Investigation of The Role of Glycan Binding on Human Norovirus Using Mass Spectrometry

Dissertation with the aim of achieving a doctoral degree at the Faculty of Mathematics, Informatics and Natural Sciences

Department of biology
of Universität Hamburg

submitted by

Hao Yan
born in Wuhan,China

Hamburg, 2021

Reviewers

Dr. Charlotte Uetrecht

Center for Structural Systems Biology (CSSB), Hamburg, Germany
Leibniz Institute for Experimental Virology (HPI), Hamburg, Germany
Deutsches Elektronen Synchrotron (DESY), Hamburg, Germany
European XFEL GmbH, Schenefeld, Germany
Faculty V: School of Life Sciences
University of Siegen, Siegen, Germany

Prof. Dr. Julia Kehr

Institute of Plant Science and Microbiology, Universität Hamburg, Germany

Disputation date: March 30th 2022
Submission of Thesis: Feb. 2021

The thesis is conducted from October 2015 to Spring 2020 under the supervision of Dr. Charlotte Uetrecht scientific group ‘Dynamics of Viral Structures’ in the Leibniz Institute for Experimental Virology (HPI). My second supervisor is Prof. Dr. Julia Kehr, Institute of Plant Science and Microbiology, the Department of Biology of Hamburg University.

Contents

Published Data

Chapter 1

Wegener, H., Mallagaray, Á., Schöne, T., Peters, T., Lockhauserbäumer, J., **Yan, H.**, Uetrecht, C., Hansman, G. S., Taub, S. Human norovirus GII.4(MI001) P dimer binds fucosylated and sialylated carbohydrates. *Glycobiology,* 2017
I involved in perform native mass spectrometry(MS) experiment and data analysis

Yan, H.*, Lockhauserbäumer, J.*, Szekeres, G. P., Mallagaray, A., Creutznacher, R., Taube, S., Peters, T., Pagel, K., Uetrecht, C. Protein secondary structure affects glycan clustering in native mass spectrometry. *Life,* 2021;*Volum*: 554
I worked in manuscript writing and performing the major part of experiment and data analysis

Chapter 2

Bücher, K. S., **Yan, H**., Creutznacher, R., Ruoff, K., Mallagaray, A., Grafmüller, A., Dirks, J. S., Kilic, T., Weickert, S., Rubailo, A. Fucose-functionalized precision glycomacromolecules targeting human norovirus capsid protein. *Biomacromolecules,* 2018;*Volum*: 3714-3724
I mainly involved in performing native experiment and data analysis as well as writing the corresponding part in the manuscript.

Chapter 3

Dülfer, J., **Yan, H.,** Brodmerkel, M. N., Creutznacher, R., Mallagaray, A., Peters, T., Caleman, C., Marklund, E. G., Uetrecht, C. Glycan-induced protein dynamics in human norovirus P dimers depend on virus strain and deamidation status. *Molecules,* 2021;*Volum*: 2125
I partially worked in manuscript writing and performing the native MS and HDX MS experiment as well as data analysis

Abstract.

Norovirus is the one leading cause of non-bacterial gastroenteritis in humans. The infection is relevant to glycan recognition during the host cell-virus interaction. P domain on the distal region of the capsid VP1 protein is known as model protein for investigating the role of glycan binding on virus infection. Several biophysics assays likely nuclear magnetic resonance (NMR) spectroscopy, surface plasmon resonance (SPR) and mass spectrometry (MS) are applied to determine the binding affinity and stoichiometry of carbohydrates of interest, e.g. histo-blood group antigen.

In the first project, native MS is applied to determine the binding strength between norovirus P dimer and blood group carbohydrates. Eight non-glycan binding protein candidates were chosen to evaluate the degree of glycan clustering. A galactose terminal non-binder ligand Gb4 is served as a negative control to compare with carbohydrate of interest. The distinct glycan clustering pattern on the reference candidates implies that the physicochemical feature of the protein affects glycan clustering during electrospray ionization via native MS. The detailed data analysis reveals the share of β-sheets structure in a protein has a strong influence on the degree of the glycan ligand clustering. Thus, large size protein with a high level of β-sheets component obtains strong ligand clustering. Thereby, a suitable non-glycan binding reference protein is pivotal in direct MS study and previous interpretation for binding constant of glycan ligands to the P dimers needs to be considered.

Fucose was previously identified as the minimal binding component during the norovirus cell attachment. X-ray crystallographic observation demonstrated that binding sites are located on the cleft region between two human norovirus (hNoV) P monomers. In the second chapter, in order to decipher whether P dimer exhibits a multiple-binding behaviour, a couple of multivalent fucose attached glycomacromolecules based on the solid phase polymer synthesis procedure are generated and applied to interact with the P dimer to achieve the global avidity. In the work, nine glycan linked structure varying from fucose/galactose number and position on the oligo backbone are constructed and

measured via native MS. The result indicates that the first generation of fucose containing structures are obtained similar binding affinity (near mM range), indicating only one fucose per glycomacromolecules occupied the binding pocket of the P dimers.

NMR and hydrogen deuterium exchange (HDX) MS determined that a post-modification site (N373) near to glycan binding region of norovirus GII.4 Saga P dimer was identified. In presence of blood group antigen carbohydrates, binding attenuation were observed on the deamidated P dimer species. In the third project, three norovirus strains, GII.4 MI001, GII.17 Kawasaki and GII.10 Vietnam206 P dimer proteins are expressed and measured using HDX MS in the presence of fucose and blood group antigen B tri-saccharides to examine whether the deamidation occurs on the P dimers of these species and how the glycan binding pattern behaves. After sequence alignment and peptide mapping. It showed that deamidated GII.4 MI001 P dimer can be detected after several months of storage in pH7. No deamidation was observed on GII.17 Kawasaki and GII.10 Vietnam206 P dimer, although the Vietnam one carries a glutamine (can be deamidated) at position 373 of glycan binding cleft similar as GII.4 Saga strains. Furthermore, a similar glycan binding dynamics was identified between GII.4 Saga and MI001 P domains due to structural and sequence similarity. GII.17 Kawasaki and GII.10 Vietnam species reflected to a different protection manner during H/D exchange. The distinct ligand occupation behaviours from different P dimer species indicate that glycan attachment induces a strain-specific manner among noroviruses.

In summary, native MS and HDX MS are applied to investigate the role of glycan binding on noroviruses. We demonstrate that non-specific clustering is controlled by intrinsic feature of the protein, the investigation of the natural glycans and designed glycomacromolecules binding on norovirus P dimer via native MS provides useful information for generating the novel anti-drug strategy to inhibit viral cell attachment. In contrast, the strain-dependent glycan attachment dynamics and deamidation process in the different norovirus strains determined from HDX MS indicates that viral particles would adapt specific infection strategy during host-virus interaction.

Zusammenfassung

Im Rahmen des ersten Projekts wird die Bindungsstärke zwischen dem Norovirus-P-Dimer und den Blutgruppenkohlenhydraten mit Hilfe der nativen MS bestimmt. Acht nicht-glykanbindende Proteine wurden für die Bewertung des Grades der Glykanclusterung ausgewählt. Ein Galaktose-terminaler Nicht-Binder-Ligand Gb4 wird als negative Kontrolle zum Vergleich mit dem zu untersuchenden Kohlenhydrat verwendet. Das deutliche Glykan-Cluster-Muster bei den Referenzkandidaten deutet darauf hin, dass die physikochemische Eigenschaft des Proteins die Glykan-Clusterbildung während der Elektrospray-Ionisierung über die native MS beeinflusst. Die ausführliche Datenanalyse ergibt, dass der Anteil der β-Sheets-Struktur in einem Protein einen erheblichen Einfluss auf den Grad der Ligandenclusterung der Glykane hat. Ein großes Protein mit einem hohen Anteil an β-Faltblättern führt zu einem starken Ligandenclustering. Somit ist ein geeignetes nicht-glykanbindendes Referenzprotein für eine direkte MS-Studie von entscheidender Bedeutung und die bisherige Interpretation der Bindungskonstante von Glykanliganden an die P-Dimere ist zu berücksichtigen.

Bislang wurde Fucose als minimale Bindungskomponente während der Zellanhaftung des Norovirus nachgewiesen. Die röntgenkristallografische Beobachtung ergab, dass sich die Bindungspunkte im Bereich des Spalts zwischen zwei humanen Norovirus (hNoV) P-Monomeren befinden. Im zweiten Teil der Studie wird untersucht, ob das P-Dimer ein Mehrfachbindungsverhalten aufweist. Mit Hilfe der Festphasen-Polymersynthese werden einige multivalente, an Fucose gebundene Glykomakromoleküle erzeugt und zur Bestimmung der globalen Avidität mit dem P-Dimer in Wechselwirkung gebracht. Dazu wurden neun glykanverknüpfte Strukturen mit unterschiedlicher Fucose/Galactose-Anzahl und Position auf dem Oligo-Rückgrat konstruiert und mit native MS analysiert. Aus dem Ergebnis geht hervor, dass die erste Generation der fucosehaltigen Strukturen eine vergleichbare Bindungsaffinität (im mM-Bereich) aufweist. Dies deutet darauf hin, dass nur eine Fucose pro Glycomakromolekül die Bindungstasche der P-Dimere einnimmt.

Durch NMR und Wasserstoff-Deuterium-Austausch (HDX) wurde eine Post-Modifikationsstelle (N373) in der Nähe der Glykan-Bindungsregion des Norovirus GII.4 Saga P-Dimers nachgewiesen. In Verbindung mit Blutgruppenantigen-Kohlenhydraten wurde eine Abschwächung der Bindung an die deamidierten P-Dimere beobachtet. Im dritten Projekt wurden drei Norovirus-Stämme,

GII.4 MI001, GII.17 Kawasaki und GII.10 Vietnam206 P-Dimer-Proteine dargestellt und mit HDX MS in Gegenwart von Fucose und Blutgruppenantigen B-Tri-Sacchariden gemessen, um zu untersuchen, ob die Deamidierung an den P-Dimeren dieser Arten erfolgt und wie das Glykan-Bindungsmuster reagiert. Nach Sequenzabgleich und Peptidzuordnung. Das Ergebnis ergab, dass das deamidierte GII.4 MI001 P-Dimer nach mehrmonatiger Lagerung bei pH7 nachgewiesen werden kann. Bei den Dimeren GII.17 Kawasaki und GII.10 Vietnam206 P wurde keine Deamidierung gefunden, obwohl das Vietnam-Dimer ein Glutamin (das deamidiert werden kann) an Position 373 des Glykanbindungsspalts trägt, ähnlich wie die GII.4 Saga-Stämme. Zudem wurde eine vergleichbare Dynamik der Glykanbindung zwischen den beiden Domänen GII.4 Saga und MI001 P aufgrund von Struktur- und Sequenzähnlichkeit festgestellt. Die Arten GII.17 Kawasaki und GII.10 Vietnam zeigen ein unterschiedliches Schutzverhalten während des H/D-Austauschs. Das unterschiedliche Ligandenbesetzungsverhalten der verschiedenen P-Dimere deutet darauf hin, dass die Glykanbindung bei Noroviren stammspezifisch ist.

Native MS und HDX MS dienen der Untersuchung der Rolle der Glykanbindung an Noroviren. Wir belegen, dass die unspezifische Clusterbildung von einer intrinsischen Eigenschaft des Proteins bestimmt wird. Durch die Untersuchung der natürlichen Glykane und der entwickelten Glykomakromoleküle, die über die native MS an das P-Dimer des Norovirus binden, werden nützliche Informationen für die Entwicklung einer neuartigen Antidrogenstrategie zur Hemmung der viralen Zellanhaftung gewonnen. Demgegenüber lässt die stammabhängige Dynamik der Glykananbindung und des Deamidierungsprozesses in den verschiedenen Norovirenstämmen, die mittels HDX-MS bestimmt wurden, darauf schließen, dass die viralen Partikel eine spezifische Infektionsstrategie während der Interaktion zwischen Wirt und Virus verfolgen.

Introduction: investigation of norovirus-glycan interaction via mass spectrometry

1.1Norovirus characteristics and evolution

Norovirus， rotavirus and gram-negative bacteria likely Campylobacter as well as some parasites, amoeba and fungus are the common cause of gastrointestinal infection(Ahmed et al., 2014). The symptom might include vomiting, diarrhea, abdominal cramping and fever. As one of the leading pathogens for acute gastroenteritis, noroviruses outbreak and annual sporadic reports worldwide have been continuously recorded in adults and especially children under 5-year-old in recent decades(Ahmed et al., 2014; Baker et al., 2010). The estimated 20% of overall gastrointestinal infectious events are identified as norovirus infection depending on epidemiological surveillance among more than 20 countries (Ahmed et al., 2014). The infectious signs often endure three to five days whereas young children and elderly group suffer a longer time period of illness, which severe cases can lead to death.

Norovirus caused gastroenteritis, previously known as winter vomiting disease, was primary documented in 1968 due to an outbreak occurred at a town called Norwalk in Ohio, United States(Maunula et al., 2005). As the first identified human infected viral agent causing gastroenteritis, Norwalk virus was sequenced and found to be associated with calicivirus family(Glass et al., 2009), this Norwalk agent is currently categorized into norovirus genogroups I. Noroviruses are small positive-strand RNA viruses belonging to the Caliciviridae family, it forms non-enveloped icosahedral viral structure ranging from 20-40nm diameter size from (Kapikian et al., 1972). Based on its VP1 sequence, ten genogroups have been determined (Ahmed et al., 2014; Bartsch et al., 2016; Chhabra et al., 2019; Glass et al., 2009). The great diversity in noroviruses consisted of more than 25 different genotypes and numerous subgroups, Norwalk virus, the first discovered strain, was entitled in genogroups I, genotype1 (GI.1). It revealed that GI, GII, GIV, GVIII and GIX infect human(Chhabra et al., 2019; Tan & Jiang, 2010). Animal noroviruses are then subdivided into other genogroups with different host species including pigs, cattle and sheep , and mice, respectively(Tan & Jiang, 2010). Although the GII genogroup

dominated majorities of epidemic and sporadic acute gastroenteritis cases(Bull et al., 2006), the identification of positive samples from wastewater and river monitor collections in Japan and Italy associated with sero-prevalence evidence suggested the emergence of genogroups IV (GIV) might be correlated to the feline and canine norovirus strains in the same genogroups (Di Martino et al., 2014; Kitajima et al., 2010; Nayak et al., 2008). The overlapped living environment between human and domestic pets, especially dogs and cats enhanced the zoonotic transmission possibility, indicating a potential norovirus evolution route. Furthermore, genome sequencing along with phylogenetic analysis on various strains revealed that viral RNA recombination and the accumulated nucleotide mutations accumulation is the main driving force for the high diversity in noroviruses(Bull et al., 2007; Nayak et al., 2008). The increasing outbreak reports evidence and the awareness of viral strains evolution in the genus Norovirus of Calicivirus family from worldwide sequencing analysis also indicating a continuous surveillance is urgently required in norovirus prevailed regions.

Baculovirus expression system was introduced to generate norovirus virus-like particle(VLP), Jiang first reported the successfully produced capsid proteins(Jiang et al., 1992). The expressed protein based on the second and third open reading frames of viral genome and self-assembled to form an empty VLP. Electron-microscopy observation indicated that the VLP presented a similar size and appearance to the native capsids, the antigenicity of the particles appeared to be identical from wild-type virions in biological examination(Jiang et al., 1992). The available recombinant norovirus VLPs facilitated the development of rapid diagnostic agents and techniques. This result also attracted scientists' attention to study of viral structure and host response. First x-ray crystal of norovirus capsid was interpreted in 1999 by B. V. V. Prasad and co-workers(Prasad et al., 1999). The particle exhibit a T =3 icosahedral symmetry.

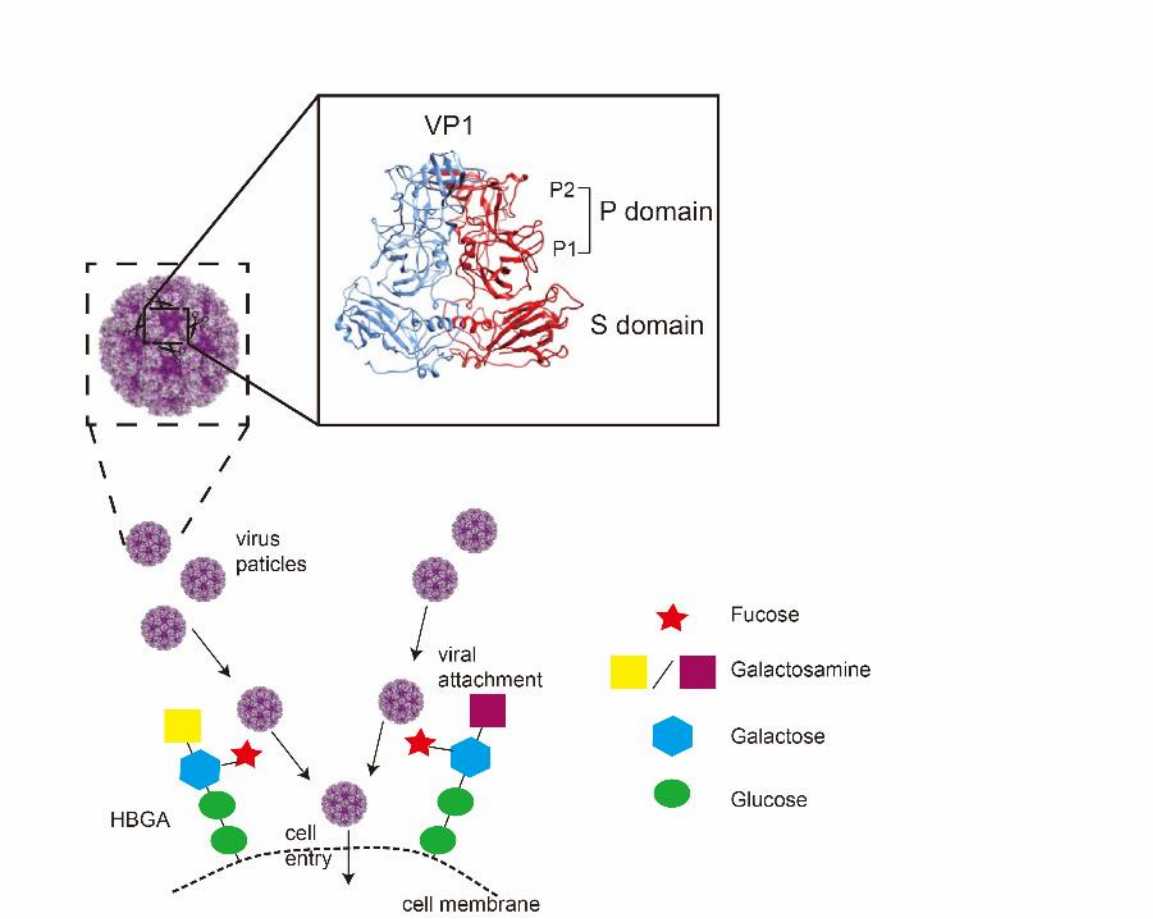


Figure1.1 schematic picture of the norovirus cell attachment. The virus recognizes the glycans of interest on the cell membranes. The PDB structure of protruding (P) domain is zoomed to demonstrate the P domain and S domain(Mallagaray et al., 2019)

It consists of 180 capsid protein units assembled into 90 dimers (Figure 1). The main capsid protein VP1 can be divided into two parts, shell (S) and protruding (P) domain(Prasad et al., 1999). Two subdomains can be further differed from P domain, in which P2 subunit, the hypervariable region of P, recognizes carbohydrates on the cell membrane. One monosaccharide, fucose, was determined as the minimal binding elements for capsid protein and norovirus particle from STD-NMR(Brigitte et al., 2012). Histo-blood group antigens (HBGA) and certain sialic acid terminal saccharides were defined as cell attachment receptors for the viral infection(Han et al., 2013; L. Han, M. Tan, et al., 2014; Wegener et al., 2017). Here, blood types are determined by the different carbohydrates linkage and structural properties of oligosaccharides expressed on the red blood cells, the discovery and explanation of ABO Blood group system were conducted by an Austrian pathologist, Karl Landsteiner at early 20th century(Yamamoto

et al., 1990). These antigens are also detected to exhibit on epithelial cell membrane in digestive, respiratory and urogenital tract mucosa(Séverine Marionneau et al., 2001). The biosynthesis of these carbohydrates structure depends on *FUT2* and *FUT3* genes with a multiple step process (Figure 1.2).

Over past three decades, norovirus GII.4 attributed to outbreak of viral gastroenteritis worldwide (Ahmed et al., 2014; Lu et al., 2015; Motomura et al., 2008; Siebenga et al., 2007; Van Beek et al., 2013). Every the outbreak season predominated by a novel epidemic variant with slightly altered HBGA glycan-receptor binding profile. It is notable that different norovirus strains possessed distinct binding preference regarding to ABO blood group, non/- secretor and Lewis antigen types(Huang et al., 2003). Human norovirus infection challenge model directly revealed that host resistance and susceptibility to the virus infection was associated with blood group antigen carbohydrate patterns (Hutson et al., 2002; Lindesmith et al., 2003; Thorven et al., 2005). Similar evidence presented among viral gastroenteritis epidemiology and structural investigation (Han et al., 2012; Tan et al., 2008; Wegener et al., 2017). Gene polymorphism characterization test also implied that ABH antigen might affect norovirus susceptibility. It implied that norovirus evolution might be driven by P domain-blood group glycan interaction(Nilsson et al., 2003).

1.2 Studies of P domain

Expression of the P2 subdomain folds constantly into dimer. The glycan binding pockets are observed inside the β-strand abundant cleft region between two recombinant P monomers(Mallagaray et al., 2015; Tan et al., 2004). Comparatively, Tan et al.,(2004)reported that using cysteines contained peptide to link the terminal amino acids of P domain, which aims to build the intermolecular disulfide bond to form P particle. The successfully produced particle is constituted by 12 P dimers (24-mer) with almost *T*=1 icosahedral organization. Electron micrograph of P particle demonstrated that P2 subdomain sits on the outer layer with an inner core assembled by P1 subunits (Tan & Jiang, 2005b). They reported that these 24-mer particles retained a dramatic

enhanced HBGA binding capability (>700-fold) during the binding affinity test(Tan & Jiang, 2005a, 2005b). However, electrospray mass spectrometry analysis of HBGA binding to the P dimer, 24-mer particle and norovirus VLP revealed that the affinity of A and B blood type carbohydrates remained in a similar mM range value to the three different constitutes (Han, E. N. Kitova, et al., 2014; Han et al., 2012; L. Han, E. N. Kitova, et al., 2014; Han et al., 2018). The outcome doubted the use of P particle as norovirus vaccine candidate. *In vivo* mice immunization injection experiment suggested that P particle only yielded low level norovirus specific IgG antibody while a single VLP inoculation can induce a high level IgG secretion in mice blood(Tamminen et al., 2012). Therefore, Whether P particle is ideal for vaccine study requires further evaluation.

In contrast, scientist found that Norovirus T =3 icosahedral symmetry organization enables S domain to bury into icosahedral contact and adapt the P domain as dimer(B. Prasad et al., 1994). The 90 dimeric P subunits cover the outer layer of capsid surface, which, in turn, is responsible for a variety of viral properties including host immune response, infectivity, tropism, host cell attachment, structural integrity. Therefore, if we can decipher the structural and functional identity of VP1, chiefly the P2 distal region, it would assist to understand the mechanism of norovirus cell entry and develop novel antiviral agents for norovirus therapy. A deletion of Norwalk virus (GI.1) VP1 N-terminal 20 residues does not affect the capsid protein to self-assemble into 38nm VLP, whereas a continuous amino acids depleting mutation in N-terminal region alters the assembly mechanism examined via X-ray crystallography(Bertolotti-Ciarlet et al., 2002).

1.3 Norovirus cell entry and receptors

The lack of applicable carbohydrates cell attachment factors in intestinal cells hinders the growth of viral particle *in vitro* assay. The cell culture of norovirus in transformed intestinal epithelium or human macrophages and dendritic cells have been proved to be a failure (Duizer et al., 2004; Lay et al., 2010). Recent progress demonstrated that human stem cell-derived, non-transformed enteric epithelium is available for the norovirus capsids expression(Ettayebi et al., 2016). The cell is isolated from intestine.

The generated monolayer intestinal enteroids consist of various cell types including enterocytes, goblet, enteroendocrine and Paneth cells(Ettayebi et al., 2016), in which certain cells with appropriate HBGA group support the norovirus replication. Recent progress showed that virus cultivation is available in human intestinal enteroid (Duizer et al., 2004; Ettayebi et al., 2021).

Murine norovirus (MNV) were discovered in immunocompromised mouse(Karst et al., 2003). The small animal model and MNV offers the great opportunity to investigate host response and identify the immune components involved in the norovirus infection even though the human norovirus cell cultivation system is limited. MNV replicate in macrophages and dendritic cells and the viral spread also propagates at mouse intestine peripheral tissues after the oral inoculation (Changotra et al., 2009; Karst et al., 2003; Mumphrey et al., 2007). Human norovirus gastroenteritis occurs at first 24 hours (hrs) after the infection. The symptom of vomiting and diarrhea among patients persist another 24 to 48 hrs. MNV replication is observed after 6 hrs viral infection. The capsid proteins translation begins after the nonstructural protein expression during the next 6 hrs via quantitative reverse transcription-PCRs examination, this phenomenon is consistent to the infection period of human norovirus gastroenteritis, In addition, the type I transmembrane protein CD3001f was recently identified as the MNV receptor(Haga et al., 2016), which provides new insight on how the virus engage the capsid protein and assist to learn the virus cell entry. Human noroviruses are known to interact with histo-blood group antigens and this interaction is important for the virus infection, but the regarding cell receptor is sofa unclear. Thus, MNV is still considered to be an ideal surrogate to compare human norovirus infection situation and explore the pathogenesis during infectious process.

It is well known that Human noroviruses bind to blood group antigen attached on the intestine cell membrane, However, the low oligosaccharides binding affinity at mM range(Mallagaray et al., 2019) implies that another unknown co-receptor might be prerequisite for viral cell entry and uncoating. The hindrance of unavailable reproducible cell culture system for human norovirus again limits the identification of specific receptor via classical biological methodology. To understand the norovirus cell entry and

uncoating, A model system related to norovirus would be feline calicivirus(FCV)(Makino et al., 2006), this virus receptors were fully studied, sialic acid containing glycomolecules, particularly, 2,6-linked structure on the upper respiratory epithelial cell membrane is used for FCV attachment and the junction adhesion molecule 1 (JAM-1)(Stuart & Brown, 2007), belonging to Ig superfamily, was characterized as the cellular binding receptor for the viral infectious entry(Kreutz et al., 1994; Makino et al., 2006; Stuart & Brown, 2007; Tan & Jiang, 2010). The collaboration of multiple interaction steps of sialic acid and immunoglobulin-like protein for internalization is not sufficient to explain the virus uncoating. A near atomic-resolution of FCV capsid from cryo-electron microscopy presented that a conformation of VP2 in FCV was observed with portal-like structure with distinctive three-fold axis organization during the JAM-1 binding(Cannon et al., 2006; Makino et al., 2006; Stuart & Brown, 2007). This structural rearrangement upon receptor engagement implies that viral genome might pass through endosome membrane via the VP2 portal-like assembled channel after the capsid protein and JAM-1 interaction(Conley et al., 2019). Unlike Human Norovirus, MNV infects macrophages and dendritic cells(Karst et al., 2003; Mumphrey et al., 2007), the virus shedding process is markedly abrogated after the addition of CD300lf antibody in the medium of murine cell line while the engineered CD300lf co-expression HEK293 cells enable to produce MNV particles and release of viral progeny(Haga et al., 2016). The change of cell susceptibility to the infection confirms that CD300 family molecules, CD300lf , is one of the receptor for MNV cell entry (Haga et al., 2016). MNV infection is pH-independent whereas viral uncoating initiation often occurs at low pH environment in endosome, comparatively FCV is sensitive to the pH value and a dramatically reduction of infectivity were observed in extracellular low pH environment(Cannon et al., 2006). The tropism difference might be a clue for the explanation that FCV attaches to the respiratory tract. MNV is similar to human norovirus, before the virions infect intestine cell and trigger the viral replication, the particles require for enduring stomach low pH circumstances. The pH-insensitive character has been identified in many viruses causing gastroenteritis (Brandenburg et al., 2007; Golden & Schiff, 2005; Pérez & Carrasco, 1993; Regan et al., 2008; Sánchez-San Martín et al., 2004; E. Sun et al., 2013). Taken together, the viral cell entry process studied on MNV and FCV is a straightforward path for the investigation on identification of unknown human norovirus cell receptor.

