]* 0 12: Unit-12.1111.

371. Yehia Mechref, Yunli Hu, Janie L. Desantos-Garcia, Ahmed Hussein, and Haixu Tang. 2013. Quantitative Glycomics Strategies. *Molecular & Cellular Proteomics* 12, 4: 874–884.

372. Yehia Mechref, Pilsoo Kang, and Milos V. Novotny. 2009. Solid-Phase Permethylation for Glycomic Analysis. In N.H. Packer and N.G. Karlsson, eds., *Glycomics*. Humana Press, Totowa, NJ, 53–64.

373. Yehia Mechref, Jan Muzikar, and Milos V. Novotny. 2005. Comprehensive assessment of N-glycans derived from a murine monoclonal antibody: a case for multimethodological approach. *Electrophoresis* 26, 10: 2034–2046.

374. Rafael A. Medina, Silke Stertz, Balaji Manicassamy, et al. 2013. Glycosylations in the Globular Head of the Hemagglutinin Protein Modulate the Virulence and Antigenic Properties of the H1N1 Influenza Viruses. *Science Translational Medicine* 5, 187: 187ra70-187ra70.

375. J. S. Mellors, V. Gorbounov, R. S. Ramsey, and J. M. Ramsey. 2008. Fully Integrated Glass Microfluidic Device for Performing High-Efficiency Capillary Electrophoresis and Electrospray Ionization Mass Spectrometry. *Analytical chemistry* 80, 18: 6881–6887.

376. Michael Melmer, Thomas Stangler, Andreas Premstaller, and Wolfgang Lindner. 2011. Comparison of hydrophilic-interaction, reversed-phase and porous graphitic carbon chromatography for glycan analysis. *Journal of Chromatography A* 1218, 1: 118–123.

377. Karsten Michelmann, Joshua A. Silveira, Mark E. Ridgeway, and Melvin A. Park. 2015. Fundamentals of trapped ion mobility spectrometry. *Journal of the American Society for Mass Spectrometry* 26, 1: 14–24.

378. Leann M. Mikesh, Beatrix Ueberheide, An Chi, et al. 2006. The utility of ETD mass spectrometry in proteomic analysis. *Biochimica Et Biophysica Acta* 1764, 12: 1811–1822.

379. D. Mikhailov, H. C. Young, R. J. Linhardt, and K. H. Mayo. 1999. Heparin dodecasaccharide binding to platelet factor-4 and growth- related protein-alpha. Induction of a partially folded state and implications for heparin-induced thrombocytopenia. *J Biol Chem* 274: 25317–29.

380. E. Mirgorodskaya, P. Roepstorff, and R. A. Zubarev. 1999. Localization of O-Glycosylation Sites in Peptides by Electron Capture Dissociation in a Fourier Transform Mass Spectrometer. *Analytical Chemistry* 71, 20: 4431–4436.

381. Yoshiaki Miura, Yasuro Shinohara, Jun-ichi Furukawa, Noriko Nagahori, and Shin-Ichiro Nishimura. 2007. Rapid and Simple Solid-Phase Esterification of Sialic Acid Residues for Quantitative Glycomics by Mass Spectrometry. *Chemistry – A European Journal* 13, 17: 4797–4804.

382. Samuel Molesworth, Sandra Osburn, and Michael Van Stipdonk. 2009. Influence of Size on Apparent Scrambling of Sequence During CID of b-Type Ions. *Journal of the American Society for Mass Spectrometry* 20, 11: 2174–2181.

383. Willy Morelle, Valegh Faid, and Jean-Claude Michalski. 2004. Structural analysis of permethylated oligosaccharides using electrospray ionization quadrupole time-of-

flight tandem mass spectrometry and deutero-reduction. *Rapid communications in mass spectrometry: RCM* 18, 20: 2451–2464.

384. Willy Morelle and Jean-Claude Michalski. 2007. Analysis of protein glycosylation by mass spectrometry. *Nature Protocols* 2, 7: 1585–1602.

385. Graham Mullard, James W. Allwood, Ralf Weber, et al. 2015. A new strategy for MS/MS data acquisition applying multiple data dependent experiments on Orbitrap mass spectrometers in non-targeted metabolomic applications. *Metabolomics* 11, 5: 1068–1080.

386. B. Mulloy. 2012. Structure and physicochemical characterisation of heparin. *Handbook of experimental pharmacology*: 77–98.

387. B. Mulloy, E. Gray, and T. W. Barrowcliffe. 2000. Characterization of unfractionated heparin: comparison of materials from the last 50 years. *Thrombosis and haemostasis* 84: 1052–6.

388. B. Mulloy, A. Heath, Z. Shriver, et al. 2014. USP compendial methods for analysis of heparin: chromatographic determination of molecular weight distributions for heparin sodium. *Analytical and bioanalytical chemistry* 406: 4815–23.

389. Amber N. Murray, Wentao Chen, Aristotelis Antonopoulos, et al. 2015. Enhanced Aromatic Sequons Increase Oligosaccharyltransferase Glycosylation Efficiency and Glycan Homogeneity. *Chemistry & Biology* 22, 8: 1052–1062.

390. Simon Mysling, Giuseppe Palmisano, Peter Højrup, and Morten Thaysen-Andersen. 2010. Utilizing Ion-Pairing Hydrophilic Interaction Chromatography Solid Phase Extraction for Efficient Glycopeptide Enrichment in Glycoproteomics. *Analytical Chemistry* 82, 13: 5598–5609.

391. National Research Council (US) Committee on Assessing the Importance and Impact of Glycomics and Glycosciences. 2012. *Transforming Glycoscience: A Roadmap for the Future*. National Academies Press (US), Washington (DC).

392. Martha I. Nelson and Edward C. Holmes. 2007. The evolution of epidemic influenza. *Nature Reviews Genetics* 8, 3: 196–205.

393. Alexey I. Nesvizhskii. 2010. A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *Journal of proteomics* 73, 11: 2092–2123.

394. C. Neyra, J. Paladino, and M. Le Borgne. 2014. Oxidation of sialic acid using hydrogen peroxide as a new method to tune the reducing activity. *Carbohydrate Research* 386: 92–98.

395. Wenqin Ni, Jonathan Bones, and Barry L. Karger. 2013. In-Depth Characterization of N-Linked Oligosaccharides Using Fluoride-Mediated Negative Ion Microfluidic Chip LC-MS. *Analytical chemistry* 85, 6: 3127–3135.

396. Eugene N. Nikolaev, Yury I. Kostyukevich, and Gleb N. Vladimirov. 2016. Fourier transform ion cyclotron resonance (FT ICR) mass spectrometry: Theory and simulations. *Mass Spectrometry Reviews* 35, 2: 219–258.

397. Nikolaos M. Nikolaidis, Mitchell R. White, Kimberly Allen, et al. 2014. Mutations flanking the carbohydrate binding site of surfactant protein D confer antiviral activity for pandemic influenza A viruses. *American Journal of Physiology - Lung Cellular and Molecular Physiology* 306, 11: L1036–L1044.

398. Koji Nishi, Tomomi Ono, Teruya Nakamura, et al. 2011. Structural insights into differences in drug-binding selectivity between two forms of human alpha1-acid glycoprotein genetic variants, the A and F1*S forms. *The Journal of Biological Chemistry* 286, 16: 14427–14434.

399. Takashi Nishikaze, Hiroki Tsumoto, Sadanori Sekiya, Shinichi Iwamoto, Yuri Miura, and Koichi Tanaka. 2017. Differentiation of Sialyl Linkage Isomers by One-Pot Sialic Acid Derivatization for Mass Spectrometry-Based Glycan Profiling. *Analytical Chemistry* 89, 4: 2353–2360.

400. Anita K. Nivedha, Spandana Makeneni, Bethany Lachele Foley, Matthew B. Tessier, and Robert J. Woods. 2014. Importance of ligand conformational energies in carbohydrate docking: Sorting the wheat from the chaff. *Journal of Computational Chemistry* 35, 7: 526–539.

401. Milos V. Novotny, William R. Alley, and Benjamin F. Mann. 2013. Analytical glycobiology at high sensitivity: current approaches and directions. *Glycoconjugate Journal* 30, 2: 89–117.

402. Marcell Olajos, Péter Hajós, Guenther K. Bonn, and András Guttman. 2008. Sample preparation for the analysis of complex carbohydrates by multicapillary gel electrophoresis with light-emitting diode induced fluorescence detection. *Analytical Chemistry* 80, 11: 4241–4246.

403. Jose A. Olivares, Nhung T. Nguyen, Clement R. Yonker, and Richard D. Smith. 1987. On-line mass spectrometric detection for capillary zone electrophoresis. *Analytical Chemistry* 59, 8: 1230–1232.

