5-4-2020

MASS SPECTROMETRY BASED ANALYSIS OF PROTEIN N-GLYCOSYLATION IN BIOMARKER DISCOVERY AND GENE THERAPY WITH THE STUDY MODELS-HEMOPHILIA A INHIBITOR DEVELOPMENT AND rAAV

Arya Aloor

MASS SPECTROMETRY BASED ANALYSIS OF PROTEIN N-GLYCOSYLATION IN BIOMARKER DISCOVERY AND GENE THERAPY WITH THE STUDY MODELS-HEMOPHILIA A INHIBITOR DEVELOPMENT AND rAAV

by

ARYA ALOOR

Under the Direction of Jun Yin, PhD

ABSTRACT

Protein glycosylation is one of the critical post-translational modifications (PTMs) and practically engaged with a wide range of physiological and biological processes. Glycosylation is the most dynamic post-translational modification and an individual's glycome changes overcome the genetic factors and get affected by environmental factors which eventually reflect his lifestyle, physiological conditions and wellbeing. The flow study, we attempted to add this information to comprehend the glycoprotein biomarker identified with inhibitor advancement and connected the glycosylation related changes to the biochemical pathway of inhibitor development against rFVIII in HA population. We performed the study with mice and human models. Plasma and IgG N-glycome examination is one of the important methodologies to identify the biomarker related to numerous conditions. The N-glycome pattern also varies in

response to the treatment. The treatment-related modifications also reaffirm the observations noted in the progression of the disease.

Similarly, the glycosylation can be a useful strategy to modify the protein-based drugs to enhance its mode of action. The variant of AAV can be a potential capsid engineering technique to alter the tropism and improve the gene delivery range of host cell range for engineering a better gene delivery system. The small amount of and glycan variants are difficult to detect in a complex biological mixture, which may require various enrichment strategies, and sample preparations help to enhance the detection sensitivity in mass spectrometry. Due to the with the development of state-of-the-art mass spectrometry (MS) technology, we tried to identify N-glycan biomarkers related to inhibitor development in HA. Also, we decided to study the response of the patient after emicizumab. Additionally, we identified N-glycosylation in rAAV-8, which can be a potential direction for future capsid engineering.

INDEX WORDS: Glycosylation biomarker, Hemophilia A, Inhibitor development, *r*AAV-8 glycosylation, Virus glycosylation, Mass spectrometry

MASS SPECTROMETRY BASED ANALYSIS OF PROTEIN N-GLYCOSYLATION IN BIOMARKER DISCOVERY AND GENE THERAPY WITH THE STUDY MODELS-HEMOPHILIA A INHIBITOR DEVELOPMENT AND rAAV

by

ARYA ALOOR

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2020

MASS SPECTROMETRY BASED ANALYSIS OF PROTEIN GLYCOSYLATION IN BIOMARKER DISCOVERY AND GENETHERAPY WITH STUDY MODEL-HEMOPHILIA A INHIBITOR DEVELOPMENT AND rAAV

by

ARYA ALOOR

Committee Chair: Jun Yin

Committee: Gangli Wang

Ming Luo

Siming Wang

Lei Li

Electronic Version Approved:

Office of Graduate Services

College of Arts and Sciences

Georgia State University

May 2020

DEDICATION

To my daughter – For her sacrifice for the absence of her mother when she wanted the most

To my parents and husband– For being my strength and support

ACKNOWLEDGEMENTS

Firstly, all my gratitude to the universal power and my family for giving me the strength and support whenever I was weak. I gratefully acknowledge my previous mentor Dr. Peng George Wang, for allowing me to start my graduate work at Georgia State University and for facilitating all the requirements for my research work during my initial time. Thus, I would like to express my great appreciation to my advisor Dr. Jun Yin and my mentor Dr. Lei Li for their valuable time, guidance, constructive suggestions, and providing facilities for the completion of my research work. I extend my thanks to Dr. Weidong Xiao and his lab members for providing a great collaboration opportunity for our research. I also thank Dr. Barbara Konkle, Dr. Carol Miao, and her lab members for providing the sample for my research study. I also recognize my previous mentors Dr. Laxmi Adhikary and Dr. Melina Soares, for encouraging me to start my Ph.D. program and sustain in the process. I want to thank my friends Dr. Madhusudan Reddy Gaddi, Dr. Pradeep Narayana Swamy, Dr. Muhammed Shukkoor Kodengaden, Dr. Ananya Paul Ms. Bibi Mathews Ms. Velama Parassery, Dr. Sreejith Ramakrishnan and Ms. Sethulakshmy Anoop for their support throughout the last four years. I also, like to offer my special thanks to my dissertation committee members, Professor Gangli Wang, Professor Ming Luo, and Dr. Siming Wang, for their patient guidance, enthusiastic encouragement, and useful critiques for the betterment of my research work. I owe special gratitude to Dr. Cheng Ma, Dr. Jing Li, Ding Liu, Dr. He Zhu, Aishwarya P, and Dr. Ebtesam Gashash for their guidance and experimental help throughout the journey. I Profound my sincere thanks to the current and former colleagues in Dr. Wang's laboratory, who contributed to this work progression. Also, I extended my gratitude to the staff members, technicians, and employees and friends at Georgia State University.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

HA	Hemophilia A
AHA	Acquired hemophilia A
HB	Hemophilia B
HC	Hemophilia C
r FVIII	Recombinant factor VIII
r AAV	Recombinant AAV
r FIX	Recombinant Factor IX
2-AA	2-aminobenzoic acid
2-AB	2-aminobenzamide
2HPM	2- Hydrazino pyramidine
2HQ	2-Hydrazino quinoline
HIV	Human immune deficiency virus
PCC	Prothrombin complex concentrate
ABC	Ammonium bicarbonate
ACN	Acetonitrile
ADCC	Antibody-dependent cellular cytotoxicity
Asn	Asparagine
PNGase F	Peptide-N4-(N-acetyl-βglucosaminyl) asparagine amidase F
PTM	Post-translational modifications
BSA	Bovine serum albumin
CE	Capillary electrophoresis
CID	Collision-induced dissociation

DTT	Dithiothreitol
ERLIC	Electrostatic repulsion hydrophilic interaction chromatography
ESI	Electrospray ionization
ETD	Electron-transfer dissociation
FA	Formic acid
FASP	Filter aided sample preparation
FDR	False discovery rate
Fuc	Fucose
Gal	Galactose
GalNAc	N-Acetyl galactosamine
GlcNAc	N-Acetylglucosamine
Neu5GC or NGNA	N-glycolyl neuraminic acid
Neu5Ac or NANA	N-acetyl neuraminic acid
Asp (D)	Aspartic acid
Glu (E)	Glutamic acid
HCD	High-energy C-trap dissociation
HILIC	Hydrophilic interaction liquid chromatography
IAA	2-Iodoacetamide
IM-MS	Ion mobility-mass spectrometry
LC	Liquid chromatography
m/z	Ratio of mass to charge
MALDI	Matrix aided laser desorption/ionization
Man	Mannose

Met	Methionine
MS	Mass spectrometry
sDHB	Super- 2-hydroxy-5-methoxybenzoic acid
NCE	Normalized collision energy
Neu5Ac	N-Acetylneuraminic acid
OSARA	One-pot solid-phase amidation and reductive amination
QTOF	Quadrupole time-of-flight
HA	Hemophilia A
HB	Hemophilia B
I	Inhibitor
I+	Inhibitor positive
I-	Inhibitor negative
SCE	Stepped collision energy
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
Ser	Serine
SIL	Stable isotope labeling
TFA	Trifluoroacetic acid
Thr	Threonine
TiO2	Titanium dioxide
ZIC-HILIC	Zwitterionic-hydrophilic interaction liquid chromatography
AFP	α-Fetoprotein
CEA	Carcinoembryonic antigen
CA 125	Cancer antigen 125
LPL	Lipoprotein lipase

HSPG	Heparan sulfate proteoglycan
2-HQ	2-Hydrazino quinoline
2-HPM	2 Hydrazino pyramidine
DMF	Dimethylformamide
SRA	Site specific relative abundance
TRA	Total relative abundance
RSD	Relative standard deviation
SD	Standard deviation
Neu5Gc	N-Glycolylneuraminic acid
Neu5Ac	N-Acetylneuraminic acid
GPI	Glycosylphosphatidylinositol

1 CHAPTER 1: INTRODUCTION TO THE GLYCOSYLATION, ITS IMPORTANCE AND MS BASED TOOL FOR ANALYSIS

1.1 Introduction

Glycosylation is one of the most common PTMs of proteins that maintaining many biological functions in the body and regulate the health and disease condition.[1] In humans, it is estimated that 50% of proteins are glycosylated.[2] Physiological changes like aging, pregnancy, malignant metastasis, disease progression, inflammatory responses, metabolic changes, bacterial and viral infections can alter the microenvironment in the cell.[3,4,5,6]. Hence, glycan profile changes can be a prognostic or diagnostic biomarker for many physiological conditions.[7] The altered glycosylation pattern can also alter the binding specificity of the glycan with other molecules and eventually change their biological functions. Even though the importance of glycosylation is known, the inability to characterize the structural details of the attached glycoforms to correlate their biological activity was not apparent for many years. Glycosylation is not a template-driven procedure, as we see in protein DNA or RNA synthesis.[8] The glycosylation process requires a panel of enzymes. The main two enzymes which have a direct role in the structural design of glycoforms are glycosyltransferases and glycosidases.[9] The enzymes responsible for sugar nucleotide synthesis also passively participate in the process.[10] There are many factors, including the expression level of enzymes and proteins participate in the glycosylation pathways, metabolic changes, availability of sugar nucleotides, that affect the overall glycosylation pattern of the protein.[11] The enzymatic machines are required to regulate the glycan structure. The exoglycosidases and glycosyltransferases are residing at the membranes of ER and Golgi. The nascent protein gets exposed to this enzyme panel; during the post/ co-translational process and encounter a competitive modification.[12] The protein can exit any time during the process of glycosylation, and, it is not possible to have a uniform /completed glycoform during the event. Hence the

resulting protein can be a heterogeneous mixture that contains glycoforms, vary in their structure, composition, linkage, and spatial arrangement. Glycosylation is typically divided into N-linked glycosylation and O-linked glycosylation. The N-linked glycosylation is happening at the Asn residue, where the glycan is covalently linked to the N atom of the Asn residue. The precursor structure for N-linked glycan synthesis is a Glc3Man9GlcNAc2 glycoform which gets added at ER to the Asn-residue of the consensus sequence and followed by the cleavage of glucose residues to form Man9GlcNAc2 .The mannosidase enzyme cleave this structure to form Man8.[13,14]At this stage the protein is taken to Golgi and it can remain as high mannose or they get further treated with mannosidase or added by different sugars sequentially from GlcNAc, Galactose and N-Acetyl neuraminic acid. The N-glycan structure usually has a typical 'pentacore' structure, which contains two N-Acetylglucosamine (GlcNAc) and three mannose (Man) residues in a biantennary form. The 'pentacore' can be extended and modified at the nonreducing terminal with different monosaccharide residues to form various structures with a variety of linkages.[15] Based on the composition, the N-glycan is divided into three groups: complex, hybrid, and high mannose. If the extension of the pentacore mannose happening with varying sugar units, including GlcNAc, galactose (Gal), and N-Acetylneuraminic acid (Neu5Ac) at both the antennae, the resulting structure is called complex sugars. They can also extend only with mannose, which is called a high mannose structure. If the core structure is extended in one antenna with different sugars and another antenna with only mannose, then the structure is called hybrid. The structures can bare fucose at innermost GlcNAc residue by Alpha1-6 linkage is called core fucosylation.[16]Each of these sugars and their percentage has a significant role in the biological system. The peptide backbone contains a consensus motif N-X-S/T, where X can be any amino acid except for proline.[17]An additional N-glycosylation sequence N-X-C was discovered first in bacteria is also occasionally found in mammalian proteins.[18,19]Similarly,

the O-linked glycosylation is occurring through the covalent linkage between the O atom of serine (Ser, S) or threonine (Thr, T) to the glycan structure.[20] For O-linked glycosylation, as shown by a typical O-glycopeptide in ***figure 1.1***, an O-GalNAc (N-Acetylglucosamine) glycan is conjugated with Ser/Thr in the peptide backbone without a consensus motif. O-linked glycosylation is much simpler structures but does not have a single biosynthetic pathway as in N-glycan. There are several kinds of O-glycans found in nature. The main ones are α linked O-GlcNAc, α linked O-GalNAc, β-linked O-xylose, α -linked O-mannose, α/ β O-glucose/ galactose, and α linked O-fucose. Due to the high biological significances, glycobiologists widely studied O-GalNAc and O-GlcNAc type of O-glycan.[21] There are mainly eight types of O-GalNAc glycan cores with diverse structures, but they all share the uniform monosaccharide residue GalNAc at the reducing end of the sugar chain.[22]

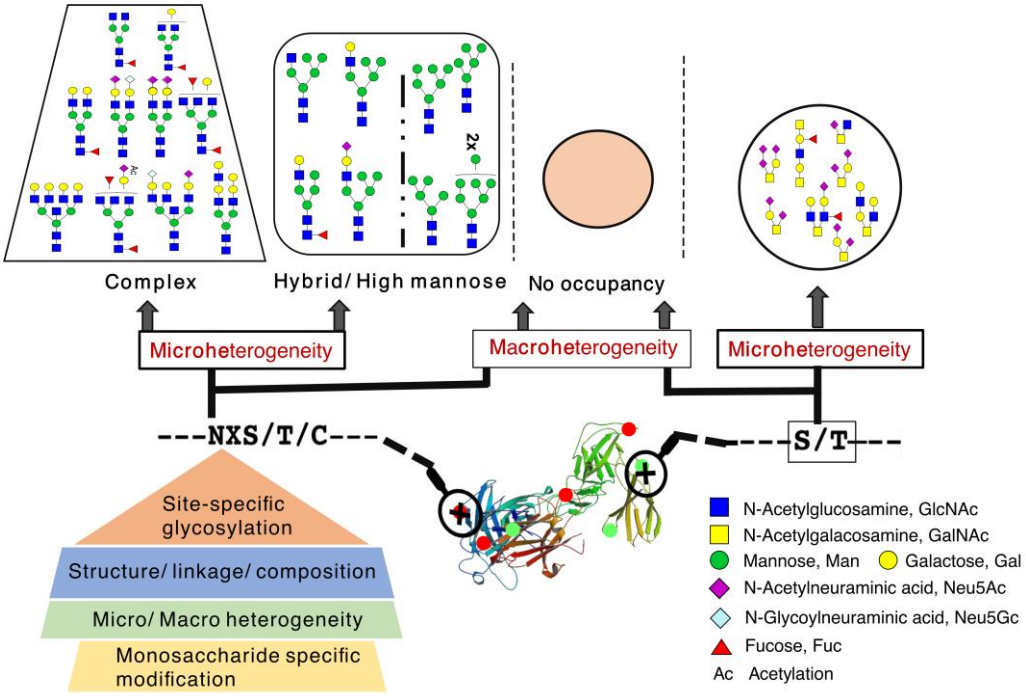


Figure 1.1-specific glycan heterogeneity
The N-glycosylation sites are labeled with red dots, and O-glycosylation sites are labeled with green dots. The potential site-specific N-glycosylation variations (at left) and O-glycosylation variations (at right) are illustrated with their micro and macro heterogeneity.

A few other glycosylations are also seen in nature. C-type glycosylation is where a mannose is added to carbon on a tryptophan side chain.[23] Glypiation features the GPI anchor that links

proteins to lipids through glycan linkages.[24]Phosphoglycans linked to phosphoserine through the phosphate group.[25]Very rarely, S-linked glycosylation, also seen in nature, refers to the linkage of glycan to the cysteine through its S-atom.[26]

1.2 Glycoprotein biomarker

In each glycoprotein, there can be multiple glycosylation sites, including N and O-glycosylation. The macro heterogeneity refers to the variability of occupancy of the glycan at each location. The structure of occupied glycan may also vary in each site, which refers to their microheterogeneity.[27] Both factors must be efficiently monitored. N- and O-linked glycans play an essential role in varieties of cellular processes, including modulating protein folding, regulating cell-cell interaction, cell-pathogen interaction, and receptor binding and signalling.

The difference in the expression of glycoprotein and their micro and macro heterogeneity can be directly correlated to some disease progression or physiological sate hence can be considered as a biomarker. There are numerous clinically approved bio markers identified with malignant growth, which are glycoproteins. For example, upregulation of the AFP[28–30] is a glycan biomarker of liver cancer, and CEA[31] is considered as a biomarker for colon cancer, CA125 is a ovarian cancer[32] biomarker, and PSA is well-known biomarker for prostate cancer.[33] Plasma protein glycan alteration can be also a biomarker of different disease conditions. For example, increased multi-antennary glycans with fucose residues in plasma are the diagnostic biomarker of hepatic cancer.[34] Similarly decreased core-fucosylated glycans, increased hybrid and multi-branched structures, decreased monoantennary, galactose, bisecting type or core fucose structures in plasma glycosylation is a diagnostic a pattern of disease progression in gastric cancer.[35]Whereas increased core fucosylation of plasma protein is a potential diagnostic biomarker of pancreatic cancer.[36]Increased plasma sialylation, high branched structures, high-mannosylation or outer-arm fucosylation and

decreased biantennary core-fucosylated glycans structures are the diagnostic biomarker of breast cancer[37].

1.3 Antibody biomarker

Antibody or immunoglobulin is one of the basic functional proteins in the body, which is essential to secure the body against outside substances, generally called antigens. At the point when an antigen enters the body, the immune system will get initiated to recognize and eliminate them through a series of biochemical responses, which is known as immune responses. The most essential part of immune activation is the generation of antibody specific to the foreign antigen by B-lymphocytes. The naïve B- lymphocyte go through the process of clonal selection and clonal expansion to produce high-binding affinity antigen which can specifically bind and attach to that type of antigen.[38] There are many glycoprotein serves as a biomarker in the body, that upregulate in response to the physiological changes. IgG is one of the dominant glycoprotein in the plasma. The IgG structures have conserved glycosylation site at CH2 domain and variable region which has a profound impact on their function.[39]

The glycosylation of human IgG1is a good example to illustrate how glycosylation change influences protein function. IgG1 is a well-studied serum glycoprotein and contains only one N-glycosylation site at 'Asn 297' of the Fc region. Antibody glycosylation is one of the best study model for correlating the change in glycosylation with the protein function. The absence of core fucosylation of IgG1 glycosylation will result in a dramatically enhanced binding of Fc to FcγRIIIa (Fc gamma receptor IIIa), promoting antibody-dependent cellular cytotoxicity (ADCC).[40] The addition of Neu5Ac at the terminal end of the glycans will inhibit FcγRIIIa ligation and increase FcγRIIIb expression, showing an anti-inflammatory effect.[40] With a decreasing degree of galactosylation, the inflammation effect will be promoted. Also, bisecting GlcNAc correlates with deceased core fucosylation and enhances ADCC.[40]

Depending on the variability in each sub type the Fc glycosylation is varying in number for example, IgG1 glycosylation site at N-180 (also sometime referred as N-297) corresponds to N176 of IgG2, N-227 of IgG3, and N-177 of IgG4. Regardless of the IgG subtype, the conserved glycosylation is generally known as CH2-84.4.[41] Different IgG subtypes have different FcR affinity. The antibody glycosylation also can decide the Fc receptor binding there by regulation of its effector functions including phagocytosis, antigen presentation, ADCC, CDC and cytokine release.[42,43] The FcRs are the on and off switch of antibody effector functions. FcγRI, FcγRII, and FcγRIII are the three major classes of FcRs in human. The receptors are of two major types as inhibitory (FcγRIIB) and activating (FcγRI, FcγRIIA, and FcγRIIIA)[44]. FcγRI generally binds to the monomeric IgG where as other FcRs binds to the multimeric antigen antibody complex with the specificity of subtypes.[44] Decreased galactosylation and increased core fucosylation is an indication of proinflammatory response in the immune system.

1.4 Glycan in virus entry to the cell.

Virus entry to the host system require specific binding capacity of virus to the host cell receptors termed as tropism. The structural sialylation is the one type which commonly exemplifies with different linkages will also influence cell-pathogen interaction. As a terminal monosaccharide residue, Neu5Ac is usually linked to Gal via α2,3-linkage or α2,6-linkage. The infection of viruses starts with the specific binding with host cells. For example, the binding of influenza A virus needs sialylated glycans as receptors on the surface of host cells. Different types of influenza A viruses have different binding specificities. Avian influenza A viruses preferentially bind to respiratory epithelial cells that contain α2,3-linked Neu5Ac moieties, while human viruses selectively bind to α2,6-linked Neu5Ac moieties.[45] Also, for enterovirus and arbovirus, they only recognize α2,3-linked Neu5Ac moieties of the host cells.[46–49] . Similarly, the host cells also carry many glycan receptors which can

specifically attach to the surface viral glycan during its entry. Many of the viruses including human pathogenic viruses, , bind to this host cell glycan receptors during their entry.[50] The specific glycans on the cell surface such as glycosaminoglycans, heparin sulphate sialic acid or galactose can serve as a receptor in virus entry to the cell. Hence the specific glycan act as a key for virus attachment and entry to the cells and there by decide the pathogenicity, transmissibility or cell tropism of the virus.

1.5 Gene therapy viral vectors

Gene therapy is therapeutic modification, introduction or removal of a genetic material to treat a genetic disorder or a disease.[51] Central point of gene therapy is the efficient delivery of DNA or RNA to the host cells. Vectors are the "molecular cargos" used to deliver the genes. There is no universal vector used for treating all diseases. The viral vectors are designed to deliver the nucleic acid in applications to gene therapy is one of the most encouraging approach in in recent years because of their immunosurveillance of the infected host system. The normal retrovirus[52], adenovirus[53], herpes simplex virus[54] and adeno-associated virus (AAV) are scientifically modified to fit for gene therapy use. AAV is an important principle gene therapy vector especially to transfer antisense RNA and ribozyme genes in pre-clinical cancer models.[55]AAVs can be used to treat some diseases, such as LPL deficiency[56], LCA (eye)[57]and hemophilia A.[58]

1.6 Importance of glycosylation analysis

The analysis of glycoprotein can be challenging due to several reasons. Primarily, the abundance of glycoproteins in a complex mixture can be relatively low. Due to the high complexity of the process and structure, it is hard to understand the glycosylation sequences, structural conformation, the isomeric pattern of monosaccharide building blocks, distinguishing their type from the same mass pattern, type of modifications, and percentage of changes using a single analytical strategy. The analytical complexities increase as the

complexity of glycosylation increases.[59] Mass spectrometry is the pivotal technique that is used to understand the details of these complexities using the combination of different methodologies. The 4S features of mass spectrometry like sensitivity, specificity, stoichiometry, and speed, make the technique compatible with characterizing the challenging features of glycoprotein.[20,60–64]However, the complete characterization requires extensive sample processing and multi-level analysis, including released glycan analysis, glycopeptide analysis(bottom-up), and glycoprotein analysis(top-down analysis), as we explained elaborately in ***figure1.2.*** The glycoproteins/glycopeptides are less efficiently ionized in comparison to the non-glycosylated version. Hence their detection is hindered in the presence of non-glycosylated peptides/proteins and must be enriched before the analysis to overcome this issue. The main challenge in glycosylation analysis under mass spectrometry is related to their structural similarity. Unlike amino acids, the different monosaccharides can have the same masses and highly similar structures. Hence structural elucidation is highly challenging.

The commonly employed technique for a detailed analysis of glycosylation study is a bottom-up technique. Wherein an endoprotease like trypsin is used to digest the protein to get a peptide pool, and glycopeptides are further enriched from this mixture. The tandem mass spectrometry is used to get structural details of glycopeptide. The analysis of glycopeptide tandem mass spectra is comparatively hard to comprehend. Here we get a mixture of fragment masses that contain both amino acid and oligosaccharide sequence information. The lack of entirely dependable software makes this analysis still complicated. However, recently developed software like pGlyco,[65,66] massytools,[67]GlycopepID,[68] Proteome Discoverer (Thermo Scientific, MA, USA), Glycomaster DB,[69] Byonics (Protein Metrics, CA, USA), GlycopepEvaluator[70] and Glycoproteome Analyzer (GPA)[71] could solve this issue to an extent with a high speed mapping of glycopeptide and false discovery rate of $\leq 1\%$. The

software is also referred to the specific enzyme existence in the organism, and already published structures related to an organism therefore species level filtration can also be possible. All these techniques are selectively used in different aspects of our studies in the biomarker discovery and rAAV glycosylation detection.

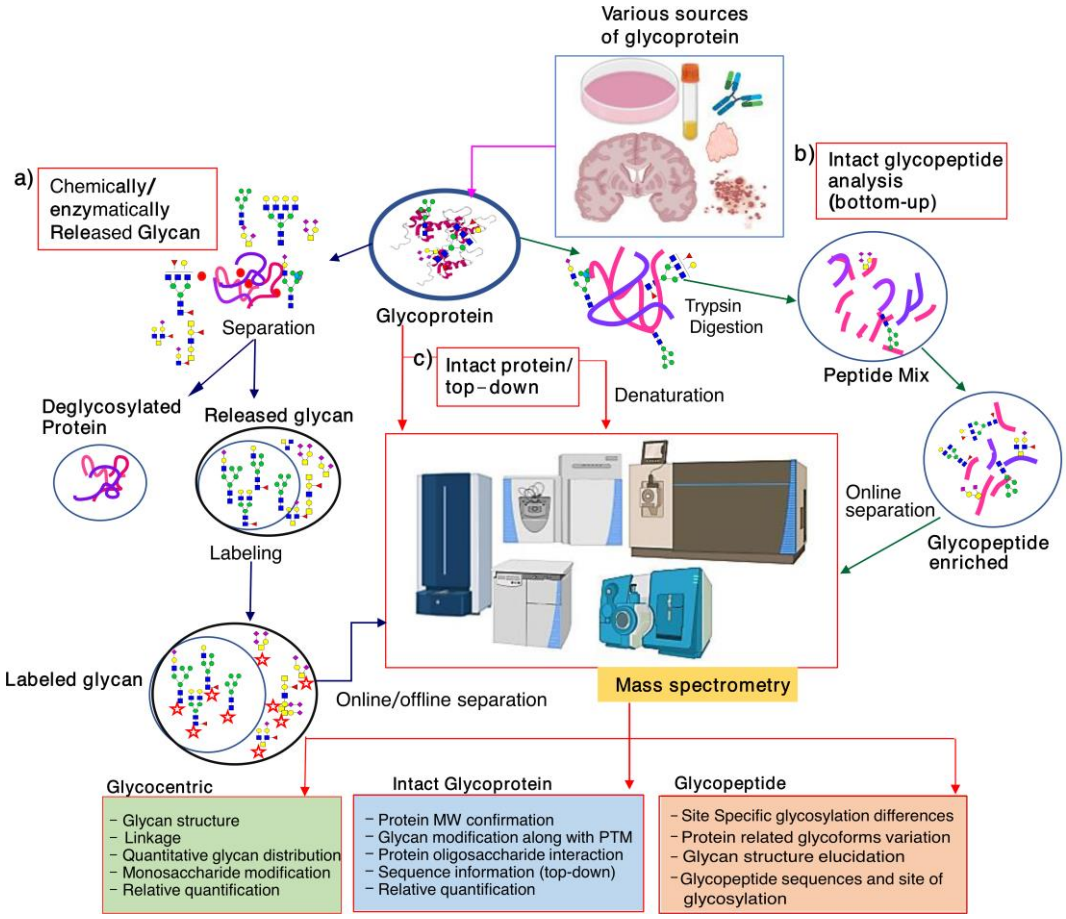


Figure 1.2:Site-specific glycan heterogeneity
A schematic overview of glycoprotein characterization using mass- spectrometry. a) glycocentric analysis where chemically/enzymatically released glycan get purified, derivatized/labeled separated and analyzed. b) glycopeptide analysis including denaturation of purified glycoproteins, trypsin digestion, enrichment and separation. c) Intact glycoprotein protein analysis after purification from complex sample.

The fundamental drawback of the bottom-up approach is the loss of information during the extensive purification, enrichment, and MS fragmentation. However, it is essential to fill the knowledge gap in each level that can be bridged with the intact glycoprotein, glycopeptide, and glycomics level information obtained by mass spectrometry.[72]

1.7 Glycocentric analysis

The glycocentric analysis is generally focused on the structural elucidation of released glycan in detail, including composition linkage and isomeric pattern of the attached oligo saccharide, their monosaccharide modification, overall distribution, and relative abundance in the sample. The analysis includes glycan release from a complex glycoprotein mixture by enzymatic or chemical method. Their purification, labeling, enrichment and analysis. The detailed workflow is illustrated in ***figure1. 3.***

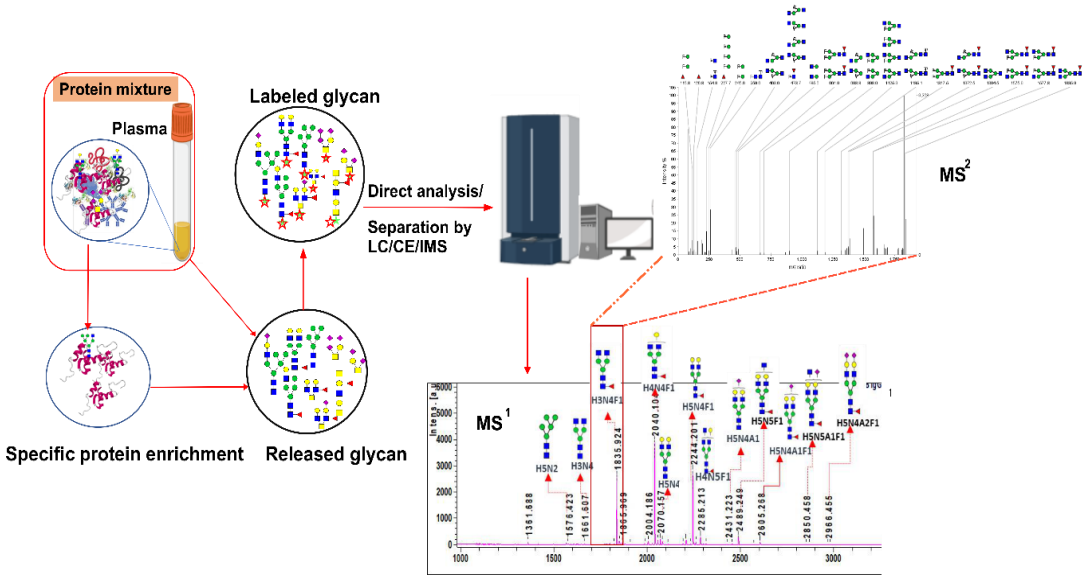


Figure 1.3:The schematic representation of release glycan analysis by MALDI-MS. The MS^1 spectra showed different glycoforms identified from the sample. MS^2 spectra showed the structural analysis of H3N4F1 sugar by tandem mass spectrometry. The structures are assigned with the help of Glycoworkbench software.

1.7.1 Glycan release

N-glycans are usually released from glycoproteins or glycopeptides by PNGase F[73] before analysis. The conventional method takes a longer time with overnight incubation period and laborious clean-up protocol. Recently Waters and Prozyme introduced an N-glycan analysis kit protocol in which they claim a rapid release of less than 10mins.Waters introduced a surfactant called RapiGest (Waters, Milford, US) which is used to denature protein molecule at higher temperature (~95ºC for 2 mins). RapiGest aided protein

denaturation makes the N-glycans accessible to rapid PNGase F (GlycoWorks, Rapid PNGase F, Waters) and helps to efficiently deglycosylate the protein at 50ºC with less than 10mins of incubation. Song et al. in 2016 developed a rapid chemical method to release N and O-glycans rapidly from different glycoproteins called oxidative release of natural glycan (ORNG)[74] by NaClO. The method is a simple and efficient approach for releasing glycans from glycoproteins and glycosphingolipids in a cost-effective manner. The O-glycan release includes β-elimination,[75–77] classical hydrazinolysis[78], O-Glycome reporter/amplification (CORA)[79] are also used as general release of O-glycan. Here we use the conventional PNGase F mediated N-glycan release. Enrichment of free glycoforms is normally accomplished by solid-phase extraction (SPE) with C18 cartridges or non-porous graphitized (non-PGC) carbon columns.[80]

1.7.2 Labeling methods of released glycan

Unlike peptides, glycans are having unique physicochemical properties including highly polar nature with poor UV absorption. They also lack a fluorophore or chromophores in their structure. These unique properties make them less ionizable in mass spectrometry as well as poorly detectable in UV or fluorescent-based standard detectors in their nascent form. Unlike nucleic acids, there are no ways to amplify glycan as we use PCR for DNA/RNA. Hence it is essential to label them before MS analysis to overcome the weak ionization nature also make them compatible with UV or fluorescent detectors by coupling with suitable tags. Therefore, different labeling strategies have been employed worldwide, based on the researcher's convenience and analytical methods employed. The most used methods are permethylation, reductive amination, hydrazide labeling, and Michael addition. The ambiguity due to the diversity in methods used by different labs is the main disadvantage when it comes to comparing the data between each lab. However, in our lab, we have applied the two widely used labeling methods in different experiments. Additionally, we developed a

novel labeling technique for carbohydrate labeling which is compatible with the existing method in a rapid and easy way.[81] The method is also described in detailed in **section 1.7.2.2.**

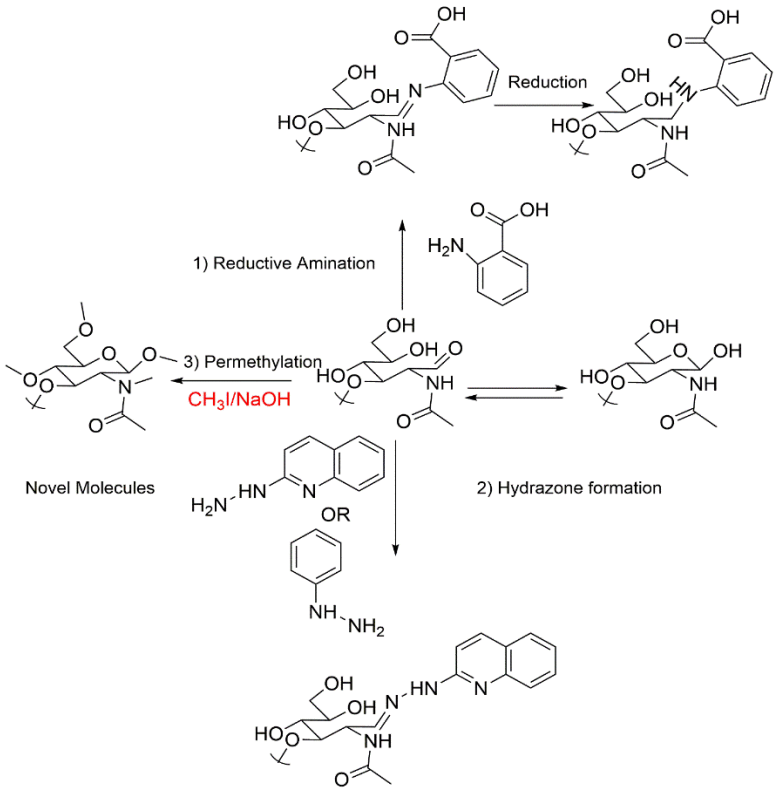


Figure 1.4 Commonly used glycan labeling reagent and their mechanism. Reductive amination is formed by a stable bond formation between the reducing terminal of sugar (1). Hydrazino formation (2) is also formed at the reducing terminal like 2-hydrazino quinoline or 2-hydrazino pyramidine through hydrazine bond. The third strategy is permethylation where every single hydrophilic groups, get methylated.

1.7.2.1 Reductive amination

Reductive amination is one among the most powerful methods for labeling the released glycan at their reducing end. The glycans with reducing end react with the primary amine or Schiff's base-containing labeling agent like 2-amino benzamide or 2-amino benzoic acid to form an unstable imine which get reduced by a reductant like sodium cyano borohydride to form stable secondary amine, (**figure1.4(1)).**[82] 2 -amino benzamide (2-AB) and 2-amino benzoic acid (2-AA) labeling is not only enhances the ionization in mass

spectrometry but it gives the fluorescent absorbance to the glycan which can uses HPLC-FD based separation and analysis. 2-aminopyridine (PA), and 2-aminonaphthalence trisulfonic acid((ANTS) and 1-aminopyrene-3,6,8-trisulfonic acid (APTS) also used for the same purpose.[83] The reagents like1-phenyl-3-methyl-5-pyrazolone (PMP) can also use in alkaline conditions, for labeling glycan at the reducing end. The method is known as Michael addition.[84] The labeled glycan can be purified by L/L extraction or SPE based purification.

1.7.2.2 Hydrazone formation

The labeling through reductive amination is efficient but time-consuming due to the extensive desalting process before MALDI analysis. In our lab, we introduced two novel reagents;2-hydrazino quinoline and 2-hydrazino pyrimidine for labeling the glycan through Schiff's base formation, which effectively enhances the signal intensity in MALDI without an extensive sample cleanup procedure. Both the reagents containing an electron-withdrawing group 2- pyridine(or quinoline,2HQ), or 2-pyrimidine (N-heterocycle, 2-HPM) can derivatize oligosaccharides at their reducing end through a stable Schiff's base within 15 min, and the derivatization can provide their better ionization in MALDI. Additionally, 2 HQ and 2-HPM can act as co-matrix or matrix, which enhances the signal intensity of the oligosaccharides in MALDI, and hence the tedious purification and desalting steps before MALDI can be avoided. We have detailed the tagging mechanism in the published literature. For 2-HPM mediated labeling, methanol with 10% glacial acetic acid and 10mg/ml reagent at 60-70 degrees for 10mins is the ideal labeling conditions.[85]

Each time we have used the ten picomoles of standards DP5 or dextrin. The method optimization was performed by varying the reaction media, type of solvent, reaction temperature, the concentration of reagent. The optimum temperature of the reaction was 37-60-degree celsius temperatures. The 2-HQ mediated labeling is further optimized with varying reaction media. The reaction media was formulated with 10% DMF and 10% glacial

acetic acid in ACN, and the reaction time was 60 minutes to 70 minutes, which is higher than 2-HPM. Each point of method optimization one parameter kept varied, and others kept constant. Derivatization rate was calculated by dividing the peak intensity of 2-HQ-derivatized DP5 by peak intensities of both underivatized and 2-HQ-derivatized DP5. The reactions were conducted in triplicate, and the average values and standard deviations were determined for each condition. The additional property of 2- HQ is its fluorescence. This can be utilized for HPLC -FD based glycan analysis as we used in the reductive amination method with 2AA or 2 AB. Since it is a novel method, the optimum fluorescent condition was set with fluorometry, and the λex was observed to be optimum at 280nm and λem at 335nm (The graph is shown in Appendix A F2). The fluorescent property was compared with conventional 2-AB by labeling one nanomole of dextrin, and injected 500picomoles in each reaction and compared their results. We have received the same fluorescent response by both the runs; hence, it is an efficient fluorescent label (Appendix A F1). The additional advantage of 2-HQ is that it can be ionized well in MALDI and quiet stable at tandem mass spectrometry. The details are shown in Appendix A F4. The 2-HQ can be used as a matrix also. The HQ labeled glycan can ionize in MALDI-MS without mixing with convention DHB mixing (Shown in Appendix A F5). Hence our inhouse developed method can also be used as an efficient and alternative labeling strategy.

1.7.2.3 Permethylation

Permethylation is widely used in mass spectrometry-based glycan analysis for enhancing the ionization potential of glycan. Hence the method is routinely performed for released glycan prior to the MALDI-MS analysis. It has number of advantages including 1) it is known to enhance the intensity of the glycans than the underivatized oligosaccharides.2) stabilize the liable sialic acid in the glycan moieties.3) The sialylated glycoforms are less efficiently ionized in mass spectrometry in ‘pos’ mode. By this method, all reactive glycan

hydrogen group like -OH, -NH and -COOH, in the glycan structures get converted to the methyl group- -OCH3, -COOCH3 and -NCH3-respectively. The charge related issues in mass spectrometry also get resolved by this method.4) better determination of the linkages and branching of different glycosidic linkages in the structure 5) conversion of hydrophilic glycoforms to hydrophobic structures which allows C18 based separation and better ionization and flying in the gaseous phase[86].Also, permethylation based MALDI-analysis offer a faster profiling than the HPLC /UPLC based techniques. ***Figure 1.4(3)*** showed the mechanism in which the simple glycan dissolved in DMSO get reacted to methyl iodide (CH3I), catalyzed by methyl-sulfinyl carbanion prepared by the NaOH in DMSO. The bases can also vary from NaOH to KOH or LiOH to form, the carbanion reagent for the reaction.[87] The permethylated hydrophobic glycans get enhanced in their ionization, which can increase the detection sensitivity of glycans in MS, including both ESI-MS and MALDI-MS.[88] The analysis is carried out in positive mode, and ions are formed in $M+Na^+$ form. Conventional permethylation can introduce an excess of salt in the mixture, which usually can purify using liquid extraction with chloroform or DCM and water.

Extraction is utilized to remove salts generated during permethylation. The method is effectively used in relative quantification and various glycan-based biomarker discovery.[89] The method is highly suitable for detection and tandem based mass spectrometric analysis due to its high signal responses. Nevertheless, it is limited for detailed analysis like linkage specific sialic acid analysis. The specific modifications can overcome the limitation of permethylation and bring up this method more suitable for the analysis and quantitation of certain sugars. The anhydrous condition and using carboxylic acid activator EDC, with the catalyst HOBt carboxylic acid get activated and binds to the adjacent OH group of galactoses to undergo a lactone formation and hence unable to label with dimethylamine. If the glycan is treated with dimethylamine in the presence of EDC and HOBt, the alpha 2,6 sialic acid gets

labeled with dimethylamine and gets modified.[90] Hence both the linkages make a 13Da difference. The mechanism is detailed in figure 1.5. The standard purification method is liquid-liquid extraction with Chloroform or DCM with water.

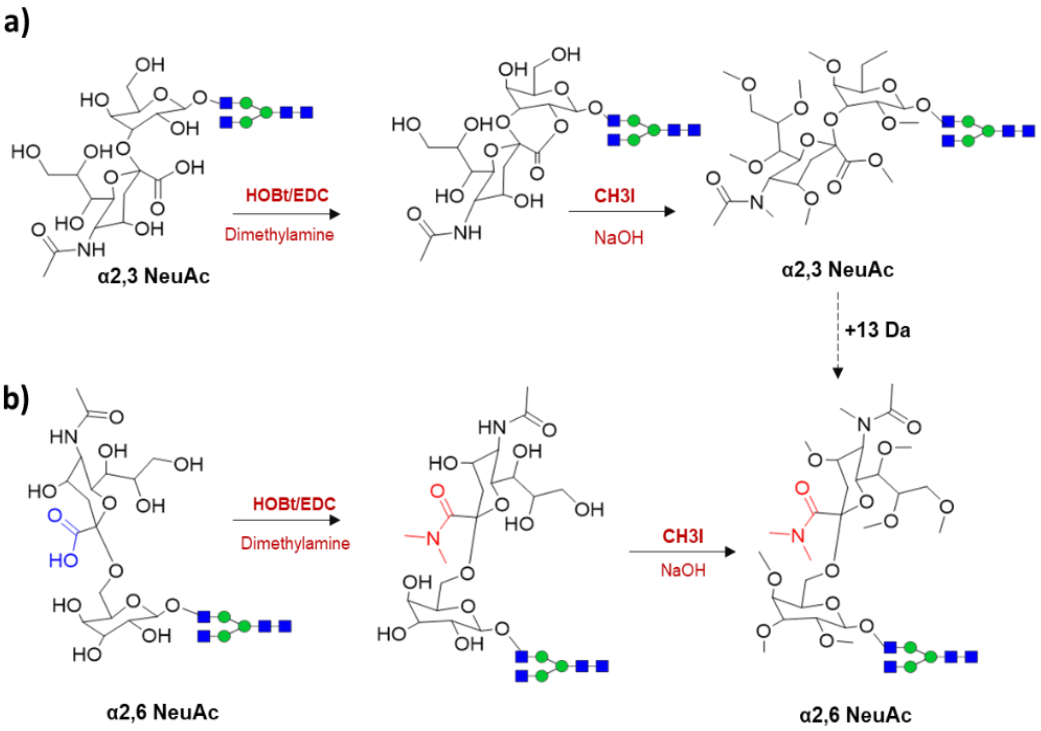


Figure 1.5:The linkage specific permethylation of alpha 2,3 and alpha 2,6 sialic acid. The alpha 2,6 linked sialic acid get activated and linked to dimethylamine made 13Da increase linkage specific analysis.

1.8 Glycopeptide analysis

The glycopeptide analysis is much more informative in comparison to glycocentric related to the functional aspect. In the glycopeptide analysis, we can get the site-specific information about the glycosylation, which allows understanding their functional aspect

related to the specific proteins. The complex proteins are isolated from the biological samples and denatured followed by endoprotease (generally trypsin) digestion to generate the peptide mixture. Since the natural abundance of glycoproteins is very low in the mixture, it is essential to enrich the glycopeptides prior to MS-based analysis.

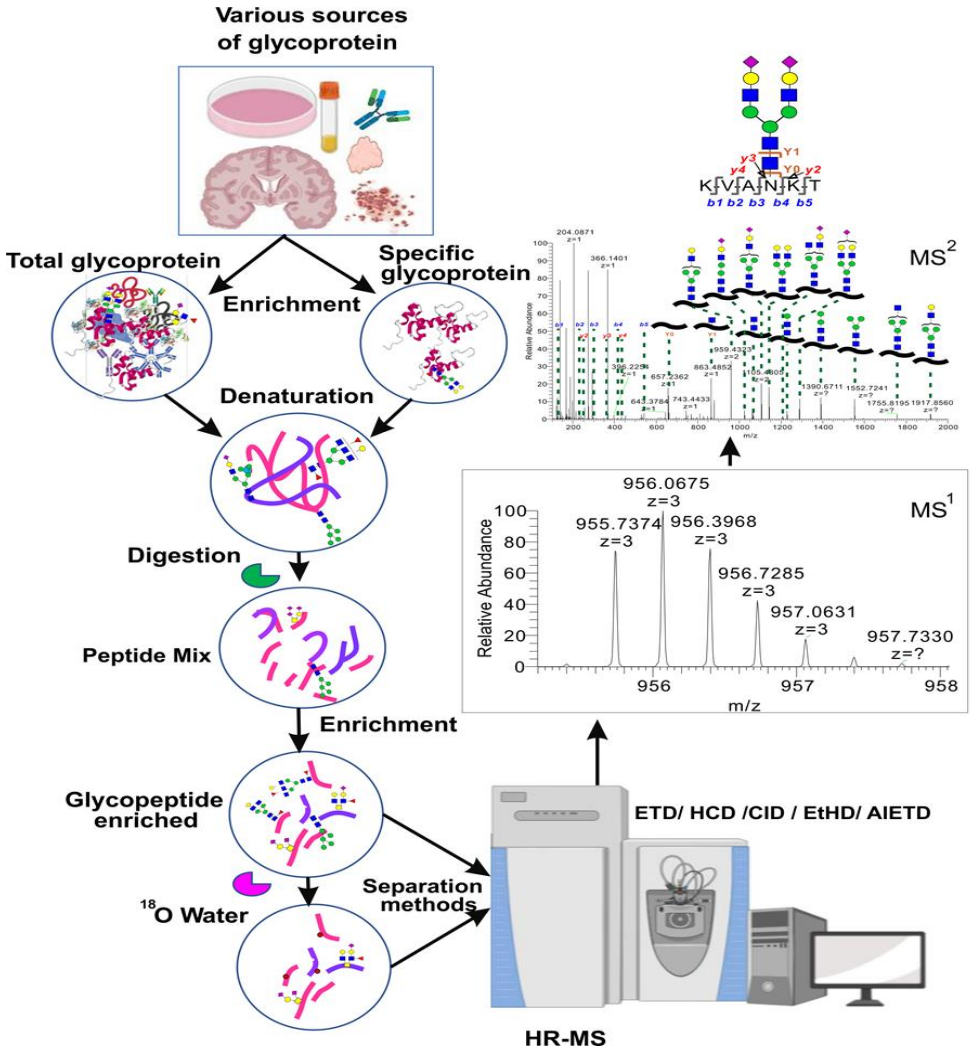


Figure 1.6 Bottom-up approach for glycoprotein analysis.
The workflow details the steps involved in the process and spectra showed the interpretation of the glycan structure from a glycopeptide.

1.8.1 Glycopeptide enrichment

The natural abundance of glycoproteins is generally very low, and if the sample is a complex mixture, the sensitivity of the analysis goes down in mass spectrometry. A practical approach to enrich glycopeptides is essential prior to mass spectrometry-based analysis. There are multiple strategies developed to enrich the glycopeptides, including chromatographic separation, biological affinity-based capturing, chemical-based capturing, enzymatic labeling, and metabolic labeling are the widely used enrichment strategies. Due to the highly polar -OH, COOH, and NH groups on the glycan structure, it makes it efficiently separable using hydrophilic interaction liquid chromatography (HILIC). The main adsorption force is H-bonding for HILIC based separation.[91] The method can also use for the enrichment of glycopeptides.[92],[93] The HILIC medium is usually Silica, cellulose, or sepharose.[92],[91],[94]The functional group attached to the basic medium can also be used for HILIC enrichment. The common functional groups are amide[95], maltose[96], diol[97], and cyclodextrin[98] are efficiently used for HILIC based glycopeptide enrichment. The modified technique like zwitterionic HILIC (ZIC-HILIC)[91],[99] and weak anion exchange-based electrostatic repulsion hydrophilic interaction chromatography (ERLIC)[100,101] also used for the glycopeptide enrichment. Specific glycan binding proteins like lectins are used for enriching a sugar-based protein and peptides.[102–104]Chelation interaction chromatography used for enriching the glycopeptide based on their charge.[105],[106],[107] Metabolic labeling mediated specific sugar enrichment is also an alternative glycopeptide enrichment method used mass spec-based glycoprotein analysis.[108]

1.8.2 Glycosite detection

The analysis of intact glycopeptide is also challenging due to their complexities. However, the detection of N-glycosylation site on the peptide is relatively simple by specific-labeling method. The PNGase F enzyme is generally used to release the N- glycosylation from

the protein. The enzyme can cleave the Asn-linked glycan by hydrolysis thereby causing a deamidation of Asn to Asp. Thus the deglycosylation can cause an addition of 0.9858 Da to the peptide at the glycosite[73] and the difference can be picked up by HR-MS. However, the deamidation of Asn residue at higher pH (more than 8) is also a possible modification (4-9%) during the trypsin digestion. Hence the procedure may lead to a false discovery of glycosylation site.[109] However, the process can be modified to label the N-glycosylation site when we carry out the reaction in the presence of ^{18}O-water.[110] The labeling by $H_2{}^{18}O$ can lead to a mass shift of 2.9883 Da, which can be discriminated from the normal deamidation[111] as shown in **figure 1.7.**

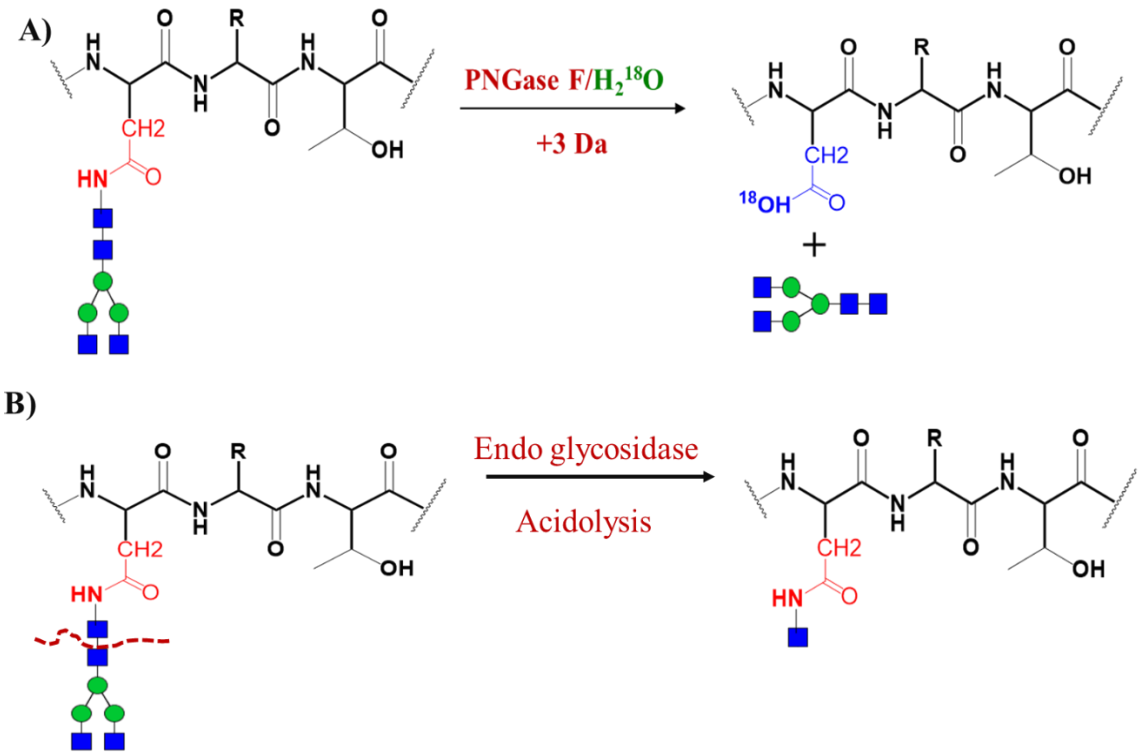


Figure 1.7: The common glycosite labeling mechanism
A) PNGase F mediated site-specific labeling of N-glycopeptide by ^{18}O-water. The partial cleavage of the glycan to leave innermost 'Asn' to the site for glycopeptide labeling

The alternating methodology for site specific labeling is possible by partially cleaving N-glycan to generate a fixed modification with GlcNAc on the N-glycosite. (Figure 1.7 B). Different Endoglycosidases like Endo H can be used to for this purpose. It breaks down the

glycoside bond between the innermost GlcNAc which binds to the ‘Asn’ and the glycan core[112], thus it can be distinguished with mass spectrometry by the addition of ‘GlcNAc’ mass to the peptide. The method can also be achieved chemically by acidolysis. The differential strength between amide and glycosidic bond can be used for the differential cleavage. The innermost GlcNAc is linked to the peptide by amide bond whereas all other sugars are linked together by glycosidic bond. Hence the relatively strong amide bond does not cleave by this method whereas glycosidic bonds get cleaves off. Microwave energy is also used to enhance the process.[113]

1.8.3 Mass spectrometry techniques for glycan analysis

The advanced mass spectrometry-based techniques with high sensitivity and compatibility are widely used to break down the complexities of the glycoprotein analysis. Mass spectrometry mainly consists of four parts which includes an ionization chamber to induce the molecular ions in gas phase, analyzer to sort the ions formed by means of time or space, detector to detect the resulting ions data analysis and a Data processor/digitizer.

Electron ionization/ electron bombard ionization causes the loss of electron in the glycan molecule which eventually help to fragment. The EI was the primordial technique to understand the monosaccharide composition of the sugar. The combination with GC/MS was efficient to analyze the sugar molecule. However, larger molecule like N-linked glycans are difficult to vaporize into the gas phase due to their high polar nature which has to be converted to permethylated glycoforms and also applied to analyze the monosaccharide analysis of the glycan after acid hydrolysis and derivatization.[114] Fast atom bombardment (FAB)was the first efficient conventional technique with gentle ionization technique to analyze negatively charged as well as neutral glycan molecule without insource fragmentation.[114]However the technique had low sensitivity and high background masses

from the matrix, which inhibit the efficient analysis of glycan from the mixture. The technique soon replaced by MALDI and ESI.

1.8.3.1 MALDI

The MALDI and ESI are the most efficient ionization techniques for analyzing the glycan molecules to date. MALDI has the advantage of gentle ionization method which keeps the oligosaccharide structure intact and ionization efficiency remains constant regardless of the size of the oligosaccharide.[114] For MALDI analysis, the glycan sample is mixed with the matrix typically as the low molecular aromatic acids like 2,5 dihydroxy benzoic acid (2,5 DHB) and allowed to co-crystallize on a MALDI-plate. The plate is then introduced into the machine for analysis. The sample spots are irradiated using a UV-laser. The matrix absorbs the energy from the laser and then transfers a small amount of energy to the analyte molecule makes them ionize??. The main disadvantage of the technique is poor sensitivity due to low ionization efficiency of carbohydrates. Some sugars like sialylated sugars are thermal labile and may degrade during the analysis. Hence, the mass obtained by MALDI- analysis alone may not be enough for identifying the structure of sugars. Normally we must confirm the identity parallel by complementary analytical techniques. The sample derivatization like permethylation can stabilize the sialic acid linkage and overcome the disadvantage mentioned above. The MALDI-is commonly interfaced with the time of flight (TOF) or TOF/TOF analyzer for fragmentation analysis.

1.8.3.2 ESI

Electron spray ionization is a gentle method of ionization that keeps oligosaccharides intact. Commonly ESI is coupled with the separation technique like HPLC, ion mobility or CE prior to the mass spec detection. The spectra are formed by spraying the glycan in liquid phase through a narrow orifice under the influence of a strong electric field(1-3kv).Difference in the potential applied between the tip of the needle and capillary is coupled with very high

temperatures, that causes finely charged droplets. The inert gas used to vaporize the solvent as they are aspirated into mass spectrometry. Both positive and negative charged ions can be produced by this method and multiply charged ions are commonly observed. Depending on the solvent used, their additives and source conditions there can be different adducts and type of ions can be formed. However, the ionization efficiency decreases as the glycan masses increases which makes the relative quantitation difficult by this method. Hence the spectra are much complicated than the MALDI-profile. The glycopeptide analysis was efficiently performed by ESI-MS

1.8.3.3 Mass analyzers

The ion separation in the early stage was performed with magnetic field sector instrument. The instruments are huge and limited to certain mass ranges, however, could produce accuracy in the mass measurements. Fourier-transfer ion cyclotron resonance (FTICR) uses a high magnetic field along with the alternating voltage to trap the ions with high resolution. In quadrupole mass analyzers, the ions are separated using a low potential and passes through centroid rods using alternating and static voltages. The linear ion trap instruments confine the ions by applying voltage at the end of rods. Paul trap is a type of quadrupole ion trap consists of two hyperbolic electrodes with their foci facing each other opposite side of a hyperbolic ring electrodes. The oscillation (RF) and static current(DC) traps the ions between the three electrodes.[115] Orbitrap analyzer provide high resolution by confining the ions by oscillating them around the central rod.[116] The Orbitrap also use FT based technique to trap the ions. However, the trapping is performed electrostatically (as opposed to magnetically) and the frequency oscillations are measured along the long axis of the trapping cell.[117] The ions are trapped in radial direction between the inner and outer electrodes by applying a DC voltage applied.

1.8.3.4 Structural analysis of carbohydrates by mass spectrometry

The compositional analysis of carbohydrates is possible by analyzing the molecular weight with the component mass. The method is useful for the purity determination or primary detection of the known compounds. The high-resolution mass spectrometry-based glycosylation analysis with tandem mass analyzers creates fragments from the peptides and glycans simultaneously. There are many fragmentation methods in MS-based glycopeptides analyses. The common ones are collision-induced dissociation (CID)[118], electron -transfer dissociation(ETD)[119] and high-energy C-trap dissociation (HCD)[120]. CID is compatible with many commercial instruments with various analyzers, including ion-trap [121] ,Q-TOF[122], triple quadrupole[123], TOF-TOF, and other hybrid MS.[124]The CID based fragmentation generally create B and Y ions[125]as shown in ***figure 1.8B***. Since the glycosidic bonds are liable than the amide bond, hence rarely produce enough b and y ions at normalized collision energy (NCE)as shown in ***figure 1.8A*.** The insufficient b and y ion generation in CID is overcome by using stepped energy collision in Q-TOF and TOF- TOF instruments. The other disadvantage of CID based method is the low molecular cutoff in CID (<300 Da) used in the detection. The advanced mass spectrometry like orbitrap instrument also used an advanced CID fragmentation method named HCD where the fragmentation takes place at collision cell instead of ion trap as in CID. It is a beam typed CID with the detection of fragment ions coupled with orbitrap analyzer.[126] HCD based fragmentation is performed with high energy than the conventional CID method to allow multiple fragmentation pathways.[127] HCD produce a higher number of oxonium ions from the glycan part than the ‘b’ and ‘y’ ion from the peptide part. The usage of SCE-HCD can improve the quality of MS^2 Spectra in glycopeptide analysis,[65] also permits low mass fragmentation detection without any low molecular cut-off mass. The orbitrap based analyzer improves the quality of the MS/MS spectra.[127]The ETD fragmentation is widely used to detect multiply charged glycopeptide by

cleaving between the N–Cα bond to produce c- and z-ions[128] (**figure 1.8A**) and localize the glycosylation site.[129] The glycan portion remain intact without fragmentation and widely used to fragment multiply charged long glycopeptides.[130] In our study the glycopeptide analysis were performed by HCD and glycosite analysis were performed by CID method using orbitrap Elite Velos hybrid high resolution mass spectrometer. Whereas the glycocentric analysis were performed using MALDI-TOF-TOF-MS.