1.4Norovirus carbohydrates recognition

Carbohydrates receptor is now regarded as a key feature for study of norovirus. animal infected noroviruses prefer to bind neuraminic acid terminal carbohydrates in the first cell entry step(Perry et al., 2009; Stuart & Brown, 2007; Tan & Jiang, 2010; Taube et al., 2009), except for bovine norovirus recognize blood group antigen(Zakhour et al., 2009), human noroviruses were thought to have rather a high binding affinity to HBGA group ligands in previous reports (Han, E. N. Kitova, et al., 2014; Han et al., 2012; Mallagaray et al., 2015; Tan & Jiang, 2005b, 2010; Wegener et al., 2017), recent glycan binding experiment suggests that HBGA is not a strong binder(mM range)(Dülfer et al., 2021; Mallagaray et al., 2019; Yan et al., 2021)

As mentioned above, HBGA is considered as a key factor during norovirus infection, many studies turn to understand the structural properties of blood group antigens. The biosynthesis of these carbohydrate structures depend on *FUT2* and *FUT3* genes with a stepwise process (Figure 1.2). The disaccharide precursor, O-ß-D-Galactopyranosyl-(1-3)-N-acetylglucosamine (Galβ1-3GlcNAc) sequentially adds with specific monosaccharide delivered by glycosyltransferase. Secretor and non-Secretor gene encode two different fucosyltransferase, *FUT3* encoded carbohydrate transferase attach e one Fucose substrate with α-1,3 or α-1,4 linkage on the acetylglucosamine of precursor molecule, the developed trisaccharide structure is named Lewis A antigen (Le^a) while individual containing *FUT2* gene translates another type of fucosyltransferase which catalyzes fucose to form a α-1,2 linkage on the precursor ß-Galactose molecule. The formed carbohydrates called H type 1 antigen. It is shown that persons with *FUT2* inactivated phenotype, namely non-secretor or secretor negative, are less susceptible to human norovirus infection.

The resistance issues have been reported in many epidemiological studies.(Costantini et al., 2015; Frenck et al., 2012; Jin et al., 2013; Nordgren et al., 2013; Tu et al., 2017; Zhang

et al., 2015), meanwhile, Incubation of Le^a antigen with VLP revealed no or extremely weak binding indicating these group of people might carry genetically advantage toward the norovirus infection(Harrington et al., 2002; Marionneau et al., 2002). However, a norovirus survey based on the inhabitants in Sweden claimed that norovirus infection is regardless of non-/secretor genetic difference(Nordgren et al., 2010). The resistance found in previous surveillance studies is mainly located in Asia. The geographical ABO system genetic mapping analysis demonstrated that Asian population contains a higher volume of A385T mutation on FUT2 exon where this missense alternation attenuates secretor phenotype. Therefore, individuals who are weak secretor positive type or non-secretor carrier in the area present no disease symptom during the norovirus induced gastroenteritis outbreak, the following study revealed that majorities of infection cases are dominated by GII.4 variants infection while norovirus infection history is highly genotype diversity. Based on these ambiguous data, system review analysis has attempted to explain the contradicted information. *FUT2* secretor negative carriers often present a nonsense G428A mutation. This is commonly found in European and individual possessing African nationally background. Approximately one fifth individuals in this characterized group are $FUT2^{-/-}$ homozygous dominant genotype(Nordgren & Svensson, 2019). H type 1 trisaccharide cannot be generated on their red blood cell membrane. The homozygous recessive persons should be protected from norovirus infection however one non-secretor was observed with asymptomatically illness affected by GII.4 strain infection during the challenge study(Frenck et al., 2012). Researchers might overemphasize the genetically protective effect of non-secretor character. Two patients from Spain and Burkina Faso who are FUT2-/- genotype were infected with typical vomiting and diarrhea symptom (Carlsson et al., 2009; Nordgren et al., 2013). These cases are retrieved from previous outbreak report. The absence of *FUT2* encoded fucosyltransferase basically enables non-secretor individuals to protect from norovirus infection. Practically, the resistance to the viral gastroenteritis might be merely norovirus genotype or strain dependent.

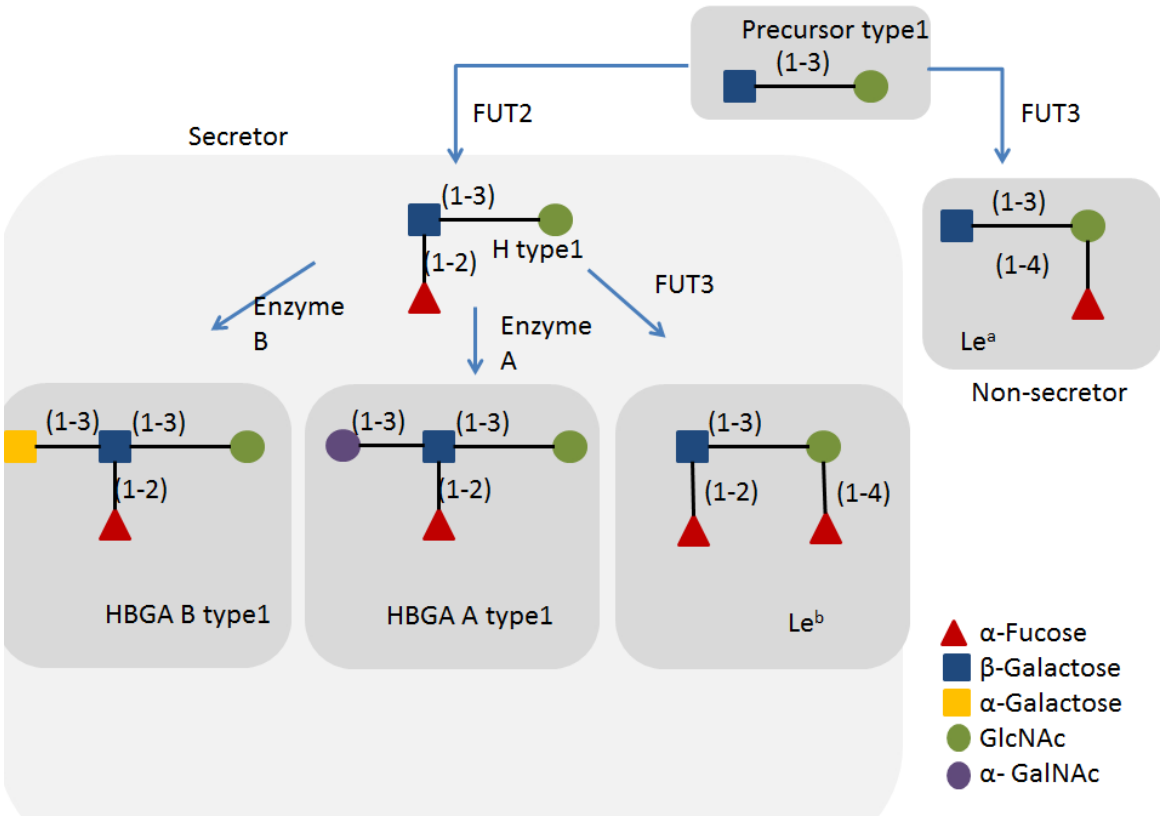


Figure1.2. The biosynthesis of the human ABH glycan according to the type 1 precursor. Histo blood group antigens(HBGAs) generated from a disaccharide precursor1 with addition of specific monosaccharides triggered by certain glycosyltransferase. The enzyme, FUT3(α-1,3 fucosyltransferase), catalyze the reaction to add a fucose monosaccharide to the disaccharide precursor with a-1,4 linkage, forming a trisaccharide, namely Lewis a (Le^a). In contrast, FUT2(α-1,2 fucosyltransferase), can add a fucose monosaccharide to the precursor with a-1,2 linkage, resulting in another type of trisaccharide called H type 1. H type 1 is a precursor for the two types of tetrasaccharides, HBGA A and B type1 and Lewis b (Le^b). Further addition of N-acetylgalactosamine (GalNAc) or a galactose at a-1,3 linkage of the H type 1, respectively, which catalyzed by two glycosyltransferases(Enzymes A, B) generates the blood group antigen A and B while FUT3 will trigger the reaction at a-1,4 linkage of H type 1 to produce Le^b

Following with the H type 1 antigen carbohydrate structure, Lewis b (Leb) antigen as well as A and B type are tetrasaccharides by addition of one more fucose into H type 1 structure with different linkage. FUT3 enzyme catalyzes α-1,4 linkage fucose monosaccharide to acetylglucosamine molecule again, Enzyme A and B provide N-acetylgalactosamine or a galactose to the terminal of trisaccharide structure with a a-1,3 linkage, respectively. A Le^b and B Le^b pentasaccharides are produced via another fucose delivery catalyzed by FUT3 encoded glycosyltransferase (stepwise carbohydrates biosynthesis process were shown below in figure1.2. Apart from HBGAs are on the cell surface of mucosal epithelium and red blood corpuscles, different blood group antigens as well present as free carbohydrates in breast milk, blood fluids, saliva on the oral

mucosa and intestine content(Dabelsteen, 2002; Tan & Jiang, 2005a). As discussed above, non-/secretor carbohydrate profile in some extent determined the norovirus infectivity, In-vitro experiments observed that ABO ligands and Lewis antigen represents different binding affinities(Han et al., 2012; Han et al., 2018; Tan et al., 2004; Wegener et al., 2017). Due to HBGA antigen is closely related to human blood type. Scientists speculated that the recently predominated GII.4 strains prefer to infect individual with secretor gene type. During the outbreak in China, Burkina Faso and Israel, the highest infection risk based on patients' blood type varied from A, AB to B, respectively(Klement & Halperin, 2008; Nordgren et al., 2013; Tan et al., 2008). However, Klement and Halperin also appealed that there is no association existed between blood type and the susceptibility to the norovirus infection during another norovirus infection outbreak which they surveyed. Similar result has also been collected during the outbreak in South Caroina, United States (Klement & Halperin, 2008; St. Clair & Patel, 2008).

1.5 Mass spectrometry

The basic mechanism of mass spectrometry (MS) was originated from a description of radiation experiments conducted by German physicist Eugen Goldstein in 1886, who described his new identified radiation "Kanalstrahlen"(El-Aneed et al., 2009), so called anode ray or positive ions, can pass through in electrical discharge glass tube during gas phase situation with a holed cathode in the opposite position. Another physicist, Wilhelm Wien, firstly illustrated that magnetic fields is available to affect the direction of positive ions and described the mass-to-charge ratios (m/z)(Grayson, 2002). The prototype of mass spectrometer was designed and developed by Professor J.J. Thomson in early 1910s(El-Aneed et al., 2009). The invention of MS and the subsequent instrumentation improvement provides a powerful device with high sensitivity and specificity in analytical chemistry in the rest of 20th century and the present. The application of this technique ranges from various scientific and applied technology field, for instance structural biology, Space Science and Astrobiology, Metabolomics, pharmaceutical screening, environmental contamination determination and food safety evaluation. It measures m/z of charged analyte(s) and enable to determine the exact

molecular weight of ionized molecule in a sample. The mass spectrometer generally consists of three main parts: ion source, analyser and detection system. The development of this analytical technique and the novel design on the instrument extends the application area from the identification of chemical properties of organic compounds at the beginning to biological macromolecules in the recent decades.

1.6 Ionization

The invention of electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) broadens the mass spectrometer ability on analysis of life science and molecular biology field. Ion source section equipped at the front part of the spectrometer where chemical substances in a sample were charged and transferred into the mass analyser. Conventional electron ionization (EI) was first ionization technique designed for mass spectrometry analysis(El-Aneed et al., 2009). The energetic electron beam were produced by tungsten filament with a magnetic field to allow the electron intersect with analyte in a certain angle inside the ion source where daughter ion is generated upon ionization and collision. This hard ionization method is widely used for the determination of chemical bond via the characterization the m/z of the fragments. However the limitation of low molecular weight range and high internal energy during the ionization impede the use on biomolecules analysis. Many other ionization types were then developed such as Atmospheric-pressure chemical ionization, Fast Atom Bombardment, Desorption Chemical Ionization. Until ESI and MADLI were introduced to connect with time of flight(TOF) mass analyser, intact protein complexes can be applied to MS analysis(Fenn et al., 1989). Electrospray ion molecule activation is characterized by applying a high voltage electrical field (1-3kV) to the liquid with biomolecule of interest, the liquid fluid will pass through the gold coated capillary to the sharp tip of a Taylor cone where aerosol spray formed. Under the atmospheric pressure, the relatively ‘bigger’ droplet shrinks when volatile buffer solution desolvation occurs. This extremely fast evaporation process (approximately 70 μs) provides a gaseous state to the targeted biomolecules and finally only the charged molecules enter into the mass analyser (Figure 1.3).

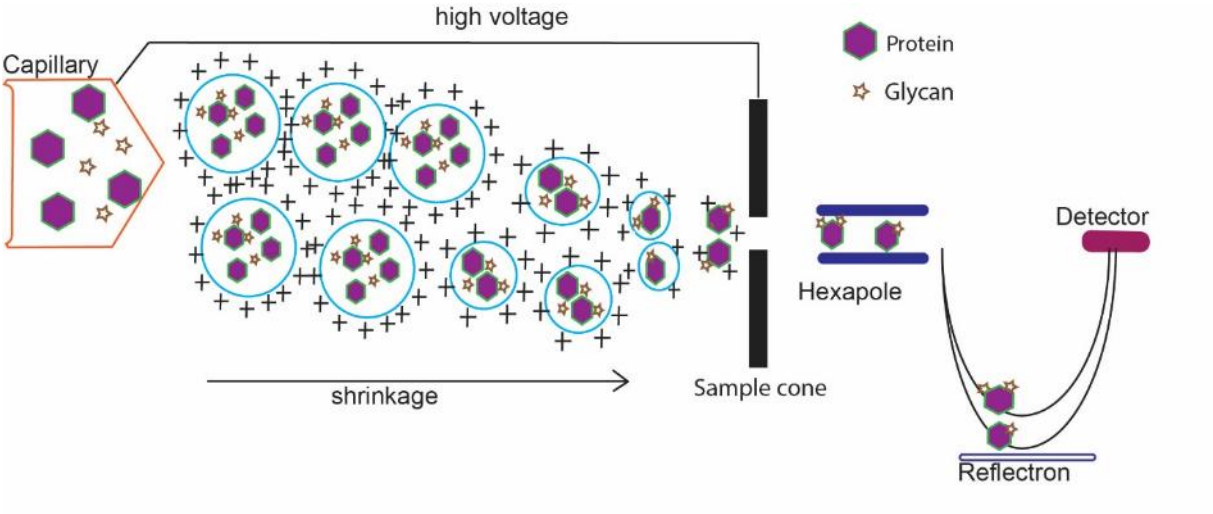


Figure 1.3: Schematic view of ESI process in positive ion mode with LCT mass spectrometer(ToF). The protein and glycan samples were mixed and injected into gold-coated capillary, the charged 'big' droplets were created and sprayed out from the opened tip of capillary. Solvent evaporation and droplet surface tension force to reduce the size of the droplet in the atmospheric pressure condition until the solvent evaporated/desolvated and only the ionized protein and glycan molecules left and then went to the sample cone of the mass spectrometer.

1.6.1 Electrospray ionization mechanism

The hypothesis of ESI basic mechanism mainly existed two different mode observed during the ESI experiments namely ion evaporation model and charged residue model. As described above, the essence of ESI process aims to transfer the biomolecule from liquid environment to gaseous state with minimal internal energy alternation. It is well known that, during the evaporation, the equilibrium collapsed when electrostatic force caused by outer surface attached ions overwhelms the strength of droplet surface tension, daughter droplet with smaller size formed and this instability was described as Rayleigh limit(Rayleigh, 1882), which the value can be quantified by droplet radius R and charges of droplet q. the equation(eq) 1.1 is written as below: where γ and ε equal to surface tension of the droplet and permittivity, respectively(Taflin et al., 1989)

$$q^2 = 64\pi^2\gamma\epsilon R^3 \qquad \text{Eq1.1}$$

It claimed that low molecular weight of biomolecules undergoes in ion evaporation mechanism (IEM)(Figure 1.4A), in which small droplet with the molecule is emitted from the multiple charged droplet when electrical field strength overcomes the surface tension. The molecule of interest moves to the edge of the droplet with random charges and this offspring droplet evaporated again until it transfers into the analyser. The characterization of this model is often influenced by chemical property of ion. The offspring droplet in IEM is commonly smaller than the one in charge residue model (CRM). Large protein complex and non-covalent bound biomolecules are frequently observed according to charge residue theory during the ESI process (Figure 1.4B). In CRM, the electrospray ejected certain size of nanodroplet with numerous charges. Theoretically, it assumes that only one biomolecule exists in a nanodroplet during the desolvation process, the solvent then evaporates and nanodroplet continuously declusters until only this ion attached biomolecule enter into the instrument. Meanwhile, during the droplet shrinkage, residues dissociation does not or rarely occur, in which it maintains non-covalent interaction state during the ionization process. Chain ejection model (CEM) was introduced to explain the emission of unfolded proteins and large polymers(Consta & Malevanets, 2012; Metwally et al., 2018)(Figure 1.4C). The compact structure of folded protein force the hydrophobic amino acids embedded into the inner part of complex while unfolded or denature one does not sustain tertiary and quaternary structure in a droplet, the exposed hydrophobic residues interact with solvent polar molecules which allows the unfolded protein ejected out the droplet in an extended shape.

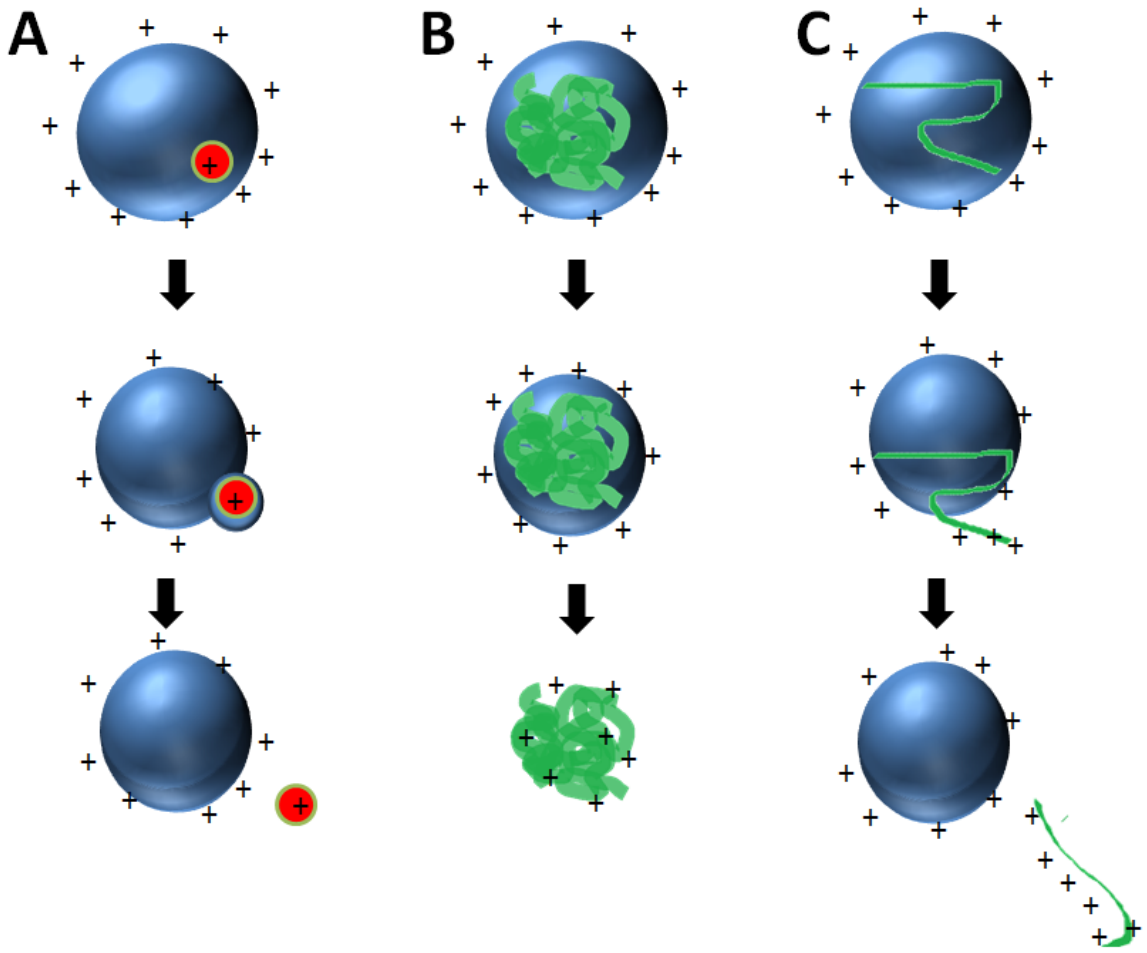


Figure 1.4: Schematic presentation of three ion molecule release mechanism during ESI process in positive ion mode. A:ion evaporation model. B: charged residue model and C chain ejection model.

Owing to low sample consumption limitation during the study in life science field, nano-flow electrospray ion source (nES) has been invented. The approximately 100 µm diameter tip for normal electrospray ion source was taken placed by gold coated borosilicate capillary with opening tips less than 10µm(Wilm & Mann, 1996). Therefore, the smaller droplets formed at the top of capillary tip are available. The high tolerance to salt and pH value as well as sustainable ionization efficiency of nES supplies accurate mass measurements in protein to protein, protein to ligand non-covalent interaction(Loo, 1997; Wilm & Mann, 1996)

Quadrupole time-of-flight (Q-ToF) mass spectrometer has been then widely spread in native MS measurement after the combination of nES. Except from that nano-spray

enhanced both positive and negative ionization efficiency, the development on m/z range extension in quadrupole allows to detect the signal of large protein complex (Lorenzen et al., 2007; Rostom & Robinson, 1999; Sobott & Robinson, 2004). in particular, HPV VLP with a molecular weight around 20MDa(Snijder et al., 2013). Other MS based method reaches to define an even higher molecular weight particle, for example, the detection of 50MDa bacteriophage P22 via charge detection mass spectrometry(Keifer et al., 2016). Recently, MS measurement with nanomechanical resonators is available to identify DNA-filled bacteriophage T5 capsid with a mass up to 105MDa(Dominguez-Medina et al., 2018). On the other hand, alternative devices likely Fourier transform-ion cyclotron resonance (FT-ICR) and recently high m/z range modified orbitrap have accessible in study the stoichiometry and assembly/disassembly of protein complex e.g. virus-like nanoparticles in high resolution and sensitivity mode (Snijder et al., 2014)

1.7 Native MS

Biological macromolecule particularly proteins is rarely analysed via MS before the introduction of novel 'soft' ionization method, namely electrospray ionization (ESI) and matrix-assisted laser desorption/ionization(MALDI), details in chapter soft ionization. Since late 1980s, scientist started to probe analysis of small molecules and protein interaction in the gas phase. The observation of the remaining associated structures between ligands and proteins as well as the MS instrumentation improvement enabled the analysis in terms of protein-protein and protein-ligands interaction studies. The gentle sample transportation technique has been gradually accepted as a promising technique for deciphering non-covalent protein-protein, protein-ligands interaction, especially integrated with other methodology including X-ray crystallography, electron microscopy (EM), small-angle X-ray scattering (SAXS), NMR spectroscopy and *Cryogenic* electron microscopy (cryo-EM) which can provide both atomic structural model and comprehensive interaction properties information.

In native MS process, bio-macromolecules are charged and transferred into gas phase with rarely chemical bond breakage to measure m/z in electrical or magnetic field. The ionization procedure affects non- or rare residue fragmentation of the tertiary and quaternary structure of biomolecules, it enables to maintain non-denatured state of protein complex and non-covalent interaction during the measurement. Native setting were firstly defined as biomolecules structure retains under pseudo-physiological state in 2004(van den Heuvel & Heck, 2004), biochemists has already probed intact protein complex by MS since the establishment of ESI at the 1990s. The breakthrough in this comparatively young scientific filed occurs around new millennium(Bakhtiar & Nelson, 2000; Fitzgerald et al., 1996; Krutchinsky et al., 2000; Przybylski & Glocker, 1996; Vanoni & Curti, 1999; Veenstra, 1999b, 1999a; Winston & Fitzgerald, 1997), for instances, Pfizer pharmaceutical analysist exhibited the spectrum of calcium-bound calmodulin-melittin complex and horse heart myoglobin by a mass spectrometer with ESI source (Finnigan MAT 900Q, Bremen Deutschland)(Loo, 1997, 2000). Renato Zenobi group applied quantitative measurement of non-covalent bound complex in both solution and gas phase(Daniel et al., 2002), It revealed that noncovalent binding is favourable to electrostatic and dipolar interaction effect during gas phase state by ESI-MS. The comprehensive assessment of binding properties via titration and competition measurement during the MS analysis illustrated that ESI or MALDI coupled MS enable to measure the stoichiometry of non-covalent macromolecule complexes. The experiment results also indicated the in certain circumstances the binding constant value is unanimously consistent between solution and gas phase situation, however the direct comparison of the binding energies in two mode is not recommendable while the release of solution abates the hydrophobic effect and gas phase strengthens the ionic interactions. The large non-covalent bound protein assembly, GroEL (approximately 800 kDa), was observed via ESI-MS as well suggested that the intact complex is amendable in gaseous state during MS measurement (Robinson et al., 1994; Robinson & Radford, 1995; Rostom & Robinson, 1999). Taken together, native MS process remains the natural folded structure of biomolecules as well as its biological properties.

Unlike to X-ray crystallography, cryo-EM and EM on interpretation of near atomic resolution structural information, native MS is designed to address the question in terms

of structural dynamics, e.g. stoichiometry, complex assembly/disassembly, subdomains composition and ligand-protein and protein-protein interaction. The minimal low sample consumption and direct measurement procedure without further chemical labelling process also facilitates the analysis of biomolecules to mimic in-vivo situation.

The core of the native MS is to examine the native state of biomolecules, whether the structure of protein or protein-ligand complex in the solution can be sustained in the gas phase is still under the investigation and discussion (Clemmer et al., 1995; Konstantin et al., 1997; Samalikova & Grandori, 2003). In the native MS measurement, the ionized biomolecule likely proteins often presented multiple charge states on the mass spectrum. The charge state distribution can therefore express the preservation of the structural complexity. It is apparent that the folded proteins obtains less charges during the ionization due to the compact structure while unfolded species accommodate a relatively more charges owing to the extended surface area(Grandori, 2010). According the experimental experience, the native protein appears a narrower range of charge state distribution, in contrast, the denatured structure presents a broader construct with higher charge states that is not shown in the regarding mass spectrum of native-like structure. The appearance of the native-like compact structure during the ESI process required the correct choice of suitable solvent during the measurement. High performance liquid chromatography is commonly coupled to ESI based mass spectrometer in the industrial area, the conventional ESI measurement buffer solution system contained methanol and acetonitrile. It provided multiple charges for the ionized protein samples in the positive ionization mode due to these agents led to the fully/partially denaturation of the protein. Afterwards, experts realized that volatile buffer solution is ease to desolvate during the ionization process. Ammonium acetate and the analogues are determined to be experimentally compatible and then widely used in native MS measurement. whereas, the common buffer system that are used in cell culture and protein expression and purification are catalogued to non-volatile solution, which disturb the ESI process for native-like structure formation in the gas phase.

The well known tobacco mosaic virus experiment intelligently supported the theory that native-like structure can be retained in the gas phase as discussed in the previous section(Siuzdak et al., 1996). The structure of virus particle is compact during the observation via electron microscopy after the electrospray process; meanwhile, the same virus species can infect the tobacco plant even it passed through the mass spectrometer. This experiment obviously explained that the capsid structure can be preserved in the gas phase. Contradictorily, similar experiment design has been applied to green fluorescent protein (GFP). fluorescence emission is considered as a core factor to evaluate whether native-state of protein can be preserved during the ESI process(Frankevich et al., 2013; Gülbakan et al., 2015). The function of GFP protein samples were examined by a fluorescence probe in the high vacuum trap after the native electrospray ionization. In the mass spectrum, GFP has been found to preserve a narrow range of charge state distribution. The collision cross-section calculation also indicated the native-like structure was retained.

1.7.1 protein-ligand interaction in native mass spectrometry

Non-covalent interaction plays a vital role in biological recognition. Binding occurs when ligands closes to the supposed specific pocket on the protein, which might induce the conformational change of the protein or trigger the exposure of another binding region for the specific ligands. Over past two decades, several instruments have been developed aiming to the study of non-covalent interaction, for instances isothermal titration calorimetry (ITC), SPR and microscale thermophoresis (MST). However, high sample consumption and contamination during the measurement weakened the use of thermodynamic instrument, the immobilization in the SPR assay and buffer intolerance as well limited it expected capability. MST is available to perform large scale ligand-protein interaction screening test, however, the high false positive rate hindered its utilization on academic area. MS approach has been established in recent years as a rapid and sensitive technique on exploring the interplay of protein and ligands. Affinity selection mass spectrometry is considered as a common method and widely used in industrial drug discovery platform. In contrast, native MS is applicable for preserving the native-like structure of protein in the gas phage, therefore, it is suited to examine the

noncovalent interactions to determine the binding occupancy and stoichiometry(Dyachenko et al., 2013; Kitova et al., 2012).