404. Petra Olivova, Weibin Chen, Asish B. Chakraborty, and John C. Gebler. 2008. Determination of N-glycosylation sites and site heterogeneity in a monoclonal

antibody by electrospray quadrupole ion-mobility time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry* 22, 1: 29–40.

405. Sara Ongay, Alexander Boichenko, Natalia Govorukhina, and Rainer Bischoff. 2012. Glycopeptide enrichment and separation for protein glycosylation analysis. *Journal of Separation Science* 35, 18: 2341–2372.

406. Sara Ongay and Christian Neusüss. 2010. Isoform differentiation of intact AGP from human serum by capillary electrophoresis-mass spectrometry. *Analytical and Bioanalytical Chemistry* 398, 2: 845–855.

407. Martin Pabst and Friedrich Altmann. 2008. Influence of Electrosorption, Solvent, Temperature, and Ion Polarity on the Performance of LC-ESI-MS Using Graphitic Carbon for Acidic Oligosaccharides. *Analytical Chemistry* 80, 19: 7534–7542.

408. Martin Pabst, Jayakumar Singh Bondili, Johannes Stadlmann, Lukas Mach, and Friedrich Altmann. 2007. Mass + Retention Time = Structure: A Strategy for the Analysis of N-Glycans by Carbon LC-ESI-MS and Its Application to Fibrin N-Glycans. *Analytical Chemistry* 79, 13: 5051–5057.

409. Martin Pabst, Josephine Grass, Stefan Toegel, Eva Liebminger, Richard Strasser, and Friedrich Altmann. 2012. Isomeric analysis of oligomannosidic N-glycans and their dolichol-linked precursors. *Glycobiology* 22, 3: 389–399.

410. Nicolle H. Packer, Margaret A. Lawson, Daniel R. Jardine, and John W. Redmond. 1998. A general approach to desalting oligosaccharides released from glycoproteins. *Glycoconjugate Journal* 15, 8: 737–747.

411. Kevin Pagel and David J. Harvey. 2013. Ion Mobility–Mass Spectrometry of Complex Carbohydrates: Collision Cross Sections of Sodiated N-linked Glycans. *Analytical Chemistry* 85, 10: 5138–5145.

412. Giuseppe Palmisano, Sara Eun Lendal, Kasper Engholm-Keller, Rikke Leth-Larsen, Benjamin L. Parker, and Martin R. Larsen. 2010. Selective enrichment of sialic acid-containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry. *Nature Protocols* 5, 12: 1974–1982.

413. Giuseppe Palmisano, Benjamin L. Parker, Kasper Engholm-Keller, et al. 2012. A Novel Method for the Simultaneous Enrichment, Identification, and Quantification of Phosphopeptides and Sialylated Glycopeptides Applied to a Temporal Profile of Mouse Brain Development. *Molecular & Cellular Proteomics* 11, 11: 1191–1202.

414. Alexandre Panchaud, Alexander Scherl, Scott A. Shaffer, et al. 2009. Precursor Acquisition Independent From Ion Count: How to Dive Deeper into the Proteomics Ocean. *Analytical Chemistry* 81, 15: 6481–6488.

415. Ioannis A. Papayannopoulos. 1995. The interpretation of collision-induced dissociation tandem mass spectra of peptides. *Mass Spectrometry Reviews* 14, 1: 49–73.

416. Thakor Patel, James Bruce, Anthony Merry, et al. 1993. Use of hydrazine to release in intact and unreduced form both N- and O-linked oligosaccharides from glycoproteins. *Biochemistry* 32, 2: 679–693.

417. Richard H. Perry, R. Graham Cooks, and Robert J. Noll. 2008. Orbitrap mass spectrometry: Instrumentation, ion motion and applications. *Mass Spectrometry Reviews* 27, 6: 661–699.

418. Andrei-J. Petrescu, Adina-L. Milac, Stefana M. Petrescu, Raymond A. Dwek, and Mark R. Wormald. 2004. Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* 14, 2: 103–114.

419. Manolo Plasencia, Dragan Isailovic, Samuel Merenbloom, Yehia Mechref, and David Clemmer. 2008. Resolving and assigning N-linked glycan structural isomers from ovalbumin by IMS-MS. *Journal of The American Society for Mass Spectrometry* 19, 11: 1706–1715.

420. Rosina Plomp, Albert Bondt, Noortje de Haan, Yoann Rombouts, and Manfred Wuhrer. 2016. Recent advances in clinical glycoproteomics of immunoglobulins. *Molecular & Cellular Proteomics*: mcp.O116.058503.

421. Matthew J. Pollard, Christopher K. Hilton, Hongli Li, Kimberly Kaplan, Richard A. Yost, and Herbert H. Hill. 2011. Ion mobility spectrometer—field asymmetric ion mobility spectrometer-mass spectrometry. *International Journal for Ion Mobility Spectrometry* 14, 1: 15–22.

422. Petr Pompach, Kevin B. Chandler, Renny Lan, Nathan Edwards, and Radoslav Goldman. 2012. Semi-Automated Identification of N-Glycopeptides by Hydrophilic Interaction Chromatography, nano-Reverse-Phase LC–MS/MS, and Glycan Database Search. *Journal of Proteome Research* 11, 3: 1728–1740.

423. A. K. Powell and D. J. Harvey. 1996. Stabilization of sialic acids in N-linked oligosaccharides and gangliosides for analysis by positive ion matrix-assisted laser desorption/ionization mass spectrometry. *Rapid communications in mass spectrometry: RCM* 10, 9: 1027–1032.

424. Steven D. Pringle, Kevin Giles, Jason L. Wildgoose, et al. 2007. An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *International Journal of Mass Spectrometry* 261, 1: 1–12.

425. Yi Pu. 2016. Applications of ion mobility spectrometry, collision-induced dissociation and electron activated dissociation tandem mass spectrometry to structural analysis of proteins, glycoproteins and glycans. Retrieved February 12, 2017 from https://open.bu.edu/handle/2144/19574.

426. Yi Pu, Mark E. Ridgeway, Rebecca S. Glaskin, Melvin A. Park, Catherine E. Costello, and Cheng Lin. 2016. Separation and Identification of Isomeric Glycans by Selected Accumulation-Trapped Ion Mobility Spectrometry-Electron Activated Dissociation Tandem Mass Spectrometry. *Analytical Chemistry* 88, 7: 3440–3443.

427. Li Qi, John C. Kash, Vivien G. Dugan, et al. 2011. The ability of pandemic influenza virus hemagglutinins to induce lower respiratory pathology is associated with decreased surfactant protein D binding. *Virology* 412, 2: 426–434.

428. Rahul Raman, Kannan Tharakaraman, Zachary Shriver, Akila Jayaraman, V. Sasisekharan, and Ram Sasisekharan. 2014. Glycan receptor specificity as a useful tool for characterization and surveillance of influenza A virus. *Trends in Microbiology* 22, 11: 632–641.

429. Rahul Raman, Maha Venkataraman, Subu Ramakrishnan, Wei Lang, S. Raguram, and Ram Sasisekharan. 2006. Advancing glycomics: Implementation strategies at the Consortium for Functional Glycomics. *Glycobiology* 16, 5: 82R–90R.

430. Rawi Ramautar, Anthonius A. M. Heemskerk, Paul J. Hensbergen, André M. Deelder, Jean-Marc Busnel, and Oleg A. Mayboroda. 2012. CE-MS for proteomics: Advances in interface development and application. *Journal of Proteomics* 75, 13: 3814–3828.

431. T N C Ramya, Eranthie Weerapana, Benjamin F Cravatt, and James C Paulson. 2013. Glycoproteomics enabled by tagging sialic acid- or galactose-terminated glycans. *Glycobiology* 23, 2: 211–221.

432. Rene Ranzinger, Stephan Herget, Claus-Wilhelm von der Lieth, and Martin Frank. 2011. GlycomeDB--a unified database for carbohydrate structures. *Nucleic Acids Research* 39, Database issue: D373–D376.

433. René Ranzinger, Stephan Herget, Thomas Wetter, and Claus-Wilhelm von der Lieth. 2008. GlycomeDB - integration of open-access carbohydrate structure databases. *BMC bioinformatics* 9: 384.

434. Anurag S. Rathore and Helen Winkle. 2009. Quality by design for biopharmaceuticals. *Nature Biotechnology* 27, 1: 26–34.

435. P. C. Reading, C. A. Hartley, R. A. B. Ezekowitz, and E. M. Anders. 1995. A Serum Mannose-Binding Lectin Mediates Complement-Dependent Lysis of Influenza Virus-Infected Cells. *Biochemical and Biophysical Research Communications* 217, 3: 1128–1136.