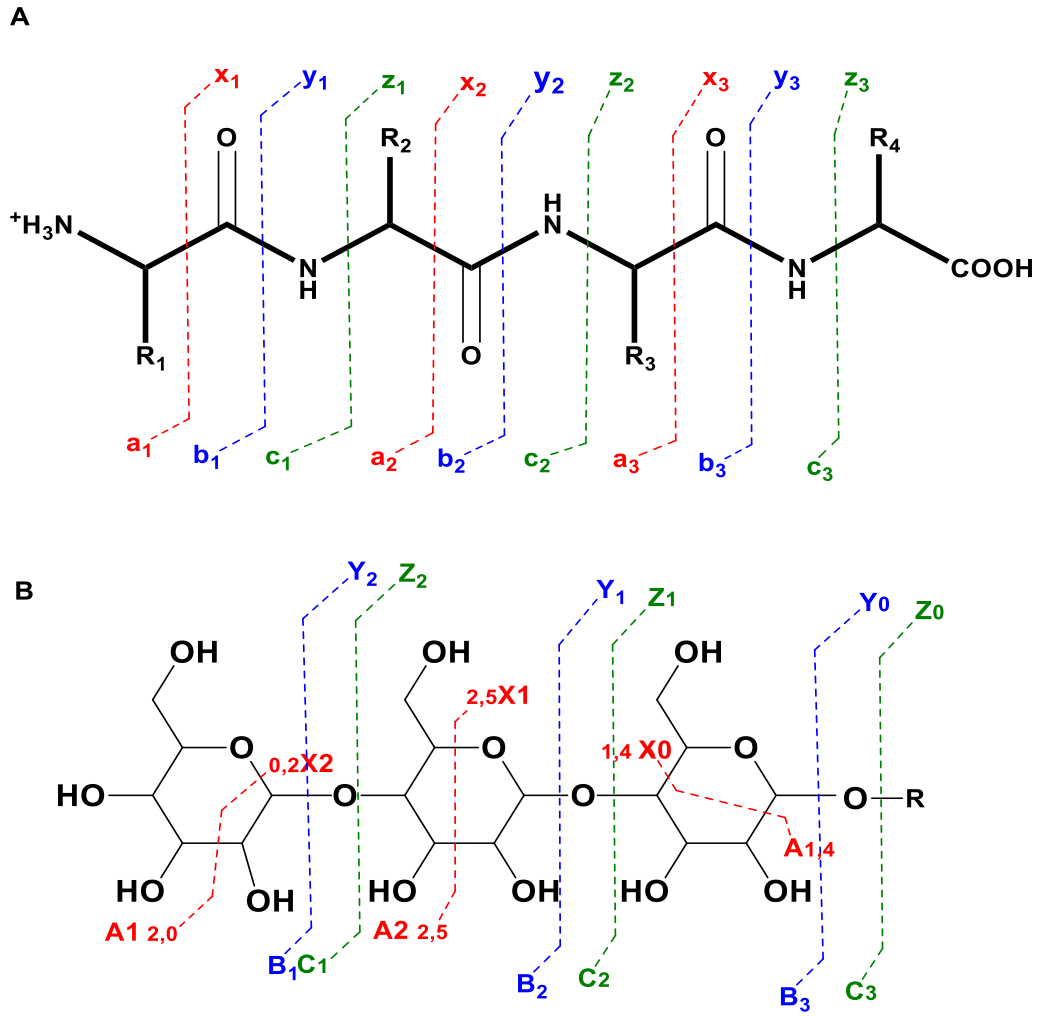


Figure 1.8:Fragmentation pattern of glycopeptide in mass spectrometry
A The fragmentation pattern in peptide and illustration of possible fragment ions. B) The possible fragment ions from glycan moieties.

1.8.4 Fragmentation pattern of glycoforms

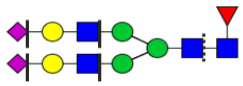
The structural details of carbohydrates by obtained by tandem mass spectrometry based on the CID fragmentation. The fragmentation pattern of the sugar labeled through

reducing terminal (reductive amination, hydrazine formation) generates B and Y ions in CID, and it is easy to interpret the monosaccharide residue by calculating the fragment ions using the formula Molecular weight-18 Da. The same type of breakage can occur any anomeric site on the other side of 'O' atom to form C and Z ions. Higher energy fragmentation can also create the cross-linked fragmentation and resulting A and X ions (Figure 1.8b). Each of these ions can be utilized for predicting sequence of the monosaccharide linkages and detecting the site-specific modifications. Monosaccharide stereochemistry and anomeric confirmations cannot be elucidated by interpretation of MS/MS data alone. To elucidate the accuracy higher status of tandem based analysis is required in conjunction with the standards and databases. The de novo structural elucidation of carbohydrate involves breaking oligosaccharides to overlapping segments of disaccharides. The data is acquired by successive rounds of CID. In this process sample is injected to the mass spectrometer to produce parent ions. The parent ions undergo fragmentation to produced disaccharides which matches to the fragmentation pattern in the databases. By comparing the disaccharide fragmentation with the standard sugar and their linkages the structures can be predicted by this method.[131] Overlapping disaccharides are piece packed together to predict the parent ion structure. Permethylation and per acetylation is widely used to increase the ionization efficiency of the glycoforms. Similarly, to the native structural ions, permethylated glycan also causes a unique signature for each monosaccharide residue, and their fragmentation is also unique at each terminus. Because of the experimental procedure and use of sodium hydroxide the theoretical mass of the permethylated sugar ions will always be in the sodiated form.[132] Hence the mass is calculated using the formula

$$\text{Theoretical mass(M/Z) of permethylated sugar} = \frac{\text{Mol. Wt} + Na^{+} \times \text{Charge state}}{[\text{Charge state}]}$$

The fragmentation pattern of permethylated N-glycoform[133] is listed in Tglable1.1 with the example of G2fS2(H5N4A2F1/54201)

Table 1.1 The fragmentation pattern of permethylated glycan



Fragmentation terminal	Structure of the fragment ions	Mass of the fragment ion (Da)
Reducing end	-GlcNAc-O-CH3	276
	-GlcNAc(-Fuc)-OCH3	451
Internal fragments	-Hex	204
	-GlcNAc	245
	-Gal-GlcNAc	449
Non – reducing(terminal) end	NeuAc	375
	Hex	219
	GlcNAc	260
	Hex-	235
	NeuAc-Gal-	596
	NeuAc-Gal-GlcNAc	824

The combination of all these techniques are used to identify, characterize and correlate the glycosylation pattern of different biological samples in our study. The Glycoworkbench software is used to predict the structure and fragmentation process.

1.9 Purpose of the study

As an essential post-translational modification, glycosylation can regulate the biological activities in the cell and body. The glycosylation can regulate the signaling

pathways, pathogenesis[17], disease progression[134] and immunogenicity[135] in biological system. Hence the health of an individual is widely dependent on perfection in the glycosylation process. The comprehensive analysis of glycosylation is essential to understand the perfection of glycosylation and their dynamic modification and aberration related to disease progression. In chapter 2, we are trying to understand the glycosylation related biomarker discovery of hemophilia A patients with inhibitor development and their profile variation in comparison with negative inhibitor hemophilia A patients. Immunoglobulin glycosylation regulates the effector functions and can be an indicator of the immune health of an individual. In this study, the plasma glycosylation profiling of inhibitor developed patients is widely studied for comprehending the root of inhibitor development. With the help of mass spectrometry techniques, we tried to understand the differences in glycoprotein expression of the patient plasma, their site-specific glycosylation differences, and the overall glycan profile of the plasma and antibody isolated from the plasma is monitored. Inhibitor positive and negative samples are compared after treating with emicizumab, a bispecific antibody used to treat for hemophilia A. Mass spectrometry-based glycan difference analysis has wide scope of analysis in biomarker discovery as well as personalized medicine development. The glycosylation variations can be linked with multi ‘omics’ studies-proteomics, genomics, transcriptomics epigenomics, lipidomic and metabolomics. We tried to understand the plasma glycosylation pattern change with the inhibitor development and after treating with emicizumab. Here we explain the conducted study related to inhibitor development in the mouse model to compare the glycosylation variation with humans. We also studied the plasma glycosylation pattern of HA mice from three different strains- BALB/c, BL-6, and SV129 -before and after rFVIII injection and inhibitor development.

The third chapter discusses the glycosylation analysis on a gene-therapy vector rAAV. Adeno associated virus (AAV) is a gene delivery tool, which has been approved as a gene therapy vector for combating genetic diseases. AAV capsid proteins are the major components that determine the tissue specificity, immunogenicity, and in vivo transduction performance of the vector. The glycosylation can be examined for AAV8 neutralization; altered receptor binding improved cell ingression and tropism in a mixture of various tissues. In this study, the AAV2, 8serotypes are monitored for the capsid glycosylation profile was systemically analyzed by peptide mass fingerprinting utilizing high-resolution mass spectrometry to determine the presence of capsid glycosylation and capsid- HCP interaction.

2 CHAPTER 2: BIOMARKER DISCOVERY IN HEMOPHILIA A INHIBITOR DEVELOPED PATIENTS

2.1 Abstract

Hemophilia A inhibitor development is a life-threatening challenge in the field of health care. The inhibitor development may relate to the immunological recognition or progression contrasts in the population. Glycan plays an essential job in the immune functions of the body, and their dynamic changes can be related to the disease progression, which implies to the immunological conditions of the body. Regardless of this fact, the intricate connection between the dynamic elements of glycosylation during inhibitor advancement is not studied well. In the present examination with the advanced mass spectrometry-based technology with a specific sample preparation strategy for glycan analysis, we researched the plasma and antibody glycosylation differences in various phases of inhibitor advancement in contrast with the inhibitor negative HA male patients. The results can be correlated with various immunological assays to comprehend the roles of glycosylation changes, during the inhibitor improvement and how it is reacting to the emicizumab treatment.

2.2 Introduction

Hemophilia is a congenital bleeding disorder caused by a deficiency of blood clotting Factors VIII, IX, and XI. The absence of FVIII causes hemophilia A(HA) and deficiency in factors IX, and XI leads to HB and HC, respectively.[136] The FVIII gene is residing on the X chromosome.[137] HA is mainly affecting the males with a prevalence of 1:5000 in the United Staes.[138]The most commonly occurring, hemophilia types are A and B. Hemophilia A occurs 85% of the total hemophilia population[139] followed by HB. Both have an X-linked recessive mode of inheritance. Whereas hemophilia C, which is less regular among all and has

an autosomal mode of inheritance.[140] The female patients with congenital hemophilia generally act as a carrier as the males are affected by their maternal X chromosome.

The deficiency in clotting factors is occurring by a genetic mutation in corresponding genes. F8 and F9 genes are more susceptible to genetic modification, and 1/3 of the cause is a spontaneous mutation without any family history.[141] Considering the measure of Factor VIII activity in the blood of HA patients, the severity of the condition is categorized into severe (<1 IU/mL), moderate (<5 IU/mL), and gentle (5-40 IU/mL). In the all, out HA population, around 60% showed severe Hemophilia, and 15% showed moderate, and the staying 25 % showed mild forms.[138] People with hemophilia A must depend on lifelong treatment strategies and medication to maintain health quality.

The main symptoms of HA include joint bleeding, muscle hematoma, and soft tissue bleeding.[142] Patients with hemophilia show prolonged bleeding when a cut or injury happens, internal bleeding for unknown reasons, muscle, joint and tissue damages, and other sequelae of bleeding. Mild conditions can have a less complicated life until they encounter severe cut, wound, surgery, or trauma. In some cases, the conditions of bleeding can lead to morbidity and mortality. The advancement of modern medicine and treatment strategies could save many such situations and improve the patients' health quality. As of now, the backbone of treatment is the replacement of FVIII with plasma derived or recombinant FVIII concentrates on accomplishing hemostasis. FVIII activity blood or plasma transfusion can transmit many infections and spread of diseases, including HIV and Hepatitis B. Hence modern medicine depends on recombinant FVIII in treating the patients.

2.2.1 Factor VIII structure

Human coagulation factor VIII (FVIII) is a critical cofactor in the coagulation pathway. Human factor VIII is one of the largest coagulation factors with 2332 amino acids containing a single chain with MW of 293kDa comprises of six domains- A1, A2, A3, B, C1, and C2. In

the activated stage, it undergoes processing to form a heavy chain consists of A1, A2, and B with 200kDa and Light chain consists of A3, C1, and C2 with a molecular weight of 80kDa linked by metal ions.[143] Factor VIII is a nonenzymatic cofactor for the enzymes-prothrombinase and tenase-in the intrinsic pathway of blood coagulation. FVIII also increases the catalytic activity of activated FIX to activate the FX in the presence of calcium ions and phospholipids. Hence the absence of variations of FVIII may cause a severe bleeding disorder called HA.[143] 'A' domain, in turn, consists of A1, A2, and A3 domains, consists of 336, 337 and 329 amino acids, respectively. The two main epitopes of 'A' domain reside in A2, which it binds to FIX a. The A1 domain contains an FX binding site. R-336 of A1 and R-562 of A2 is the binding site of activated protein C binding site, which inhibits coagulation.[144–148] The functions of the B domain are not entirely elusive.[149] The modified rFVIII has B domain deleted, and with the heavy chain of 90,000 MW, and a light chain of 80,000 MW, likewise, end up being similarly functional as the full-length FVIII.[150] C domains are C1 and C2 consists of 153 and 160 amino acids, respectively .C2 domain has phospholipid linkage site, addition to that C2 also connected to C1 and A1 domains.C2 domain also contains thrombin and factor Xa domain binding sites.C1 does not have a direct impact on FVIII activity. However, the researchers suggest that it can influence on C2 domain to von Willebrand factor linkage strengthening.[151]

2.2.1 The Inhibitor development

Replacement therapy with rFVIII is successful in HA treatment, except if a patient builds up an alloantibody (inhibitor) against the exogenous FVIII. [152] The studies by Scandella D with inhibitor developed patients' plasma has used immunoprecipitation assays for epitope mapping indicated that mainly (around 70%) anti FVIII antibodies are binding to various sites of A2 as well as the C2 domains of the Factor VIII proteins.[148] The results of neutralization assays indicate that there is a third critical inhibitor epitope within the light

chain outside C2. Anti-A2 antibodies prevent normal function of the factor X ase complex of the intrinsic pathway of blood coagulation. Anti-C2 antibodies prevent the binding of FVIII to phospholipid and to von Willebrand factor, both of which are important for normal FVIII function.[148]Around 25-30% of patients develop high-titer, neutralizing anti-factor VIII (FVIII) antibodies, which binds to the functional domains of FVIII and inhibit their action (inhibitors),causes a significant snag in HA treatment.[153] The epitope mapping studies also proved that there are multiple antibodies, bind to various active domains of FVIII collectively called as inhibitors. The essential process of FVIII immunology is based on the T

helper cell-dependent pathway (TH cells), in which B-cells segregate into B memory cells and FVIII antibody-secreting plasma cells. Briefly, the FVIII antigens are recognized by the antigen-presenting cells and degrade the FVIII proteins into peptides, and the antigens are displayed on their MHC II for recognition by CD4+ TCR and ultimately leading to the TH cell activation and inhibitor production. The real mechanism and cause of inhibitor development of some HA cohort are still not precise.

The studies with patients, who develop inhibitors proved the incomplete or complete absence of the adequacy of replacement therapy.[153] The studies with patients, who develop inhibitors proved the incomplete or complete absence of the adequacy of replacement therapy.[154] The inhibitors are classified based on their titer or the historic titer based on the FVIII exposure.[155] The inhibitors are high responding if the titer is >5BU/mL and low responding if the titer is <5BU/mL in patients' blood.[156–158] Inhibitor development is not only limited to congenital hemophilia patients. Sometimes, individuals with no history of hemophilia or genetic conditions can also develop unexpected bleeding disorder caused by the autoimmune response of their body against the FVIII. This condition is called acquired hemophilia. Hence patients with Factor VIII deficiency along with inhibitor development have variable severity make it challenging to select the adequate treatment strategy. All the

above treatments are highly expensive. The treatment with the monoclonal antibody drug, emicizumab, is widely used as a novel treatment strategy since it can mimic functional aspects of the FVIII in the body. The FVIII coagulation cascade interacts with FIX to activate FX. Emicizumab is a bispecific, genetically engineered monoclonal antibody that can bind to both FIX and FX. Hence it can overcome the FVIII deficiency by bridging FIX and FX to complete the hemostasis. It is essential to unwind the mechanism of inhibitor development for the betterment of strategizing the therapeutic measures. The FVIII coagulation cascade interacts with FIX to activate FX. Emicizumab is a bispecific, genetically engineered monoclonal antibody that can bind to both FIX and FX. Hence it can overcome the FVIII deficiency by bridging FIX and FX to complete the hemostasis. It is essential to unwind the mechanism of inhibitor development for the betterment of maneuvering the therapeutic measures.

2.2.2 Glycan as a biomarker

The studies revealed that in humans, around 20%-50% of proteins are glycosylated.[159] Approximately1-2% of the genome is assigned to produce the glycosylation machinery protein components in various cells.[160] Even though the glycosylation process occurs in the endoplasmic reticulum and Golgi apparatus, the process of nucleotide-sugar donor-synthesis generally happens inside the cytosol except for CMP-sialic acid, which is synthesized inside the nucleus.[161,162] Different factors can regulate the structural determination of glycan in glycoproteins. The main factors are the expression level of glycosyltransferases, glycosidases, and other regulatory proteins and their availabilities.[163,164] The second factor is the substrate availability and the competition between several glycosyltransferases and glycan acceptors for the same substrate [160]. The third factor is the transportation of essential enzymes and nucleotide sugars to the endoplasmic reticulum and Golgi apparatus. The fourth factor is the protein quaternary structure and folding [164,165] , and

the fifth factor is the microenvironmental variations in the like oxygen supply and pH .[166] The protein glycosylation varies in the multilevel of life – cellular, tissue, individual, and species level. Even though they match at the protein level, the glycan species may change, which inurn alter their function. Since the glycan play an interconnection between the cells and molecules, the differences in the micro and macro heterogeneity of the proteins may relate to the health and disease of the individual. They also vary according to the physiological changes like aging[4], pregnancy[5], malignant metastasis[6], disease progression, inflammatory responses, metabolic changes, bacterial and viral infections that caused alteration of the microenvironment in the cell.[3] Hence, glycan profile changes can be a prognostic or diagnostic biomarker for many physiological conditions.[7] The altered glycosylation pattern can also alter the binding specificity of the cells or proteins with other molecules and eventually change their biological functions. There are many plasma proteins like AFP[167], PSA[33] , and CA125[31] also be related to certain cancers which is detailed in **section1.1.** Plasma antibody glycosylation is also a very good indicator of many disease conditions and immune ageing.[168],[169] The structural and functional importance of immunoglobulin and glycosylation is detailed under **section 2.2.3**

It is realized that improper protein glycosylation may modify the cellular trafficking and flag , proteins for degradation instead of secretion. Aberrant glycoforms may bring about protein misfolding because of missing oligosaccharide markers, prompting less stable, non-utilitarian protein adaptations. Subsequently, the host cell glycosylation condition won't just affect FVIII glycosylation, yet additionally impact its resistant framework on inhibitor development. Hence we tried to understand the glycosylation variation in inhibitor developed patients over regular HA patients. Since glycosylation is a dynamic modification, we also wanted to study the glycosylation pattern changes of N-glycome of plasma and antibody with inhibitor development and in response to emicizumab treatment. The standard strategies of

glycoprotein analysis are illustrated in ***figure 1.1.*** Glycosylation is the both species and tissue specific. The tissue specific glycosylation may required for their functional specification. Plasma is a mixture of various secretary protein produced by different tissues. Hence analysing plasma N-glycome gives overall differences in the glycosylation differences between the individuals during disease progression.[170] Where as IgG glycosylation can tell about the B-cell specific glycosylation modification during the disease progression.

2.2.3 IgG glycosylation

Immunoglobulins are the most significant and thoroughly comprehended glycoproteins of our adaptive immune system.[164] In the normal functioning of the antibody involves recognition of the foreign pathogens and allergens by initiating a series of biochemical reactions (immunological reactions) to eliminate them from the body. Nevertheless, in rare cases, the immune system can be focused against the individual's own epitopes which may lead to an autoimmune disease.[171] Their production types and existence vary among the individuals based on their acquired immunity. All immunoglobulins are glycoproteins and the existence of glycan , their structure and interaction between the glycan and amino acids can regulate their effector function.[118] Immunoglobulins are primarily divided into 5 different types as Ig G, M, A, E and D. The heavy chain isoforms are of five types as γ, μ, α, ε, and δ. The light chains can be of two different isoforms as κ and λ.[118] IgG is the most abundant immunoglobulin in the body which itself divided into four types as IgG1, G2, G3 and G4. All the IgG's have a conserved N-glycosylation site generally termed as Asn-297 at their heavy chain, Fc region of the CH2 domain.[171] The conserved N-glycans are very important for the IgG to bind its Fc γ receptors and regulate their effector functions.[171] For example, the capping of Fc-N-glycoform with alpha 2,6 sialic acid can impair the antibody binding to its Fc-receptor, and thus leading them to the anti-inflammatory response.[172] The anti-inflammatory property will sustain if it contains the glycoforms

terminated with galactose.[172] Also, the increased terminal galactosylation can enhance the complement C1q binding and thereby increases the CDC.[173] Whereas the same antibody containing a shorter N-glycan terminated with GlcNAc residue leads to the binding of FCγRIIIa receptors which get enhanced with non-fucosylation and proinflammatory response.[172] Fucosylation is generally known as a 'safety switch' for ADCC in IgG.[172] As the glycosylation is directly regulating the antibody effector functions, it may act as a linking between innate and adaptive immune response; the topic is a point of discussion in the specified field of study.[174]

IgG glycome is a biomarker for various pathological and physiological state of an individual. The IgG glycome show the highest microheterogeneity among other plasma proteins.[175] The IgG related glycoform variation is a biomarker in many diseases. The de-galactosylated N-glycan is a common biomarker related to primary osteoarthritis, RA[172] ,ulcerative colitis, Crohn's disease and some malignancies.[176] However some physiological changes during the disease condition reverse the glycosylation pattern. For example it is proved that during pregnancy, RA woman showed higher galactosylation than in the pre-pregnant condition.[177] Also, it is noted that the IgG glycosylation changes in response to antigen encountering like vaccination.[165]

2.2.4 The IgG subtype expression in different immunological conditions

The expression of each IgG subtypes is also related to the antigen it encounters at the time of clonal selection. IgG1 is highly expressed against soluble protein antigen as it contains the highest percentage in plasma. IgG1 deficiency can cause hypogammaglobulinemia, and its lack of other protein can also be related to certain infections.[178] IgG2 subtype is generally expressed against the bacterial polysaccharide, blood group antigen, and other carbohydrate antigens, and their absence can be related to the

deficiency of anti-carbohydrate antibody in the blood.[179] IgG3 is a proinflammatory antibody, as it is expressed first in the course of infections .[179],[180]Some allergic reactions can be an inducer of IgG1 and 4 (Figure 2.2).

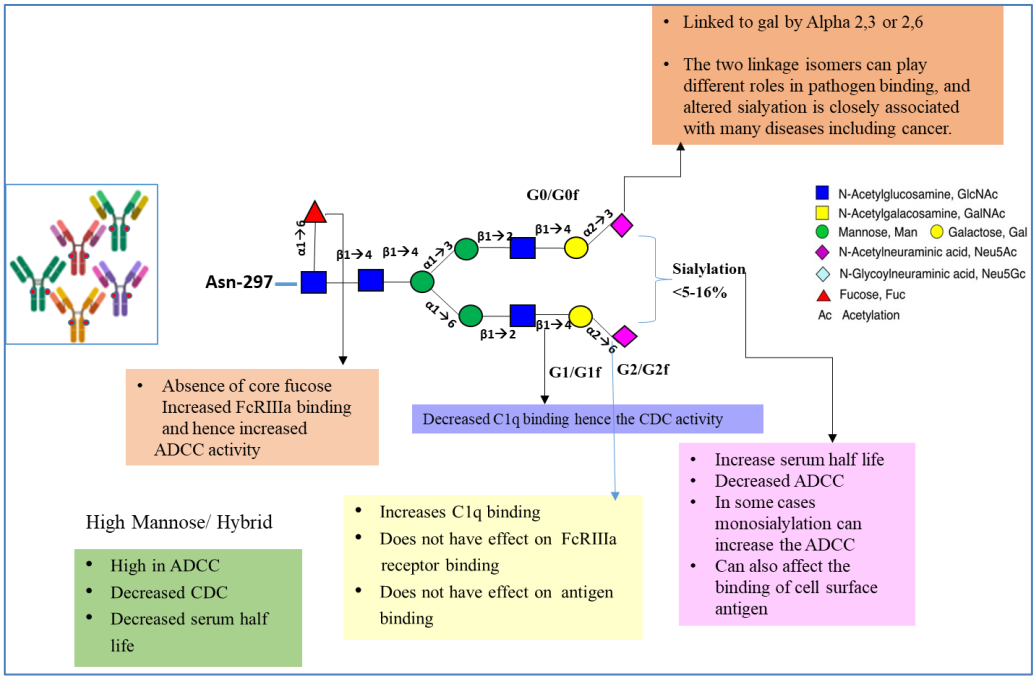


Figure 2.1The IgG site-specific glycan and their functional influence of its effector function

IgG4 is upregulated as an immune response against the long term exposure of an antigen , against some allergens[181] or in response to some parasite infections[181]. IgG4 can be dominantly expressed against the long term expression of therapeutic proteins including FVIII and FIX.[182–184]The different IgG subtypes have different affinity for the FcRs in the cells.[185] Furthermore, the equivalent FcR can be connected to various signaling molecules when connected by various cell types varies in their immunological functions.[180] Decreased N-linked galactosylation and sialylation of IgG known to decrease the CDC activity , and this is a biomarker in many autoimmune disease like in RA.[186]The immune response balance is occurred through the action of different activating FcRs and inhibitory FcRs .The effect of different N-glycosylation on effector functions of antibody is shown in **figure 2.1**.

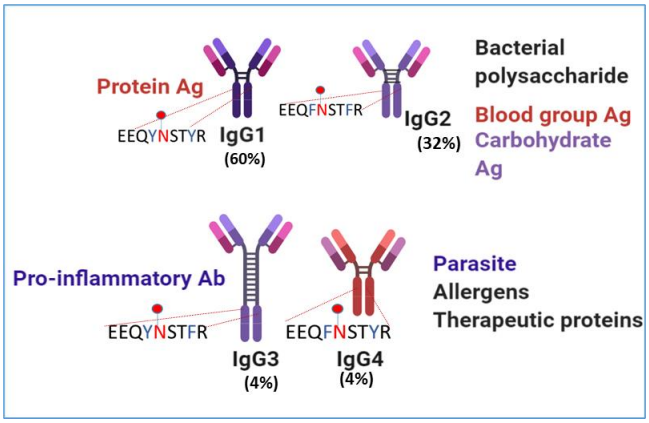


Figure 2.2 IgG subtyping and their traits of inflammatory response
The CH2.84.2 conserved glycopeptide sequence (glycosite is highlighted in red) is slightly different in all the subtypes (shown in blue). The general percentage of distribution from total IgG in normal conditions are given in bracket. The specificities of their antigenic

2.2.5 Mouse plasma glycan biomarker

The mouse models are widely used for biomarker discovery and drug discovery due to their easiness of life support at lab environments in a controlled way with a short life cycle.[187,188] Mouse models are widely used for protein specific biomarker discovery in various diseases.[189] Even though the glycosylation has some species-specific differences between human and mice model the disease specific aberrations can be monitored much efficiently with mice models.[189]

Hence to make a controlled lab atmosphere the glycosylation differences with respect to hemophilia A is to monitor the glycosylation differences with inhibitor development with more related lab grown murine system.[173] There are species-related differences between human and mouse models in their immunology[190] and glycosylation. Nevertheless, the humanized murine models are widely used to study the antibody effector functions.[191] There are three main IgG subtypes detected in mice that are divided as IgG1, IgG 2b, and IgG 3. In addition to that IgG 2a/2c is also been seen related to a specific strain. The site-specific glycosylation is widely studied for various effector functionalities.

2.3 Scope of the current study

Glycomics and glycoproteomics based biomarkers are widely explored relating to the field of disease progression in the recent past. It is reported that the MS-based analysis for glycomics studies are reliable as they used globally and compared between the different labs and endorsed by A multi-institutional assessment of glycomics methodologies coordinated by the Human Disease Glycomics/Proteome Initiative (HGPI) and Human Proteome Organization (HUPO)[86] The disease progression in individuals is majorly influenced by their genetic background and the environmental effect mainly caused by their habits of life. In 2012, U.S national academy has declared that glycans are directly involved in the pathophysiology of every major diseases and Extra information from Glycosciences will be expected to understand the objectives of personalized medication and to exploit significant interest in human genome and proteome research and its effect on human wellbeing.[170] Glycans and glycoproteins are already reported as the essential molecules in health and disease.[164]In our lab, we are trying to understand the mechanism of inhibitor development by studying glycosylation differences in factor VIII as a potential cause of immunogenic response in the inhibitor developed patients. As we realize that the inhibitor development is occurring just among specific individuals, we needed to comprehend the distinctions of their immune system and protein expression. Numerous studies impressively fortify that the knowledge of glycomics flavors and sometime overturned the traditional understanding of protein functions. The in depth studies are conducted in cancer and autoimmune diseases.[9,177] Here we tried to explore the glycosylation differences among plasma protein in congenital HA patients without inhibitor development and the HA patient with inhibitor development using mass spectrometry based glycan analyses. The changes in the expression level of glycoprotein, especially the immunoglobulin subtypes and the modification of their glycosylation, may lead to an indicator regarding the unique mechanism of inhibitor

development. We also planned to study the treatment specific glycan alteration with emicizumab treatment, hence we can discriminate the healthy and proinflammatory specific glycan pattern. Incorrect glycoforms may result in protein misfolding due to altered oligosaccharide markers, leading to less-stable, non-functional or mal functional protein conformations. Therefore, the host cellular glycosylation environment may influence its immune system on inhibitor formation. Since the glycosylation is dynamic the treatment strategies may improve the glycosylation pattern ,hence reverse the healthy pattern of glycosylation. Both the cases are important to understand the underlying pathways of inhibitor mechanism. Mass spectrometry based glycocentric and glycopeptide analysis help us to understand the overall N-glycome variation between HA , inhibitor developed and control population . It also tell us about the micro and macro heterogeneity related to Inhibitor development through bottom- up analysis of glycoproteins. Additionally mass spectrometry based relative quantitation can tell us about the inhibitor specific upregulation of any glycoproteins in the sample. The variations can be compared with other biological assays like cytokine analysis and eventually help us to correlate with the immunological alterations relating to glycosylation changes.

Hence, the comprehensive analysis of the glycosylation pattern of human antibody and plasma can give us an indication of the type of effector function of the immune system and intensity of inflammation. The biological role of glycoproteins is mainly regulated by the type of glycan attached to the glycosylation sites. The micro and macro heterogeneity of glycosylation can relate to the genetic and environmental variability within the cells. The glycoprotein expression level can also vary in response to specific physiological conditions. We anticipate that the microenvironment of inhibitor developed patient plasma changes due to the difference in the immune response against FVIII. The site-specific antibody glycosylation of antibody classes and subclasses are also determining the type of antigen it

encountered by their immune system **(figure 2.2)**. All the trials are replicated in the HA mice models from 3 different strains. The observations are correlated with human studies.

In summary, we tried to understand the glycosylation variation in inhibitor developed patients with regular HA patients. The human study was extended with mass spectrometry related plasma glycosylation changes with emicizumab treatment. All these steps with organized analysis can provide concrete evidence for the facts of inhibitor development.

2.4 The experimental design

In this experiment, we decided to understand the glycosylation pattern of inhibitor developed patients and the typical HA patient without inhibitor development. The released glycan analysis of plasma protein can provide the details of glycosylation structure, linkage, and overall distribution. We investigated the glycosylation differences from total plasma and antibody isolated from the different HA patients with inhibitor development and compared them with HA, inhibitor negative cases within similar age group. The investigation was performed in three different stages. The first stage was for primary screening. For the initial discovery, we investigated healthy individuals without HA considering the variability in ethnicity, age, and blood group differences to understand the possible causes of variabilities. All the individuals considered for the study were male. We parallelly examined the HA patients with and without inhibitor advancement. In the subsequent stage, we inspected a patient's plasma at his pre-and post-treatment with emicizumab. The patient was inhibitor positive, and the two stages of treatments were compared with a control plasma without inhibitor development but within the same group. We expanded our study for comprehensive analysis at glycocentric and glycopeptide level. In the third stage, we increased the number of individuals within the same age group with and without inhibitor development and studied their glycosylation pattern change after the emicizumab treatment. We additionally analyzed

control samples without Inhibitor development to prove the effect of emicizumab glycosylation in overall profile.

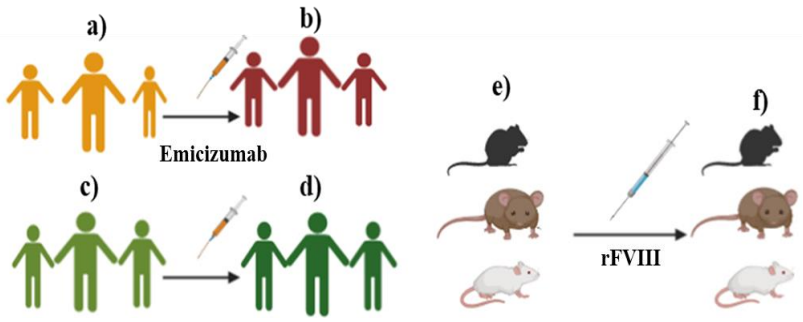


Figure 2.3: experimental plan for analytical cohort selection.
a) Inhibitor developed individual from same age group at emicizumab pre-treatment stage. b) 'a' at emicizumab post-treatment. c) Control HA patient without inhibitor development d) 'c' at emicizumab post-treatment e) BL-6, BALB/c and SV-129 strains before treatment with rFVIII and inhibitor development f) e) BL-6, BALB/c and SV-129 strains after treatment with rFVIII and inhibitor development

For the detailed study of antibody and plasma glycosylation profile, we enriched the antibody from the plasma, which detailed in **figure2.3**. We have performed a glycosylation analysis of plasma and antibody independently to distinguish alpha 2,3 and alpha 2,6 sialic acid linkage analysis. The two-step permethylation will be supplementary analysis for the one-step process, which also gives additional information about the sialic acid linkage. The site-specific glycosylation analysis is performed at the glycopeptide level using glycopeptide mapping by HR-MS. Site-specific glycosylation differences, protein related glycoforms variation, glycan structure elucidation with glycopeptide sequences and the site of glycosylation can be the additionally obtained from the glycopeptide analysis.

We additionally analyzed the N-glycome pattern differences in HA male mice before and after FVIII injection and inhibitor development. We analyzed three different strains - SV129, BALB/c, and BL-6 CD.45.1- of lab mice to confirm the glycosylation alteration during inhibitor development in the mice. The HA mice plasma was collected in the 11th week for the baseline analysis. After that, the mice were treated with rFVIII and allowed to develop the inhibitor. In the 15th week, after inhibitor development, the plasma was collected again, and

their glycomics and glycoproteomics patterns were analyzed and the study can supplement the observations we get it in human samples hence prove the chances of glycan biomarker related to inhibitor development. As a background analysis, we also compared the HA mice sample, and wildtype mice sample with BL-6 stain to confirm no variation in their plasma glycome during the HA condition.

2.5 Materials and methods

Plasma from a HA patient drawn before and after emicizumab treatment and inhibitor values (citrated) obtained from Washington Center for Bleeding Disorders at Bloodworks Northwest. The hemophilic mouse plasma was obtained from Seattle children's hospital(citrated). The glycan standard 4 4 0 0 1 were produced in-house. Super clean™ ENVI-Carb™ SPE Tube (Millipore/ sigma), Thermo Scientific™ HyperSep™ C18 Cartridges , magnetic stand, and Pierce™ Protein A/G Magnetic Beads (Thermo Scientific, MA, US), Formic acid (FA), Human serum, Trifluoracetic (TFA) (LC-MS grade) acid and super-DHB matrix for MALDI-MS (≥99.0%) were purchased from Sigma-Aldrich (ST. Louis, MO, USA). Tris-HCl buffer was purchased from US Biological (Swampscott, MA, USA). 10X TBS was purchased from Bio-Rad. HPLC grade acetonitrile (ACN) and methanol were purchased from J. T. Baker® Chemicals (Avantor Performance Materials, Inc. Center Valley, PA, USA). Deionized water was produced using a Milli-Q A 10 system from Millipore (Bedford, MA, USA). Microcon-30kDa Centrifugal Filter units (YM-30, 0.5 mL) with Ultracel® low-binding regenerated cellulose membrane was purchased from Millipore. BEH Amide Column, 130Å, 5 µm was purchased from Waters Corporation (MI, USA). Other materials, including sodium dodecyl sulfate (SDS), urea (UA), ammonium bicarbonate (ABC; NH4HCO3), sodium hydroxide beads (NaOH, 20-40 mesh, 97%), iodomethane ICH3 (≥99.0%), dimethyl sulfoxide (DMSO, anhydrous, ≥99.9%), dichloromethane (DCM, HPLC grade), and ammonium hydroxide solution (NH4OH, 30%), dimethylamine solution (2 M in methanol), dimethyl sulfoxide

(DMSO),1-hydroxybenzotriazole (HOBt) hydrate,1-ethyl-3-(3-(dimethyl amino)propyl)-carbodiimide (EDC), Tween 20 detergent and BSA were purchased from Sigma-Aldrich (ST. Louis, MO, USA). 1.5 mL and 2 mL microcentrifuge tubes purchased from Eppendorf, PNGase F were purchased from purchased from New England BioLabs (Ipswich, MA), 2-Iodoacetamide (IAA) and Dithiothreitol (DTT) was obtained from Alfa Aesar (Ward Hill, MA). N glycosidase F (PNGase F) was purchased from New England BioLabs (Ipswich, MA). Sequencing grade porcine trypsin was purchased from Promega (Madison, WI). LC-MS grade quality acetonitrile (ACN), formic acid (FA), and water were purchased from Thermo Fisher Scientific, Inc. (Waltham, MA).

2.5.1 Antibody enrichment

The IgG enrichment of plasma protein was performed by the manufactures protocol with minute modifications. Approximately 20µl of blood plasma is diluted with 480µl of binding buffer (1 X TBS contains 0.1%tween 20 and 0.1% BSA). Approximately 150µl of protein A/G beads were taken in a separate 1.5mL microfuge tube. The beads were washed with 150µl of binding buffer and gently mixed and collected on the magnetic stand. The supernatant was discarded. The beads were equilibrated with 1ml of binding buffer and repeated the previous step.

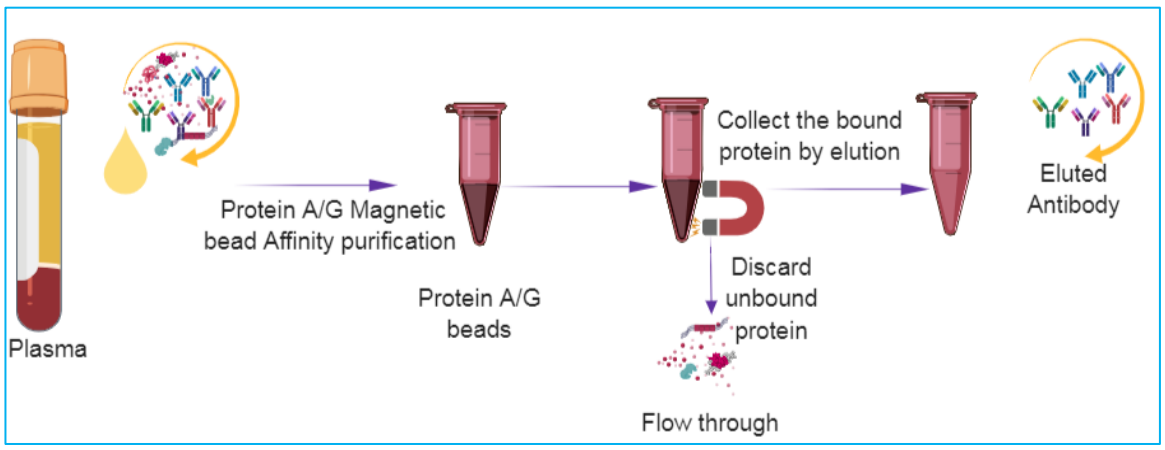


Figure 2.4 The schematic representation of antibody enrichment from plasma using Protein A/G magnetic beads

The diluted plasma sample was added to the prewashed tube and incubate at room temperature for 1hr with slow mixing. The beads were collected using the magnetic stand and washed with binding buffer without BSA. To the collected beads, add 150µl of elution buffer and followed by 10mins incubation without mixing. The beads were collected again and the supernatant containing IgG was neutralized using 15µl of 1M tris buffer pH 8.0. 5µl of enriched antibody and 0.3 µl of human plasma was separated on the reducing gel (10% SDS). The gels were stained after the separation using Coomassie stain. We repeated the analysis with mouse stains following the same protocol.

2.5.2 Released N- glycan analysis of plasma and antibody

2.5.2.1 N-glycan release and purification

For the release of N-glycan, around 6µl of plasma (mice or human) and the enriched antibody from 10µl of the plasma (human) were taken separately. The volume was reduced below 10µl and then diluted it further with the denaturation buffer and boiled the mixture at 90 °C for 10 min. After cooling it down, 1µl of 10% of NP-40 and 1µl of PNGase F (500U) were added to the mixture using manufactures protocol with minute modification in the incubation time upto16 hrs. The enzymatically released N-glycans obtained from plasma and antibody were purified using a connected cartridges system where envicarb on bottom and C18-6ml on top. The cartridges were pre-conditioned with 5 mL of acetonitrile (ACN)containing 0.01% TFA followed by 3ml of 25% ACN in 0.1% TFA and finally with 6ml of 5% ACN with 0.1%TFA water. The sample was diluted with 5% ACN in 0.1% TFA water and, passed through the cartridge system and washed it again with additional 6ml of 5% ACN in 0.1% TFA water. The C18 cartridge was removed at this stage and the separated glycans were bounded to envicarb cartridge. The enriched glycans were eluted out with 6ml of 25%ACN in 0.1. % TFA water. The purified glycans were aliquoted into two vials and dried for the following derivatization.

2.5.2.2 Glycan permethylation (one step method)

The base for permethylation is prepared by mixing around 350µl of 50% NaOH in DMSO and transferred into a screw-capped glass tube before quickly adding 700 µL anhydrous methanol. The mixture was vortexed well and 4ml of anhydrous DMSO was added to the tube. The mixture was vortexed well and centrifuged at 2000rpm for 1 min. The supernatant was discarded and 4mL of fresh DMSO was added again to the NaOH gel to dissolve the base. The base was washed with DMSO (3×4mL) and finally dissolved with 2ml of fresh DMSO to use as a reaction base for permethylation. One aliquot of dried glycans was redissolved in 200µl of DMSO and 250µl of freshly prepared reaction base was added to it. 150µl of iodomethane was added to the mixture and sonicated for sharp 10mins. The reaction was stopped by adding 500µl of water in a dropwise manner. Then, the permethylated glycans were extracted twice with 400 µL DCM by liquid -liquid extraction. The mixture was vortexed well and centrifuged at10000rpm for 5mins.The supernatant was discarded and additionally 800µl of water was added and repeated the previous step. The washing was continued for additional 4 times and make sure that the supernatant is at pH 7. The bottom layer was transferred to a fresh tube and kept it open under the hood for 4 hrs. The dried glycans were resuspended in 20µl of 50% methanol for further MALDI analysis.

2.5.2.3 Glycan permethylation (two steps method)

The experiment was performed with previously published method with minute modification.[90] A fraction of isolated glycan from plasma proteins was mixed with 25 µL of reaction mixture composed of 250 mM dimethylamine, 250 mM EDC and 500 mM HOBt in DMSO and incubated at 60 ºC for 1 h .The samples were kept in a speed Vac for additional 20 mins at 60 ºC for the removal of excess dimethylamine. The resulting sample was diluted with 200µl of DMSO and continued the permethylation process as described in

section2.5.2.2 The permethylated sample was dried and dissolved in 50% ACN/50% H2O for MALDI-TOF analysis.

2.5.2.4 N-glycan analysis by MALDI-TOF-MS

One microliter of permethylated glycan was mixed to an equal volume of super-DHB (10 mg/mL in 50% ACN), 1 μL of the mixture was spotted on the target plate (Bruker Daltonics, MTP 384 polished steel) and allowed for crystallization. The samples were spotted in triplicate along with the permethylated dextran. The dextran serves as an external calibrant. We have used Ultra Flextreme MALDITOF-MS (Bruker Daltonics, Bremen, Germany) equipped with 1 kHz Smart beam-II laser operated by flex Control 3.4 software (Bruker Daltonics). The data was acquired in the reflector positive mode with accuracy up to 50 ppm. For each spot, the laser was shot up to 20,000 shots at a laser frequency of 666.7 Hz, using a complete-sample random walk with 2000 shots per raster spot. The mass range detection was 1000- 5000 Da. The laser intensity was set based on the sample concentration, ionization of the glycan species, and depends on the instrument condition while the monoisotopic peak was still clearly defined for all detectable glycan masses. The m/z range was monitored from 1000 to 5000. Tandem mass spectrometry (MALDI-MS/MS) was performed for the possible most intense glycoforms using CID.

2.5.2.5 Data analysis using flex analysis v3.4 (Bruker daltonics)

The glycoforms distribution in plasma and antibody was calculated independently. The details of data analysis include data acquisition, raw data, data clean up, including peak picking, baseline correction, scaling, normalization, and quality assessment.[192]In details, the resultant MALDI-MS spectra were smoothed, and baseline-subtracted across all the m/z range and internally calibrated using a set of common complex glycan masses which is present in the human serum listed in appendix B table 1 . For the mice sample the

glycosylation variations are considered and spectra was calculated with the common glycoforms present in human plasma structures were listed in the same table **.** The glycan masses are in single charges in $M+Na^+$ form. Masses were picked in the spectra using monoisotopic heights as intensity. The S/N intensity was set up to 10. The listed masses were assigned using a software, Glycoworkbench v.2 (http://www.eurocarbdb.org/applications/ms-tools) to find the corresponding glycoforms and their compositions. The MS/MS was performed for the ambiguous structures, and the most probable structures were reconfirmed using the fragmentation pattern by the same software.

The glycan masses were calculated as $[M+Na]^+$ precursor ions. The dextrin was used to calibrate the system. The known masses in the region of N-glycan mass 1500-5000 Da was also used for internal calibration in the flex analysis software. The calibrant list was given in the Appendix B table 1. Some of the common calibrant masses are H5N2 (m/z 1579.7826), H3N4 (m/z 1661.8357),H6N2 (m/z 1783.8824), H4N4F1 (m/z 1835.9249), and (m/z 1906.9620) ,H4N4F1 (m/z 2040.0247), H3N5F1 (m/z 2081.0512),H4N5 (m/z 2111.0618), H5N4F1 (m/z 2244.1245), and H4N5F2 (m/z 2459.2402)). The sialylation samples were not considered due to their differences in sialylation in mice samples. The calibration performed with 50 ppm tolerance (H = Hexose, N = N-Acetyl hexosamine, and F = Fucose, G = N-glycolyl neuraminic acid and A = N-acetyl neuraminic acid. The number of residues given after the letter). The intensity of each peak was calculated considering the monoisotopic ionic mass of each glycan. The fucosylation (mono, bi and multi), sialylation (mono, di and multi sialylation) were calculated by percentage calculation by adding the individual glycoforms with specification in total glycoforms intensities. Galactosylation index were calculated by the ratio calculation using the following formula.[6]

$$GI = \frac{G0f}{(G1f + 2 \times G2f))}$$

In which GI is galactosylation index, G0f represents non-galactosylated glycoforms, G1f represents mono-galactosylated glycoforms, and G2f represents di-galactosylated glycoforms. The data has been compiled together based on their structure specification as terminal galactosylation, and fucosylation. The plasma glycans are segregated based on their structural specification as in **figure 2.8**. The glycans were generally categorized into complex hybrid, high mannose and categorized based on their specific sugar which has functional impact like fucose and sialic acid. This segregation was again subdivided based on the number of fucose /sialic acid like total fucosylated / sialic acid containing species(total), mono, di and multi. The percentage calculation is done by taking the ratio of the sum of specific sugar intensities to the total identified sugar intensities. Similarly, to the galactosylation index calculation, sialylation index can also be calculated using the similar formula based on 54000 (G2) glycoform. In addition to that, the linkage specificity was also calculated based on the two-steps permethylation data, using the formula

SI=G2/G2S1+G2S2×2. In which SI is sialylation index, G2 represents non-sialylated glycoforms, G2S1 represents mono-sialylated glycoforms, and G2S2 represents di-sialylated glycoforms The sialylated species will vary based on the specific linkage calculation. The hybrid G2S2 structure (Which has both type of linkages) will be added to G2S1 as the specific sialic acid content was only one in the structure. The formula is adapted to mice SI calculation.

2.5.2.6 Statistical analysis

Before applying the statistics, the intensities across m/z range specific mass spectrum were smoothed and baseline subtracted. The relative intensities of each glycoforms were calculated using 5 4 1 0 0 (G2fS1, H5N4A1) as a reference peak. For the antibody glycosylation 4 4 0 0 1 (G1f, H4N4F1) is taken as the reference peak. The quality of overall

signal intensity of each MADI-MS spectrum was confirmed by their individual peaks S/N and maintained minimum to 10.

Differences in each type of hemophilia and control were analyzed across different glycoforms from the spectra and categorized based on their specifications, including glycoform structure, type, complexity, fucosylation, galactosylation, and sialylation. For that, all the glycoforms were summed together with their intensities in the spectra and the percentage were calculated with total glycoforms intensities. The statistical analyses were performed using GraphPad Prism version for Windows (GraphPad Software, La Jolla, California, USA). Statistical analysis was performed with the two-tailed student's t-tests for unpaired data. The standard deviation, p value testing and graph visualization was performed using the same software. For the statistical testing, in which a p-value of < 0.05 was considered statistically significant.

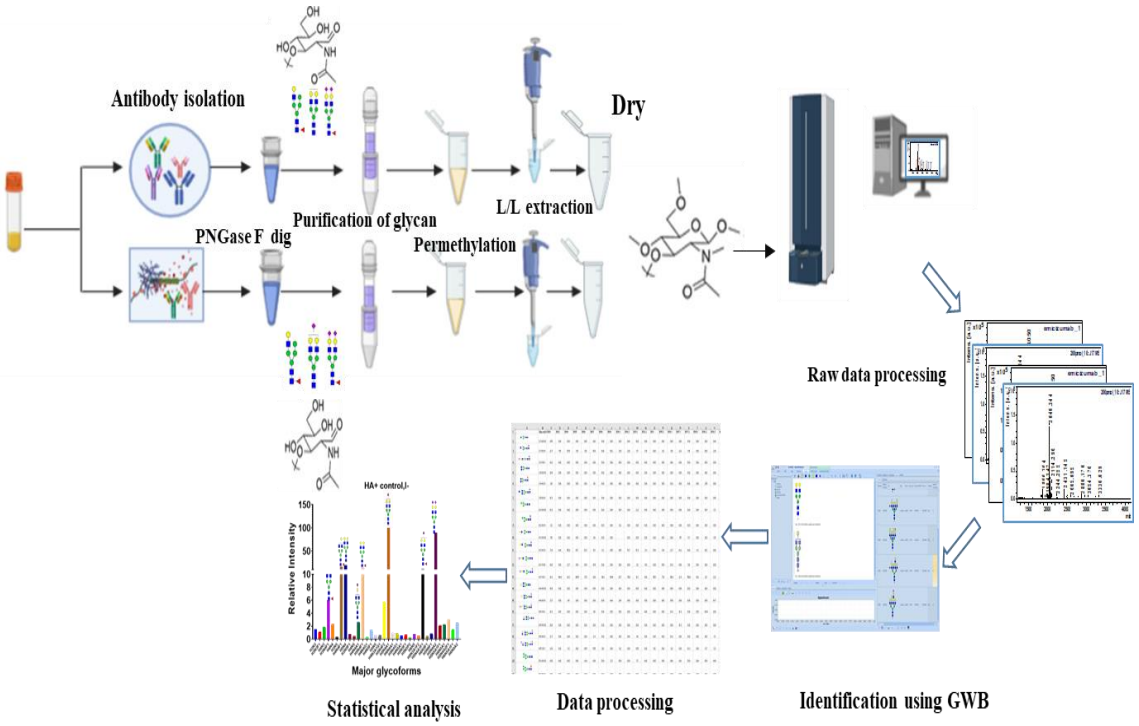


Figure 2.5 The workflow conducted for glycomic profiling of human and mice plasma

2.5.2.7 MALDI-MS method reliability study (spiking study)

The MADI-based permethylated glycan analysis is already published as a reliable method with the biopharmaceuticals studies and their accuracy , repeatability and intraday-intra-day variability and also showed the permethylated glycan by MALDI can be a potential application of glycosylation analysis for glycan biomarker studies combined with orthogonal methods.[86] Here we took 8 µl of commercially available human serum (which contain approximately 1mg of protein calculated by Nanodrop). The N-glycans were separated as described in the **section 2.5.2.1.** The isolated glycans were equally split into 5 vials (each contains glycans from approximately 200µg of proteins)50, 100, 150 and 200 picomoles of 4 4 0 0 1 standard was spiked individually. The final tube was kept as a control baseline tube with no spiking. The tubes were continued for the further sample processing and data acquisition **(Section 2.5.2.2, 2.5.2.4-2.5.2)**.Each sample was spotted in triplicate and the relative intensity of the standard 4 4 0 0 1 was calculated with the inherent glycoform 5 4 1 0 0 in the serum sample. The linearity of 4 4 0 0 1 was also calculated with total intensity percentage. The standard deviation of the relative intensity and individual percentage at each point of 4 4 0 0 1 were calculated and plotted separately to calculate the %RSD and linear regression.

2.5.3 Glycopeptide analysis and (Bottom-Up Identification)

The glycopeptide level examination is additionally named as a bottom-up approach. We can recognize the glycoprotein using the sequence of the peptide and its identification by the database search. Likewise, we can get the site-specific glycan structure, percentage of occupancy along with the amino acid sequence analysis. Glycopeptide examination requires multi-step sample processing, which includes protein purification, trypsin (or endoprotease) digestion followed by glycopeptide enrichment together to determine the quality of spectra.

Separation and identification using adequate mass spectrometry methods and comprehending the complex data using adequate software are also essential for the precision of the analysis.

2.5.3.1 Trypsin digestion of plasma protein

The plasma protein concentration was determined by a nanodrop and approximately 1 mg of total proteins from human or mouse plasma was subjected to FASP procedures.[193]The protein was denatured using the lysis buffer (4% SDS, 100 mM DTT, 20 mM Tris-HCl, pH 7.6) at 95°C for 10 min. The denatured protein was diluted to 1 mL with UA1 buffer (8 M urea, 20 mM Tris-HCl, pH 8.5). The solution was transferred to a centrifugal 30kDa filter gradually buffer exchanged with UA1 for 3 times at 150000 rpm for 15 mins each time. The volume of the sample was maintained at 20µl and 100 µL of IAA solution (50 mM IAA in UA solution) was added to the filter and incubated for 30 min in the darkness. The mixture was centrifugated at 15,000 rpm for 10 min. and buffer exchanged with UA solution a followed by 200 µL of digestion buffer (100 µL of 50 mM NH_4HCO_3 each time). At last, 20 µg of trypsin was added into the filter and the pH was adjusted to 8.0. The filter was incubated at 37 °C for 12 h. The digested tryptic peptides were collected from the filter by passing 100 µL of 50 mM NH_4HCO_3 through the filter for six times. The concentration of the collected peptides was measured at 280nm using a Nanodrop. The peptides were dried and reconstituted in 100µl of 80% ACN in 0.1% TFA/ water for glycopeptide enrichment.

2.5.3.2 Enrichment of glycopeptides

The glycopeptides enrichment was performed using an NP-HPLC with Bridge BEH Amide Column, 130Å, 5 µm 600 µL of 80% ACN containing 0.1% TFA. The column was equilibrated with 80% ACN containing 1% TFA. The glycopeptides was injected and allowed to separate on the column. The fractions of glycopeptides were collected from 66.5-

40% of ACN. The fractions were collected and dried using a centrifugal evaporator and injected to nano RP HPLC-MS for further analysis.

2.5.3.3 *Nano LC-MS/MS analysis*

The glycopeptide sample was injected into an LTQ-Orbitrap Elite mass spectrometer (Thermo Fisher) equipped with an EASY-spray source connected to nano-LC UltiMate 3000 high-performance liquid chromatography system (Thermo Fisher). The mass spectra were recorded in a data-dependent mode in a full scan survey of m/z range 375 Da-2000Da. The fragmentation was undergone in higher-energy C trap dissociation (HCD) in orbitrap. The resolution at 400 m/z is 60,000 MS experiment and automatic gain control target, 1,000,000 ions. The maximum accumulation time is 50 ms. The ten most intense ions were fragmented with optimized stepped collision energy (SCE) of 30 ± 15% NCE. The glycopeptides were separated on RP column at the flow rate of 300 nL/min, EASY-Spray pepmap C18 (75 μm × 50 cm, 2 μm, Thermo Fisher, US) using a linear gradient of mobile phase 3% B to 40% B for 120mins (MP A: 1.95% ACN, 97.95% H2O, 0.1% FA; MP B: 79.95% ACN, 19.95% H2O, 0.1% FA .

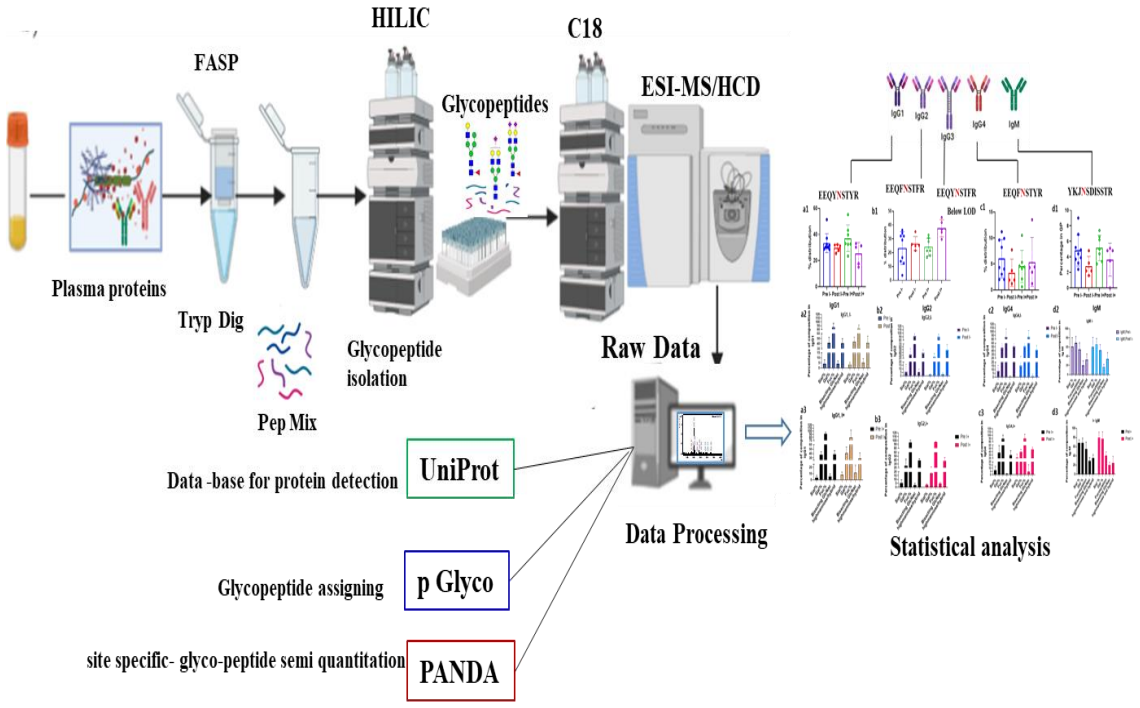


Figure 2.6. The workflow for the site-specific identification and quantitation of glycopeptide/ glycoprotein by bottom up approach.