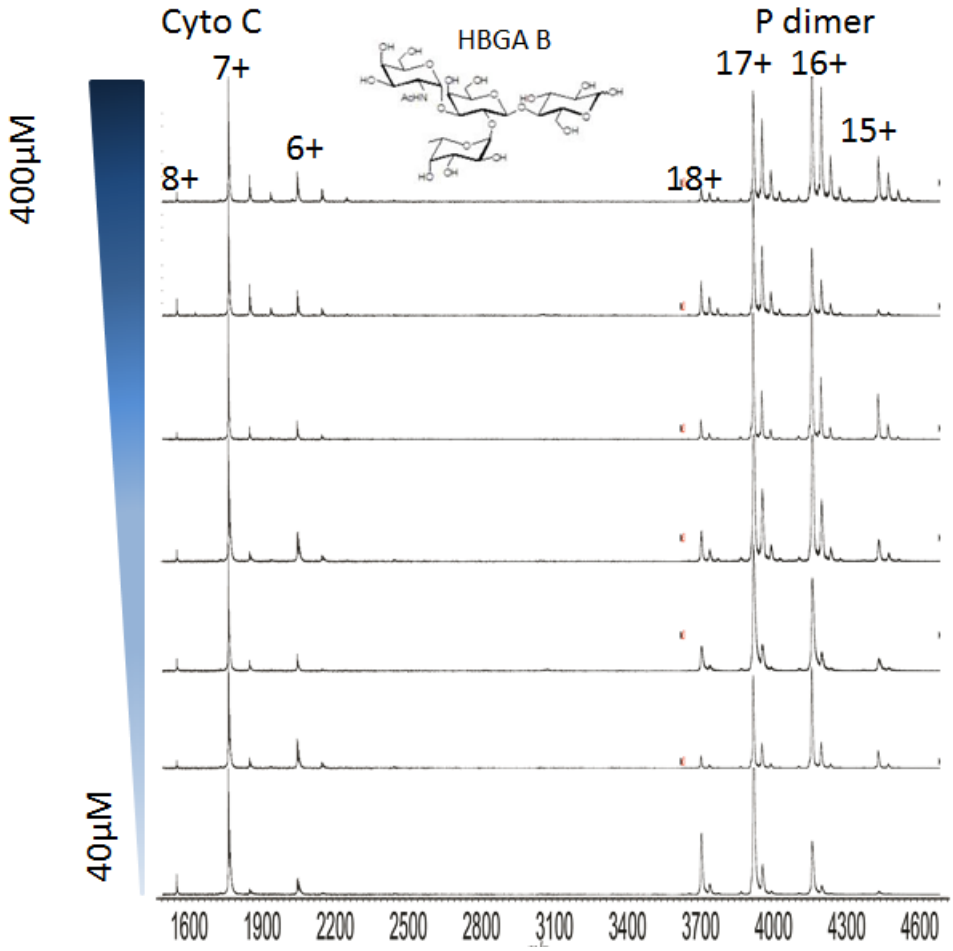


Figure1.5. titration experiment of glycan binding on P dimer and Cytocrhome c (Cyt c) non-covalent interaction between protein and glycan ligands can be observed on the mass spectrum. The native protein reflected to multiple charged state distributed on the certain range of mass to charge ratio as Cyt c or P dimer.

During the native MS measurement, non-covalent binding events can be directly resolved on the mass spectrum (Figure 1.5). The bound ligands information is transferred into the mass of the specific ligands added up to the total mass of the proteins. Due to charge state distribution of the protein species, the analysis of the binding stoichiometry started with the calculation of the bound ligand molecules on the each of charge states related peaks. However, native MS has its own limitation for the

investigation of protein-ligand interaction, e.g, unbound ligand might coincidently stay with the protein molecule in one droplet during the solvent evaporation procedure, which might lead to inappropriate data interpretation on the binding events.

MS experts then invented a novel method to eliminate unspecific clustering through introduction of a reference protein during the ESI-MS analysis (Kitova et al., 2012; Sun et al., 2006; J. Sun et al.)(Figure2.1). It considered that the reference protein candidate do not possess specific ligand binding pocket(s) and would not interact with the target protein to form protein-protein complex. Thereby, the added mass on the charge states of the reference protein is deemed as non-specific binding, assuming that both reference candidate and protein of interest obtain same ionization efficiency, the amounts of non-specific clustering for the target protein can be deducted by the ligand attachment on reference protein. This direct measurement finally enabled to determine the correct bound ligand numbers on the protein-ligand complex.

Apart from resolving the binding events, the determination of binding strength is as well a pivotal point on the study of protein-ligand interaction. Based on the reference protein candidate method, the unbound protein and ligand protein complex is assumed to be ionized with identical efficiency, the ratio(R) between the unbound protein (P) peaks and bound protein-ligand complex (PL) peaks can be determined by Eq 1.2

$$R = \frac{PL}{P} \qquad \text{Eq1.2}$$

For the reversible reaction of protein-ligand noncovalent interaction, when the general equilibrium reaches, the equation of the binding process can be describe as Eq1.3 and the corresponding binding constant can be deduced as Eq1.4.

$$P + L \leftrightarrow PL \qquad \text{Eq1.3}$$

$$K_a = \frac{[PL]}{[P][L]} \qquad \text{Eq1.4}$$

Considering of simplest binding event as the condition described below that one ligand only binds to one binding pocket of the protein. The equilibrium concentration of [PL]

and [P] can be determined by the initial protein concentration added in the buffer solution(Eq1.5).

$$[P_{initial}] = [P] + [PL] \quad \text{Eq1.5}$$

Assuming that the spray and detection efficiencies are identical for the unbound protein and ligand bound protein complex during the ESI process and inside mass spectrometer and no internal energy transferred during the passage of ion molecules through the Hexapole and ToF instrument. The ratio of P and PL can be acquired by the collection of sustainable sprayed ion molecules in a period of time reflected on the chromatograph panel. In other words, the peak area of the unbound P and ligand bound PL shown on the mass spectrum is regard as the ration between the concentration of [P] and the regarding [PL] in the buffer solution at equilibrium situation

Similar equation can be used to calculate the final equilibrium concentration [L] by the beginning concentration of the target ligand[$L_{initial}$]. The Eq 1.6 can be written as below.

$$[L] = [L_{initial}] - [PL] \quad \text{Eq1.6}$$

The binding affinity of ligand binding on the protein is commonly evaluated by the reciprocal of the binding constant, namely dissociation constant (K_D). Therefore, according to the equation Eq1.4, Eq1.5 and Eq1.6. The K_D is resolved in the equation7 and 8(Eq1.7 and Eq1.8) as described

$$K_D == \frac{([P_{initial}] - [PL])([L_{initial}] - [PL])}{[PL]} \quad \text{Eq1.7}$$

$$K_D = \frac{[L_{initial}] - \frac{R[P_{initial}]}{1+R}}{R} \quad \text{Eq1.8}$$

For the multiple binding events, ligand molecules can be bound subsequently to proteins. The additional binding event(s) have an influence on the previous binding constant, the reversible reaction and the corresponding dissociation constant calculation

will be more complicated. The series of protein-ligand interactions reaction can be expressed as protein(P) binds to N number of ligand(s)(Eq1.9)

$$P + nL \leftrightarrow PL + PL_2 + \cdots PL_n \qquad \text{Eq1.9}$$

Thus, the abundance of ion molecules for all the species that observed in the mass spectrum are still equivalent to the equilibrium concentrations of the regarding ligand bound complexes, , and the equation is shown as below (Eq1.10).

$$[P_{initial}] = [P] + [PL] + [PL_2] + \cdots [PL_n] \qquad \text{Eq1.10}$$

The regarding dissociation constant can be found from equation 1.10 and 1.11

$$K_{D2} = \frac{[PL][L]}{[PL_2]} \qquad \text{Eq1.11}$$

$$K_{Dn} = \frac{[PL_{n-1}][L]}{[PL_n]} \qquad \text{Eq1.12}$$

And the Eq1.12 can be reconstructed by the initial concentration of the free ligands and the protein of interest. The Binding affinity for the multiple binding events is then acquired by Eq1.13 (A. El-Hawiet et al., 2012):

$$K_{Dn} = \frac{R_{n-1}([L_{initial}] - \frac{(R_1 + 2R_2 + \cdots nR_n)[P_{initial}]}{1 + R_1 + R_2 + \cdots R_n}}{R_n} \qquad \text{Eq1.13}$$

Although native MS data can be interpreted through the mathematics equation to decipher the binding affinity of the target ligand, during the multiple binding events for one ligand, it is noticeable that the direct ESI-MS method has the limitation to determine the microscopic dissociation constants for each of the binding event occurs on the different binding pocket, therefore, the dissociation constant resolved in the study is macroscopic.

1.8 Hydrogen deuterium exchange mass spectrometry

The early recorded experiment of Hydrogen-deuterium exchange(HDX) started at 1950s(Hvidt & Linderstrøm-Lang, 1954). Site-resolved kinetics determination experiment has been established in NMR before HDX technique introduced in MS field(Johnson & Walsh, 1994; Neubert et al., 1997; Ohguro et al., 1994). MS based method has been established and introduced to determine the HDX exchange rates by counting the mass of deuterated peptides on Liquid chromatography mass spectrometry(LC-MS). The so-called bottom-up method was then used to digest the protein into peptides for sequence mapping and measuring the HDX exchange ratio on the peptide fragments. Online digestion system and automatic analysis software facilitated the HDX-MS development. automated sample preparation device generated from LabTex and Waters corporation coupled with ultra-high performance liquid chromatography(UPLC)(Houde et al., 2009; Wu et al., 2006) as well aid the investigation of HDX exchange ratio on protein structure dynamics(Houde et al., 2009), antibody bio-similarity quantification(Houde et al., 2011), protein-protein and protein-ligand interaction(Mallagaray et al., 2019).

1.8.1 Hydrogen deuterium exchange mechanism

Hydrogen exchange is a common phenomenon occurring on hydroxyl; N-H and S-H group of amino acids in proteins(Konermann et al., 2011; Marciano et al., 2014), however, MS is not available to detect the exchange due to no mass change can be recorded. D_20 were considered as an ideal solvent, the one Da increases can be detected on mass spectrum to evaluate exchange event and therefore the protein structure behaviour can be traced by HDX MS, however, it is noteworthy that the HDX exchange is a reversible reaction. The replaced deuterium(D) on the moiety of the amino acid can be reversed again by the hydrogen, which is so called back exchange or exchange out(Masson et al., 2019).

During the HDX process, Hydrogen at side chain of the amino acids would be exchanged in a fast mode that the reversible reaction of hydrogen(H) to D attained the equilibrium in seconds, which hardly can be quenched and detected through the high sensitivity MS(Wei et al., 2014). In contrast, hydrogen at amide group of the amino acid backbone extended the period of time to reach equilibrium. It has been found the corresponding

HDX rate is approximately two orders of magnitudes slower than the hydrogen on the sidechain and N-/ C- terminal residues(Konermann et al., 2011). The gradual exchange rate endures in minutes to several hours, even some days, which, in turns, allows MS to detect the mass difference on the backbone amide hydrogen(Konermann et al., 2011; Wales & Engen, 2006).

In the microenvironment of the intermolecular interaction, the mechanism of the overall exchange can be simplified as the shift between two states of hydrogen molecule: open and close transitions, it assumed that in the beginning moment, the hydrogen on the amide group stays as a closed (inactivated) state, if the exchange proceeds, the closed state needs to shift in to open (activated) conformation to allow the deuterium to take place of the position(Konermann et al., 2008). The corresponding kinetics can be written as:

$$\mathrm{H}_{cl} \underset{K_{cl}}{\overset{K_{op}}{\rightleftarrows}} \mathrm{H}_{op} \xrightarrow{D_2O} D_{OP} \underset{K_{cl}}{\overset{K_{op}}{\rightleftarrows}} \mathrm{D}_{cl} \qquad \text{Eq1.14}$$

The subscripts cl(closed) and op(open) regard to the two transitions for the hydrogen and Deuterium. The rate constant of the open and closed states are expressed as K_{op} and K_{cl}. In details, the hydrogen bonds constantly break and successively re-close owing to protein thermal fluctuation. For example, when backbone amide hydrogens transfer to opening state in a moment, it is ready to be replaced by deuterium with same open transition state in the solvent, the exchange subsequently occurs. The backbone amide deuterium re-closes to the closing state of deuterium. Therefore, in the deuterium abundant situation likely D_20 solvent, the reaction will be forced to move into deuterium closing state. the amount of hydrogen that ready to change into deuterium can be calculated as described in Lars Konermann's work(Konermann et al., 2008)

$$\mathrm{H}_{op} + \mathrm{H}_{cl} = \mathrm{H}_{ex} \qquad \text{Eq1.15}$$

And the K_{ex} equals to the chemical exchange constant when the reaction goes to in the right direction in Eq1.15 that Deuterium start to occupy the original position of the hydrogen in the backbone amide group.

$$\frac{dH}{dt} = -K_{ex}H_{op} \quad \text{Eq1.16}$$

Assuming that generation and reduction of H_{OP} reach to the equivalent speed in the moment during the exchange process, Therefore, the rate constant of HDX equals to (Eq1.17):

$$\frac{K_{cl}K_{ex}}{K_{ex} + K_{cl} + K_{op}} = K_{HDX} \quad \text{Eq1.17}$$

During the solvent D_20 Labelling experiment, the chemical reaction of HDX attributes to D_{cl} state on the protein samples. It indicated that $K_{cl} \gg K_{op}$, where

$$\frac{K_{cl}K_{ex}}{K_{ex} + K_{cl}} = K_{HDX} \quad \text{Eq1.22}$$

Here, two limiting model can be categorized, if $K_{cl} \gg K_{ex}$, the above model can be modified(Eq1.23)

$$\frac{K_{op}}{K_{cl}}K_{ex} = K_{HDX} \quad \text{Eq1.23}$$

Then, the HDX process will follow EX2 Kinetics (Figure 1.6), which is often shown in the native state of the protein or protein-ligand interaction. A mild structural fluctuation can lead to the exposure of the hydrogen to the solvent, on the mass spectrum, it is obvious that the mass shift occurs gradually when deuterium solvent incubation time period extends(Weis et al., 2006).

If $K_{cl} \ll K_{ex}$, the HDX process will be attributed to EX1 Kinetics(Figure 1.6), in the limiting model, the proton is considered to be exchanged during the global structural fluctuation, the scenario is commonly shown in general denaturing situation that the hydrogen is deeply buried in the protein structure and exposed after an complete HDX(Weis et al., 2006). The following equation is regarded as below (Eq1.24)

$$K_{op} = K_{HDX} \qquad \text{Eq1.24}$$

On the mass spectrum, in the ideal situation, two m/z peaks distribution will be emerged with a separated low and high mass population .

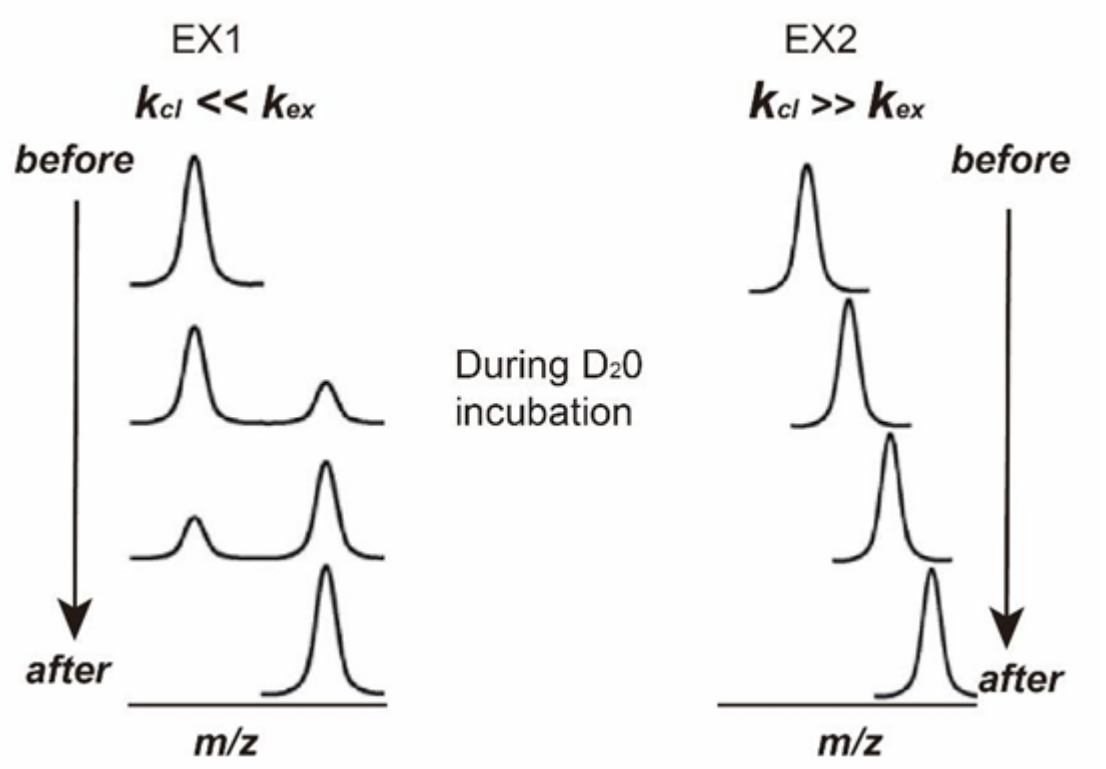


Figure 1.6: The scheme presentation of hydrogen deuterium exchange at two Kinetics model, EX1 and EX2, respectively. During EX1 Kinetics, global structural fluctuation occurs, ,hydrogen is buried in the protein structure and exposed after H/D exchange. An obvious mass shift can be seen by MS. In EX2 Kinetics, a mild structural fluctuation happens, mass shift occurs gradually when incubation time period extends(Weis et al., 2006).

Since EX1 and EX2 kinetics is measurable on the mass spectra, some group of scientists pioneered the method to study the protein dynamics and conformational changes(Bai et al., 1993; Weis, 2016; Weis et al., 2006). Before The primary construction of the protein is amino acid sequence. The backbone amide hydrogen is accessible for deuterium occupation in D_20 solvent(Kaltashov et al., 2012). However, the secondary structure and tertiary structure will hinder the process of the HDX. Here, to simplify the microenvironment of the molecular interaction(Masson et al., 2019; Weis, 2016). We considered hydrogen bond as the main intermolecular force, the compact structure of protein contains numbers of secondary structures likely α-helix, β-sheet, α-turns, the backbone amide hydrogen is circled by intramolecular Hydrogen bonds, D_20 solvent has limited access to induce the Hydrogen breaking on the backbone of the amino acids sequence. The HDX behaviour in this region is radically reduced to 10^6 as experiment determined(Konermann et al., 2008). Therefore, the HDX rate change can be considered as a parameter to evaluate the protein dynamics, which the folded regions reduced the HDX rate, in contrast, the loop or dynamics region have relatively higher exchange ratio.

In the protein measurement, the exchange rate constant is influenced by hydrogen ion, hydroxyl group and the water solution. It is known that the K_{ex} is pH dependent(Masson et al., 2019). Which enables to set up HDX experiment at low exchange rate condition in acidic pH value , however the exchange rate increases gradually after it reaches the bottom at pH 2-3(Bai et al., 1993). Thereby, the quenching process during the experiment is mainly set up at the pH2.3-2.5. Temperature is another factor affects the exchange halftime. Based on the Arrhenius equation, the K_{ex} reduces dramatically when temperature decreases, the halftime can be dropped in a magnitude when the temperature decreases from 25 to 0 °C(Masson et al., 2019).

1.8.2 HDX MS application

As mentioned in the previous chapter, native MS(Deng et al., 2013; Han et al., 2013), ITC(Sun et al., 2006), MST and SPR(Amr El-Hawiet et al., 2012) and other conventional methods have been developed to probe protein-protein, protein-ligand interactions. Apart from direct MS analysis of the mass increase at intact protein level from charge

state distribution, some MS experts also pioneered to measure HDX process during the ligand binding and protein conformational change on MS(Kaltashov et al., 2012; Sun et al., 2006). For instance, Dülfer et al. (2021) hypothesized that the binding pocket on the protein is occupied by the bound ligand, the occupation leads intermolecular interaction between the ligand and amino acid on the protein and subsequently attenuates the deuterium replacement on the protein backbone amide hydrogen which it can be detected by MS. In contrast, ligand bound might cause the protein structure un-protected or extended. It allows a higher rate of HDX exchange. Similar to the protein structural change situation, the protein conformational change induce the exposure state of the protein in comparison to the compact closing state.

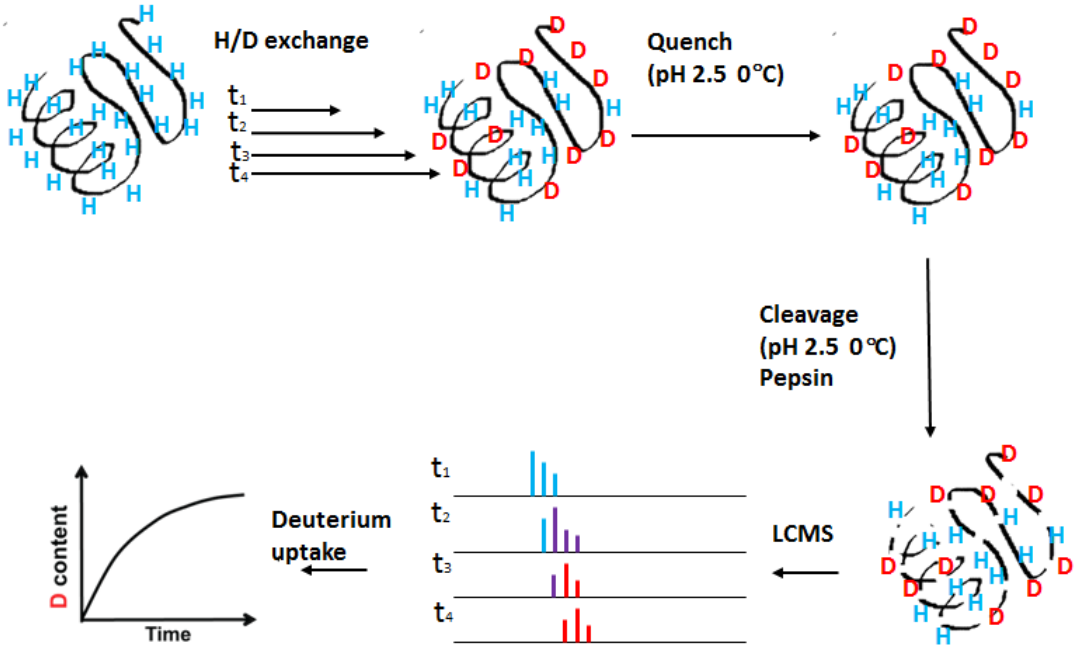


Figure 1.7 The scheme presentation of HDX experiment workflow. HDX begins with D_20 incubation, the hydrogen (H, blue) on the sample is exchanged by deuterium (D, red). Quenching process is applied at the different time point at 0 °C icy water in pH2.5. After that, the sample is then digested by enzyme and the peptides are separated and measured by LCMS.

In the experiment procedure, the procedure can be divided into two aspects(Figure x), the protein sample initially are incubated in D_20 solvent for the deuterium labelling procedure. The following step is to quench the HDX process in the different time point at 0 °C icy water in pH2.5 during titration measurement. The quenched sample solution can

be directly injected into the LC-MS system to measure the mass shift at intact protein level. The same protocol can be applied to protein ligand mix, However, the method is merely allowed to view whether the binding occurs on the protein via comparing the HDX rate difference between the protein and protein-ligand complex, the enzyme digestion and the introduction of bottom up strategy from data-dependent analysis(DDA) provided structural information via sequence mapping for the study of non-covalent ligand-protein interaction. The binding occurs either at HDX rates decreased area (protected region) or the exchange rate increased areas (unprotected region) according to the bound complex and non-bound species HDX sequence mapping result. The applied enzyme in the HDX workflow decides whether the digested peptides can cover the potential binding region to preserve global sequence coverage. In addition, due to the harsh condition for the quenching step, the selection of the enzyme must be active in low pH range. Pepsin is active at pH 2-4, which matches the requirement of HDX experiment.. It used previously to mix the pepsin with the protein of interest in a ratio of 1:1 for minutes to maintain the good performance digestion during the off-line protocol, the recent development prefer to immobilize the pepsin onto the specific column to fulfil the on-line digestion process to preserve high quality and consistent data. The sample was firstly inject into LC manual sampler, the solvent pump provide the acidic water(mobile phase: 0.23% Formic acid, pH 2.4) to transport sample from sample loop to the pepsin column for enzyme digestion, the peptides fragments were then flowed through and remained in the trap column through valve switching,. Acetonitrile solvent from binary solvent pump subsequently washed through the trapping column to carry the remaining peptide to the final C18 reversed-phase analytical column to separate the peptides of interest via the retention time difference. Each of peptides would be detected and interpreted by tandem MS analyser.

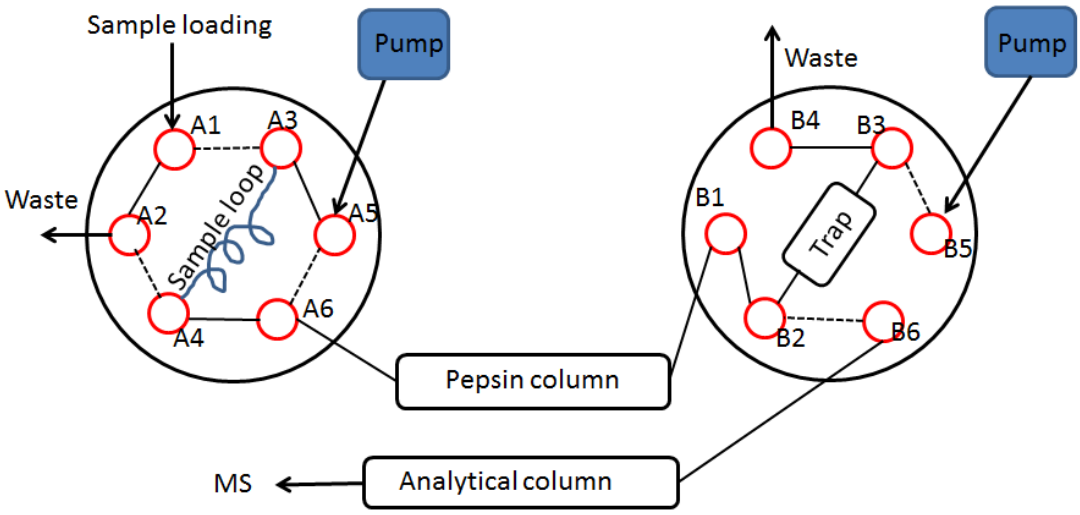


Figure1.8 The scheme view of HDX digestion module: The solid and dotted lines present the solvent flow route. The module can be divided into three successive steps. A) sample loading: the protein sample flows to sample loop (pathway: A1-A3-A4). B) Enzyme digestion and trapping. After switching valve, the solvent pump pushes the sample solution to pepsin column for digestion and then the solution is trapped in trap column (pathway: A5-A3-A4-A6-B1-B3). C) MS detection: binary solvent pump pushes the digested sample solution from the trap column to (C18 reversed-phase) analytical column to separate the peptide via different retention time and then the peptides are detected by MS(pathway: B5-b3-B4-b6).

2.Aim of thesis

The scope of thesis focused on explore the role of glycan binding on human norovirus via native MS and HDX MS

Native MS is often used to determine the binding strength during the protein-glycan interaction(Amr El-Hawiet et al., 2010; Han et al., 2018; Jecklin et al., 2010; Sun et al., 2006; Wegener, 2017), The binding affinity of several groups of carbohydrates were measured towards the human norovirus P dimer, protein complex and VLP. (Dülfer et al., 2021; Han et al., 2013; Han et al., 2018; Mallagaray et al., 2019; Mallagaray et al., 2015; Wegener et al., 2017), However, the ambiguous K_D value from the NMR and ESI-MS techniques motivated our present work to decipher the gap on glycan binding affinity determination on human norovirus VP1 protein P domains. In the direct ESI-MS method, reference protein candidates are considered as valuable standard to adjust the non-specific clustering during native MS measurement. In the chapter 3, eight reference candidates were chosen and measured. We plan to illustrate the impact of reference protein on the clustering issue. Several different protein physiochemical characters are explored to address the question on which factor(s) correlate the glycan clustering in the gas phase.

Fucose is the minimal binding component to P dimer(Brigitte et al., 2012) and binding pockets locate at the cleft region formed by two P2 domains on the distal area of viral capsid protein(Hansman et al., 2011). In the chapter 4, a group of glycomacromolecules were generated with a peptide-like backbone chain upon different number of fucose/galactose attaching on the functionalized moiety group. The designed structure of these molecule is aimed for achieve multivalent binding during the interaction to P dimer. Native MS is then used to identify which glycomacromolecule enables to reach a strong binding avidity to compete with natural glycan. A non-binder control with only galactose attached structure is also measured to examine the non-specific clustering.

Recently, HDX MS and NMR identified a post-modification site (N373) near to canonical glycan binding region on P2 domain of GII.4 Saga strains (Dülfer et al., 2021; Mallagaray et al., 2019). The 1Da mass shift of asparagine to iso-asparate led to carbohydrate binding attenuation on deamidated P dimers in presence of blood group antigen b type. Epidemiological surveys also reported that an increased GII.17 strains infection cases while GII.10 Vietnam was isolated from sporadic infection cases (Ahmed et al., 2014; Chan et al., 2017). Sequence alignment revealed that these two strains do not carry an asparagine at position of 373 on the P2 domain. In addition, GII.4 MI001 strains P dimers shows high structural similarity compared to Saga P dimer. Thereby, HDX MS analysis is applied to decipher the binding profile of these three P species. We also aim to investigate whether deamidation can be observed on GII.17 and GII.10 strain and how the glycan binding patterns on deamidated and non-deamidated MI001 P dimer species.