436. Patrick C Reading, Danielle L Pickett, Michelle D Tate, Paul G Whitney, Emma R Job, and Andrew G Brooks. 2009. Loss of a single N-linked glycan from the hemagglutinin of influenza virus is associated with resistance to collectins and increased virulence in mice. *Respiratory Research* 10, 1: 117.

437. Erin A. Redman, Nicholas G. Batz, J. Scott Mellors, and J. Michael Ramsey. 2015. Integrated microfluidic capillary electrophoresis-electrospray ionization devices with online MS detection for the separation and characterization of intact monoclonal antibody variants. *Analytical Chemistry* 87, 4: 2264–2272.

438. Karli R. Reiding, Dennis Blank, Dennis M. Kuijper, André M. Deelder, and Manfred Wuhrer. 2014. High-Throughput Profiling of Protein N-Glycosylation by MALDI-TOF-MS Employing Linkage-Specific Sialic Acid Esterification. *Analytical Chemistry* 86, 12: 5784–5793.

439. Mario dos Reis, Asif U. Tamuri, Alan J. Hay, and Richard A. Goldstein. 2011. Charting the Host Adaptation of Influenza Viruses. *Molecular Biology and Evolution* 28, 6: 1755–1767.

440. Jian Min Ren, Tomas Rejtar, Lingyun Li, and Barry L. Karger. 2007. N-Glycan Structure Annotation of Glycopeptides Using a Linearized Glycan Structure Database (GlyDB). *Journal of Proteome Research* 6, 8: 3162–3173.

441. H. E. Revercomb and E. A. Mason. 1975. Theory of plasma chromatography/gaseous electrophoresis. Review. *Analytical Chemistry* 47, 7: 970–983.

442. Nicholas M. Riley, Matthew J. P. Rush, Christopher M. Rose, et al. 2015. The Negative Mode Proteome with Activated Ion Negative Electron Transfer Dissociation. *Molecular & Cellular Proteomics*: mcp.M115.049726.

443. Daniel R. Roe and Thomas E. Cheatham. 2013. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation* 9, 7: 3084–3095.

444. P. Roepstorff and J. Fohlman. 1984. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical Mass Spectrometry* 11, 11: 601.

445. G. N. Rogers, J. C. Paulson, R. S. Daniels, J. J. Skehel, I. A. Wilson, and D. C. Wiley. 1983. Single amino acid substitutions in influenza haemagglutinin change receptor binding specificity. *Nature* 304, 5921: 76–78.

446. Christopher M. Rose, Jason D. Russell, Aaron R. Ledvina, et al. 2013. Multipurpose dissociation cell for enhanced ETD of intact protein species. *Journal of the American Society for Mass Spectrometry* 24, 6: 816–827.

447. Philip L. Ross, Yulin N. Huang, Jason N. Marchese, et al. 2004. Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents. *Molecular & Cellular Proteomics* 3, 12: 1154–1169.

448. Vsevolod V. Rostovtsev, Luke G. Green, Valery V. Fokin, and K. Barry Sharpless. 2002. A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective "Ligation" of Azides and Terminal Alkynes. *Angewandte Chemie International Edition* 41, 14: 2596–2599.

449. Pauline M. Rudd, Tama Endo, Cristina Colominas, et al. 1999. Glycosylation differences between the normal and pathogenic prion protein isoforms. *Proceedings of the National Academy of Sciences* 96, 23: 13044–13049.

450. L. R. Ruhaak, G. Zauner, C. Huhn, C. Bruggink, A. M. Deelder, and M. Wuhrer. 2010. Glycan labeling strategies and their use in identification and quantification. *Analytical and Bioanalytical Chemistry* 397, 8: 3457–3481.

451. L. Renee Ruhaak, André M. Deelder, and Manfred Wuhrer. 2009. Oligosaccharide analysis by graphitized carbon liquid chromatography-mass spectrometry. *Analytical and Bioanalytical Chemistry* 394, 1: 163–174.

452. L. Renee Ruhaak, Sandra L. Taylor, Carol Stroble, et al. 2015. Differential N-Glycosylation Patterns in Lung Adenocarcinoma Tissue. *Journal of Proteome Research* 14, 11: 4538–4549.

453. Neil G. Rumachik, Graeme C. McAlister, Jason D. Russell, Derek J. Bailey, Craig D. Wenger, and Joshua J. Coon. 2012. Characterizing Peptide Neutral Losses Induced by Negative Electron-Transfer Dissociation (NETD). *Journal of the American Society for Mass Spectrometry* 23, 4: 718–727.

454. B. T. Ruotolo, S. J. Hyung, P. M. Robinson, K. Giles, R. H. Bateman, and C. V. Robinson. 2007. Ion mobility-mass spectrometry reveals long-lived, unfolded

intermediates in the dissociation of protein complexes. *Angewandte Chemie-International Edition* 46: 8001–8004.

455. Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman J. C Berendsen. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* 23, 3: 327–341.

456. E. Sachon, S. Mohammed, N. Bache, and O. N. Jensen. 2006. Phosphopeptide quantitation using amine-reactive isobaric tagging reagents and tandem mass spectrometry: application to proteins isolated by gel electrophoresis. *Rapid Communications in Mass Spectrometry* 20, 7: 1127–1134.

457. Andrej Šali, Liz Potterton, Feng Yuan, Herman van Vlijmen, and Martin Karplus. 1995. Evaluation of comparative protein modeling by MODELLER. *Proteins: Structure, Function, and Bioinformatics* 23, 3: 318–326.

458. Romelia Salomon-Ferrer, Andreas W. Götz, Duncan Poole, Scott Le Grand, and Ross C. Walker. 2013. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *Journal of Chemical Theory and Computation* 9, 9: 3878–3888.

459. Miloslav Sanda and Radoslav Goldman. 2016. Data Independent Analysis of IgG Glycoforms in Samples of Unfractionated Human Plasma. *Analytical Chemistry* 88, 20: 10118–10125.

460. K Sandra, B Devreese, J Van Beeumen, I Stals, and M Claeyssens. 2004. The Q-Trap mass spectrometer, a novel tool in the study of protein glycosylation. *Journal of the American Society for Mass Spectrometry* 15, 3: 413–423.

461. R. Sasisekharan and Z. Shriver. 2009. From crisis to opportunity: A perspective on the heparin crisis. *Thrombosis and haemostasis* 102: 854–8.

462. Richard Alexander Scheltema, Jan-Peter Hauschild, Oliver Lange, et al. 2014. The Q Exactive HF, a Benchtop Mass Spectrometer with a Pre-filter, High-performance Quadrupole and an Ultra-high-field Orbitrap Analyzer. *Molecular & Cellular Proteomics* 13, 12: 3698–3708.

463. Matthew R. Schenauer, John K. Meissen, Youjin Seo, James B. Ames, and Julie A. Leary. 2009. Heparan Sulfate Separation, Sequencing, and Isomeric Differentiation: Ion Mobility Spectrometry Reveals Specific Iduronic and Glucuronic Acid-Containing Hexasaccharides. *Analytical Chemistry* 81, 24: 10179–10185.

464. K. Schmid, R. B. Nimerg, A. Kimura, H. Yamaguchi, and J. P. Binette. 1977. The carbohydrate units of human plasma alpha1-acid glycoprotein. *Biochimica Et Biophysica Acta* 492, 2: 291–302.

465. Benjamin L. Schulz, Nicolle H. Packer, and Niclas G. Karlsson. 2002. Small-Scale Analysis of O-Linked Oligosaccharides from Glycoproteins and Mucins Separated by Gel Electrophoresis. *Analytical Chemistry* 74, 23: 6088–6097.

466. Irene T. Schulze. 1997. Effects of Glycosylation on the Properties and Functions of Influenza Virus Hemagglutinin. *The Journal of Infectious Diseases* 176, s1: S24–S28.

467. Michaela Scigelova, Martin Hornshaw, Anastassios Giannakopulos, and Alexander Makarov. 2011. Fourier transform mass spectrometry. *Molecular & Cellular Proteomics*: mcp.O111.009431.

468. Nichollas E Scott, Benjamin L Parker, Angela M Connolly, et al. 2011. Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, higher energy collisional dissociation, and electron transfer dissociation MS applied to the N-linked glycoproteome of Campylobacter jejuni. *Molecular & cellular proteomics: MCP* 10, 2: M000031-MCP201.

469. Barbara A. Seaton, Erika C. Crouch, Francis X. McCormack, James F. Head, Kevan L. Hartshorn, and Richard Mendelsohn. 2010. Review: Structural determinants of pattern recognition by lung collectins. *Innate Immunity* 16, 3: 143–150.

470. Roberto Sebastiano, Martha E. Mendieta, Nunzia Contiello, Attilio Citterio, and Pier Giorgio Righetti. 2009. An N-methylpolyvinylpyridinium cationic polymer for capillary coating in electrophoresis of proteins and peptides. *Electrophoresis* 30, 13: 2313–2320.