2.5.3.4 Data analysis

The data was analyzed using p Glyco 3.0 (an improved version of p Glyco 2.0 software) was used to identify N-glycopeptides. In the glycan database, we have a code in a short form in the format “H N A G F’’. “H” is for galactose and mannose, “N” is for N-acetyl hexosamine, “A” for Neu5Ac and “G” for Neu5Gc and “F” is for fucose. The glycopeptides were assigned based on the UniProt-Swiss human protein database (20,417 reviewed entries on March 20, 2019) and the N-glycan database. Similarly, for the mice were assigned based on the UniProt-Swiss mouse database (17,032 reviewed entries on August 1,2019). The fixed modification was given as carbamidomethyl (Cys) and variable modifications are deamidation of N and Q and methionine oxidation. The trypsin digestion with maximum of 2 missed cleavages were allowed for identification. The mass tolerance for the precursor ion was set to 10ppm and fragment ion was set to 20 ppm. The total false discovery rate (FDR) of 1% was applied to the data. The identified glycoproteins were annotated using UniProt ID

mapping. The quantitation of the glycopeptides were performed using ‘PANDA’ which is annexure software of pGlyco. The PANDA search was performed using the pGlyco search result and the original raw data with allowed FDR of 1%. The intensity with respect to the scan number and retention time can be compared and manually annotated for the glycosite specific glycoform analysis of specific protein. The software extracted intensity of specific glycopeptide related to the antibody was used for calculating site-specific relative abundance (SRA) of each IgG subtypes with total relative abundance (TRA) of all N-glycoforms related to that site. For SRA, the glycopeptide intensity cutoff was set to 3(S:N>3). Then, the peak intensity of each glycoform was normalized to the total signal at each specific site. Similarly, the total glycoprotein composition within the sample was also calculated based on their glycopeptide intensities normalized to the total glycopeptide intensities within the plasma sample.

2.6 Results

The Antibody from the human plasma was successfully isolated and confirmed the enrichment by SDS-PAGE analysis. Whereas, we could not efficiently enrich the antibody from the mice sample. Hence, the IgG specific analysis was restricted to glycopeptide SRA. The N-glycosylation pattern was also compared with the normal patient to screen any possible differences. The HA without inhibitor development and with inhibitor development showed a remarkable difference in galactosylation especially in antibody galactosylation. The patients with lesser galactosylation showed an improvement in galactosylation percentage after treating with emicizumab. The trend was not shown in sample who has no differences in galactosylation. We additionally considered the plasma of two distinctive hemophilia A patients, one with and the other without inhibitor advancement, as the controls for this examination. We also analyzed the mice samples between BALB/c, SV129 and BL-6 CD54.1 before and after inhibitor development, detailed in **section 2.6.4**

2.6.1 Antibody enrichment

The antibody enrichment could get away many proteins from plasma and we can get a plasma cell specific glycosylation at the time of infection. As an underlying investigation, we efficiently separated the IgG from blood plasma by protein A/G conjugated magnetic beads using manufacture's protocol with minute modifications. Further we confirmed the purity of isolated IgG using SDS-PAGE. The results are shown in **figure 2.7.** The blood plasma sample was separated on 10% SDS gel and stained with Coomassie stain.

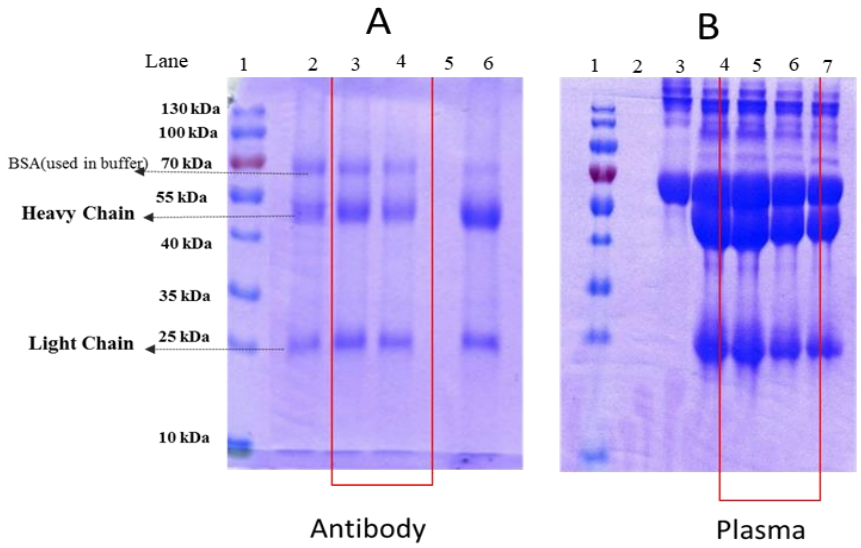


Figure 2.7 SDS-PAGE separation of antibody and plasma
The gel A corresponds to the antibody isolated from plasma. The 50kDa corresponds to the heavy chain and 25kDa corresponds to the light chain. The additional band ~ 70kda band is the BSA in the binding buffer. The well 2 corresponds to HA patient plasma. 3,4 corresponds to plasma with inhibitor development and 6 corresponds to the healthy individual. The gel B corresponds to plasma of different plasma sample the well 3 is BSA. The well 4 corresponds to HA patient plasma. 5,6 corresponds to plasma with inhibitor development and 7 corresponds to the healthy individual. The inhibitor developed one are highlighted.

We could not effectively enrich the mouse antibody from the plasma due to poor binding of mice antibody to protein A/G beads as well as some nonspecific binding. Hence, we limited the antibody enrichment study to only human samples.

2.6.2 MALDI-MS method reliability study and standard spiking study

Reliable glycan analysis work flows should be having a fast sample processing, reliability with , less sample to-sample variation, and increased speed and efficiency. MALDI-TOF-MS

is known to have high coefficients of variation (CVs) for quantification when compared to the commonly used UHPLC methods[86]. It was reported that the S/N ration of the individual should be maintained minimum limit of 10 which is considered as the limit of quantitation of the analytical instrument. The limit of quantitation of different analytical instrument may change due to their variation in sensitivity however when we perform the analysis we have to maintain the S/N 10 which is preferable for quantitation purposes. In all our current analysis we cross verified the individual glycoform which we took for quantitation and relative comparison were above S/N 10. The LOD and LOQ of the permethylated glycans were calculated previously from the lab by Kuan Jiang and detailed in previously published work.[90]The studied with permethylated glycan standard using the same MALDI, and standardized method proves that the glycan is detectable at as low as 50 fmol levels and showed a good signal-to-noise ratio (>10).[90] While checking the data quality during peak picking, it is essential to check the S/N ratio of each peak. This was just to get the clarity about instrument related sensitivity. Here we performed the spiking study was conducted with the common glycoform G1F standard (4 4 0 0 1) to the commercially available human serum. The human serum N-glycome which inherently having a small percentage of 4 4 0 0 1 which we spiked to get the linearity. The relative intensity of the spiked standard was calculated with the 5 4 1 0 0 (G2S1), the inherent glycoform within the serum. The spiking study was showing linearity in their percentage calculation (considering the total glycan intensity). We have used RI calculation of the significant glycoforms concerning the most abundant 5 4 1 0 0 glycoform in the plasma. We adapted the RI calculation of the spiked standard for the common glycoform present in the human serum 5 4 1 0 0. The strength of the linear relation between the values were estimated using R-squared value, also called the coefficient of determination. For this case, in relative intensity of the standard was obtained with R-squared = 0.9761 (Appendix B F3 a1) and the percentage in total intensity was

obtained with R-squared = 0.9995(Appendix B F3 a2). So, we could conclude that the linear fitting is strong in this experiment. 200 picomole spike was out of linear range but considered for SD calculation (shown in the bar graph at APPENDIX BF6). All the acquisition was in triplicate trials, and %RSD was within 14% which in turn prove the method reliability for the current study. All the acquisition was in triplicate trials and %RSD was within 14%. The experiment is proving the consistency in individual measurements where absolute S/N depends on instruments/methods which is previously reported that less than 50 fmol[90]. Hence, it prove that the method is suitable to analyze the biological sample

2.6.3 N-glycan analysis by MALDI-TOF-MS

In the primary screening, we analyzed 11 healthy males varying in their age ethnicity and blood group. The antibody and plasma glycans were compared. Parallelly we have analyzed 4 HA patient with inhibitors and 2 without inhibitors. Three patients were showing with differences in the antibody galactosylation. The antibody galactosylation decreases in response to age and inhibitor development in most of the cases. We have identified around 82 different glycosylation by plasma one-step permethylation in **Appendix B. table 2**. The glycoform mass and their intensity obtained through MALDI-MS spectra after smoothening the spectra and baseline is subtracted. The total glycan species intensity was also calculated from the all the glycoforms identified from the spectra. The percentage distribution was calculated by summing up each glycan species based on their structural specificity (Appendix B table -4).

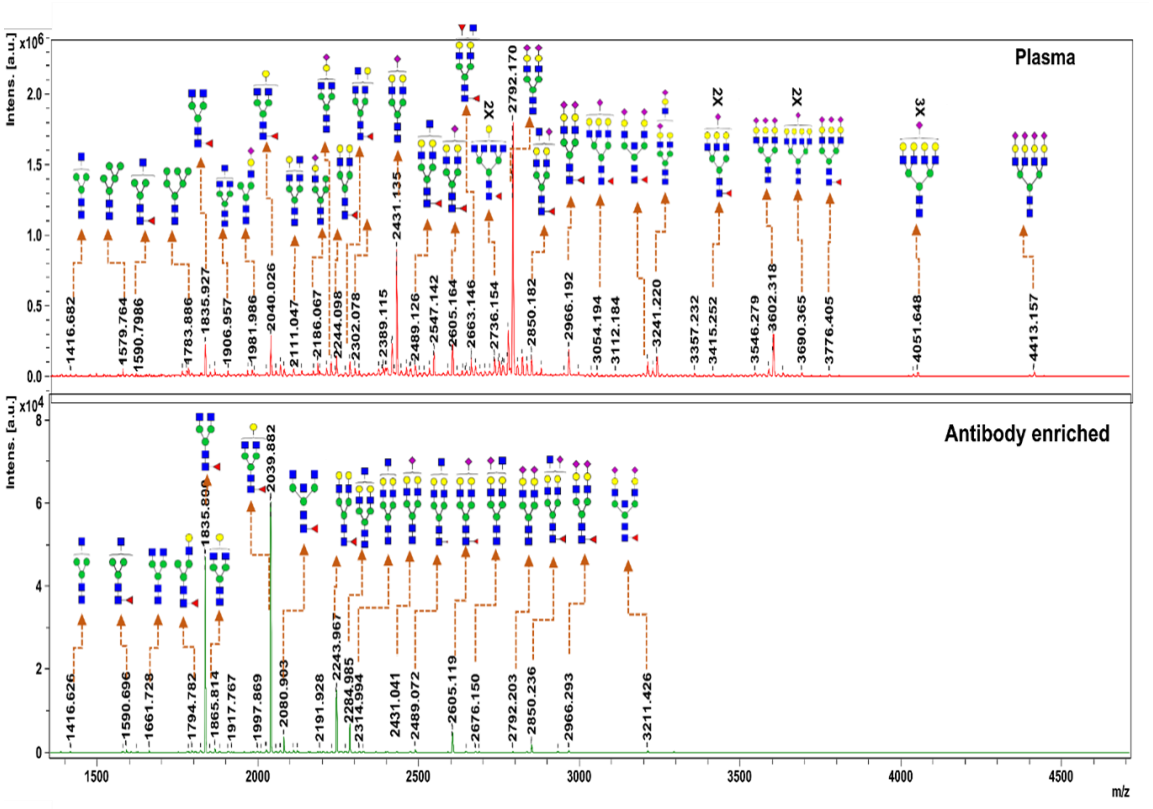


Figure 2.8 MALDI-MS profile of FC N-glycoforms derived from plasma and antibody ■ N-acetylglucosamine (GlcNAc), ■ N-acetyl galactosamine (GalNAc), ● mannose (Man), ● Galactose (Gal), and ◆ N-acetylneuraminic acid (Neu5Ac).

The major glycoforms were identified from the purified antibody is listed in (appendix B. table 3). The fucosylation (mono and multi) galactosylation and sialylation (mono, di and multi) percentages were calculated based on the structural specification identified by MALDI-MS and the peaks are assigned by Glycoworkbench software which is listed in appendix B table 1 and table 2. The major differences were observed in galactosylation. Hence the galactosylation index was calculated using a formula G0f/(G1f+G2f X 2)[176] in which G0f is 3 4 0 0 1 and G1f is 4 4 0 0 1 and G2f is 5 4 0 0 1. The ratio will give the conversion rate of the most abundant glycoform from its agalactosylated states to its possible galactosylated forms.[176] If the ratio is higher, it indicates lower conversion rate of galactosylation hence an indication of aberrant glycan which can be a potential biomarker. The initial study was conducted with a patient with inhibitor and the same patient after treating with emicizumab within 4 years gap. The initial study proved that

as the age increases the galactosylation decreases in antibody. The galactosylation increased in response to the emicizumab treatment. The control plasma sample was selected from the HA patient within the age group of inhibitors developed sample. After emicizumab treatment, the sample was showing a significant increase in galactosylation, which is comparable to control value (Figure 2.9A). Due to the significant differences which we observed in this pair of samples, we further studied the additional analysis including NP-HPLC after treating the glycan with neuraminidase and derivatized with 2-AB. The increase in G2 peak was indicating the percentage of sialic acid-containing sugar in the real sample (Figure2.11 C), and the technique was complementing the MALDI-MS analysis. The two-step permethylation process was complementing the one-step result and separating the sialic acid containing sugars based on their linkage specificities (Figure2.11 D). Both α-2,3 and α-2,6 sialylations were showing a significant decrease in the pretreatment stage when compared to post-treatment and control samples. All these results were complementing each other and showing a significant difference in the N-glycome of plasma at the pretreatment stage. The results also gave us the MALDI-MS method reliability to distinguish the differences between the samples.

To confirm the observation, we increased the number of sample pairs and performed the antibody galactosylation index study. The results did not show consistent variation between pre and post treatment samples (Figure 2.9B). We could not conclude the results because each pair was varying in the inhibitor value. However, the inhibitor negative(control) samples never showed any changes in galactosylation values after treating with emicizumab. This observation was proving that the emicizumab treatment itself cannot have any effect on galactosylation improvement. We further arranged the data based on the inhibitor development and emicizumab treatment. We could see that half of the population with

inhibitor development showing higher galactosylation index whereas the inhibitor negative and emicizumab treated sample does not show this trend.

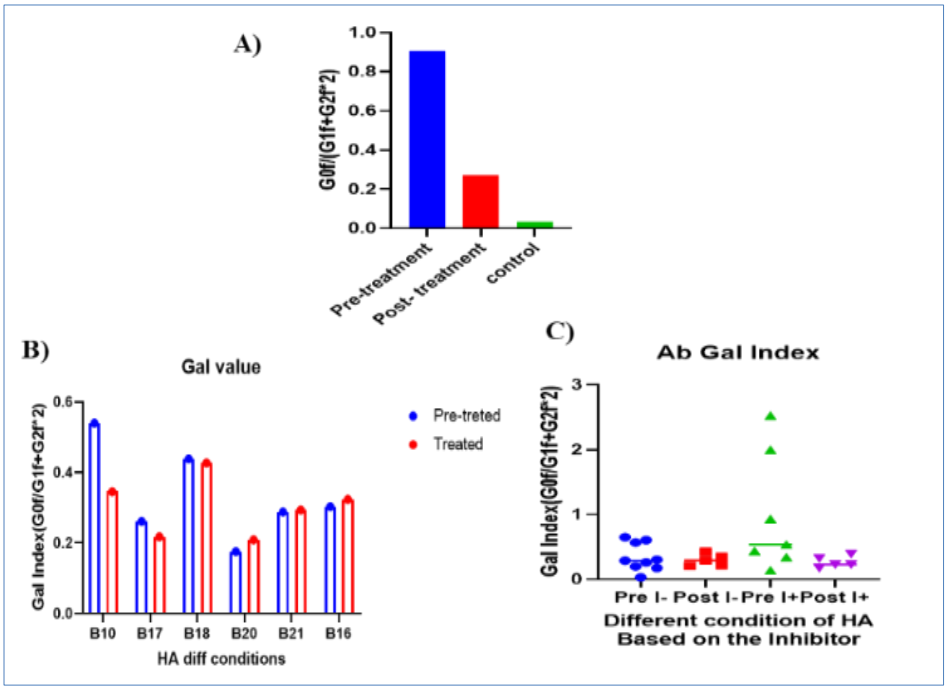


Figure 2.9 Protein A/G enriched galactosylation index (G I) of the same person before and after emicizumab treatment.
A) The Galactosylation Index of one patient with high proinflammatory signal with inhibitor development before and after emicizumab B) The study repeated with more pair of emicizumab treatment with varying inhibitor development and age. C) the galactosylation value arranged based on Inhibitor development and emicizumab treatment with a greater number of samples.

2.6.4 Human plasma N- glycome analysis

In this study, we categorized the relative intensity of the most intense 33 glycoforms, and their intensities were plotted using Graph Pad Prism 8.0 software. The data acquisitions were made in triplicate to confirm the repeatability of the spectra. The sample, which showed a significant increase in the antibody galactosylation index, was again showing a significant difference in plasma glycosylation profile. In the pretreatment stage, the glycoforms were less galactosylated and sialylated. After treating with emicizumab the glycosylation pattern was improved with higher galactosylation and sialylation and became comparable to the control sample. In the pretreatment stage, fucosylated glycoforms were higher in amount which get decreased after emicizumab treatment. We also observed less amount of complex multi-

branched structures, which improved after post-treatment. Due to the significant differences which we observed in this pair of samples, we further studied the additional analysis including NP-HPLC after treating the glycan with neuraminidase and derivatized with 2-AB. The increase in G2 peak was indicating the percentage of sialic acid-containing sugar in the real sample (Figure2.11 C), and the technique was complementing the MALDI-MS analysis. The two-step permethylation process was complementing the one-step result and separating the linkage specific differences of sialylation (Figure2.11D). Both α-2,3 and α-2,6 sialylations were showing a significant decrease in the pretreatment stage when compared to post-treatment and control samples. All these results were complementing each other and showing a significant difference in the N-glycome of plasma at the pretreatment stage. The results also gave us the MALDI-MS method reliability to distinguish the differences between the samples. However, we wanted to further confirm the reliability of the method with a spiking study using the commercially available serum. The 4 4 0 0 1 standard spiked study was showing linearity in their percentage calculation (considering the total glycan intensity) from 1800-4500 da mass range. We have used RI calculation of the significant glycoforms concerning the most abundant 5 4 1 0 0 glycoform in the plasma. We adapted the RI calculation of the spiked standard for the common glycoform present in the human serum 5 4 1 0 0. The linearity of the percentage was obtained with an R^2 value of 0.995. The Linearity of RI was obtained with an R2 value of 0.976. 200 picomole spike was out of linear range but considered for SD calculation (shown in the bar graph at APPENDIX BF6). All the acquisition was in triplicate trials, and %RSD was within 14%

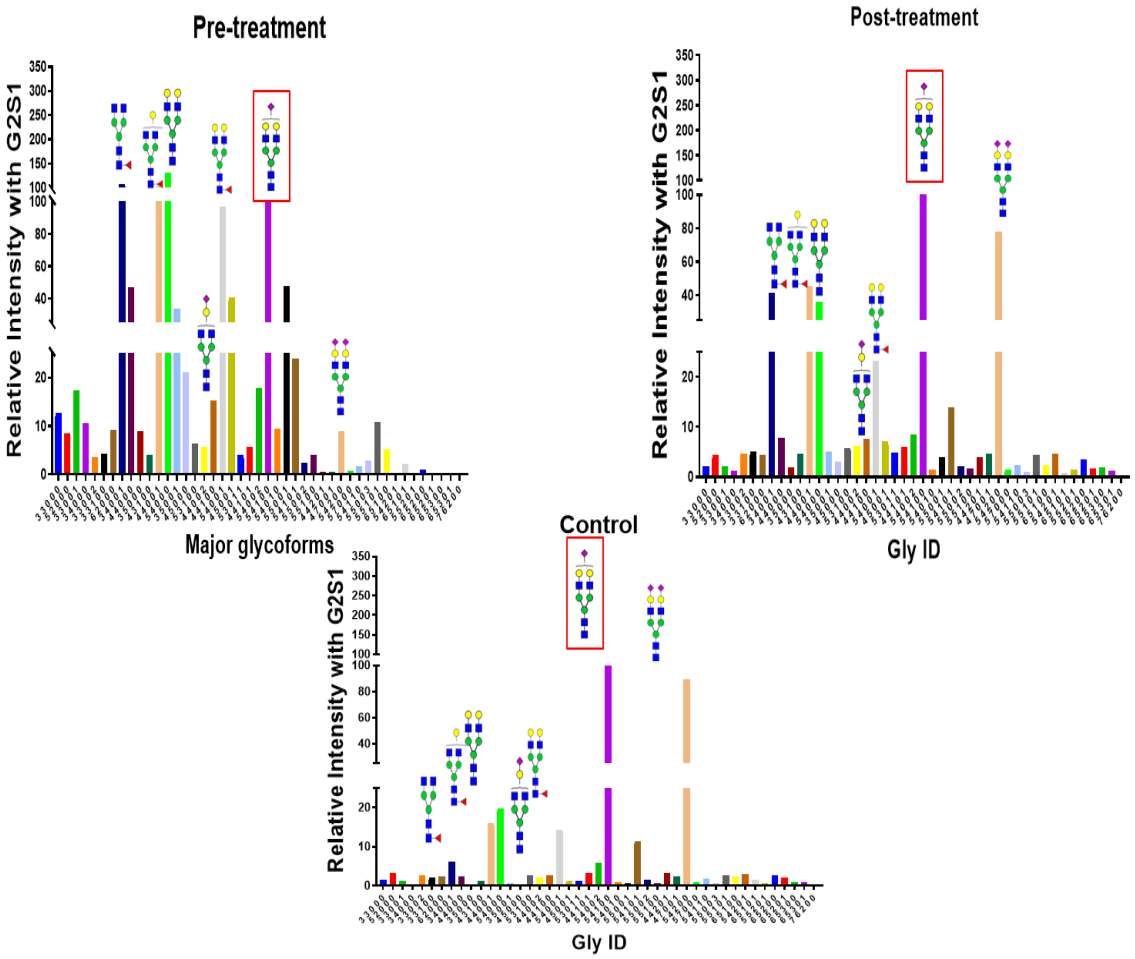


Figure 2.10 MALDI-MS profile of FC N-glycoforms derived from plasma of the same patient before and after emicizumab treatment in comparison with control without Inhibitor and same age.

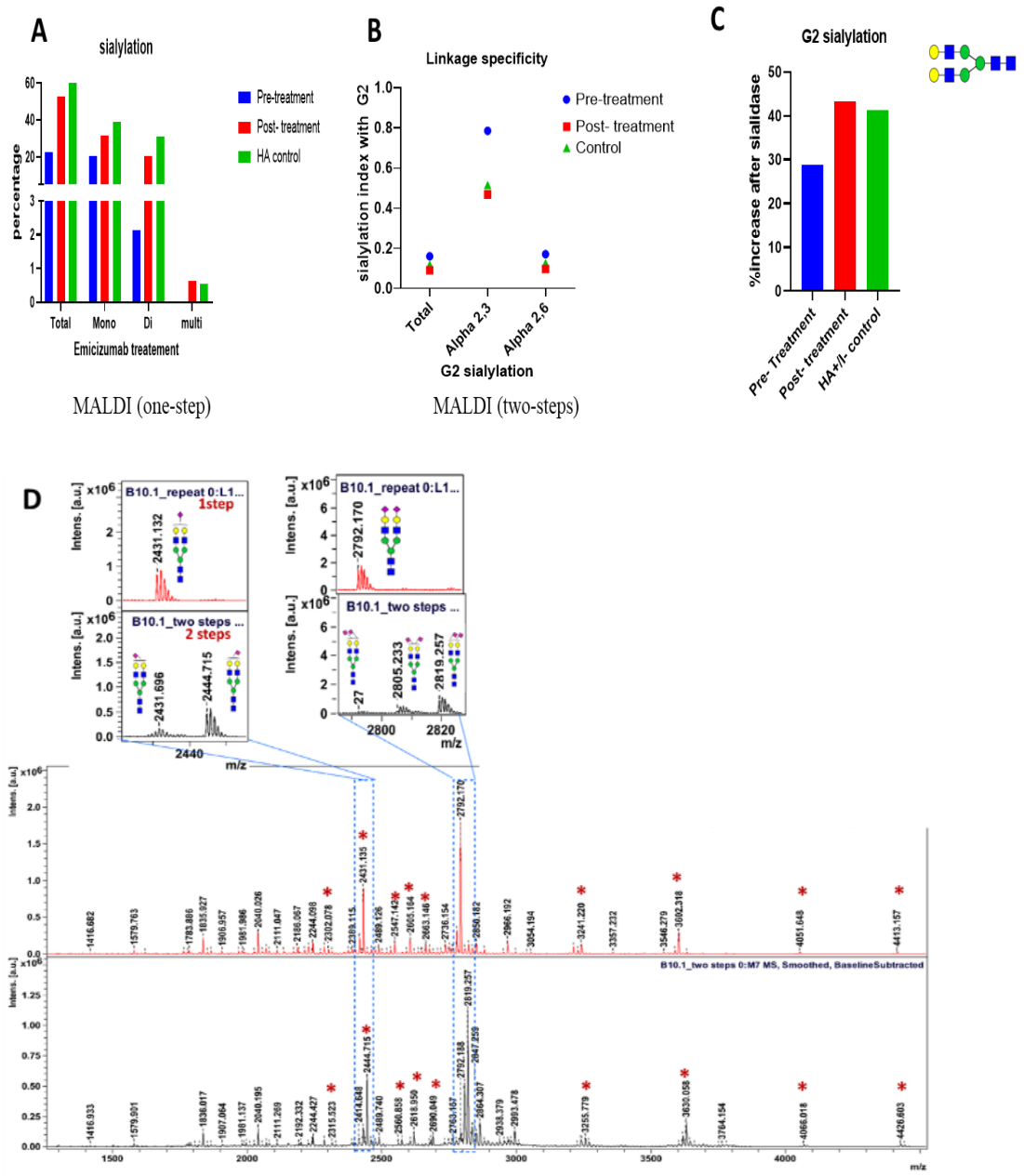


Figure 2.11The glycosylation pattern differences calculated by different analytical techniques using G2 sialylation as an example.

A)The plasma sialylation, identified based on the intensity of each glycoform obtained by permethylation analysis using MALDI-MS .B and D) linkage specificity of sialic acid by twostep method based on 54000 B indicate the SI value of each sample and D indicate the spectra showing differences in 1 step and 2 step permethylation differences wrf 54000 (G2) glycoform. C) The sialylation percentage was also estimated with NP-HPLC by injecting each sample before and after neuraminidase digestion. The percentage of G2 peak was estimated after subtracting the amount in before sialidase digestion. The study also indicate pretreatment stage shows significant decrease in sialylation.

Hence, we increased the sample number to relate the changes with inhibitor development. In contrast to the previous trial with one pair, the variations were insignificant in larger cohort (Figure 2.13) and individual dependent. Hence, we could not relate to the emicizumab treatment and Inhibitor development. Refer appendix B table1-4 and Figure 1 and 2.

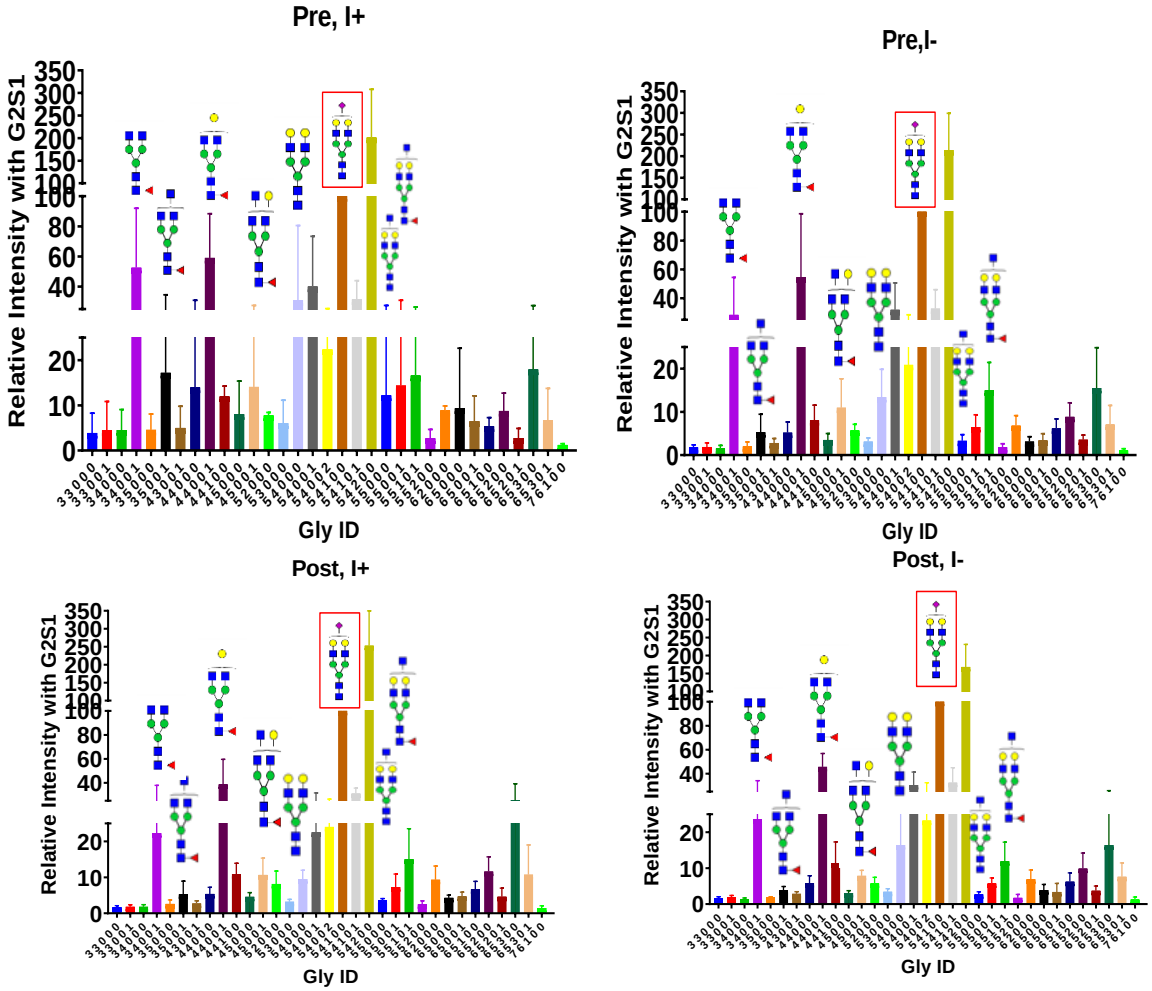


Figure 2.12 Changes of N-glycome composition human plasma from different HA patients based on the emicizumab treatment and inhibitor development.
The most abundant 33 glycoforms are compared between different HA condition based on the Inhibitor treatment. The relative abundance was calculated with reference to G2S1(5 4 1 0 0). The patient with inhibitors is I+ and without inhibitors I-. The changes in glycoforms before and after the treatment is denoted as pretreatment(pre) and post treatment (Post).

The relative intensities of individual glycoform were calculated with the intensity of 5 4 1 0 0 glycoform and compared the result before and after emicizumab treatment in inhibitor negative and positive cohort. We additionally examined the glycoforms based on

their structural specificity and then categorized each glycoform into the primary type like complex, hybrid, and high mannose. The glycoforms were further categorized based on their specific sugar composition and linkage. The main categories are based on fucosylation, sialylation, and their specificity in linkage as well as the number of existences. The specific sialic acid linkage was calculated based on G2 glycoform based analysis using the formula SI= G2/ G2S1+G2S2 X 2). The human categorization is listed based on the **Appendix B table 4** and the statistical correlation is listed in **figure 2.13.**

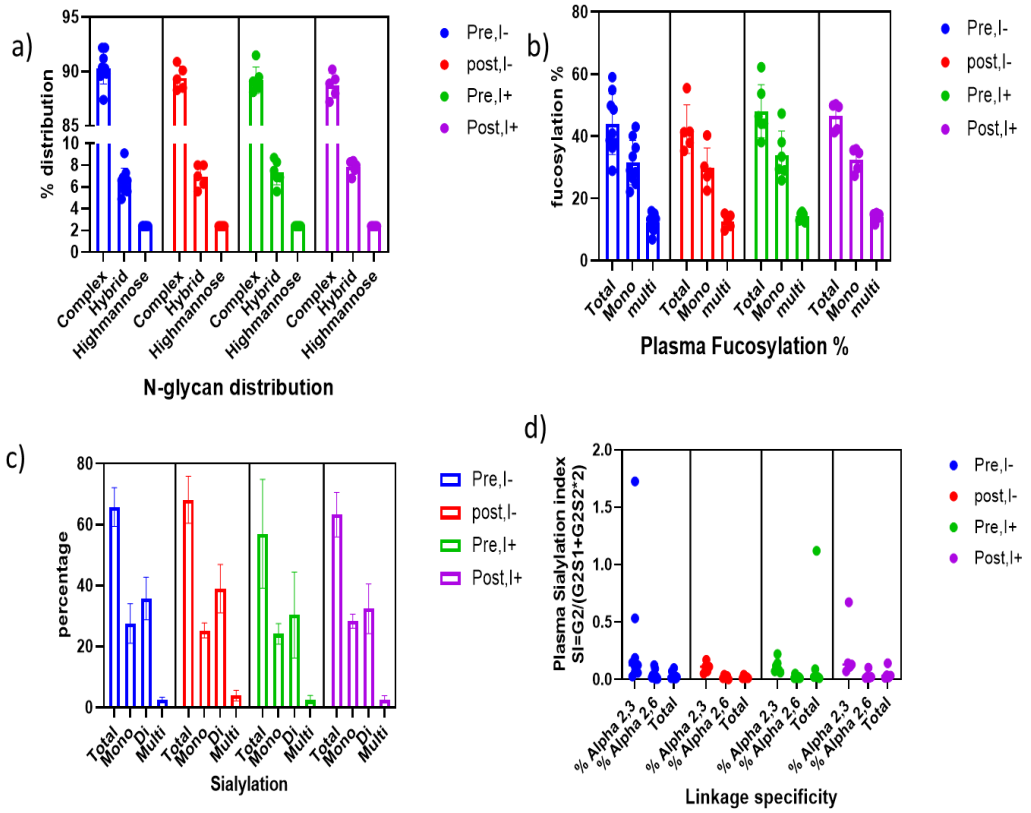


Figure 2.13Changes of N-glycome composition based on the types.
a) N-glycans based on their major three category complex, hybrid and high mannose. b) based on the fucosylation c) based on sialylation percentage d) plasma sialylation index calculated based on G2 glycoform in plasma using the formula SI=G2/(G2S1+G1S2x2).

2.6.5 Mouse plasma N- glycome analysis

The plasma glycome analysis was identified similarly like in human and compiled in **Appendix B table 10.** We have identified around 89 different glycoforms by one-step permethylation without considering the sialic acid linkage specificities. The glycoforms

varies in their percentage of composition between the stains. The glycoforms were also compiled based on the percentage of composition between the stains. We have identified the glycoforms similarly in all the three strains. However, the proportions were slightly varying.

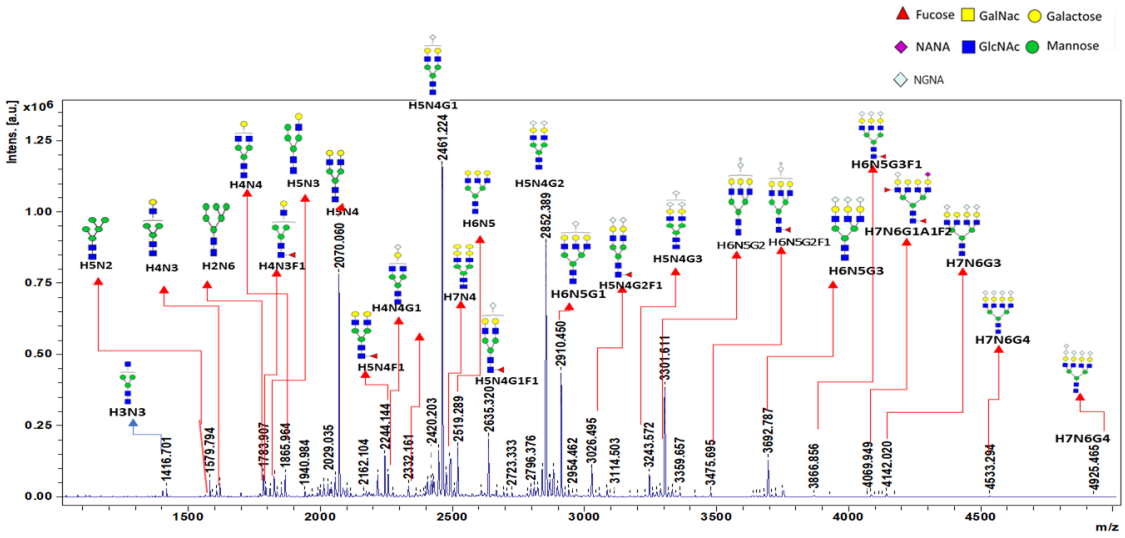


Figure 2.14 MALDI-MS profile of N-glycoforms of mouse plasma

In this study, the spectra were more closely matching between their baseline-inhibitor development studies in comparison to human trials. The main differences between human and mice plasma N-glycome were the presence of α-gal and specific sialic acid, Neu5Gc, instead of Neu5Ac in the structure. The specific sugar compositions are also different between the species as the difference increases the glycoform signatures also changes Mainly, human glycoforms have a more non-fucosylated complex structure than in mice. However, we compared the glycoforms similarly as we strategized for human samples.

The N-glycome pattern changes during inhibitor development and between the strains were monitored. The data acquisitions were made in triplicate to confirm the repeatability of the spectra. As a background analysis, we compared the HA wildtype mice sample with BL-6 strain to confirm there is no variation in their plasma glycome during HA condition. Also, 11 weeks and 15-weeks of HA mice plasma galactosylation were studied to confirm the age-related and HA related ambiguity of the data. In both the cases of age and HA does not show

any significant differences in glycosylation (data is not shown). In general, only a slight change was observed between the strains. However, the existing differences were more evident in BALB/c mice strain than the other two mice strains (Appendix B F17, Figure 2.15). The differences were not statistically significant (P value < 0.05). The galactosylation, fucosylation, and sialylation were also considered for the comparison. The mice sample also showed increased inhibitor development (Figure 2.15a). The fucosylation between the strains and during the inhibitor development was also monitored (Figure 2.15 b, c, d, e), which did not show any significant differences. However, we could see individual-based and strain-based differences (Appendix B F7). Henceforth, the distinctions were not sufficiently able to relate with the inhibitor improvement. The sialylation, both of percentage and the linkage specificity did not indicate a huge contrast in inhibitor development (Figure 2.16 and Figure 2.17).

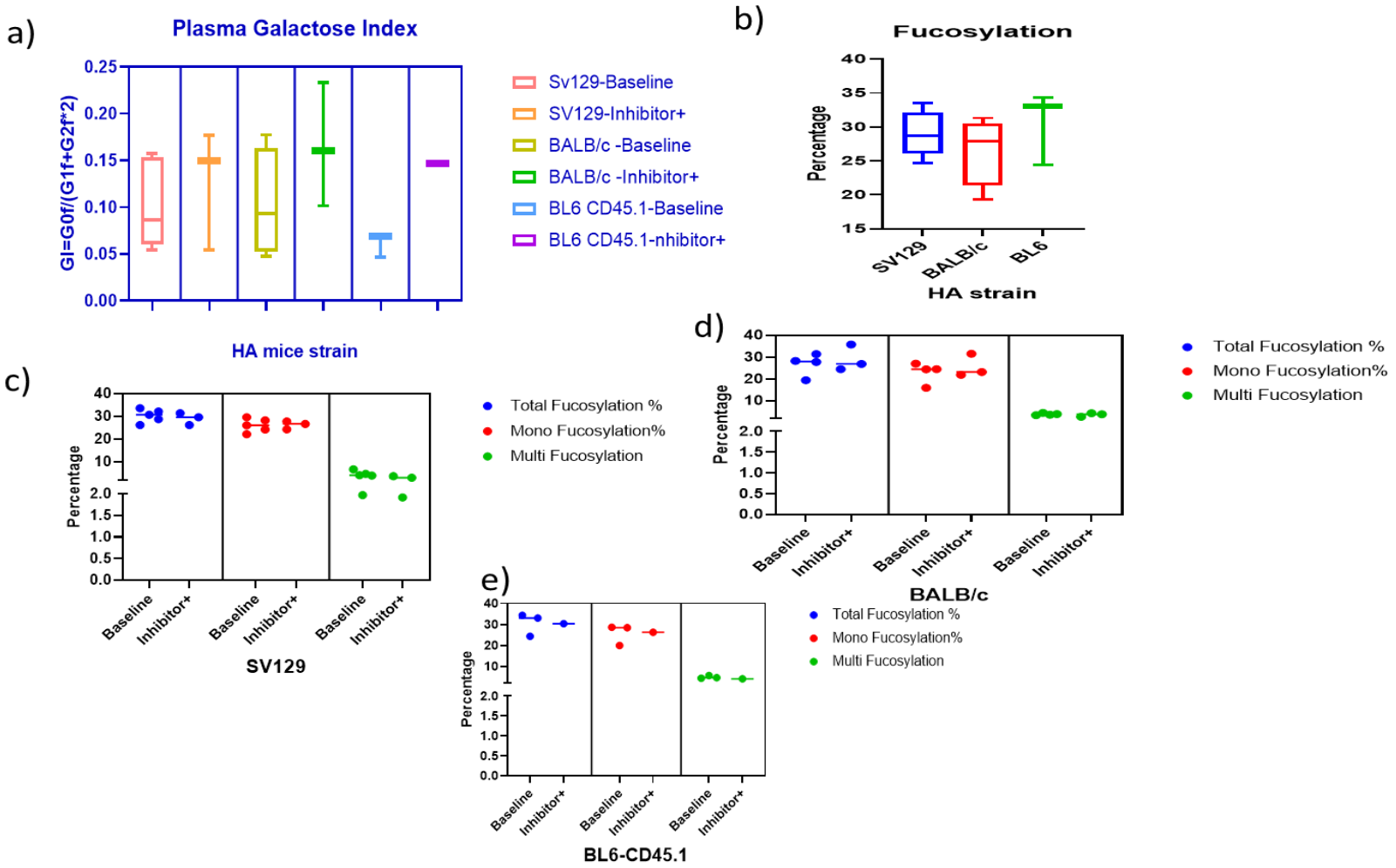


Figure 2.15 Mice plasma N-Glycan profile change compiled based on specific sugars
a) Plasma galactosylation index b) Total fucosylated species percentage in inter-strain .c) effect of inhibitor development in total fucosylation in plasma in SV129 d) BALB/c e) BL6-CD.45.1

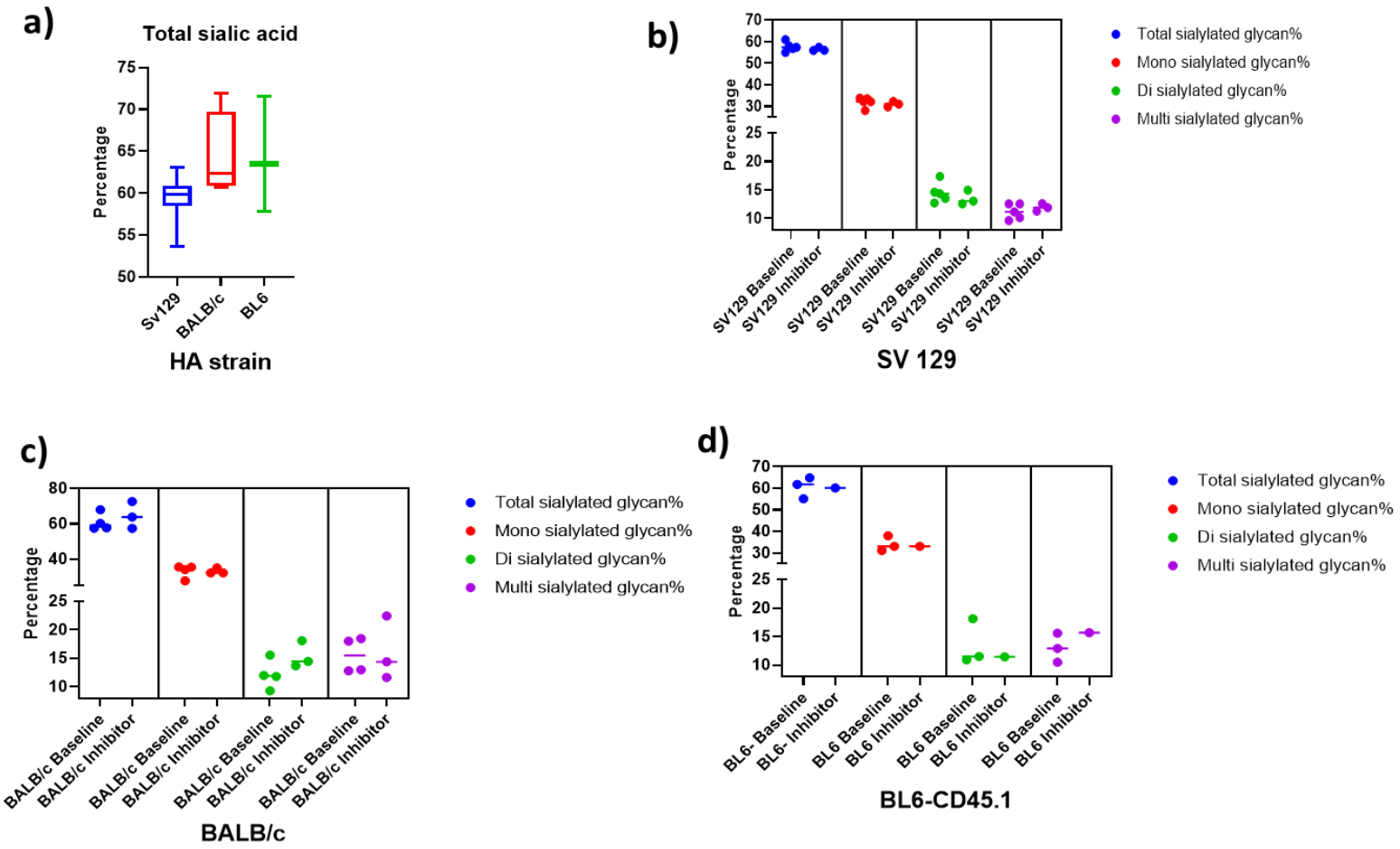


Figure 2.16 Mice sialylation changes during inhibitor development.
a) Total sialylation percentages between the strains. b) Sialylation changes in SV129 strain c) BALB/c strain and d)BL-6 CD45.1 .

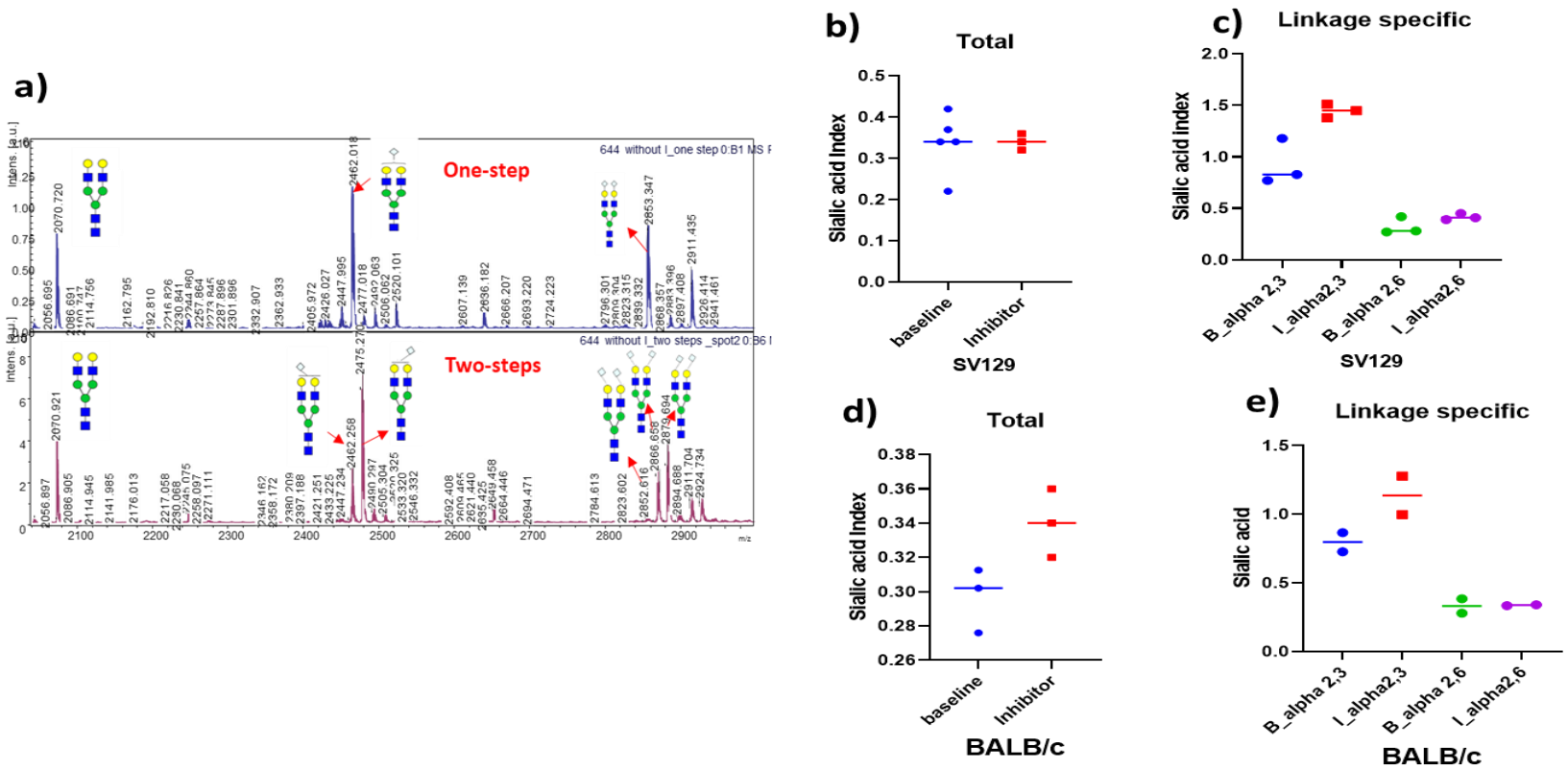


Figure 2.17 Sialic acid composition analysis between the strains during Inhibitor development
a) Sialic acid linkage identification differences based on one step and two steps permethylation process labeled with G2S1 and G2S2 b) Sialic acid index calculated based on the equation G2/G2S1+G2S2*2 by one step equation in SV129 strain. c) Sialic acid index calculated based on the equation G2/G2S1+G2S2*2 by two steps permethylation to distinguish alpha 2,3 and alpha 2,6 linkage analysis in SV129 strain. d) Sialic acid index calculated based on the equation G2/G2S1+G2S2*2 by one step equation in BALB/c strain. c) Sialic acid index calculated based on the equation G2/G2S1+G2S2*2 by two steps permethylation to distinguish alpha 2,3 and alpha 2,6 linkage analysis in BALB/c strain.

2.6.6 Human glycopeptide analysis and bottom-up approach

Plasma glycoproteins were subjected to tryptic digestion by FASP. Then, the glycopeptides were enriched by HILIC chromatography and injected into orbitrap. The peptides are separated on an RP column prior to connecting to the orbitrap machine, and the spectra were further analyzed by peptide fragment fingerprinting (PFF) using pGlyco software using Uniprot human database. The identified peptides were quantitated using the PANDA software using the unique scan number and retention time differences. Hence we could semiquantitate the glycoprotein present in the sample based on the abundance of glycopeptide related to specific proteins in the sample.We couldnot see a significant differences related to their composition

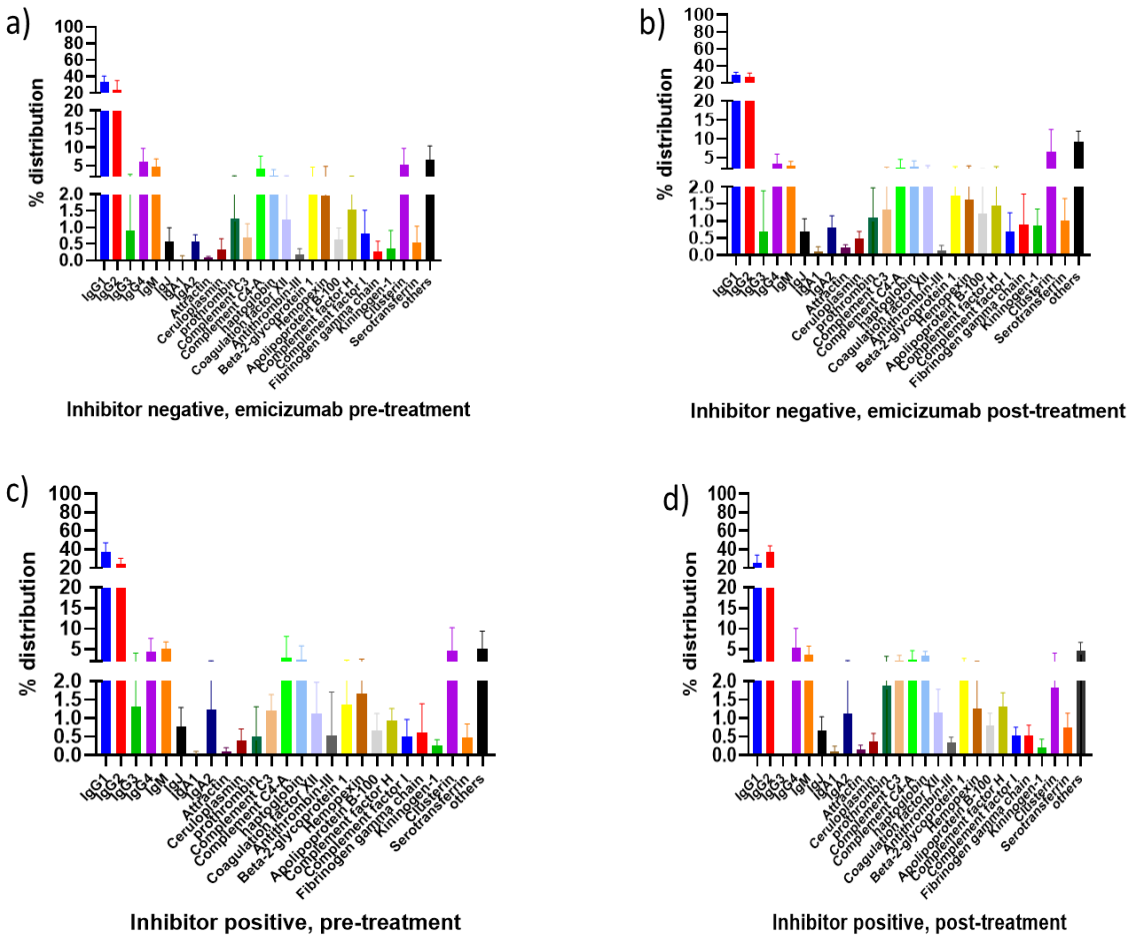


Figure 2.18 The glycoprotein profiling based on the bottom-up approach in each category of plasma protein analysis.
a) inhibitor negative and emicizumab pre-treatment b) inhibitor negative and emicizumab post- treatment. c) inhibitor positive pre-treatment and inhibitor positive post-treatment

2.6.7 Microheterogeneity identification at IgG subtyping level

The site-specific glycan heterogeneity was performed through trypsin-proteolysis followed by offline HILIC-enrichment, leading to RP-HR-MS/MS through HCD fragmentation. The site-specific assignment of the glycan on the glycopeptide to segregate the changes in IgG subtype level, which may give more insight to the functional changes happening through the inhibitor development also during emicizumab treatment. The study was conducted to provide more insight into the initial observation received at glycan analysis during both the cases. The spectra were analyzed using pGlyco software, which detected the glycoproteins also provided the glycopeptides and their site specific microheterogeneity. The

Annexure Manual characterization was simultaneously carried out on N- and O-glycopeptide HCD spectra exhibiting signature fragments of oxonium ions. All observed N- and O-glycoform compositions are summarized in Table 1.

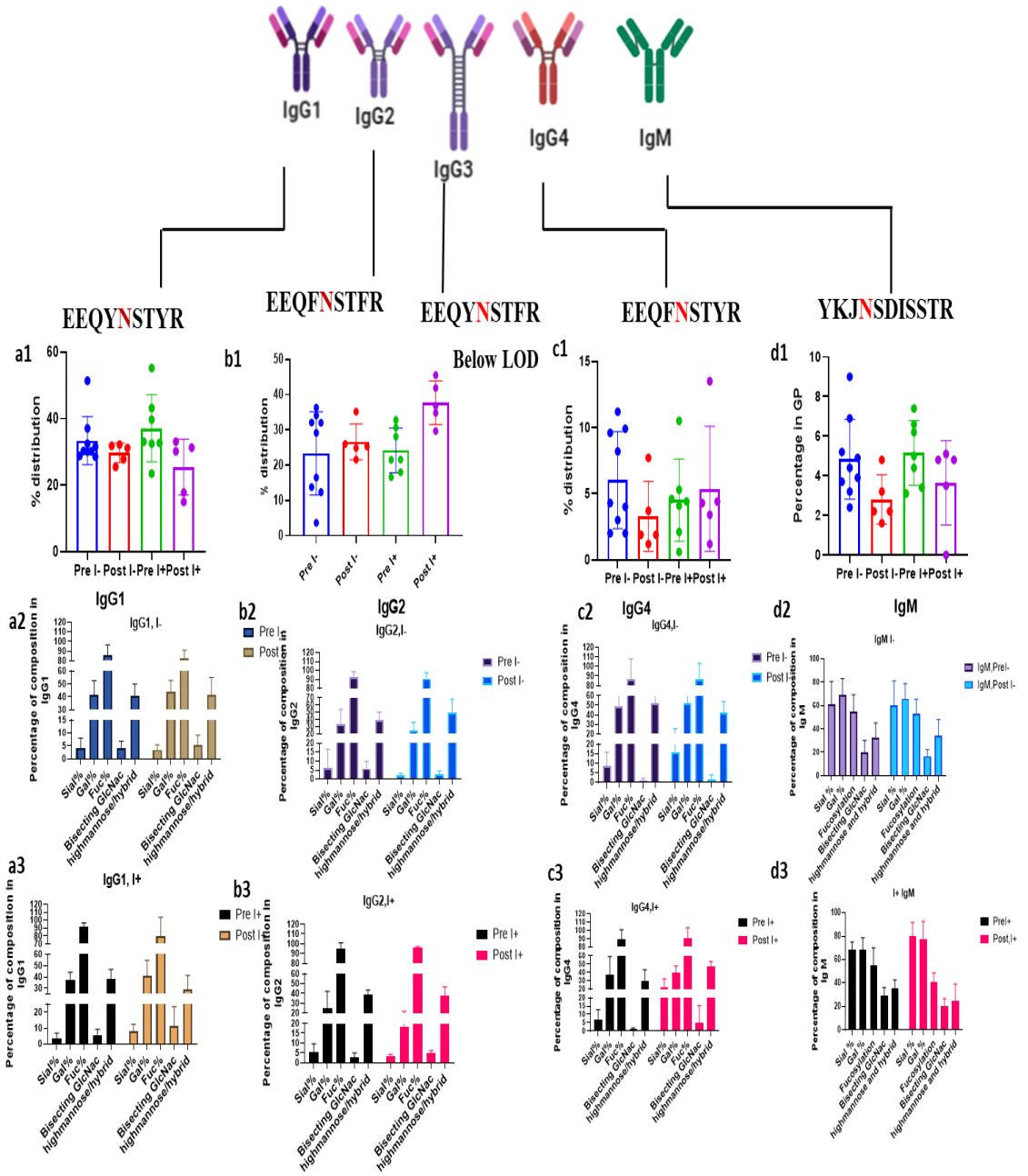


Figure 2.19 IgG subtyping expression and their microheterogeneity compilation
a1-d1)The IgG subtyping based on the bottom-up approach we compared IgG1,2 ,4 based on the glycopeptide composition in plasma .IgG3 was below LOD. IgM was also included based on their N-glycopeptide composition. a2-d2) The comparison of site-specific glycosylation differences of emicizumab treatment by comparing the glycosylation pattern in inhibitor negative samples. a3-d3) The comparison of site-specific glycosylation differences of emicizumab treatment by comparing the glycosylation pattern in inhibitor positive samples.