In summary, native MS and HDX MS measurement of different P dimer species-glycan interaction could assist us to understand the role of glycan binding during the virus infection. The determination of the binding affinity of natural carbohydrates and synthesized glycan attached structures aims to provide useful information for developing the anti-viral drug platform for the potential pharmaceutical study. The investigation of the binding profile of different norovirus strains would facilitate us to track the virus evolution and address the issue on how glycan binding influences on virus infection.

3. Protein secondary structure affects glycan clustering during native MS analysis of norovirus-carbohydrate interaction

3.1 Introduction

This chapter is largely based on the published manuscript (Yan et al., 2021). I performed native MS analysis of glycans binding to norovirus GII.4 P dimers and commercial available reference protein candidates to understand the influence of glycan clustering.

During glycan binding studies, most groups use the dimers of the protruding domain (P dimers) of the major capsid protein VP1. Notably, L-galactose derived from L-fucose by substituting one hydrogen atom for a hydroxyl group at C6 is known not to bind (Fiege et al.; Mallagaray et al., 2019). Because the determination of dissociation constants and binding stoichiometry plays a key role to understanding protein–ligand interactions, a number of studies reported such data for the binding of HBGAs to hNoVs based on different experimental approaches. Sun et al. previously established an MS-based method to determine specific glycan binding to proteins(Sun et al., 2006), which relies on the addition of reference proteins to the sample mixture and simultaneously analysing the constituents using native MS. This approach allows the quantitation and elimination of unspecific ligand clustering during electrospray ionization (ESI). Clustering arises from the statistical presence of free ligands within the same droplet as the free or ligand-bound proteins, which can then dry down to the protein surface upon droplet evaporation. This process becomes relevant at elevated ligand concentrations and is independent of protein molecular weight (Gulbakan et al., 2018; Jecklin et al., 2008; Pacholarz et al., 2012). Calculations to correct for this effect are based on total peak areas per mass species within the same spectrum to ensure identical ionization conditions.

A couple of groups including ours have employed this method to characterize glycan binding to hNoV P dimers using distinct reference proteins(Han, M. Tan, et al., 2014; Han

et al., 2013; Han et al., 2018; Mallagaray et al., 2019; Mallagaray et al., 2015). Notably, results disagreed (Han et al., 2018; Mallagaray et al., 2019), raising the question of whether the selection of a reference protein influences data interpretation. Moreover, with the exception of glycan mimetics (Bücher et al., 2018), the search for non-binding ligand or P dimer controls for native MS was unsuccessful in our hands. This is in stark contrast to NMR data, which show that certain P dimers and glycans do not interact. This suggests a severe issue with the native MS approach. Strikingly, direct MS analysis determined mM K_Ds for binding of several sialic acid containing carbohydrates hNoV P dimers (L. Han, M. Tan, et al., 2014; Han et al., 2018; Mallagaray et al., 2015; Wegener et al., 2017), whereas orthogonal saturation transfer difference (STD) and protein-based chemical shift perturbation (CSP) NMR experiments clearly revealed no binding of sialic acids to hNoV P dimers or virus-like particles (VLPs)(Brigitte et al., 2012; Han et al., 2013; Mallagaray et al., 2015). This questions the validity of the results from direct MS measurements.

Additionally, the reported K_D values from carbohydrate- binding studies to P dimers are not comparable in STD NMR, native MS and isothermal titration calorimetry (ITC) (Han et al., 2018; Ling et al., 2015; Mallagaray et al., 2015). Importantly, K_Ds obtained for active pharmaceutical agents based on different biophysical assays such as native MS, ITC, surface plasmon resonance (SPR) and circular dichroism (CD) were equivalent in numerous other cases (Gulbakan et al., 2018; Jecklin et al., 2010; Ling et al., 2015). However, these compounds showed higher binding affinities (µM range) for the target protein and exhibited in most cases no carbohydrate-like structures [17]. In contrast, glycan binding affinity is often in the high µM to mM range. This also holds for hNoV–carbohydrate interactions (Han et al., 2013; L. Han, M. Tan, et al., 2014; Han et al., 2018; Ling et al., 2015; Mallagaray et al., 2015; Wegener et al., 2017). In former studies, the origin of the discrepancies between NMR and native MS data for P dimer–glycan interaction was not in focus (Mallagaray et al., 2015). The problem became evident when comparing non-deamidated (wildtype), and deamidated GII.4 Saga P dimers. In this GII.4 P dimer, an asparagine residue flanking the binding site is specifically and spontaneously converted into an iso-aspartate. The deamidated P dimer has been shown to have greatly reduced glycan binding affinity in NMR and hydrogen/deuterium

exchange (HDX) MS, roughly by an order of magnitude for HBGA B trisaccharide and fucose compared to the wildtype (Bücher et al., 2018; Mallagaray et al., 2019).

Here, wildtype and deamidated P dimers are compared using native MS to shed light on potential methodological issues. Moreover, Gb4, an all-galactose tetrasaccharide, is employed as a negative control. Using this information, we hypothesize that clustering depends on physicochemical properties of the proteins. To deduce what obscures the binding studies, multiple reference proteins varying in properties are compared. The results point to an influence of β-sheet content. This theory is corroborated by ion mobility MS (IMMS) measurements on P dimers in presence of glycans and additional data on other P dimers (e.g., from MNV P dimers), which were recently shown by NMR not to bind glycans at all (R. Creutznacher et al., 2021). Our results indicate that reference proteins need to be chosen carefully to match the structural properties of the target protein for glycan binding studies, and, crucially, they suggest the additional influence of structural dynamics that preclude glycan-binding studies in native MS for hNoVs.

3.2 Results

3.2.1. Glycan Clustering on Norovirus P Dimers

Here, the wildtype and deamidated GII.4 Saga P dimers are used as positive and negative protein-binding control (Mallagaray et al., 2019), respectively. The tetrasaccharide of HBGA B type 1 is employed as a glycan known to bind to the wildtype and Gb4 as an all galactose glycan non-binder (Han et al., 2013; L. Han, M. Tan, et al., 2014). Thereby, the applicability of the MS approach to investigate binding equilibria in presence of a reference protein is verified. As can be seen in Figure 3.1A, the reference protein cytochrome c (Cyt c) picks up similar amounts of clustered glycans in both spectra indicating similar spray conditions. Surprisingly, the wildtype and deamidated P dimers also reveal a comparable pattern of glycan attachment although a reduced amount of glycans is expected to stick to the deamidated protein. After correcting for clustering to

the reference protein, even more glycans are supposedly specifically bound to the deamidated P dimer, which serves as a low- or non-binding control, than to the wildtype. Moreover, occupancy is much higher than would be expected based on the K_Ds for the wildtype determined by NMR (12 mM for B tetrasaccharide (R. Creutznacher et al., 2021)), suggesting an intrinsic issue with the measurement approach. Additionally, the non-binding glycan Gb4 shows a pattern similar to HBGA B after correction (Figure 3.1B). But with more glycan attached to the wildtype protein

3.2.2 Glycan clustering pattern on norovirus P dimers

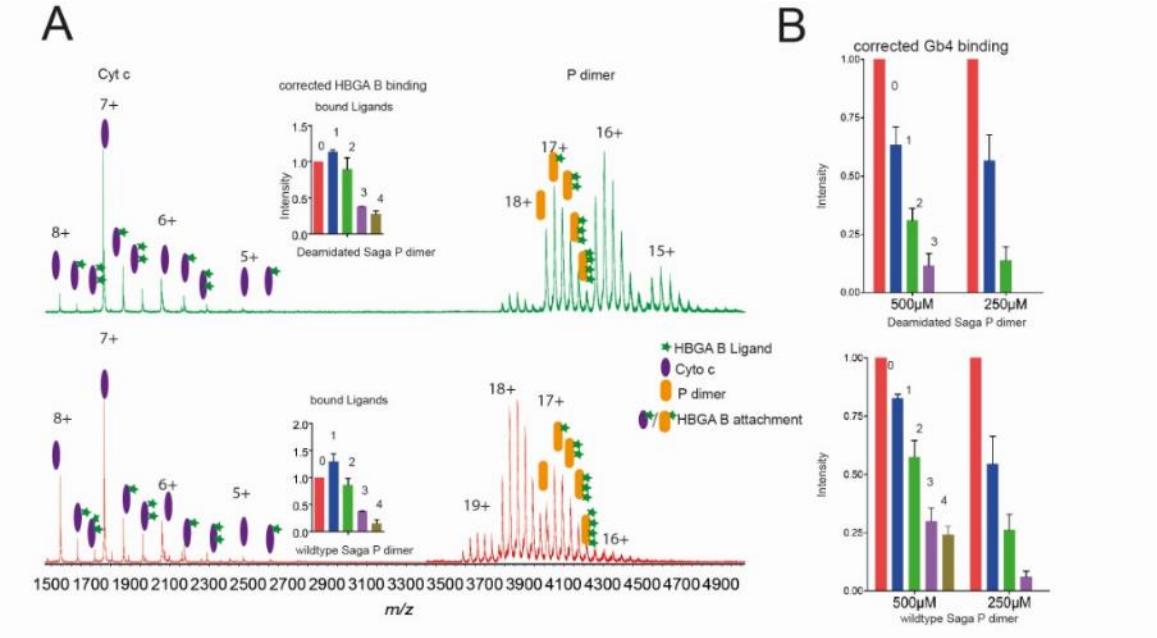


Figure 3.1 The impact of deamidation on glycan clustering to norovirus P dimers. (A) Native mass spectra of cytochrome c (Cyt c) with wildtype (red, bottom) or deamidated (green, top) GII.4 Saga P dimer and 500 µM HBGA B ligand in 150 mM ammonium acetate solution at pH 7. Signal intensity was normalized to the base peak in the spectra. The corrected HBGA B occupancy is shown as insets. (B) The corrected binding to the wildtype and deamidated GII.4 Saga P dimer is shown for 500 µM and 250 µM Gb4.

Additional support for largely unspecific interactions stems from a mutated hNoV GII.4 MI001 P dimer. Here, the supposed glycan-binding pocket is mutated resulting in altered dimerization behaviour; hence, the P monomer and dimer signals. The corrected data

shows equal clustering behaviour for HBGA B, A and 3′-sialyllactose (GM3) on the monomer and dimer (Figure 3.2). Interestingly, Gb4 shows basically no binding after correction in this case. Furthermore, the negatively charged GM3 shows similar patterns compared to the neutral HBGA glycans, both on the monomer and dimer resulting in similar occupancy after correction. This suggests that charged glycans also suffer from similar problems in the direct MS approach. Notably, correction for clustering based on the P monomer also results in no specific binding to the P dimer for the other three glycans. This suggests that Cyt c is not a suitable reference and that glycan clustering does depend on the biophysical properties of the proteins. Therefore, we compare several potential reference proteins that are commercially available and exhibit distinct properties.

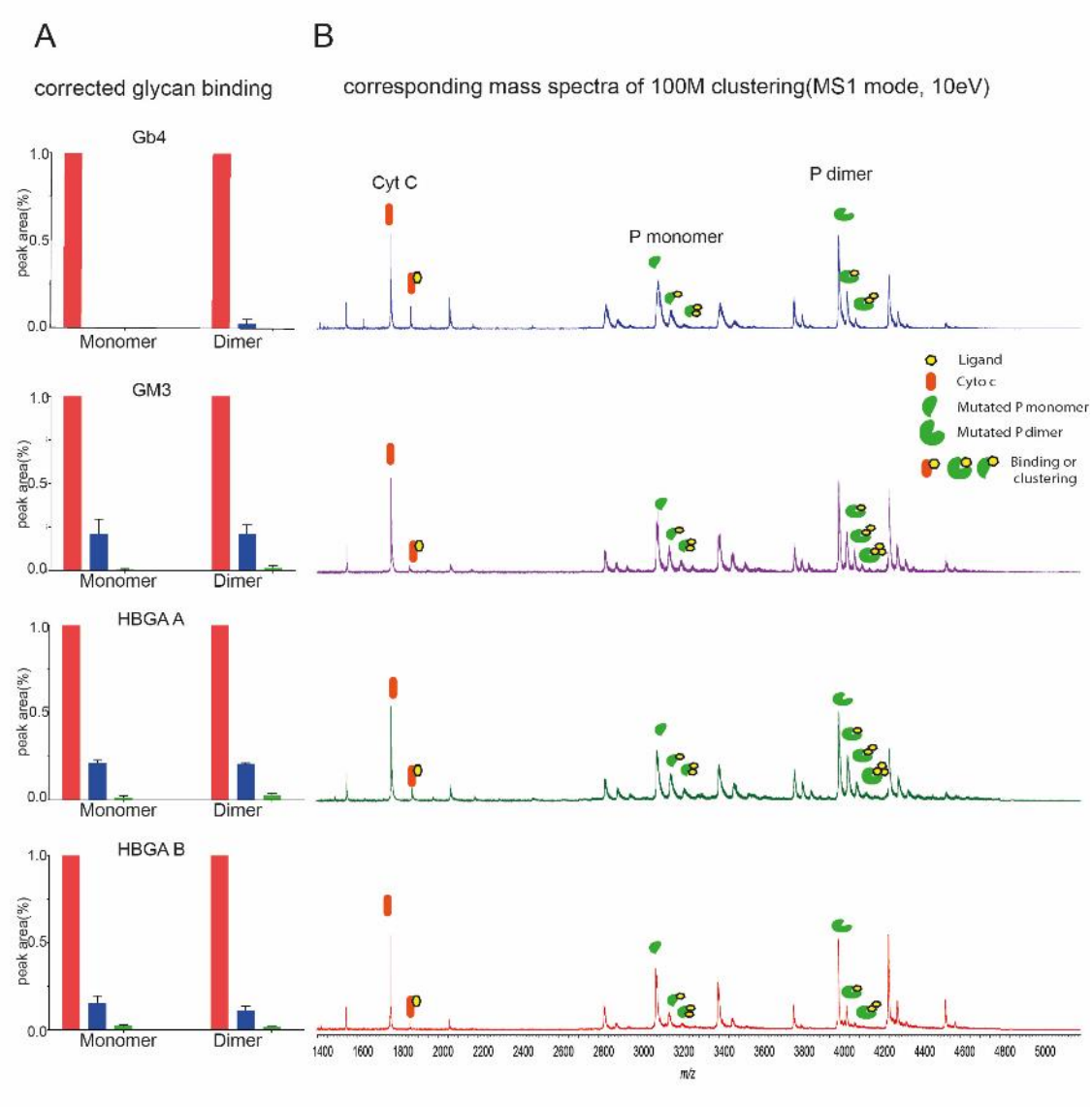


Figure3.2 glycan clustering on mutated hNoV GII.4 MI001 P dimer. (A) The corrected binding of Gb4, GM3, HBGA A and B tetrasaccharides to mutated hNoV MI001 P dimmer is presented from top to bottom, respectively at 100μM glycan concentration. (B) Native mass spectra of Cyt c and mutated GII.4 MI001 P dimer with four glycan ligands at 150 mM ammonium acetate buffer solution at pH7. From top to bottom: Gb4 (dark blue), GM3 (purple), HBGA A (green), HBGA B (red). Signal intensity is normalized to the base peak of the proteins. P dimer: 3800-5000*m*/*z*, P monomer: 2800-3700*m*/*z*,Cyt c: 1400-2600*m*/*z*,

3.2.3 Glycan clustering on different reference proteins

The ratio, R, between the free and ligand-associated reference protein is used to eliminate non-specific clustering to the target protein. In total, eight different reference candidates differing in mass, size and structural composition display very different R values (Table 3.1). For example, the interaction of carbohydrates with the Saga P dimer and four reference proteins (Cyt c, GFP, CA, and ADH) is shown (Figure 3.3). The data illustrated similar binding patterns of HBGA B to Saga P dimers but vastly different glycan clustering to the reference proteins under identical measurement conditions. The dimeric ADH displays the strongest ligand clustering with ADH–ligand (1:1) becoming the base peak similar to the P dimer. CA and GFP show similar patterns with slightly less clustering, whereas Cyt c presents a unique profile with lower clustering ratios and a non-Gaussian charge state distribution.

Table 3.1 The calculation of Glycan clustering ratio *R* and K_{D1} during P dimer-HBGA B interaction via native MS. Data correction is based on the reference protein (Ref. protein) method

Ref. Protein	Mass (kDa)	HBGA B glycan ligand clustering ratio					K_{D1} of glycan-P complex for HBGA B (mM)
		100μM	200μM	300μM	400μM	500μM	
Myo	17.0	0.14±0.03	0.18±0.02	0.19±0.06	0.31±0.04	0.36±0.05	0,25±0,10
Cyt c	13.2	0.14±0.01	0.15±0.01	0.19±0.03	0.29±0.02	0.35±0.03	0,16±0 ,05
Ubq	8.9	0.04±0.01	0.04±0.01	0.10±0.03	0.18±0.01	0.25±0.00	0,20±0,13
GFP	26.8	0.30±0.02	0.41±0.02	0.44±0.04	0.57±0.01	0.67±0.02	0,90±0,35
CA	29.1	0.22±0.03	0.32±0.05	0.40±0.04	0.60±0.05	0.75±0.07	0,40±0,10

ADH	147.0	0.42±0.03	0.64±0.06	0.97±0.08	1.02±0.08	1.45±0.11	24,00±16,00
apo-TFF	79.6	0.37±0.03	0.56±0.05	0.62±0.03	0.74±0.01	0.82±0.03	n.a .
LDH	146.8	0.38±0.03	0.54±0.05	0.69±0.02	0.77±0.05	1.02±0.04	n.a.

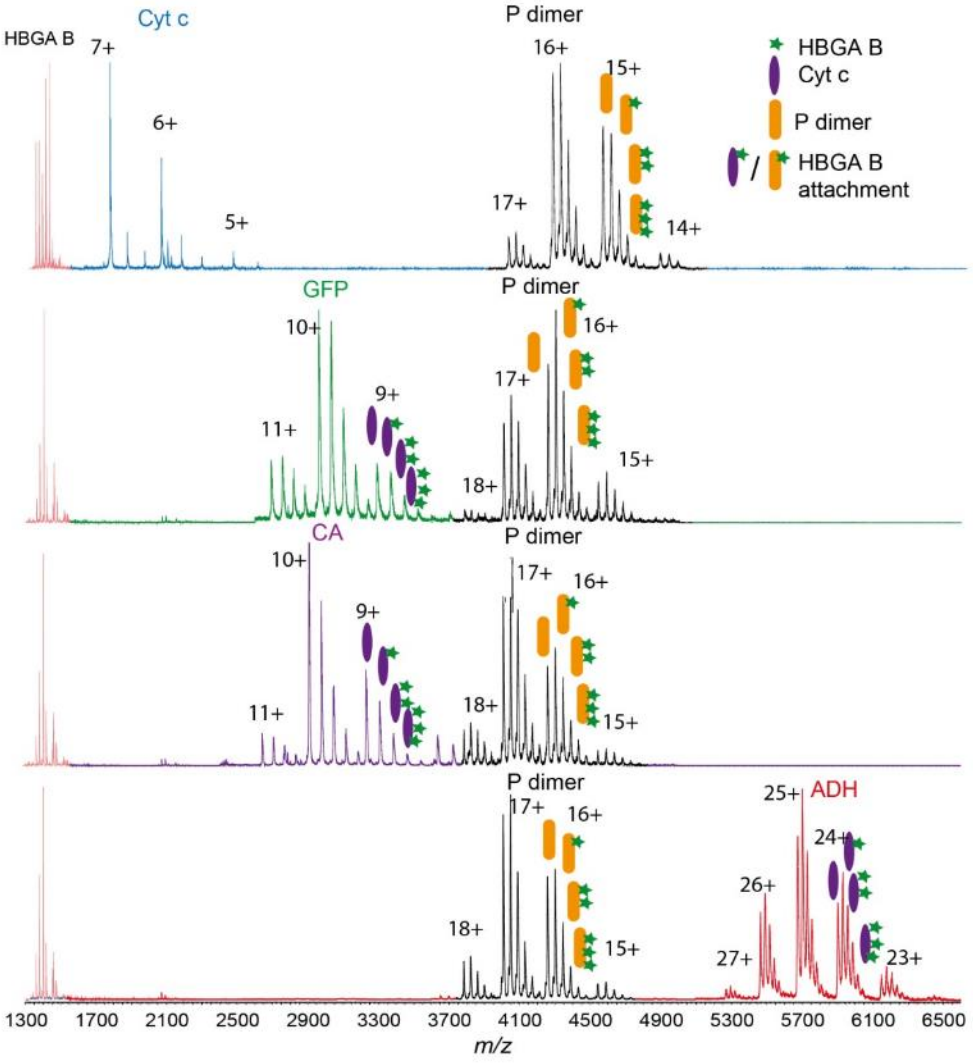


Figure 3.3 Native mass spectra of four reference protein candidates. Wildtype Saga P dimer (1 µM) and four reference proteins (3 µM) with 500 µM HBGA B ligand at 150 mM ammonium acetate solution at pH 7. From top to bottom, Cyt c(blue), GFP(green), carbonic anhydrase II(CA, purple), ADH(red). Signal intensity is normalized to the base peak of the proteins. P dimer: 3800-5000*m/z*, Cyt c: 1400-2600*m/z*, GFP: 2700-3500*m/z*, CA: 2600-3700*m/z*, ADH: 5200-6500*m/z*.

In line with the observed discrepancies in clustering ratio R, the resulting K_D values for the same protein-ligand interaction vary strongly between a few 100 µM and 24 mM depending on the choice of the reference protein. Notably, the K_D obtained with ADH as reference protein is in line with the expected value of 12 mM from NMR data [18]. Differences in R and K_D are also consistent in five glycan concentration titration (Table S1). When the clustering ratio R is plotted over the glycan concentration (Figure 3.4), all reference proteins show the expected proportionality of R and glycan concentration for the three tested glycans but with different slopes. The plots clearly reveal the strongest clustering to ADH in all conditions. On the other hand, the small-sized proteins (Myo, Cyt c and Ubq, approx. 17, 13 and 9 kDa, respectively) stick out with much lower clustering. The remaining proteins display intermediate R values. This may suggest an influence of molecular weight on the reference protein despite other reports (Han et al., 2018; Sun et al., 2006). While all plots fit linear regression, the plots based on the reciprocal R value reveal that clustering to Ubq is not linearly dependent on glycan concentration (Figure 3.4, grey line). Therefore, Ubq is excluded from the following analysis

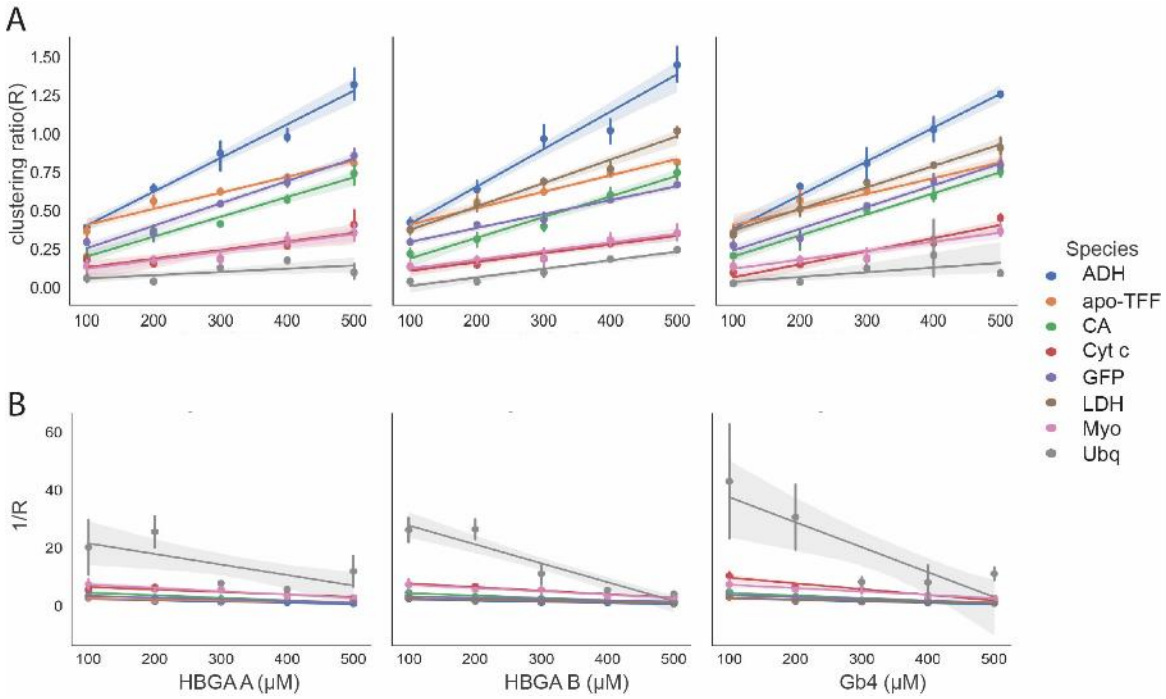


Figure 3.4 Correlating glycan concentrations to glycan clustering ratios. (A) R value calculated based on the peak area of the reference proteins (ADH; CA, Cyt c, GFP; Myo, Ubq, LDH, apo-TFF). (B) Corresponding reciprocal value. Titration experiments were performed with P dimers (1 µM, monomer) and reference proteins (3 µM, monomer) in 150 mM ammonium acetate solution pH

7 with ligand concentrations ranging from 100 µM to 500 µM for HBGA A, B and Gb4. The shaded areas represent the standard error of the slope for the linear fit.

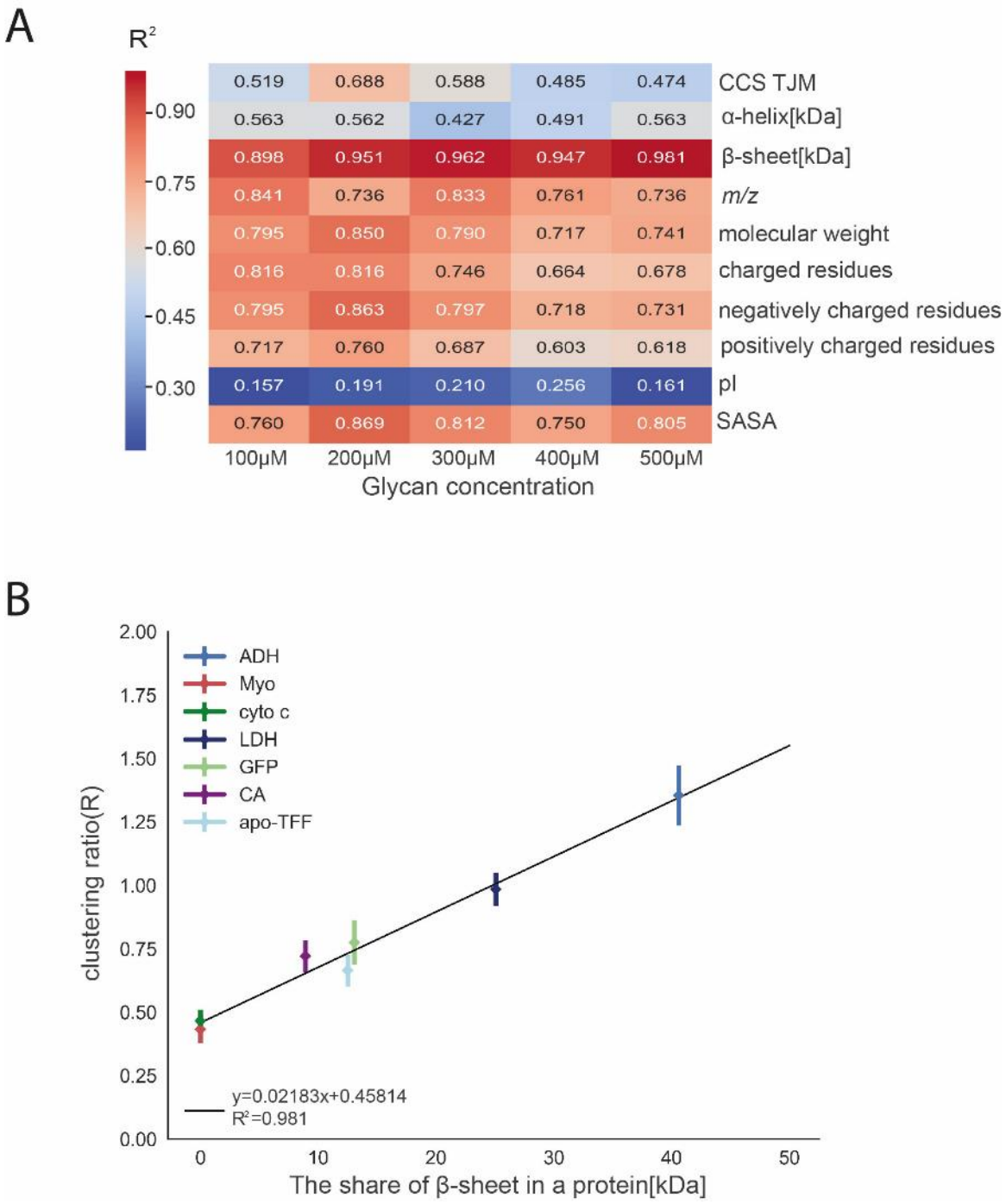


Figure 3.5 Correlating of clustering ratios R to various protein properties. (A) Heatmap of the R^2 values obtained from the linear regression at the indicated HBGA B glycan concentration for the listed protein properties. (B) The correlation of β-sheet amount in kDa to unspecific glycan

clustering ratios R at 500 µM HBGA B for seven reference proteins (ADH, CA, Cyt c, GFP; Myo, LDH, apo-TFF). The black line represents the linear regression and the resulting equation is given.