471. Zaneer M. Segu and Yehia Mechref. 2010. Characterizing protein glycosylation sites through higher-energy C-trap dissociation. *Rapid Communications in Mass Spectrometry* 24, 9: 1217–1225.

472. Sadanori Sekiya, Yoshinao Wada, and Koichi Tanaka. 2005. Derivatization for stabilizing sialic acids in MALDI-MS. *Analytical Chemistry* 77, 15: 4962–4968.

473. Michael W. Senko, Steven C. Beu, and Fred W. McLafferty. 1995. Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *Journal of the American Society for Mass Spectrometry* 6, 4: 229–233.

474. Michael W. Senko, Philip M. Remes, Jesse D. Canterbury, et al. 2013. Novel Parallelized Quadrupole/Linear Ion Trap/Orbitrap Tribrid Mass Spectrometer Improving Proteome Coverage and Peptide Identification Rates. *Analytical Chemistry* 85, 24: 11710–11714.

475. Youjin Seo, Armann Andaya, and Julie A. Leary. 2012. Preparation, Separation, and Conformational Analysis of Differentially Sulfated Heparin Octasaccharide Isomers Using Ion Mobility Mass Spectrometry. *Analytical Chemistry* 84, 5: 2416–2423.

476. Robert E. Settlage, Paul S. Russo, Jeffrey Shabanowitz, and Donald F. Hunt. 1998. A novel μ-ESI source for coupling capillary electrophoresis and mass spectrometry: Sequence determination of tumor peptides at the attomole level. *Journal of Microcolumn Separations* 10, 3: 281–285.

477. J. C. Severs and R. D. Smith. 1997. Characterization of the microdialysis junction interface for capillary electrophoresis/microelectrospray ionization mass spectrometry. *Analytical Chemistry* 69, 11: 2154–2158.

478. Quazi Shakey, Brian Bates, and Jiang Wu. 2010. An Approach to Quantifying N-Linked Glycoproteins by Enzyme-Catalyzed 18O3-Labeling of Solid-Phase Enriched Glycopeptides. *Analytical Chemistry* 82, 18: 7722–7728.

479. N. Sharon and H. Lis. 2004. History of lectins: from hemagglutinins to biological recognition molecules. *Glycobiology* 14: 53r–62r.

480. Jared B. Shaw, Wenzong Li, Dustin D. Holden, et al. 2013. Complete Protein Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation. *Journal of the American Chemical Society* 135, 34: 12646–12651.

481. Evgenia Shishkova, Alexander S. Hebert, and Joshua J. Coon. 2016. Now, More Than Ever, Proteomics Needs Better Chromatography. *Cell Systems* 3, 4: 321–324.

482. J. E. Shively and H. E. Conrad. 1976. Formation of anhydrosugars in the chemical depolymerization of heparin. *Biochemistry* 15: 3932–42.

483. Archana Shubhakar, Karli R. Reiding, Richard A. Gardner, Daniel I. R. Spencer, Daryl L. Fernandes, and Manfred Wuhrer. 2014. High-Throughput Analysis and Automation for Glycomics Studies. *Chromatographia* 78, 5–6: 321–333.

484. Alexandre A. Shvartsburg and Richard D. Smith. 2008. Fundamentals of Traveling Wave Ion Mobility Spectrometry. *Analytical chemistry* 80, 24: 9689–9699.

485. Joshua A. Silveira, Mark E. Ridgeway, and Melvin A. Park. 2014. High resolution trapped ion mobility spectrometery of peptides. *Analytical Chemistry* 86, 12: 5624–5627.

486. Charandeep Singh, Cleidiane G. Zampronio, Andrew J. Creese, and Helen J. Cooper. 2012. Higher Energy Collision Dissociation (HCD) Product Ion-Triggered Electron Transfer Dissociation (ETD) Mass Spectrometry for the Analysis of N-Linked Glycoproteins. *Journal of Proteome Research* 11, 9: 4517–4525.

487. J. J. Skehel and D. C. Wiley. 2000. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annual Review of Biochemistry* 69: 531–569.

488. Lekha Sleno and Dietrich A. Volmer. 2004. Ion activation methods for tandem mass spectrometry. *Journal of mass spectrometry: JMS* 39, 10: 1091–1112.

489. David F Smith and Richard D Cummings. 2014. Investigating virus–glycan interactions using glycan microarrays. *Current Opinion in Virology* 7: 79–87.

490. David F. Smith, Xuezheng Song, and Richard D. Cummings. 2010. Use of glycan microarrays to explore specificity of glycan-binding proteins. *Methods in Enzymology* 480: 417–444.

491. Richard D. Smith, Jose A. Olivares, Nhung T. Nguyen, and Harold R. Udseth. 1988. Capillary zone electrophoresis-mass spectrometry using an electrospray ionization interface. *Analytical Chemistry* 60, 5: 436–441.

492. Hangtian Song and Kristina Håkansson. 2012. Electron Detachment Dissociation and Negative Ion Infrared Multiphoton Dissociation of Electrosprayed Intact Proteins. *Analytical Chemistry* 84, 2: 871–876.

493. R. Burke Squires, Jyothi Noronha, Victoria Hunt, et al. 2012. Influenza research database: an integrated bioinformatics resource for influenza research and surveillance. *Influenza and Other Respiratory Viruses* 6, 6: 404–416.

494. Pamela Stanley, Harry Schachter, and Naoyuki Taniguchi. 2009. N-Glycans. In A. Varki, R.D. Cummings, J.D. Esko, et al., eds., *Essentials of Glycobiology*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY).

495. Gregory O. Staples, Michael J. Bowman, Catherine E. Costello, et al. 2009. A chip-based amide-HILIC LC/MS platform for glycosaminoglycan glycomics profiling. *PROTEOMICS* 9, 3: 686–695.

496. Gregory O. Staples, Hicham Naimy, Hongfeng Yin, et al. 2009. Improved Hydrophilic Interaction Chromatography LC/MS of Heparinoids Using a Chip with Postcolumn Makeup Flow. *Anal. Chem.* 82, 2: 516–522.

497. Gregory O. Staples, Xiaofeng Shi, and Joseph Zaia. 2010. Extended N-Sulfated Domains Reside at the Nonreducing End of Heparan Sulfate Chains. *Journal of Biological Chemistry* 285, 24: 18336–18343.

498. Kathrin Stavenhagen, Hannes Hinneburg, Morten Thaysen-Andersen, et al. 2013. Quantitative mapping of glycoprotein micro-heterogeneity and macro-heterogeneity: an evaluation of mass spectrometry signal strengths using synthetic peptides and glycopeptides. *Journal of Mass Spectrometry* 48, 6: 627–639.

499. Kathrin Stavenhagen, Daniel Kolarich, and Manfred Wuhrer. 2014. Clinical Glycomics Employing Graphitized Carbon Liquid Chromatography–Mass Spectrometry. *Chromatographia*: 1–14.

500. Kathrin Stavenhagen, Rosina Plomp, and Manfred Wuhrer. 2015. Site-Specific Protein N- and O-Glycosylation Analysis by a C18-Porous Graphitized Carbon–Liquid Chromatography-Electrospray Ionization Mass Spectrometry Approach Using Pronase Treated Glycopeptides. *Analytical Chemistry* 87, 23: 11691–11699.

501. David A. Steinhauer. 2013. Influenza: Pathways to human adaptation. *Nature* 499, 7459: 412–413.

502. James Stevens, Ola Blixt, Li-Mei Chen, Ruben O. Donis, James C. Paulson, and Ian A. Wilson. 2008. Recent Avian H5N1 Viruses Exhibit Increased Propensity for Acquiring Human Receptor Specificity. *Journal of Molecular Biology* 381, 5: 1382–1394.

503. James Stevens, Ola Blixt, James C. Paulson, and Ian A. Wilson. 2006. Glycan microarray technologies: tools to survey host specificity of influenza viruses. *Nature Reviews Microbiology* 4, 11: 857–864.

504. James Stevens, Ola Blixt, Terrence M. Tumpey, Jeffery K. Taubenberger, James C. Paulson, and Ian A. Wilson. 2006. Structure and Receptor Specificity of the Hemagglutinin from an H5N1 Influenza Virus. *Science* 312, 5772: 404–410.

505. Sean R. Stowell, Connie M. Arthur, Ryan McBride, et al. 2014. Microbial glycan microarrays define key features of host-microbial interactions. *Nature Chemical Biology* 10, 6: 470–476.

506. Weston B. Struwe, Kevin Pagel, Justin L. P. Benesch, David J. Harvey, and Matthew P. Campbell. 2016. GlycoMob: an ion mobility-mass spectrometry collision cross section database for glycomics. *Glycoconjugate Journal* 33, 3: 399–404.