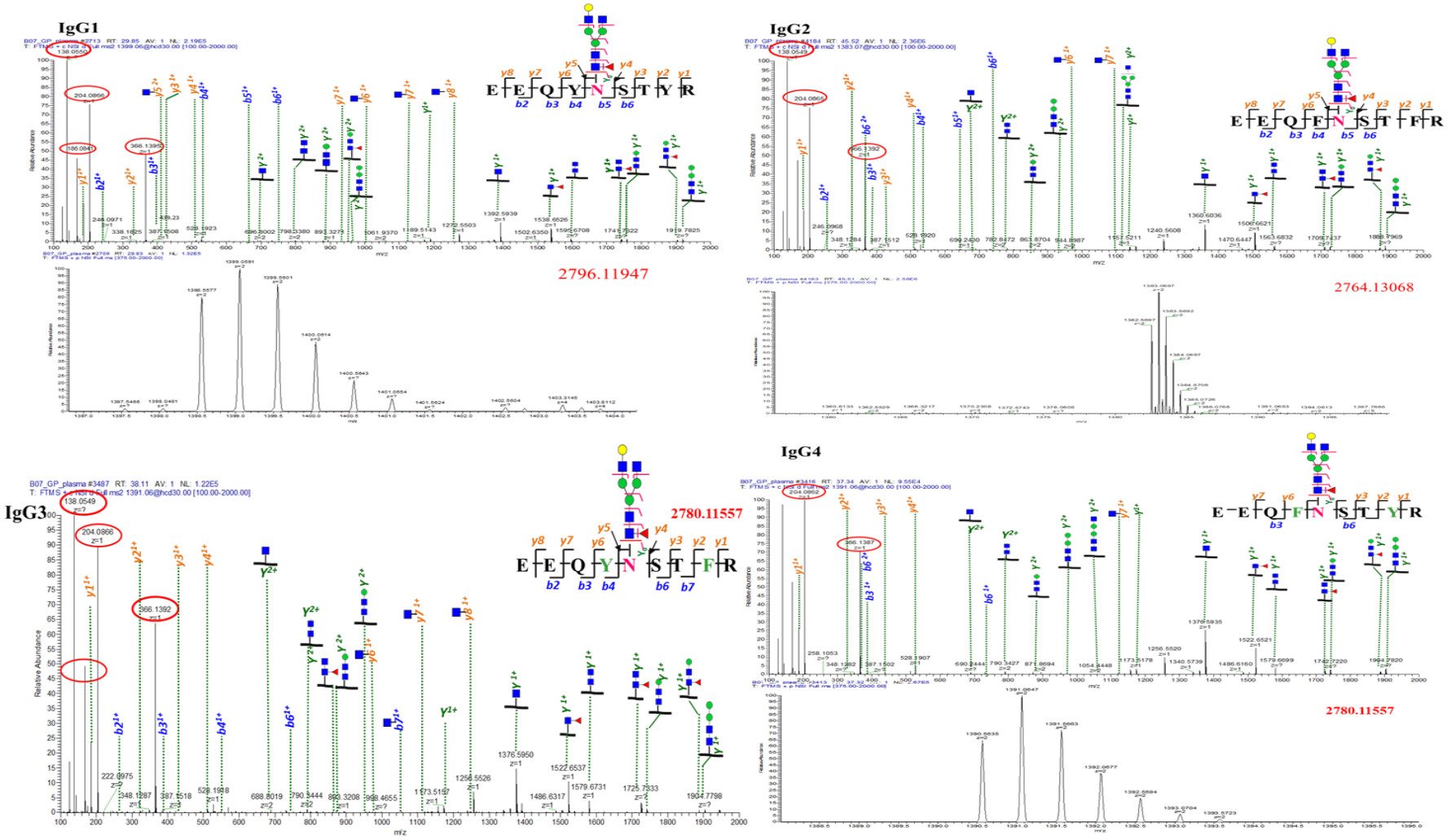


Figure 2.20 HCD fragmentation pattern of the human IgG subtyping(IgG1,2,3 and 4)
identification based on their subtle differences in their peptide back bone. Here we selected a common glycoform for showing similarity and differences in fragmentation process.

2.6.8 Mouse glycopeptide analysis (Bottom Up-Approach)

Plasma glycoproteins were subjected to tryptic digestion by FASP. Then, the glycopeptides were enriched by HILIC chromatography and injected into orbitrap. The peptides are separated on an RP column prior connecting to the orbitrap machine, and the spectra were further analyzed by peptide fragment fingerprinting (PFF) using pGlyco software using UniProt human database. The identified peptides were quantitated using the PANDA software using the unique scan number and retention time differences. Hence we could semiquantitate the glycoprotein present in the sample based on the abundance of glycopeptide related to specific proteins in the sample.we could see a significant differences in IgG related glycopeptides in their expression

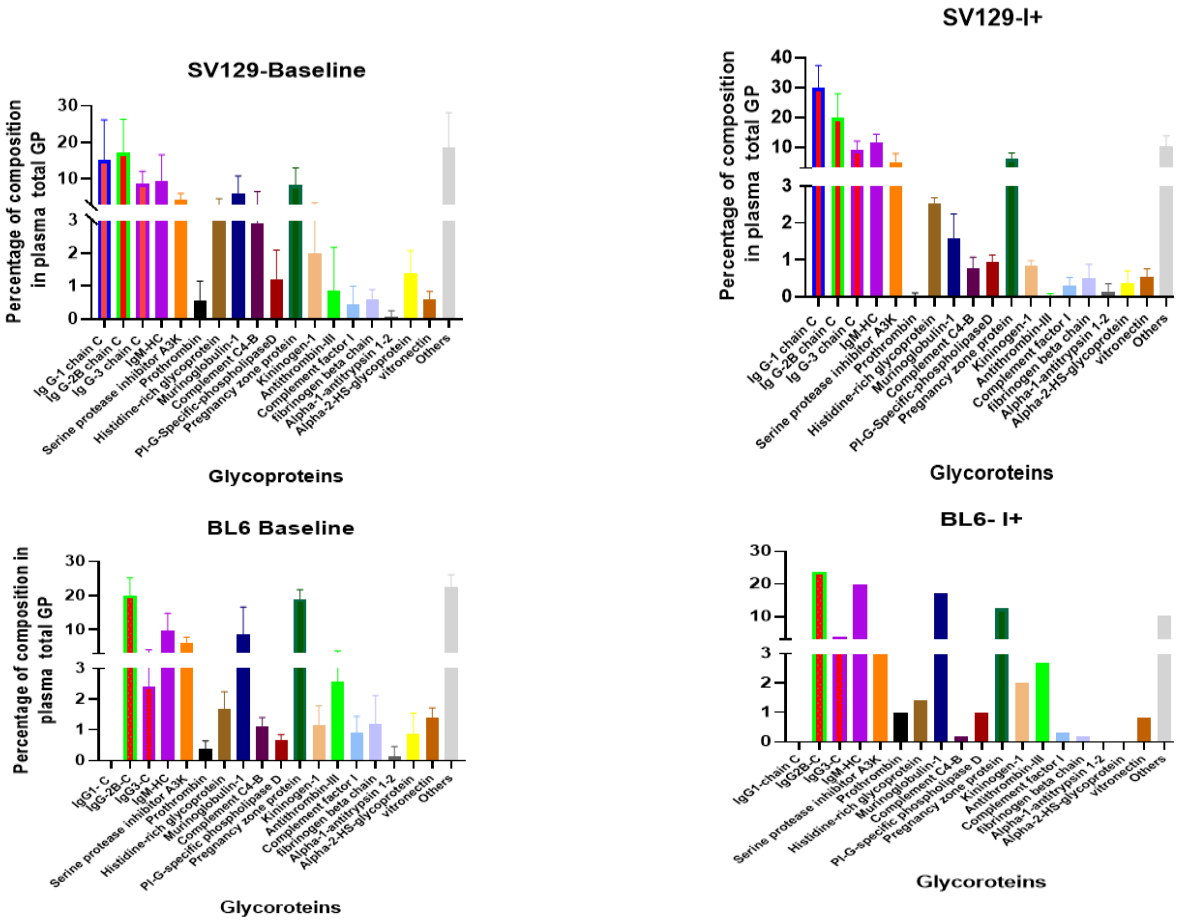


Figure 2.21The overall glycoprotein detection and quantitation based on bottom up approach.
Here in both the strains there is an increase related to IgG and IgM. BL-6 did not the IgG1 expression in all the individual sample may related to strain specificity.

There is a strain-specific trend observed in BL6 and SV129. IgG1 was showing below the LOD level in BL6. BALB/c was showing significant improvement in the expression of the IgG and IgM. Hence we concentrated on BALB/c strain to study the microheterogeneity of the IgG subtypes, detailed in **Figure 2.22** .

Here the baseline/inhibitor negative (I-) samples are showing more sialylation, especially monosialylation in comparison to the inhibitor developed sample (I+). The result was constant in IgG1,2B and 3 subtypes. The observation was matching with the human samples as sialylation and galactosylation are anti-inflammatory glycan relating to antibody. The tandem mass spectrometry profile was (Figure **2.23)** relating to IgG subtyping N-glycopeptides.

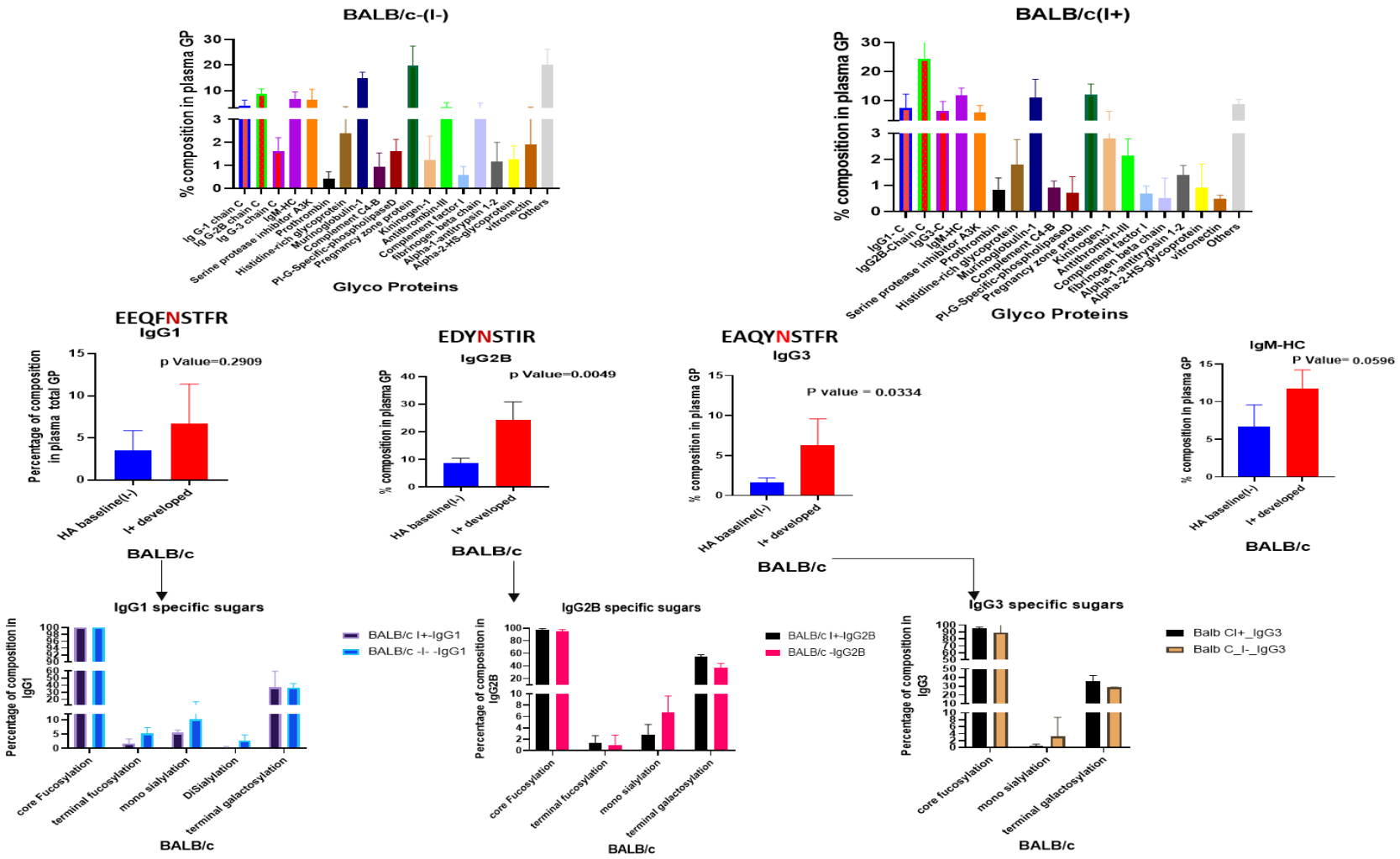


Figure 2.22 a1-d1) The IgG subtyping based on the bottom-up approach.
we compared IgG1,2b,3 and IgM composition based on their N-glycopeptide composition in plasma. IgG3 was below LOD. IgM was also included based on their N-glycopeptide composition. a2-d2) The comparison of site-specific glycosylation differences of emicizumab treatment by comparing the glycosylation pattern in inhibitor negative samples. a3-d3) The comparison of site-specific glycosylation differences of emicizumab treatment by comparing the glycosylation pattern in inhibitor positive samples.

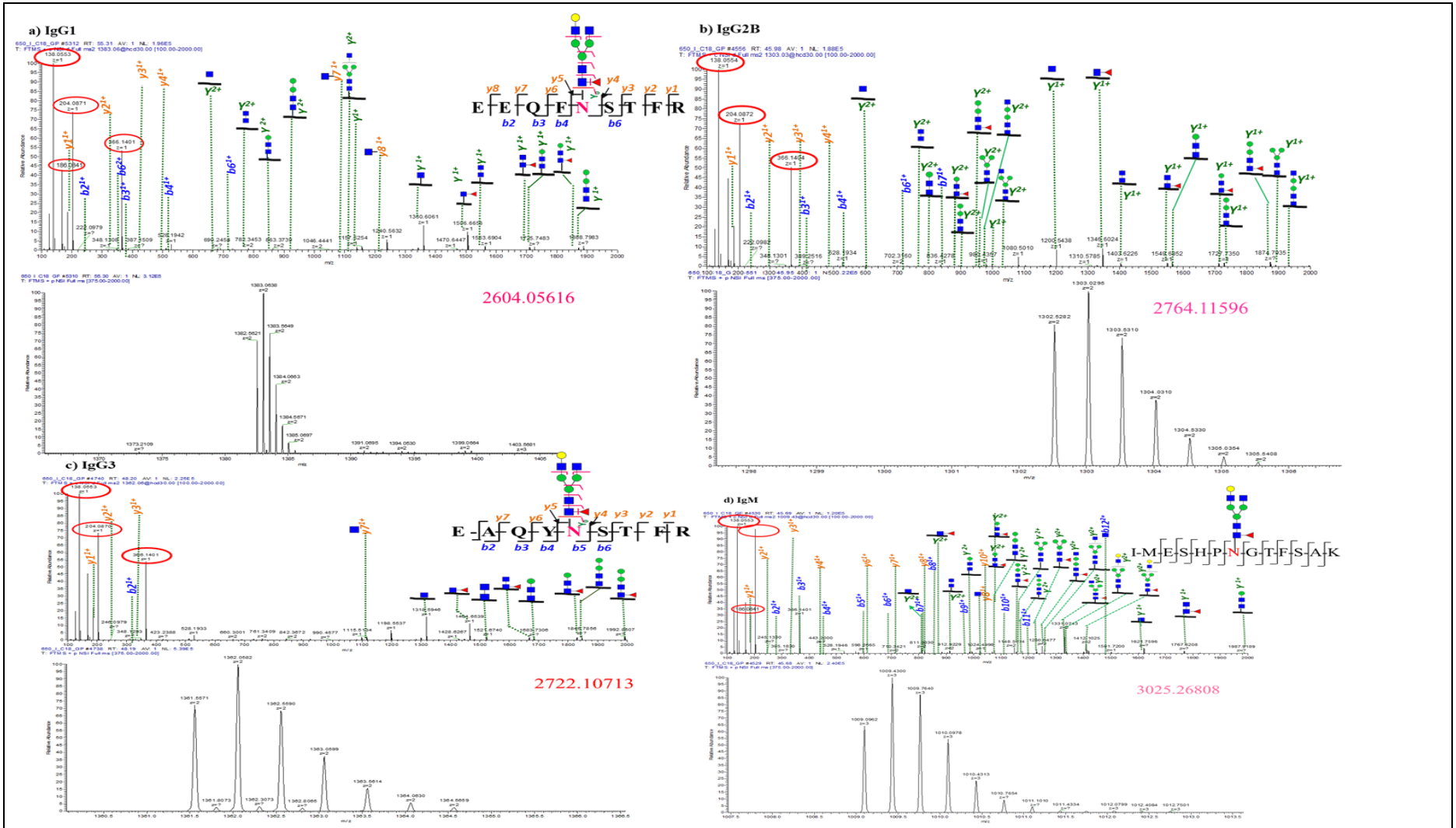


Figure 2.23 HCD fragmentation pattern of the mouse IgG subtyping.
IgG1,2,3 and IgM identified based on their subtle differences in their peptide back bone. Here we selected a common glycoform 4 4 0 0 1 for showing similarity and differences in fragmentation process. a)IgG1,b)IgG2B c)IgG3 and d)IgM

2.7 Discussion

FVIII neutralizing antibodies, or inhibitors, happens to 25-30% of severe HA patients receiving rFVIII replacement therapy. The inhibitor development against rFVIII is challenging in the treatment perspective, and the causes of inhibitor development is still unknown. It has been realized the distinctions in the glycomics can be a biomarker of the physiological changes occurring because of the disease progression and in light of the treatment. The recent development of understanding glycosylation reveals the importance of glycan in antigenicity immunogenicity, physiological changes, disease progression, and personalized medicine development. In order to investigate the effect of the host glyco environment on inhibitor development, based on our developed MS-based approach, glycosylation pattern of total immunoglobulins, and overall glycome changes of the host plasma environment was thoroughly characterized in a number of HA patients' samples, .The advancement in the mass spectrometry analyzers, software, and databases helps glycobiologists analyze complex biological samples and understand the role of glycans in the natural system. The improvement in the analytical techniques, including sample preparation and separation technologies, help promote the application of glycobiology in prognostic and diagnostic aspects of diseases. All these help us understand the role of glycosylation changes during the inhibitor development and build up a corresponding structural library of glycoforms. Here, we identified plasma and antibody N-glycome of human HA patients who developed inhibitors and their negative control of the average HA population. We could not see any unusual glycoforms related to inhibitor development. Hence, we decided to relative quantitate the sugars with respect to the most intense glycoform G2S1(5 4 1 0 0 /H5N4A1) in plasma and G1f (4 4 0 0 1/H4N4F1) in the antibody. We checked the reliability of the MALDI-MS method by spiking the 4 4 0 0 1 glycoform to human plasma and calculating the relative intensity with 5 4 1 0 0 glycoforms,

which already present in the serum. We also calculated the linearity of the percentage of spiked standards. After confirming the reliability of the method to semi-quantitate the glycoform in test plasma, we monitored the variability of most intense 33 glycoforms in the plasma. Additionally, we categorized the identified glycan based on their structural specificities, and their percentage was calculated by summing up the intensities of all the glycoforms within each category and divided it with total intensities and converting it to percentage. The sialic acid linkage specificities were also calculated with the conversion of G2 glycoform to its sialylated forms based on their linkage. We could see some individuals who developed inhibitor showed lesser galactosylation, especially reflected in antibody galactosylation index calculation. We conducted a background analysis to study the other factors which affect the glycosylation changes. Factors like age, hemophilia A, inhibitor development, blood group, and ethnicity were tested to check the N-glycome change. The main elements which influenced the N-glycome profile were age and inhibitor advancement. Both the cases indicated a huge in galactosylation and sialylation decrease of their antibody. Here, we noted the fact that the test study must be correlated with the individuals of similar age groups. We additionally analyzed the Plasma N-glycome of the inhibitor developed individual who showed a huge difference in the plasma and antibody glycosylation pattern at the pre-treatment and post-treatment stage of emicizumab treatment. The examination indicated a momentous change in the glycosylation profile after emicizumab treatment and the profile was coordinating with the control HA patient with a similar age group and who is inhibitor negative. Hence, we developed the glycopeptide analysis strategy with the same patient who showed a huge difference in their IgG sub type-expression and their site-specific glycosylation pattern (Appendix B F6). The study was complementing with glycocentric results. It was showing a lower number of fucosylated glycopeptides and higher number of Sialylated glycopeptides after emicizumab treatment which was nearing to the range of control sample. Subsequently, we broadened the data

analysis methodology with the assistance of PANDA software and quantitated the intensity of each glycopeptide within the sample, which initially identified with the assistance of pGlyco software. We merged both the results and compiled them to calculate the micro-heterogeneity at the N-glycosite of IgG's based on their variability in the peptide sequences. The tandem mass spectrometry profile of human and mice IgG subtyping is illustrated with an example in figure 2.20 and figure 2.23, respectively. The quantitation of glycoforms at each site demonstrated a noteworthy increment of galactosylation, which specifically happening at IgG1 and IgG4 antibodies. Thus, the examination was supplementing the glycocentric investigation as well as giving an understanding of glycan variety at IgG subtyping level. The IgG specific glycosylation differences were also compared with the cytokine analysis of inhibitor developed sample, and indicated the over expression of pro-inflammatory cytokine IP-10 in the pre-treatment stage. The difference is comparable to glycoform changes and reverse back after emicizumab treatment. Hence, we increased the number of samples for the analysis to confirm our observations.

We analyzed additionally 6 pairs of HA plasma samples which contains with and without inhibitors at their pre and post treatment stages of emicizumab. We analyzed 11 samples of Inhibitor developed HA samples and 14 samples of inhibitor negative HA samples. Out of that, 5 samples were treated with emicizumab in each category. The compiled results were showing that the patients who had developed inhibitors were susceptible to have a plasma N-glycome differences, especially a decrease in galactosylation at their antibody glycosylation but not necessarily reflected in all the cases. However, the patients with glycosylation pattern changes, is dynamic and positively responded to emicizumab treatment and became normal after treating with emicizumab. Since emicizumab itself is an antibody with glycosylation, we analyzed the drug glycosylation pattern to exclude the confusion about the emicizumab interference in the change at antibody level. The emicizumab showed a higher galactosylation index value (very

less amount of galactosylation) itself, which in turn prove that there is no interference of emicizumab in the plasma/antibody galactosylation index value. Hence, it is proving that the emicizumab administration is improving the immune aging profile of antibody isolated from the plasma of individuals with inhibitors with higher galactosylation index. However, the patients without a higher galactosylation index did not show any IgG glycan change relating to emicizumab treatment.

We additionally performed the studies with mice model. Here, we analyzed three different strains of HA mice at its 11th week, followed by rFVIII injection to them to induce the inhibitor and further isolated the plasma at 15th week to identify the glycosylation changes specific to inhibitor development. By this study, we additionally analyzed the interstain glycome variations at baseline and inhibitor development stage. Also, we compared the plasma N-glycome of mice with human by analyzing them using the same methodologies. The N-glycome results were showing the differences at individual level. The structural based glycoform pattern analysis showed a specific strain related variation are also seen in the plasma glycomics and glycoproteomics level. The changes related to inhibitor development were not significant. However, the inhibitor development related changes were more evident in BALB/c strain when compared to SV129 and BL-6 CD.45.1. in mice model, we could see an improvement in IgG and IgM production during inhibitor development. The distinctions were measurably critical. The site-specific glycan micro-heterogeneity indicated a decrease in sialylation (particularly mono sialylation) at the inhibitor advancement stage (figure2.22). The pattern was seen in all IgG subtypes identified (IgG1,2B and 3).

Further we extended the study of glycocentric and glycopeptide data of human and mice model with statistical analysis. Both the results were complementing with each other. However, the human data were not statistically significant due to some cases who did not vary their glycome profile with inhibitor development. This also indicating how the glycomics and

immunogenicity of an individual varies. The observations related to this study was also giving an insight about the importance of glycome analysis of patients in personalized medicine.

3 CHAPTER 3 SITE-SPECIFIC N-GLYCOSYLATION ON THE AAV8 CAPSID PROTEIN

3.1 Abstract

Adeno associated virus (AAV) is a versatile gene delivery tool, which has been approved as a human gene therapy vector for combating genetic diseases. AAV capsid proteins are the major components that determine the tissue specificity, immunogenicity and in vivo transduction performance of the vector. In this study, the AAV8 capsid glycosylation profile was systemically analyzed by peptide mass fingerprinting utilizing high-resolution mass spectrometry to determine the presence of capsid glycosylation. We identified N-glycosylation on the amino acid N499 of the capsid protein. We characterized the overall glycan profile for vector produced in 293 cells. Multiple N-glycosylated host-cell proteins (HCPs) copurified with AAV8 vectors and were identified by analyzing LC-MS data utilizing a human database and proteome discoverer search engine. The N-glycosylation analysis by MALDI-TOF MS, highlighted the probability of AAV8 interaction with terminal galactosylated N-glycans within the HCPs.

3.2 Introduction

Adeno associated virus (AAV) is a depend parvovirus that developed as a gene therapy vector for treating a variety of genetic disorders and acquired diseases. Lack of pathogenicity, low immunogenicity and differential tropism to multiple cell type make AAV a versatile gene delivery system[194]. Lack of pathogenicity, low immunogenicity and differential tropism to multiple cell type make AAV a versatile gene delivery system. This non-enveloped virus is approximately 25 nm in diameter and has a unique linear single-stranded DNA genome[195–197]. It has a 4.8kilobase genome flanked with two copies of 145 bp inverted terminal repeats. The two AAV open reading frames comprised of rep and cap gene.

Rep gene encodes four different replicating proteins (Rep78, Rep 68, Rep 52 and Rep 40) and cap gene encodes the three VPs (VP1, VP2 and VP3) which are translated from different start codons at the same ORF[198]. AAP is also identified, which facilitate AAV packaging[199,200]. The three VPs are expressed from the same ORF region by alternating splicing of mRNA, hence contains a common C- terminal domain. In spite of the common region at C-terminal, VP1 contains extra N-terminal sequence contrasted with VP2 and the VP2 contains an extra amino acid sequence at its N-terminal, contrasted with VP3 [201] . Capsid proteins of AAV assemble to form a T1 icosahedral virion with sixty units of VP1, VP2 and VP3 in a ratio of 1:1:10 [202,203]. The structural studies of the intact capsid with cryo-electron microscopy, X-ray crystallography and image reconstructions revealed that the N-terminal regions of VP1 and VP2 got enfold inside the capsid structure[204,205] ; consequently, blocked off from their binding activities [206].Hence, the common C-terminal VP3 region(~530aa) determines the receptor binding of the virus. To date, around 13 distinct serotypes of AAV has been used widely for gene therapy[207]. AAV serotype display 55-99% sequence homology,[208,209] however, are different in their tissue tropism[210–212]. Amino acids in a particular serotype determines its tissue specificity[213]. AAV serotype 8 (AAV8) was isolated from rhesus monkey widely known for its high performance in liver transduction[205]. It is a vector of choice for the treatment of genetic disorders utilizing hepatocytes as the target organ.[195,214–219]. AAV8 is generally utilized in quality treatment research for the hemophilia A, hemophilia B, familial hypercholesterolemia and glycogen storage disease type II[58,220–225].AAV8 accounted for crossing blood-cerebrum obstruction subsequently transferred the genes in cardiac and skeletal muscle in hamster and mice[226].

Typically, non-enveloped viruses are less commonly glycosylated. However, the capsid protein of hepatitis E, fiber protein of adenovirus-2 are glycosylated[213,227]. Such glycosylation has a drastic impact on viral properties especially regarding tropism. Moreover,

glycosylation may affect the immunogenicity of viruses if it is part of the capsid components. Site specific modifications of capsid proteins like tyrosine phosphorylation is reported to promote a ubiquitination and degradation of AAV2 capsid protein leading to decreased tropism[228,229]. Similarly, AAV tropism may have been varied by the presence of glycosylation on the capsid protein. To date, AAV has widely been deemed a non-glycosylated DNA virus. Here we presented direct evidence of N-glycosylation on the $NNS_{499\text{-}501}$ AAV8.

3.3 Experimental design

AAV has a complicate capsid structure consisting of multiple units of VP1, VP2 and VP3. **Figure 3.1** showed the flowchart of our methodology for characterizing AAV capsid glycosylation. This comprehensive approach incorporates in purification of rAAV from media (as in a secreted form) and intracellular virus. The purified virus were separated on SDS-PAGE and differential glycoprotein staining technique used to identify the glycoproteins in the mixture. The purity of the mixture further checked by Coomassie staining. The differential fractions were taken for the N-glycan analysis using MALDI-MS. The viral proteins were used for site specific glycosylation analysis by HCD and CID after PNGase F mediated ^{18}O labeling using CID analysis using HR-MS. Both HCP and AAV capsid glycosylation are investigated by analyzing with AAV capsid proteome database incorporated to the UniProt human database.

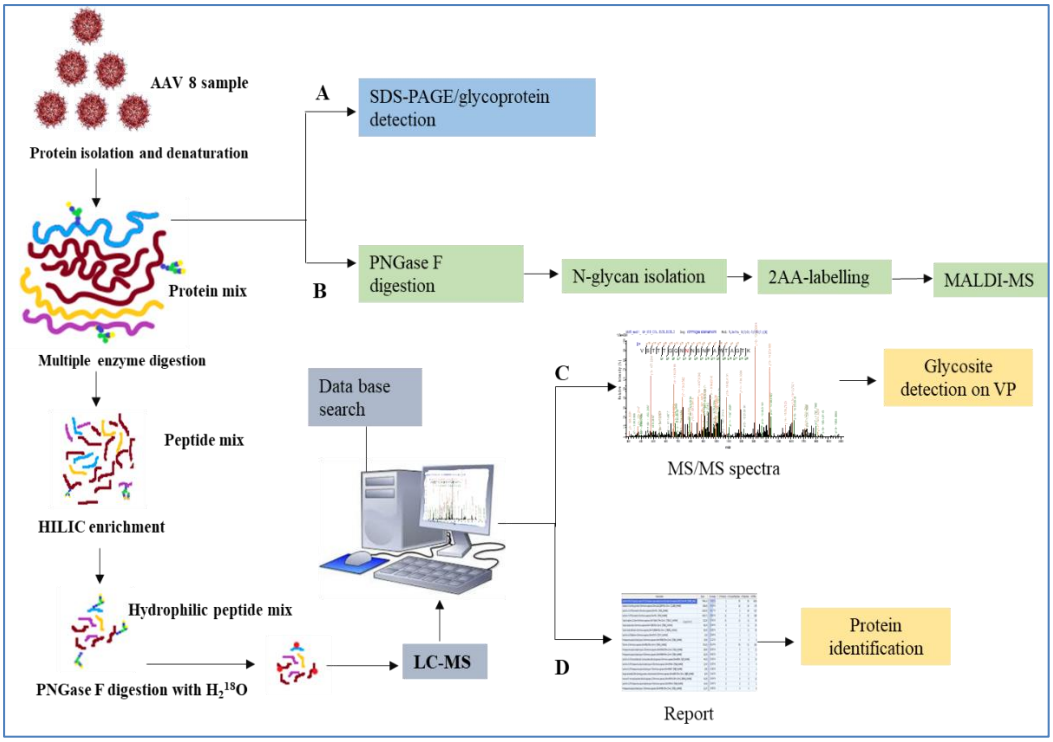


Figure 3.1 The general workflow of deciphering the glycosylation identification in AAV8 sample. Intracellular or secreted AAV8 were processed parallelly and analyzed using multiple techniques. A) The glycoproteins are visualized on the SDS-PAGE gel by glycoprotein labeling. B) N-glycan analysis using MALDI-MS identifies the type of glycosylation. The peptides generated by Glu C or trypsin digestion were enriched using HILIC cartridges. The deglycosylation of enriched glycopeptide in the presence of $H_2{}^{18}O$ to label the glycosite with ^{18}O.C) The samples are injected to LC-MS and using the database of AAV8 VPs, the N-glycosite was identified D) The same data is searched against the human database to find out the HCPs interacted to AAV8 from host cell system.

3.4 Materials and Methods

The AAV8 vector was expressed using HEK 293 cell line and purified by density gradient centrifugation. Both secreted and intracellular AAV8 were purified from same AAV batch using differential purification from Sol Sherry Thrombosis Research Center (Temple university, Philadelphia, PA). The alkylating agent, iodoacetamide (IAM) and reducing agent DTT (dithio thritol) were purchased from Acros Organics (Morris Plains, NJ, USA). 30% acrylamide Bis -acrylamide solution obtained from Bio-Rad. Glycoprotein detection kit, BCA protein assay kit, and protein markers purchased from Thermo Fischer. Sequencing grade endoproteases trypsin and Glu C were purchased from Promega (Madison, WI, USA). Peptide-N-glycosidase F (PNGase F) bought from New England Biolabs (Ipswich, MA, USA). Absolute alcohol purchased from Decon Labs, Mass spec grade ACN was purchased from J.T Baker chemicals (Avantor Performance Materials, Inc., Center Valley, PA, USA).

Click Mal (5µ, 100 Å), the HILIC material is purchased from ACCHROM (Beijing, China). Microcon-10 kDa(YM-10,0.5ml) centrifugal filters obtained from Millipore. The 3M Empore C8 disk purchased from Bioanalytical Technologies (St. Paul, MN, USA). Urea, SDS, sodium cyanoborohydride, anthranilic acid (2AA), Tris base, tris HCl dihydrogen monoxide) and $H_2{}^{18}O$ (97% ^{18}O) were purchased from sigma-Aldrich (St. Louis, MO). TEMED, ammonium persulphate, ethyl acetate, urea, sodium dodecyl sulfate (SDS) ammonium bicarbonate (ABC)and other reagents were also purchased from Sigma-Aldrich (St. Louis, MO).

The samples were concentrated using speed vac to minimize the water content, and the proteins were precipitated using ice-cold ethanol. The protein pellet, dissolved in urea buffer (6M urea in 0.1 M Tris/HCl, pH 8.5) the concentration of the protein mixture was calculated using BCA protein detection kit according to the kit protocol using urea buffer as a blank. The protein mixture was denaturated by adding 1M DTT (100:1 v/v) in the reaction mixture and heated at 95 °C for five minutes and cooled and added around 1M IAM in the ratio 50:1, v/v at 37 °C for 1h. The reduced and alkylated samples desalted, and buffer exchanged using 50mM ammonium bicarbonate pH 8.0 buffer and aliquoted to three parts.

SDS–PAGE and visualization of glycoprotein band and in-gel digestion. The capsid proteins of AAV8 particles were separated (around 10µg of protein) on 4- 12 % Bis-tris by SDS-PAGE. The glycoprotein detection kit purchased from thermo fischer scientific (catalog number:24562) was utilized to stain the glycoprotein band concretely on the SDS-PAGE gel. The glycoprotein staining was performed according to the manufacture's protocol. In summary, When the protein gel treated with periodic acid (oxidizing reagent), glycols present in the sugar moieties of glycoproteins are oxidized to aldehydes. During this reaction, a magenta pink band is developed wherever the glycans are present. Rest of the protein remains invisible. The gel was further stained by Coomassie blue to visualize the protein bands. The

gels were stained exclusively with Coomassie blue, for the in-gel digestion and further mass spectrometric analysis. The corresponding bands identified from glycoprotein staining, was excised for in-gel digestion referred to pre-established protocol. The 10µg of PNGase F treated samples were loaded adjacently with the same amount of control samples to compare the protein profile and band shift after de-N- glycosylation[230].

3.4.1 N-glycan analysis

The intracellular and secreted AAV8 (~100µg each) were treated with ~50U of PNGase F. The sample was incubated at 37° C for 16. The released released N-glycans were separated by ethanol precipitation. The precipitated proteins were separated by centrifugation and the supernatant contains the released glycan. The N-glycans were dried using vacufuge and tagged with anthranilic acid according to the previously published protocol (30, 31). An excess of labeling agent got removed by ethyl acetate wash. The 2AA-labeled N-glycans were reconstituted in MS grade water and drop dialyzed on nitrocellulose membrane filter 0.05µm VMWP in water for 30 mins. The dialyzed samples were concentrated to 1 µl in a Speed-Vac Vac and commixed with 1µl of a saturated solution of DHB matrix (70% ACN in dihydrogen monoxide). A 1µl of the sample spotted on the MALDI- plate and allowed to dry and form crystals. The plate is installed in the instrument, and the sample spot was bombarded with 32.5% high energy laser power utilizing a Bruker Daktronics MALDI-TOF-MS (UltrafleXtreme, Bruker Daltonics; Bremen, Germany) system to acquire all the MS spectra. The data acquisition was in negative ionization using reflectron mode. The spectra were generated with the uniform signal intensity. The glycoworkbench (http://code.google.com/p/glycoworkbench/)was availed to annotate the m/z values in the spectra and give the structural identification cognate to the human system. The MS/ MS spectra of the major glycoforms, confirmed the structures. The tandem mass spectrometry was performed by ‘LIFT’ mode (negative ionization) bombarded with high energy laser. The

pattern of fragmentation was attesting the structure and composition of each oligosaccharide moiety. The relative intensity of the glycan masses calculated by Flex Analysis software (Bruker Daltonics) to produce the final spectra.

3.4.2 Peptide mapping analysis

The third aliquot of the sample around 100µg was treated with sequencing grade trypsin in an enzyme protein ratio of 1:50 (w/w) and incubated at 37ºC for 16h. The reaction was ceased by keeping the tube in a boiling water bath for five mins. Approximately 50µg of the peptide was saved for further analysis and the other aliquot was digested with Glu C in an enzyme protein ratio of 1:50 (w/w) and incubated at 37ºC for overnight. These samples were further subjected to glycopeptide enrichment, HCD analysis, N-glycosite occupancy, and intact glycopeptide analysis.

3.4.3 Glycopeptide enrichment

The Click Mal, a HILIC media was used to enrich the glycopeptide from the virus sample according to previously published method with minute modifications[231].An in-house HILIC-SPE column was prepared by inserting a small C8 disc into a 200 µL tip. Around 6 mg of HILIC media was weighed out and washed with pure ACN and then transferred to the microtip to make the HILIC-SPE column. The column was then washed by passing with 100 µL of 10% ACN containing 0.1% FA and then equilibrated with binding buffer (80% ACN containing 1.0% FA) for 3-5 times. The proteolytically (trypsin and trypsin followed by Glu C) digested peptides were dried and resuspended in 10µL of binding buffer and introduced to the microcolumn and allowed to bind to the column for 10 mins at room temperature. The unbounded peptides were washed off by passing 100µl of binding buffer and repeated the step for five times. The column bound glycopeptide were then eluted out utilizing 200µL of the elution buffer (EB; dihydrogen monoxide contacting 1.0% FA). The HILIC-enriched

glycopeptide were then dried using speed vac and resuspended in 5 µL of 2% ACN containing 0.1% FA in H_2O and injected to the nano-LC orbitrap MS-system.

3.4.4 N-glycosite detection (^{18} O labeling)

A fraction of (approximately 40µg) proteolytic enzymes treated intact peptides were subjected to glycosylation site identification. The digest was thoroughly dried in a speed Vac and resuspended in 10µl of ABC buffer (ammonium bicarbonate prepared in $H_2{}^{18}O$, pH 8.0). Asn linked N-glycans was removed by 1µl of PNGase F (50U), and the mixture was incubated at 37 °C for 16 h.

3.4.5 Peptide analysis by LC−MS/MS

The intact, enriched and ^{18}O labeled peptides were dried and resolved in 2% ACN containing 0.1% FA. 5 µL of the intact glycopeptides or 3 µL of the ^{18}O-labeled N-deglycosylated peptides were injected to Dionex Ultimate 3000 RSLC nano System (Thermo Fisher, Waltham, MA, USA). The LC system has Nano Trap column packed with Acclaim Pep Map100 C18 (2 cm × 75 µm I.D, 3 µm). The flow rate was adjusted to 5 µL/min with mobile phase A (2% ACN, 0.1% FA) for 10 min for sample trapping, then washing and followed by the separation on C18 Column (15 cm × 75 µm I.D., 3 µm, 100 Å). The separation was accomplished by a 120 mins linear gradient (3 to 40% Mobile Phase B (80% ACN, 0.1% FA) at the flow rate of 300 nL/min. The column was washed for 10 min with 99% B and reconditioned with 1.0% B for 5 mins for the next run[232].

Elite mass spectrometer with the spray source(1.6kV) of (Thermo Fisher) LTQ-Orbitrap is integrated with the LC system. The LTQ Orbitrap mass spectrometer was adjusted to data-dependent mode with an alternating MS1 and MS 2 acquisition. The Orbitrap mass analyser MS scan was performed with the mass range, m/z 400−1600; resolution at m/z 400, 6×104; automatic gain control target (AGC), 106 ions; maximum ion accumulation time, 50 ms. The MS1 ions of the ten most intense species, were subjected to MS/MS CID in the ion

trap analyser. The MS/MS scan model was performed in centroid scan model. CID-MS parameters were set by giving default charge state as 3, activation Q was 0.25 with an activation time 5.0 ms and isolation width was set to m/z 3.0. The normalized collision energy was set up to 35%.

For the HCD mode, orbitrap analyser was set at a resolution of 15,000 at m/z 400; AGC was 10,000 ions, and maximum ion accumulation time was increased to 200 ms. All the CID parameters were set remained same except the parameters of activation time which set upto 0.5 ms and the isolation width of m/z 2.0. The /MS parameter was performed at 27% NCE. MS/MS Data Interpretation For LC−CID−MS data analysis of deglycosylated peptides, pFind software 2.8 (http://pfind.ict.ac.cn)was used [233],[65]. FASTA sequence of AAV8 sequence were created using the already published sequence of VP1, VP2 and VP3(13). Tolerance of peptide mass was set to 20 ppm, and the fragment ion tolerance was 0.5 Da. Since the proteins were reduced and alkylated the fixed modification was carboxyamidomethylation of cysteine (+57.021 Da), and the variable modifications were deamidation for N and Q (+0.984 Da) and oxidation of methionine (+15.995 Da). Two maximum missed cleavage sites were selected for trypsin (KR-C) and Glu C (DE). For the glycosite identification: CID−MS/MS data analysis of ^{18}O- labeled de-N-glycosylated peptides were performed by setting the precursor ion mass between 350 and 6000 Da. Along with above described search parameters, we have included N-deamidated with ^{18}O (+2.988 Da) and specified as variable modifications with FDR < 0.1. The theoretical glycopeptide masses were obtained from the online server protein ExPAsy (http://web.expasy.org/peptide_mass) and GlycoMod tool (http://web.expasy. org/glycomod) possibility of glycosylations and sites of glycosylation by consensus sequence predicted by NetNglyc (http://www.cbs.dtu.dk/services/NetNGlyc) were further related manually to the data which we obtained from N-glycan independent analysis[234].

3.4.6 HCP identification by LC-MS/MS analysis

In the MS analysis of the peptide mapping data were sanctioned to detect the HCP glycoprotein present in the AAV 8 sample. The multiple protein identification in the sample was performed by Proteome Discoverer^{TM} (Thermo Fischer Scientific). Human database was downloaded from UniProt (http://www. uniprot.org) which can also detect even the process induced, common contaminant proteins like keratin. The database was edited by adding AAV8 VP3 which is the most abundant protein in the sample. AAV 8 VP sequence was taken from a previously published article and fasta files were generated and uploaded to the search engine [231,235,236]. The database was used to identify the other proteins apart from AAV8 in the sample especially host cell proteins (HCP) which is interacted to AAV 8. The precursor mass tolerance was set to 20 ppm and fragment mass tolerance was set to 0.8 Da. The data was generated from trypsin digested sample hence the enzyme entered used for the given search was trypsin. All the possible dynamic modifications like glutamine pyroglutamine conversion on any N- terminus(-17.027Da), methionine oxidation (+15.995 Da), acetylation (K, +42.011 Da) and deamidation (N, Q/+0.984) were considered in the specification. Since we have performed ^{18}O labeling, we included specific deamidation (N/+2.988 Da) also. Static modification is set to carbamidomethyl (C/+57.021 Da). Strict target false discovery rate was set to 0.01-0.05.

3.5 Results:

3.5.1 Detection of glycosylation by SDS-PAGE

The AAV8 samples were denatured and separated on 12% SDS-PAGE. The capsid proteins, VP1,VP2, and VP3, are segregated in the region of 50-100kDa determined by the protein marker[237,238] . We have identified magenta pink colored feeble bands in between the VP region indicated the VP glycosylation (figure 2 a). A dark magenta pink band was observed around 100kDa in the AAV8 secreted sample suggested the presence of host cell

glycoprotein in the sample. The VP glycosylation staining showed similar results in the region of VP2 and VP3. There was some additional glycoprotein band observed around 100kDa in secreted AAV8. The gel was further stained by Coomassie brilliant blue to visualize the total protein profile on the SDS-PAGE **(figure3.2 b).** Side by side analysis of AAV8 secreted sample and AAV8(secreted) subjected to PNGase F digestion, confirmed the type of VP glycosylation on the capsid proteins **(figure 3.2).** We could likewise affirm some extra bands vanished in PNGase F digested sample compares to the host cell proteins promote validation of N- glycosylation in the given sample (figure 2C). The difference in glycoprotein bands from intracellular and secreted AAV8 can supplementally designate the different intermolecular interactions of VPs with various glycoprotein which may copurify with AAV particle in the purification procedures. We have seen multiple protein bands which are unrelated to the capsid proteins in protein profiling. Some of them are found to be heavily glycosylated and stained positively by glycoprotein staining procedure. The host cell proteins can copurify along with the VPs if they have a similar molecular weight which may lead to ambiguity in the analysis. Hence this technique has considered only for preliminary examination, and peptide mapping designates protein identity of the HCPs by high-resolution

LC-MS with the aid of database search.

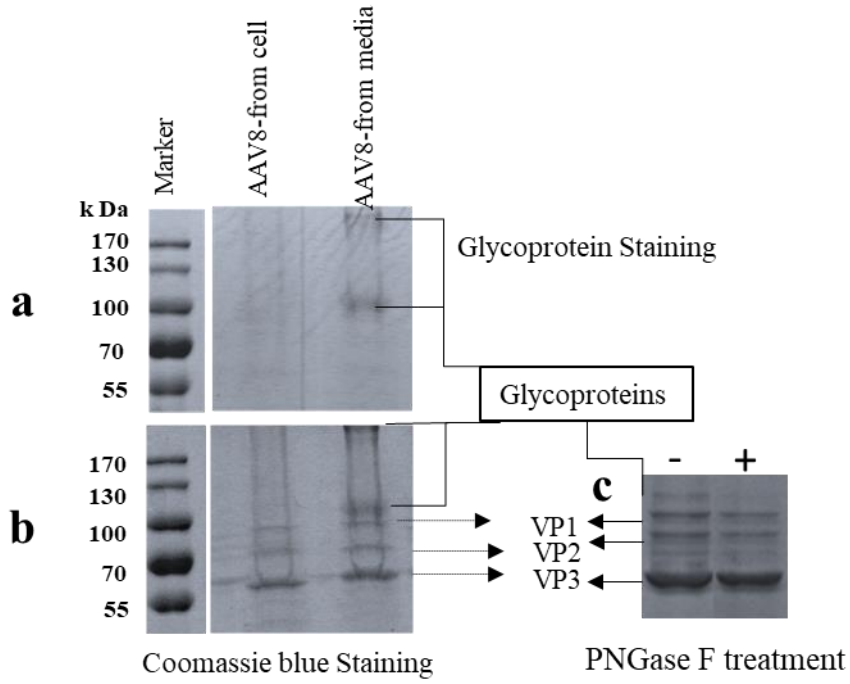


Figure 3.2 AAV8 Glycosylation detection on SDS-PAGE by differential staining method. AAV8 derived from the cell(intracellular), and media (secreted) are separated on SDS-PAGE a) showed the protein bands visualized after glycoprotein staining. Glycosylated bands were observed near the VP region (55-100 kDa). The non-related VP bands (~200kDa, and 110kDa) can be an indicator of glycosylated HCP in the sample. Gel 'b' is identically tantamount gel stained with Coomassie brilliant blue to visualize the whole protein profile. c) is an independent gel running after de-N- glycosylation along with its reaction control. (+ and – betoken the addition of PNGase F). The loaded protein amount kept constant. The gel profile is after Coomassie staining. The glycosylated bands between the VPs have vanished after PNGase F treatment. Some weak bands in the VP regions in-between the main VP bands were mildly stained by the glycoprotein staining which may be cognate to VP glycosylation.

3.5.2 The N-glycan profile of intracellular derived AAV8 and media derived AAV8

The intracellular and secreted AAV8 samples, parallelly treated with PNGase F for N-glycome analysis. The N-glycans of corresponding samples were isolated and tagged with 2 Aminobenzoic acid (2-AA) and spotted on MALDI-plate to provide unique spectra of the N-glycan profile for corresponding samples. The resulted spectra were smoothed, and baseline subtracted. The MS/MS of high-intensity peaks were performed in 'Lift mode' to confirm the monosaccharide composition. The peak annotation is done using the Glycoworkbench software after verified by CFG database **(figure 3.3)**. Since the previous protein analysis suggested the presence of host cell glycoproteins in the sample, the N-glycosylation profile of the secreted and intracellular AAV8 cannot be directly related to the

AAV8 capsid protein (VP). However, the high intensity of secreted AAV8 glycan spectra is correlating with the glycoprotein detection data and suggest the possibility of AAV8 interaction with host cell glycans. We could identify that the majority of the N-glycans in the sample were terminally galactosylated, indicating the capsid protein affinity to the terminally galactosylated HCPs. The glycoforms are identical in structure in both samples (Appendix C Table 1). The difference in proportion and intensity of each glycoform is noticeable (Figure3.1).

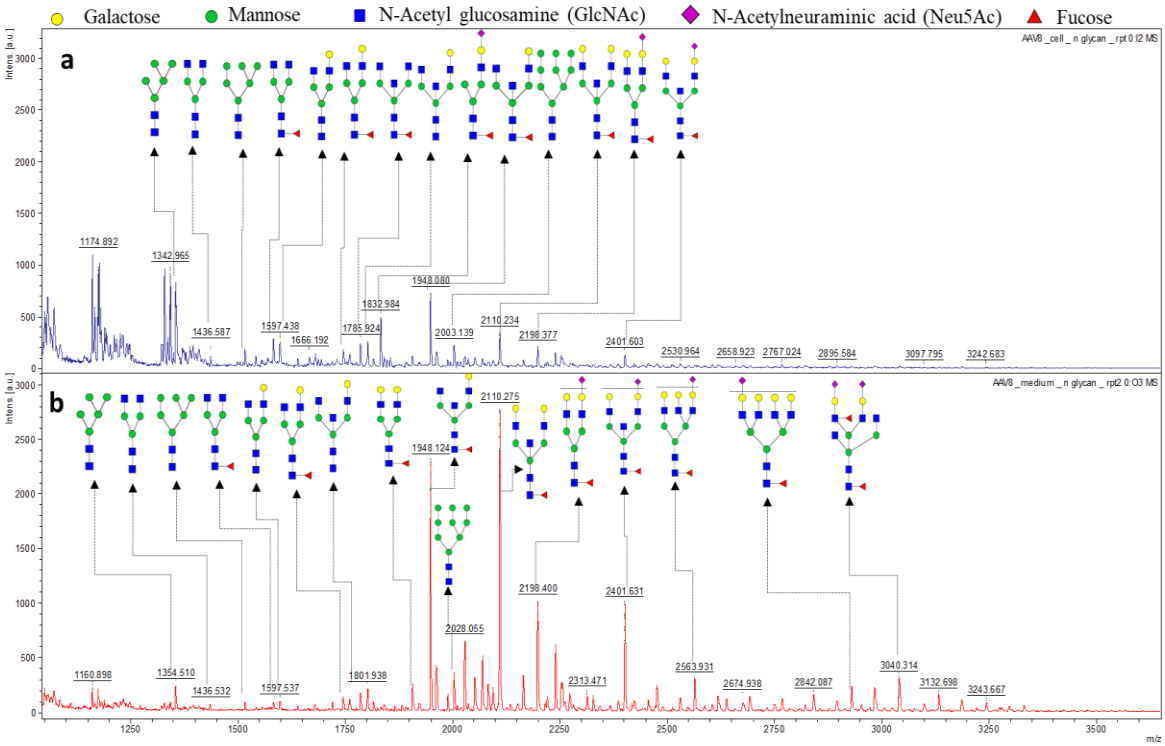


Figure 3.3 MALDI- MS spectra of 2-AA derived N-glycan in negative mode [M−H] ⁻.
a) the whole glycome profile of intracellular AAV8 b) The whole glycome analysis of secreted AAV8. The most intense peaks from both the spectra (m/z2110.234,1948.088 ,2401.633,1832.984 were further confirmed by MS/MS fragmentation spectra in MALDI-TOF MS in lift mode (supplemental figure 2). The uniform spectra were generated by keeping constant analytical parameters and sample amount.

3.5.3 The glycosylation analysis in peptide level

The Peptide mass fingerprinting data by orbitrap was performed by multiple ways, including in-gel and in-solution digestion. In-gel digested data showed very less coverage (less than 10% sequence coverage). Hence, we tried to perform in-solution digestion which increased the coverage to 77-85%. The sequence search by the 'GlycoMod'

tool suggested that AAV8 has six putative sites for N- glycosylation; five are within the part of common region of VPs ($NGT_{263-265}$, $NLT_{338-340}$,$NNS_{499-501}$ and $NQS_{665-667}$ and an exclusive site near to the N- terminal (NLS_{14-16}) region of VP1 (figure 4). Hence, we focused the search of glycopeptide variants in HCD and CID spectra to find the glycosylated sites.

To characterize the N-glycosylation site of capsid proteins, we focused on HCD mass spectra of tryp/Glu C digested proteins of intracellular and secreted AAV8. The peptides were indicating N-glycosylation on multiple peptides at different retention time with the marker fragment ions in MS2 spectra (oxonium ions at m/z 366 (HexHexNAc1,1 +), m/z 292 (Neu5Ac, 1+), m/z 204 (HexNAc, 1+), m/z 162 (Hex, 1+), and sub fragment ions at m/z 186 (HexNAc- H_2O, 1+), m/z 168 (HexNAc-$2H_2O$, 1+), m/z 138 (HexNAc-2 $H_2O-CH3OH$, 1+), m/z 126 (HexNAc-$2H_2O-CH3COH$, 1+), and m/z 274 (Neu5Ac− H_2O, 1+)).The analysis can also refer to O-glycosylation, but the spectra was not confirming any O-glycosylation. To enhance the intensity of detected glycopeptides, which is covalently linked to multiple glycoforms, the HILIC-enriched glycoptides were deglycosylated by PNGase F in the presence of ^{18} O water. Thus, the weak signals generated from scattering the peptides over the reverse phase column because of differential compositions of glycoforms covalently linked to the same peptide is eliminated. Furthermore, the signals from peptides derived from host cell proteins (HCPs) is hindering the glycopeptide signals. Therefore, we switched to CID analysis of the deglycosylated peptides, which have equal chance of vapour ionization in the mass analyser as non-glycosylated peptides. The shift of 2.988 Da (O^{18}-deamidated Asn) in the MS spectrum clearly confirms to the glycosylation site.

We could withal confirm the glycosylation in capsid protein on N_{499}, identified and attested by MS/MS analysis. Since the amount of glycosylated peptide was below the LOD, in normal enzymatic digest (trypsin and Glu C as well as trypsin alone) analysis, trypsin and Glu C trials were failed to pick up the glycopeptides; but improved the

coverage 77.0% to 81.38%. The glycopeptide was identified from media isolated AAV8 sample, only after HILIC enrichment. The entire experiment was repeated with an independent batch of AAV8 and the result was reproducible (**Appendix C F1**). These sample showed the presence of glycosylated as well as non-glycosylated peptide. As enrichment process may cause the loss of non-glycosylated peptide, we could not quantitate the percentage of glycosylation. Repeatability of our results fortifies the current discovery.

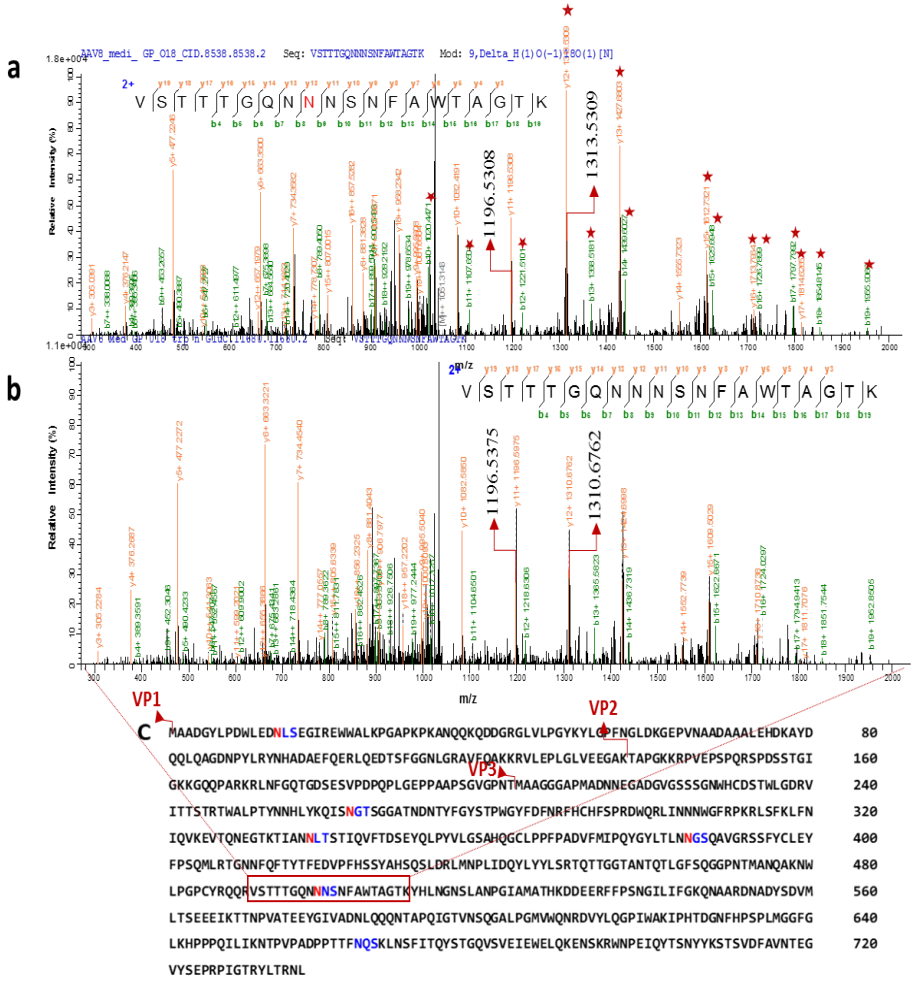


Figure 3.4: Glycosite identification on AAV 8 capsid protein.
Two distinct tandem mass spectra of peptide sequence ‘VSTTTGQNNNSNFAWTAGTK’ of the AAV8 capsid protein (in the common region of VPs).