3.2.4 Biophysical properties impacts on glycan clustering

Ten different protein characteristics that could have affected glycan clustering are plotted against the R values of the remaining 7 reference protein candidates to dissect possible correlations: the absolute amount of α-helix or β-sheet in kDa, number of total charged residues, total number of positively or negatively charged residues, isoelectric point (pI), m/z values, molecular weight, solvent accessible surface area (SASA), and collision cross-section calculated from trajectory method (CCS TJM) (Figure 3.5A, TableS2 and S3). Glycan protein associations are often mediated by CH–π interactions with tryptophan residues. Such interactions are expected to be weak in the gas phase; hence number of tryptophan is not included. Notably, the SAGA P dimer contains only 8 tryptophans, which is below average for its size, clearly showing that tryptophan content cannot explain the huge clustering ratio for the P dimer. R^2 is used to assess the quality of the linear fits. The only parameter providing strong correlations at all ligand concentrations is the β-sheet amount of the reference proteins (Figure 3.5)

The β-sheet amount and glycan clustering grow proportionally. The absolute share of β-sheets in a protein is also connected to the molecular weight. The larger the protein, the more amino acids engage within β-sheets, which is the likely explanation for the apparent impact of protein size. Notably, the amount of α-helix shows no anti-proportional or any other correlation, indicating that these contribute marginally to glycan clustering. Some of the parameters have R^2 values between 0.7 and 0.9, i.e., significantly lower than the β-sheet amount; however, a contribution to overall glycan clustering cannot be excluded for SASA, charged residues, molecular weight and m/z. CCS TJM and pI on the other hand are below 0.5, like the α-helices, and clearly do not contribute(Figure S9 and S10).

3.2.5 P dimer-glycan interaction via Ion mobility MS

Native ion mobility MS (IMMS) can reveal conformational changes in proteins, e.g., upon ligand binding, and has also been used to investigate P dimer glycan binding (Han et al., 2013). In IMMS, changes in arrival time above 3–5% are generally considered significant, which are not expected here based on from HDX-MS and NMR (Mallagaray et al., 2019). Hence, the binding of a small ligand like HBGA B should only cause marginal changes. However, in a previous report, a strong size increase upon glycan addition to SAGA P dimers was observed in IMMS (Han et al., 2013). We hypothesize that interaction of the glycans with the β-sheets lead to the melting of these structures in the gas phase, which could be observed in IMMS. In Figure 3.6, the arrival-time distributions are shown for the wildtype GII.4 Saga P dimer together with its complexes with HGBA B (Figure 3.6A) and Gb4 (Figure 3.6B) for all detectable stoichiometry, as well as the respective single complexes of 1 P dimer (0–3 HGBA B (Figure 3.6C)) and 1 P dimer: (0–3 Gb4 1 molecules (Figure 3.6D)). A small difference in the mean values and the peak widths is observed among the panels A and B, and C and D in Figure 3.6, which is due to changes in environmental conditions, such as temperature, and the slightly different tuning parameters used to optimize the signal. To ensure that no conformational changes are induced by the different conditions, the rotationally averaged collision cross-section (CCS) for the pure P dimer is calculated: in A and C, the CCS is 3928 ± 14 $Å^2$, while in B and D, the CCS is 3867 ± 12 $Å^2$. The difference between these two CCS values is ~1.6%, which is well below the significance threshold of the instrument, and is normal for experiments with large molecules. It could be seen that the arrival-time distribution curves preserve their near-Gaussian, single-peak profile at all P dimers. The ligand stoichiometry suggests no major conformational changes upon complex formation and also no evidence of structural destabilization in the gas phase. Even though the arrival-time distributions for all complexes (Figure 3.6A,B) do not seem to follow an obvious trend, the difference between the distribution of the P dimer (black traces) and the sample with 500 µM of the respective ligands (green traces) suggest that at high-ligand concentration, the relative concentration of the free P dimer decreased. As can be seen in the distributions of selected complexes (Figure 3.6C,D), with increasing number of bound ligands, the arrival-time distribution curves shift slightly towards higher values.

The data clearly show that the binding- and the non-binding ligands trigger the same behaviour, which we largely attribute to clustering.

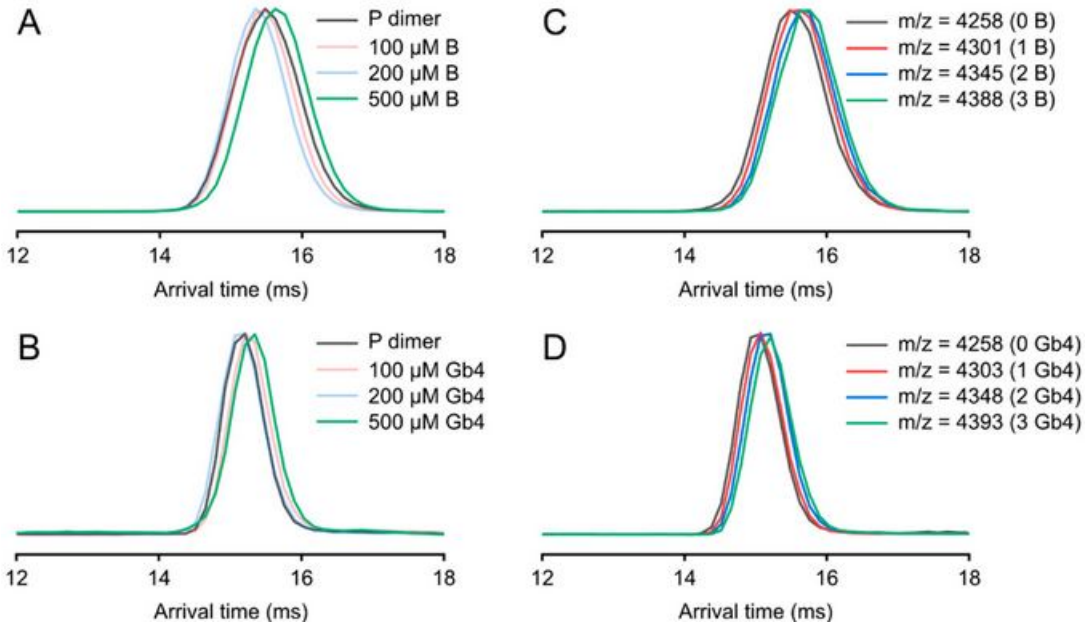


Figure3.6 Glycan clustering to wildtype GII.4 P dimers in IMMS. Changes in the arrival time distribution upon increasing HBGA B (A) and Gb4 (B) concentration are depicted for the 16^{+} ion. Next to these are the extracted arrival-time distributions for free P dimer and P dimer plus 1–3 ligands from a single measurement with HBGA B (C) and Gb4 (D) revealing an increasing trend with increasing number of ligands.

3.3discussion

Glycan clustering was previously observed in native MS experiments (Han et al., 2013; L. Han, M. Tan, et al., 2014; Han et al., 2018; Sun et al., 2006). It occurs during ionization when solvent evaporates and the ions transfer to the gas phase. Due to an excess of ligand, free glycans statistically end up in the same droplet and dry down onto the protein next to specifically interacting glycans. The addition of a reference protein enables the determination of non-specific ligand clustering and hence correction to allow direct MS analysis of the binding occupancy and K_D. Therefore, this method has been widely used (Bücher et al., 2018; Mallagaray et al., 2015; Wegener et al., 2017)to study protein–ligand interactions. Glycans pose a specific problem as the interactions are often of low affinity in the mM range and require vast ligand excess to occupy binding sites.

Using deamidated GII.4 Saga P dimers as well as MNV and MI001 P monomers as negative controls for P dimer-glycan interaction, we reveal inherent problems with the direct MS approach for employing reference proteins. This is further corroborated by complementary experiments using the non-binding all galactose glycan Gb4. The binding incompetent monomers show the same extent of glycan association as the respective P dimers do. Of note, in these cases, binding to the P dimers is also not expected due to mutations and MNV P dimers being unable to bind glycans at all (R. Creutznacher et al., 2021). The results indicate that these problems are not limited to neutral glycans but also occurred for sialylated glycans. Nevertheless, for some glycans, MS yields even higher affinities for deamidated GII.4 Saga P dimers than for the wildtype, contradicting results from NMR and HDX-MS, which showed increased flexibility in the deamidated P dimers, suggesting that the structure affects glycan clustering. Here, we re-examine the degree of glycan attachment to different reference proteins to elucidate its origin and general suitability in native MS.

We select eight reference proteins differing in properties. The clustering on ADH appears similar to the glycan distribution on the Saga P dimers while cyt c only presents a small amount of clustering. This further confirms that glycan clustering is influenced by the protein's physicochemical properties. While most proteins show a linear correlation between clustering ratio and glycan concentration, Ubq behaves differently. In general, folded proteins are thought to ionize via the charged residue model (CRM). However, the non-linear glycan clustering behaviour of Ubq hints at ionization following another model, as previously suggested by MD simulations. According to these simulations small proteins can also be ionized through the ion evaporation model (IEM) (Aliyari & Konermann, 2020), which would explain the peculiar behaviour of Ubq and may also play a role for other small proteins. The Cyt c clustering patterns have a non-Gaussian character, which could be caused by varying ESI efficiencies. In contrast to Han et al., 2013 and 2018 (Han et al., 2013; Han et al., 2018), these results indicate an unexpectedly different carbohydrate-clustering propensity, resulting in K_Ds spanning two orders of magnitude. The question arises: which biophysical characteristics caused this effect?

Previous work suggested that the mass and size of the reference proteins did not affect the correction procedure significantly (Han et al., 2013; Sun et al., 2006). Since most carbons in glycans carry hydroxyl groups, glycans have a much higher hydrogen bonding capacity than other small molecules, which could affect the interaction with the protein surface during ESI. Therefore, inspection of the structure of the selected reference proteins reveals distinct ratios of secondary structure elements. Various criteria are tested for correlation to the clustering ratio with the strongest correlation observed for β-sheet contents. We hypothesize that the net-like hydrogen bonding pattern in the β-sheets favours interaction with the glycans in contrast to α-helical structures. This could imply that intercalation occurred during ionization when glycans attached to the protein surface. Furthermore, β-sheets are more labile during the ESI process(Seo et al., 2016), which could facilitate glycan intercalation. Notably, a reasonable correlation is also observed for molecular weight and SASA, which could be related to the higher probability of containing a significant amount of β-sheets with increasing size.

The direct MS approach with a reference protein is therefore not well suited for studying low-affinity glycan binding. It was originally developed for higher-affinity interactions where agreement in KD to other methods was observed (Sun et al., 2006). It has also proven invaluable for other ligand types (Garcia-Alai et al., 2018; Gulbakan et al., 2018; Testa et al., 2013). IMMS demonstrates consistent results for HBGA B and Gb4, implying similar glycan clustering. No major changes in arrival-time distributions are observed upon glycan addition in accordance with small ligands being added, and little structural changes were observed in NMR and HDX-MS (Mallagaray et al., 2019). This contrasts with a previous report (Han et al., 2018) where the experimental and data acquisition parameters differed. Conformational changes were reported for glycan charge states 17^+ and 18^+ up to 800 µM, and we have also checked HBGA B and the 17^+ charge state at 800 µM but find no indication of conformational changes. Notably, we use a modified instrument containing a drift tube as opposed to the travelling-wave ion-mobility device employed by the other group, which could have caused over-activation and unfolding. Another explanation could be an isobaric contaminant with a larger gas phase structure. However, in our case, we are certain that isobaric contaminants are not present and

over-activation of the structures does not occur due to the single near-Gaussian shape of the IMMS peaks.

Overall, our results suggest that reference proteins with similar properties to the protein of interest should be used. Moreover, small proteins that could be affected by IEM upon ESI should be avoided. This explains some of the observed KD discrepancies in the literature. We had previously used Cyt c, both small and mostly α-helical, which overestimated the binding affinity (Han et al., 2018; Mallagaray et al., 2015). Others used a small-sized monoclonal antibody single-chain fragment (scFv 26 kDa, (Han et al., 2013; Han et al., 2018; Sun et al., 2006)), which mostly consisted of β-sheets. While the latter was much better suited, our data show that protein dynamics, as in the deamidated P dimer, further influence clustering (Dülfer et al., 2021; Mallagaray et al., 2019).

The introduction of a non-binding ligand control is an additional assessment parameter to confirm the specific binding for low-affinity ligands, which has also been performed for glycan mimetics (Bücher et al., 2018). After an ADH based protein clustering correction, the non-binder Gb4 expressed no specific binding in any listed glycan concentration, and HBGA B shows a single binding event at 500 µM concentration (compare to the stoichiometry information(Figure S11 and Table S2). The calculated K_D (24 mM) is in accordance with the result from NMR (K_D: HBGA B-tetrasaccharide type 1 12 mM, B trisaccharide 6.7 mM, fucose 22 mM) (R. Creutznacher et al., 2021; Mallagaray et al., 2019). Hence, an appropriate reference candidate is crucial for the K_D determination for low-affinity glycans.

Taken together, in low-affinity glycan binding studies, the K_D calculation is heavily dependent on the degree of glycan clustering on the reference protein. The protein's structure and dynamics seem to heavily influence the degree of glycan clustering. The quantification of direct binding affinity requires the careful selection of a reference protein with similar β-sheet content and even similar structural dynamics to obtain accurate binding occupancy and affinities. The recent introduction of submicron emitters could be a way to reduce or circumvent the problem(Báez Bolivar et al., 2021).

3.4 Contributions

H. Yan and J. Lockhauserbäumer performed native MS measurements: The peak area curve fitting and linear fit analysis was conducted by H. Yan. G. Szekeres performed IMMS analysis. A. Mallagaray and R. Creutznacher provided the NMR data support and norovirus P dimer protein preparation H. Yan wrote the manuscript and J. Lockhauserbäumer revised the manuscript for the publication(Yan et al., 2021)

4. Native MS analysis on binding affinity of fucose attached glycomacromolecules towards human norovirus P-dimer

4.1 Introduction

In this chapter, I performed native MS analysis to determine the stoichiometry of the first generation glycomacromolecules to the GII.4 NoV P dimer which published in Bücher et al. (2018).

Recently, a third and fourth α1,2-linked fucose binding pockets (termed fucose sites 3 and 4 (Figure 4.1)) located between the two outer canonical binding sites were first discovered by X-ray crystallography (Koromyslova et al., 2015)and STD NMR as well as native MS(Mallagaray et al., 2015). The atomic distances between the four-fucose binding pockets are estimated to be 11 Å (fucose sites 1 and 3), 17 Å (fucose sites 1 and 4) and 27 Å (pockets 1 and 2). It has been shown that the terminal α-L-fucose moiety of HBGAs plays a key role in binding onto P-dimer for many NoV genotypes, where the responsible residues at binding pockets 1 and 2 are highly conserved(Hansman et al., 2011). However, the binding affinity to L-fucose is rather weak and in the low millimolar range. In order to increase binding by avidity, multivalent structures presenting several fucose ligands can be used. Guiard and Rademacher have shown that fucosylated polymers (60-100 kDa) had 106 fold enhanced binding to NoV -like particles(Guiard et al., 2011; Rademacher et al., 2011). However, little is known about the underlying mechanism of the ligand-virus binding and whether it is possible for a multivalent fucose polymer to bind simultaneously at all four binding sites.

Therefore, in this study, we report on the synthesis of a first generation of fucosylated precision glycomacromolecules and their use as multivalent model structures for the investigation of binding to NoV P-dimers. Precision glycomacromolecules were synthesized following previously established protocols for the stepwise assembly of

tailor-made building blocks on solid support, giving an oligo (amidoamine) scaffold that allows for the conjugation of sugar ligands in the side chains(Ebbesen et al., 2017; Geng et al., 2007; Gerke et al., 2017; Igde et al., 2017; Ponader et al., 2014; Ponader et al., 2012). By choosing the sequence of building blocks during solid phase assembly, the number and position of fucose ligands along the scaffold can be controlled. The binding of the fucosylated glycomacromolecules on NoV P-dimers were investigated using native MS

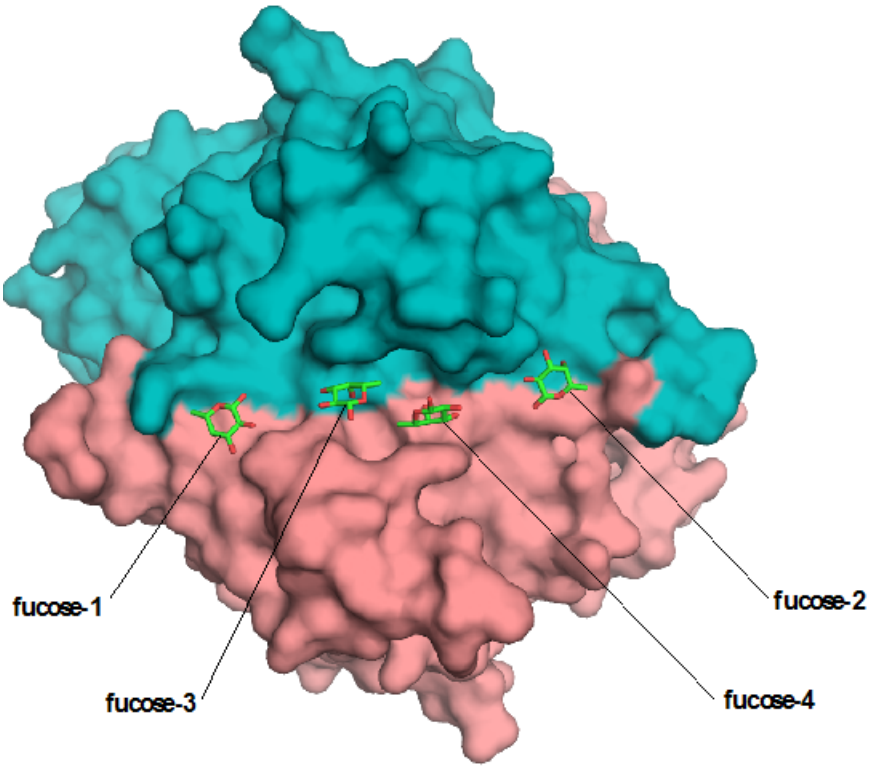


Figure 4.1 Structure of NoV GII.10 P-dimer (monomers surface representation) in complex with α-L-methylfucose units (green sticks). Binding of methyl-α-L-fucopyranoside to the four binding sites in P-dimer was demonstrated to be a dose-dependent and step-wise process, in which pockets 1 and 2 (exhibiting a distance of about 27 Å) displayed the highest affinities towards methyl-α-L-fucopyranoside(Koromyslova et al., 2015).

4.2 Result

4.2.1Synthesis of the designed fucose/galactose attached glycomacromolecules

The focus of this study was the synthesis of fucose-presenting precision macromolecules and their use as multivalent model structures to investigate binding to NoV GII.4 P dimers. We applied previously reported stepwise assembly of tailor-made building blocks on solid support, or so-called solid phase polymer synthesis (Ponader et al., 2014; Ponader et al., 2012) for the sequence-controlled attachment of α-L-fucose in the side chain of monodisperse oligo (amidoamine) scaffolds. Two different building blocks were used a hydrophilic spacer building block (EDS) and an alkyne functionalized building block (TDS)(Ponader et al., 2012)(Figure 4.2). All building blocks possess a free carboxyl and an Fmoc-protected amine group allowing for chain elongation via standard Fmoc-peptide coupling protocols. The sequence of building blocks during the chain elongation thereby gives the primary sequence of the macromolecular scaffold and allows for variations of the number and position of sugar ligands, as well as the overall length of the scaffold. After assembly of the scaffold, the alkyne side chains are conjugated with azido-functionalized α-L-fucose or D-galactose derivatives via Cu (I)-mediated 1,3-dipolar cycloaddition (CuAAC) to give the final precision glycomacromolecules (Figure S12). The azido-functionalized D-galactose was obtained following literature protocols(Geng et al., 2007) and used as anomeric mixture with an α/β-ratio of 1:4 as both anomers could serve as a negative control for binding. Azido-functionalized α-L-fucose derivative, 1-azidoethyl-2,3,4-tri-O-acetyl-α-L-fucopyranoside, was also synthesized, adapting a protocol applying H2SO4-silica catalyst for α-L-fucose with an azido ethyl linker(Hoshino et al., 2012). Here, isolation of the α-anomer is important, since this is the “native type” of monosaccharide that binds onto the P-dimer(Rademacher et al., 2011).

Overall, nine precision glycomacromolecules were synthesized presenting up to four fucose ligands (Figure 4.2). As monovalent ligand 1, an oligomeric backbone with three EDS building blocks to each side of the fucose side chain was synthesized in order to account for a similar overall chain length in comparison to the other glycomacromolecules of this study. For divalent glycomacromolecules, a series with varying interligand distance going from zero to three EDS spacing building blocks between the fucose carrying building blocks was synthesized (glycomacromolecule 2-5). Assuming an all-stretched conformation of the oligomeric backbone, the distance

between two neighboring fucose ligands with no additional spacer building block would be ~ 31 Å and thus correspond roughly to the distance between fucose sites 1 and 2 (~ 27 Å) (Rademacher et al., 2011). However, we should rather assume a coiled structure of the glycomacromolecules in solution, since this coiled structure was previously discovered using fluorescence correlation spectroscopy (Ponader et al., 2014).

The glycomacromolecules have been isolated as crude products after deprotection of the carbohydrate side chains and obtained after cleavage from the solid support with purities of 90-95% (analyzed by integration of UV signal at 214 nm using RP-HPLC). All glycomacromolecules were then further purified by ion exchange chromatography (quaternary ammonium, acetate form)(Sun et al., 2006) followed by semi-preparative RP-HPLC and obtained with final purities ≥97% (analyzed by integration of UV signal at 214 nm using RP-HPLC (data not shown)) and the structures were confirmed by MS (Bücher et al., 2018).

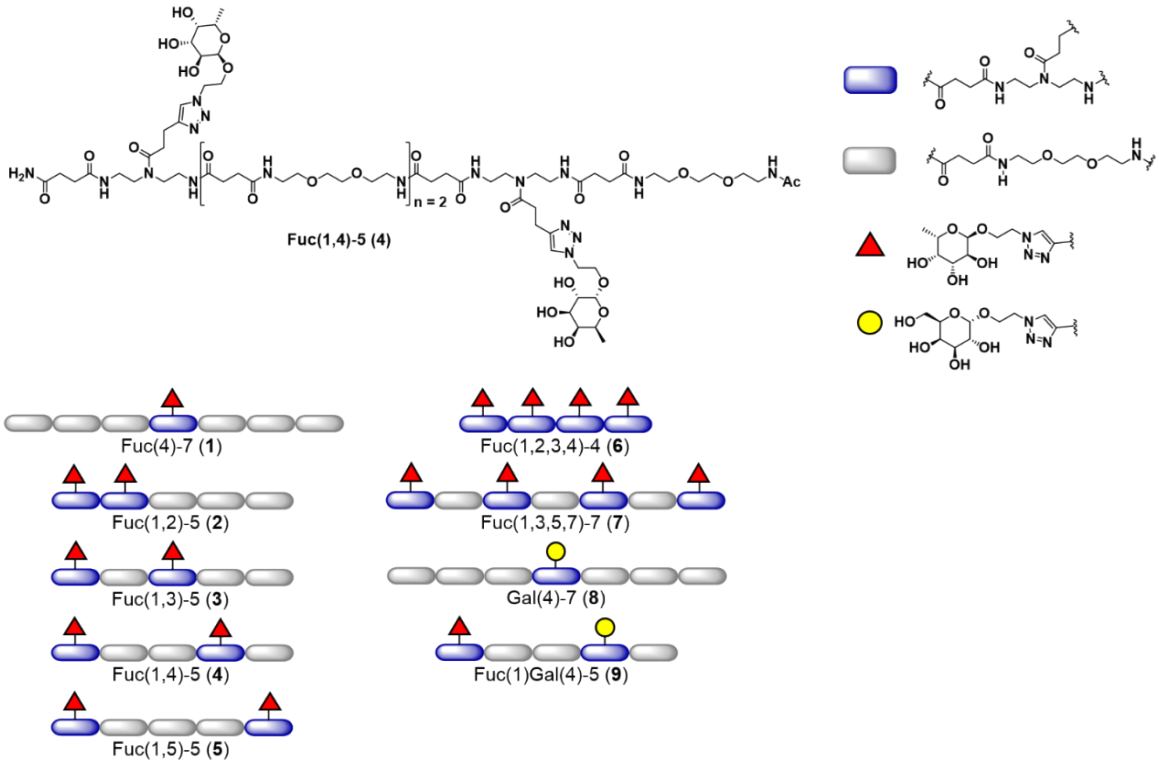


Figure 4.2: Overview of the structures of precision glycomacromolecules 1-9 presenting α-L-fucose (red, triangle) and D-galactose (yellow, round). Hydrophilic spacer building block (EDS) and (ii) an alkyne functionalized building block (TDS) coloured in grey and blue cube, respectively.

To obtain insights into the conformation of fucosylated glycomacromolecules in solution, MD simulations were performed. Distance distributions of neighboring fucose ligands (center of mass of focuses) for glycomacromolecules with different spacing range from approximately 10 Å to 20 Å (see Figure 4.3b).

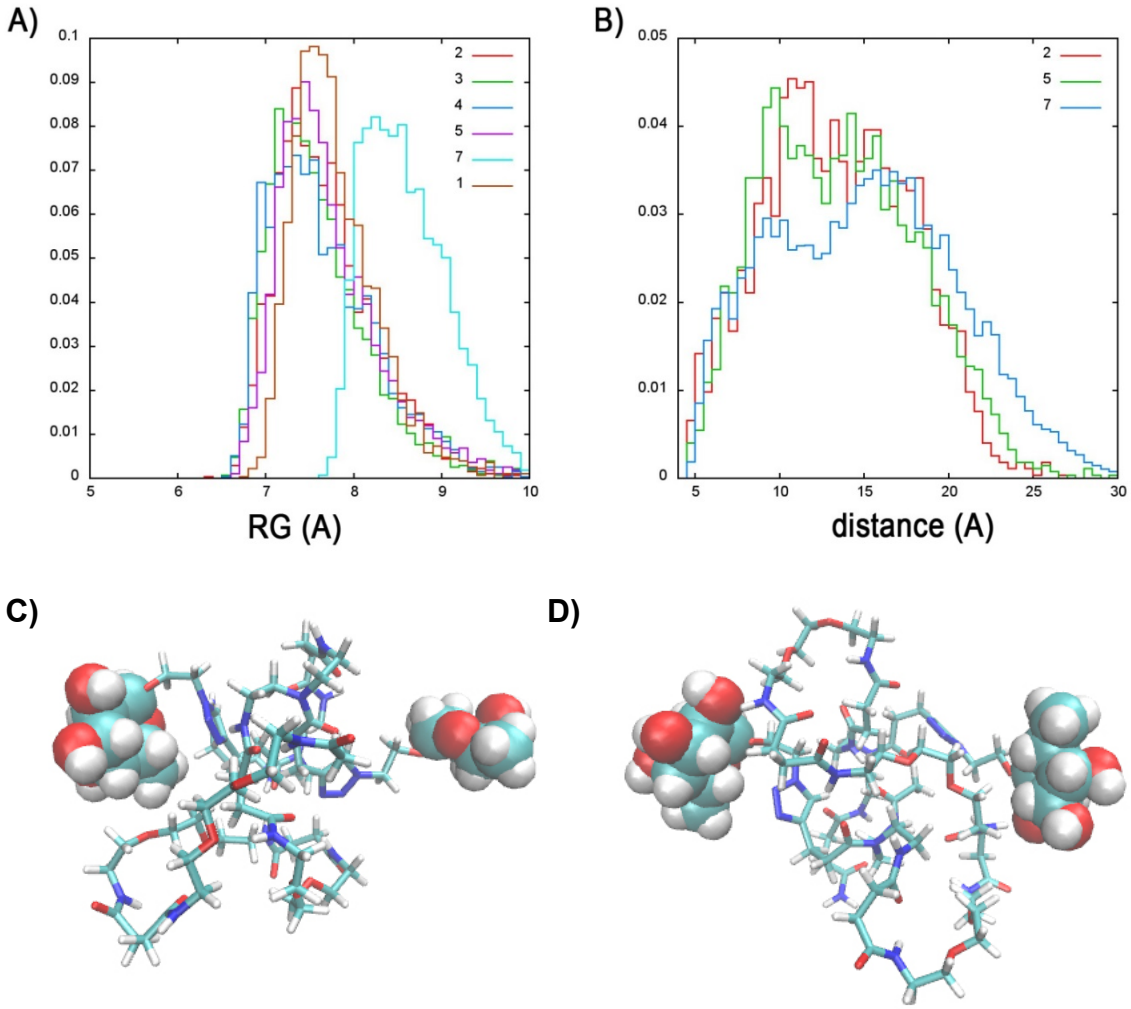


Figure 4.3 Glycomacromolecule simulations. A) distribution of the radius of gyration for molecules 1, 2, 3, 4, 5 and 7; B) distribution of the distance between fucose ligands for molecules 2 (purple), 5 (blue) and 7 (green); C) snapshot of glycomacromolecule 2 and D) snapshot of glycomacromolecule 5.