507. Zeynep Sumer-Bayraktar, Terry Nguyen-Khuong, Roxana Jayo, et al. 2012. Micro- and macroheterogeneity of N-glycosylation yields size and charge isoforms of human sex hormone binding globulin circulating in serum. *Proteomics* 12, 22: 3315–3327.

508. Xiangjie Sun, Akila Jayaraman, Pavithra Maniprasad, et al. 2013. N-Linked Glycosylation of the Hemagglutinin Protein Influences Virulence and Antigenicity of the 1918 Pandemic and Seasonal H1N1 Influenza A Viruses. *Journal of Virology* 87, 15: 8756–8766.

509. Shamil Sunyaev, Adam J. Liska, Alexander Golod, Anna Shevchenko, and Andrej Shevchenko. 2003. MultiTag: Multiple Error-Tolerant Sequence Tag Search for the Sequence-Similarity Identification of Proteins by Mass Spectrometry. *Analytical Chemistry* 75, 6: 1307–1315.

510. H. Suzuki, O. Müller, A. Guttman, and B. L. Karger. 1997. Analysis of 1-aminopyrene-3,6,8-trisulfonate-derivatized oligosaccharides by capillary electrophoresis with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Analytical Chemistry* 69, 22: 4554–4559.

511. Yasuo Suzuki, Toshihiro Ito, Takashi Suzuki, et al. 2000. Sialic Acid Species as a Determinant of the Host Range of Influenza A Viruses. *Journal of Virology* 74, 24: 11825–11831.

512. Danielle L. Swaney, Graeme C. McAlister, Matthew Wirtala, Jae C. Schwartz, John E. P. Syka, and Joshua J. Coon. 2007. Supplemental Activation Method for High-Efficiency Electron-Transfer Dissociation of Doubly Protonated Peptide Precursors. *Analytical Chemistry* 79, 2: 477–485.

513. John E. P. Syka, Joshua J. Coon, Melanie J. Schroeder, Jeffrey Shabanowitz, and Donald F. Hunt. 2004. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* 101, 26: 9528–9533.

514. Zoltan Szabo, András Guttman, Tomas Rejtar, and Barry L. Karger. 2010. Improved sample preparation method for glycan analysis of glycoproteins by CE-LIF and CE-MS. *Electrophoresis* 31, 8: 1389–1395.

515. Keqi Tang, Alexandre A. Shvartsburg, Hak-No Lee, et al. 2005. High-Sensitivity Ion Mobility Spectrometry/Mass Spectrometry Using Electrodynamic Ion Funnel Interfaces. *Analytical Chemistry* 77, 10: 3330–3339.

516. Michelle D. Tate, Emma R. Job, Yi-Mo Deng, Vithiagaran Gunalan, Sebastian Maurer-Stroh, and Patrick C. Reading. 2014. Playing Hide and Seek: How

Glycosylation of the Influenza Virus Hemagglutinin Can Modulate the Immune Response to Infection. *Viruses* 6, 3: 1294–1316.

517. P. Tatrai, K. Egedi, A. Somoracz, et al. 2010. Quantitative and Qualitative Alterations of Heparan Sulfate in Fibrogenic Liver Diseases and Hepatocellular Cancer. *Journal of Histochemistry & Cytochemistry* 58: 429–441.

518. T. Tecle, M. R. White, E. C. Crouch, and K. L. Hartshorn. 2007. Inhibition of influenza viral neuraminidase activity by collectins. *Archives of Virology* 152, 9: 1731–1742.

519. K. Thalassinos, S. E. Slade, K. R. Jennings, et al. 2004. Ion mobility mass spectrometry of proteins in a modified commercial mass spectrometer. *International Journal of Mass Spectrometry* 236, 1–3: 55–63.

520. Kannan Tharakaraman, Rahul Raman, Nathan W. Stebbins, Karthik Viswanathan, Viswanathan Sasisekharan, and Ram Sasisekharan. 2013. Antigenically intact hemagglutinin in circulating avian and swine influenza viruses and potential for H3N2 pandemic. *Scientific Reports* 3.

521. Kannan Tharakaraman, Rahul Raman, Karthik Viswanathan, et al. 2013. Structural Determinants for Naturally Evolving H5N1 Hemagglutinin to Switch Its Receptor Specificity. *Cell* 153, 7: 1475–1485.

522. Morten Thaysen-Andersen and Nicolle H. Packer. 2012. Site-specific glycoproteomics confirms that protein structure dictates formation of N-glycan type, core fucosylation and branching. *Glycobiology* 22: 1440–1452.

523. Morten Thaysen-Andersen and Nicolle H. Packer. 2014. Advances in LC–MS/MS-based glycoproteomics: Getting closer to system-wide site-specific mapping of the N- and O-glycoproteome. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1844, 9: 1437–1452.

524. Morten Thaysen-Andersen, Nicolle H. Packer, and Benjamin L. Schulz. 2016. Maturing glycoproteomics technologies provide unique structural insights into the N-glycoproteome and its regulation in health and disease. *Molecular & Cellular Proteomics*: mcp.O115.057638.

525. Roger Théberge, Giuseppe Infusini, Weiwei Tong, Mark E. McComb, and Catherine E. Costello. 2011. Top-Down Analysis of Small Plasma Proteins Using an LTQ-Orbitrap. Potential for Mass Spectrometry-Based Clinical Assays for Transthyretin and Hemoglobin. *International Journal of Mass Spectrometry* 300, 2–3: 130–142.

526. Andrew Thompson, Jürgen Schäfer, Karsten Kuhn, et al. 2003. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Analytical Chemistry* 75, 8: 1895–1904.

527. Kristina A. Thomsson, Niclas G. Karlsson, and Gunnar C. Hansson. 1999. Liquid chromatography–electrospray mass spectrometry as a tool for the analysis of sulfated oligosaccharides from mucin glycoproteins. *Journal of Chromatography A* 854, 1–2: 131–139.

528. Sarah M. Totten, Christa L. Feasley, Abel Bermudez, and Sharon J. Pitteri. 2017. Parallel Comparison of N-Linked Glycopeptide Enrichment Techniques Reveals Extensive Glycoproteomic Analysis of Plasma Enabled by SAX-ERLIC. *Journal of Proteome Research*.

529. John Treanor. 2004. Influenza Vaccine — Outmaneuvering Antigenic Shift and Drift. *New England Journal of Medicine* 350, 3: 218–220.

530. M J Treuheit, C E Costello, and H B Halsall. 1992. Analysis of the five glycosylation sites of human alpha 1-acid glycoprotein. *Biochemical Journal* 283, Pt 1: 105–112.

531. Shweta Tripathi, Mitchell R. White, and Kevan L. Hartshorn. 2015. The amazing innate immune response to influenza A virus infection. *Innate Immunity* 21, 1: 73–98.

532. G.A. Turner. 1992. N-glycosylation of serum proteins in disease and its investigation using lectins. *Clinica Chimica Acta* 208, 3: 149–171.

533. Iva Turyan, Xiaoping Hronowski, Zoran Sosic, and Yelena Lyubarskaya. 2014. Comparison of two approaches for quantitative O-linked glycan analysis used in characterization of recombinant proteins. *Analytical Biochemistry* 446: 28–36.

534. C. Uetrecht, R. J. Rose, E. van Duijn, K. Lorenzen, and A. J. R. Heck. 2010. Ion mobility mass spectrometry of proteins and protein assemblies. *Chemical Society Reviews* 39: 1633–1655.

535. Jakub Ujma, Kevin Giles, Michael Morris, and Perdita E. Barran. 2016. New High Resolution Ion Mobility Mass Spectrometer Capable of Measurements of Collision Cross Sections from 150 to 520 K. *Analytical Chemistry* 88, 19: 9469–9478.

536. Sara Ullsten, Aida Zuberovic, Magnus Wetterhall, Emilia Hardenborg, Karin E. Markides, and Jonas Bergquist. 2004. A polyamine coating for enhanced capillary electrophoresis-electrospray ionization-mass spectrometry of proteins and peptides. *Electrophoresis* 25, 13: 2090–2099.

537. UniProt Consortium. 2015. UniProt: a hub for protein information. *Nucleic Acids Research* 43, Database issue: D204-212.

538. Martin Van Eijk, Mitchell R White, Joseph J Batenburg, et al. 2004. Interactions of Influenza A Virus with Sialic Acids Present on Porcine Surfactant Protein D. *American Journal of Respiratory Cell and Molecular Biology* 30, 6: 871–879.