Characterized CID spectra of glycopeptide after ^{18}O mediated digestion. A) the position of de-N-glycosylated asparagine 499(^{18}O-incorporated aspartic acid) was attested by mass increment of 2.8547 Da (theoretical mass difference 2.98 Da) to the series of b and y ion (the precursor ion mass: m/z 2101.93193). B) The position of Non glycosylated asparagine 499 in the sequence was confirmed by y12/b8 ions (the precursor ion mass: m/z 2099.0766 Da). The highlighted ions (marked with a star) shows the mass differences in both the spectra. y11+ (highlighted) ions are same in both the spectra. C) Capsid protein amino acid sequence and possible N-glycosylation sites predicted by NetNglyc software based on the consensus sequence (NXT/S, ‘X’ can be any amino acid except proline). Different N-terminal sequence of VP1, VP2 and VP3 are marked in the sequence. The glycosite peptide is marked on the sequence.

Consensus sequence search of N- glycosylation NXT/S through GlycoMod revealed 6 possible glycosite; N_{14}LS, N_{263}GT, N_{338}LT, N_{385}GS, N_{499}NS, and N_{665}QS (figure 4C). Aside from N_{14} every single other site is in the common capsid proteins. N_{14} , reported to be in the N-terminal domain and looped inside the capsid assembly[206]. Hence, the analysis for glycopeptide were mainly focused on VP3 because of the presence of all feasible glycosylation sites and abundance of peptide in the total sample.

We also identified various host cell glycoproteins in the sample that were copurified with AAV8. The virus vector isolated from the media and cells were analyzed utilizing proteome discover[1.4] software. The search result was listed in **Table3.1.** We listed only the proteins showed more than 20% coverage. We could identify fifteen HCPs from intracellular AAV8 and thirteen HCPs from secreted AAV8 (**Figure 3.5).** Most of the proteins identified by the data search were heavily N-glycosylated. The presence of Lam R, G3BP (galectin 3 binding protein), fibronectin etc., are found to be the common proteins in co-purifying with AAV8 samples.

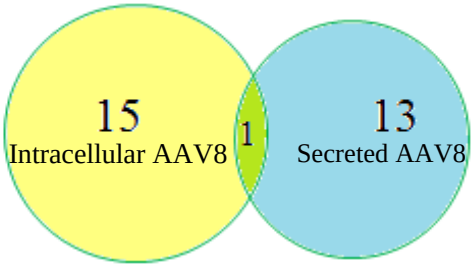


Figure 3.5 HCPs identification in intracellular and secreted AAV8 using Proteome Discoverer[1.4]
The protein indicated over 20% sequence coverage is recorded in Table 1 and compared between differentially purified samples. Galectin-3-restricting protein is the main protein, regularly found in both samples.

Table 3.1: HCP identification using Proteome Discoverer[1.4]

UniProt Accession No	**Protein identity**	**Coverage%**	
		Intracellular	**Secreted**
P62158	Calmodulin (Human)	38.93	ND
O60637-3	Tetraspanin-3 (Human)	34.39	ND
P06748-3	Nucleophosmin (Human)	34.36	ND
Q92542-2	Nicastrin (Human)	33.53	ND
Q9BY67-2	Cell adhesion molecule 1 (Human)	33.33	ND
P13473	Lysosome-associated membrane glycoprotein 2 (Human)	32.44	ND
Q5ZPR3-3	Isoform 3 of CD276 (Human)	29.61	ND
P11279	Lysosome-associated membrane glycoprotein 1 (Human)	29.5	ND
Q14108	Lysosome membrane protein 2 (Human)	28.24	ND
Q13740	CD166 (Human)	27.1	ND
Q6PCB8-2	Embigin (Human)	25.27	ND
Q08380	Galectin-3-binding protein (Human)	23.93	36.75
P60900	Proteasome subunit alpha type-6 (Human)	22.76	ND
P05556-2	Isoform 2 of Integrin beta-1 (Human)	22.18	ND
Q13162	Peroxiredoxin-4 (Human)	21.03	ND
P12268	Inosine-5'-monophosphate dehydrogenase 2 (Human)	ND	19.65
P68363-2	Tubulin alpha-1B chain	ND	41.49
P01834	Ig kappa chain C region (Human)	ND	35.85
P02751-5	Isoform 5 of Fibronectin (Human)	ND	33.7
P01620	Ig kappa chain V-III region (Human)	ND	31.19
P23142-4	Isoform C of Fibulin-1	ND	30.75
P23142	Fibulin-1	ND	30.3
Q16222-2	Isoform AGX1 of UDP-N-acetyl hexosamine pyro phosphorylase	ND	29.31
P07437	Tubulin beta chain	ND	26.35
P0CG48	Polyubiquitin-C	ND	23.65
Q04837	Single-stranded DNA-binding protein, mitochondrial	ND	22.3
P35556	Fibrillin-2 (Human)	ND	20.81

**ND -Not detected*

3.6 Discussion

Host cell glycosylation plays significant roles in AAV viral entry, tissue selection, and infectivity. These roles are partly documented by identified AAV receptors, which are often glycans. Heparin sulphate, N-glycans terminated with galactose, and sialic acid are the well-known primary receptors of various AAV serotypes.[212] The essential receptor for AAV2 and 3 is Heparan Sulphate Proteoglycan (HSPG). AAV1 and 5 utilize N-linked sialic acid cell surface receptors for the viral entry while AAV9 utilizes N-linked galactose.[224] The essential receptor of AAV4 is known to be O-linked sialic acid in the host cell surface.[225] Likewise, AAV6 uses N-linked sialic acid and HSPG. The secondary receptor of AAV8 is a laminin receptor (Lam R), a host cell surface glycoprotein.[239]

However, AAV glycosylation status is still unclear. The initial study of AAV prototypes was based on AAV serotype 2, which was not perceived to have any glycosylation even though the primary sequences of capsid proteins suggest presumptive glycosylation sites on capsid protein.[237]Since there are many AAV serotypes that have been identified and characterized, some AAV serotypes have notable differences in terms of *in vivo* performances and secretion efficiency, which are characteristic feature that may have been influenced by glycosylation. Here we utilized the high-resolution mass spectrometry with well-designed sample preparation to explore potential glycosylation in AAV8 capsid protein.

We designed this study to enrich the glycosylated AAV from both intracellular and secreted AAV8. It is anticipated that AAV8 may be glycosylated in the secretion pathway by the glycosylation machinery. The SDS-PAGE specific staining (figure 2a) indicated to glycoproteins in the AAV8 capsid proteins. The capsid proteins VP1,VP2 and VP3 have molecular weights around 87 kDa,73 kDa and 60 kDa, respectively[213]. The darkly stained glycoprotein bands around 100 kDa in the profile was suggesting the presence of copurified host cell glycoprotein in the secreted AAV8 sample. There are feeble glycoprotein bands in

the capsid protein region(55-100kDa) indicated capsid protein glycosylation. The PNGase F treated sample did not demonstrate any glycoprotein in the protein profile, which affirm the presence of N-glycosylation in the capsid proteins.

The N-glycome analysis of intracellular and secreted vectors confirmed the presence of N-glycosylation in the overall viral sample. Even though the examination intended to separate and characterize the N-glycans derived from purified capsid proteins, the parallel analyses confirmed the presence of host cell glycoproteins in both samples and diverted the outcome significantly connecting to the HCP glycosylation. The N- glycan profile of differentially purified sample demonstrates the distinction in extents and limit of glycosylation. Although the method is not an absolute quantitative technique yet, the N-glycosylation pattern in terms of proportion and MS profile is unique for each differentially purified virus and showed the presence of high amount of N-glycoforms in secreted AAV virus. The result correlated to SDS-PAGE gel examination for glycoprotein detection. The glycoform analysis indicates that both the AAV samples contain high galactosylated species as the major sugar. The protein ID verified the presence of G3BP (Galectin 3-binding protein), a heavily galactosylated glycoprotein backup the affinity of AAV8 to the galactose sugar, which is reflected in the N-glycan analysis. The interaction of G3BP with AAV-6 has been reported earlier.[229] This galactose binding protein was eluted at the same region of darkly stained glycoproteins in the SDS-PAGE gel. Both AAV samples showed different glycosylation profiles that may arise due to the variation in the host cell glycoprotein present in each sample.

The peptide level investigation of virus samples utilizing high resolution mass spectrometry confirms the previously mentioned observations. The HCD information of in-gel processed 70 kDa band (which can be a glycosylated VP3) demonstrated signature

masses of carbohydrate moieties covalently linked to AAV peptides. However, HCD spectra alone was plausible, and the sequence coverage was very less for further investigation (data not given)[228]. To enhance the coverage, as well as the signal intensity of glycopeptide, we performed in-solution digestion followed by HILIC enrichment of glycosylated peptides. The HCD spectra of HILIC-enriched samples were suggesting many glycopeptides with signature fragments with enhanced signal intensities. Although HILIC enrichment increases the probability of fishing glycosylated peptides and thus enhance the glycopeptide detection, the intensity of a glycopeptide is far low compared to non-glycosylated peptide. Also, HCD experiments showed the presence of many unrelated glycopeptides. Thus, to overcome the analytical challenge of detecting the viral peptides and avoid the column-based segregation of modest glycopeptides related to the capsid proteins, the N-glycan release was performed in the presence of $H2^{18}O$, by which the glycosylated site is labeled and localized easily.

The glycosylation was found on the overlapping peptide of trypsin and double digested sample, VSTTTGQNNNSNFAWTAGTK at $NNS_{499\text{-}501}$ of AAV8 capsid proteins. The amide (CO-NH2) group of N-glycan linked N_{499} is deamidated to carboxylic group (CO-OH) thus to D_{499} with ^{18}O atom in $H2^{18}O$ media. This replacement causes an increment of +2.98 Da, which is easily identified by mass spectrometry. The mass differences of 2.8547 Da from y^{12+}-y^{17+} ions in the MS/MS spectra (figure 4) clearly distinguished the glycosylated peptide from the most abundant non-glycosylated peptide. The experiment was repeated with another batch of AAV8 and CID spectra showed a reproducible spectrum with the mass difference of 3.03 Da from y^{12+}-y^{17+} ions in the MS/MS spectra (figure S1). The fragmentation pattern of non-glycosylated peptide does not show any mass differences in y^{12+}-y^{17+} ions as shown in the deglycosylated peptide tandem MS spectrum. The CID fragmentation has confirmed the presence of negligible amount of glycosylated variant of AAV8. The enormous amount of non-glycosylated peptide substantiates the majority

population of AAV8 are non-glycosylated virus. The superlative method of glycopeptide enrichment and analytical strategy could identify the inadequate glycosylated variants of the capsid protein, indicating the minor population of glycosylated AAV8. Repeatability of our results fortifies the current discovery. Since glycosylation has the essential biological functions such as cell recognition, ingression, and immunogenicity, this discovery is significant for studying the effect of AAV8 N-glycosylation for manipulating their therapeutic application in various monogenic disorders. From here, further research focus is in need.

The enormous glycopeptides present in the peptide mapping data was unrelated to virus capsid protein. The component proteins apart from major VPs detected in current samples were also a censorious subject to be considered. Since the expression system was HEK-293 cells, we searched the generated CID spectra of AAV8 peptides against combined fasta files of the human protein database and AAV8 VP3. In summary, the LC-MS/MS peptide map fingerprinting recognizes the best matching protein belongs to the human database or AAV8 VP3. We have reported only the recommended protein which showed more than 20% coverage for better reliability. Since many non-glycosylated peptides were lost through HILIC enrichment, expecting 100% coverage of a glycoprotein is unrealistic. Majority of the reported proteins are glycosylated. The presence of G3BP emphasis the galactose binding properties of AAV8. The interaction of the galactose binding protein with AAV-6 has been reported earlier and found to be segregated between 75-100 kDa on SDS-PAGE[227]. Detection of Lam R, a secondary receptor of AAV8, also confirms the close associations of reported HCPs with AAV8 vector and may explain the copurification.

In summary, we studied different aspects of N-glycosylation in AAV8 recombinant vector and confirmed the presence of N-glycans on N_{499} on 'VSTTTGQNNNSNFAWTAGTK' peptide, localized on the common region of AAV8 capsid

protein. The glycosylated peptide and non-glycosylated variant coexisted in the AAV preparations suggest that the existence of multiple pathway for AAV8 maturation post AAV packaging. It would be important to study how glycosylation of AAV and its copurified proteins would affect the neutralizing AAV antibody and *in vivo* performance. The close interaction of AAV8 with different host cell glycoprotein also suggest providing addition clue on viral tissue tropism.

4 CONCLUSIONS

Life is comprised of four significant classifications of molecules, including nucleic acid, protein, lipids, and carbohydrates. Be that as it may, the significance of glycans is an understudied field. The detailed analysis of protein glycosylation can connect numerous fields of biological science to see how a protein changes its functions during various physiological conditions utilizing a dynamic glycan arm. The glycosylation can get altered by genetic, environmental, epigenetic factors of an individual. Hence, the altered glycoforms may result in malfunctioning due to protein misfolding due to micro and macro heterogeneity of their oligosaccharide part. Therefore, the host N-glycome can influence its immune system on disease progression. The changes can also revert during the treatment. Hence, the glycosylation can be potential candidate for altering the biological specificity between the molecule and alter between healthy and disease state of the body .

The glycosylation is an essential field in gene therapy-based vector designing. The glycan may increase tropism and create a specific channel of gene-delivery to the host cell. Glycan can be used as a unique key for delivering the gene into a distinctive tissue type based on their special receptors on their cell surface. Nevertheless, the glycosylation analysis is still challenging for a novice because of the diversities of analytical methods available in the field and the limitations of each method. However, glycan analysis methods evolve rapidly from simple monosaccharide analysis strategy, which provides only compositional information for intact glycoproteins and glycopeptides, and gynocentric analysis strategies that provide structural, compositional, isomeric, and linkage specific information. The latest developments in mass spectrometry and its complementary analytical techniques help to advance the glycoprotein interpretation efficiently. The advancement in the mass spectrometry analyzers, software, and databases helps glycobiologists analyze complex biological samples and understand the role of glycans in the biological system. The advancement in the analytical

techniques, including sample preparation and separation technologies, helps promote the application of glycobiology in prognostic and diagnostic aspects of diseases. In Chapter 1, we tried to give an overview of the standard techniques and method for characterizing the protein glycosylation, which commonly used for the project. All these techniques help us understand the role of glycosylation changes during a specific biological condition and build up a corresponding structural library of glycoforms. The recent developments in mass-spectrometry-based glycoprotein characterization are providing an enormous scope of glycobiology applications in clinical and research aspects.

In **Chapter 2,** we tried to build up the knowledge of the mass spectrometry-based glycoprotein analysis to elucidate the biomarker discovery relating to inhibitor development against rFVIII. Here we analyzed the plasma and antibody N-glycome change during inhibitor improvement, taking HA-mice and HA-human as an investigation model. The glycosylation can change from according to the biological organization and hence vary from human to mouse model. However, we assumed a trend that can be correlated related to inhibitor development. We addressed the complexity of glycan analysis by approaching the study using multilevel by different mass spectrometry-based methods primarily as a released N-Glycan analysis(glycocentric) and glycopeptide analysis(bottom-up).

The released N-glycan analysis was conducted in two levels by one-step and two steps permethylation followed by MALDI-MS based data acquisition. In humans, we could successfully enrich the antibody from plasma. Whereas in mouse, we could not enrich the antibody; hence, the glycan analysis was limited to plasma N-glycome. Eventhough, we considered three different strains of HA-mice and the induced for inhibitor development by treating with rFVIII. The plasma was taken before and after the rFVIII treatment and inhibitor development. At the pretreatment stage, the mice were at 11 th week, and inhibitor developed stage the mice were at 15th week. We did a background study to see the variation

between 11th and 15-week mice. Additionally, we analyzed HA and wild type mice to confirm there is no interference of age and HA in the N-glycome analysis of the current cohorts. For the human sample, we did a preliminary analysis to see the difference in the plasma N-glycome of individuals who vary in their age, HA condition, blood group, and ethnicity. Since HA is an X-linked recessive disorder, we considered only male candidates for the study. We noted that the galactosylation and sialylation change in plasma, mainly in the antibody, as the age increases. The changes were noted in two categories, in response to age and inhibitor development. In both cases, galactosylation and sialylation decreased as the value increases. Hence, we concluded that we must conduct the study within the same age group. However, considering the fact that unknown factors like diet, health, other diseases, and the physiological difference can also alter the glycosylationprofile[240],[241] we assumed inter-species that comparison could fortify our study.

In the next level, we reconfirmed the observation with an individual with inhibitor development and the same individual after treating with emicizumab, which is the functional analog of rFVIII. The study was conducted to see how dynamic the N-glycome profile and how it changes by just a treatment protocol. Here we could see a drastic change in both plasma and IgG glycome profile. We could see a significant improvement in sialylation (in both the linkages), galactosylation, and complex multi antennary glycoform formation, in response to emicizumab treatment. The post-treatment profile was positively correlated to the range of the analytical control sample. The control sample was the plasma taken from the HA individual from the same age group but without inhibitor development. By one-step analysis, we identified 80-90 glycoforms in different human and mice samples. There were no specific sugar structures identified with inhibitor development. But the discussed differences were calculated based on their relative intensity-based glycan profiling. By the two-step permethylation analysis, we could further divide the sialylated glycoforms based on their

linkages. Here we calculated the difference based on the conversion of 5 4 0 0 0 glycoforms (di- galactosylated biantennary non-fucosylated complex sugar, H5N4) to its sialylated glycoform 5 4 1 0 0 and 5 4 2 0 0 by calculating the sialylation index. The approach was decided to discriminate the linkage specific sialylated glycoforms from the one-step study, which in turn confirms the variabilities related to the individual profile. The glycocentric analysis can provide better structural elucidation, linkage analysis, and relative quantitation of glycan profiling. The permethylated glycoforms are analyzed using the glycoworkbench software, which in turn linked with glycan databases (CFG, Carbbank, Glycome DB, and Glycosciences). We extended our study to glycopeptide analysis, where we could identify the glycoproteins by a bottom-up approach using pGlyco software. We additionally, semi quantitated the microheterogeneity of the site-specific glycan related to IgG subtyping. We could see that IgG productions are higher in inhibitor developed plasma in comparison to the negative inhibitor cohort. We could also see a decrease in IgG galactosylation at IgG subtype level. According to our result, the significant decline in galactosylation is happening at IgG1 and 4, whereas the differences are negligible at IgG2. According to the glycocentric analysis, the plasma sialylation was also less at pre-treatment and increased after emicizumab treatment, which again in the range of control sample. Here we could see a more significant number of sialylated glycopeptides in the post-treatment stage and HA control sample compared to the pre-treatment stage. The glycocentric analysis also showed a substantial increase in fucosylated glycoforms at the pre-treatment stage, which reduced after post-treatment and matched to the control result. The glycopeptide analysis again fortifies this observation. All the mentioned results were complementing between the glycocentric and glycopeptide analysis. Hence, we set our analytical strategies, and we extended our studies with a higher number of emicizumab treatment sample pairs and inhibitor-positive samples to confirm our results. Here the differences were not very significant in response to inhibitor

development in both glycocentric and glycopeptide analysis. However, the number of patients who showed lesser antibody galactosylation was all from the inhibitor developed set. We also observed that the changes get vanished, and galactosylation improved after treating these individuals with emicizumab. Those were showing negligible changes were not showing any differences after treating with emicizumab. The human-based analysis was also giving an insight into the importance of personalized medicine and how glycosylation plays an essential role in individual-based treatment strategies.[241]

Additionally, from our study, we could see the species level distinction between the human and mice plasma glycome. The variability in the percentage of glycoforms was also noted at different strains of HA mice sample. However, we could not enrich the antibody from mice plasma. Therefore, the analysis was limited at plasma glycome level. Here the differences between the individual were higher, so we could not correlate the change specific to inhibitor development. Nevertheless, there was a slight decrease in plasma galactosylation, noted after inhibitor development. To get a more precise result, we analyzed the plasma glycopeptide analysis using orbitrap-MS with HCD- based fragmentation. The bottom-up approach results were showing a statistically significant increase in the expression level of IgGs and IgM. The site-specific heterogeneity showed that there is a slight decrease in monosialylation, responding to the inhibitor development at all the IgG subtype level.

Both the models are showing a difference at their IgG glycosylation. Both their trends were matching, primarily by reducing sialylation/ galactosylation in response to inhibitor development. The decrease in IgG galactosylation/ sialylation indicates a pro-inflammatory response, whereas the increase in galactosylation/ sialylation is indicating the anti-inflammatory response of the same antibody. Hence it is clear that patient who develops inhibitors are containing pro-inflammatory antibodies whereas the inhibitor negative individuals containing anti-inflammatory antibody in their blood. We also observed that the

emicizumab treatment could convert the immunological status of an individual from pro-inflammatory -to anti-inflammatory at a personalized level.

In chapter 3, we used the advanced mass spectrometric techniques to identify the glycosylated variants of a gene therapy vector which in turn will be useful in the field of capsid engineering. The challenge and new aspect of the study was the identification of a very small variant of glycoprotein from a complex protein mixture using different mass spectrometric techniques. Adeno Associated virus plays a significant role as a gene therapy vector to treat multiple genetic disorders. In the present study, the scope of AAV8 capsid glycosylation was systemically examined using high-resolution mass spectrometry and confirm the N-glycosylation site of the capsid protein at N499. We analyzed the overall sugar profile of purified vector produced in 293 cells, using MALDI-TOF-MS and correlated with proceeding results. We also identified the presence of co-purified host cell glycoproteins in differentially purified AAV8 vectors, highlighted the probability of AAV8 interaction with terminal galactosylated N-glycans within the HCPs.

The viral samples were expressed in HEK-293 cell line and purified separately as secreted and intra cellular fraction using density gradient centrifugation. The glycoprotein profiling was carried out by SDS-PAGE and differentially stained using pierce glycoprotein staining kit. The PNGase F treated sample was run as a negative control to confirm the N-glycosylation. The overall protein profile and purity of the viral protein were determined by Coomassie staining. The N-glycan profile by 2AA-labelling and MALDI-MS analysis identified from the differentially purified samples. The glycopeptides were enriched from the samples and analyzed using orbitrap-MS. The site of N-glycosylation was confirmed by labelling it with ^{18}O. The co-purified host cell glycoproteins were identified analyzing the peptide map data using proteome discoverer search engine and human database.

Virus capsid glycosylation is a potential PTM which may influence the vector immunogenicity, intracellular trafficking, host tissue specificity, and tissue tropism. The post-translational modification of AAV capsid protein is an underserved area. Here we identified the existence of natural N-glycosylation on the AAV8 capsid protein. The protein profiling by SDS-PAGE, indicated the glycosylated variants of capsid protein in a feeble amount in the AAV8 sample. We identified the presence of other glycoproteins which possibly from the host cell protein interaction, which further confirmed by peptide mass fingerprinting. PNGase F digested capsid proteins were analyzed side by side on the same gel also indicated the presence of heavy molecular weight variants of VP1, VP2, and VP3 in the control samples. To confirm the glycosylation, we performed peptide mass fingerprinting after enriching the glycosylated peptides. We also performed the PNGase F mediated deglycosylation in ^{18}O water, which specifically deamidate the N-glycan linked Asn to Asp labeled with ^{18}O and verified by the mass addition of 2.98 Da The peptide mass fingerprinting and tandem mass spectrometry of ^{18}O labeled glycopeptides confirmed the presence of N-linked glycan at N499 in "VSTTTGQNNNSNFAWTAGTK" peptide, localized in the common region of AAV8 capsid proteins. We also identified nonglycosylated version in ample amount implies the existence of multiple pathways for AAV8 maturation and packaging. The total protein identifications in AAV8 sample indicated the presence of different glycoproteins from host cell through AAV8-host cell interactions. The N-glycan analysis of the differentially purified AAV8 samples was performed independently to monitor the structural features of glycan present in the sample. The isolated N-glycans were labelled with 2 aminobenzoic acid and analyzed using MALDI-MS. Each glycoform was well characterized by tandem mass spectrometry and further confirmed by Glycoworkbench analysis tool. The high amount of terminally galactosylated glycoforms and presence of galectin binding proteins pointing out

the high affinity of AAV8 to the terminal galactose containing glycoforms. These results provide new insights into AAV8 vector production and capsid engineering process.

In summary, by using the advanced techniques of mass spectrometry and sample preparation, we could explore the field of glycobiology to study the significance of protein N-glycome in biomarker discovery and gene therapy using Hemophilia A inhibitor development and rAAV8 as study models.

5 REFERENCES

(1) van Belkum, A.; Welker, M.; Erhard, M.; Chatellier, S. Biomedical Mass Spectrometry in Today's and Tomorrow's Clinical Microbiology Laboratories. *J. Clin. Microbiol.* **2012**, *50* (5), 1513–1517. https://doi.org/10.1128/JCM.00420-12.

(2) Khoury, G. A.; Baliban, R. C.; Floudas, C. A. Proteome-Wide Post-Translational Modification Statistics: Frequency Analysis and Curation of the Swiss-Prot Database. *Sci. Rep.* **2011**, *1* (1), 90. https://doi.org/10.1038/srep00090.

(3) Shade, K.-T.; Conroy, M. E.; Anthony, R. M. IgE Glycosylation in Health and Disease. In *Current Topics in Microbiology and Immunology*; 2019; Vol. 423, pp 77–93. https://doi.org/10.1007/82_2019_151.

(4) Vanhooren, V.; Dewaele, S.; Libert, C.; Engelborghs, S.; De Deyn, P. P.; Toussaint, O.; Debacq-Chainiaux, F.; Poulain, M.; Glupczynski, Y.; Franceschi, C.; et al. Serum N-Glycan Profile Shift during Human Ageing. *Exp. Gerontol.* **2010**, *45* (10), 738–743. https://doi.org/10.1016/j.exger.2010.08.009.

(5) Adamczyk, B.; Albrecht, S.; Stöckmann, H.; Ghoneim, I. M.; Al-Eknah, M.; Al-Busadah, K. A. S.; Karlsson, N. G.; Carrington, S. D.; Rudd, P. M. Pregnancy-Associated Changes of IgG and Serum N-Glycosylation in Camel (Camelus Dromedarius). *J. Proteome Res.* **2016**, *15* (9), 3255–3265. https://doi.org/10.1021/acs.jproteome.6b00439.

(6) Li, M.; Song, L.; Qin, X. Glycan Changes: Cancer Metastasis and Anti-Cancer Vaccines. *J. Biosci.* **2010**, *35* (4), 665–673. https://doi.org/10.1007/s12038-010-0073-8.

(7) Chen, J.; Li, X.; Edmondson, A.; Meyers, G. D.; Izumi, K.; Ackermann, A. M.; Morava, E.; Ficicioglu, C.; Bennett, M. J.; He, M. Increased Clinical Sensitivity and Specificity of Plasma Protein N-Glycan Profiling for Diagnosing Congenital Disorders

of Glycosylation by Use of Flow Injection-Electrospray Ionization-Quadrupole Time-of-Flight Mass Spectrometry. *Clin. Chem.* **2019**, *65* (5), 653–663. https://doi.org/10.1373/clinchem.2018.296780.

(8) Mechref, Y.; Hu, Y.; Garcia, A.; Hussein, A. Identifying Cancer Biomarkers by Mass Spectrometry-Based Glycomics. *Electrophoresis* **2012**, *33* (12), 1755–1767. https://doi.org/10.1002/elps.201100715.

(9) Powers, T.; Holst, S.; Wuhrer, M.; Mehta, A.; Drake, R. Two-Dimensional N-Glycan Distribution Mapping of Hepatocellular Carcinoma Tissues by MALDI-Imaging Mass Spectrometry. *Biomolecules* **2015**, *5* (4), 2554–2572. https://doi.org/10.3390/biom5042554.

(10) Rejzek, M.; Hill, L.; Hems, E. S.; Kuhaudomlarp, S.; Wagstaff, B. A.; Field, R. A. Profiling of Sugar Nucleotides. In *Methods in Enzymology*; Elsevier Inc., 2017; Vol. 597, pp 209–238. https://doi.org/10.1016/bs.mie.2017.06.005.

(11) Del Val, I. J.; Polizzi, K. M.; Kontoravdi, C. A Theoretical Estimate for Nucleotide Sugar Demand towards Chinese Hamster Ovary Cellular Glycosylation. *Sci. Rep.* **2016**, *6* (June), 1–15. https://doi.org/10.1038/srep28547.

(12) Lodish H, Berk A, Zipursky SL, et al. Protein Glycosylation in the ER and Golgi Complex. *Mol. Cell Biol.* **2000**, *4th editio* (Section 17.7), 21744.

(13) Roth, Z.; Yehezkel, G.; Khalaila, I. Identification and Quantification of Protein Glycosylation. *Int. J. Carbohydr. Chem.* **2012**, *2012*, 1–10. https://doi.org/10.1155/2012/640923.

(14) Ellgaard, L.; Molinari, M.; Helenius, A. Setting the Standards: Quality Control in the Secretory Pathway. *Science (80-.).* **1999**, *286* (5446), 1882–1888. https://doi.org/10.1126/science.286.5446.1882.

(15) Schachter, H. The Joys of HexNAc. The Synthesis and Function of N- and O-Glycan Branches. *Glycoconj. J.* **2000**, *17* (7–9), 465–483. https://doi.org/10.1023/A:1011010206774.

(16) Takahashi, M.; Kuroki, Y.; Ohtsubo, K.; Taniguchi, N. Core Fucose and Bisecting GlcNAc, the Direct Modifiers of the N-Glycan Core: Their Functions and Target Proteins. *Carbohydr. Res.* **2009**, *344* (12), 1387–1390. https://doi.org/10.1016/j.carres.2009.04.031.

(17) Moran, A. P.; Gupta, A.; Joshi, L. Sweet-Talk: Role of Host Glycosylation in Bacterial Pathogenesis of the Gastrointestinal Tract. *Gut* **2011**, *60* (10), 1412–1425. https://doi.org/10.1136/gut.2010.212704.

(18) Faid, V.; Denguir, N.; Chapuis, V.; Bihoreau, N.; Chevreux, G. Site-Specific N - Glycosylation Analysis of Human Factor XI: Identification of a Noncanonical NXC Glycosite. *Proteomics* **2014**, *14* (21–22), 2460–2470. https://doi.org/10.1002/pmic.201400038.

(19) Nita-Lazar, M.; Wacker, M.; Schegg, B.; Amber, S.; Aebi, M. The N-X-S/T Consensus Sequence Is Required but Not Sufficient for Bacterial N-Linked Protein Glycosylation. *Glycobiology* **2005**, *15* (4), 361–367. https://doi.org/10.1093/glycob/cwi019.

(20) Zauner, G.; Kozak, R. P.; Gardner, R. A.; Fernandes, D. L.; Deelder, A. M.; Wuhrer, M. Protein O-Glycosylation Analysis. *Biol. Chem.* **2012**, *393* (8), 687–708. https://doi.org/10.1515/hsz-2012-0144.

(21) You, X.; Qin, H.; Ye, M. Recent Advances in Methods for the Analysis of Protein O-Glycosylation at Proteome Level. *J. Sep. Sci.* **2018**, *41* (1), 248–261. https://doi.org/10.1002/jssc.201700834.

(22) *Brockhausen, I.; Schachter, H.; Stanley, P., O-GalNAc Glycans. In Essentials of Glycobiology , Nd; Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Eds. Cold Spring Harbor (NY), 2009.*, 2nd ed.; Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Ed.; Cold Spring Harbor (NY),: NY, 2009; Vol. Chapter 9.

(23) Shcherbakova, A.; Tiemann, B.; Buettner, F. F. R.; Bakker, H. Distinct C-Mannosylation of Netrin Receptor Thrombospondin Type 1 Repeats by Mammalian DPY19L1 and DPY19L3. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (10), 2574–2579. https://doi.org/10.1073/pnas.1613165114.

(24) Christoph, M. Biomedical Applications of Glycosylphosphatidylinositol-Anchored Proteins. 1–26.

(25) S. Defaus,ab P. Gupta, b D. A. and R. G.-G. Mammalian Protein Glycosylation – Structure versus Function. **2014**, 2944–2967. https://doi.org/10.1039/c3an02245e.

(26) Maynard, J. C.; Burlingame, A. L.; Medzihradszky, K. F. Cysteine S-Linked N-Acetylglucosamine (S-GlcNAcylation), a New Post-Translational Modification in Mammals. *Mol. Cell. Proteomics* **2016**, *15* (11), 3405–3411. https://doi.org/10.1074/mcp.M116.061549.

(27) Banazadeh, A.; Veillon, L.; Wooding, K. M.; Zabet-moghaddam, M.; Mechref, Y. Recent Advances in Mass Spectrometric Analysis of Glycoproteins. *Electrophoresis*

2017, *38* (1), 162–189. https://doi.org/10.1002/elps.201600357.

(28) Gupta, S.; Bent, S.; Kohlwes, J. Test Characteristics of Alpha-Fetoprotein for Detecting Hepatocellular Carcinoma in Patients with Hepatitis C. A Systematic Review and Critical Analysis. *Ann. Intern. Med.* **2003**, *139* (1), 46–50. https://doi.org/10.7326/0003-4819-139-1-200307010-00012.

(29) Saito, S.; Ojima, H.; Ichikawa, H.; Hirohashi, S.; Kondo, T. Molecular Background of α-Fetoprotein in Liver Cancer Cells as Revealed by Global RNA Expression Analysis. *Cancer Sci.* **2008**, *99* (12), 2402–2409. https://doi.org/10.1111/j.1349-7006.2008.00973.x.

(30) Saffroy, R.; Pham, P.; Reffas, M.; Takka, M.; Lemoine, A.; Debuire, B. New Perspectives and Strategy Research Biomarkers for Hepatocellular Carcinoma. *Clin. Chem. Lab. Med.* **2007**, *45* (9), 1169–1179. https://doi.org/10.1515/CCLM.2007.262.

(31) Guo, J.; Yu, J.; Song, X.; Mi, H. Serum CA125, CA199 and CEA Combined Detection for Epithelial Ovarian Cancer Diagnosis: A Meta-Analysis. *Open Med.* **2017**, *12* (1), 131–137. https://doi.org/10.1515/med-2017-0020.

(32) Pepin, K.; Carmen, M. del; Brown, A.; Dizon, D. S. CA 125 and Epithelial Ovarian Cancer: Role in Screening, Diagnosis, and Surveillance. *Am. J. Hematol. / Oncol.* **2014**, *10* (6), 22–29.

(33) Pérez-Ibave, D. C.; Burciaga-Flores, C. H.; Elizondo-Riojas, M. Á. Prostate-Specific Antigen (PSA) as a Possible Biomarker in Non-Prostatic Cancer: A Review. *Cancer Epidemiol.* **2018**, *54* (March), 48–55. https://doi.org/10.1016/j.canep.2018.03.009.

(34) Miyahara, K.; Nouso, K.; Dohi, C.; Morimoto, Y.; Kinugasa, H.; Wada, N.; Takeuchi, Y.; Kuwaki, K.; Onishi, H.; Ikeda, F.; et al. Alteration of N-Glycan Profiles in Patients with Chronic Hepatitis and Hepatocellular Carcinoma. *Hepatol. Res.* **2015**, *45* (9), 986–993. https://doi.org/10.1111/hepr.12441.

(35) Qin, R.; Zhao, J.; Qin, W.; Zhang, Z.; Zhao, R.; Han, J.; Yang, Y.; Li, L.; Wang, X.; Ren, S.; et al. Discovery of Non-Invasive Glycan Biomarkers for Detection and Surveillance of Gastric Cancer. *J. Cancer* **2017**, *8* (10), 1908–1916. https://doi.org/10.7150/jca.17900.

(36) Zhao, Y.-P.; Zhou, P.-T.; Ji, W.-P.; Wang, H.; Fang, M.; Wang, M.-M.; Yin, Y.-P.; Jin, G.; Gao, C.-F. Validation of N-Glycan Markers That Improve the Performance of CA19-9 in Pancreatic Cancer. *Clin. Exp. Med.* **2017**, *17* (1), 9–18. https://doi.org/10.1007/s10238-015-0401-2.

(37) Saldova, R.; Asadi Shehni, A.; Haakensen, V. D.; Steinfeld, I.; Hilliard, M.; Kifer, I.;

Helland, Å.; Yakhini, Z.; Børresen-Dale, A. L.; Rudd, P. M. Association of N-Glycosylation with Breast Carcinoma and Systemic Features Using High-Resolution Quantitative UPLC. *J. Proteome Res.* **2014**, *13* (5), 2314–2327. https://doi.org/10.1021/pr401092y.

(38) Britannica, T. E. of E. Antibody. *Encycl. Br.* **2020**, *2011* (2865), 1–9.

(39) Vestrheim, A. C.; Moen, A.; Egge-Jacobsen, W.; Reubsaet, L.; Halvorsen, T. G.; Bratlie, D. B.; Paulsen, B. S.; Michaelsen, T. E. A Pilot Study Showing Differences in Glycosylation Patterns of Igg Subclasses Induced by Pneumococcal, Meningococcal, and Two Types of Influenza Vaccines. *Immun. Inflamm. Dis.* **2014**, *2* (2). https://doi.org/10.1002/iid3.22.

(40) Russell, A.; Adua, E.; Ugrina, I.; Laws, S.; Wang, W. Unravelling Immunoglobulin G Fc N-Glycosylation: A Dynamic Marker Potentiating Predictive, Preventive and Personalised Medicine. *Int. J. Mol. Sci.* **2018**, *19* (2), 390. https://doi.org/10.3390/ijms19020390.

(41) Maverakis, E.; Michiko Shimoda, K.; Gershwin, M.; Patel, F.; Wilken, R.; Raychaudhuri, S.; Ruhaak, R.; Lebrilla, C. Glycans In The Immune System and The Altered Glycan Theory of Autoimmunity: A Critical Review. *J. Autoimmun.* **2015**, *48* (916), 1–13. https://doi.org/10.1097/MPG.0b013e3181a15ae8.Screening.

(42) Hulett, M. D.; Hogarth, P. M. Molecular Basis of Fc Receptor Function. In *Advances in immunology*; 1994; Vol. 57, pp 1–127. https://doi.org/10.1016/S0065-2776(08)60671-9.

(43) Daëron, M. Fc Receptor Biology. *Annu. Rev. Immunol.* **1997**, *15* (1), 203–234. https://doi.org/10.1146/annurev.immunol.15.1.203.

(44) Ravetch, J. V.; Bolland, S. IgG Fc Receptors. *Annu. Rev. Immunol.* **2001**, *19* (1), 275–290. https://doi.org/10.1146/annurev.immunol.19.1.275.

(45) Stevens, J.; Blixt, O.; Paulson, J. C.; Wilson, I. A. Glycan Microarray Technologies: Tools to Survey Host Specificity of Influenza Viruses. *Nat. Rev. Microbiol.* **2006**, *4* (11), 857–864. https://doi.org/10.1038/nrmicro1530.

(46) Lipton, H. L.; Kumar, A. S. M.; Hertzler, S.; Reddi, H. V. Differential Usage of Carbohydrate Co-Receptors Influences Cellular Tropism of Theiler's Murine Encephalomyelitis Virus Infection of the Central Nervous System. *Glycoconj. J.* **2006**, *23* (1–2), 39–49. https://doi.org/10.1007/s10719-006-5436-x.

(47) Zocher, G.; Mistry, N.; Frank, M.; Hähnlein-Schick, I.; Ekström, J. O.; Arnberg, N.; Stehle, T. A Sialic Acid Binding Site in a Human Picornavirus. *PLoS Pathog.* **2014**, *10*

(10), 2–10. https://doi.org/10.1371/journal.ppat.1004401.

(48) Nokhbeh, M. R.; Hazra, S.; Alexander, D. A.; Khan, A.; McAllister, M.; Suuronen, E. J.; Griffith, M.; Dimock, K. Enterovirus 70 Binds to Different Glycoconjugates Containing 2,3-Linked Sialic Acid on Different Cell Lines. *J. Virol.* **2005**, *79* (11), 7087–7094. https://doi.org/10.1128/jvi.79.11.7087-7094.2005.

(49) Liu, Y.; Sheng, J.; Baggen, J.; Meng, G.; Xiao, C.; Thibaut, H. J.; Van Kuppeveld, F. J. M.; Rossmann, M. G. Sialic Acid-Dependent Cell Entry of Human Enterovirus D68. *Nat. Commun.* **2015**, *6*, 1–7. https://doi.org/10.1038/ncomms9865.

(50) Chen S1, Kapturczak M, Loiler SA, Zolotukhin S, Glushakova OY, Madsen KM, Samulski RJ, Hauswirth WW, Campbell-Thompson M, Berns KI, Flotte TR, Atkinson MA, Tisher CC, A. A. Efficient Transduction of Vascular Endothelial Cells with Recombinant Adeno-Associated Virus Serotype 1 and 5 Vectors. *Hum Gene* **2005**, *16* (2), 235–247.

(51) Gonçalves, G. A. R.; Paiva, R. de M. A. Gene Therapy: Advances, Challenges and Perspectives. *Einstein (Sao Paulo).* **2017**, *15* (3), 369–375. https://doi.org/10.1590/S1679-45082017RB4024.

(52) Yi, Y.; Jong Noh, M.; Hee Lee, K. Current Advances in Retroviral Gene Therapy. *Curr. Gene Ther.* **2011**, *11* (3), 218–228. https://doi.org/10.2174/156652311795684740.

(53) Wilson, J. M.; Engelhardt, J. Adenovirus Vectors for Gene Therapy. *Biotechnol. Adv.* **1997**, *15* (3–4), 769. https://doi.org/10.1201/9780824758608.ch2.

(54) Hukkanen, V. Editorial: [Herpesvirus Vectors in Gene Therapy]. *Open Virol. J.* **2010**, *4* (3), 94–95. https://doi.org/10.2174/1874357901004030094.

(55) Lewin, A. S.; Hauswirth, W. W. Ribozyme Gene Therapy: Applications for Molecular Medicine. *Trends Mol. Med.* **2001**, *7* (5), 221–228. https://doi.org/10.1016/S1471-4914(01)01965-7.

(56) Naso, M. F.; Tomkowicz, B.; Perry, W. L.; Strohl, W. R. Adeno-Associated Virus (AAV) as a Vector for Gene Therapy. *BioDrugs* **2017**, *31* (4), 317–334. https://doi.org/10.1007/s40259-017-0234-5.

(57) Hareendran, S.; Balakrishnan, B.; Sen, D. Adeno-Associated Virus (AAV) Vectors in Gene Therapy : Immune Challenges and Strategies to Circumvent Them. **2013**, No. September, 399–413. https://doi.org/10.1002/rmv.

(58) Sarkar, R.; Tetreault, R.; Gao, G.; Wang, L.; Bell, P.; Chandler, R.; Wilson, J. M.; Kazazian, H. H. Total Correction of Hemophilia A Mice with Canine FVIII Using an

AAV 8 Serotype. *Blood* **2004**, *103* (4), 1253–1260. https://doi.org/10.1182/blood-2003-08-2954.

(59) Liu, H.; Zhang, N.; Wan, D.; Cui, M.; Liu, Z.; Liu, S. Mass Spectrometry-Based Analysis of Glycoproteins and Its Clinical Applications in Cancer Biomarker Discovery. *Clin. Proteomics* **2014**, *11* (1), 1–9. https://doi.org/10.1186/1559-0275-11-14.

(60) Dodds, E. D. Gas-Phase Dissociation of Glycosylated Peptide Ions. *Mass Spectrom. Rev.* **2012**, *31* (6), 666–682. https://doi.org/10.1002/mas.21344.

(61) Pabst, M.; Altmann, F. Glycan Analysis by Modern Instrumental Methods. *Proteomics* **2011**, *11* (4), 631–643. https://doi.org/10.1002/pmic.201000517.

(62) Sansing, L. H.; Harris, T. H.; Welsh, F. A.; Kasner, S. E.; Hunter, C. A.; Kariko, K. Toll-like Receptor 4 Contributes to Poor Outcome after Intracerebral Hemorrhage. *Ann. Neurol.* **2011**, *70* (4), 646–656. https://doi.org/10.1002/ana.22528.

(63) Quenin, F. Le Dépôt Légal de La Bibliothèque Nationale. *Bibliogr. la Fr.* **1979**, *a. 168* (20), 10–14. https://doi.org/10.1134/S0006297913070031.Mass.

(64) Patrie, S. M.; Roth, M. J.; Kohler, J. J. Introduction to Glycosylation and Mass Spectrometry. *Methods Mol. Biol.* **2013**, *951*, 1–17. https://doi.org/10.1007/978-1-62703-146-2_1.

(65) Liu, M.-Q.; Zeng, W.-F.; Fang, P.; Cao, W.-Q.; Liu, C.; Yan, G.-Q.; Zhang, Y.; Peng, C.; Wu, J.-Q.; Zhang, X.-J.; et al. PGlyco 2.0 Enables Precision N-Glycoproteomics with Comprehensive Quality Control and One-Step Mass Spectrometry for Intact Glycopeptide Identification. *Nat. Commun.* **2017**, *8* (1), 438. https://doi.org/10.1038/s41467-017-00535-2.

(66) Zeng, W. F.; Liu, M. Q.; Zhang, Y.; Wu, J. Q.; Fang, P.; Peng, C.; Nie, A.; Yan, G.; Cao, W.; Liu, C.; et al. PGlyco: A Pipeline for the Identification of Intact N-Glycopeptides by Using HCD- and CID-MS/MS and MS3. *Sci. Rep.* **2016**, *6* (May), 1–10. https://doi.org/10.1038/srep25102.

(67) Jansen, B. C.; Reiding, K. R.; Bondt, A.; Hipgrave Ederveen, A. L.; Palmblad, M.; Falck, D.; Wuhrer, M. MassyTools: A High-Throughput Targeted Data Processing Tool for Relative Quantitation and Quality Control Developed for Glycomic and Glycoproteomic MALDI-MS. *J. Proteome Res.* **2015**, *14* (12), 5088–5098. https://doi.org/10.1021/acs.jproteome.5b00658.

(68) Irungu, J.; Go, E. P.; Dalpathado, D. S.; Desaire, H. Simplification of Mass Spectral Analysis of Acidic Glycopeptides Using GlycoPep ID. *Anal. Chem.* **2007**, *79* (8),

3065–3074. https://doi.org/10.1021/ac062100e.

(69) He, L.; Xin, L.; Shan, B.; Lajoie, G. A.; Ma, B. GlycoMaster DB: Software to Assist the Automated Identification of N-Linked Glycopeptides by Tandem Mass Spectrometry. *J. Proteome Res.* **2014**, *13* (9), 3881–3895. https://doi.org/10.1021/pr401115y.

(70) Zhu, Z.; Su, X.; Go, E. P.; Desaire, H. New Glycoproteomics Software, Glycopep Evaluator, Generates Decoy Glycopeptides de Novo and Enables Accurate False Discovery Rate Analysis for Small Data Sets. *Anal. Chem.* **2014**, *86* (18), 9212–9219. https://doi.org/10.1021/ac502176n.

(71) Park, G. W.; Kim, J. Y.; Hwang, H.; Lee, J. Y.; Ahn, Y. H.; Lee, H. K.; Ji, E. S.; Kim, K. H.; Jeong, H. K.; Yun, K. N.; et al. Integrated GlycoProteome Analyzer (I-GPA) for Automated Identification and Quantitation of Site-Specific N-Glycosylation. *Sci. Rep.* **2016**, *6* (June 2015), 1–12. https://doi.org/10.1038/srep21175.

(72) Kocurek, K. I.; Griffiths, R. L.; Cooper, H. J. Ambient Ionisation Mass Spectrometry for in Situ Analysis of Intact Proteins. *J. Mass Spectrom.* **2018**, *53* (7), 565–578. https://doi.org/10.1002/jms.4087.

(73) Tarentino, A. L.; Gomez, C. M.; Plummer, T. H. Deglycosylation of Asparagine-Linked Glycans by Peptide: N-Glycosidase F. *Biochemistry* **1985**, *24* (17), 4665–4671. https://doi.org/10.1021/bi00338a028.

(74) Xuezheng Song, Hong Ju, Yi Lasanajak, Matthew R. Kudelka, David F. Smith, and R. D. C. Oxidative Release of Natural Glycans for Functional Glycomics. *Nat Methods* **2016**, *13* (6), 528–534. https://doi.org/doi:10.1038/nmeth.3861.

(75) Zanto, T. P.; Hennigan, K.; Östberg, M.; Clapp, W. C.; Gazzaley, A. NIH Public Access. **2011**, *46* (4), 564–574. https://doi.org/10.1016/j.cortex.2009.08.003.Predictive.

(76) Huang, Y.; Konse, T.; Mechref, Y.; Novotny, M. V. Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Compatible β-Elimination of O-Linked Oligosaccharides. *Rapid Commun. Mass Spectrom.* **2002**, *16* (12), 1199–1204. https://doi.org/10.1002/rcm.701.

(77) Carlson, D. M. Oligosaccharides Isolated from Pig Submaxillary Mucin. *J. Biol. Chem.* **1966**, *241* (12), 2984–2986.

(78) Merry, A. H.; Neville, D. C. A.; Royle, L.; Matthews, B.; Harvey, D. J.; Dwek, R. A.; Rudd, P. M. Recovery of Intact 2-Aminobenzamide-Labeled O-Glycans Released from Glycoproteins by Hydrazinolysis. *Anal. Biochem.* **2002**, *304* (1), 91–99.

https://doi.org/10.1006/abio.2002.5620.

(79) Kudelka, M. R.; Antonopoulos, A.; Wang, Y.; Duong, D. M.; Song, X.; Seyfried, N. T.; Dell, A.; Haslam, S. M.; Cummings, R. D.; Ju, T. Cellular O-Glycome Reporter/Amplification to Explore O-Glycans of Living Cells. *Nat. Methods* **2016**, *13* (1), 81–86. https://doi.org/10.1038/nmeth.3675.

(80) Morelle, W.; Michalski, J.-C. Analysis of Protein Glycosylation by Mass Spectrometry. *Nat. Protoc.* **2007**, *2* (7), 1585–1602. https://doi.org/10.1038/nprot.2007.227.

(81) Ruhaak, L. R.; Zauner, G.; Huhn, C.; Bruggink, C.; Deelder, A. M.; Wuhrer, M. Glycan Labeling Strategies and Their Use in Identification and Quantification. *Anal. Bioanal. Chem.* **2010**, *397* (8), 3457–3481. https://doi.org/10.1007/s00216-010-3532-z.

(82) Anumula, K. R.; Dhume, S. T. 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* **1998**, *8* (7), 685–694. https://doi.org/10.1093/glycob/8.7.685.

(83) Pabst, M.; Kolarich, D.; Pöltl, G.; Dalik, T.; Lubec, G.; Hofinger, A.; Altmann, F. Comparison of Fluorescent Labels for Oligosaccharides and Introduction of a New Postlabeling Purification Method. *Anal. Biochem.* **2009**, *384* (2), 263–273. https://doi.org/10.1016/j.ab.2008.09.041.

(84) You, J.; Sheng, X.; Ding, C.; Sun, Z.; Suo, Y.; Wang, H.; Li, Y. Detection of Carbohydrates Using New Labeling Reagent 1-(2-Naphthyl)-3-Methyl-5-Pyrazolone by Capillary Zone Electrophoresis with Absorbance (UV). *Anal. Chim. Acta* **2008**, *609* (1), 66–75. https://doi.org/10.1016/j.aca.2007.12.022.

(85) Jiang, K.; Aloor, A.; Qu, J.; Xiao, C.; Wu, Z.; Ma, C.; Zhang, L.; Wang, P. G. Rapid and Sensitive MALDI MS Analysis of Oligosaccharides by Using 2-Hydrazinopyrimidine as a Derivative Reagent and Co-Matrix. *Anal. Bioanal. Chem.* **2017**, *409* (2), 421–429. https://doi.org/10.1007/s00216-016-9690-x.

(86) Shubhakar, A.; Kozak, R. P.; Reiding, K. R.; Royle, L.; Spencer, D. I. R.; Fernandes, D. L.; Wuhrer, M. Automated High-Throughput Permethylation for Glycosylation Analysis of Biologics Using MALDI-TOF-MS. **2016**, No. v. https://doi.org/10.1021/acs.analchem.6b01639.

(87) Phillips, L. R.; Fraser, B. A. Methylation of Carbohydrates with Dimsyl Potassium in Dimethyl Sulfoxide. *Carbohydr. Res.* **1981**, *90* (1), 149–152.

https://doi.org/10.1016/S0008-6215(00)85623-2.

(88) Zaia, J. Mass Spectrometry of Oligosaccharides. *Mass Spectrom. Rev.* **2004**, *23* (3), 161–227. https://doi.org/10.1002/mas.10073.

(89) Mechref Y1, Hu Y, Garcia A, Zhou S, Desantos-Garcia JL, H. A. Defining Putative Glycan Cancer Biomarkers by Mass Spectrometry. *Bioanalysis* **2013**, *4* (20), 2457–2469. https://doi.org/10.4155/bio.12.246.Defining.

(90) Jiang, K.; Zhu, H.; Li, L.; Guo, Y.; Gashash, E.; Ma, C.; Sun, X.; Li, J.; Zhang, L.; Wang, P. G. Sialic Acid Linkage-Specific Permethylation for Improved Profiling of Protein Glycosylation by MALDI-TOF MS. *Anal. Chim. Acta* **2017**, *981*, 53–61. https://doi.org/10.1016/j.aca.2017.05.029.

(91) Selman, M. H. J.; Hemayatkar, M.; Deelder, A. M.; Wuhrer, M. Cotton HILIC SPE Microtips for Microscale Purification and Enrichment of Glycans and Glycopeptides. *Anal. Chem.* **2011**, *83* (7), 2492–2499. https://doi.org/10.1021/ac1027116.

(92) Sun, S.; Shah, P.; Eshghi, S. T.; Yang, W.; Trikannad, N.; Yang, S.; Chen, L.; Aiyetan, P.; Höti, N.; Zhang, Z.; et al. Comprehensive Analysis of Protein Glycosylation by Solid-Phase Extraction of N-Linked Glycans and Glycosite-Containing Peptides. *Nat. Biotechnol.* **2016**, *34* (1), 84–88. https://doi.org/10.1038/nbt.3403.

(93) Buszewski, B.; Noga, S. Hydrophilic Interaction Liquid Chromatography (HILIC)-a Powerful Separation Technique. *Anal. Bioanal. Chem.* **2012**, *402* (1), 231–247. https://doi.org/10.1007/s00216-011-5308-5.

(94) Fu, Q.; Liang, T.; Li, Z.; Xu, X.; Ke, Y.; Jin, Y.; Liang, X. Separation of Carbohydrates Using Hydrophilic Interaction Liquid Chromatography. *Carbohydr. Res.* **2013**, *379*, 13–17. https://doi.org/10.1016/j.carres.2013.06.006.

(95) Zauner, G.; Koeleman, C. A. M.; Deelder, A. M.; Wuhrer, M. Protein Glycosylation Analysis by HILIC-LCMS of Proteinase K-Generated N- and Oglycopeptides. *J. Sep. Sci.* **2010**, *33* (6–7), 903–910. https://doi.org/10.1002/jssc.200900850.

(96) Yu, L.; Li, X.; Guo, Z.; Zhang, X.; Liang, X. Hydrophilic Interaction Chromatography Based Enrichment of Glycopeptides by Using Click Maltose: A Matrix with High Selectivity and Glycosylation Heterogeneity Coverage. *Chem. - A Eur. J.* **2009**, *15* (46), 12618–12626. https://doi.org/10.1002/chem.200902370.

(97) Tanaka, H.; Zhou, X.; Masayoshi, O. Characterization of a Novel Diol Column for High-Performance Liquid Chromatography. *J. Chromatogr. A* **2003**, *987* (1–2), 119–125. https://doi.org/10.1016/S0021-9673(02)01949-0.

(98) Zhao, Y.; Yu, L.; Guo, Z.; Li, X.; Liang, X. Reversed-Phase Depletion Coupled with

Hydrophilic Affinity Enrichment for the Selective Isolation of N-Linked Glycopeptides by Using Click OEG-CD Matrix. *Anal. Bioanal. Chem.* **2011**, *399* (10), 3359–3365. https://doi.org/10.1007/s00216-011-4652-9.

(99) Takegawa Y1, Deguchi K, Ito H, Keira T, Nakagawa H, N. S. Simple Separation of Isomeric Sialylated N-Glycopeptides by a Zwitterionic Type of Hydrophilic Interaction Chromatography. *J Sep Sci.* **2006**, *29* (16), 2533–2540. https://doi.org/10.1002/jssc.200600133.

(100) Zhang, H.; Guo, T.; Li, X.; Datta, A.; Park, J. E.; Yang, J.; Lim, S. K.; Tam, J. P.; Sze, S. K. Simultaneous Characterization of Glyco- and Phosphoproteomes of Mouse Brain Membrane Proteome with Electrostatic Repulsion Hydrophilic Interaction Chromatography. *Mol. Cell. Proteomics* **2010**, *9* (4), 635–647. https://doi.org/10.1074/mcp.M900314-MCP200.

(101) Hao, P.; Guo, T.; Sze, S. K. Simultaneous Analysis of Proteome, Phospho- and Glycoproteome of Rat Kidney Tissue with Electrostatic Repulsion Hydrophilic Interaction Chromatography. *PLoS One* **2011**, *6* (2), e16884. https://doi.org/10.1371/journal.pone.0016884.

(102) Hardman, K. D.; Ainsworth, C. F. Myo-Inositol Binding Site of Concanavalin A. *Nat. New Biol.* **1972**, *237* (71), 54–55. https://doi.org/10.1038/newbio237054a0.

(103) Kornfeld K, Reitman ML, K. R. The Carbohydrate-Binding Specificity of Pea and Lentil Lectins. Fucose Is an Important Determinant. *J Biol Chem* **1981**, *256* (13), 6633–6640.

(104) Nicolson, G. L.; Blaustein, J.; Etzler, M. E. Characterization of Two Plant Lectins from Ricinus Communis and Their Quantitative Interaction with a Murine Lymphoma. *Biochemistry* **1974**, *13* (1), 196–204. https://doi.org/10.1021/bi00698a029.

(105) Larsen, M. R.; Jensen, S. S.; Jakobsen, L. A.; Heegaard, N. H. H. Exploring the Sialiome Using Titanium Dioxide Chromatography and Mass Spectrometry. *Mol. Cell. Proteomics* **2007**, *6* (10), 1778–1787. https://doi.org/10.1074/mcp.M700086-MCP200.

(106) Ravi Chand Bollineni, Christian Jeffrey Koehler, Randi Elin Gislefoss, J. H. A. & B. T. Large-Scale Intact Glycopeptide Identification by Mascot Database Search Ravi Chand Bollineni, Christian Jeffrey Koehler, Randi Elin Gislefoss, Jan Haug Anonsen & Bernd Thiede Scientific Reports Volume 8, Article Number: 2117 (2018). *Sci. Rep.* **2018**, *8*, 2117.