Thus, different spacing of ligands along the backbone does not become evident as differences in ligand spacing in the coiled conformation (Figure 4.3c and d). However, upon contact to a protein receptor a change in the scaffold's conformation might occur to accommodate binding sites of the receptor (Erba & Zenobi, 2011; Sun et al., 2006). Furthermore, MD simulations indicate that longer backbones lead to an increase in the

radius of gyration of the overall glycomacromolecule (Figure 4.3a). Therefore, in order to keep the overall chain length constant, EDS building blocks were added to the backbone to obtain an overall chain length of five building blocks for all divalent macromolecules. Additionally, two tetravalent glycomacromolecules were synthesized having zero or one spacer building block between the sugar side chains 6 and 7. As control, a monovalent galactose functionalized glycomacromolecule 8 was synthesized as well as a first hetero-divalent structure presenting one fucose and one galactose ligand 9.

4.2.2 Binding studies of glycomacromolecules towards NoV GII.4 P dimer

In order to decipher the potential multivalent binding of fucose-presenting glycomacromolecules to P-dimer, native MS measurements were performed using GII.4 P dimers. In short, native MS employs ESI to preserve non-covalent complexes in the gas phase and therefore allows analysis of the number of glycomacromolecules bound to the P-dimer protein(Han et al., 2013; Mallagaray et al., 2015). Since binding affinities of P-dimers for glycans are low, the reference protein method is used to correct for unspecific clustering of ligands during the ESI process(Sun et al., 2006). It has previously been shown that the size of the reference protein does not affect the unspecific clustering (Wang et al., 2005), we therefore chose cytochrome c to avoid any spectral overlaps. As small ligands only marginally influence ionization efficiency, the peak areas of bound and unbound P-dimer can be translated into concentrations and the KD be directly retrieved from the law of mass action. As is well described in the literature, multivalent ligands such as the precision glycomacromolecules can undergo different binding modes when binding to a multivalent protein(Cecioni et al., 2015). In the native MS experiment, the intermolecular complex formation can be observed and particularly the stoichiometry of several ligands binding to one protein receptor can be detected(Uetrecht & Heck, 2011). Furthermore, it is possible to ramp ligand concentrations like in a titration to deduce cooperativity. In this case, we have analyzed 2-3 concentrations, which gave consistent KD value (Table 4.1), indicating no strong cooperative effects.

Figure 4.4 shows an exemplary model after correction for unspecific clustering and normalization to the highest unbound protein peak recorded for glycomacromolecules 5 and 6 binding to GII.4 P-dimer at different concentrations. For higher concentrations (200 µM), it can be seen that up to two glycomacromolecules 5 can bind to GII.4 P-dimer at the concentrations tested (p2) (Figure 4.4). From the experiment we cannot say, which of the fucose ligands binds to the protein or which of the protein binding sites is occupied. However, an apparent binding constant K_D for the first binding event (p1) can be derived from the intensities of the different complexes giving an indication for the overall affinity of the glycomacromolecules towards GII.4 P dimer.

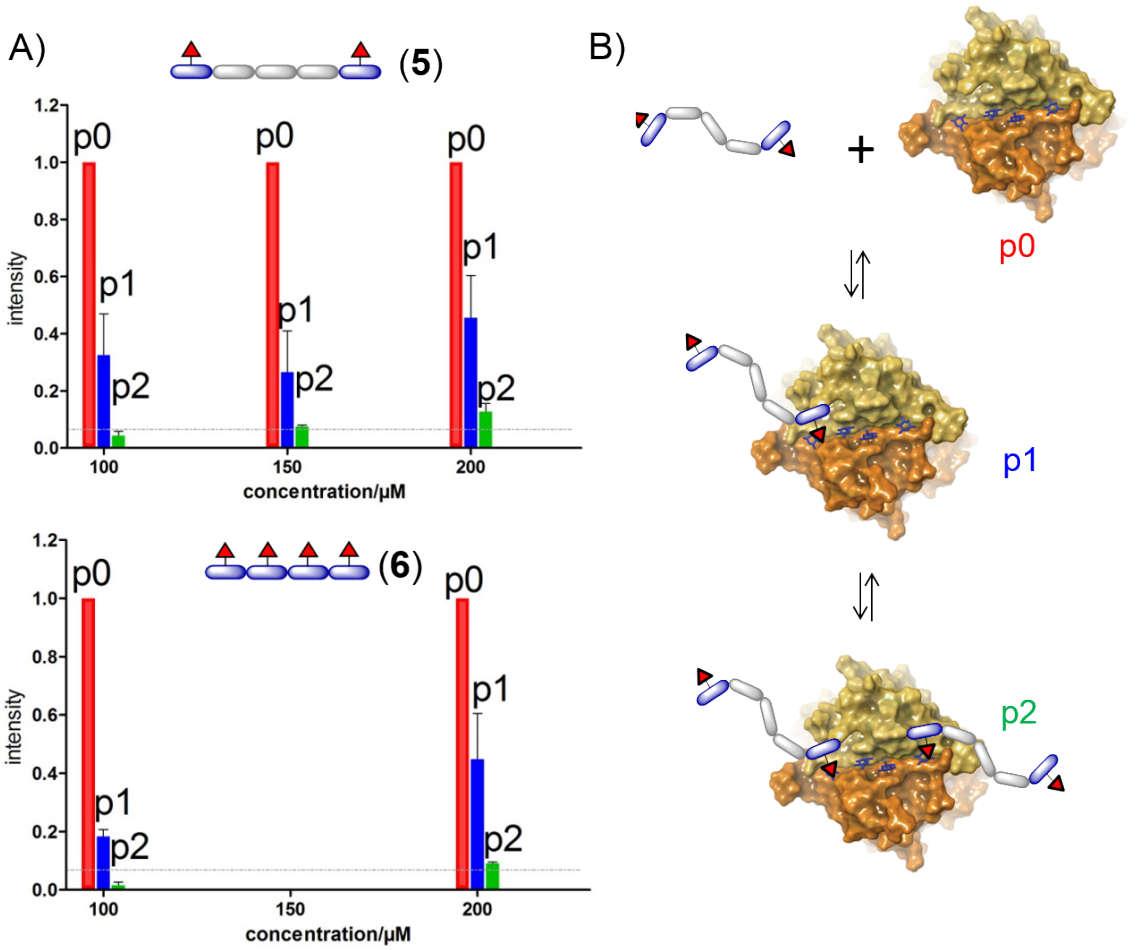
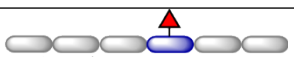
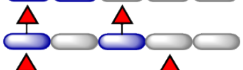
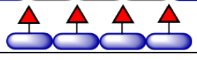


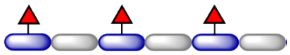
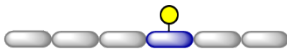
Figure 4.4 Native MS analysis of glycomacromolecule binding. (A) analysis of interaction of NoV GII.4 P-dimer with glycooligomers 5 and 6 at indicated concentrations in 300 mM AmAc & 20 mM TEAA pH 7; (B) proposed binding of divalent glycomacromolecule 5 with P dimer.

4.3 Discussion

As shown in Table 4.1, we obtained the data for the binding strength of this group of glycomacromolecules using GII.4 P dimers. Initially, we examined the negative control, the galactose carrying structure 8. We observed some residual binding at higher concentration with weak binding affinities (K_{D1} = 2.4+/-0.6 mM). Based on the available STD NMR data, we know there is no interaction between NoV and galactose(Dülfer et al., 2021). The observed binding therefore is considered a result from variation in electrospray quality or marginal backbone contribution. Signals of this residual binding do not exceed 10% including standard deviation for p1 and are therefore defined as threshold for stoichiometry determination. Next, the number of glycomacromolecules was examined that can bind to GII.4 P-dimer in dependence of their valences. We found that at the highest concentration evaluated, ligand-protein complexes with up to two glycomacromolecules per protein were detected for divalent systems 2-5. All other glycomacromolecules show only binding of one glycomacromolecule per protein above threshold. This is particularly interesting for the higher valent structures 6 and 7, as they present four instead of two fucose units but show no enhanced binding and formation of only single ligand-protein complexes. At this point, we assume this is an effect of the overall size of the glycomacromolecules as both simulation and experimental data have shown increasing size with increasing number of fucose units. Larger and in this case higher valent glycomacromolecules could be sterically hindered to access more than one binding site on the P-dimer.

Table 4.1 Results of native MS measurements of glycomacromolecules **1-9** as well as HBGA B tetrasaccharide as positive controls binding to GII.4 P-dimer

Entry	Glycomacromolecule	K_{D1} (µM)[a]	Concentration glycomacromolecule		
			100	150	200
	Schematic structures		Maximum number of glycooligomers per P-dimer		
1		230 ± 50	1	1	1
2		310 ± 90	1	1	2
3		240 ± 60	1	2	2
4		340 ± 130	1	1	2
5		290 ± 90	1	1	2
6		380 ± 100	1		1

7		330 ± 80	1	1	1
8		2400 ± 600	1	1	1
9		370 ± 90	1		1
HBGA B		110 ± 30	2	3	3

[a]Average value for the dissociation constant for the first glycomacromolecule bound, errors represent the standard deviation.

When we look at the K_{D1} values derived from the native MS experiments, all fucose containing glycomacromolecules have similar affinities (K_{D1} of 200-400 µM). In order to compare K_{D1} values with known ligands of P-dimer, HBGA B tetrasaccharide type 1 was tested with GII.4 P-dimers. The natural ligand has a slightly higher affinity (K_{D1} of 110 ± 30 µM). The monosaccharide ligand, α-L-methylfucose, could not be used as reference as the mass difference is too small to allow for resolution between bound and unbound species. If the glycomacromolecules would be able to address more than one binding site of P-dimer simultaneously, we would expect an increase in affinity. We therefore conclude that only single fucose ligands per glycomacromolecule bind and no avidity effects are observed in the different multivalent structures as is also supported by the heterovalent species 9 behaving similar to the fucose only structures.

Based on the ligand distances derived from the modeling and light scattering study(Bücher et al., 2018), glycomacromolecules in their coiled conformation could bridge binding site 1 with fucose binding sites 3 (also sites 2 and 4) (Figure 4.1), although as previously discussed this is not supported by similar K_{D1} values in native MS experiments. In theory, multivalent ligands based on flexible scaffolds such as the glycomacromolecules could also undergo a conformational change paying an entropic penalty in order to increase ligand-protein interactions(Braun et al., 2011; Hoshino et al., 2012). Thereby glycomacromolecules could theoretically also bridge the outer binding sites 1 and 4. To test this, divalent glycomacromolecule 4 was used in an electron spin resonance (ESR) experiment similar to previously described experiments by Wittmann and Drescher.[17] No significant differences in the absence and presence of GII.4 P-dimer were observed suggesting that the conformational ensemble of the glycomacromolecule remains unaltered upon interaction with P-dimer(Bücher et al., 2018).

Recent studies based on native MS(Han et al., 2018) and simple docking(Fiege et al., 2017) show that larger oligosaccharides such as blood group A or B tri- and tetrasaccharides do not bind into all four binding sites simultaneously. Along those lines, monovalent binding of glycomacromolecules could be attributed to steric effects resulting from the scaffold blocking binding of a second ligand. Since there are no indications of simultaneous binding of several fucose residues of the same glycomacromolecule, at this time, we cannot exclude the influence of other structural parameters leading to an overall monovalent binding mode, such as an improper spacing of ligands along the scaffold or the conformational flexibility of the scaffold that are well-known to affect multivalent ligand-receptor binding(Braun et al., 2011; Cecioni et al., 2015; Ponader et al., 2014).

4.4 Contributions

R.Creutznacher produced and purified the proteins and K. Bücher designed and synthesized the eleven glycomacromolecules. H. Yan conducted the native MS analysis of glycomacromolecules-Saga P dimer interaction. KSB wrote the main content of the manuscript for the publication(Bücher et al., 2018). H. Yan and R.Creutznacher contributed the regarding parts of the manuscript writing.

5. HDX MS analysis of strain-dependent glycan binding manner on human norovirus P dimer species

5.1Introduction

In the chapter 5, I performed native MS analysis for norovirus P dimers quality control and HDX measurement on the three P dimer species in and absence of ligand of interest. The data is published in Dülfer et al. (2021)

NMR measurements identified a spontaneous deamidation of N373 with subsequent formation of an iso-aspartate (iD) in GII.4 Saga P dimers that strongly attenuates glycan binding(Mallagaray et al., 2019). This deamidation appears to be site specific and occurs in GII.4 MI001 P dimers as well, whereas it is absent in GII.10 Vietnam 026 and GII.17 Kawasaki 308 P dimers, which carry a Gln or Asp at the equivalent position [15]. HDX-MS measurements showed binding site protection only in the wildtype P dimer, confirming the loss of HBGA B trisaccharide binding in deamidated P dimers, which correlated with increased flexibility in the P2 domain compared to the wildtype P dimer. In the chapter, I mainly conducted native MS analysis for P dimer quality control and participated in HDX experiment to examine whether glycan binding or deamidation can induce distinct structural dynamics on P dimers in (Dülfer et al., 2021)

HDX-MS measures the exchange of protein backbone hydrogens to deuterium in solution. As this exchange strongly depends on solvent accessibility and hydrogen bonding patterns, the method can provide information about regions involved in ligand binding as well as changes in protein dynamics in solution. This makes it a valuable technique for identification of glycan induced structural dynamics in different strains as well as elucidation of altered protein dynamics in deamidated P dimers. While P dimers across strains are structurally highly similar, their glycan binding behavior and infectivity

is highly variable, leading to the hypothesis that varying structural dynamics are linked to the different profiles.

Therefore, we specifically investigated binding of HBGA B trisaccharide and L-fucose to P dimers of GII.4 Saga, GII.4 MI001, GII.17 Kawasaki 308 and GII.10 Vietnam 026, significantly extending previous NMR studies (R. Creutznacher et al., 2021; Mallagaray et al., 2019). GII.4 MI001 infects humans and mice [24] and has been chosen as comparison to the almost identical strain GII.4 Saga. GII.17 Kawasaki 308 is an emerging strain and the less abundant GII.10 Vietnam 026 is capable of binding four fucose molecules per P dimer. Our results reveal protection of the canonical binding site in all wildtype P dimers. While glycan binding behavior in GII.4 Saga and GII.4 MI001 strains is identical, distinct glycan induced dynamics are observed in GII.17 Kawasaki 308 and GII.10 Vietnam 026.

5. 2 result

5.2.1 Assessment of monomer/dimer ratio on different P dimer species.

Prior to HDX-MS analysis P dimers were subjected to native MS for quality control. Furthermore, ion exchange separated wildtype and fully deamidated GII.4 Saga P dimers were measured for comparison. GII.17 Kawasaki, GII.10 Vietnam, wildtype GII.4 MI001 and wildtype GII.4 Saga P domains showed dimers with the expected molecular masses, apart from a small fraction of tetramers likely formed during the electrospray ionization (ESI) process (Figure 5.1). Interestingly, both deamidated GII.4 P domains were also present as monomers. Increased monomer fractions correlate with the extent of N373 deamidation: 16% monomers are detected for the 64% deamidated GII.4 MI001 sample and 32% monomers are found for the 100% deamidated GII.4 Saga sample (Table S4).

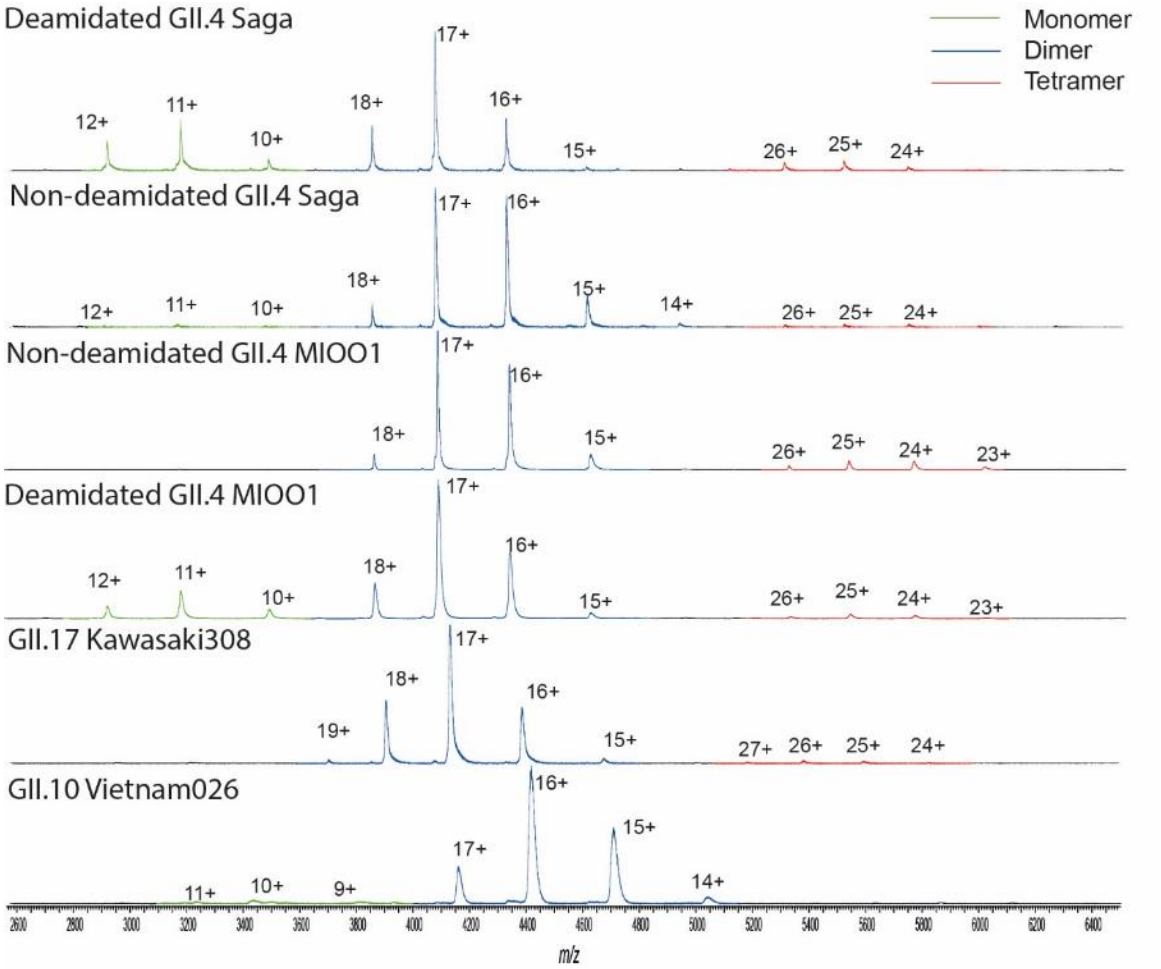


Figure5.1 Native MS spectra of different norovirus P dimers. Norovirus P domain (3μM) from non-deamidated/wild type and deamidated GII.4 Saga, non-deamidated/wild type and deamidated GII.4 MI001, GII.17 Kawasaki and GII.10 Vietnam strains, respectively, measured at 300 mM ammonium acetate solution at pH 7. non-deamidated/wild type and deamidated GII.4 Saga, non-deamidated/wild type and deamidated GII.4 MI001, GII.17 Kawasaki and GII.10 Vietnam P dimers mainly formed as dimers, monomers was presented in deamidated Saga and MI001 species. The formed tetramers of the P dimers were might due to ESI process.

5.2.2 HDX MS analysis of glycan binding induced changes on different P dimers species

It is known that Fucose and HBGA binding pocket is located at cleft region between two P monomers(Mallagaray et al., 2015), the expected HDX exchange were found as well in this area during the fucose and B tri-saccharides binding test, respectively. The binding of the glycan ligands would lead to the hindrance of the deuterium replacement to Hydrogen molecule and demonstrated that the regarded binding pockets are occupied

by glycans. Protected regions in wildtype GII.4 MI001 P dimers are highly similar to GII.4 Saga P dimers (Mallagaray et al., 2019) for both glycans (canonical binding site G443, Y444 and residues 283–303) indicating dataset validity. In addition, protection of a β-sheet region in the top cleft of the P2 domain (residues 333–353) can be detected in GII.4 MI001 in presence of HBGA B trisaccharide. Chemical shift perturbations (CSP) in this region could also be seen in NMR experiments with GII.4 Saga P dimers in presence of glycans(Mallagaray et al., 2019). Overall, protected regions in GII.4 MI001 match with regions showing CSPs in GII.4 Saga NMR data, suggesting that both strains respond similarly to glycan binding.

Protection of residues 333–353 in the P2 domain can be seen in all strains. For GII.17 Kawasaki P dimers, significant protection in presence of HBGA B trisaccharide is only present in this specific region. When incubated with fucose, additional protection of the canonical glycan binding site (G443, Y444) and residues 269–286, located in the protein centre below the P2 domain, can be detected (Figure 5.2A). GII.10 Vietnam P dimers also show protection in their canonical binding site (G451, Y452) and the β-sheet region in the binding cleft of the P2 domain (residues 337–364 and 379–399) is protected as well (Figure 5.2D). All differences depicted in Figure 3 can only be seen in the second, highly deuterated peak distribution. The lowly deuterated peak distribution showed no significant differences between the unbound and the glycan-bound state in any of the strains indicating that either only the highly deuterated species can bind glycans or labelling time was too short to detect deuteration differences in already strongly protected regions.

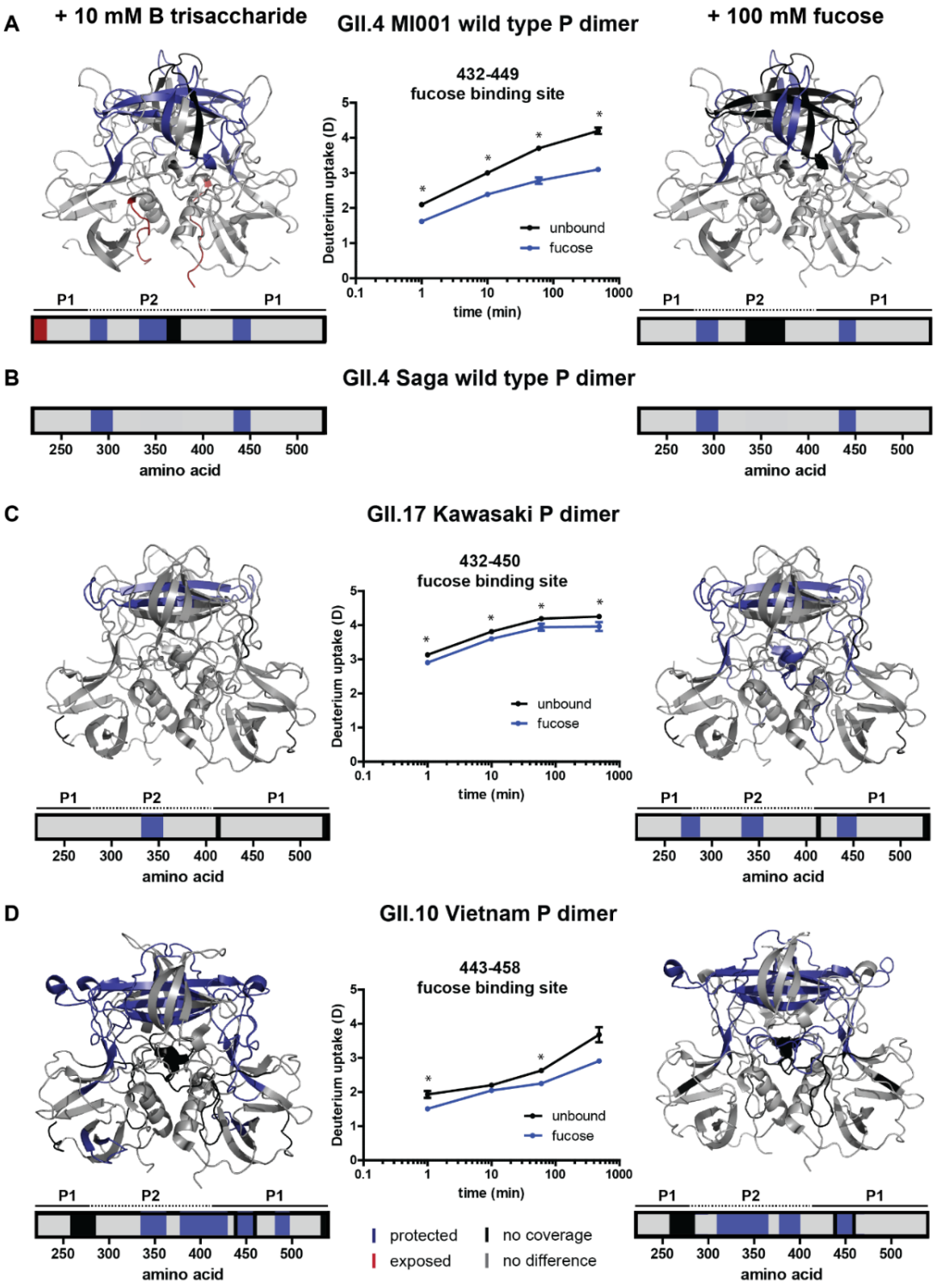


Figure 5.2 HDX differences upon glycan binding in human norovirus P dimers (A) GII.4 MI001; (B) GII.4 Saga; (C) GII.17 Kawasaki 308 and (D) GII.10 Vietnam. Protected regions in wildtype GII.4 Saga P dimers are shown for comparison(Mallagaray et al., 2019). Depicted are protein regions with significant deuterium uptake differences in the second (main) peak distribution between unbound P dimers and P dimers with either 10 mM HBGA B trisaccharide or 100 mM fucose ($p < 0.05$, Student's T-test and $\Delta D > 2x$ pooled average SD). Deuterium uptake plots show

significant (*) protection of the canonical fucose binding site. The deuteration difference at the 8 h time point of GII.10 Vietnam is not considered significant based on the applied criteria because the fucose state only represents a single measurement and thus no Student's T-test can be applied. Bar graphs and coloured structures illustrate regions of P dimers, which get significantly more protected (dark blue) or exposed (red) upon interaction with glycans. Areas coloured in grey showed no significant difference in the chosen HDX time regime and black areas have no peptide coverage. P1/P2 refers to the two domains of the P dimer (shown in Figure 1).

5.3 Discussion

In this chapter, we addressed the differences in structural responses to glycan binding of norovirus P dimers of the Asian epidemic strain GII.17 Kawasaki 308, the rarely detected strain GII.10 Vietnam 026 and the GII.4 MI001 strain, which belongs to the highly epidemic GII.4 genotype and has been shown to infect mice as well (Taube et al., 2013).

For GII.4 MI001, regions protected from H/D exchange were almost identical to the earlier investigated GII.4 Saga P dimer (canonical binding site G443, Y444 and residues 283–303) (Mallagaray et al., 2019), apart from additional protection in the upper P2 binding cleft (residues 333–353). Involvement of this region has been seen in NMR data of GII.4 Saga P dimers as well (Mallagaray et al., 2019). Furthermore, a recent NMR study suggests identical glycan binding behavior of both GII.4 strains (R. Creutznacher et al., 2021). The same study also shows that MNV P dimers do not bind HBGAs, underscoring that infectivity of GII.4 MI001 in mice cannot be explained by different glycan-induced dynamics between GII.4 Saga and MI001 in line with our observations.

GII.17 Kawasaki P dimer crystal structures with fucose and HBGA A trisaccharide show backbone interactions in T348 and G443 and side chain interactions in R349, D378 and Y444 (R. Creutznacher et al., 2021; Koromyslova et al., 2017). When incubated with HBGA B trisaccharide and fucose, protection from HDX is observed for residues 333–353 corresponding to interactions with T348 and R349. In the presence of 100 mM fucose, the canonical binding site (G443, Y444) is protected, as well as residues 269–286, which

cannot be explained by the known interactions from the crystal structures. This region is located below the glycan binding cleft in the protein center, so protection from HDX could rather be the result of a long-distance structural change than of direct interaction with fucose. It would be interesting to see how long-distance structural changes would further propagate into the S domain in VLPs and if they would influence the dynamic P domain lift off from the S domain that has been seen for different norovirus strains (Jung et al., 2019; Smith & Smith, 2019; Tubiana et al., 2017).