539. Csaba Váradi, Stefan Mittermayr, Silvia Millán-Martín, and Jonathan Bones. 2016. Quantitative twoplex glycan analysis using (12)C6 and (13)C6 stable isotope 2-aminobenzoic acid labelling and capillary electrophoresis mass spectrometry. *Analytical and Bioanalytical Chemistry* 408, 30: 8691–8700.

540. A Varki. 1993. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* 3, 2: 97–130.

541. Ajit Varki. 2006. Nothing in Glycobiology Makes Sense, except in the Light of Evolution. *Cell* 126, 5: 841–845.

542. Ajit Varki. 2016. Biological roles of glycans. *Glycobiology*.

543. Ajit Varki, Richard D. Cummings, Jeffrey D. Esko, et al., eds. 2009. *Essentials of Glycobiology*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY).

544. Anatoli N. Verentchikov, Werner Ens, and Kenneth G. Standing. 1994. Reflecting time-of-flight mass spectrometer with an electrospray ion source and orthogonal extraction. *Analytical Chemistry* 66, 1: 126–133.

545. Anamika Verma, Mitchell White, Vinod Vathipadiekal, et al. 2012. Human H-Ficolin Inhibits Replication of Seasonal and Pandemic Influenza A Viruses. *Journal of immunology (Baltimore, Md. : 1950)* 189, 5: 2478–2487.

546. David J. Vigerust and Virginia L. Shepherd. 2007. Virus glycosylation: role in virulence and immune interactions. *Trends in Microbiology* 15, 5: 211–218.

547. Antonio Villalobo, José Antonio Horcajadas, Sabine André, and Hans-J. Gabius. 1996. Glycobiology of Signal Transduction. In H.-J. Gabius and S. Gabius, eds., *Glycosciences*. Wiley-VCH Verlag GmbH, 485–496.

548. R.I. Viner, S. Snovida, E. Bodnar, H. Perreault, and J. Saba. 2010. A Novel Workflow for Glycopeptide Analysis Using Cellulose-Based Separation Cartridges, TMT-Labeling and LTQ Orbitrap ETD. *Journal of Biomolecular Techniques : JBT* 21, 3 Suppl: S25.

549. Rui Vitorino, Renato Alves, António Barros, et al. 2010. Finding new posttranslational modifications in salivary proline-rich proteins. *Proteomics* 10, 20: 3732–3742.

550. Valery G. Voinov, Samuel E. Bennett, and Douglas F. Barofsky. 2015. Electron-induced dissociation of peptides in a triple quadrupole mass spectrometer retrofitted with an electromagnetostatic cell. *Journal of the American Society for Mass Spectrometry* 26, 5: 752–761.

551. Valery G. Voinov, Samuel E. Bennett, Joseph S. Beckman, and Douglas F. Barofsky. 2014. ECD of tyrosine phosphorylation in a triple quadrupole mass spectrometer with a radio-frequency-free electromagnetostatic cell. *Journal of the American Society for Mass Spectrometry* 25, 10: 1730–1738.

552. Valery G. Voinov, Max L. Deinzer, Joseph S. Beckman, and Douglas F. Barofsky. 2011. Electron capture, collision-induced, and electron capture-collision induced dissociation in Q-TOF. *Journal of the American Society for Mass Spectrometry* 22, 4: 607–611.

553. Valery G. Voinov, Peter D. Hoffman, Samuel E. Bennett, Joseph S. Beckman, and Douglas F. Barofsky. 2015. Electron Capture Dissociation of Sodium-Adducted Peptides on a Modified Quadrupole/Time-of-Flight Mass Spectrometer. *Journal of the American Society for Mass Spectrometry* 26, 12: 2096–2104.

554. Gerda C. M. Vreeker and Manfred Wuhrer. 2016. Reversed-phase separation methods for glycan analysis. *Analytical and Bioanalytical Chemistry*: 1–20.

555. Robert P. de Vries, Erik de Vries, Berend Jan Bosch, Raoul J. de Groot, Peter J. M. Rottier, and Cornelis A. M. de Haan. 2010. The influenza A virus hemagglutinin glycosylation state affects receptor-binding specificity. *Virology* 403, 1: 17–25.

556. T. Wachs, R. L. Sheppard, and J. Henion. 1996. Design and applications of a self-aligning liquid junction-electrospray interface for capillary electrophoresis-mass spectrometry. *Journal of Chromatography. B, Biomedical Applications* 685, 2: 335–342.

557. S. Hunter Walker, Januka Budhathoki-Uprety, Bruce M. Novak, and David C. Muddiman. 2011. Stable-Isotope Labeled Hydrophobic Hydrazide Reagents for the Relative Quantification of N-Linked Glycans by Electrospray Ionization Mass Spectrometry. *Analytical Chemistry* 83, 17: 6738–6745.

558. S. Hunter Walker, Laura M. Lilley, Monica F. Enamorado, Daniel L. Comins, and David C. Muddiman. 2011. Hydrophobic derivatization of N-linked glycans for

increased ion abundance in electrospray ionization mass spectrometry. *Journal of the American Society for Mass Spectrometry* 22, 8: 1309–1317.

559. S. Hunter Walker, Amber D. Taylor, and David C. Muddiman. 2013. Individuality Normalization when Labeling with Isotopic Glycan Hydrazide Tags (INLIGHT): A Novel Glycan Relative Quantification Strategy. *Journal of the American Society for Mass Spectrometry* 24, 9: 1376–1384.

560. Yonghui Wang, Shiaw-Lin Wu, and William S. Hancock. 2006. Approaches to the study of N-linked glycoproteins in human plasma using lectin affinity chromatography and nano-HPLC coupled to electrospray linear ion trap--Fourier transform mass spectrometry. *Glycobiology* 16, 6: 514–523.

561. Chih-Jen Wei, Jeffrey C. Boyington, Kaifan Dai, et al. 2010. Cross-Neutralization of 1918 and 2009 Influenza Viruses: Role of Glycans in Viral Evolution and Vaccine Design. *Science Translational Medicine* 2, 24: 24ra21-24ra21.

562. Chad R. Weisbrod, Michael R. Hoopmann, Michael W. Senko, and James E. Bruce. 2013. Performance Evaluation of a Dual Linear Ion Trap-Fourier Transform Ion Cyclotron Resonance Mass Spectrometer for Proteomics Research. *Journal of proteomics* 88: 109–119.

563. Craig D. Wenger and Joshua J. Coon. 2013. A Proteomics Search Algorithm Specifically Designed for High-Resolution Tandem Mass Spectra. *Journal of Proteome Research* 12, 3: 1377–1386.

564. Mitchell R. White, Patrick Boland, Tesfaldet Tecle, et al. 2010. Enhancement of Antiviral Activity of Collectin Trimers through Cross-Linking and Mutagenesis of the Carbohydrate Recognition Domain. *Journal of Innate Immunity* 2, 3: 267–279.

565. Mitchell R White, Mona Doss, Patrick Boland, Tesfaldet Tecle, and Kevan L Hartshorn. 2008. Innate immunity to influenza virus: implications for future therapy. *Expert Review of Clinical Immunology* 4, 4: 497–514.

566. Mitchell R. White, Ruth Kandel, Shweta Tripathi, et al. 2014. Alzheimer's Associated β-Amyloid Protein Inhibits Influenza A Virus and Modulates Viral Interactions with Phagocytes. *PLoS ONE* 9, 7: e101364.

567. Julia Wiesner, Thomas Premsler, and Albert Sickmann. 2008. Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications. *Proteomics* 8, 21: 4466–4483.

568. Jonathan P. Williams, Andrew J. Creese, David R. Roper, Brian N. Green, and Helen J. Cooper. 2009. Hot Electron Capture Dissociation Distinguishes Leucine

from Isoleucine in a Novel Hemoglobin Variant, Hb Askew, β54(D5)Val→Ile. *Journal of the American Society for Mass Spectrometry* 20, 9: 1707–1713.

569. Jonathan P. Williams, Megan Grabenauer, Richard J. Holland, et al. 2010. Characterization of simple isomeric oligosaccharides and the rapid separation of glycan mixtures by ion mobility mass spectrometry. *International Journal of Mass Spectrometry* 298, 1–3: 119–127.

570. M. Wilm and M. Mann. 1996. Analytical properties of the nanoelectrospray ion source. *Analytical Chemistry* 68, 1: 1–8.

571. Ian A. Wilson, Robert C. Ladner, John J. Skehel, and Don C. Wiley. 1983. The structure and role of the carbohydrate moieties of influenza virus haemagglutinin. *Biochemical Society Transactions* 11, 2: 145–147.

572. Steven Ray Wilson, Mikolai Jankowski, Milairn Pepaj, et al. 2007. 2D LC separation and determination of bradykinin in rat muscle tissue dialysate with on-line SPE-HILIC-SPE-RP-MS. *Chromatographia* 66: 469–474.