(107) Ma, W. F.; Li, L. L.; Zhang, Y.; An, Q.; You, L. J.; Li, J. M.; Zhang, Y. T.; Xu, S.; Yu, M.; Guo, J.; et al. Ligand-Free Strategy for Ultrafast and Highly Selective Enrichment

of Glycopeptides Using Ag-Coated Magnetic Nanoarchitectures. *J. Mater. Chem.* **2012**, *22* (45), 23981–23988. https://doi.org/10.1039/c2jm35196j.

(108) Wen, L.; Zheng, Y.; Jiang, K.; Zhang, M.; Kondengaden, S. M.; Li, S.; Huang, K.; Li, J.; Song, J.; Wang, P. G. Two-Step Chemoenzymatic Detection of N-Acetylneuraminic Acid-α(2-3)-Galactose Glycans. *J. Am. Chem. Soc.* **2016**, *138* (36), 11473–11476. https://doi.org/10.1021/jacs.6b07132.

(109) Hao, P.; Ren, Y.; Alpert, A. J.; Siu, K. S. Detection, Evaluation and Minimization of Nonenzymatic Deamidation in Proteomic Sample Preparation. *Mol. Cell. Proteomics* **2011**, *10* (10), 1–11. https://doi.org/10.1074/mcp.O111.009381.

(110) Gonzalez, J.; Takao, T.; Hori, H.; Besada, V.; Rodriguez, R.; Padron, G.; Shimonishi, Y. A Method for Determination of N-Glycosylation Sites in Glycoproteins by Collision-Induced Dissociation Analysis in Fast Atom Bombardment Mass Spectrometry: Identification of the Positions of Carbohydrate-Linked Asparagine in Recombinant α-Amylase by Treat. *Anal. Biochem.* **1992**, *205* (1), 151–158. https://doi.org/10.1016/0003-2697(92)90592-U.

(111) Zhang, W.; Cao, W.; Huang, J.; Wang, H.; Wang, J.; Xie, C.; Yang, P. PNGase F-Mediated Incorporation of (18)O into Glycans for Relative Glycan Quantitation. *Analyst* **2015**, *140* (4), 1082–1089. https://doi.org/10.1039/c4an02073a.

(112) Robbins, P. W.; Trimble, R. B.; Wirth, D. F.; Hering, C.; Maley, F.; Maley, G. F.; Das, R.; Gibson, B. W.; Royal, N.; Biemann, K. Primary Structure of the Streptomyces Enzyme Endo-Beta-N-Acetylglucosaminidase H. *J. Biol. Chem.* **1984**, *259* (12), 7577–7583.

(113) Ma, C.; Qu, J.; Meisner, J.; Zhao, X.; Li, X.; Wu, Z.; Zhu, H.; Yu, Z.; Li, L.; Guo, Y.; et al. Convenient and Precise Strategy for Mapping N-Glycosylation Sites Using Microwave-Assisted Acid Hydrolysis and Characteristic Ions Recognition. *Anal. Chem.* **2015**, *87* (15), 7833–7839. https://doi.org/10.1021/acs.analchem.5b02177.

(114) Harvey, D. J. Mass Spectrometric Analysis of Glycosylated Viral Proteins. *Expert Rev. Proteomics* **2018**, *15* (5), 391–412. https://doi.org/10.1080/14789450.2018.1468756.

(115) Paul, W.; Steinwedel, H. Notizen: Ein Neues Massenspektrometer Ohne Magnetfeld. *Zeitschrift für Naturforsch. A* **1953**, *8* (7), 448–450. https://doi.org/10.1515/zna-1953-0710.

(116) Rajawat, J.; Jhingan, G. *Mass Spectroscopy*; Elsevier Inc., 2019. https://doi.org/10.1016/B978-0-12-816548-5.00001-0.

(117) Hecht, E. S.; Scigelova, M.; Eliuk, S.; Makarov, A. *Fundamentals and Advances of*

Orbitrap Mass Spectrometry; 2019. https://doi.org/10.1002/9780470027318.a9309.pub2.

(118) Maverakis, E.; Kim, K.; Shimoda, M.; Gershwin, M. E.; Patel, F.; Wilken, R.; Raychaudhuri, S.; Ruhaak, L. R.; Lebrilla, C. B. Glycans in the Immune System and The Altered Glycan Theory of Autoimmunity: A Critical Review. *J. Autoimmun.* **2015**, *57*, 1–13. https://doi.org/10.1016/j.jaut.2014.12.002.

(119) Singh, C.; Zampronio, C. G.; Creese, A. J.; Cooper, H. J. Higher Energy Collision Dissociation (HCD) Product Ion-Triggered Electron Transfer Dissociation (ETD) Mass Spectrometry for the Analysis of N-Linked Glycoproteins. *J. Proteome Res.* **2012**, *11* (9), 4517–4525. https://doi.org/10.1021/pr300257c.

(120) Segu, Z. M.; Mechref, Y. Characterizing Protein Glycosylation Sites through Higher-Energy C-Trap Dissociation. *Rapid Commun. Mass Spectrom.* **2010**, *24* (9), 1217–1225. https://doi.org/10.1002/rcm.4485.

(121) Wang, D.; Hincapie, M.; Rejtar, T.; Karger, B. L. Ultrasensitive Characterization of Site-Specific Glycosylation of Affinity-Purified Haptoglobin from Lung Cancer Patient Plasma Using 10 Mm i.d. Porous Layer Open Tubular Liquid Chromatography-Linear Ion Trap Collision-Induced Dissociation/Electron Transf. *Anal. Chem.* **2011**, *83* (6), 2029–2037. https://doi.org/10.1021/ac102825g.

(122) Hinneburg, H.; Stavenhagen, K.; Schweiger-Hufnagel, U.; Pengelley, S.; Jabs, W.; Seeberger, P. H.; Silva, D. V.; Wuhrer, M.; Kolarich, D. The Art of Destruction: Optimizing Collision Energies in Quadrupole-Time of Flight (Q-TOF) Instruments for Glycopeptide-Based Glycoproteomics. *J. Am. Soc. Mass Spectrom.* **2016**, *27* (3), 507–519. https://doi.org/10.1007/s13361-015-1308-6.

(123) Jiang, H.; Desaire, H.; Butnev, V. Y.; Bousfield, G. R. Glycoprotein Profiling by Electrospray Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2004**, *15* (5), 750–758. https://doi.org/10.1016/j.jasms.2004.01.009.

(124) Sandra, K.; Devreese, B.; Van Beeumen, J.; Stals, I.; Claeyssens, M. The Q-Trap Mass Spectrometer, a Novel Tool in the Study of Protein Glycosylation. *J. Am. Soc. Mass Spectrom.* **2004**, *15* (3), 413–423. https://doi.org/10.1016/j.jasms.2003.11.003.

(125) Reinhold, V. N.; Reinhold, B. B.; Costello, C. E. Carbohydrate Molecular Weight Profiling, Sequence, Linkage, and Branching Data: ES-MS and CID. *Anal. Chem.* **1995**, *67* (11), 1772–1784. https://doi.org/10.1021/ac00107a005.

(126) Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. Higher-Energy C-Trap Dissociation for Peptide Modification Analysis. *Nat. Methods* **2007**, *4*

(9), 709–712. https://doi.org/10.1038/nmeth1060.

(127) Jedrychowski, M. P.; Huttlin, E. L.; Haas, W.; Sowa, M. E.; Rad, R.; Gygi, S. P. Evaluation of HCD- and CID-Type Fragmentation within Their Respective Detection Platforms for Murine Phosphoproteomics. *Mol. Cell. Proteomics* **2011**, *10* (12), 1–9. https://doi.org/10.1074/mcp.M111.009910.

(128) Roepstorff, P.; Fohlman, J. Letter to the Editors - Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. *Biol. Mass Spectrom.* **1984**, *11* (11), 601–601. https://doi.org/10.1002/bms.1200111109.

(129) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and Protein Sequence Analysis by Electron Transfer Dissociation Mass Spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (26), 9528–9533. https://doi.org/10.1073/pnas.0402700101.

(130) Zhu, Z.; Su, X.; Clark, D. F.; Go, E. P.; Desaire, H. Characterizing O -Linked Glycopeptides by Electron Transfer Dissociation: Fragmentation Rules and Applications in Data Analysis. *Anal. Chem.* **2013**, *85* (17), 8403–8411. https://doi.org/10.1021/ac401814h.

(131) Hsu, H. C.; Liew, C. Y.; Huang, S. P.; Tsai, S. T.; Ni, C. K. Simple Method for de Novo Structural Determination of Underivatised Glucose Oligosaccharides. *Sci. Rep.* **2018**, *8* (1), 1–12. https://doi.org/10.1038/s41598-018-23903-4.

(132) Viseux, N.; Costello, C. E.; Domon, B. Post-Source Decay Mass Spectrometry: Optimized Calibration Procedure and Structural Characterization of Permethylated Oligosaccharides. *J. Mass Spectrom.* **1999**, *34* (4), 364–376. https://doi.org/10.1002/(SICI)1096-9888(199904)34:4<364::AID-JMS787>3.0.CO;2-P.

(133) Lapadula, A. J.; Hatcher, P. J.; Hanneman, A. J.; Ashline, D. J.; Zhang, H.; Reinhold, V. N. Congruent Strategies for Carbohydrate Sequencing. 3. OSCAR: An Algorithm for Assigning Oligosaccharide Topology from MS n Data. *Anal. Chem.* **2005**, *77* (19), 6271–6279. https://doi.org/10.1021/ac050726j.

(134) Kirwan, A.; Utratna, M.; O'Dwyer, M. E.; Joshi, L.; Kilcoyne, M. Glycosylation-Based Serum Biomarkers for Cancer Diagnostics and Prognostics. *Biomed Res. Int.* **2015**, *2015*. https://doi.org/10.1155/2015/490531.

(135) An, Y.; Cipollo, J. F. An Unbiased Approach for Analysis of Protein Glycosylation and Application to Influenza Vaccine Hemagglutinin. *Anal. Biochem.* **2011**, *415* (1), 67–80. https://doi.org/10.1016/j.ab.2011.04.018.

(136) PHB, B.-M.; KJ, P. Haemophilias A and B. *Lancet* **2003**, *361 North* (9371), 1801-1809 9p.

(137) ZOLL, P. M. et al 1956. The New England Journal of Medicine Downloaded from Nejm.Org at Karolinska Institutet University Library on April 23, 2015. For Personal Use Only. No Other Uses without Permission. Copyright © 1994 Massachusetts Medical Society. All Rights Reserved. **1956**, *297* (15), 800–802.

(138) Soucie, J. M.; Evatt, B.; Jackson, D. Occurrence of Hemophilia in the United States. The Hemophilia Surveillance System Project Investigators. *Am. J. Hematol.* **1998**, *59* (4), 288–294.

(139) Carcao, M.; Goudemand, J. INHIBITORS IN HEMOPHILIA : A PRIMER Fifth Edition. *World Fed. Hemoph.* **2018**, *November* (7), 1–24.

(140) Dicaprio, M. R. Systemic Disorders. *Orthop. Knowl. Updat. 12* **2018**, 169–170. https://doi.org/10.5005/jp/books/12784_5.

(141) Srivastava, A.; Brewer, A. K.; Mauser-Bunschoten, E. P.; Key, N. S.; Kitchen, S.; Llinas, A.; Ludlam, C. A.; Mahlangu, J. N.; Mulder, K.; Poon, M. C.; et al. Guidelines for the Management of Hemophilia. *Haemophilia* **2013**, *19* (1). https://doi.org/10.1111/j.1365-2516.2012.02909.x.

(142) Konkle, B. A.; Ebbesen, L. S.; Erhardtsen, E.; Bianco, R. P.; Lissitchkov, T.; Rusen, L.; Serban, M. A. Randomized, Prospective Clinical Trial of Recombinant Factor VIIa for Secondary Prophylaxis in Hemophilia Patients with Inhibitors. *J. Thromb. Haemost.* **2007**, *5* (9), 1904–1913. https://doi.org/10.1111/j.1538-7836.2007.02663.x.

(143) Mazurkiewicz-Pisarek, A.; Płucienniczak, G.; Ciach, T.; Plucienniczak, A. The Factor VIII Protein and Its Function. *Acta Biochim. Pol.* **2016**, *63* (1), 11–16. https://doi.org/10.18388/abp.2015_1056.

(144) O'Brien LM1, Mastri M, F. P. Regulation of Factor VIIIa by Human Activated Protein C and Protein S: Inactivation of Cofactor in the Intrinsic Factor Xase. *Blood* **2000**, *95* (5), 1714–1720.

(145) Fay, P. J.; Koshibu, K. The A2 Subunit of Factor VIIIa Modulates the Active Site of Factor IXa. *J. Biol. Chem.* **1998**, *273* (30), 19049–19054. https://doi.org/10.1074/jbc.273.30.19049.

(146) Healey, J. F.; Lubin, I. M.; Nakai, H.; Saenko, E. L.; Hoyer, L. W.; Scandella, D.; Lollar, P. Residues 484-508 Contain a Major Determinant of the Inhibitory Epitope in the A2 Domain of Human Factor VIII. *J. Biol. Chem.* **1995**, *270* (24), 14505–14509. https://doi.org/10.1074/jbc.270.24.14505.

(147) Bajaj, S. P.; Schmidt, A. E.; Mathur, A.; Padmanabhan, K.; Zhong, D.; Mastri, M.; Fay, P. J. Factor IXa:Factor VIIIa Interaction. *J. Biol. Chem.* **2001**, *276* (19), 16302–16309. https://doi.org/10.1074/jbc.M011680200.

(148) Scandella, D. Human Anti-Factor VIII Antibodies: Epitope Localization and Inhibitory Function. *Vox Sang.* **1996**, *70* (1), 9–14. https://doi.org/10.1159/000462135.

(149) Fijnvandraat, K.; Berntorp, E.; ten Cate, J. W.; Johnsson, H.; Peters, M.; Savidge, G.; Tengborn, L.; Spira, J.; Stahl, C. Recombinant, B-Domain Deleted Factor VIII (r-VIII SQ): Pharmacokinetics and Initial Safety Aspects in Hemophilia A Patients. *Thromb. Haemost.* **1997**, *77* (02), 298–302. https://doi.org/10.1055/s-0038-1655957.

(150) Ga, V.; Keyt, B.; Eaton, D.; Rodriguez, H.; Dp, O. B.; Rotblat, F.; Oppermann, H.; Wi, W.; Rn, H.; Eg, T.; et al. Structure of Human Factor VIII . 6438527. https://doi.org/10.1016/j.str.2008.03.001.

(151) Pratt, K. P.; Shen, B. W.; Takeshima, K.; Davie, E. W.; Fujikawa, K.; Stoddard, B. L. Structure of the C2 Domain of Human Factor VIII at 1.5 Å Resolution. *Nature* **1999**, *402* (6760), 439–442. https://doi.org/10.1038/46601.

(152) Meeks, S. L.; Batsuli, G. Hemophilia and Inhibitors: Current Treatment Options and Potential New Therapeutic Approaches. *Hematology* **2016**, *2016* (1), 657–662. https://doi.org/10.1182/asheducation-2016.1.657.

(153) Whelan, S. F. J.; Hofbauer, C. J.; Horling, F. M.; Allacher, P.; Wolfsegger, M. J.; Oldenburg, J.; Male, C.; Windyga, J.; Tiede, A.; Schwarz, H. P.; et al. Distinct Characteristics of Antibody Responses against Factor VIII in Healthy Individuals and in Different Cohorts of Hemophili A Patients. *Blood* **2013**, *121* (6), 1039–1048. https://doi.org/10.1182/blood-2012-07-444877.

(154) Peyvandi, F.; Garagiola, I.; Young, G. The Past and Future of Haemophilia: Diagnosis, Treatments, and Its Complications. *Lancet* **2016**, *388* (10040), 187–197. https://doi.org/10.1016/S0140-6736(15)01123-X.

(155) Shah, S.; Patel, T.; Bhatnagar, N.; Gajjar, M.; Shah, M.; Tripathi, S. "Prevalence of Inhibitors in Hemophilia Patients and Its Clinical Implications": A Study of 276 Patients in Western India. *Glob. J. Transfus. Med.* **2019**, *4* (2), 168. https://doi.org/10.4103/GJTM.GJTM_35_19.

(156) Castaman, G. The Role of Recombinant Activated Factor VII in the Haematological Management of Elective Orthopaedic Surgery in Haemophilia A Patients with Inhibitors. *Blood Transfus.* **2017**, *15* (5), 478–486. https://doi.org/10.2450/2017.0369-16.

(157) Wójcik, C.; Schymik, M. L.; Cure, E. G. Activated Prothrombin Complex Concentrate Factor VIII Inhibitor Bypassing Activity (FEIBA) for the Reversal of Warfarin-Induced Coagulopathy. *Int. J. Emerg. Med.* **2009**, *2* (4), 217–225. https://doi.org/10.1007/s12245-009-0125-8.

(158) Négrier, C.; Lienhart, A.; Numerof, R.; Stephens, D.; Wong, W. Y.; Baghaei, F.; Yee, T. T. SURgical Interventions with FEIBA (SURF): International Registry of Surgery in Haemophilia Patients with Inhibitory Antibodies. *Haemophilia* **2013**, *19* (3). https://doi.org/10.1111/hae.12080.

(159) Khoury, G. A.; Baliban, R. C.; Floudas, C. A. Proteome-Wide Post-Translational Modification Statistics: Frequency Analysis and Curation of the Swiss-Prot Database. *Sci. Rep.* **2011**, *1* (1), 90. https://doi.org/10.1038/srep00090.

(160) Etzler ME, E. V. A. C. R. E. J. F. H. S. P. B. C. H. G. *Nematoda*, 2nd editio.; Etzler ME, V. A. C. R. E. J. F. H. S. P. B. C. H. G., Ed.; Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009, 2009. https://doi.org/doi.org/10.1038/nrm3383.

(161) Münster, A. K.; Eckhardt, M.; Potvin, B.; Mühlenhoff, M.; Stanley, P.; Gerardy-Schahn, R. Mammalian Cytidine 5′-Monophosphate N-Acetylneiiraminic Acid Synthetase: A Nuclear Protein with Evolutionary Conserved Structural Motifs. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (16), 9140–9145. https://doi.org/10.1073/pnas.95.16.9140.

(162) Moremen, K. W.; Tiemeyer, M.; Nairn, A. V. Vertebrate Protein Glycosylation: Diversity, Synthesis and Function. *Nat. Rev. Mol. Cell Biol.* **2012**, *13* (7), 448–462. https://doi.org/10.1038/nrm3383.

(163) Anugraham, M.; Jacob, F.; Nixdorf, S.; Everest-Dass, A. V.; Heinzelmann-Schwarz, V.; Packer, N. H. Specific Glycosylation of Membrane Proteins in Epithelial Ovarian Cancer Cell Lines: Glycan Structures Reflect Gene Expression and DNA Methylation Status. *Mol. Cell. Proteomics* **2014**, *13* (9), 2213–2232. https://doi.org/10.1074/mcp.M113.037085.

(164) Almeida, A.; Kolarich, D. The Promise of Protein Glycosylation for Personalised Medicine. *Biochim. Biophys. Acta - Gen. Subj.* **2016**, *1860* (8), 1583–1595. https://doi.org/10.1016/j.bbagen.2016.03.012.

(165) Selman, M. H. J.; De Jong, S. E.; Soonawala, D.; Kroon, F. P.; Adegnika, A. A.; Deelder, A. M.; Hokke, C. H.; Yazdanbakhsh, M.; Wuhrer, M. Changes in Antigen-Specific IgG1 Fc N-Glycosylation upon Influenza and Tetanus Vaccination. *Mol. Cell. Proteomics* **2012**, *11* (4), 1–10. https://doi.org/10.1074/mcp.M111.014563.

(166) Tredan, O.; Galmarini, C. M.; Patel, K.; Tannock, I. F. Drug Resistance and the Solid Tumor Microenvironment. *JNCI J. Natl. Cancer Inst.* **2007**, *99* (19), 1441–1454. https://doi.org/10.1093/jnci/djm135.

(167) Melms, J. C.; Thummalapalli, R.; Shaw, K.; Ye, H.; Tsai, L.; Bhatt, R. S.; Izar, B. Alpha-Fetoprotein (AFP) as Tumor Marker in a Patient with Urothelial Cancer with Exceptional Response to Anti-PD-1 Therapy and an Escape Lesion Mimic. *J. Immunother. Cancer* **2018**, *6* (1), 4–8. https://doi.org/10.1186/s40425-018-0394-y.

(168) de Haan, N.; Reiding, K. R.; Wuhrer, M.; Boeddha, N. P.; Ekinci, E.; Driessen, G. J.; Emonts, M.; Hazelzet, J. A. Differences in IgG Fc Glycosylation Are Associated with Outcome of Pediatric Meningococcal Sepsis. *MBio* **2018**, *9* (3), 1–13. https://doi.org/10.1128/mBio.00546-18.

(169) Matsumoto, A.; Shikata, K.; Takeuchi, F.; Kojima, N.; Mizuochi, T. Autoantibody Activity of IgG Rheumatoid Factor Increases with Decreasing Levels of Galactosylation and Sialylation. *J. Biochem.* **2000**, *128* (4), 621–628. https://doi.org/10.1093/oxfordjournals.jbchem.a022794.

(170) Council, N. R. *Transforming Glycoscience*; National Academies Press: Washington, D.C., 2012. https://doi.org/10.17226/13446.

(171) Wuhrer, M.; Stam, J. C.; Van De Geijn, F. E.; Koeleman, C. A. M.; Verrips, C. T.; Dolhain, R. J. E. M.; Hokke, C. H.; Deelder, A. M. Glycosylation Profiling of Immunoglobulin G (IgG) Subclasses from Human Serum. *Proteomics* **2007**, *7* (22), 4070–4081. https://doi.org/10.1002/pmic.200700289.

(172) Parekh, R. B.; Dwek, R. A.; Sutton, B. J.; Fernandes, D. L.; Leung, A.; Stanworth, D.; Rademacher, T. W.; Mizuochi, T.; Taniguchi, T.; Matsuta, K.; et al. Association of Rheumatoid Arthritis and Primary Osteoarthritis with Changes in the Glycosylation Pattern of Total Serum IgG. *Nature* **1985**, *316* (6027), 452–457. https://doi.org/10.1038/316452a0.

(173) Peschke, B.; Keller, C. W.; Weber, P.; Quast, I.; Lünemann, J. D. Fc-Galactosylation of Human Immunoglobulin Gamma Isotypes Improves C1q Binding and Enhances Complement-Dependent Cytotoxicity. *Front. Immunol.* **2017**, *8* (31), 646. https://doi.org/10.3389/fimmu.2017.00646.

(174) Gudelj, I.; Lauc, G.; Pezer, M. Immunoglobulin G Glycosylation in Aging and Diseases. *Cell. Immunol.* **2018**, *333* (January), 65–79. https://doi.org/10.1016/j.cellimm.2018.07.009.

(175) Pučić, M.; Knežević, A.; Vidič, J.; Adamczyk, B.; Novokmet, M.; Polašek, O.; Gornik,

O.; Šupraha-Goreta, S.; Wormald, M. R.; Redžic, I.; et al. High Throughput Isolation and Glycosylation Analysis of IgG-Variability and Heritability of the IgG Glycome in Three Isolated Human Populations. *Mol. Cell. Proteomics* **2011**, *10* (10), 1–15. https://doi.org/10.1074/mcp.M111.010090.

(176) Qian, Y.; Wang, Y.; Zhang, X.; Zhou, L.; Zhang, Z.; Xu, J.; Ruan, Y.; Ren, S.; Xu, C.; Gu, J. Quantitative Analysis of Serum IgG Galactosylation Assists Differential Diagnosis of Ovarian Cancer. *J. Proteome Res.* **2013**, *12* (9), 4046–4055. https://doi.org/10.1021/pr4003992.

(177) Reiding, K. R.; Vreeker, G. C. M.; Bondt, A.; Bladergroen, M. R.; Hazes, J. M. W.; van der Burgt, Y. E. M.; Wuhrer, M.; Dolhain, R. J. E. M. Serum Protein N-Glycosylation Changes with Rheumatoid Arthritis Disease Activity during and after Pregnancy. *Front. Med.* **2018**, *4* (January). https://doi.org/10.3389/fmed.2017.00241.

(178) JEFFERIS, R.; KUMARARATNE, D. S. Selective IgG Subclass Deficiency: Quantification and Clinical Relevance. *Clin. Exp. Immunol.* **2008**, *81* (3), 357–367. https://doi.org/10.1111/j.1365-2249.1990.tb05339.x.

(179) Ferrante, A.; Beard, L. J.; Feldman, R. G. IgG Subclass Distribution of Antibodies to Bacterial and Viral Antigens. *Pediatr. Infect. Dis. J.* **1990**, *9* (8 Suppl), S16-24. https://doi.org/10.1097/00006454-199008001-00004.

(180) Vidarsson, G.; Dekkers, G.; Rispens, T. IgG Subclasses and Allotypes: From Structure to Effector Functions. *Front. Immunol.* **2014**, *5* (OCT), 1–17. https://doi.org/10.3389/fimmu.2014.00520.

(181) Nouri-Aria, K. T.; Wachholz, P. A.; Francis, J. N.; Jacobson, M. R.; Walker, S. M.; Wilcock, L. K.; Staple, S. Q.; Aalberse, R. C.; Till, S. J.; Durham, S. R. Grass Pollen Immunotherapy Induces Mucosal and Peripheral IL-10 Responses and Blocking IgG Activity. *J. Immunol.* **2004**, *172* (5), 3252–3259. https://doi.org/10.4049/jimmunol.172.5.3252.

(182) Iizuka, A.; Nagao, T. Analysis of IgG Heavy Chain Subclasses of Alloantibodies to Factor IX by Crossed Immunoelectrophoresis of Factor IX Using the Intermediate Gel Technique. *Br. J. Haematol.* **1983**, *53* (4), 687–688. https://doi.org/10.1111/j.1365-2141.1983.tb07323.x.

(183) van Helden, P. M. W.; van den Berg, H. M.; Gouw, S. C.; Kaijen, P. H. P.; Zuurveld, M. G.; Mauser-Bunschoten, E. P.; Aalberse, R. C.; Vidarsson, G.; Voorberg, J. IgG Subclasses of Anti-FVIII Antibodies during Immune Tolerance Induction in Patients with Hemophilia A. *Br. J. Haematol.* **2008**, *142* (4), 644–652.

https://doi.org/10.1111/j.1365-2141.2008.07232.x.

(184) ANDERSEN, B. R.; TERRY, W. D. Gamma G4-Globulin Antibody Causing Inhibition of Clotting Factor VIII. *Nature* **1968**, *217* (5124), 174–175. https://doi.org/10.1038/217174a0.

(185) Nagae, M.; Yamaguchi, Y. Function and 3D Structure of the N-Glycans on Glycoproteins. *Int. J. Mol. Sci.* **2012**, *13* (7), 8398–8429. https://doi.org/10.3390/ijms13078398.

(186) Axford, J. S. Glycosylation and Rheumatic Disease. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1999**, *1455* (2–3), 219–229. https://doi.org/10.1016/S0925-4439(99)00057-5.

(187) Yoshida, Y.; Furukawa, J. I.; Naito, S.; Higashino, K.; Numata, Y.; Shinohara, Y. Quantitative Analysis of Total Serum Glycome in Human and Mouse. *Proteomics* **2016**, *16* (21), 2747–2758. https://doi.org/10.1002/pmic.201500550.

(188) Kim, J.; Coffey, D. M.; Creighton, C. J.; Yu, Z.; Hawkins, S. M.; Matzuk, M. M. High-Grade Serous Ovarian Cancer Arises from Fallopian Tube in a Mouse Model. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (10), 3921–3926. https://doi.org/10.1073/pnas.1117135109.

(189) Juan, H.; Chen, J.-H.; Hsu, W.; Huang, S.; Chen, S.-T.; Yi-Chung Lin, J.; Chang, Y.; Chiang, C.; Wen, L.; Chan, D.; et al. Identification of Tumor-Associated Plasma Biomarkers Using Proteomic Techniques: From Mouse to Human. *Proteomics* **2004**, *4* (9), 2766–2775. https://doi.org/10.1002/pmic.200400785.

(190) Mestas, J.; Hughes, C. C. W. Of Mice and Not Men: Differences between Mouse and Human Immunology. *J. Immunol.* **2004**, *172* (5), 2731–2738. https://doi.org/10.4049/jimmunol.172.5.2731.

(191) Schwab, I.; Lux, A.; Nimmerjahn, F. Pathways Responsible for Human Autoantibody and Therapeutic Intravenous IgG Activity in Humanized Mice. *Cell Rep.* **2015**, *13* (3), 610–620. https://doi.org/10.1016/j.celrep.2015.09.013.

(192) Uh, H.-W.; Klarić, L.; Ugrina, I.; Lauc, G.; Smilde, A. K.; Houwing-Duistermaat, J. J. Choosing Proper Normalization Is Essential for Discovery of Sparse Glycan Biomarkers. *Mol. Omi.* **2020**. https://doi.org/10.1039/c9mo00174c.

(193) Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal Sample Preparation Method for Proteome Analysis. *Nat. Methods* **2009**, *6* (5), 359–362. https://doi.org/10.1038/nmeth.1322.

(194) Mueller, C.; Flotte, T. R. Clinical Gene Therapy Using Recombinant Adeno-

Associated Virus Vectors. *Gene Ther.* **2008**, *15* (11), 858–863. https://doi.org/10.1038/gt.2008.68.

(195) Daya, S.; Berns, K. I. Gene Therapy Using Adeno-Associated Virus Vectors. *Clin. Microbiol. Rev.* **2008**, *21* (4), 583–593. https://doi.org/10.1128/CMR.00008-08.

(196) Pillay, S.; Zou, W.; Cheng, F.; Puschnik, A. S.; Meyer, N. L.; Ganaie, S. S.; Deng, X.; Wosen, J. E.; Davulcu, O.; Yan, Z.; et al. AAV Serotypes Have Distinctive Interactions with Domains of the Cellular Receptor AAVR. *J. Virol.* **2017**, *91* (18), JVI.00391-17. https://doi.org/10.1128/JVI.00391-17.

(197) Lauriel F. Earley, a John M. Powers, b Kei Adachi, b Joshua T. Baumgart, b Nancy L. Meyer, c Qing Xie, c Michael S. Chapman, c Hiroyuki NakaiLauriel F. Earley, a John M. Powers, b Kei Adachi, b Joshua T. Baumgart, b Nancy L. Meyer, c Qing Xie, c Michael S, c H. N. Crossm Activating Protein Is Not an Essential Requirement for Capsid Assembly Of. **2017**, *91* (3), 1–21. https://doi.org/DOI: 10.1128/JVI.01980-16.

(198) Vance, M. A.; Mitchell, A.; Samulski, R. J. AAV Biology, Infectivity and Therapeutic Use from Bench to Clinic. In *Gene Therapy - Principles and Challenges*; InTech, 2015. https://doi.org/10.5772/61988.

(199) Sonntag, F.; Kother, K.; Schmidt, K.; Weghofer, M.; Raupp, C.; Nieto, K.; Kuck, A.; Gerlach, B.; Bottcher, B.; Muller, O. J.; et al. The Assembly-Activating Protein Promotes Capsid Assembly of Different Adeno-Associated Virus Serotypes. *J. Virol.* **2011**, *85* (23), 12686–12697. https://doi.org/10.1128/JVI.05359-11.

(200) Sonntag, F.; Schmidt, K.; Kleinschmidt, J. A. A Viral Assembly Factor Promotes AAV2 Capsid Formation in the Nucleolus. *Proc. Natl. Acad. Sci.* **2010**, *107* (22), 10220–10225. https://doi.org/10.1073/pnas.1001673107.

(201) Trempe, J. P.; Carter, B. J. Alternate MRNA Splicing Is Required for Synthesis of Adeno-Associated Virus VP1 Capsid Protein. **1988**, *62* (9), 3356–3363.

(202) Nam, H.-J.; Lane, M. D.; Padron, E.; Gurda, B.; McKenna, R.; Kohlbrenner, E.; Aslanidi, G.; Byrne, B.; Muzyczka, N.; Zolotukhin, S.; et al. Structure of Adeno-Associated Virus Serotype 8, a Gene Therapy Vector. *J. Virol.* **2007**, *81* (22), 12260–12271. https://doi.org/10.1128/JVI.01304-07.

(203) Gerlach, B.; Kleinschmidt, J. A.; Böttcher, B. Conformational Changes in Adeno-Associated Virus Type 1 Induced by Genome Packaging. *J. Mol. Biol.* **2011**, *409* (3), 427–438. https://doi.org/10.1016/j.jmb.2011.03.062.

(204) Padron, E.; Bowman, V.; Kaludov, N.; Levy, H.; Nick, P.; Mckenna, R.; Muzyczka, N.; Chiorini, J. A.; Baker, T. S.; Agbandje-mckenna, M.; et al. Structure of Adeno-

Associated Virus Type 4. *J. Virol.* **2005**, *79* (8), 5048–5058. https://doi.org/10.1128/JVI.79.8.5047.

(205) Xie, Q.; Bu, W.; Bhatia, S.; Hare, J.; Somasundaram, T.; Azzi, A.; Chapman, M. S. The Atomic Structure of Adeno-Associated Virus (AAV-2), a Vector for Human Gene Therapy. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (16), 10405–10410. https://doi.org/10.1073/pnas.162250899.

(206) Kronenberg, S.; Böttcher, B.; Der, C. W. Von; Bleker, S.; Kleinschmidt, J. a; Bo, B.; Lieth, C. W. Von Der. A Conformational Change in the Adeno-Associated Virus Type 2 Capsid Leads to the Exposure of Hidden VP1 N Termini. *J. Virol.* **2005**, *79* (9), 5296–5303. https://doi.org/10.1128/JVI.79.9.5296.

(207) Drouin, L. M.; Agbandje-McKenna, M. Adeno-Associated Virus Structural Biology as a Tool in Vector Development. *Future Virol.* **2013**, *8* (12), 1183–1199. https://doi.org/10.2217/fvl.13.112.

(208) Gurda, B. L.; DiMattia, M. A.; Miller, E. B.; Bennett, A.; McKenna, R.; Weichert, W. S.; Nelson, C. D.; Chen, W. -j.; Muzyczka, N.; Olson, N. H.; et al. Capsid Antibodies to Different Adeno-Associated Virus Serotypes Bind Common Regions. *J. Virol.* **2013**, *87* (16), 9111–9124. https://doi.org/10.1128/JVI.00622-13.

(209) Gao, G.-P.; Alvira, M. R.; Wang, L.; Calcedo, R.; Johnston, J.; Wilson, J. M. Novel Adeno-Associated Viruses from Rhesus Monkeys as Vectors for Human Gene Therapy. *Proc. Natl. Acad. Sci.* **2002**, *99* (18), 11854–11859. https://doi.org/10.1073/pnas.182412299.

(210) Burger, C.; Gorbatyuk, O. S.; Velardo, M. J.; Peden, C. S.; Williams, P.; Zolotukhin, S.; Reier, P. J.; Mandel, R. J.; Muzyczka, N. Recombinant AAV Viral Vectors Pseudotyped with Viral Capsids from Serotypes 1, 2, and 5 Display Differential Efficiency and Cell Tropism after Delivery to Different Regions of the Central Nervous System. *Mol. Ther.* **2004**, *10* (2), 302–317. https://doi.org/10.1016/j.ymthe.2004.05.024.

(211) Gao, G.; Alvira, M. R.; Somanathan, S.; Lu, Y.; Vandenberghe, L. H.; Rux, J. J.; Calcedo, R.; Sanmiguel, J.; Abbas, Z.; Wilson, J. M. Adeno-Associated Viruses Undergo Substantial Evolution in Primates during Natural Infections. *Proc. Natl. Acad. Sci.* **2003**, *100* (10), 6081–6086. https://doi.org/10.1073/pnas.0937739100.

(212) Gao, G.; Vandenberghe, L. H.; Alvira, M. R.; Lu, Y.; Calcedo, R.; Zhou, X.; Wilson, J. M. Clades of Adeno-Associated Viruses Are Widely Disseminated in Human Tissues. *J. Virol.* **2004**, *78* (12), 6381–6388. https://doi.org/10.1128/JVI.78.12.6381-

6388.2004.

(213) Rayaprolu, V.; Kruse, S.; Kant, R.; Venkatakrishnan, B.; Movahed, N.; Brooke, D.; Lins, B.; Bennett, A.; Potter, T.; McKenna, R.; et al. Comparative Analysis of Adeno-Associated Virus Capsid Stability and Dynamics. *J. Virol.* **2013**, *87* (24), 13150–13160. https://doi.org/10.1128/JVI.01415-13.

(214) Kattenhorn, L. M.; Tipper, C. H.; Stoica, L.; Geraghty, D. S.; Wright, T. L.; Clark, K. R.; Wadsworth, S. C. Adeno-Associated Virus Gene Therapy for Liver Disease. **2016**, *27* (12), 947–961. https://doi.org/10.1089/hum.2016.160.

(215) Shen, X.; Storm, T.; Kay, M. A. Characterization of the Relationship of AAV Capsid Domain Swapping to Liver Transduction Efficiency. *Mol. Ther.* **2007**, *15* (11), 1955–1962. https://doi.org/10.1038/sj.mt.6300293.

(216) Mingozzi, F.; Schüttrumpf, J.; Arruda, V. R.; Liu, Y.; Liu, Y.-L.; High, K. A.; Xiao, W.; Herzog, R. W. Improved Hepatic Gene Transfer by Using an Adeno-Associated Virus Serotype 5 Vector. *J. Virol.* **2002**, *76* (20), 10497–10502. https://doi.org/10.1128/JVI.76.20.10497.

(217) Bartel, M. A.; Weinstein, J. R.; Schaffer, D. V. Directed Evolution of Novel Adeno-Associated Viruses for Therapeutic Gene Delivery. *Gene Ther.* **2012**, *19* (6), 694–700. https://doi.org/10.1038/gt.2012.20.

(218) Bhatt, D. L.; Bode, C.; Burton, P.; Ph, D.; Cohen, M.; Cook-bruns, N.; Fox, K. A. A.; Ch, B.; Goto, S.; Murphy, S. A.; et al. New England Journal. **2012**, 9–19. https://doi.org/10.1056/NEJMoa1201161.

(219) Hastie, E.; Samulski, R. J. Adeno-Associated Virus at 50: A Golden Anniversary of Discovery, Research, and Gene Therapy Success—A Personal Perspective. *Hum. Gene Ther.* **2015**, *26* (5), 257–265. https://doi.org/10.1089/hum.2015.025.

(220) Manuscript, A.; Nanobiomaterials, B. NIH Public Access. **2013**, *6* (8), 1–16. https://doi.org/10.1021/nn300902w.Release.

(221) Nienhuis, A. W.; Nathwani, A. C.; Davidoff, A. M. Gene Therapy for Hemophilia. *Mol. Ther.* **2017**, *25* (5), 1163–1167. https://doi.org/10.1016/j.ymthe.2017.03.033.

(222) Chir, B.; Chowdary, P.; Riddell, A.; Sc, B.; Pie, A. J.; Davidoff, A. M. Adenovirus-Associated Virus Vector-Mediated Gene Transfer in Hemophilia B. *N. Engl. J. Med.* **2012**, *365* (25), 2357–2365. https://doi.org/10.1056/NEJMoa1108046.Adenovirus-Associated.

(223) Nathwani, A. C. Long-Term Safety and Efficacy of Factor IX Gene Therapy in Hemophilia B. **2015**, *371* (21), 1994–2004.

https://doi.org/10.1056/NEJMoa1407309.Long-Term.

(224) Lebherz, C.; Gao, G.; Louboutin, J.-P.; Millar, J.; Rader, D.; Wilson, J. M. Gene Therapy with Novel Adeno-Associated Virus Vectors Substantially Diminishes Atherosclerosis in a Murine Model of Familial Hypercholesterolemia. *J. Gene Med.* **2004**, *6* (6), 663–672. https://doi.org/10.1002/jgm.554.

(225) Sun, B.; Zhang, H.; Franco, L. M.; Young, S. P.; Schneider, A.; Bird, A.; Amalfitano, A.; Chen, Y. T.; Koeberl, D. D. Efficacy of an Adeno-Associated Virus 8-Pseudotyped Vector in Glycogen Storage Disease Type II. *Mol. Ther.* **2005**, *11* (1), 57–65. https://doi.org/10.1016/j.ymthe.2004.10.004.

(226) Wang, Z.; Zhu, T.; Qiao, C.; Zhou, L.; Wang, B.; Zhang, J.; Chen, C.; Li, J.; Xiao, X. Adeno-Associated Virus Serotype 8 Efficiently Delivers Genes to Muscle and Heart. *Nat. Biotechnol.* **2005**, *23* (3), 321–328. https://doi.org/10.1038/nbt1073.

(227) Cauet, G.; Strub, J.; Leize, E.; Wagner, E.; Dorsselaer, A. Van; Lusky, M. Identification of the Glycosylation Site of the Adenovirus Type 5 Fiber Protein. **2005**, 5453–5460. https://doi.org/10.1021/bi047702b.

(228) Zhong, L.; Li, B.; Mah, C. S.; Govindasamy, L.; Agbandje-McKenna, M.; Cooper, M.; Herzog, R. W.; Zolotukhin, I.; Warrington, K. H.; Weigel-Van Aken, K. A.; et al. Next Generation of Adeno-Associated Virus 2 Vectors: Point Mutations in Tyrosines Lead to High-Efficiency Transduction at Lower Doses. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (22), 7827–7832. https://doi.org/10.1073/pnas.0802866105.

(229) Liu, Y.; Kim, Y. J.; Ji, M.; Fang, J.; Siriwon, N.; Zhang, L. I.; Wang, P. Enhancing Gene Delivery of Adeno-Associated Viruses by Cell-Permeable Peptides. *Mol. Ther. - Methods Clin. Dev.* **2014**, *1* (October 2013), 1–13. https://doi.org/10.1038/mtm.2013.12.

(230) Jin, X.; Liu, L.; Nass, S.; O'Riordan, C.; Pastor, E.; Zhang, X. K. Direct Liquid Chromatography/Mass Spectrometry Analysis for Complete Characterization of Recombinant Adeno-Associated Virus Capsid Proteins. *Hum. Gene Ther. Methods* **2017**, *X* (X), hgtb.2016.178. https://doi.org/10.1089/hgtb.2016.178.

(231) Ma, C.; Zhao, X.; Han, H.; Tong, W.; Zhang, Q.; Qin, P.; Chang, C.; Peng, B.; Ying, W.; Qian, X. N-Linked Glycoproteome Profiling of Human Serum Using Tandem Enrichment and Multiple Fraction Concatenation. *Electrophoresis* **2013**, *34* (16), 2440–2450. https://doi.org/10.1002/elps.201200662.

(232) Gashash, E. A.; Aloor, A.; Li, D.; Zhu, H.; Xu, X. Q.; Xiao, C.; Zhang, J.; Parameswaran, A.; Song, J.; Ma, C.; et al. An Insight into Glyco-Microheterogeneity

of Plasma von Willebrand Factor by Mass Spectrometry. *J. Proteome Res.* **2017**, *16* (9), 3348–3362. https://doi.org/10.1021/acs.jproteome.7b00359.

(233) Le-heng Wang1, 3, 4, De-Quan Li1, 3, 4, Yan Fu1, 3, Hai-Peng Wang1, 3, 4, J.-F. Z.; Zuo-Fei Yuan1, 3, 4, Rui-Xiang Sun1, 3, Rong Zeng2, Si-Min He1, 3* and Wen Gao1, 3*. PFind 2.0: A Software Package for Peptide and Protein Identification via Tandem Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2007**, *21* (21), 2985–2991. https://doi.org/DOI: 10.1002/rcm.3173 pFind.

(234) Wilkins, M. R.; Lindskog, I.; Gasteiger, E.; Bairoch, A.; Sanchez, J. C.; Hochstrasser, D. F.; Appel, R. D. Detailed Peptide Characterization Using PEPTIDEMASS - A World-Wide-Web-Accessible Tool. *Electrophoresis* **1997**, *18* (3–4), 403–408. https://doi.org/10.1002/elps.1150180314.

(235) Duncan, M.; Fung, K.; Wang, H.; Yen, C.; Cios, K. Identification of Contaminants in Proteomics Mass Spectrometry Data. *Proc. 2003 IEEE Bioinforma. Conf. CSB 2003* **2003**, 409–410. https://doi.org/10.1109/CSB.2003.1227348.

(236) Cottrell, J. S. Protein Identification Using MS/MS Data. *J. Proteomics* **2011**, *74* (10), 1842–1851. https://doi.org/10.1016/j.jprot.2011.05.014.

(237) Murray, S.; Nilsson, C. L.; Hare, J. T.; Emmett, M. R.; Korostelev, A.; Ongley, H.; Marshall, A. G.; Chapman, M. S. Characterization of the Capsid Protein Glycosylation of Adeno-Associated Virus Type 2 by High-Resolution Mass Spectrometry. *J. Virol.* **2006**, *80* (12), 6171–6176. https://doi.org/10.1128/JVI.02417-05.

(238) Lei, Y.; Yu, H.; Dong, Y.; Yang, J.; Ye, W.; Wang, Y.; Chen, W.; Jia, Z.; Xu, Z.; Li, Z.; et al. Characterization of N-Glycan Structures on the Surface of Mature Dengue 2 Virus Derived from Insect Cells. *PLoS One* **2015**, *10* (7). https://doi.org/10.1371/journal.pone.0132122.

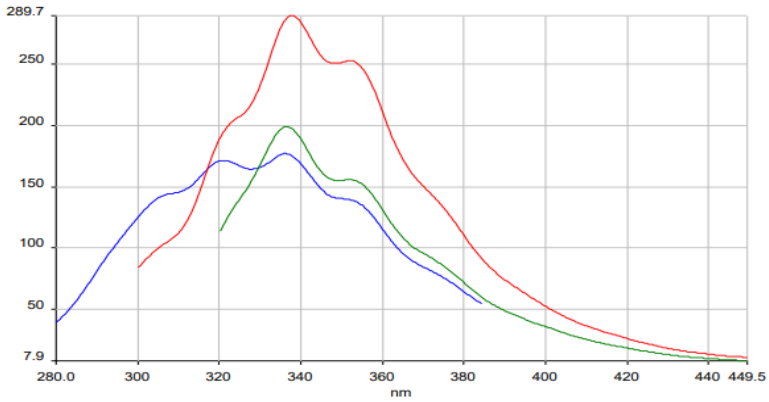
(239) Akache, B.; Grimm, D.; Pandey, K.; Yant, S. R.; Xu, H.; Kay, M. A. The 37/67-Kilodalton Laminin Receptor Is a Receptor for Adeno-Associated Virus Serotypes 8, 2, 3, and 9. *J. Virol.* **2006**, *80* (19), 9831–9836. https://doi.org/10.1128/JVI.00878-06.

(240) Miura, Y.; Endo, T. Glycomics and Glycoproteomics Focused on Aging and Age-Related Diseases - Glycans as a Potential Biomarker for Physiological Alterations. *Biochim. Biophys. Acta - Gen. Subj.* **2016**, *1860* (8), 1608–1614. https://doi.org/10.1016/j.bbagen.2016.01.013.

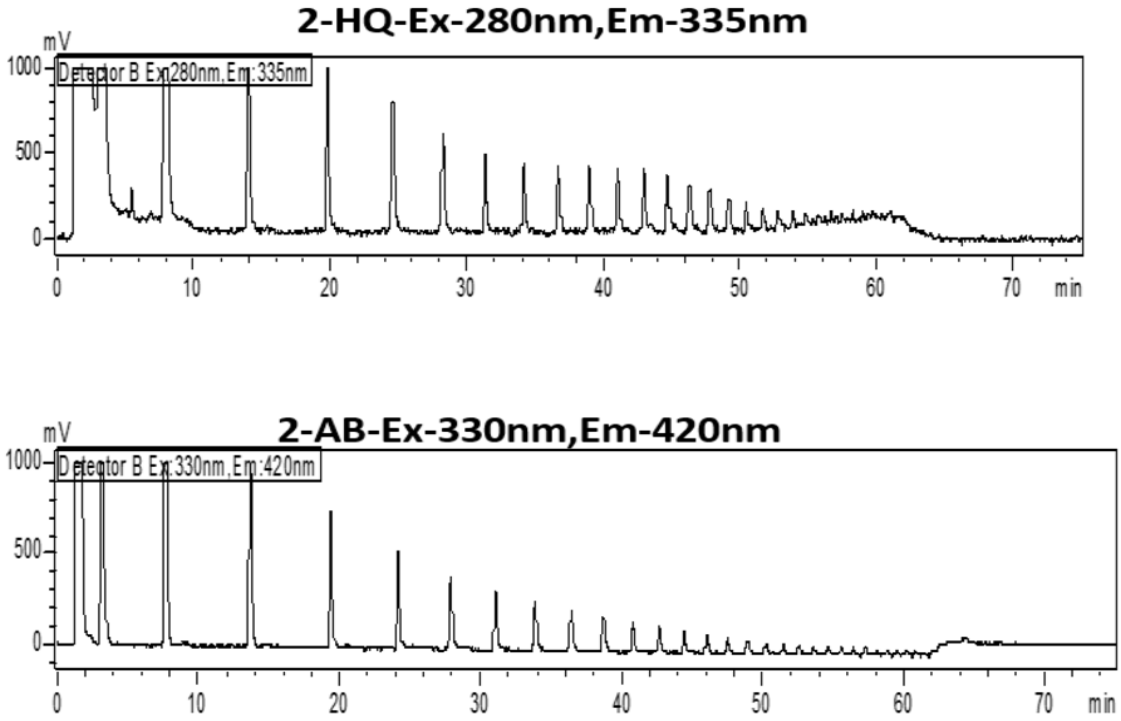
(241) Han, L.; Costello, C. E. Mass Spectrometry of Glycans. *Biochem.* **2013**, *78* (7), 710–720. https://doi.org/10.1134/S0006297913070031.

APPENDICES

Appendix A

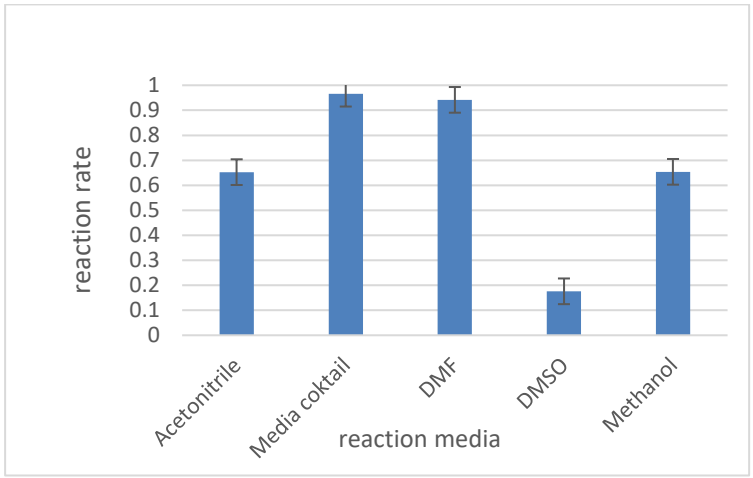


Appendix A F.1: The fluorescent property of 2HQ labeled DP5 λ em=335nm was determined at λex=280nm using fluorometry.

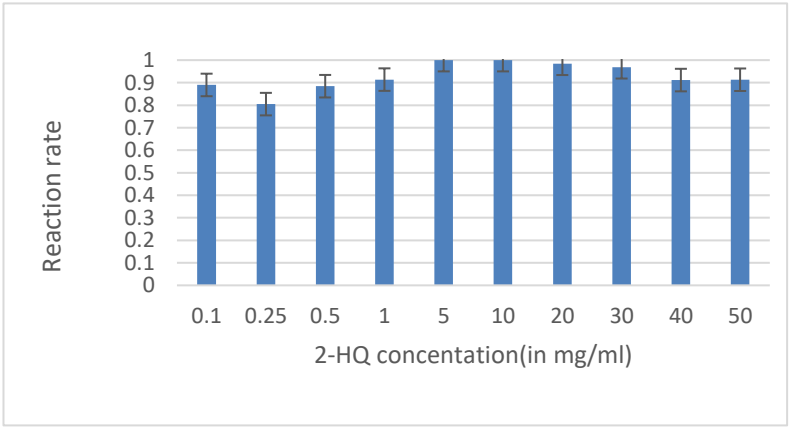


Appendix A F.2: The fluorescent property of 2HQ
Labeled dextrin separated on Amide column using the FLD detector with excitation wave length λex=280nm and emission wavelength λ em=335nm compared with conventional 2AB labeled dextrin separated with the same method but FD parameter was set at λex=330nm and λem was set at λ em=420 n

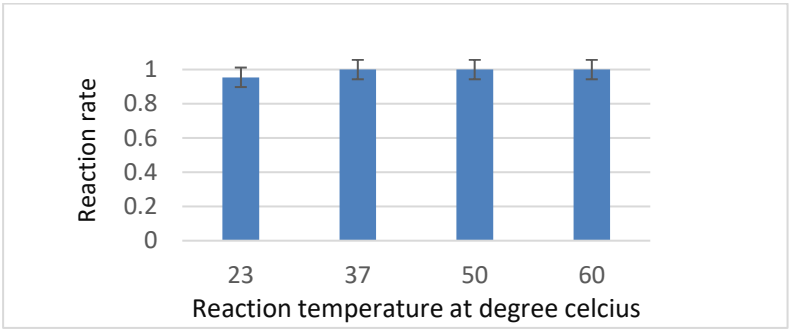
A)



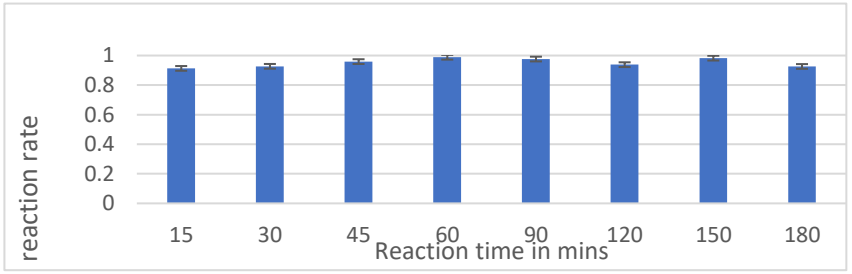
B)



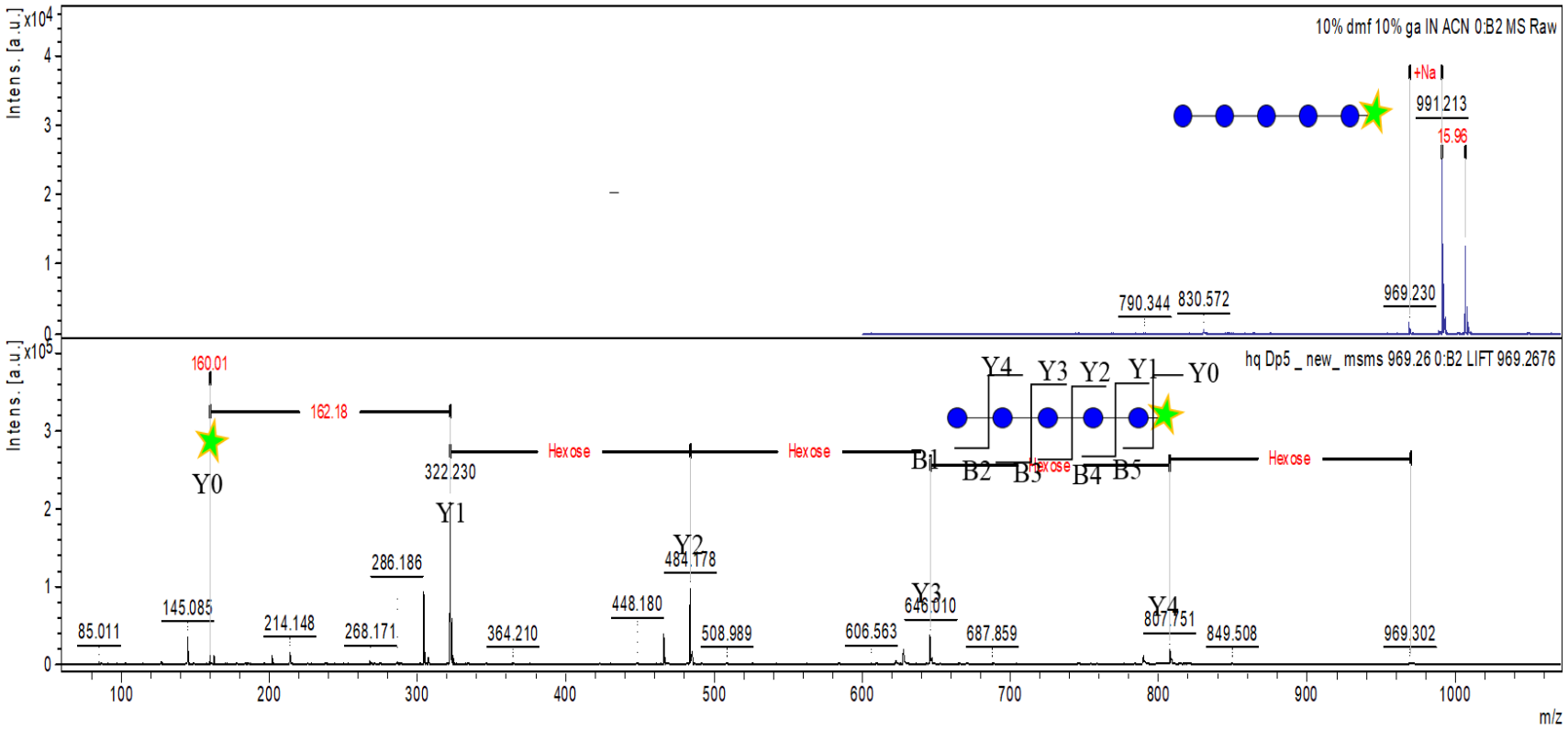
C)



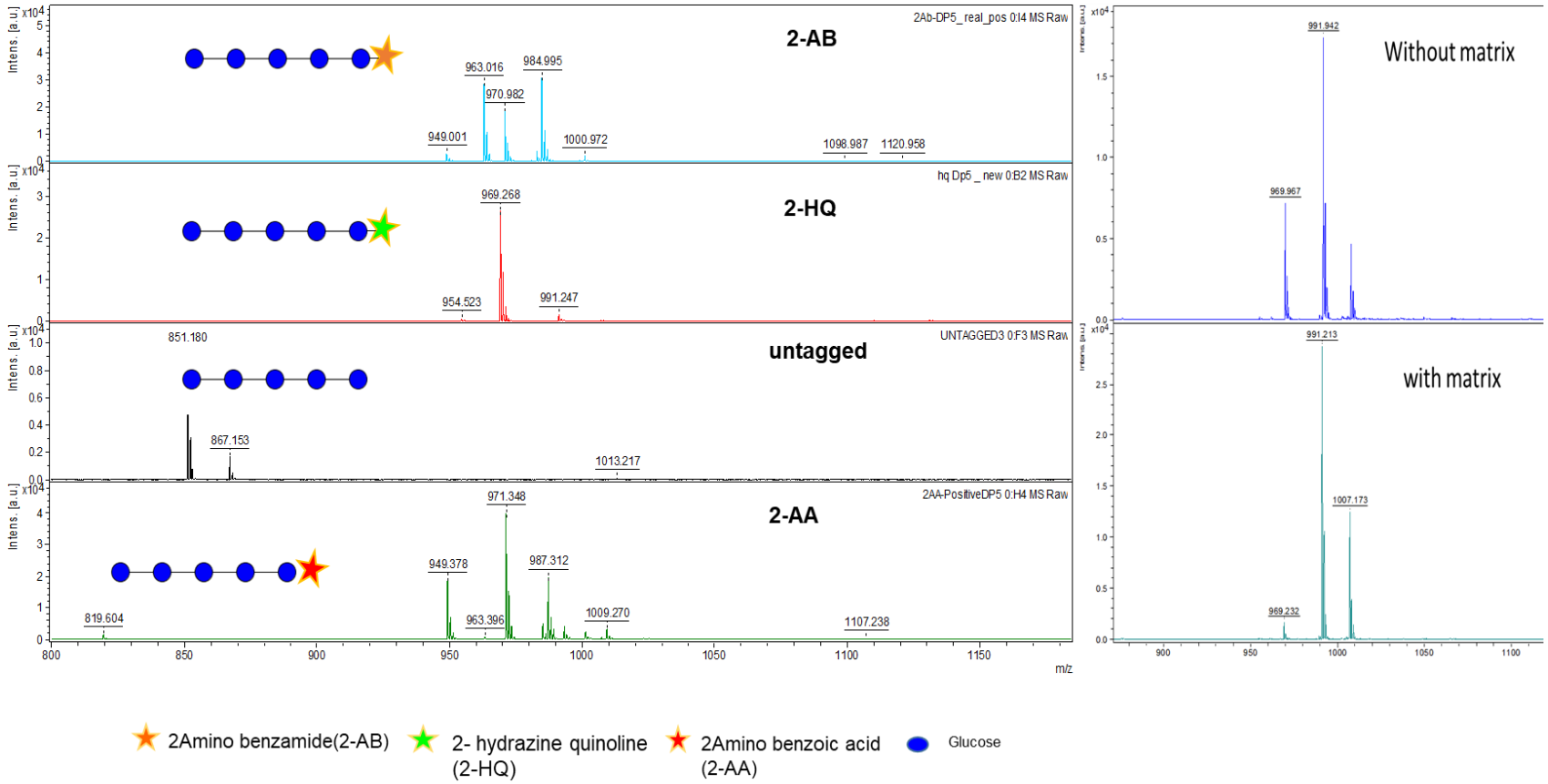
D



Appendix A F.3: The optimization parameters with triplicate analysis.
A)Reaction mediaAt varying reaction temperature) B) conversion time C) reaction media and D) HQ concentration



Appendix A F.4: The MALDI-MS profile of 30picomoles of DP5 and MS/MS profile with 2-HQ.