For the GII.10 Vietnam strain, binding of two HBGA B trisaccharide molecules and up to four fucose molecules has been seen in crystal structures (G. S. Hansman et al., 2011; A. D. Koromyslova et al., 2015). Compared to GII.4 MI001 and GII.17 Kawasaki, we see protection in more protein areas for both HBGA B trisaccharide and fucose, which mainly corresponds to the known glycan interactions summarized in Table 1. Interestingly, we see a protection of several residue stretches that cannot be explained by known glycan interactions. Residues 311–336 belong to an unstructured region below the P2 binding cleft. In presence of HBGA B trisaccharide, protection of the aforementioned residues is not present under the chosen conditions. A possible explanation could be that these changes in dynamics are triggered by occupation of binding sites 3 and 4 in the P2 cleft, which so far has not been seen for HBGA B trisaccharide at similar concentrations (A. D. Koromyslova et al., 2015). HBGA B trisaccharide binding is mainly mediated by the fucose residue, with an additional interaction of galactose with G451 and some water mediated interactions (G. S. Hansman et al., 2011). In our data we detect protection of residues 483–496 on the bottom of the P dimer in addition, which could be a long-range effect not triggered by fucose alone.

Taken together, P dimers of all investigated strains showed protection of the upper P2 binding cleft (residues 333–353) underscoring the importance of this region for glycan binding as well as the consistence between datasets. Protection of the canonical glycan binding site (G443, Y444 for GII.4 and GII.17; G451, Y452 for GII.10) was also detected in all strains and for all glycans apart from HBGA B trisaccharide binding with GII.17 Kawasaki P dimers. HBGA B trisaccharide could have a lower binding affinity in GII.17

Kawasaki compared to the other strains that leads to smaller deuteration changes that are below the detection limit in the current experimental setup. We also noticed that the GII.17 Kawasaki datasets have a higher back exchange (D/H) than the other datasets so that small glycan induced deuteration changes are more likely to be lost during the measurement. GII.4 P dimers show protection of residues 285–298, which is absent in GII.17 and GII.10 P dimers. Interestingly, P dimers of the more prevalent strains GII.4 and GII.17 (Chan et al., 2017; de Graaf et al., 2015) show less changes in HDX upon glycan binding compared to GII.10 Vietnam, which is rarely detected in patients (G. S. Hansman et al., 2011).

5.4 Contributions

R. Creutznacher and A. Mallagaray produced and purified the HNoV GII.4 Saga, GII.4 MI001 P dimers proteins, R. Hansman provide the HNoV GII.17 Kawasaki and GII.10 Vietnam P dimers. H. Yan and J. Dülfer performed the quality control of the protein species via native MS. J. Dülfer performed the HDX MS measurement of HNoV GII.4 Saga, GII.4 MI001, GII.10 P dimers proteins in the presence of two respective glycan ligands while H. Yan conducted the GII.17 Kawasaki P dimer. J. Dülfer wrote the manuscript for the final publication

6 Material and Methods

6.1 Expression and purification of P dimers

GII.10 Vietnam 026 (residues 224–538, Genbank number AF504671), GII.4 MI001 (residues 225–530, Genbank number KC631814) and GII.17 Kawasaki 308 2015 (residues 225–530, Genbank number LC037415) P domains protein, respectively, were expressed via E. coli system and purified by size-exclusion chromatography (GE Healthcare column: Superdex 26/600 75 pg column). The protein sample was eluted in 20mM buffer at pH 7.3 (A. Mallagaray et al., 2019). Furthermore, the non-deamidated GII.4 MI001 P dimers

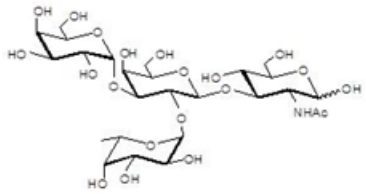
were stored in 4°C at 20 mM NaAc, 100 mM NaCl pH 4.9. GII.10 Vietnam P dimer was stored at 25 mM TrisHCl, 300 mM NaCl buffer solution at pH 7.3 and GII.17 Kawasaki P protein were stored at 4°C with 20 mM NaAc, 100 mM NaCl buffer solution at pH 4.9.

6.2 Reference proteins

The used reference protein candidates were purchased in Sigma-Aldrich (Merck, Germany), the CAS number is listed below (Cyt c, 9007-43-6), (Ubq, 75986-22-4), (CA, 9001-03-0), (ADH, 9031-72-5), (LDH, 9001-60-9), (Myo, 100684-32-0), (apo-TFF, 11096-37-0). Green fluorescent protein(GFP). The GFP protein were gifted and stored in the Lab, all the proteins were stored at 4°C for the further experiment

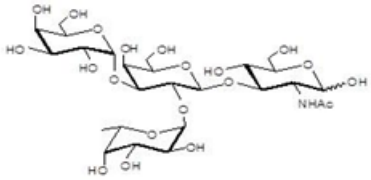
6.3 Glycan of interest

The Three glycans were purchased at Elicityl-Oligotech, dissolved in milliQ water. (1) HBGA A tetra-saccharide type 1 (GalNAcα1-3(Fucα1-2)Galβ1-3GlcNAc), (2) HBGA B tetrasaccharide type 1(Galα1-3(Fucα1-2)Galβ1-3GlcNAc). (3) Gb4 (GalNAc β1-3 Galα1-4Galβ1-4Glc)(Figure 3.6).



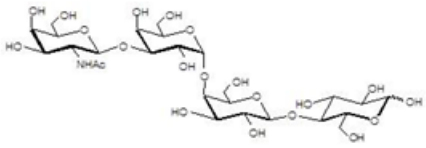
HBGA A

GalNAcα1-3(Fucα1-2)Galβ1-3GlcNAc



HBGA B

Galα1-3(Fucα1-2)Galβ1-3GlcNAc



Gb4

GalNAc β1-3 Galα1-4Galβ1-4Glc

Figure6.1 Chemical structure of three glycan ligands. HBGA A and B are the defined binders of hNoV, Gb4 is served a non-binder control.

6.4 Native MS experiment

1 µM Saga P dimer and 3 µM of the reference proteins were mixed with glycan ligands as the experimental set-up for the native MS analysis. 150 mM ammonium acetate (Sigma-Aldrich, Merck, 98%) at pH7 were used for buffer exchange via Vivaspin 500 (MWCO 10000, Sartorius) at 12000 g at 4°C. Mass spectra were acquired in positive ion mode on a Waters/Micromass LCT mass spectrometer (Waters, UK and MS Vision, the Netherlands) with a nano-ESI ion source. The capillary and cone voltages as well as pressures were adjusted for glycan-protein interaction measurement. The capillary and cone voltage were set to 1.2KV and 240V, respectively. The argon gas pressure was applied to 6.5 x 10-2 mbar to 6.8 x 10-2 mbar in the source to minimize the possible protein-glycan complex in-source dissociation. Masslynx (Waters corporation) software was used to record the raw data and OriginPro 2016 SR2 (OriginLab) software was utilized for peak area calculation and peak curve fitting. The non-specific glycan ligand clustering was assessed based on the method(Sun et al., 2006). Python program including Pandas, Numpy and Matplotlib and seaborn modules were imported to generate script for the raw mass spectra data analysis. The sklearn.linear model was selected according to the scatter point obtained from the analysed data for the linear fit. The heat-map plot was generated via heatmap package from Matplotlib module.

6.5 Titration experiment of glycan ligand-P dimer interaction and K_D determination

To determine the K_D value of glycan ligand-P dimer interaction, the titration measurement is set with a fix concentration of GII.4 Saga P dimers varying from different concentration of glycan ligands (100µM to 500µM). The obtained raw mass spectra data are analysed in 10min to avoid the pH change influence during the nano-spray. LCT mass spectrometer is designed to reduce the in-source dissociation risk through a short length Hexapole device coupling with ToF mass analyser. HBGA A, HBGA B and non-binder control (Gb4) were tested, respectively, to determine the K_D of the first binding event. The calculation method is based on equation listed in the chapter 1.6.1. as described in (Sun et al., 2006). In general, the binding constant is related to the concentration of the protein, ligand and protein-ligand complex in solution. It assumed that positive mode electrospray ionization and detection efficiency is identical for protein-glycan ligand

complex species and the protein molecule. Therefore, the rate between the peak area of the non-bound protein and bound protein complex obtained from the raw spectra is considered to be equal to ratio of free protein and protein-ligand complex in solution. In contrast, same parameter is applied for quality control of on purified GII.4 Saga(deamidated and non-deamidated), GII.10 Vietnam206, GII.4 MI001 (deamidated and non-deamidated) and GII.17 Kawasaki P protein, respectively. MassLynx software (Waters) was applied to peak assignment of charge states and protein molecular weight estimation. Peak fitting software (OriginPro 2016 SR2) was utilized to calculate the rate of P monomer/dimer acquired for the raw mass spectra data of each P species.

6.6 Ion mobility mass spectrometry

IMMS measurement is performed in a modified Waters Synapt G2-S mass spectrometer from the collaborator group. The instrument is customized with a linear drift-tube ion mobility cell and Helium gas is applied under 2.4mbar pressure. 3µM GII.4 Saga P dimer and 0, 100, 200 and 500µM HBGA B and Gb4 tetrasaccharide, respectively were mixed and injected into Pd/Pt-coated borosilicate glass capillary. In order to mimic the experimental condition performed during native MS investigation of P dimer-glycan interaction, the instrumental parameters is applied with the following values: 1200V in capillary voltage, 15V in cone voltage, 15V in source offset. The gas flow of the nano-ESI source is set to 0.0mL/min. In addition, the backing pressure is 0.1 bars and the source temperature is 0.1 bars to maintain the soft ionization condition, which native-like protein state could be preserved during the ionization spray process. Each single test is conducted at least twice to address reproducibility. The arrival-time distribution is shown and the corresponding data is acquired at Helium cell DC of 35V and a bias of 55V.

6.7 Glycomacromolecules synthesis

The synthesis of fucose/galactose attached Glycomacromolecules was applied by solid phase polymer synthesis (SPPoS)(Wojcik et al., 2012). The following synthesis steps is written below based on Bücher et al((Bücher et al., 2018))the backbone of the

glycomacromolecules were assembled stepwisely with alkyne-functionalized building block (TDS) and ethylene glycol building block (EDS) (Figure4.2).(Ponader et al., 2012) The glycomacromolecule backbones were functionalized with fucose or galactose ligands on defined positions by using azido-functionalized α-L-fucopyranoside (Fuc-N_3) in a Cu-mediated alkyne-azide cycloaddition (CuAAC) on solid support adapted from the protocol in (Wojcik et al., 2012). The final overall structures (glycomacromolecules A-I, Figure 2) were cut and collect from the resin and then purified by preparative HPLC.

6.8 Glycomacromolecules purification process

The glycomacromolecules have been isolated as crude products after de-protection of the carbohydrate side chains and obtained after cleavage from the solid support with purities of 90-95% (UV detection at 214nm). All glycomacromolecules were then further purified by ion exchange chromatography followed by Agilent 1200 reverse-phase semi-preparative high-performance liquid chromatography at 25°C (preparative RP-HPLC) instrument with Varian Persuit semi-preparative column (C18, 250x10.0 mm). HPLC pure water and acetonitrile (ACN) were used in mobile phase to elute the glycomolecules with a flow rate at 20 mL/min. The purity of glycol-oligomers was assessed by reversed phase - high pressure liquid chromatography - mass spectrometry (RP-HPLC/MS) at 25°C. Agilent 1260 Infinity system is coupled with 6120 Quadrupole LC/MS. the HPLC column is Poroshell 120 EC-C18 (3.0×50 mm, 2.5 μm). The measurement was performed with a linear gradient ladder starting with 100% mobile phase A (H_2O:ACN, 95:5, 0.1%Formic acid) as well as no mobile phase B(H_2O:ACN, 5:95, 0.1%Formic acid) until ending 50% mobile phase A and B in 30 min with a flow rate 0.4 mL/min. The final purity of the obtained molecules are at least higher than 95% (UV detection at 214 nm via RP-HPLC) (Bücher et al., 2018)

6.9 glycomacromolecules binding determination

Native MS were conducted based on the previously described reference protein method(Sun et al., 2006)using 3 μM(dimer) hNoV GII.4 Saga P-dimer and 10 μM reference protein cytochrome c (C7752 Sigma-Aldrich). Vivaspin500 (MWCO 10000, Sartorius) were used to buffer exchange the purified P dimer sample and reference protein Cyt c into 300 mM ammonium acetate and 20 mM triethylammonium acetate (TEAA) buffer solution (90358 Sigma). The freeze-dried glycomacromolecule powder was

dissolved in MilliQ water and the titration measurement was conducted in 100µM; 150 µM and 200 µM. Mass spectra were obtained in positive ion mode on Micromass LCT mass spectrometer (MSvision, the Netherlands) with a nano-electrospray ionization (ESI) source. In-house produced capillaries were used for direct sample infusion. Capillary and sample cone voltages were 1.20 kV to 1.30 kV and 260 V, respectively. Cesium iodide solution (25 mg/ml was collected in the same measurement time period for calibration of raw data using the MassLynx software (Waters, UK). OriginPro 2016 SR2 (OriginLab, United States) software were used to determine the bound ligand and non-bound ligand peak area. The K_D calculation method is followed by the equation illustrated in section 6.4

6.10 HDX MS and experiment design

P dimers (30–50 pmol) were mixed with glycans at tenfold of the final concentration (final: 10 mM HBGA B trisaccharide, 100 mM fucose) and directly diluted 1:9 in 99% deuterated 20 mM Tris buffer (pH 7, 150 mM NaCl, 25 °C) to start the exchange reaction. After various time points (1 min, 10 min, 1 h, 8 h for most datasets, detailed information is given in the SI, HDX summary tables) the exchange reaction was quenched by 1:1 addition of ice-cold quench buffer (300 mM phosphate buffer, pH 2.3, 6 M urea), which decreased the pH to 2.3 and frozen in liquid nitrogen. As a fully deuterated (FD) control, P dimers were diluted 1:9 in 99% deuterated 20 mM Tris buffer with 150 mM NaCl and 6 M urea at pH 7, labelled for 24–72 h at room temperature and quenched as described above.

Samples were thawed, centrifuged to remove aggregates and injected onto a cooled (2 °C) HPLC System (Agilent Infinity 1260, Agilent Technologies, Santa Clara, CA, USA) equipped with a home packed pepsin column (IDEX guard column with an internal volume of 60 µL, Porozyme immobilized pepsin beads, Thermo Scientific, Waltham, MA, USA) in a column oven (25 °C), a peptide trap column (OPTI-TRAP for peptides, Optimize Technologies, Oregon City, OR, USA) and a reversed-phase analytical column (PLRP-S for Biomolecules, Agilent Technologies, Santa Clara, CA, USA). Pepsin digestion was performed online at a flow rate of 75 µL/min (0.4% formic acid in water) and peptides were trapped in the trap column. Peptides were eluted and separated on the analytical column using a 7 min gradient of 8–40% solvent B (solvent A: 0.4% formic acid in water, solvent B: 0.4% formic acid in acetonitrile) at 150 µL/min. After each run the analytical

column was washed with a high percentage of solvent B. MS was performed using an Orbitrap Fusion Tribrid in positive ESI MS only mode (Orbitrap resolution 120,000, 4 micro-scans, Thermo Scientific, Waltham, MA, USA).

All time points were performed in three technical replicates, apart from the 8 h time point of GII.10 Vietnam with fucose, which only represents a single measurement. The triplicate measurement of GII.4 MI001 P dimer was influenced by increased peptide carryover (Figure 2), which overlaid with the lower deuterated peak distribution and led to a falsely high intensity. Therefore, a separate single-replicate measurement with additional pepsin column washing (2 M urea, 2% acetonitrile, 0.4% formic acid, pH 2.5) between sample injections was performed to minimize carryover and only deuteration differences, which are present in both datasets, are considered real.

6.11 Peptide sequence mapping and PTM determination

Identification of peptides and post-translational modifications (PTM) was performed on non-deuterated samples using a 27 min elution gradient of 8–40% solvent B in data-dependent MS/MS acquisition mode (Orbitrap resolution 120000, 1 micro-scan, HCD 30 with dynamic exclusion). Precursor and fragment ions were searched and matched against a local protein database just containing the protein of interest in MaxQuant (version 1.5.7.0) using the Andromeda search engine (Tyanova et al., 2016). The digestion mode was set to “unspecific” and N-terminal acetylation, deamidation, oxidation and disulfide bond formation were included as variable modifications with a maximum number of 5 modifications per peptide. Peptides between 5 and 30 amino acids length were accepted. The MaxQuant default mass tolerances for precursor (4.5 ppm) and fragment (20 ppm) ions defined for the Thermo Orbitrap instrument were used for data search. The minimum score for successful identifications was set to 20 for unmodified and 40 for modified peptides. The elution gradient was chosen in a way that wildtype and deamidated peptides could be clearly separated, as shown in a previous study (Mallagaray et al., 2019). For these peptides, spectra were checked manually and chromatographic peak areas where calculated in Xcalibur (Thermo Scientific, Waltham, MA, USA) to obtain a wildtype/deamidated peptide ratio.

6.12 HDX data analysis

DeutEx software (peterslab.org) was used to determine the deuterium uptake via centroid analysis (Trcka et al., 2019). Excel (Microsoft), GraphPad Prism (GraphPad Software, Inc., San Diego, CA, USA), OriginPro 2016 (Origin Lab Corporation, Northampton, MA, USA) and PyMOL (Schrödinger, New York, NY, USA) software were used for data visualization and statistical analysis. For comparison of triplicate data, a two-sided Student's T-test using deuteration difference from centroid analysis was used with the α-value set to 0.05. A peptide was only considered to have a significant HDX difference if the peptide passed the T-test and ΔD exceeded 2x the pooled average standard deviation (Hageman & Weis, 2019; Houde et al., 2011) of the dataset either for several time points or for the same time point in overlapping peptides. As there were some inconsistencies in the preparation of fullz deuteriated (FD) controls between the datasets, FD controls were not used for normalization and are not shown in the deuterium uptake plots. Only for comparison of the unbound wildtype and deamidated MI001 P dimer, the ratio of the FD controls (which were prepared under identical conditions) from both measurements was used for normalization. Additionally, a higher cut-off of ΔD > 0.42 (99% percentile calculated according to (Arora et al., 2015)) was used to account for possible day-to-day variation in the experimental conditions. Regions with significant deuterium uptake differences were mapped to existing P dimer crystal structures or the homology model (GII.4 MI001).

To compare relative intensities of both distributions in different states, an average over all bimodal time points for both distributions in each state was calculated for several peptides. Because of the unknown degree of carry-over in the GII.17 Kawasaki and GII.10 Vietnam experiments, relative intensities are not compared across different strains, but only between different states in the same experiment, in which the amount of carryover is considered to be unchanged. Averaged relative intensities of the first peak distribution in different peptides are presented as bar plots in Figure 3. The statistical significance of relative intensity differences of the first peak distribution in different states were analyzed with a two-sided Student's T-test for each pair of states (unbound vs. ligand-bound) in an individual experiment ($p < 0.05$). Peptide coverage maps, indicating the effective peptide coverage in each HDX experiment, were plotted with MS Tools (Kavan & Man, 2011)

7. Perspectives

Glycan clustering is recognized as an artefact during direct MS analysis of protein-ligand interactions (Han et al., 2013; Han, Tan, et al., 2014a; Han et al., 2018; J. Sun et al., 2006). Solvent evaporation process during ESI allows glycans to cluster around protein surfaces. Previously, reference proteins were introduced during direct MS analysis to correct for the non-specific clustering and determine the binding affinity of glycans to proteins of interest (Bücher et al., 2018; Sun et al., 2006; Wegener, 2017). However, we found that it is difficult to distinguish between clustering and specific binding at low affinity via native MS.

In chapter 3, we examined several different P dimer species including GII.4 Saga wildtype and deamidated P dimers as well as mutated MI001 P monomer and dimer to analyze the glycan clustering. The obtained data showed that more glycans cluster to deamidated P dimer than wildtype in native MS. In conjunction with HDX-MS data indicating higher structural flexibility for deamidated P dimers, it suggests flexibility promotes ligand clustering. Therefore, we examined the degree of glycan clustering for eight reference proteins with distinct biophysical characteristics. It revealed that glycan clustering on ADH (large size protein) and Cyt c (small size protein) varies significantly (Figure 3.3), however, previous reports suggested that the mass and nature of the reference proteins did not affect the glycan clustering (Sun et al., 2006; Wang et al., 2005). Our experimental results cast doubt on the selection of the reference protein during the direct MS analysis: i) Apart from Ubq, the reference protein candidates demonstrate a proportional behavior when plotting the clustering to the glycan concentration (Figure3.4). ii) Proteins with native-like structure are supposed to be ionized via the charge residue model (CRM) (Consta & Malevanets, 2012). The obtained Ubq data does not obey to the linear mode implying it could follow another ionization model. MD simulations also indicated that small folded proteins can be ionized through the ion evaporation model (IEM) (Aliyari & Konermann, 2020). iii) Cyt c also presents a non-Gaussian peak distribution upon glycan addition (Figure 3.3). iv) Clustering is also vastly different for large and mid-sized reference protein candidates. It seems protein properties play a role upon glycan clustering. This hints at the calculation of binding constants using reference proteins for glycan clustering elimination being inaccurate and insufficient. In order to address this question, a non-binding control Gb4, an all galactose glycan, was applied to compare with the binding of blood group antigens. Galactose was also employed as negative control for glycomacromolecule binding in chapter 4 (Bücher et al., 2018). When comparing to the stoichiometry information from Figure S10 and Table S2, HBGA B presents a

weak single binding event at 500 µM concentration while non-binder Gb4 shows no specific binding at all listed concentration after the ADH based protein clustering correction. Furthermore, the determined K_{D1} (24 mM) based on native MS analysis is consistent to the binding affinity value (K_D: HBGA B-tetrasaccharide type 1, 12 mM) previously acquired from NMR (Creutznacher et al., 2021; Mallagaray et al., 2019). Thereby, it indicates that an appropriate reference candidate is key for K_D calculation during direct MS measurement.

For K_D determination, our result presented a two orders of magnitude difference compared to the value acquired in previous reports (Han et al., 2013; Han et al., 2018; Mallagaray et al., 2019). This raises another question whether some protein intrinsic parameters are involved in glycan clustering. After the examination of several biophysical properties, we revealed that the share of β-sheets presents strongest correlation to clustering ratio (Figure 3.5A). In addition, the characteristics of SASA and molecular weight also show comparatively good correlation (Figure S1 and S2). In the crystallographic observation, parallel pleated β-sheets often form the core of the globular protein, NH and CO groups of the peptide bond engage in linear hydrogen bonding networks between individual β-strands. In contrast, in α-helices, the hydrogen bonds twist around the longitudinal axis of the helix. Glycan structure has multiple hydroxyl moieties, which could engage in hydrogen bonding during protein-ligand clustering in β-sheets. IMMS was then applied to investigate if binding of HBGA B (binder) and Gb4 (non-binder) to the P dimer alters structure significantly as suggested in earlier study(Han et al., 2018) and to examine whether intercalation of glycans into the protein structure is detectable. It demonstrated that these two glycans behave similarly. The arrival-time distributions present no major changes after the addition of the two glycans, respectively. This result is also in accordance with the observation from NMR and HDX MS(Mallagaray et al., 2019). Previously, others reported occurrence of conformational changes at 17+ and 18+ charge states of P dimer upon 800µM HBGA B binding(Han et al., 2018). However, we do not observe a similar arrival-time distribution shift using the same P dimer species and HBGA B glycan. Possible explanation could be the instrument employed in the experiment is a modified device, which might lead to over-activation and unfolding, but the single near-Gaussian peak of the P dimer in IMMS can exclude the artefact of over-activation. Taken together, direct MS analysis via reference proteins for clustering elimination is now considered unsuitable in glycan binding studies, especially for low affinity ligands. Although the original approach is established and the calculated K_D values are verified through other techniques, this method is merely applicable in higher-affinity binding studies. Additional proof comes from mutated MI001 P domains and MNV P domains that form

monomers and dimers (Figure 3.2, Yan et al., 2021). P domain monomers lack an intact glycan binding groove and are hence binding incompetent. It was also found that HBGA does not bind to MNV (R. Creutznacher et al., 2021). Furthermore, deamidation apparently affected clustering (Mallagaray et al., 2019; Yan et al., 2021), which warrants further investigation.

In chapter 4, The K_D value of HBGA B tetrasaccharide type 1 was used to compare with glycomacromolecules. It was previously demonstrated that HBGA B has a slightly high affinity (near 100µM) (Mallagaray et al., 2015). This K_D value is obtained after elimination of the non-specific clustering based on Cyt c (Bücher et al., 2018). As discussed above, Cyt c is a small, α-helical protein that is now concerned inappropriate for data correction according to chapter 3.

Nonetheless, Cyt c was applied as reference protein in the glycomacromolecule binding study. In order to decipher whether glycomacromolecule structures achieve avidity through multiple binding sites on the P dimer, nine glycomacromolecules are examined. The binding strengths of galactose carrying glycomacromolecule 8 (negative control) is ten-fold lower than fucose containing mono- or divalent structures (Table 4.1), which were within error indifferent. It revealed that the designed macromolecules can bind to P dimers but do not achieve multivalent binding. The calculated binding affinities of both monovalent and divalent structures are higher than the recent K_D value (12mM, NMR) determined for fucose alone (Creutznacher et al., 2021). However, it shows that the native MS approach can be used if i) a negative control is employed and ii) the ligand is largely not a glycan. The glycans on the glycomacromolecules are presented on a peptide-like backbone suggesting that e.g. peptide binding studies are still amenable. iii) The designed structure with a comparatively large size and mass could also weaken clustering itself owing by steric effect. In order to reach multivalent binding, further modification of the ligand structure could be an option, for example, exploring the variations of the macromolecular scaffold. Interestingly, the binding of bile acid is recently identified to trigger MNV P dimer conformation change (Creutznacher et al., 2021; Williams et al., 2021). An alternative plan could be scheduled for the design and synthesis of bile presenting mimetics for MNV P dimer binding. If this designed molecule could initiate similar allosteric structural change, it would provide useful information on the scaffold for the next generation of fucose attached glycomacromolecules. On the other hand, the glycomacromolecule binding study also emphasized the significance of the non-binder control and careful assessment of reference protein candidate during the direct native MS measurement. . For studying of next generation of glycomacromolecules, it is valuable to re-examine whether the nature of the volatile salt or

the ionic strength affects the results. It is advised to use the non-binding control for correction and KD determination instead of a reference protein system.

In chapter 5, other norovirus strains including GII.4 MI001, GII.10 Vietnam and GII.17 Kawasaki P dimers were investigated via HDX MS to explore differences in structural effects of glycan binding. A similar protection manner during H/D exchange was detected on GII.4 MI001 strain (compared to previously identified Saga P dimer in (Mallagaray et al., 2019)) but not on GII.10 Vietnam and GII.17 Kawasaki P dimers. It has been reported that GII.4 dominated the outbreak cases worldwide in last two decades (Eden et al., 2013). The deamidation that occurs in GII.4 strains could be an evolutionary benefit during the host-virus interaction. In contrast, GII.17 Kawasaki strain infection cases were gradually increasing in a recent epidemiological survey (Chan et al., 2017). However, no deamidation is observed in Kawasaki P dimers and the sequence analysis showed that the glycan binding region on the P2 cleft does not possess an asparagine (Ahmed et al., 2014; de Graaf et al., 2015). Combination of the findings on GII.4 and GII.17 strains, it implies that the presence or absence of deamidation can both be considered as an infection strategy during the virus-host interaction. In addition, we observed stronger protection during glycan binding in GII.10 Vietnam P dimers using HDX MS (Figure 5.2). The distinctive protection pattern could indicate an unknown mechanism during infection. Therefore, the identified deamidation process could be one of many strategies that the virus adopted during host interaction.

We investigated glycan binding pattern differences on the P dimers of three norovirus strains, but little is known on how the virus particles survive from acidic environment at stomach and enter into the intestinal tract and the role of glycans during this processes. Pogan et al. (2018) examined norovirus capsid stability in different pH conditions using native MS and Dülfer et al. (2021) investigated the glycan binding on P dimer of the GII.4 MI001, GII.10 Vietnam and GII.17 Kawasaki strains via HDX MS. Further work can be moved forward to investigate the VLP structural dynamics upon glycan binding via HDX MS at different physicochemical conditions. In contrast, the role of deamidation on norovirus capsid is another issue needs to be addressed.

Thus, binding affinity determination is still a hot topic in the study of norovirus-glycan interaction. Other techniques likely MST or SPR would be an option. However, we should be cautious that the fluorophore labelling on the protein during MST test or covalent coupling via CM5 chip might not preserve the native state of the protein structure. The buffer selection and

the amount of added DMSO in the sample solution could be another potential issue. Nevertheless, these two techniques could at least assist to screen potential strong binders from a new generation of glycomacromolecules with different spacing or ligand types and numbers.