573. D. Wittmer, B. K. Luckenbill, H. H. Hill, and Y. H. Chen. 1994. Electrospray-Ionization Ion Mobility Spectrometry. *Analytical Chemistry* 66: 2348–2355.

574. Jeremy J. Wolff, Tatiana N. Laremore, Alexander M. Busch, Robert J. Linhardt, and I. Jonathan Amster. 2008. Electron detachment dissociation of dermatan sulfate oligosaccharides. *Journal of the American Society for Mass Spectrometry* 19, 2: 294–304.

575. Jeremy J. Wolff, Tatiana N. Laremore, Alexander M. Busch, Robert J. Linhardt, and I. Jonathan Amster. 2008. Influence of Charge State and Sodium Cationization on the Electron Detachment Dissociation and Infrared Multiphoton Dissociation of Glycosaminoglycan Oligosaccharides. *Journal of the American Society for Mass Spectrometry* 19, 6: 790–798.

576. Jeremy J. Wolff, Franklin E. Leach, Tatiana N. Laremore, et al. 2010. Negative Electron Transfer Dissociation of Glycosaminoglycans. *Analytical chemistry* 82, 9: 3460–3466.

577. Philip SH Wong and R. Graham Cooks. 1997. Ion trap mass spectrometry. *Current separations* 16: 85–92.

578. Christina M. Woo, Anthony T. Iavarone, David R. Spiciarich, Krishnan K. Palaniappan, and Carolyn R. Bertozzi. 2015. Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact N- and O-glycopeptide discovery and analysis. *Nature Methods* 12, 6: 561–567.

579. Carrie L. Woodin, David Hua, Morgan Maxon, Kathryn R. Rebecchi, Eden P. Go, and Heather Desaire. 2012. GlycoPep Grader: A Web-Based Utility for Assigning the Composition of N-Linked Glycopeptides. *Analytical Chemistry* 84, 11: 4821–4829.

580. Carrie L. Woodin, Morgan Maxon, and Heather Desaire. 2013. Software for automated interpretation of mass spectrometry data from glycans and glycopeptides. *The Analyst* 138, 10: 2793.

581. A. S. Woods, M. Ugarov, T. Egan, et al. 2004. Lipid/peptide/nucleotide separation with MALDI-ion mobility-TOF MS. *Analytical Chemistry* 76: 2187–2195.

582. Sz-Wei Wu, Suh-Yuen Liang, Tsung-Hsien Pu, Fang-Yu Chang, and Kay-Hooi Khoo. 2013. Sweet-Heart — An integrated suite of enabling computational tools for automated MS2/MS3 sequencing and identification of glycopeptides. *Journal of Proteomics* 84: 1–16.

583. Zhengliang L. Wu, Xinyi Huang, Andrew J. Burton, and Karl A. D. Swift. 2015. Probing Sialoglycans on Fetal Bovine Fetuin with Azido-sugars Using Glycosyltransferases. *Glycobiology*: cwv109.

584. Manfred Wuhrer, Arjen R. de Boer, and André M. Deelder. 2009. Structural glycomics using hydrophilic interaction chromatography (HILIC) with mass spectrometry. *Mass Spectrometry Reviews* 28, 2: 192–206.

585. Manfred Wuhrer, M. Isabel Catalina, André M. Deelder, and Cornelis H. Hokke. 2007. Glycoproteomics based on tandem mass spectrometry of glycopeptides. *Journal of Chromatography B* 849, 1–2: 115–128.

586. Manfred Wuhrer, Carolien A. M. Koeleman, and André M. Deelder. 2009. Hexose Rearrangements upon Fragmentation of N-Glycopeptides and Reductively Aminated N-Glycans. *Analytical Chemistry* 81, 11: 4422–4432.

587. Manfred Wuhrer, Carolien A. M. Koeleman, Cornelis H. Hokke, and André M. Deelder. 2006. Mass spectrometry of proton adducts of fucosylated N-glycans: fucose transfer between antennae gives rise to misleading fragments. *Rapid Communications in Mass Spectrometry* 20, 11: 1747–1754.

588. Peter Wujek, Elizabeth Kida, Marius Walus, Krystyna E. Wisniewski, and Adam A. Golabek. 2004. N-glycosylation is crucial for folding, trafficking, and stability of human tripeptidyl-peptidase I. *The Journal of Biological Chemistry* 279, 13: 12827–12839.

589. Vicki H. Wysocki, Katheryn A. Resing, Qingfen Zhang, and Guilong Cheng. 2005. Mass spectrometry of peptides and proteins. *Methods (San Diego, Calif.)* 35, 3: 211–222.

590. T. Wyttenbach, G. von Helden, J. J. Batka, D. Carlat, and M. T. Bowers. 1997. Effect of the long-range potential on ion mobility measurements. *Journal of the American Society for Mass Spectrometry* 8: 275–282.

591. Thomas Wyttenbach, Gert von Helden, and Michael T. Bowers. 1996. Gas-Phase Conformation of Biological Molecules: Bradykinin. *Journal of the American Chemical Society* 118, 35: 8355–8364.

592. Haopeng Xiao, Johanna M. Smeekens, and Ronghu Wu. 2016. Quantification of tunicamycin-induced protein expression and N -glycosylation changes in yeast. *Analyst* 141, 12: 3737–3745.

593. Haopeng Xiao and Ronghu Wu. 2017. Quantitative investigation of human cell surface N -glycoprotein dynamics. *Chemical Science* 8, 1: 268–277.

594. Rui Xu, Robert P. de Vries, Xueyong Zhu, et al. 2013. Preferential Recognition of Avian-Like Receptors in Human Influenza A H7N9 Viruses. *Science* 342, 6163: 1230–1235.

595. Ying Xu, Ulla-Maja Bailey, and Benjamin L. Schulz. 2015. Automated measurement of site-specific N-glycosylation occupancy with SWATH-MS. *PROTEOMICS* 15, 13: 2177–2186.

596. K. Yamashita, H. Ideo, T. Ohkura, et al. 1993. Sugar Chains of Serum Transferrin from Patients with Carbohydrate-Deficient Glycoprotein Syndrome - Evidence of Asparagine-N-Linked Oligosaccharide Transfer Deficiency. *Journal of Biological Chemistry* 268: 5783–5789.

597. Shuang J. Yang and Hui Zhang. 2012. Glycan Analysis by Reversible Reaction to Hydrazide Beads and Mass Spectrometry. *Analytical Chemistry* 84, 5: 2232–2238.

598. Shuang Yang, Meiyao Wang, Lijun Chen, et al. 2015. QUANTITY: An Isobaric Tag for Quantitative Glycomics. *Scientific Reports* 5: 17585.

599. Shuang Yang, Wei Yuan, Weiming Yang, et al. 2013. Glycan analysis by isobaric aldehyde reactive tags and mass spectrometry. *Analytical Chemistry* 85, 17: 8188–8195.

600. Shuang Yang and Hui Zhang. 2012. Solid-phase glycan isolation for glycomics analysis. *Proteomics. Clinical applications* 6, 0: 596–608.

601. Ziping Yang and William S. Hancock. 2004. Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column. *Journal of Chromatography A* 1053, 1–2: 79–88.

602. Ziping Yang and William S. Hancock. 2005. Monitoring glycosylation pattern changes of glycoproteins using multi-lectin affinity chromatography. *Journal of Chromatography A* 1070, 1–2: 57–64.

603. Liyan Yu, Haibing He, Zhifei Hu, and Zhongjun Ma. 2016. Comprehensive quantification of N-glycoproteome in Fusarium graminearum reveals intensive glycosylation changes against fungicide. *Journal of Proteomics* 142: 82–90.

604. Shin-Yi Yu, Sz-Wei Wu, and Kay-Hooi Khoo. 2006. Distinctive characteristics of MALDI-Q/TOF and TOF/TOF tandem mass spectrometry for sequencing of permethylated complex type N-glycans. *Glycoconjugate Journal* 23, 5–6: 355–369.

605. Xiang Yu, Yan Jiang, Yajie Chen, Yiqun Huang, Catherine E. Costello, and Cheng Lin. 2013. Detailed Glycan Structural Characterization by Electronic Excitation Dissociation. *Analytical Chemistry* 85, 21: 10017–10021.

606. Xiang Yu and Wendy Zhong. 2016. Differentiation of Norvaline and Valine in Peptides by Hot Electron Capture Dissociation. *Analytical Chemistry* 88, 11: 5914–5919.

607. Lucia F. Zacchi and Benjamin L. Schulz. 2016. N-glycoprotein macroheterogeneity: biological implications and proteomic characterization. *Glycoconjugate Journal* 33, 3: 359–376.