Appendix A F.5: The MALDI-MS profile comparison with reductive amination,2HQ labelling and untagged sugars acquired in MALDI-MS .

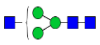
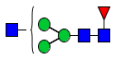
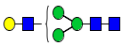
The results showed that using the same method and their intensities were compared.it showed that 2HQ can give similar ionization with reductive amination and 3 times more intensity than the untagged glycan.

Appendix B

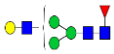
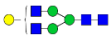
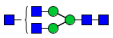
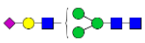
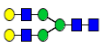
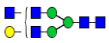
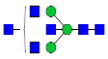
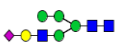
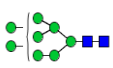
Appendix B Table 1 Internal calibration mass list used for MALDI-TOF-MS

No.	Peak label(Na+)	Mass signal (m/z)	Tolerance (ppm)	calibrant
1	33000	1416.7094	50	✓
2	34001	1835.924	50	✓
3	44001	2040.0247	50	✓
4	54001	2244.1245	50	✓
5	54101	2605.2981	50	✓
6	54201	2966.4718	50	✓

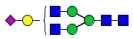
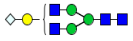
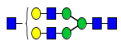
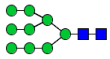
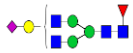
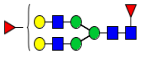
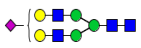
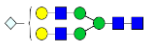
Appendix B Table 2 The **g**lycoforms detected from human plasma

No.	Structure	Gly ID	Theoretical m/z (M+Na+)	Experimental m/z (M+Na+)	mass error (ppm)	Main-Type	Subtype
1		3 3 0 0 0	1416.709	1416.682	-19.34	Hybrid	Non- Fucose/agalactosylated/
2		5 2 0 0 0	1578.783	1578.768	-9.25	High mannose	High mannose
3		3 3 0 0 1	1590.799	1590.784	-9.18	Hybrid	Core-fucose/agalactosylated
4		4 3 0 0 0	1620.809	1620.798	-6.84	Hybrid	Non- Fucose/galactosylated
5		3 4 0 0 0	1661.836	1661.832	-2.23	Complex	Non- Fucose/agalactosylated
6		3 3 0 0 2	1764.888	1763.867	-578.72	complex	Multifucosylated/galactosylated
7		3 3 1 0 0	1777.883	1777.822	-34.31	complex	Non- Fucose/Mono sialylated
8		6 2 0 0 0	1783.882	1783.876	-3.59	High mannose	High mannose

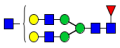
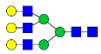
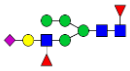
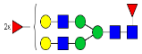
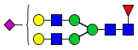
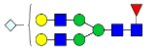
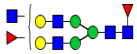
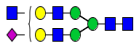
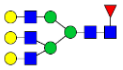
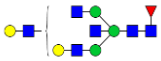
…. Continues

No.	Structure	Gly ID	Theoretical m/z (M+Na+)	Experimental m/z (M+Na+)	mass error (ppm)	Main-Type	Subtype
9		4 3 0 0 1	1794.8984	1793.879	-568.27	Hybrid	Core-Fucose/galactosylated
10		5 3 0 0 0	1824.9089	1824.893	-8.71	Hybrid	Non- Fucose/galactosylated
11		3 4 0 0 1	1835.9249	1835.933	4.41	Complex	Core-Fucose/agalactosylated
12		4 4 0 0 0	1865.9355	1865.939	1.875	Complex	Non- Fucose/galactosylated
13		3 5 0 0 0	1906.962	1906.961	-0.52	Complex	Non- Fucose/agalactosylated
14		4 3 0 0 2	1968.9876	1967.97 30	-515.56	Complex	Multifucosylated/galactosylated
15		4 3 1 0 0	1981.9828	1981.974	-4.44	Hybrid	Non-fucosylated/Mono sialic
16		4 4 0 0 1	2040.0247	2040.023	-0.83	Complex	Core-Fucose/galactosylated
17		5 4 0 0 0	2070.0352	2070.036	0.39	Complex	Non- Fucose/galactosylated
18		3 5 0 0 1	2081.0521	2081.06	3.8	Complex	Core-Fucose/agalactosylated
19		4 5 0 0 0	2111.0618	2111.059	-1.33	Complex	Non- Fucose/galactosylated
20		3 6 0 0 0	2154.06	2154.062	0.93	Complex	Non- Fucose/agalactosylated
21		5 3 0 0 2	2173.0873	2172.066	-470.2	Hybrid	Multifucosylated/galactosylated
22		5 3 1 0 0	2186.0826	2186.068	-6.68	Hybrid	Non-fucosylated/Mono sialylated
23		8 2 0 0 0	2192.0819	2192.066	-7.25	High mannose	High mannose

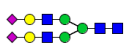
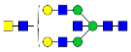
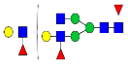
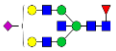
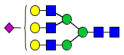
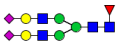
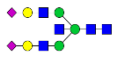
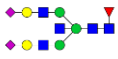
…. continues

No.	Structure	Gly ID	Theoretical m/z (M+Na+)	Experimental m/z (M+Na+)	mass error (ppm)	Main-Type	Subtype
24		4 4 0 0 2	2214.1139	2213.084	-465.37	Complex	Multifucosylated/galactosylated
25		4 4 1 0 0	2227.1091	2227.094	-6.78	Complex	Non-fucosylated/Mono sial
26		5 4 0 0 1	2244.1245	2244.115	-4.23	Complex	Core-fucose/galactosylated
27		4 4 0 1 0	2257.1197	2257.109	-4.74	Complex	Non fucosylated/Mono sial/NGNA
28		4 5 0 0 1	2285.151	2285.143	-3.5	Complex	Core-fucosylated/galactosylated
29		5 5 0 0 0	2315.1616	2315.135	-11.49	Complex	Non-fucosylated/galactosylated
30		6 3 1 0 0	2390.1824	2390.163	-8.12	Hybrid	Non fucosylated/Mono sial
31		9 2 0 0 0	2396.1824	2396.189	3.05	High mannose	High mannose
32		4 4 1 0 1	2401.1983	2401.149	-20.53	Complex	Core-fucosylated/mono sial
33		5 4 0 0 2	2418.2137	2417.17	-431.79	Complex	Multifucosylated/galactosylated
34		5 4 1 0 0	2431.2089	2431.215	2.51	Complex	Non-fucosylated/mono sial/galactosylated
35		4 0 1 0	2461.2195	2461.221	0.61	Complex	Non-fucosylated/mono sial/NGNA/galactosylated
36		5 1 0 0	2472.2355	2472.246	4.25	Complex	Non-fucosylated/mono sialylated

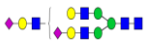
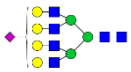
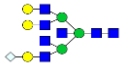
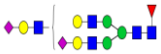
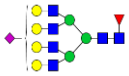
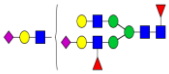
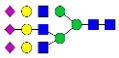
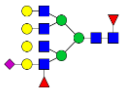
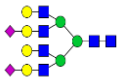
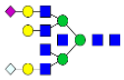
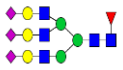
…. continues

No.	Structure	Gly ID	Theoretical m/z (M+Na+)	Experimental m/z (M+Na+)	mass error (ppm)	Main-Type	Subtype
37		5 5 0 0 1	2489.2508	2489.252	0.48	Complex	Core-fucosylated/galactosylaated
38		6 5 0 0 0	2519.2613	2519.269	3.06	Complex	Non-fucosylated/galactosylated
39		5 3 1 0 2	2534.261	2533.242	-402.25	Hybrid	Multifucosylated/monosial
40		5 3 2 0 0	2547.2563	2547.242	-5.61	Hybrid	Non-fucosylated/disial
41		5 4 0 0 3	2592.3029	2591.263	-401.31	Complex	Multifucosylated/galactosylated
42		5 4 1 0 1	2605.2981	2605.282	-6.18	Complex	Core-fucosylated/monosial/gal
43		5 4 0 1 1	2635.3087	2635.299	-3.68	Complex	Core-fucosylated/monosial/gal/NGNA
44		4 5 1 0 1	2646.3246	2646.28	-16.85	Complex	Core-fucosylated/monosial
45		5 5 0 0 2	2663.314	2663.306	-3	Complex	Multifucosylated/agalactosylated
46		5 5 1 0 0	2676.3352	2676.309	-9.79	Complex	Non-fucosylated/monosial
47		6 5 0 0 1	2693.3506	2693.31	-15.07	Complex	Core-fucosylated/galactosylated
48		5 6 0 0 1	2734.3117	2736.305	728.46	Complex	Core-fucosylated/galactosylated
49		6 3 2 0 0	2751.3157	2750.326	-359.85	Hybrid	Non-fucosylated/disialylated
50		4 4 2 0 1	2762.372	2762.308	-23.17	Complex	Core-fucosylated/disial

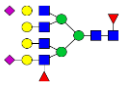
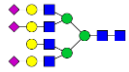
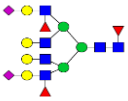
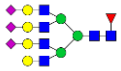
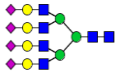
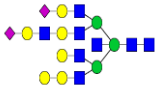
…. continues

No.	Structure	Gly ID	Theoretical m/z (M+Na+)	Experimental m/z (M+Na+)	mass error (ppm)	Main-Type	Subtype
51		4 7 0 0 1	2775.4056	2776.32	329.36	Complex	Core-fucosylated/galactosylated
52		5 4 2 0 0	2792.3862	2792.385	-0.43	Complex	Non-fucosylated/disialylated
53		5 7 0 0 0	2805.4142	2806.393	348.78	Complex	Non-fucosylated/galactosylated
54		5 4 1 1 0	2822.3942	2822.427	11.62	Complex	Non-fucosylated/disial/NGNA
55		5 5 0 0 3	2837.4232	2836.435	-348.4	Complex	Multifucosylated/galactosylated
56		5 5 1 0 1	2850.4244	2850.457	11.44	Complex	Core-fucosylated/monosial
57		6 5 1 0 0	2880.435	2880.456	7.29	Complex	Non-fucosylaated/monosial
58		5 4 2 0 1	2966.4718	2966.485	4.45	Complex	Core-fucosylated/ disial
59		5 4 1 1 1	2996.4824	2996.5	5.87	Complex	Core-fucosylated/ disial/NGNA
60		5 5 1 0 2	3024.5137	3023.52	-328.66	Complex	Multifucosylated/Monosial
61		5 5 2 0 0	3037.5089	3037.477	-10.5	Complex	Non fucosylaated/disial
62		6 5 1 0 1	3054.5242	3054.533	2.88	Complex	Core-fucosylated/monosial
63		5 5 2 0 1	3211.5981	3211.594	-1.28	Complex	Core-fucosylated/disial
64		5 8 0 0 1	3224.6297	3226.555	596.7	Complex	Core- fucosylated/ galactosylated

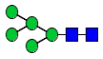
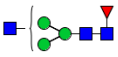
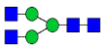
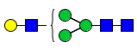
…. continues

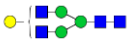
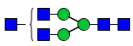
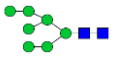
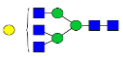
No.	Structure	Gly ID	Theoretical m/z (M+Na+)	Experimental m/z (M+Na+)	mass error (ppm)	Main-Type	Subtype
65		6 5 2 0 0	3241.6081	3241.607	-0.34	Complex	Non-fucosylation/disial/galactosylated
66		7 6 1 0 0	3329.6611	3329.633	-8.68	Complex	Non-fucosylated/monosial//galactosylated
67		6 7 0 1 0	3400.6982	3400.662	-10.64	Complex	Non-fucosylated/monosial/NGNA/galactosylated
68		6 5 2 0 1	3415.6979	3415.691	-2.02	Complex	Core-fucosylated/disial/galactosylated
69		7 6 1 0 1	3503.7503	3503.766	4.48	Complex	Core-fucosylated/monosial/galactosylated
70		6 5 2 0 2	3589.7871	3588.719	-297.63	Complex	Multi fucosylated/disial/galactosylated
71		6 5 3 0 0	3602.7823	3602.762	-5.65	Complex	Non-fucosylated/multisial
72		7 6 1 0 2	3677.8395	3677.845	1.5	Complex	Multi-fucosylated/Monosial/galactosylated
73		7 6 2 0 0	3690.8348	3690.813	-5.91	Complex	Non-fucosylated/disial/galactosylated
74		6 7 1 1 0	3761.8791	3762.889	268.38	Complex	Non-fucosylated/disial/NGNA/Galactosylated
75		6 5 3 0 1	3776.8716	3776.862	-2.54	Complex	Core-fucosylated/multi sial
76		7 6 2 0 1	3864.924	3864.92	-1.03	Complex	Core-fucosylated/disial/galactosylared

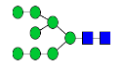
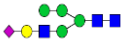
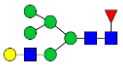
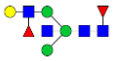
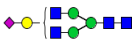
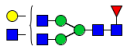
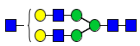
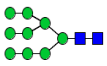
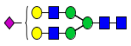
…. Continue

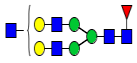
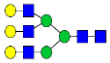
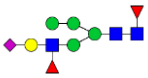
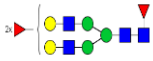
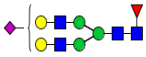
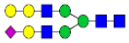
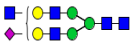
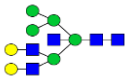
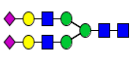
No.	Structure	Gly ID	Theoretical m/z (M+Na+)	Experimental m/z (M+Na+)	mass error (ppm)	Main-Type	Subtype
77		7 6 2 0 2	4039.0312	4040.055	253.41	Complex	Multifucosylated/disial/galactosylated
78		7 6 3 0 0	4052.0084	4052.01	0.39	Complex	Non-fucosylated/multisial/galactosylated
79		7 6 2 0 3	4213.1024	4213.183	19.13	Complex	Multi-fucosylated/disial/ galactosylated
80		7 6 3 0 1	4226.0976	4226.105	1.75	Complex	Core-fucosylated/multisial/Galactosylated
81		7 6 4 0 0	4413.1821	4413.229	10.63	Complex	Non-fucosylated/multisial
82		9 8 2 0 0	4589.287	4589.468	39.43	Complex	Non-fucosylated/Disial/Galactosylated

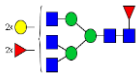
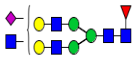
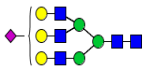
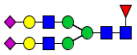
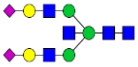
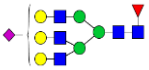
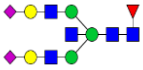
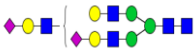
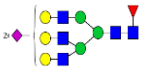
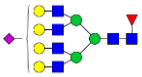
Appendix B Table 3 The **glycoforms detected from Protein A/G enriched IgG fraction**

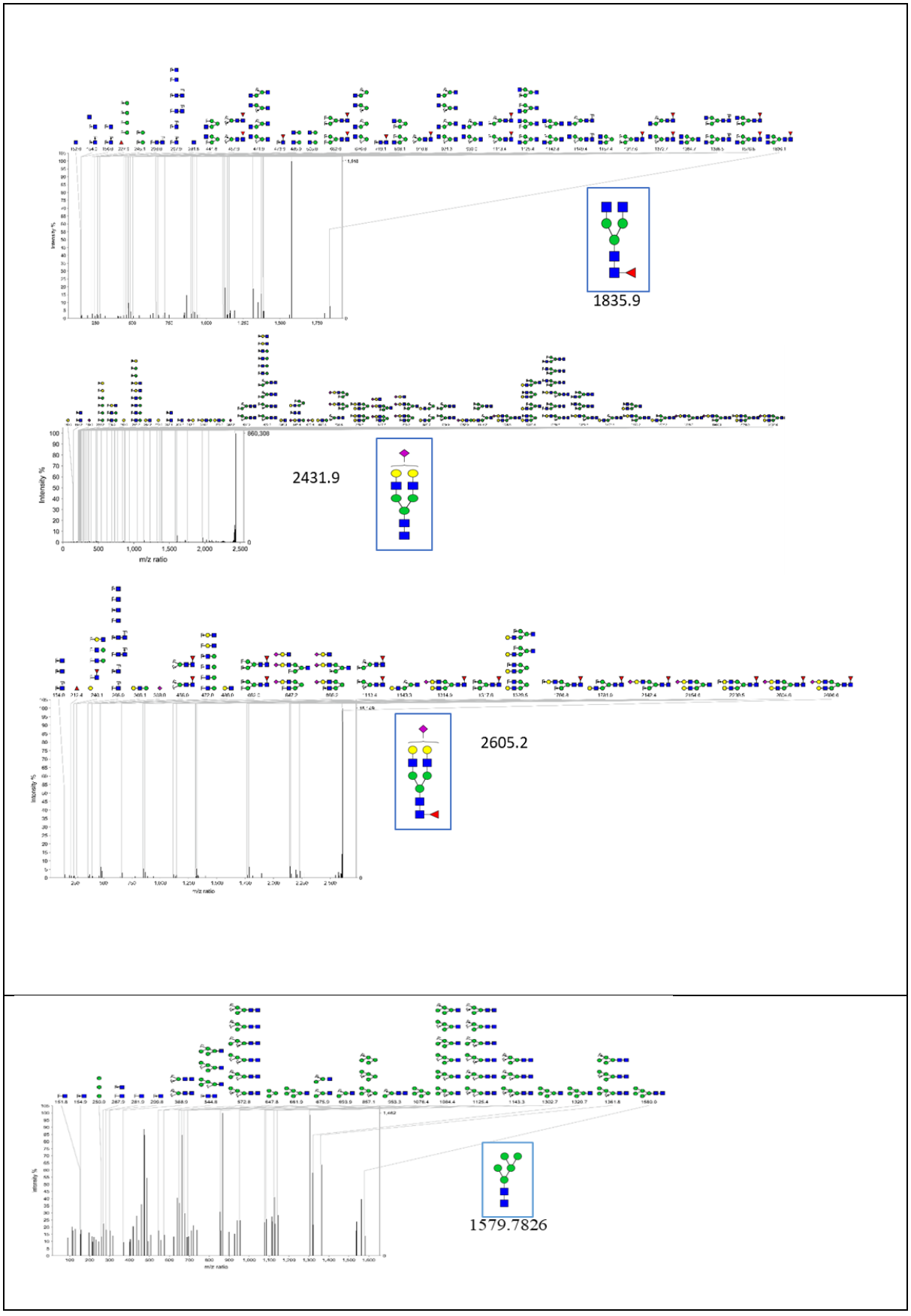
No	Structure	Gly ID (H,N,A,G,F)	Theoretical (M+Na+)	Experimental (M+Na+)	Error in ppm
1		3 3 0 0 0	1416.7094	1416.75	28.7
2		5 2 0 0 0	1578.7826	1578.83	30.0
3		3 3 0 0 1	1590.7986	1590.844	28.5
4		4 3 0 0 0	1620.8091	1620.854	27.7
5		3 4 0 0 0	1661.8357	1661.881	27.3
6		5 2 0 0 1	1753.8718	1753.93	33.2

7		6 2 0 0 0	1783.8824	1783.934	28.9
8		4 3 0 0 1	1794.8984	1794.95	28.7
9		5 3 0 0 0	1824.9089	1824.959	27.5
10		3 4 0 0 1	1835.9249	1835.973	26.2
11		4 4 0 0 0	1865.9355	1865.988	28.1
12		3 5 0 0 0	1906.9620	1905.993	-508.4
13		6 2 0 0 1	1957.9716	1957.035	-478.6
14		7 2 0 0 0	1987.9821	1988.035	26.6
15		4 4 0 0 1	2040.0247	2040.073	23.7
16		5 4 0 0 0	2070.0352	2070.09	26.5
17		3 5 0 0 1	2081.0521	2081.104	24.9
18		4 5 0 0 0	2111.0618	2111.107	21.4
19		4 3 1 0 1	2156.0720	2156.123	23.7

20		5 3 1 0 0	2186.0826	2185.127	-437.3
21		8 2 0 0 0	2192.0819	2192.137	25.1
22		6 3 0 0 1	2203.0971	2202.157	-426.9
23		4 4 0 0 2	2214.1139	2213.121	-448.6
24		4 4 1 0 0	2227.1091	2227.132	10.3
25		5 4 0 0 1	2244.1245	2244.177	23.4
26		4 4 0 1 0	2257.1197	2258.172	466.0
27		4 5 0 0 1	2285.151	2285.205	23.6
28		5 5 0 0 0	2315.1616	2315.202	17.4
29		3 6 0 0 1	2326.1176	2326.794	290.7
30		9 2 0 0 0	2396.1817	2396.235	22.2
31		4 4 1 0 1	2401.1983	2401.254	23.2
32		5 4 0 0 2	2418.2137	2417.228	-407.8
33		5 4 1 0 0	2431.2089	2431.267	23.9
34		4 5 1 0 0	2472.2355	2472.255	7.9

35		5 5 0 0 1	2489.2508	2489.307	22.6
36		6 5 0 0 0	2519.2613	2518.293	-384.5
37		5 3 1 0 2	2534.261	2533.302	-378.6
38		6 3 1 0 1	2564.2716	2563.322	-370.5
39		5 4 0 0 3	2592.3029	2591.339	-372.0
40		5 4 1 0 1	2605.2981	2605.36	23.8
41		6 4 1 0 0	2635.3087	2635.362	20.2
42		4 5 1 0 1	2646.3246	2646.385	22.8
43		5 5 1 0 0	2676.3352	2676.394	22.0
44		3 6 1 0 1	2687.3512	2686.927	-157.9
45		7 5 0 0 0	2723.3611	2722.392	-356.0
46		5 4 2 0 0	2792.3862	2792.458	25.7

47		5 5 0 0 3	2837.4232	2836.478	-333.2
48		5 5 1 0 1	2850.4244	2850.501	26.9
49		6 5 1 0 0	2880.435	2880.502	23.3
50		5 4 2 0 1	2966.4718	2966.562	30.4
51		5 5 2 0 0	3037.5087	3037.142	-120.7
52		3 6 2 0 1	3048.5247	3048.08	-145.9
53		6 5 1 0 1	3054.5242	3054.619	31.0
54		5 5 2 0 1	3211.5981	3211.722	38.6
55		6 5 2 0 0	3241.6087	3241.6087	0.0
56		6 5 2 0 1	3415.6979	3416.866	341.9
57		7 6 1 0 1	3503.7503	3503.269	-137.4



Appendix B F1 MS/MS structure confirmation of human plasma derived sugars

Appendix B Table 4 The percentage of glycan composition in plasma proteins based on specific sugar structure**.**

Hemophilia sample manifest			Plasma Glycan								
			Sialic acid%				Fucosylation%			Galactosylation %	Gal Index
Sample	I(BU)	Age	Total	Mono	Di	Multi	Total	Mono	Multi		G0f/(G1f+G2f*2)
B01	11	62	58.9	22.8	32.2	3.9	45.3	37.8	7.5	44.8	0.02
B04	43	65	64.1	22.2	39.0	2.9	43.4	33.3	10.1	46.9	0.02
B05	<0.6	34	56.0	22.8	30.0	3.2	51.7	43.1	8.6	52.0	0.02
B06	<0.6	18	57.9	22.2	33.1	2.6	49.1	40.0	9.2	51.1	0.02
B07**	<0.6	27	21.3	19.4	1.9	0.0	51.6	47.5	4.1	73.5	1.12
B08**	<0.6	31	52.0	31.8	19.5	0.6	41.2	35.6	5.6	59.9	0.14
B09	<0.6	25	70.7	39.2	30.9	0.5	26.1	21.4	4.7	59.9	0.07
B10-1	<0.6	31	64.1	27.9	33.5	2.8	39.6	29.3	10.4	48.7	0.02
B10-2	past	25	54.7	26.2	26.0	2.5	49.5	40.5	9.0	56.2	0.04
B13	<0.6	34	63.7	28.2	33.4	2.0	46.1	33.5	12.6	50.6	0.03
B17-1	<0.6	49	60.9	24.6	33.6	2.7	45.6	36.1	9.5	51.6	0.03
B17-2	<0.6	58	71.6	26.7	42.4	2.4	39.1	29.8	9.3	45.2	0.01
B18-1	6	19	63.7	24.5	36.5	2.7	39.4	28.8	10.6	46.2	0.02
B18-2	<0.6	34	68.4	27.0	37.6	3.7	38.2	26.8	11.3	47.7	0.02
B19-2	36	5	70.9	27.6	40.0	3.3	35.5	25.6	9.9	45.3	0.01
b11-2	3	21	61.6	27.6	31.2	2.8	47.0	34.3	12.7	52.0	0.03
b12-2	3	20	68.2	25.0	39.6	3.7	39.8	27.2	12.5	46.6	0.02
b20-1	<0.6	35	67.1	37.9	27.5	1.6	33.2	25.7	7.4	62.8	0.10
b20-2	<0.6	37	74.2	21.1	46.3	6.8	35.1	28.1	7.0	39.0	0.01

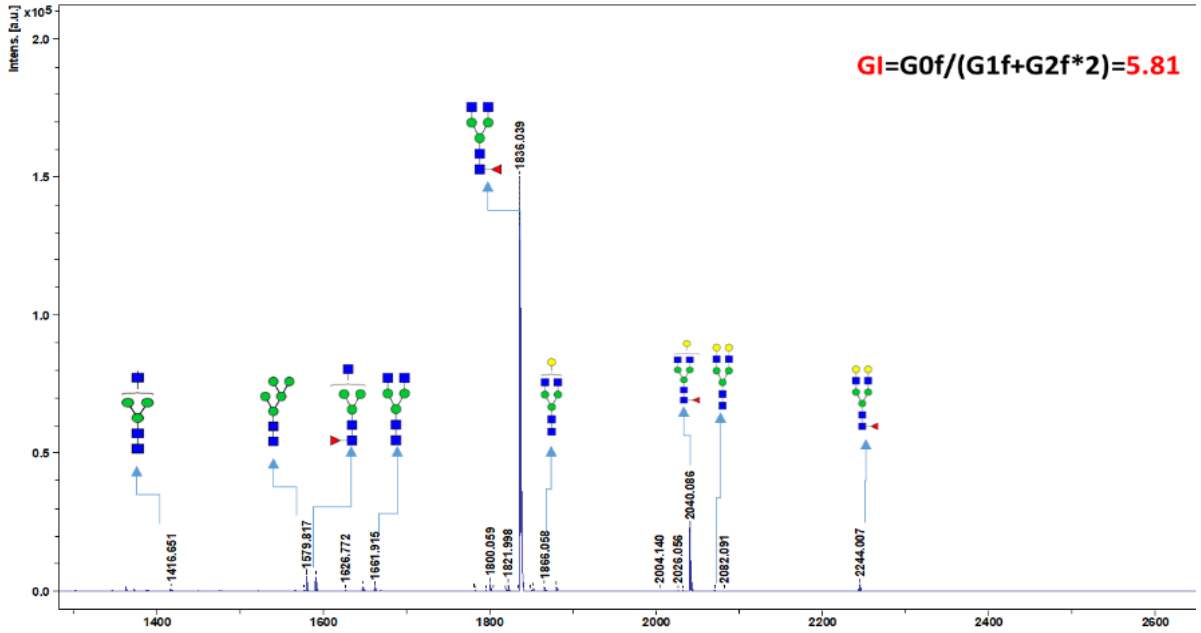
Appendix B Table 4…Continued

Hemophilia sample manifest			Plasma Glycan								
			Sialic acid%				Fucosylation%			Galactosylation %	Gal Index
Sample	I(BU)	Age	Total	Mono	Di	Multi	Total	Mono	Multi		G0f/(G1f+G2f*2)
b21-1	<0.6	27	73.3	24.4	45.3	3.6	36.2	28.3	7.9	42	0.01
b21-2	<0.6	33	72.9	24.1	44.4	4.4	33.0	22.2	10.8	41.9	0.01
b15-2	11	38	63.1	28.2	31.4	3.5	46.8	35.3	11.5	54	0.04
b22-1	<0.6	22	71.8	22	46.8	3	34.9	24.3	10.6	40.6	0.01
b16-1	<0.6	23	71.7	25.3	43.1	3.2	38.5	28.8	9.7	44	0.01
B16-2	<0.6***	33	71.7	28.4	40.6	2.8	38.4	29.5	8.9	45.4	0.01

*Patient with acquired inhibitor, that is he did not have inherited hemophilia, but late in life developed an antibody to FVIII

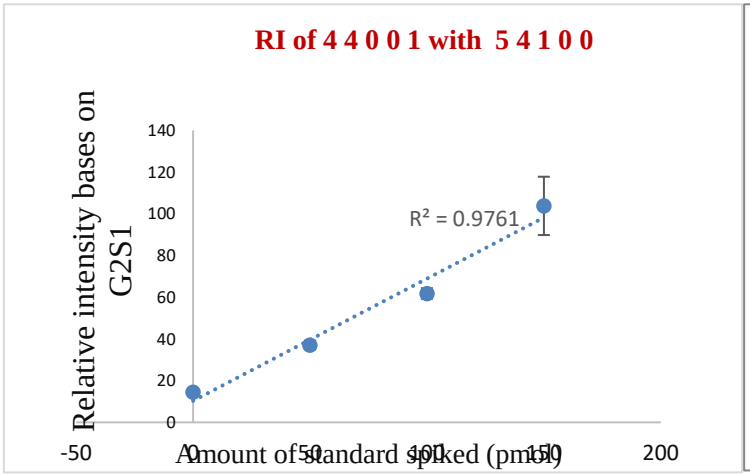
**These are on the same individual pre (B07) and post (B08) emicizumab. He had a history of a high titre inhibitor but did not have recent FVIII exposure at the time of these lab draws and did not have a measurable inhibitor. He was later given FVIII and his inhibitor titre peaked at 50 BU one month after the B08 specimen was drawn.

*** This specimen was drawn prior to a FVIII infusion that showed that the FVIII half-life was decreased. So he has a functional inhibitor below the level of detection. He has a history of a high inhibitor titre

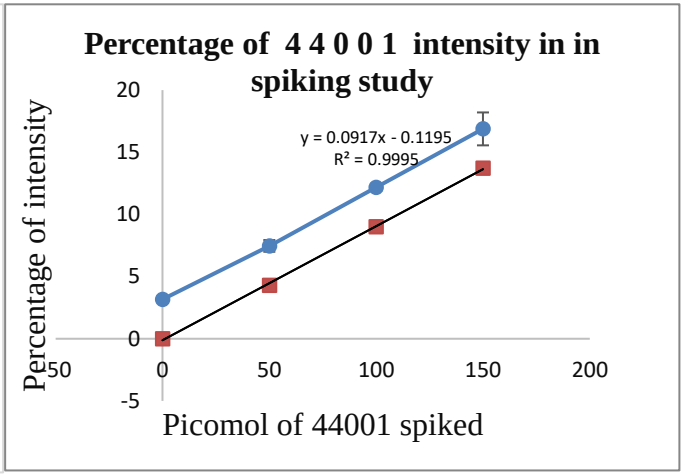


Appendix B F2 Glycan profile of emicizumab
The galactosylation index is calculated using the formula GI=G0f/G1f+G2f*2 and found to be 5

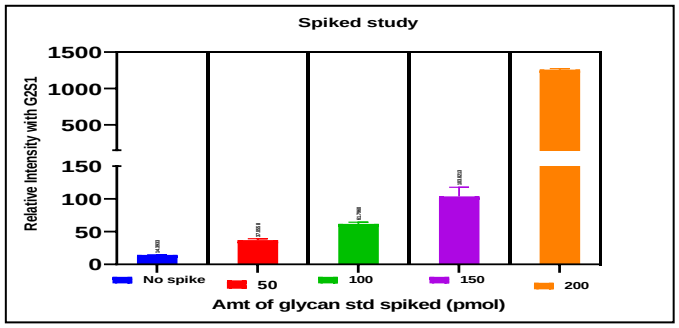
a1)



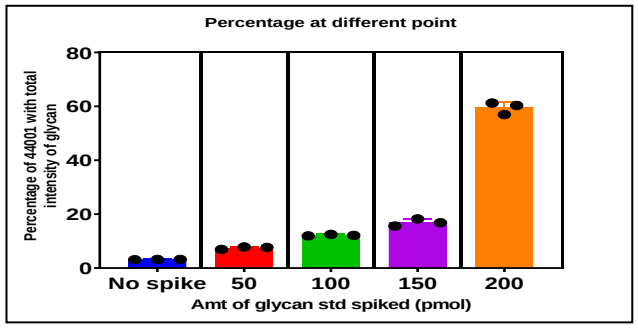
a2)



b1)

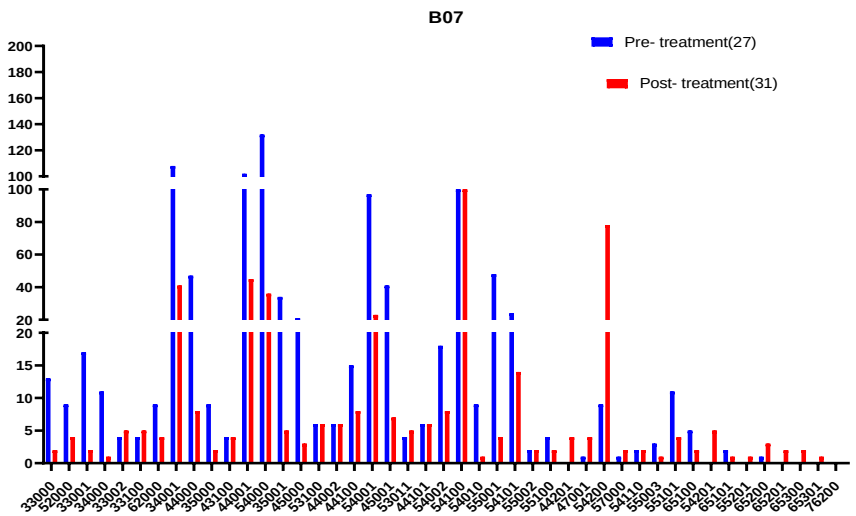


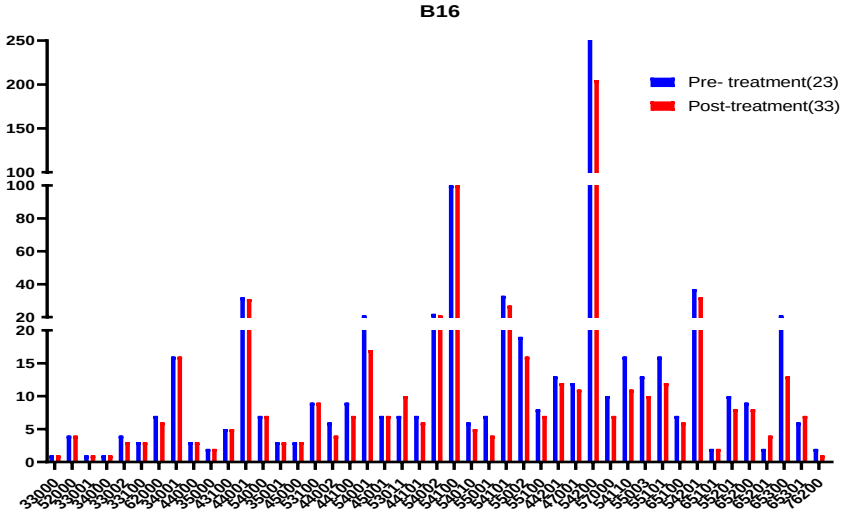
b2)

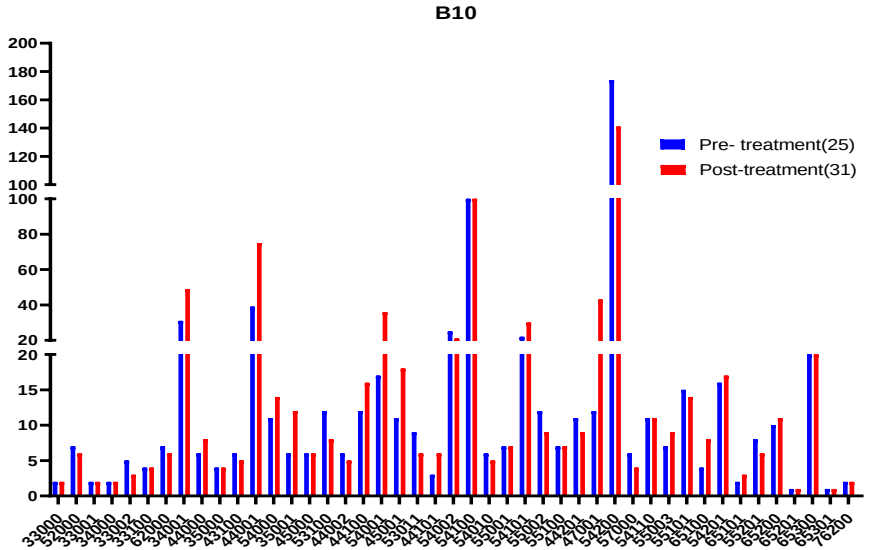


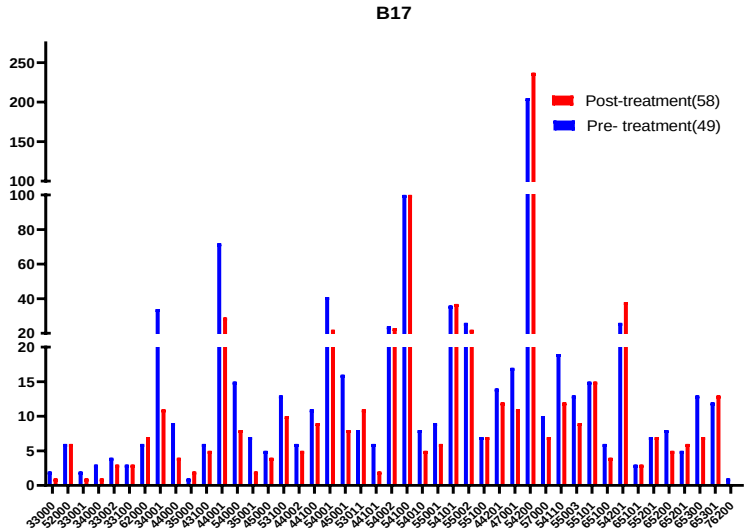
Appendix B F3 Method reliability analysis of MALDI MS by standard spiking.

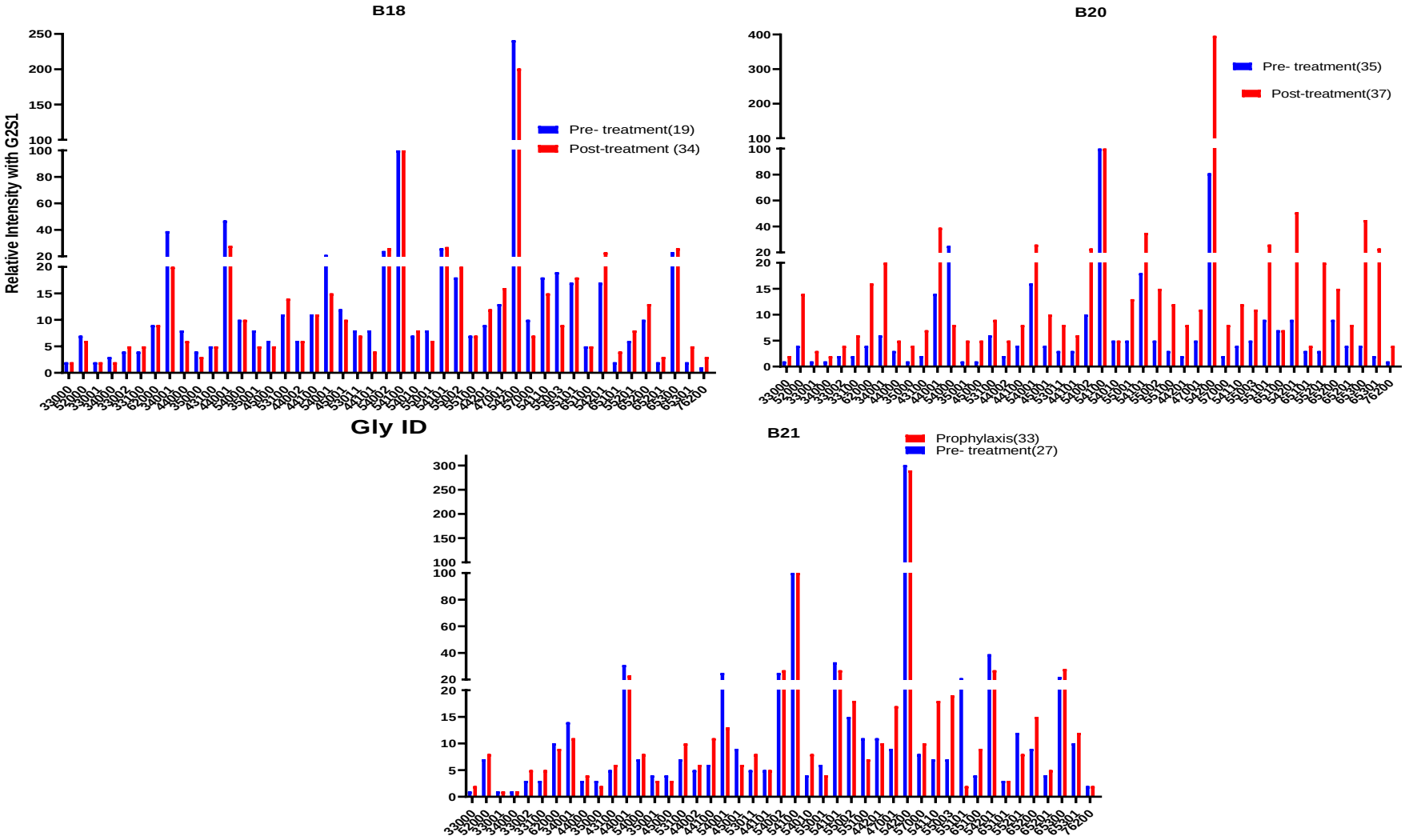
Glycan isolated from 8µl of human serum was aliquoted equally to 5 tubes equally and 50,100,150 and 200 picomoles of 4 4 0 0 1(H4N4F1 or G1f) was added to them individually and the last one was kept as control tube without spiking. The relative intensity of the spiked standard was calculated with the 5 4 1 0 0 (G2S1) inherent glycoform within the serum (a1 and b1). The percentage of intensity of the spiked standard at each point (a2 and b2) were calculated. The linearity was calculated even after reducing the intensity from the no spike sample(a2, orange spots). The SD with triplicate values was plotted in bar graph at b1 and b2 for the spiking study at different points . 200 picomole spike was out of linear range but considered for SD calculation (shown in bar graph). All the acquisition was in triplicate trials and %RSD was within 14%.


B07
Pre- treatment(27)
Post- treatment(31)

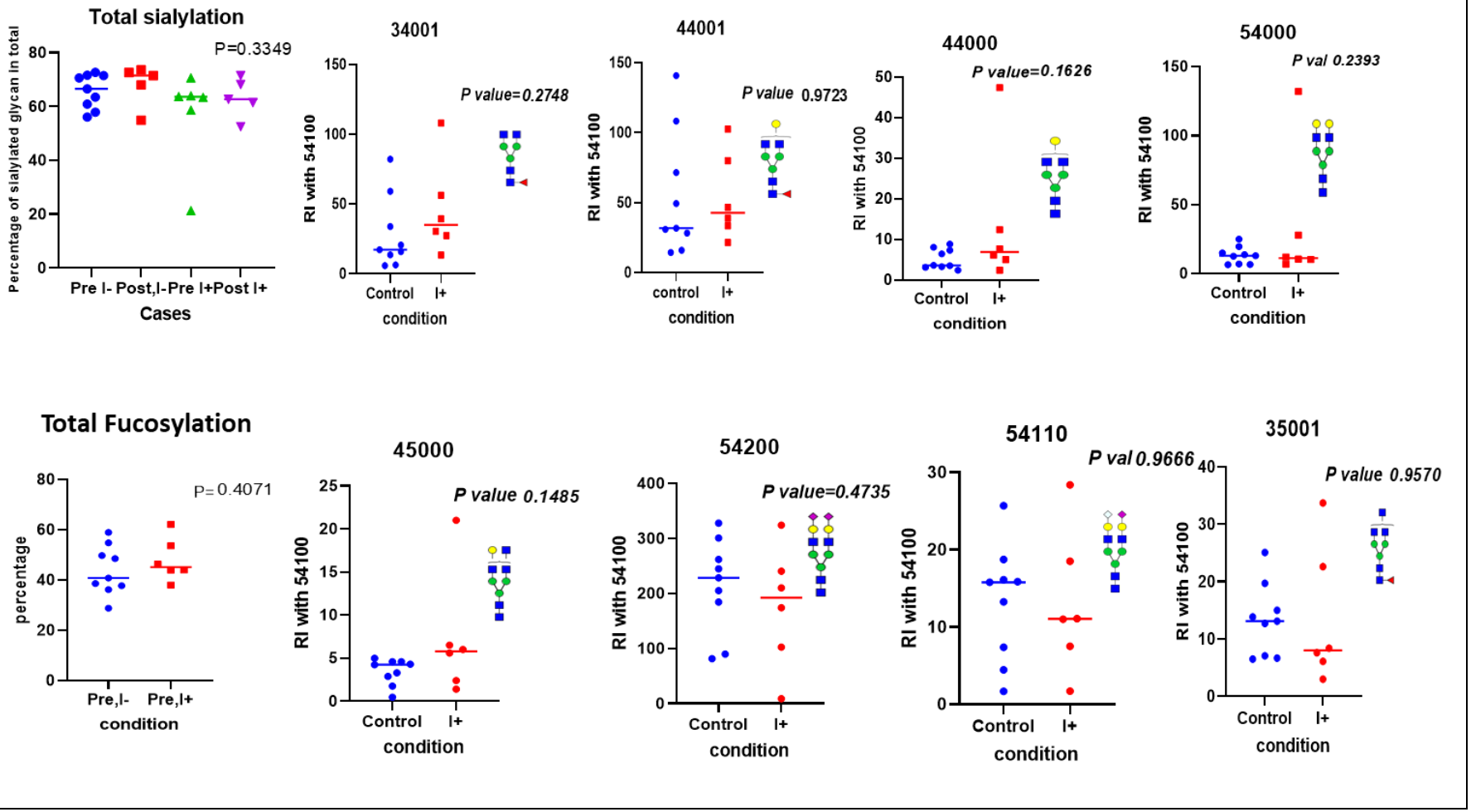

B16
Pre- treatment(23)
Post-treatment(33)


B10
Pre- treatment(25)
Post-treatment(31)


B17
Post-treatment(58)
Pre- treatment(49)

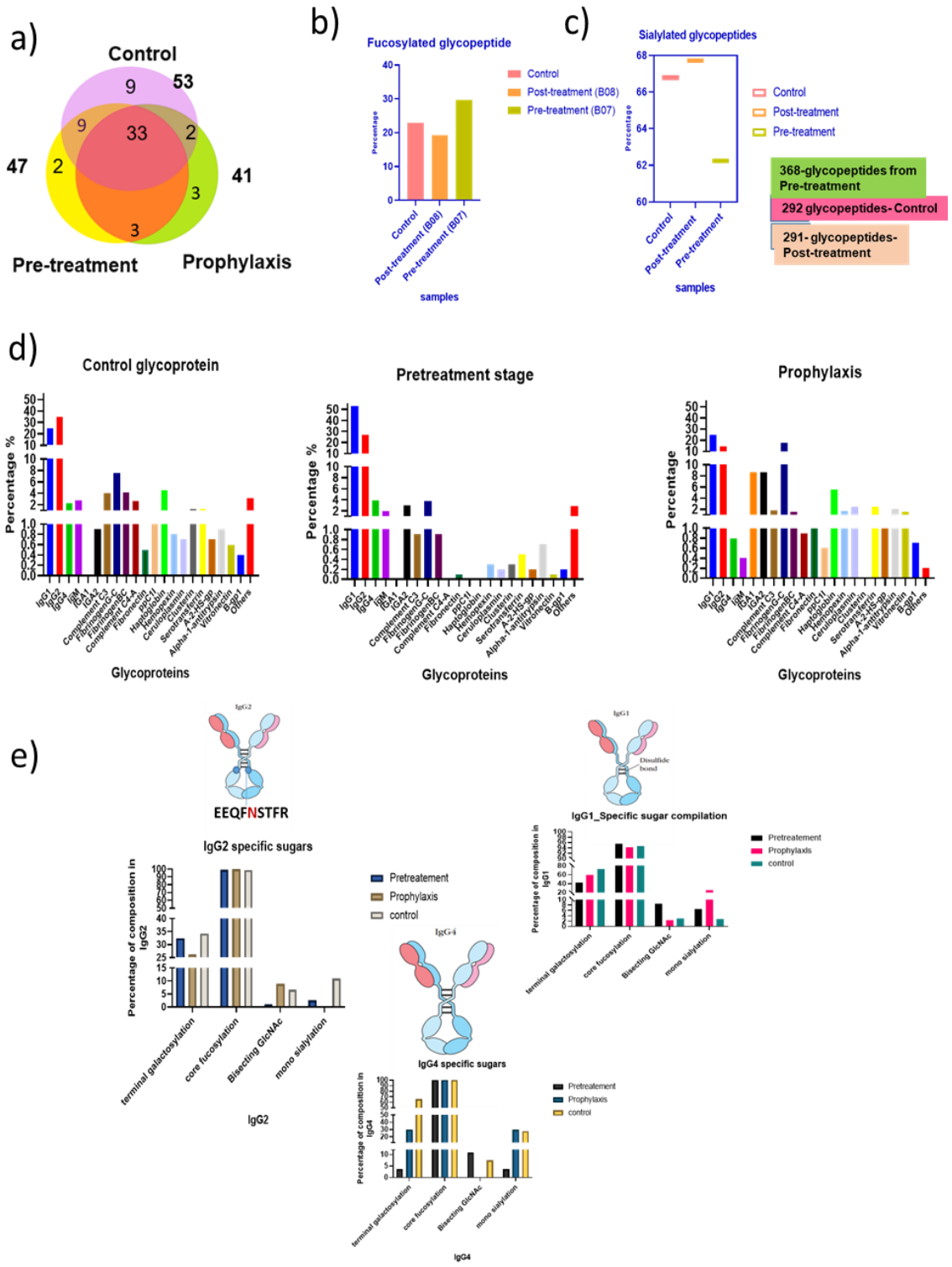


Appendix B F4 Effect of emicizumab on the patient with HA individual patient compilation
B07 and B10 is the patient with I in the past. B18-2 has inhibitor development and the others are inhibitor negative.



Appendix B F5 Comparison of the most different glycoforms between I+ and I- sets with p values.

The most varied glycoforms from I+ and I- HA were compiled and unpaired t test were performed. The p values of total sialylation and fucosylation were calculated between the inhibitor negative and inhibitor positive before and after emicizumab treatment with anova. All were showing. p value is >0.05, statistically insignificant



Appendix B F6 Site-specific glycan subtyping of a single batch showed maximum differences in antibody glycan index and its improvement after emicizumab treatment.

The study was conducted in comparison with the normal HA patient (I-)from the same age group. a) bottom-up approach study and most abundant glycoprotein profiling and comparison between control, pre-treatment and post-treatment samples. b) fucosylated glycopeptides comparison between samples and their percentage calculated with total number of glycopeptide detected c) sialylated glycopeptides comparison between the samples and their percentage calculated with total number of glycopeptide detected d) the glycoprotein profiling of most abundant glycoproteins in the sample calculated with their percentage of composition calculated with their intensities e) IgG subtyping -N-glycosite microheterogeneity calculated based on the structural categorization of glycan.

Appendix B Table 5 The most abundant glycoproteins identified in human plasma sample by bottom up approach

Uniprot id	**common Protein identified in human plasma**
P0DOX5	Immunoglobulin gamma-1 heavy chain
P01859	Immunoglobulin heavy constant gamma 2
P01861	Immunoglobulin heavy constant gamma 4
P01024	Complement C3
P02679	Fibrinogen gamma chain
P02675	Fibrinogen beta chain
P0C0L4	Complement C4-A
P02751	Fibronectin
P02748	Complement component C9
P04114	Apolipoprotein B-100
P04004	Vitronectin
Q96PD5	N-acetylmuramoyl-L-alanine amidase
P01591	Immunoglobulin J chain
P05155	Plasma protease C1 inhibitor
P02750	Leucine-rich alpha-2-glycoprotein
P0DOX2	Immunoglobulin alpha-2 heavy chain
P00738	Haptoglobin
P01009	Alpha-1-antitrypsin
P29622	Kallistatin
P00734	Prothrombin
P00748	Coagulation factor XII
P19823	Inter-alpha-trypsin inhibitor heavy chain H2
P01871	Immunoglobulin heavy constant mu
P02790	Hemopexin
P00450	Ceruloplasmin
P00751	Complement factor B
P05546	Heparin cofactor 2
P10909	Clusterin
P02787	Serotransferrin
P13671	Complement component C6
O75882	Attractin
P03952	Plasma kallikrein
P02749	Beta-2-glycoprotein 1
P02765	Alpha-2-HS-glycoprotein
P22792	Carboxypeptidase N subunit 2
P19827	Inter-alpha-trypsin inhibitor heavy chain H1
P02763	Alpha-1-acid glycoprotein 1
P43652	Afamin
P25311	Zinc-alpha-2-glycoprotein
P01008	Antithrombin-III
P08603	Complement factor H
P06681	Complement C2
P01042	Kininogen-1
Q92954	Proteoglycan 4
P08185	Corticosteroid-binding globulin
Q14624	Inter-alpha-trypsin inhibitor heavy chain H4
P01031	Complement C5
P04196	Histidine-rich glycoprotein
P01023	Alpha-2-macroglobulin
P01876	Immunoglobulin heavy constant alpha 1
P27169	Serum paraoxonase/arylesterase 1
P04217	Alpha-1B-glycoprotein
P01011	Alpha-1-antichymotrypsin
P02743	Serum amyloid P-component
P05156	Complement factor I
P01880	Immunoglobulin heavy constant delta
Q96IV0	Peptide-N(4)-(N-acetyl-beta-glucosaminyl)asparagine amidase
P04278	Sex hormone-binding globulin

Appendix B Table 6 Site specific microheterogeneity of IgG1 compiled based on specific sugar percentage of existence

IgG1 specific glycan	Pre-I+							Post I-					Pre-I-									Post I+				
	B01	B02	B04	B07	B10_1	B18.1	B19	B10_2	B17.2	B18.2	B20.2	B21.2	B05	B06	B09	B13	B16.1	B17.1	B20.1	B21.1	B22.1	B08	B11_2	B12_2	B15.2	B16.2
Sial%	0.1	8.9	5.5	2.7	0.2	5.6	3.1	3.09	5.18	5.62	1.29	2.26	3.5	2.3	0.2	9.6	11.8	2.9	1.5	0.4	4.0	5.12	1.73	11.82	8.48	12.44
terminal gal%	39.9	46.1	30.5	44.2	32.3	40.2	30.0	54.02	52.23	40.21	40.58	33.76	55.0	46.5	32.3	49.1	35.3	55.9	40.2	22.8	36.0	30.50	46.91	43.38	59.57	24.92
Fucosylation	85.7	95.8	96.4	89.6	91.2	88.4	96.7	86.36	68.13	88.38	84.81	86.89	96.4	93.1	91.2	87.2	92.7	67.4	88.2	89.7	69.8	39.29	92.93	94.22	81.63	92.66
bisecting	5.0	3.3	4.1	7.2	13.5	3.4	2.9	0.80	3.17	7.64	10.46	4.38	1.63	4.29	2.88	4.78	3.77	10.27	5.01	3.26	0.00	29.9	1.8	4.7	4.4	17.4
Hybrid high mannose	29.0	26.6	37.7	38.9	43.5	45.2	48.5	25.8	33.0	38.9	48.4	61.0	31.7	34.8	50.6	29.3	31.1	40.9	45.0	57.3	43.7	44.0	36.2	31.9	13.5	18.9

Appendix B Table 7 Site specific microheterogeneity of IgG2 compiled based on specific sugar percentage of existence

Microheterogeneity of IgG2	Pre I+							Post I-					Pre I-									Post I+				
	B01	B02	B04	B07	B10_1	B18.1	B19	B10_2	B17.2	B18.2	B20.2	B21.2	B05	B06	B09	B13	B16.1	B17.1	B20.1	B21.1	B22.1	B08	B11_2	B12_2	B15.2	B16.2
Sial%	0.8	9.6	11.7	0.3	7.4	4.2	2.7	3.0	0.6	1.7	3.6	0.7	3.2	1.9	2.5	4.3	4.8	0.6	6.1	0.9	33.5	2.2	4.4	3.5	3.9	1.5
Gal%	15.5	29.9	52.3	8.2	42.0	11.1	19.5	13.1	24.4	17.9	30.6	40.1	20.2	14.8	26.7	25.4	17.7	24.4	44.4	54.4	76.4	12.8	18.2	19.2	22.8	18.3
Fuc%	85.0	98.6	86.5	98.1	98.2	98.9	98.3	98.0	90.1	92.0	94.0	80.3	97.3	93.4	96.9	92.5	97.3	90.1	84.5	84.1	100.0	94.6	96.9	96.9	94.5	97.0
Bisecting	1.1	1.1	0.9	4.1	5.9	1.5	5.1	2.5	3.2	2.4	5.3	0.3	2.4	4.5	6.0	2.5	1.8	5.2	6.5	11.2	13.0	3.4	6.9	3.4	4.3	5.3
high mannose hybrid	38.5	33.2	44.0	46.9	36.2	35.3	34.7	31.1	28.8	48.0	68.8	68.2	50.7	41.9	38.8	24.2	38.3	39.9	52.8	46.4	18.9	47.6	42.9	24.1	37.5	37.3

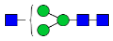
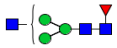
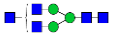
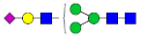
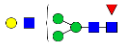
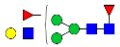
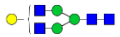
Appendix B Table 8 Site specific microheterogeneity of IgG4 compiled based on specific sugar percentage of existence

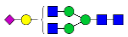
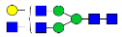
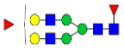
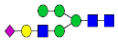
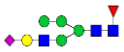
	Pre I+						Post I-					Pre I-									Post I+				
Microheterogeneity of IgG4	B01	B02	B04	B07	B10_1	B19	B10_2	B17.2	B18.2	B20.2	B21.2	B05	B06	B09	B13	B16.1	B17.1	B20.1	B21.1	B22.1	B08	B11_2	B12_2	B15.2	B16.2
Sial%	0.0	10.6	3.9	3.2	9.4	15.6	14.3	24.7	10.6	3.8	26.3	3.8	2.6	7.3	14.4	16.3	2.9	1.8	21.8	5.6	21.8	22.9	36.7	5.8	20.5
Gal%	18.6	73.2	18.2	23.5	54.1	34.3	56.9	69.6	51.5	29.5	55.0	37.2	41.1	30.3	54.3	41.7	69.0	28.7	82.9	59.5	44.8	44.1	28.2	47.5	32.9
Fuc%	84.8	100.0	83.8	100.0	97.1	71.1	100.0	100.0	59.0	82.3	90.6	98.7	96.2	100.0	100.0	91.7	33.7	70.3	91.5	94.0	100.0	100.0	100.0	82.2	73.8
Bisecting GlcNac	2.2	1.8	0.3	1.4	0.0	0.0	0.0	0.0	3.0	4.9	0.0	0.0	2.1	1.1	1.1	0.0	0.7	3.2	0.0	0.9	0.0	1.6	0.0	0.0	23.4
high mannose/hybrid	20.5	34.0	52.0	26.0	10.7	33.4	51.3	56.1	35.7	40.2	27.1	33.5	52.8	49.1	55.5	25.5	80.1	38.9	61.5	73.3	48.3	41.8	48.8	56.2	38.9

Appendix B Table 9 Site specific microheterogeneity of IgM compiled based on specific sugar percentage of existence

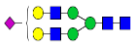
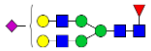
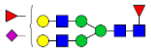
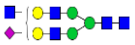
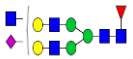
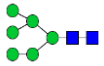
Microheterogeneity at IgM glycopeptide	Pre I+						Post I-					Pre I-									Post I+			
	B01	B02	B04	B07	B10_1	B19	B10_2	B17.2	B18.2	B20.2	B21.2	B05	B06	B09	B13	B16.1	B17.1	B20.1	B21.1	B22.1	B08	B11_2	B12_2	B15.2
Sial %	66	77	60.2	75.4	62.5	68.2	72.7	69.8	80.9	48.7	29.6	79.9	75.5	53.9	67.2	82	78	38.3	32.9	42.1	75.5	65.3	92.8	85.3
Gal %	74.7	75.5	73.6	67.9	71.2	49.1	79.1	63	54.8	52.4	79.6	83.6	61.4	52.2	72.7	83.3	84	62.2	76.1	48.1	62	67.9	91.3	89.8
Fucosylation	67.1	42.5	71	51.2	65.2	34.1	40.5	45.7	66.8	46.6	65.8	49.3	56.4	69.1	74.9	43.9	28.9	68.1	54.8	49.5	44.8	43.3	28.2	45.7
Bisecting GlcNac	37.5	20.1	35.6	23.1	29.5	31.5	12.2	21.9	13.7	11.7	23.9	25.1	25.5	19.4	41.1	22.2	10.7	14.5	13.3	6.9	16.7	15.7	17.5	30.1
high mannose and hybrid	28.9	28.9	37.4	34	35	48.8	26.6	36.8	33.2	55.8	19.7	21.3	39.5	45.9	30.6	17.3	21.2	41.1	23.8	52.3	37.7	36.7	13	12.7

Appendix B Table 10 Site specific microheterogeneity in N- glycosite relating to major immunoglobulins

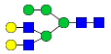
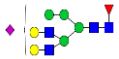
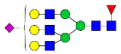
No	Structures identified	Gly ID(H,N,A,G,F)	IgG1	IgG2	IgG4	IgM
1		3 3 0 0 0	Yes	yes	yes	yes
2		3 3 0 0 1	Yes	yes	yes	yes
3		3 3 1 0 1	no	no	no	yes
4		3 4 0 0 0	Yes	Yes	Yes	Yes
5		3 4 0 0 1	Yes	Yes	Yes	Yes
6		3 5 0 0 0	Yes	Yes	no	no
7		3 5 0 0 1	Yes	Yes	Yes	Yes
8		4 2 0 0 0	no	no	no	yes
9		4 2 0 0 1	no	no	no	no
10		4 3 0 0 0	Yes	Yes	Yes	Yes
11		4 3 1 0 0	no	no	no	yes
12		4 3 0 0 1	Yes	Yes	Yes	Yes
13		4 3 0 0 2	Yes	No	No	No
14		4 3 1 0 1	Yes	yes	No	yes
15		4 4 0 0 0	Yes	Yes	Yes	Yes

No	Structures identified	Gly ID(H,N,A,G,F)	IgG1	IgG2	IgG4	IgM
16		4 4 0 0 1	Yes	Yes	Yes	Yes
17		4 4 0 0 2	Yes	No	No	No
18		4 4 1 0 0	Yes	Yes	Yes	No
19		4 4 1 0 1	Yes	Yes	Yes	Yes
20		4 5 0 0 0	Yes	Yes	No	Yes
21		4 5 0 0 1	Yes	Yes	Yes	Yes
22		4 5 0 0 3	Yes	No	No	No
23		4 5 1 0 1	Yes	No	No	No
24		5 2 0 0 0	Yes	Yes	No	yes
25		5 2 0 0 1	No	No	No	yes
26		5 4 0 0 2	No	yes	No	No
27		5 3 0 0 1	Yes	Yes	No	Yes
28		5 3 1 0 0	No	No	No	yes
29		5 3 1 0 1	Yes	Yes	No	Yes
30		5 4 0 0 0	Yes	Yes	Yes	Yes

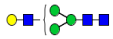
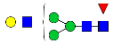
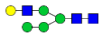
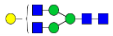
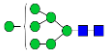
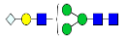
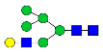
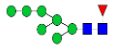
……continues

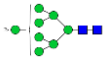
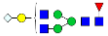
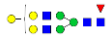
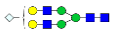
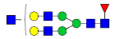
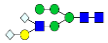
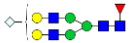
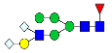
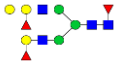
No	Structures identified	Gly ID(H,N,A,G,F)	IgG1	IgG2	IgG4	IgM
31		5 4 0 0 1	Yes	Yes	Yes	Yes
32		5 4 1 0 0	Yes	Yes	yes	yes
33		5 4 1 0 1	Yes	yes	yes	yes
34		5 4 1 0 2	No	No	No	yes
35		5 4 2 0 1	Yes	yes	yes	yes
36		5 5 0 0 0	No	No	No	yes
37		5 5 0 0 1	Yes	yes	No	yes
38		5 5 1 0 0	No	No	No	yes
39		5 5 1 0 1	Yes	Yes	No	yes
40		5 5 1 0 2	No	No	No	yes
41		6 2 0 0 0	Yes	yes	No	yes

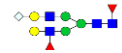
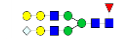
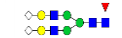
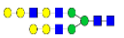
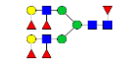
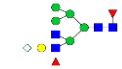
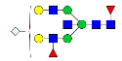
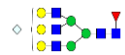
….. continues

No	Structures identified	Gly ID(H,N,A,G,F)	IgG1	IgG2	IgG4	IgM
42		6 3 0 0 0	**No**	**No**	**No**	yes
43		6 3 1 0 0	No	No	No	**yes**
44		**6 3 1 0 1**	**Yes**	No	No	yes
45		6 4 0 0 0	No	yes	No	yes
46		6 4 0 0 1	No	No	No	yes
47		6 4 1 0 0	No	No	No	yes
48		6 4 1 0 1	No	yes	No	yes
49		6 5 1 0 1	No	No	No	yes

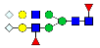
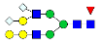
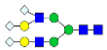
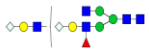
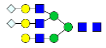
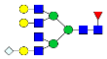
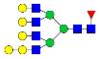
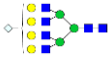
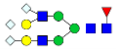
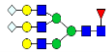
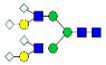
Appendix B Table 11 Mouse plasma glycome detected by one-step permethylation

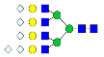
No	Structure	Gly ID	Theoretical Mass (Na+)	Experimental mass(Na+)	Glycan main type	Subtype
1		3 3 0 0 0	1416.709	1416.701	Complex	Non- Fucose/agalactosylated/
2		5 2 0 0 0	1579.783	1579.794	High mannose	High mannose
3		4 3 0 0 0	1620.809	1620.826	Hybrid	Non- Fucose/galactosylated
4		6 2 0 0 0	1783.8824	1783.907	High mannose	High mannose
5		4 3 0 0 1	1794.8984	1793.906	Hybrid	Core-Fucose/galactosylated
6		3 3 0 1 0	1807.8936	1807.92	complex	Non- Fucose/Mono sial
7		5 3 0 0 0	1824.9098	1824.9370	Hybrid	Non- Fucose/galactosylated
8		3 4 0 0 1	1835.9249	1835.949	Complex	Core-Fucose/agalactosylated
9		4 4 0 0 0	1865.9355	1865.964	Complex	Non- Fucose/galactosylated
10		7 2 0 0 0	1987.9821	1988.011	High mannose	High mannose
11		5 3 0 0 1	1998.9981	1998.006	Hybrid	Fucosylated /galactosylated
12		4 3 0 1 0	2011.995	2012.017	complex	Non- Fucose/Mono sial
13		6 3 0 0 0	2029.0089	2029.035	Hybrid	Non- Fucose/galactosylated
14		4 4 0 0 1	2040.0247	2040.045	Complex	core fucosylated/galactosylated
15		5 4 0 0 0	2070.0352	2070.06	Complex	Non- Fucose/galactosylated
16		7 2 0 0 1	2162.0712	2162.104	High mannose	High mannose/fucosylated
17		5 4 0 0 1	2244.1245	2244.144	complex	galactosylated/ core-fucosylated
18		4 4 0 1 0	2257.1197	2257.146	complex	Non- Fucose/Mono sial

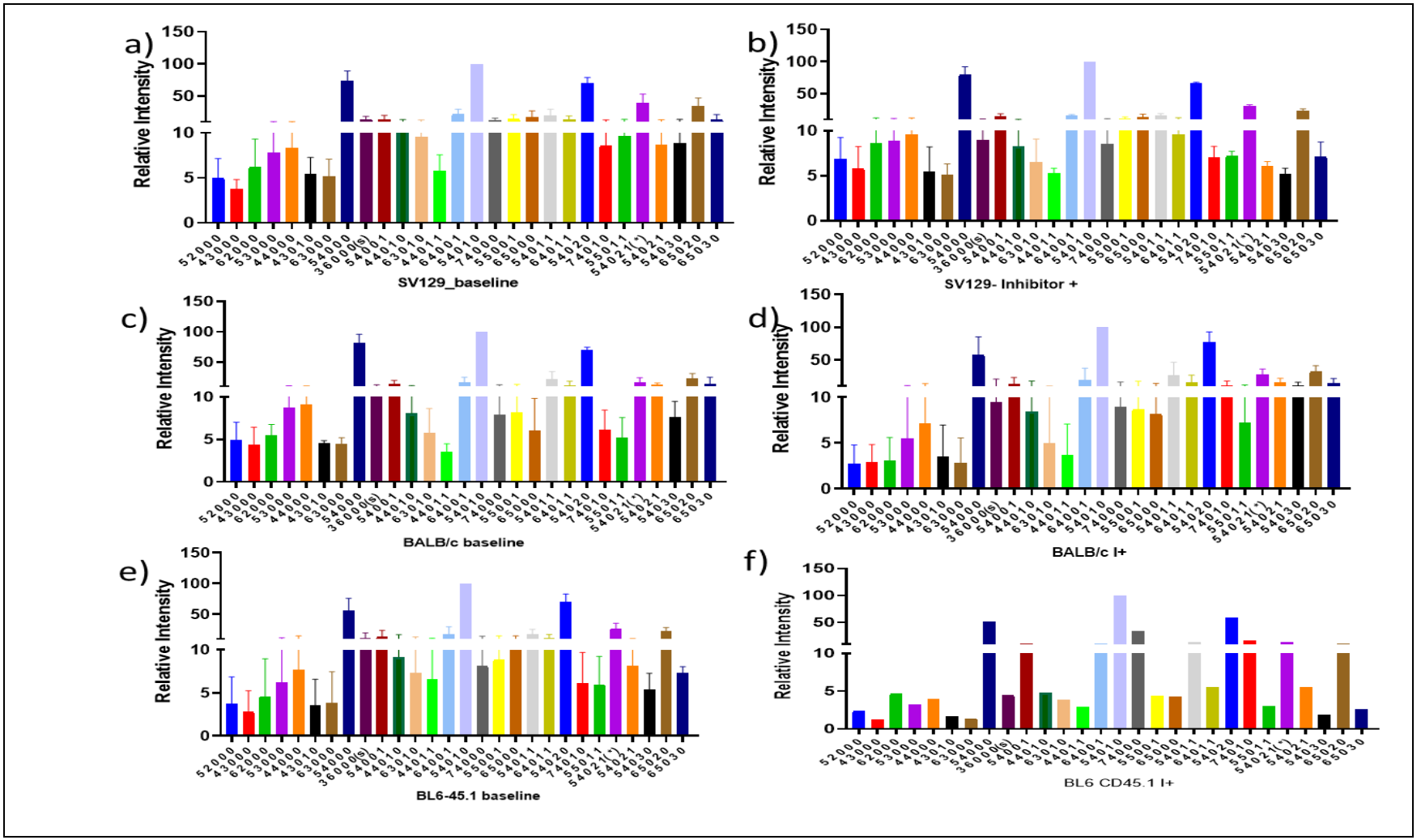
No	Structure		Theoretical Mass (Na+)	Experimental mass(Na+)	Glycan main type	Subtype
19		6 4 0 0 0	2274.135	2274.15	complex	Non- Fucose/galactosylated/alpha-gal
20		6 3 0 1 0	2362.1874	2362.172	Hybrid	Non-fucosylated/ monosial/core-fucose/Neuraminic acid
21		5 3 0 1 1	2490.1824	2390.184	Hybrid	Non-fucosylated/ monosial/core-fucose
22		9 2 0 0 0	2396.1817	2396.193	High mannose	High mannose
23		6 3 0 1 0	2420.1929	2420.203	Hybrid	Non-fucosylated/ Monosial/NGNA
24		4 4 0 1 1	2431.2089	2431.2	Complex	Core-fucosylated/ monosial
25		6 4 0 0 1	2448.2242	2447.207	Complex	galactosylated/ core-fucosylated/alpha-gal
26		5 4 0 1 0	2461.2195	2461.224	Complex	Non-fucosylated/ monosial/NGNA
27		7 4 0 0 0	2478.2378	2477.228	Complex	Non-fucosylated/ galactosylated/alpha-gal
28		5 5 0 0 1	2489.2508	2491.263	Complex	core-fucosylated/galactosylated
29		6 5 0 0 0	2519.2613	2519.289	Complex	Non-fucosylated/ galactosylated
30		5 3 0 2 0	2607.2774	2607.297	Hybrid	Non-fucosylated/ Disialylated
31		5 4 0 1 1	2635.308	2635.32	Complex	core-fucosylated/galactosylated/monosial
32		6 4 0 1 0	2665.3193	2665.331	Complex	Non-fucosylated/ galactosylated/alpha-gal/mono-sial
33		6 5 0 0 1	2693.3506	2692.332	Complex	core-fucosylated/galactosylated
34		5 5 0 1 0	2706.3456	2706.353	Complex	Non-fucosylated/ galactosylated/mono-sial
35		7 5 0 0 0	2723.3615	2723.333	Complex	Non-fucosylated/ galactosylated/alpha-gal
36		5 3 0 2 1	2781.3666	2781.36	Hybrid	Core-fucosylated/disial
37		6 4 0 0 3	2796.4026	2795.372	Complex	multi fucosylated/galactosylated/ alpha-gal

No	Structure		Theoretical Mass (Na+)	Experimental mass(Na+)	Glycan main type	Subtype
38		5 4 0 1 2	2809.3979	2808.367	Complex	multi fucosylated/galactosylated/ monosial
39		5 4 1 1 0	2822.39314	2822.371	Complex	Non-fucosylated/Di monosial
40		6 4 0 1 1	2839.40846	2838.377	Complex	Core-fucosylated/galactosylated/alpha- gal/ monosial
41		5 4 0 2 0	2852.4037	2852.389	Complex	Non-fucosylated/disial
42		6 5 0 0 2	2867.4398	2866.407	Complex	multi-fucosylated/galactosylated
43		7 4 0 1 0	2869.419	2868.394	Hybrid	non-fucosylated/ galactosylated/monosial
44		5 5 0 1 1	2880.435	2882.423	Complex	core-fucosylated/ galactosylated/ mono sial
45		7 5 0 0 1	2897.435	2896.427	Complex	Core-fucosylated/ galactosylated/alpha-gal
46		5 4 0 2 1	2910.285	2910.45	Complex	Core-fucosylated/disial
47		8 5 0 0 0	2927.4609	2926.437	Complex	non-fucose/galactosylated/alpha-gal
48		5 4 0 0 5	2940.4813	2940.459	Complex	multi-fucosylated/galactosylated
49		5 4 1 0 3	2954.2743	2954.462	Complex	multi-fucosylated/galactosylated/mono-sial
50		7 6 0 0 0	2968.487	2968.473	Complex	non-fucosylated/galactosyated
51		6 4 0 1 2	3013.4917	3012.468	Hybrid	multi-fucosylated/monosial
52		5 4 0 2 1	3026.4929	3026.495	Complex	core-fucosylated/di-sial
53		6 5 0 0 3	3041.529	3042.471	Complex	multi-fucosylated/galactosylated
54		5 5 0 1 2	3054.5242	3056.503	Complex	multi-fucosylated/galactosylated/monosial
55		6 5 0 1 1	3084.5348	3084.522	Complex	core-fucosylated/galactosylated/monosial

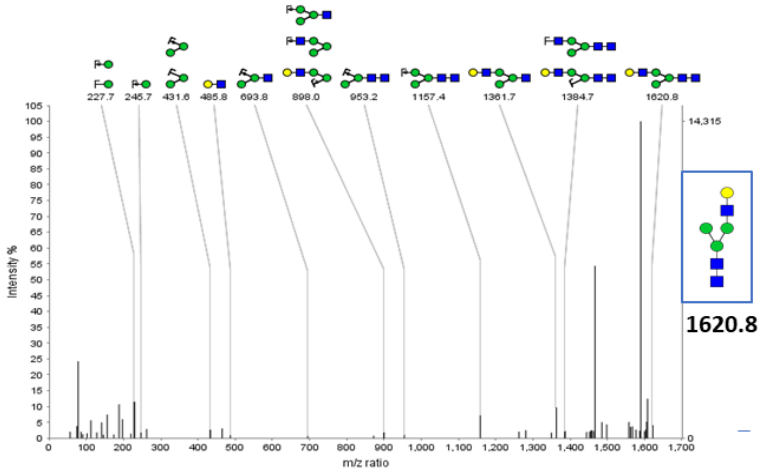
…. Continued

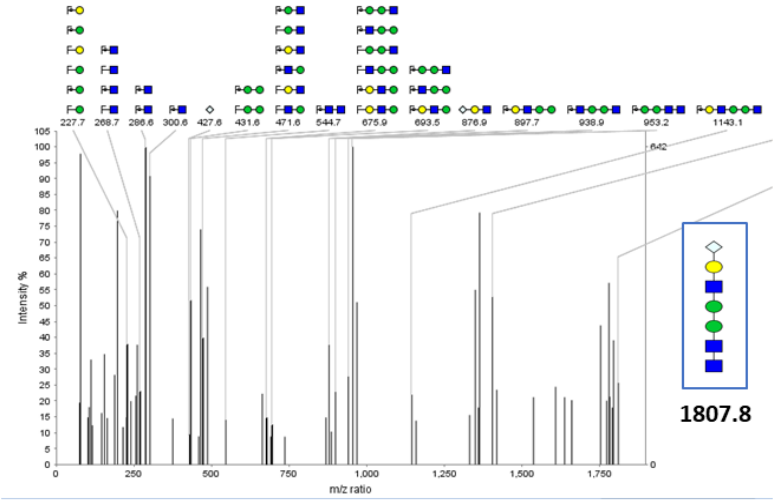
No	Structure	Gly ID	Theoretical Mass (Na+)	Experimental mass(Na+)	Glycan main type	Subtype
56		5 5 0 2 0	3097.53	3097.511	Complex	non-fucosylated/disial
57		7 5 0 1 0	3114.453	3114.503	complex	non-fucosylated/galactosylated/alpha-gal/mono-sial
58		8 6 0 0 0	3172.5872	3172.527	Complex	non-fucosylated/galactosylated/alpha-gal
59		5 4 0 2 2	3200.5821	3200.545	Complex	multi-fucosylated/disial
60		6 4 0 2 1	3228.6134	3229.56	Complex	core-fucosylated/disial/galactosylated/alpha-gal
61		5 4 0 3 0	3743.5879	3243.572	Complex	non-fucosylated/multi-sial
62		6 5 0 1 2	3258.624	3257.582	Complex	Multi-fucosylated/galactosylated/mono-sial
63		5 5 0 2 1	3271.6192	3271.589	Complex	Fucosylated /Di-sial
64		6 6 0 0 3	3286.6553	3286.589	complex	multi-fucosylated/galactosylated
65		6 5 0 2 0	3301.6298	3301.611	Complex	non-fucosylated/di-sial- galactosylated
66		7 6 0 0 2	3316.6659	3315.614	Complex	multi-fucosylated/galactosylated
67		6 6 0 1 1	3329.6611	3331.632	Complex	core-fucosylated/galactosylated/mono-sialylated
68		8 6 0 0 1	3346.6764	3345.628	Complex	core-fucosylated/galactosylated/ alpha-gal
69		7 6 0 1 0	3359.6717	3359.657	complex	Non-fucosylated/ galactosylated/ mono-sial
70		5 4 0 3 1	3417.6717	3417.667	complex	core-fucosylated/multi-sial
71		6 5 0 2 1	3476.689	3476.689	complex	core-fucosylated/galactosylatedmulti-sial
72		5 4 0 4 0	3634.7722	3634.738	complex	core-fucosylated/multi-sial

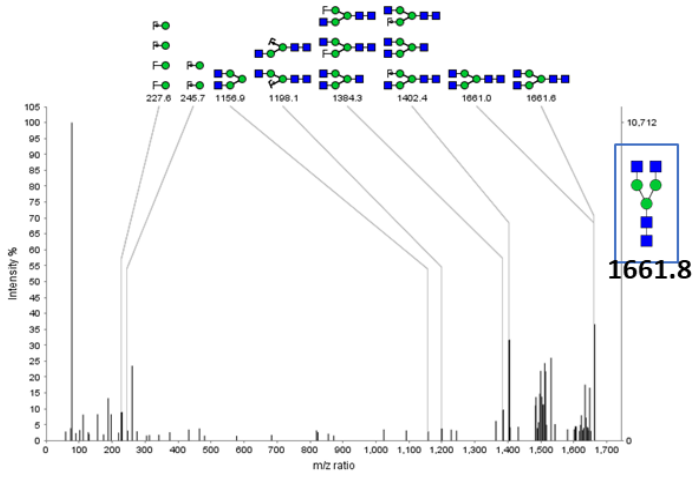
No	Structure	Gly Id	Theoretical Mass (Na+)	Experimental mass(Na+)	Glycan main type	Subtype
73		6 5 0 2 2	3649.8082	3648.737	complex	Multi-fucosylated/ multi sialylated /galactosylated
74		5 5 0 3 1	3662.8035	3662.764	complex	core-fucosylated/ multi sialylated
75		7 5 0 2 1	3679.8188	3678.769	complex	core-fucosylated/galactosylated/ multi sialylated
76		6 5 0 3 0	3692.814	3692.787	complex	Non-fucosylated/ multi sialylated
77		7 6 1 1 0	3720.8453	3722.812	complex	Non-fucosylated/ multi sialylated
78		7 6 0 2 0	3750.8559	3751.839	complex	Non-fucosylated/ multi sialylated
79		6 5 0 3 1	3866.9032	3866.856	complex	core-fucosylated/ multi sialylated
80		7 5 0 2 1	3924.9451	3924.89	complex	core-fucosylated/ multi sialylated
81		7 6 1 1 2	4069.0238	4069.949	complex	Multi-fucosylated/ multi sialylated
82		6 5 0 4 0	4083.9983	4083.976	complex	Non-fucosylated/ multi sialylated
83		7 6 0 2 2	4099.0343	4097.973	complex	Multi-fucosylated/ multi sialylated
84		6 6 0 3 1	4112.0296	4113.008	complex	core-fucosylated/ multi sialylated
85		7 4 0 2 1	4129.0449	4127.985	complex	core-fucosylated/ multi sialylated /galactosylated/alpha-gal
86		7 6 0 3 0	4142.0401	4142.02	complex	Non-fucosylated/ multisial
87		7 7 0 2 1	4170.0714	4171.985	complex	core-fucosylated/multi-sial/galactosylated
88		7 6 0 4 0	4533.2244	4533.294	complex	Non-fucosylated/ multi sialylated
89		7 6 0 5 0	4924.4086	4925.465	complex	Non-fucosylated/ multi sialylated

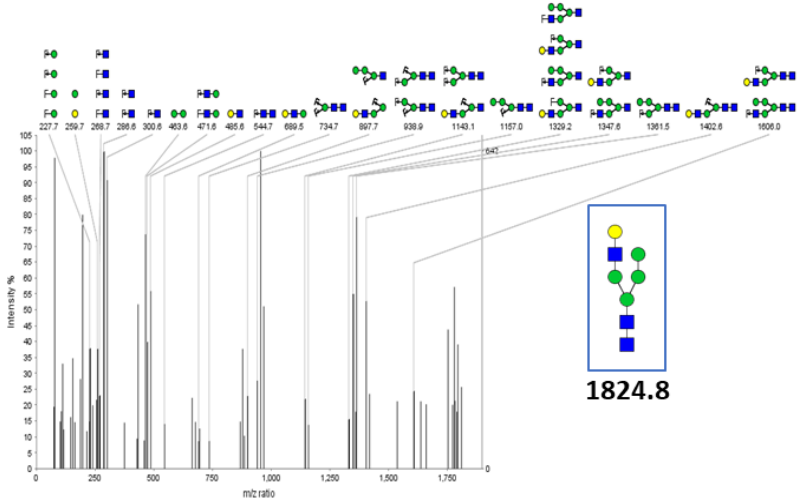


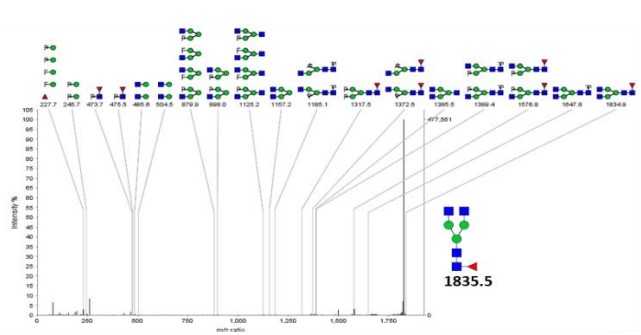
Appendix B F7 Changes of N-glycome composition based on the major 28 glycoforms in mice plasma from different HA patients based on the inhibitor development. The most abundant 33 glycoforms are compared between different HA mice strains. a) SV129 baseline plasma analysis b) SV129 Inhibitor development c) BALB/c baseline d)BALB/c-after inhibitor development e) BL6- baseline f) BL-6 inhibitor development .RI calculated based on G2S1(H5N4G1, 5 4 0 1 0) glycoform

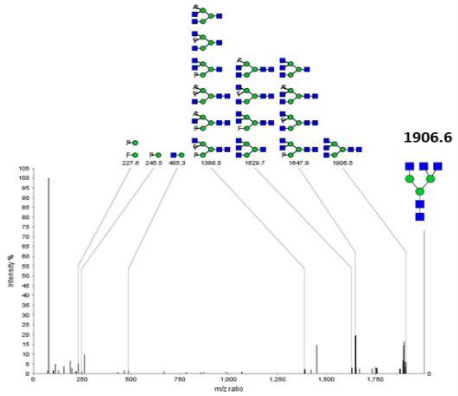

Intensity %
m/z ratio
1620.8

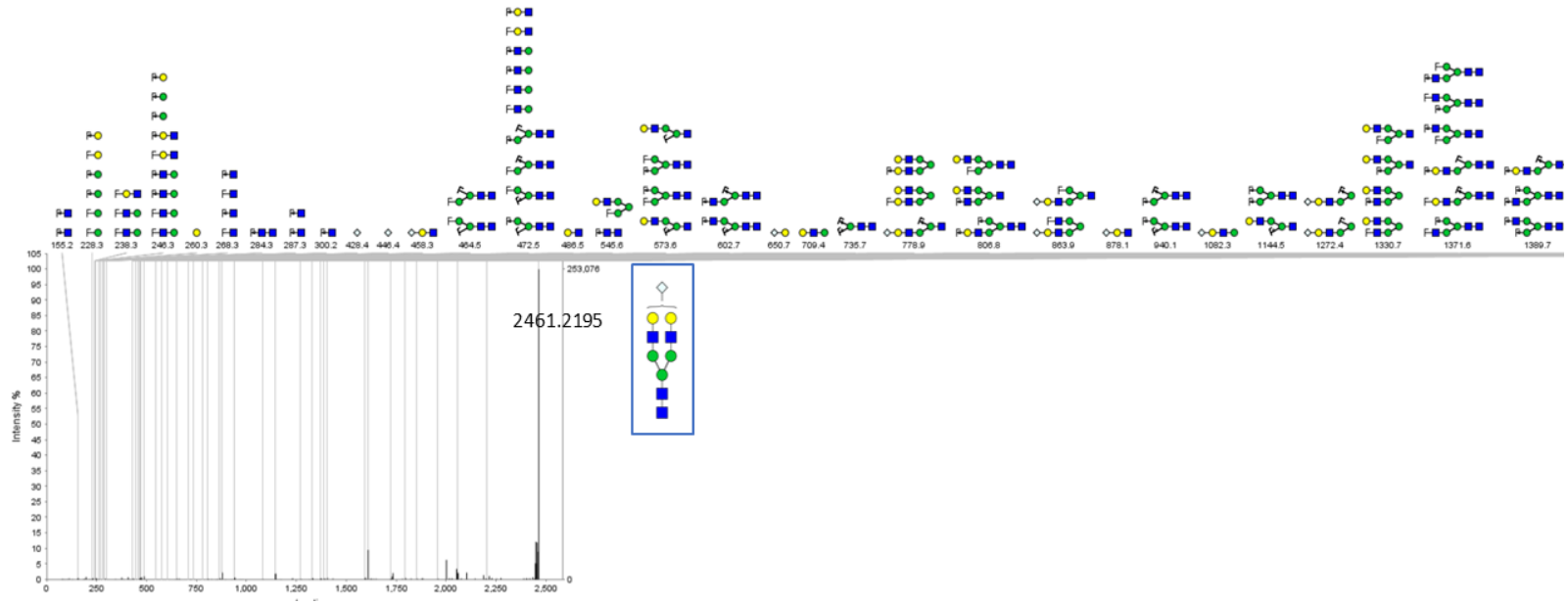

Intensity %
m/z ratio
1807.8

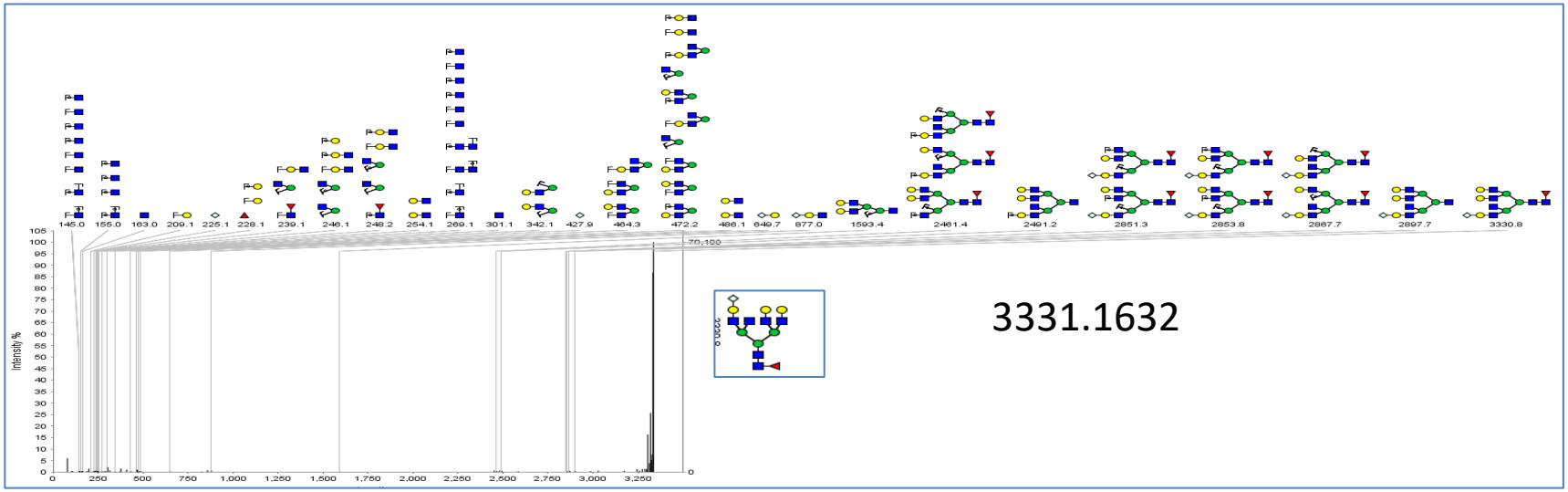

Intensity %
m/z ratio
1661.8


Intensity %
m/z ratio
1824.8


1835.5


1906.6


2461.2195



Appendix B F8 MS/MS structure confirmation of mice plasma derived sugars

Appendix B Table 12 Mouse Plasma glycoprotein detected by bottom -up approach across different mouse strain and inhibitor negative(B) and inhibitor positive(I) cases

Protein	662I+ (BL6-1+)	BL6-1(B)	BL6-2(B)	BL6-3 (B)	BL6-4 (B)	BALB/c1 (B)	BALB/c2 (B)	BALB/c3 (B)	BALB/c4 (B)	649I (BALB/c) (I+)	650I(BALB/c1) (I+)	660I(BALB/c1) (I+)	SV129_1 (B)	SV129_2 (B)	SV129_3 (B)	645B(SV129)	646B(SV129)	647B(SV129)	SV129_1(I+)	645I(SV129)	646I (SV129
Serine protease inhibitor A3K	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Prothrombin	Yes	Yes	Yes	Nil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Nil	Yes	Nil	Yes	Nil	Nil
Histidine-rich glycoprotein	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Nil	Yes	Yes
Murinoglobulin-1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fibronectin	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Alpha-1-antitrypsin 1-2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Nil	Yes	Nil	Yes	Yes	Nil
Immunoglobulin heavy constant mu	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Complement C4-B	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Vitamin D-binding protein	Yes	Yes	Yes	Yes	Yes	Nil	Yes	Yes	Nil	Yes	Yes	Yes	Yes	Yes	Yes	Nil	Nil	Yes	Yes	Yes	Nil
Phosphatidylinositol-glycan-specific phospholipase D	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Haptoglobin	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Nil	Nil	Yes	Yes	Yes	Nil
Ig gamma-2B chain C	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Apolipoprotein C-IV	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Nil	Nil	Nil	Nil	Nil	Nil
Vitronectin	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Nil	Yes	Yes	Yes	Yes
Carboxypeptidase N subunit 2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Nil	Nil	Nil	Nil	Nil	Nil
Ig gamma-3 chain C region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Complement component C9	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Nil	Nil	Nil	Yes	Nil	Nil
Pregnancy zone protein	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Transthyretin	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Heparin cofactor 2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Hemopexin	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Nil	Yes	Yes	Nil	Nil	Nil	Yes	Yes	Yes
H-2 class I histocompatibility antigen, Q10 alpha chain	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Epidermal growth factor receptor	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Nil	Yes	Yes	Yes	Yes	Yes

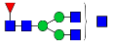
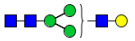
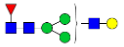
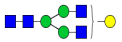
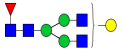
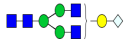
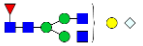
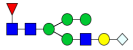
Protein	662I+ (BL6-1+)	BL6-1(B)	BL6-2(B)	BL 6-3 (B)	BL 6-4 (B)	BALB/c1 (B)	BALB/c2 (B)	BALB/c3 (B)	BALB/c4 (B)	649I (BALB/c) (I+)	650I(BALB/c1) (I+)	660I(BALB/c1) (I+)	SV129_1 (B)	SV129_2 (B)	SV129_3 (B)	645B(SV129)	646B(SV129)	647B(SV129)	SV129_1 (I+)	645I(SV129)	646I (SV129
Leukemia inhibitory factor receptor	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Kininogen-1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Antithrombin-III	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Nil	Nil	Yes	Yes	Yes	Nil
Afamin	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clusterin	Yes	Yes	Yes	Yes	Yes	Yes	Nil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Nil	Nil	Yes	Nil	Nil	Yes	Nil
Complement C5	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Nil	Nil	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Serotransferrin	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Alpha-1-acid glycoprotein 1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Plasma kallikrein	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Nil	Yes	Yes
Fibrinogen beta chain	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Nil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Nil
Inter-alpha-trypsin inhibitor heavy chain H1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Complement factor H	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Carboxylesterase 1C	Yes	Yes	Yes	Yes	Yes	Nil	Nil	Yes	Nil	Yes	Yes	Nil	Nil	Yes	Yes	Nil	Nil	Nil	Nil	Nil	Nil
Serum paraoxonase/arylesterase 1	Yes	Nil	Nil	Yes	Nil	Yes	Nil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Nil	Nil	Nil	Nil	Nil	Nil
Protein Z-dependent protease inhibitor	Yes	Nil	Nil	Yes	Nil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Nil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Complement factor I	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Immunoglobulin J chain	Yes	Nil	Nil	Yes	Nil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Nil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Carboxypeptidase B2	Yes	Nil	Yes	Nil	Nil	Nil	Nil	Yes	Nil	Yes	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil
Apolipoprotein B-100	Yes	Nil	Yes	Yes	Yes	Yes	Nil	Yes	Yes	Yes	Yes	Nil	Yes	Yes	Yes	Nil	Nil	Nil	Nil	Nil	Nil
C4b-binding protein	Nil	Nil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fetuin-B	Nil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Nil	Nil	Nil	Nil	Nil	Nil
Alpha-2-HS-glycoprotein	Nil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Interleukin-1 receptor accessory protein	Nil	Yes	Nil	Nil	Nil	Yes	Yes	Yes	Yes	Yes	Nil	Nil	Yes	Yes	Yes	Nil	Nil	Nil	Nil	Nil	Nil

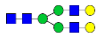
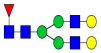
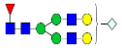
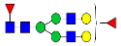
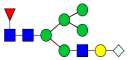
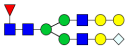
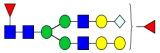
Protein	662I+ (BL6-1+)	BL6-1(B)	BL6-2(B)	BL6-3 (B)	BL 6-4 (B)	BAL B/c1 (B)	BAL B/c2 (B)	BAL B/c3 (B)	BAL B/c4 (B)	649I (BA LB/c) (I+)	650I(B ALB/c1) (I+)	660I(B ALB/c1) (I+)	SV1 29_ 1 (B)	SV129_2 (B)	SV12 9_3 (B)	645B(SV129)	646B(SV129)	647B(SV129)	SV129 _1(I+)	645I(S V129)	646I (SV12 9
Serum amyloid P-component	Nil	Yes	Yes	Yes	Nil	Yes	Nil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Nil
Sulfhydryl oxidase 1	Nil	Nil	Nil	Yes	Nil	Yes	Yes	Yes	Nil	Nil	Yes	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Sulfhydryl oxidase 2	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Beta-2-glycoprotein 1	Nil	Yes	Yes	Yes	Nil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Nil	Nil	Nil	Nil	Nil	Nil
Complement component C8 beta chain	Nil	Yes	Yes	Yes	Nil	Yes	Yes	Yes	Yes	Yes	Yes	Nil	Yes	Yes	Yes	Nil	Nil	Nil	Nil	Nil	Nil
Serine protease inhibitor A3M	Nil	Nil	Yes	Nil	Nil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Nil	Yes	Nil	Yes	Nil	Nil	Nil	Yes	Nil
Inter alpha-trypsin inhibitor, heavy chain 4	Nil	Nil	Yes	Yes	Ye s	Yes	Yes	Yes	Yes	Yes	Yes	Nil	Nil	Yes	Yes	Yes	Nil	Yes	Nil	Yes	Nil
Hepatocyte growth factor activator	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Yes	Nil	Yes	Yes	Nil	Nil	Yes	Yes	Nil	Nil	Nil	Nil	Nil	Yes
Phosphatidylcholine-sterol acyltransferase	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Yes	Nil	Yes	Nil	Nil	Yes	Yes	Yes	Nil	Yes	Nil	Yes	Nil	Yes
Coagulation factor XIII B chain	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Yes	Nil	Yes	Yes	Nil	Nil	Yes	Nil	Yes	Yes	Yes	Yes	Yes	Nil
Microfibril-associated glycoprotein 4	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Yes	Nil	Yes	Yes	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Killer cell lectin-like receptor 4	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Ceruloplasmin	Nil	Nil	Yes	Nil	Nil	Yes	Nil	Yes	Yes	Yes	Yes	Nil	Nil	Yes	Nil	Yes	Yes	Yes	Yes	Yes	Nil
Complement component C8 gamma chain	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Yes
Zinc-alpha-2-glycoprotein	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Yes	Nil	Yes	Yes	Yes	Nil	Yes	Yes	Nil	Nil	Nil	Nil	Nil	Nil
Odorant-binding protein 1a	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Cellulose synthase-like protein D2	Nil	Yes	Yes	Nil	Nil	Yes	Yes	Yes	Nil	Yes	Yes	Yes	Nil	Yes	Yes	Nil	Yes	Yes	Yes	Yes	Yes
G2 and S phase-expressed protein 1	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil
Mannan-binding lectin serine protease 1	Nil	Nil	Nil	Yes	Ye s	Yes	Yes	Yes	Nil	Yes	Yes	Nil	Nil	Nil	Nil	Yes	Yes	Yes	Yes	Yes	Nil
Pentatricopeptide repeat-containing protein At2g32630	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Proteoglycan 4	Nil	Nil	Nil	Yes	Ye s	Nil	Nil	Yes	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Acyl-CoA-binding domain-containing protein 1	Nil	Nil	Nil	Yes	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Periostin	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Collagen alpha-1(I) chain	Nil	Nil	Yes	Yes	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil

Protein	662I+ (BL6-1+)	BL6-1(B)	BL6-2(B)	BL6-3 (B)	BL6-4 (B)	BALB/c1 (B)	BALB/c2 (B)	BALB/c3 (B)	BALB/c4 (B)	649I (BALB/c) (I+)	650I(BALB/c1) (I+)	660I(BALB/c1) (I+)	SV129_1 (B)	SV129_2 (B)	SV129_3 (B)	645B(SV129)	646B(SV129)	647B(SV129)	SV129_1(I+)	645I(SV129)	646I (SV129
Lumican	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Carboxypeptidase N catalytic chain	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Yes	Yes	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil
Ig gamma-1 chain C region secreted form	Nil	Nil	Nil	Nil	Nil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Nil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Platelet glycoprotein V	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Thrombospondin-1	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Yes	Nil	Yes	Yes	Nil	Nil	Yes	Nil	Yes	Yes	Nil
Serum albumin	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Zinc finger protein 638	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Beta-taxilin	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Yes	Yes	Nil	Nil	Nil	Nil	Nil	Nil
Coagulation factor IX	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Alpha-1-acid glycoprotein 1(Mus caroli (Ryukyu mouse(Mus caroli (Ryukyu mouse)	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Yes	Yes	Nil	Yes	Nil	Yes	Nil	Nil	Yes	Yes	Nil
Insulin-like growth factor-binding protein complex acid labile subunit	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Yes	Nil	Yes	Yes	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Yes	Nil	Nil
Ig heavy chain V region AC38 15.3	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Complement factor B	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Corticosteroid-binding globulin	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Plasma protease C1 inhibitor	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Testis-expressed protein 15	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Kinesin-like protein KIN-14Q	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Zinc finger and BTB domain-containing protein 8A	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil
Prenylcysteine oxidase	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Zinc finger protein ENHYDROUS	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Armadillo repeat protein deleted in velo-cardio-facial syndrome homolog	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Two-component response regulator-like APRR2	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
LON peptidase N-terminal domain and RING finger protein 3	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil

Protein	662I+ (BL6-1+)	BL6-1(B)	BL6-2(B)	BL 6-3 (B)	BL 6-4 (B)	BAL B/c1 (B)	BAL B/c2 (B)	BAL B/c3 (B)	BAL B/c4 (B)	649I (BAL B/c) (I+)	650I(BA LB/c1) (I+)	660I(BA LB/c1) (I+)	SV12 9_1 (B)	SV12 9_2 (B)	SV12 9_3 (B)	645B(S V129)	646B(S V129)	647B(S V129)	SV129_ 1(I+)	645I(S V129)	646I (SV129
Complement C1q subcomponent subunit A	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Yes	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Gamma-tubulin complex component 4	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Transcription initiation factor TFIID subunit 6	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil
DDT domain-containing protein DDR4	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Complement C1s-A subcomponent	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Ankyrin repeat and EF-hand domain-containing protein 1	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Inhibitor of carbonic anhydrase	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Dedicator of cytokinesis protein 5	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Coagulation factor V	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Yes	Nil	Nil	Yes	Nil
Collagen alpha-1(III) chain	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Yes	Nil	Nil	Nil	Nil	Yes	Nil	Nil
N-acetylmuramoyl-L-alanine amidase	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Dynein heavy chain 8, axonemal	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Apolipoprotein D	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Integrin beta-2-like protein	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Coagulation factor XII	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Yes	Nil	Nil
DUF724 domain-containing protein 2	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil
Protein AMBP	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Yes	Nil	Nil	Nil
Cartilage oligomeric matrix protein	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil
Fibrinogen gamma chain	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Yes	Nil	Yes	Yes
Ig heavy chain V region MC101	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil
Complement C1r-A subcomponent	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Yes	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil	Nil
Totally identified	42	45	58	54	44	56	57	81	51	71	66	53	50	71	56	40	40	42	47	48	34
Total glycopeptide(less than1%FDR	358	438	596	410	495	546	398	873	546	664	756	586	337	877	680	364	506	530	670	565	365

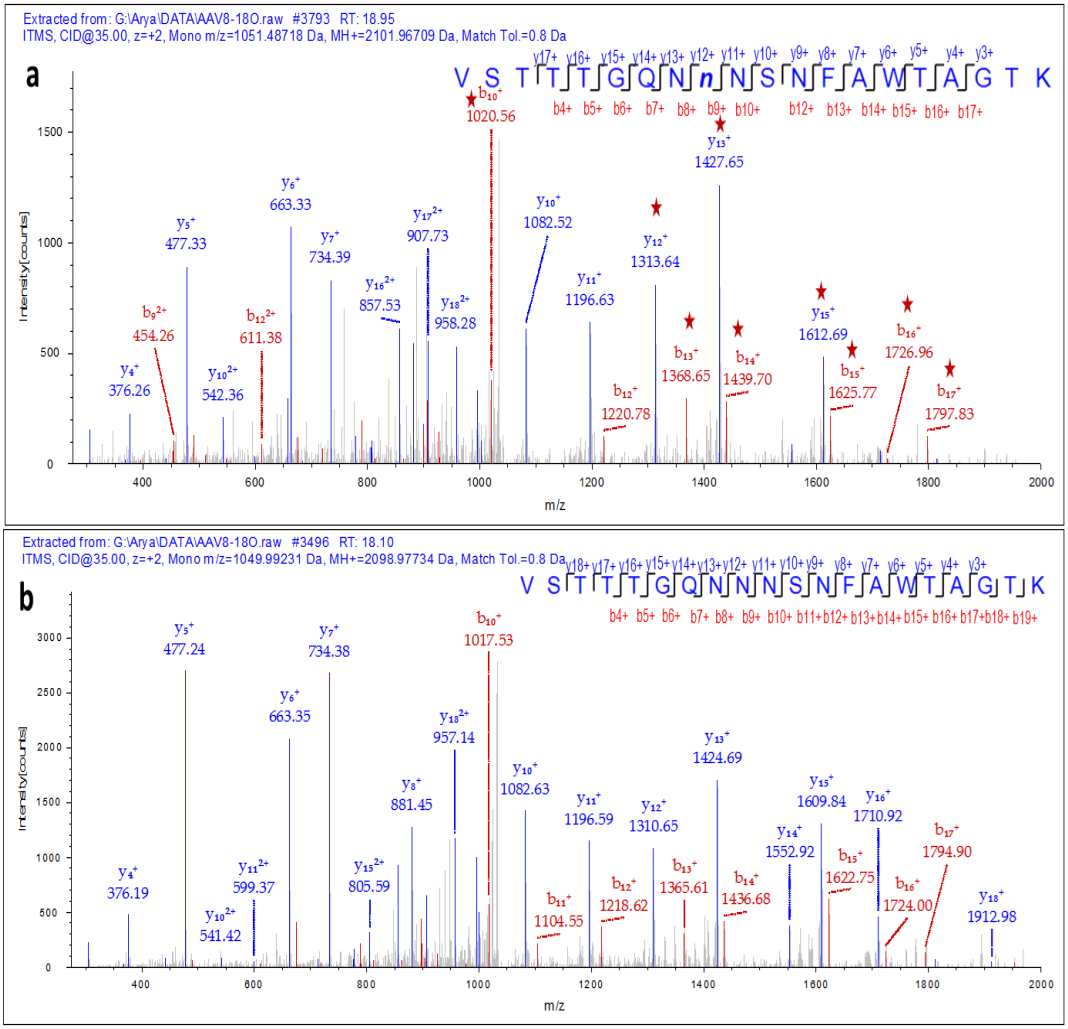
Appendix B Table 13 Mouse IgG subtyping site specific microheterogeneity arranged bases on BALB/c data

Trial No	Glycan Id (HNAGF)	Cartoon structure	IgG1	IgG2B	IgG3
1	3 3 0 0 0		yes	yes	yes
2	3 3 0 0 1		yes	yes	yes
3	3 4 0 0 0		yes	yes	yes
4	3 4 0 0 1		yes	yes	yes
5	3 5 0 0 1		yes	no	no
6	4 3 0 0 0		yes	no	no
7	4 3 0 0 1		yes	yes	yes
8	4 3 0 1 1		yes	yes	yes
9	4 4 0 0 0		yes	yes	no
10	4 4 0 0 1		yes	yes	yes
11	4 4 0 1 0		yes	no	no
12	4 4 0 1 1		yes	yes	yes
13	5 2 0 0 0		yes	no	no
14					
15	5 3 0 1 1		yes	no	no

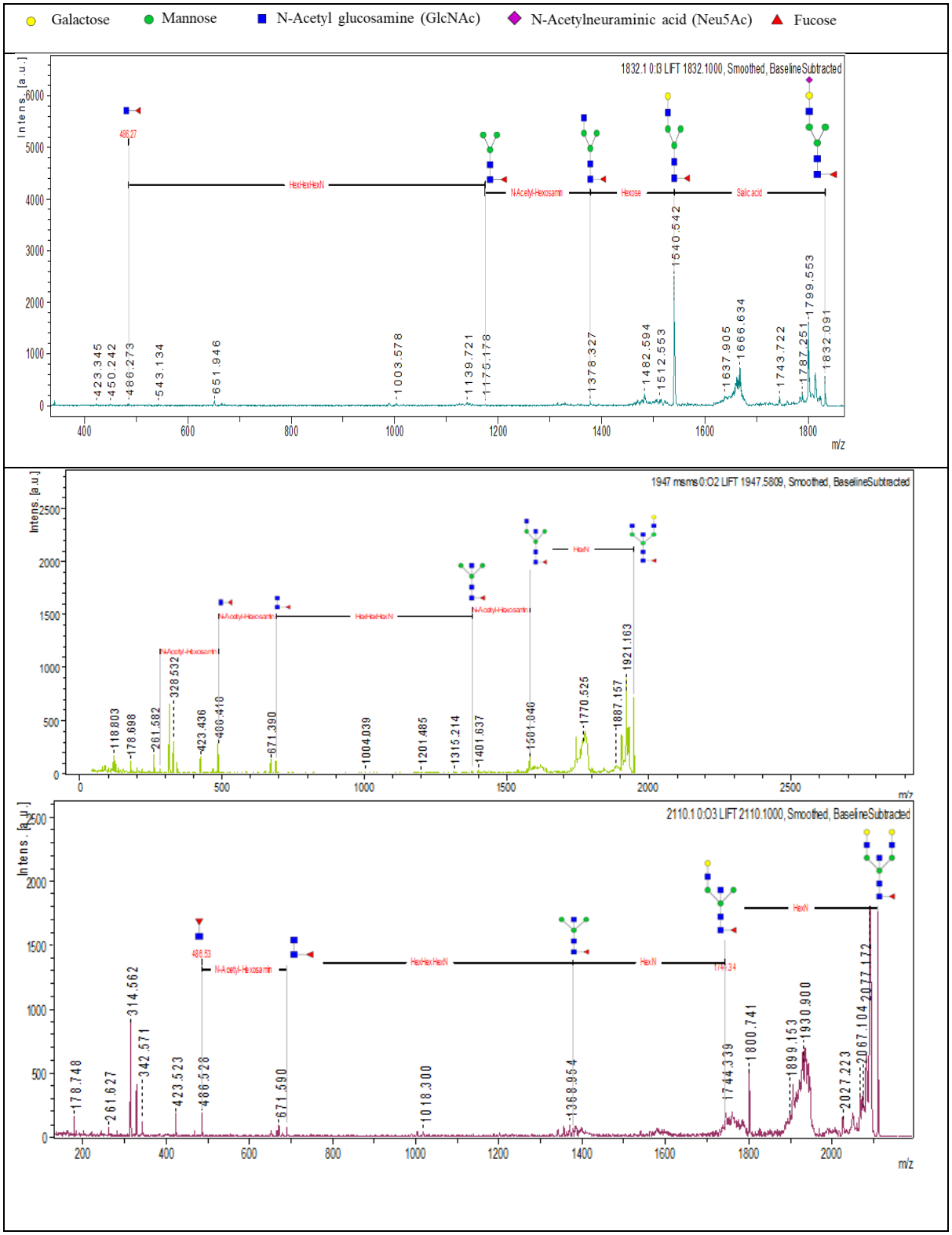
Glycan Id (HNAGF)	Cartoon structure	IgG1	IgG2B	IgG3	Glycan Id (HNAGF)
16	5 4 0 0 0		yes	no	no
17	5 4 0 0 1		yes	yes	yes
18	5 4 0 1 1		yes	yes	yes
19	5 4 0 0 2		yes	yes	yes
20	5 4 0 2 1		yes	yes	no
21	6 4 0 0 2		yes	yes	no
22	6 3 0 1 1		yes	no	no
23	6 4 0 1 1		yes	yes	no
24	6 4 0 1 2		yes	yes	no

NB:- The glycoforms are common in both baseline (B) and Inhibitor (I) developed sample. The main differences were in their intensities.

Appendix C



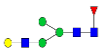
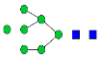
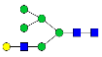
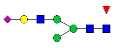
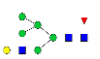
Appendix C F1 Re-confirmation of N-glycosite identification on AAV8 glycopeptide by MS/MS. The batch is expressed purified, processed and analysed independently from the previous analysis. a) CID spectra of the ^{18}O labeled glycopeptide b) CID spectra of nonglycosylated peptide in a single run.

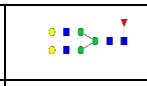
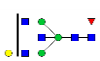
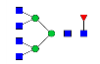
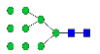
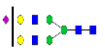
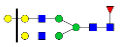
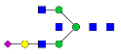
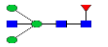
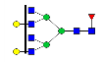
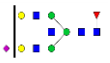
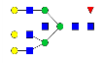
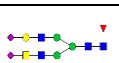
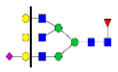
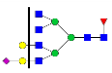
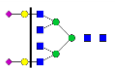
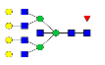
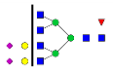


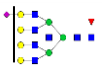
Appendix C F2 Structural confirmation of major glycoforms detected from the AAV8 secreted fraction

Appendix C Table 1 The Major glycoforms AAV8 samples.
The Major glycoforms are confirmed by MS/MS and minor species are identified by GlycoWorkbench and CFG database search.

Galactose Mannose N-Acetyl glucosamine (GlcNAc) N-Acetylneuraminic acid (Neu5Ac) Fucose

	Experimental m/z		Theoretical	Structure	Composition	Type
No	**Secreted**	**Intracellular**	**m/z)**			
1	1354.510	1354.458	1354.4789		H5N2	High mannose
2	1436.532	1436.587	1436.532		H3N4	Complex
3	1516.606	1516.618	1516.5317		H6N2	High mannose
4	1541.654	1541.576	1541.5631		H4N3F1	Complex
5	1558.431	1557.538	1557.5873		H5N3	Hybrid
6	1582.710	1582.658	1582.5872		H3N4F1	Complex
7	1597.537	1597.438	1598.5848		H4N4	Complex
8	1678.516	1679.583	1678.5846		H7N2	High mannose
9	ND	1686.738	1686.6009		H4N3S1	complex
10	1719.742	ND	1719.611		H6N3	Hybrid
11	1745.000	1744.814	1744.6427		H3N4F1	Complex
12	1760.959	1760.903	1760.6377		H4N4	Complex
13	1785.973	1785.924	1785.669		H3N5F1	Complex
14	1801.938	1801.882	1801.6642		H4N5	Complex
15	1833.035	1832.984	1832.5688		H4N3S1F1	Complex
16	1841.019	1841.849	1840.6304		H8N2	High mannose
17	1848.866	1848.956	1848.6537		H5N3S1	Hybrid
18	1865.875	ND	1865.669		H6N3F1	Hybrid
19	1881.924	ND	1881.664		H7N3	Hybrid
20	1890.984	ND	1889.680		H4N4F1	Complex

21	1907.067	1907.089	1906.696		H5N4F1	Complex
22	1948.124	1948.080	1947.722		H4N5F1	Complex
23	1989.172	1989.183	1988.749		H3N6F1	Complex
24	2003.070	2003.139	2002.690		H9N2	High mannose
25	2028.055	2027.958	2027.722		H7N3F1	Hybrid
26	2052.218	2053.157	2051.733		H5N4S1	Complex
27	2069.126	2069.191	2068.748		H6N4F1	Complex
28	2092.760	2094.338	2092.760		H4N5S1	Complex
29	2110.275	2110.234	2109.775		H5N5F1	Complex
30	2198.400	2198.377	2197.791		H5N4S1F1	Complex
31	2313.471	2313.569	2312.854		H5N6F1	Complex
32	2401.631	2401.603	2400.870		H5N5S1F1	Complex
33	2475.764	2475.78	2474.907		H6N6F1	Complex
34	2489.599	2489.867	2488.886		H5N4S2F1	Complex
35	2530.676	2530.964	2529.913		H4N5S2F1	Complex
36	2563.931	2563.828	2562.923		H6N5S1F1	Complex
37	2604.750	2604.881	2603.950		H5N6S1F1	Complex
38	2691.961	2693.006	2691.966		H5N5S2F1	Complex
39	2750.676	ND	2748.987		H5N6S2	Complex
40	2767.100	2767.024	2766.002		H6N6F1S1	Complex
41	2842.087	ND	2840.039		H7N7F1	Complex
42	2894.984	2895.584	2895.045		H5N6S2F1	Complex

43	2930.047	2930.072	2928.055		H7N6S1F1	Complex
44	3098.125	3097.795	3098.124		H5N7S2F1	Complex
45	3132.698	3132.202	3131.132		H7N7S1F1	Complex
46	3185.324	3186.702	3186.141		H5N6S3F1	Complex