608. Joseph Zaia. 2008. Mass Spectrometry and the Emerging Field of Glycomics. *Chemistry & Biology* 15, 9: 881–892.

609. Joseph Zaia. 2009. On-line separations combined with MS for analysis of glycosaminoglycans. *Mass Spectrometry Reviews* 28, 2: 254–272.

610. Joseph Zaia. 2010. Mass Spectrometry and Glycomics. *OMICS : a Journal of Integrative Biology* 14, 4: 401–418.

611. Joseph Zaia, Kshitij Khatri, Joshua A. Klein, Chun Shao, Yuewei Sheng, and Rosa Viner. 2016. Complete Molecular Weight Profiling of Low Molecular Weight Heparins Using Size Exclusion Chromatography-Ion Suppressor-High Resolution Mass Spectrometry. *Analytical Chemistry*.

612. Gerhild Zauner, Carolien A. M. Koeleman, André M. Deelder, and Manfred Wuhrer. 2010. Protein glycosylation analysis by HILIC-LC-MS of Proteinase K-generated N- and O-glycopeptides. *Journal of Separation Science* 33, 6–7: 903–910.

613. Ai-hua Zhang, Hui Sun, Guang-li Yan, Ying Han, and Xi-jun Wang. 2013. Serum Proteomics in Biomedical Research: A Systematic Review. *Applied Biochemistry and Biotechnology* 170, 4: 774–786.

614. J. H. Zhang, L. L. Lindsay, J. L. Hedrick, and C. B. Lebrilla. 2004. Strategy for profiling and structure elucidation of mucin-type oligosaccharides by mass spectrometry. *Analytical Chemistry* 76: 5990–6001.

615. Jing Zhang, Lei Xin, Baozhen Shan, et al. 2012. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Molecular & cellular proteomics: MCP* 11, 4: M111.010587.

616. Ming Zhang, Brian Gaschen, Wendy Blay, et al. 2004. Tracking global patterns of N-linked glycosylation site variation in highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin. *Glycobiology* 14, 12: 1229–1246.

617. Qianqian Zhang, Xi Chen, Zhijia Zhu, et al. 2013. Structural Analysis of Low Molecular Weight Heparin by Ultraperformance Size Exclusion Chromatography/Time of Flight Mass Spectrometry and Capillary Zone Electrophoresis. *Analytical Chemistry* 85, 3: 1819–1827.

618. Shu Zhang, Xiaohui Liu, Xiaonan Kang, et al. 2012. iTRAQ plus 18O: A new technique for target glycoprotein analysis. *Talanta* 91: 122–127.

619. Terry Zhang, Rosa Viner, Zhiqi Hao, and Vlad Zabrouskov. 2011. Analysis of Glycopeptides Using Porous Graphite Chromatography and LTQ Orbitrap XL ETD Hybrid MS. In A.R. Ivanov and A.V. Lazarev, eds., *Sample Preparation in Biological Mass Spectrometry*. Springer Netherlands, 535–546.

620. Yan Zhang, Jiping Zhu, Yongtao Li, et al. 2013. Glycosylation on Hemagglutinin Affects the Virulence and Pathogenicity of Pandemic H1N1/2009 Influenza A Virus in Mice. *PLoS ONE* 8, 4: e61397.

621. Zhongqi Zhang. 2011. Prediction of Collision-Induced-Dissociation Spectra of Peptides with Post-translational or Process-Induced Modifications. *Analytical Chemistry* 83, 22: 8642–8651.

622. Cheng Zhao, Bo Xie, Shiu-Yung Chan, Catherine E. Costello, and Peter B. O'Connor. 2008. Collisionally Activated Dissociation and Electron Capture Dissociation Provide Complementary Structural Information for Branched Permethylated Oligosaccharides. *Journal of the American Society for Mass Spectrometry* 19, 1: 138–150.

623. Jia Zhao, Diane M. Simeone, David Heidt, Michelle A. Anderson, and David M. Lubman. 2006. Comparative serum glycoproteomics using lectin selected sialic acid glycoproteins with mass spectrometric analysis: application to pancreatic cancer serum. *Journal of Proteome Research* 5, 7: 1792–1802.

624. Peng Zhao, Rosa Viner, Chin Fen Teo, Geert-Jan Boons, David Horn, and Lance Wells. 2011. Combining High-Energy C-Trap Dissociation and Electron Transfer Dissociation for Protein O-GlcNAc Modification Site Assignment. *Journal of Proteome Research* 10, 9: 4088–4104.

625. Xueyun Zheng, Xing Zhang, Nathaniel S. Schocker, et al. 2016. Enhancing glycan isomer separations with metal ions and positive and negative polarity ion mobility spectrometry-mass spectrometry analyses. *Analytical and Bioanalytical Chemistry*: 1–10.

626. Xuefei Zhong, Zhengwei Chen, Sergei Snovida, Yan Liu, John C. Rogers, and Lingjun Li. 2015. Capillary Electrophoresis-Electrospray Ionization-Mass Spectrometry for Quantitative Analysis of Glycans Labeled with Multiplex Carbonyl-Reactive Tandem Mass Tags. *Analytical Chemistry* 87, 13: 6527–6534.

627. Shiyue Zhou, Xue Dong, Lucas Veillon, Yifan Huang, and Yehia Mechref. 2016. LC-MS/MS analysis of permethylated N-glycans facilitating isomeric characterization. *Analytical and Bioanalytical Chemistry*: 1–14.

628. Shiyue Zhou, Yunli Hu, and Yehia Mechref. 2016. High-temperature LC-MS/MS of permethylated glycans derived from glycoproteins. *Electrophoresis* 37, 11: 1506–1513.

629. Shiyue Zhou, Yunli Hu, Lucas Veillon, et al. 2016. Quantitative LC-MS/MS Glycomic Analysis of Biological Samples Using AminoxyTMT. *Analytical Chemistry*.

630. Wen Zhou and Kristina Håkansson. 2013. Electron Capture Dissociation of Divalent Metal-adducted Sulfated N-Glycans Released from Bovine Thyroid Stimulating Hormone. *Journal of the American Society for Mass Spectrometry* 24, 11.

631. Feifei Zhu, David E. Clemmer, and Jonathan C. Trinidad. 2017. Characterization of lectin binding affinities via direct LC-MS profiling: implications for glycopeptide enrichment and separation strategies. *Analyst* 142, 1: 65–74.

632. Feifei Zhu, Sunyoung Lee, Stephen J. Valentine, James P. Reilly, and David E. Clemmer. 2012. Mannose7 Glycan Isomer Characterization by IMS-MS/MS Analysis. *Journal of The American Society for Mass Spectrometry* 23, 12: 2158–2166.

633. Feifei Zhu, Jonathan C. Trinidad, and David E. Clemmer. 2015. Glycopeptide Site Heterogeneity and Structural Diversity Determined by Combined Lectin Affinity Chromatography/IMS/CID/MS Techniques. *Journal of The American Society for Mass Spectrometry*: 1–11.

634. Zhikai Zhu, Xiaomeng Su, Daniel F. Clark, Eden P. Go, and Heather Desaire. 2013. Characterizing O-Linked Glycopeptides by Electron Transfer Dissociation: Fragmentation Rules and Applications in Data Analysis. *Analytical Chemistry* 85, 17: 8403–8411.

635. Zhikai Zhu, Xiaomeng Su, Eden P. Go, and Heather Desaire. 2014. New Glycoproteomics Software, GlycoPep Evaluator, Generates Decoy Glycopeptides de Novo and Enables Accurate False Discovery Rate Analysis for Small Data Sets. *Analytical Chemistry* 86, 18: 9212–9219.

636. Roman Zubarev. 2006. Protein primary structure using orthogonal fragmentation techniques in Fourier transform mass spectrometry. *Expert Review of Proteomics* 3, 2: 251–261.

637. Roman A Zubarev. 2004. Electron-capture dissociation tandem mass spectrometry. *Current Opinion in Biotechnology* 15, 1: 12–16.

638. Roman A. Zubarev, Neil L. Kelleher, and Fred W. McLafferty. 1998. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *Journal of the American Chemical Society* 120, 13: 3265–3266.

639. Roman A. Zubarev and Alexander Makarov. 2013. Orbitrap Mass Spectrometry. *Analytical Chemistry* 85, 11: 5288–5296.

640. 1998. Lovenox (enoxaparin sodium) package insert. .

641. 2009. Recommendations for Symbolism and Nomenclature for Mass Spectroscopy. *Pure and Applied Chemistry* 50, 1: 65–74.

642. 2010. Docket No. FDA-2003-P-0273. .

643. 2016. Understanding the Chemistry and Biology of Glycosylation with Glycan Synthesis. *Annual Review of Biochemistry* 85, 1: 599–630.

644. Biomolecular structural separations by ion mobility–mass spectrometry - Springer. .

CURRICULUM VITAE