Supplement

S1 Material

S1.1Instruments

Instrument	Manufacturer
pH-meter 913	Metrohm (Herisau, Switzerland)
Spectrophotometer DS-11 FX	DeNovix (Wilmington, USA)
Sputter coater Q150R	Quorum Technologies (Lewes, UK)
Centrifuge 5417R	Eppendorf (Hamburg, Germany)
Centrifuge 5430R	Eppendorf (Hamburg, Germany)
Mass spectrometer LCT	Waters Corporation (Milford, USA) MS Vision (Almere, The Netherlands)
Mass spectrometer QToF2	Waters Corporation (Milford, USA) MS Vision (Almere, The Netherlands)

S1.2Chemicals

Chemical	CAS number	Distributor
Acetic acid	64-19-7	Th. Geyer (Renningen, Germany)
Acetonitrile	61-12-3	Th. Geyer (Renningen, Germany)
Ammonium acetate	631-61-8	Sigma-Aldrich (St. Louis, USA)
Argon	7440-37-1	SOL Germany (Gersthofen, Germany)
Cesium Iodide	7789-17-5	Sigma-Aldrich (St. Louis, USA)
Formic acid	64-18-6	Sigma-Aldrich (St. Louis, USA)
Triethylammonium acetate	5204-74-0	Sigma-Aldrich (St. Louis, USA)

S1.3Commercial protein list

Chemical	CAS number	Distributor
Alcohol Dehydrogenase (ADH) from *Saccharomyces cerevisiae*	9031-72-5	Sigma-Aldrich (St. Louis, USA)
apo-Transferrin human	11096-37-0	Sigma-Aldrich (St. Louis, USA)

Carbonic anhydrase II from human erythroCyt es	9001-03-0	Sigma-Aldrich (St. Louis, USA)
Concanavalin A from *Canavalia ensiformis*	11028-71-0	Sigma-Aldrich (St. Louis, USA)
Cyt ochrome c from(Bovine)		Sigma-Aldrich (St. Louis, USA)
Cyt ochrome c from equine heart	9007-43-6	Sigma-Aldrich (St. Louis, USA)
L-Glutamic Dehydrogenase (GDH) from bovine liver	9029-12-3	Sigma-Aldrich (St. Louis, USA)
L-Lactic dehydrogenase from rabbit muscle	9001-60-9	Sigma-Aldrich (St. Louis, USA)
Myoglobin from equine heart	100684-32-0	Sigma-Aldrich (St. Louis, USA)
Pyruvate Kinase from rabbit muscle	9001-59-6	Sigma-Aldrich (St. Louis, USA)
Ubiquitin(bovine erythroCyt es)	75986-22-4	Sigma-Aldrich (St. Louis, USA)

S1.4Consumables

Consumable	**Distributor**
Centrifugal filter units Vivaspin 500 (10,000; 30,000; MWCO)	Sartorius (Göttingen, Germany)
Desalting columns Micro Bio-Spin™ 6	Bio-Rad Laboratories, Inc. (Hercules, USA)
Desalting columns Zeba™, (7K and 40K MWCO)	Thermo Fisher Scientific (Waltham, USA)
Glass capillaries, 100 mm, 1.2 mm outer diameter, 0.68 mm inner diameter	World Precision Instruments (Sarasota, USA)
Gold Target (57mm Ø x 0.5mm)	LOT-QuantumDesign (Darmstadt, Germany)

S2 Supplementary Figures

S1 The influence of protein solvent accessible area to glycan clustering

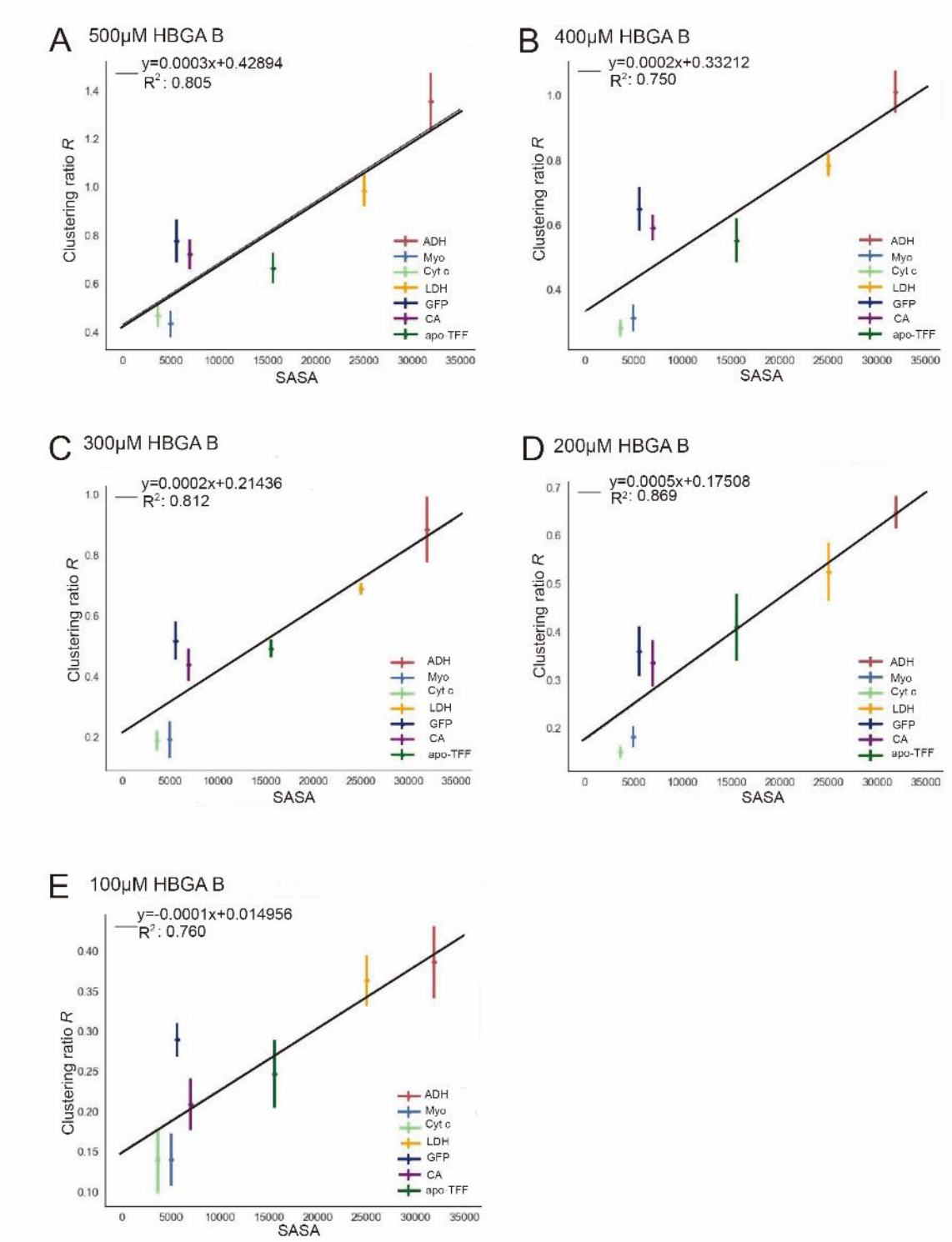


Figure S1 The relationship of solvent accessible area (SASA) to glycan clustering ratio. The clustering ratio *R* value is calculated from native MS spectra during titration measurement of P

dimer and seven listed reference proteins(ADH; CA; Cyt c; GFP; Myo; LDH; apo-TFF), respectively, with HBGA B ligand at five concentrations ranging from 100μM to 500μM in 150 mM ammonium acetate solution in pH7. The linear-fit results at different concentrations are listed as: A) 500μM; B) 400μM; C) 300μM; D) 200μM; E) 100μM. The black line indicates the linear regression result and the corresponding linear fit equation is shown.

S2 the influence of protein mass to glycan clustering

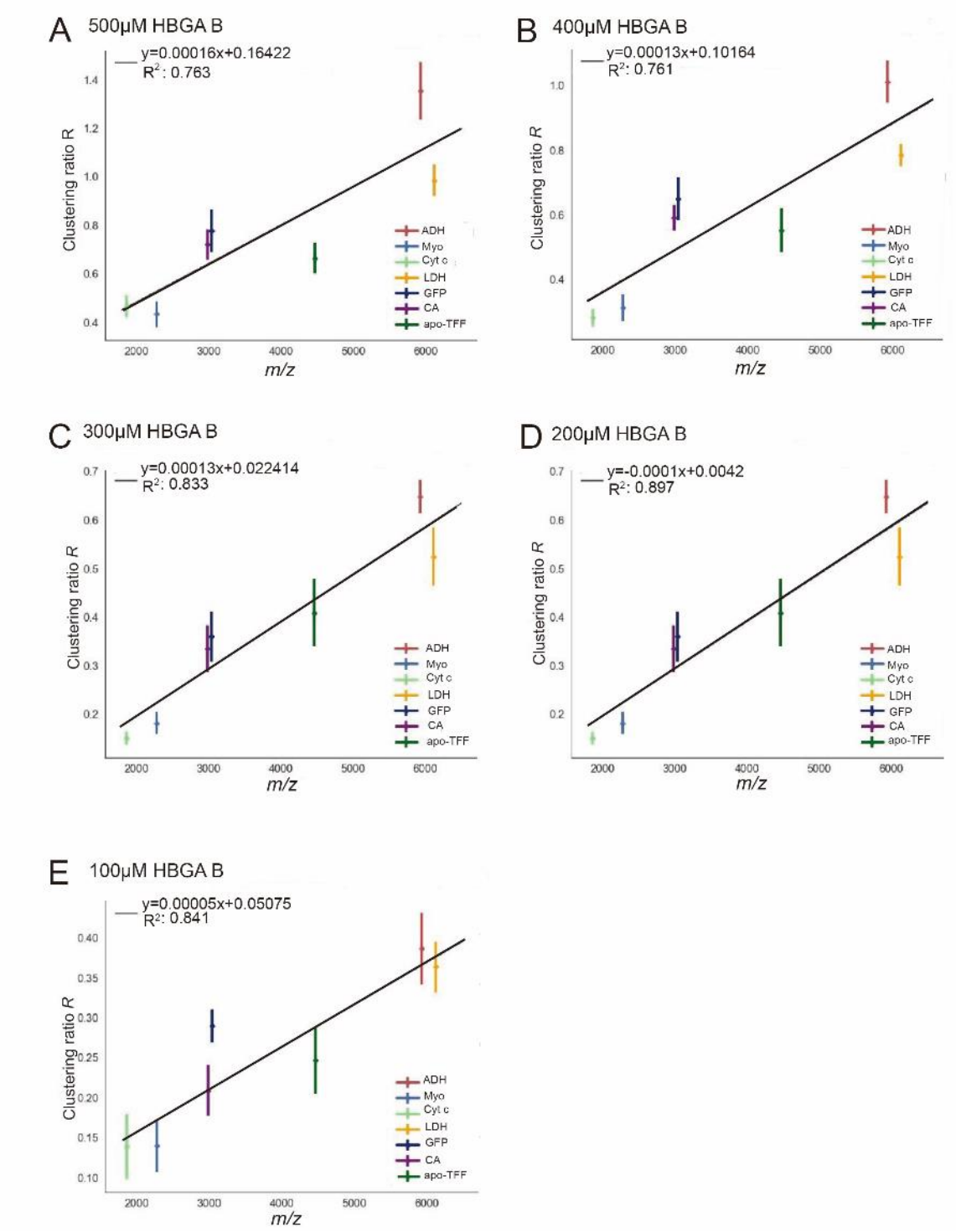


Figure S2 The relationship of protein mass to glycan clustering ratio. The clustering ratio *R* value is calculated from native MS spectra during titration measurement of P dimer and seven listed reference proteins(ADH; CA; Cyt c; GFP; Myo; LDH; apo-TFF), respectively, with HBGA B ligand at five concentrations ranging from 100μM to 500μM in 150 mM ammonium acetate solution in pH7. The linear-fit results at different concentrations are listed as: A) 500μM; B) 400μM; C) 300μM; D) 200μM; E) 100μM. The black line indicates the linear regression result and the corresponding linear fit equation is shown.

S3 the correlation of mass to charge ratio of reference protein to glycan clustering ratio

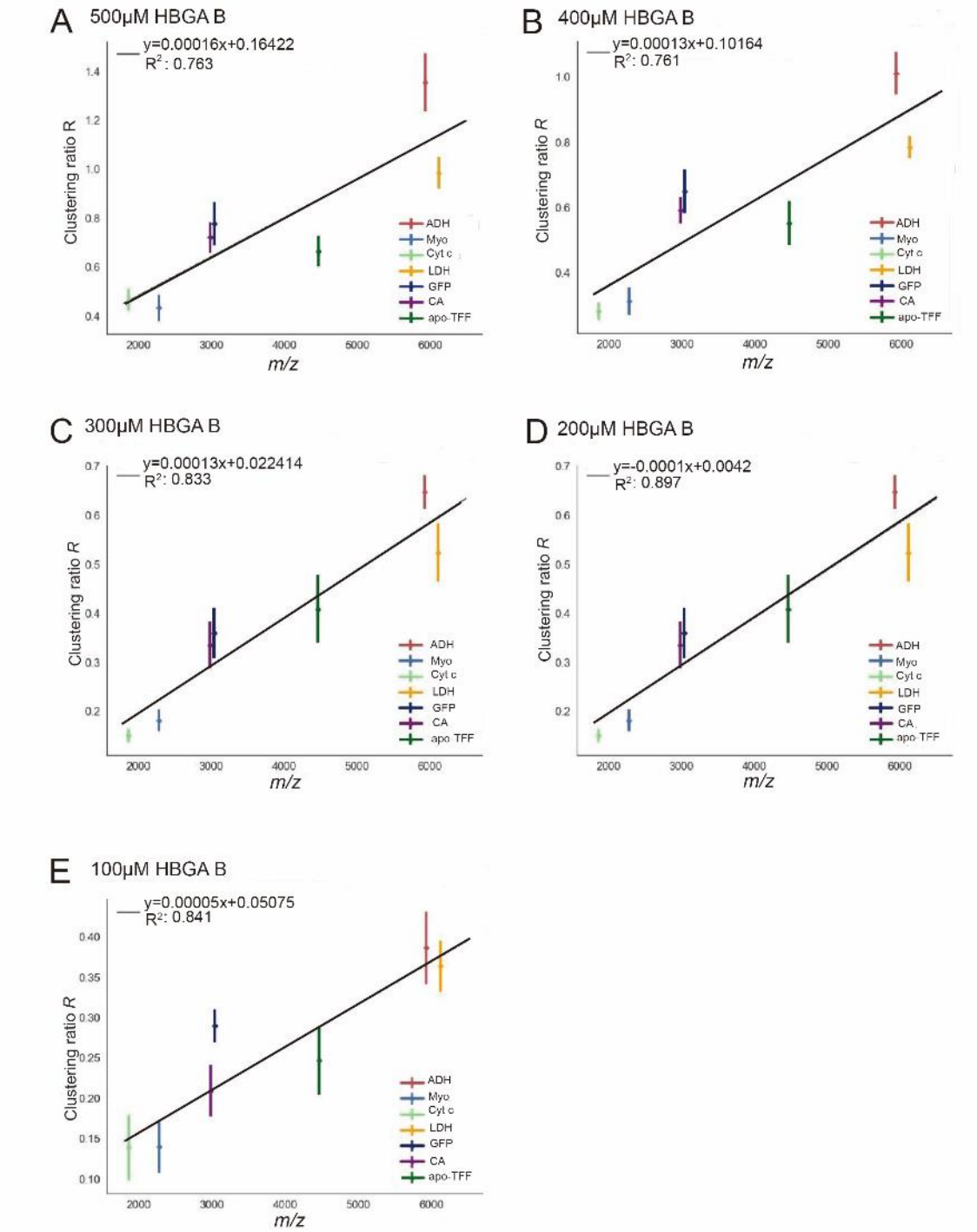


Figure S3 The relationship of mass to charge ratio (base peak) of reference protein to glycan clustering ratio. The clustering ratio *R* value is calculated from native MS spectra during titration measurement of P dimer and seven listed reference proteins(ADH; CA; Cyt c; GFP; Myo; LDH; apo-TFF), respectively, with HBGA B ligand at five concentrations ranging from 100μM to 500μM in 150 mM ammonium acetate solution in pH7. The linear-fit results at different concentrations are listed as: A) 500μM; B) 400μM; C) 300μM; D) 200μM; E) 100μM. The black line indicates the linear regression result and the corresponding linear fit equation is shown.

S4 the correlation of charged residues to glycan clustering ratio

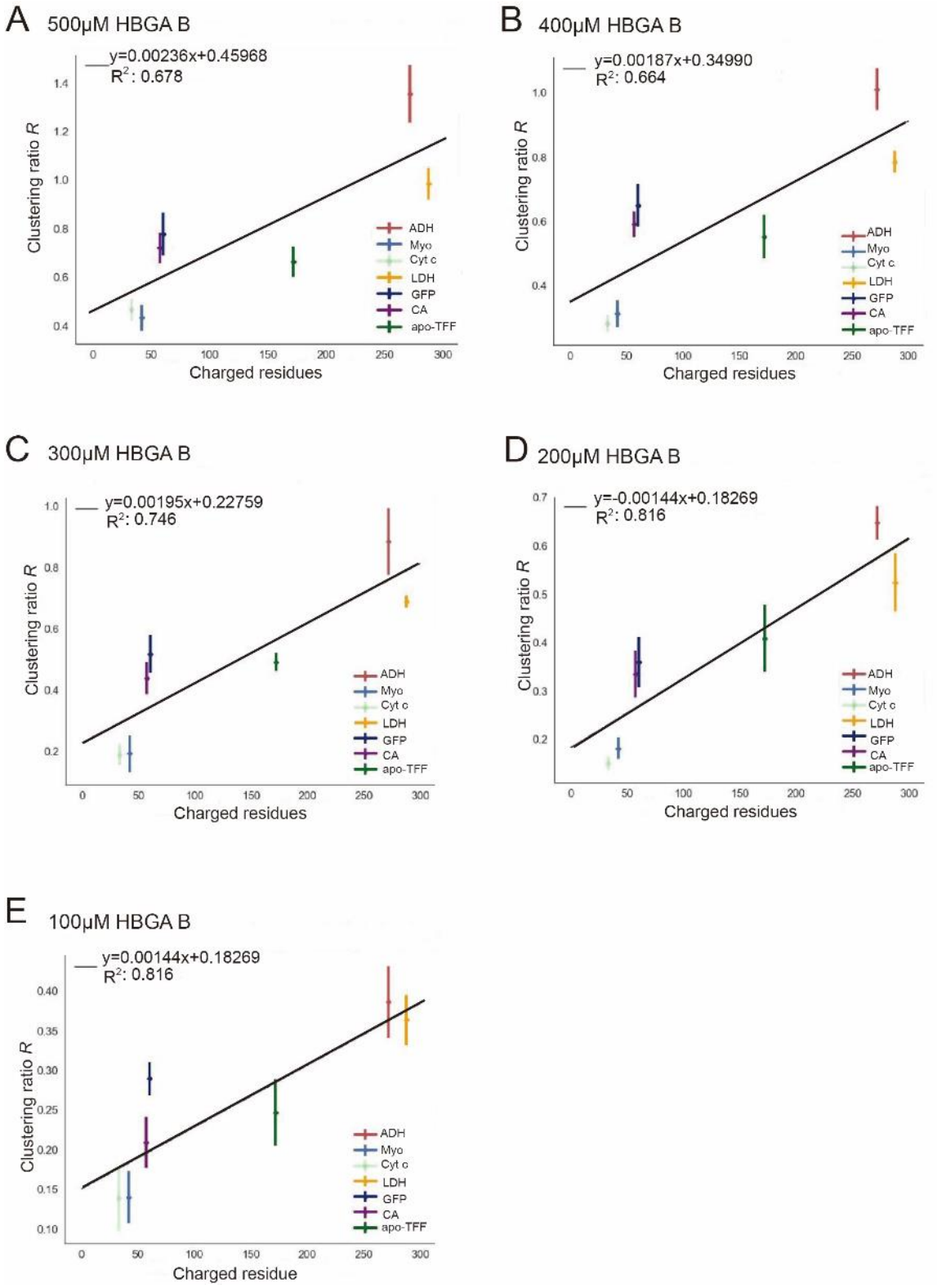


Figure 4 The relationship of charged residues of reference protein to glycan clustering ratio. The clustering ratio *R* value is calculated from native MS spectra during titration measurement of P dimer and seven listed reference proteins(ADH; CA; Cyt c; GFP; Myo; LDH; apo-TFF), respectively, with HBGA B ligand at five concentrations ranging from 100μM to 500μM in 150 mM ammonium acetate solution in pH7. The linear-fit results at different concentrations are

listed as: A) 500μM; B) 400μM; C) 300μM; D) 200μM; E) 100μM. The black line indicates the linear regression result and the corresponding linear fit equation is shown.

S5 the correlation of positively charged residues to glycan clustering ratio

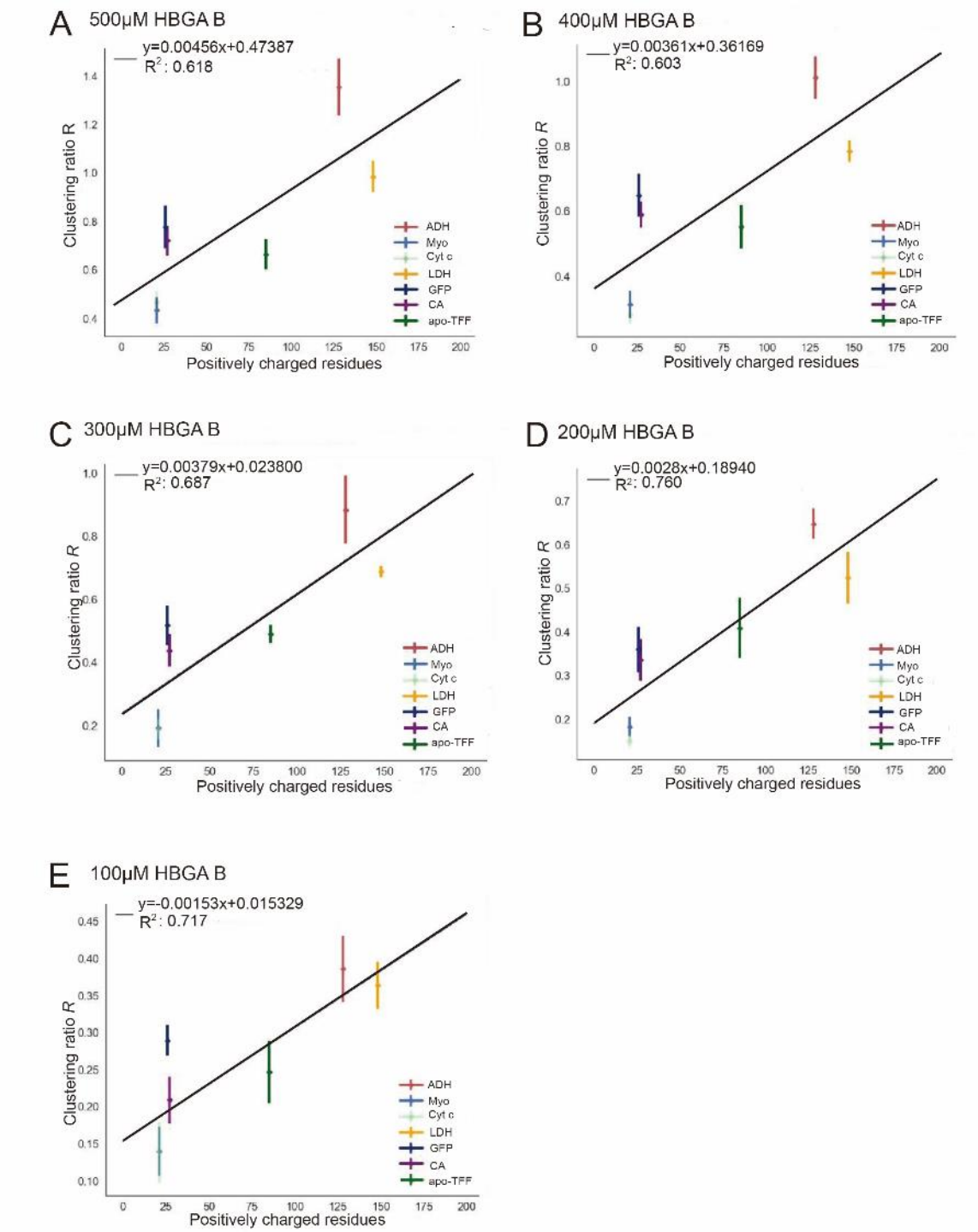


Figure S5 The relationship of positively charged residues of reference protein to glycan clustering ratio. The clustering ratio *R* value is calculated from native MS spectra during titration measurement of P dimer and seven listed reference proteins(ADH; CA; Cyt c; GFP; Myo; LDH; apo-TFF), respectively, with HBGA B ligand at five concentrations ranging from 100μM to

500μM in 150 mM ammonium acetate solution in pH7. The linear-fit results at different concentrations are listed as: A) 500μM; B) 400μM; C) 300μM; D) 200μM; E) 100μM. The black line indicates the linear regression result and the corresponding linear fit equation is shown.

S6 the correlation of negatively charged residues to glycan clustering ratio

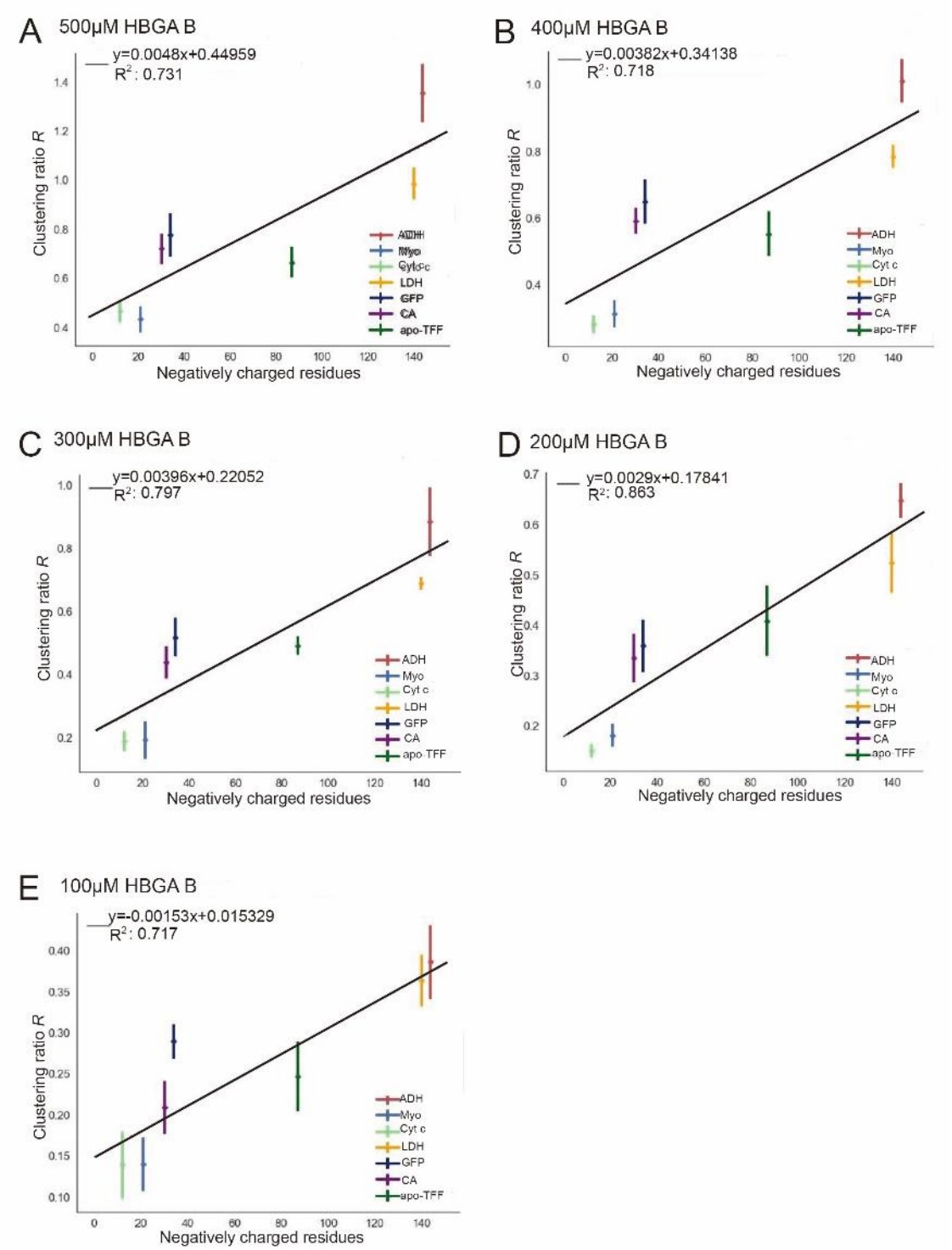


Figure S6 The relationship of negatively charged residues of reference protein to glycan clustering ratio. The clustering ratio *R* value is calculated from native MS spectra during titration measurement of P dimer and seven listed reference proteins(ADH; CA; Cyt c; GFP; Myo; LDH; apo-TFF), respectively, with HBGA B ligand at five concentrations ranging from 100μM to

500μM in 150 mM ammonium acetate solution in pH7. The linear-fit results at different concentrations are listed as: A) 500μM; B) 400μM; C) 300μM; D) 200μM; E) 100μM. The black line indicates the linear regression result and the corresponding linear fit equation is shown.

S7 Correlating the share of α-helix in a protein to glycan clustering ratio

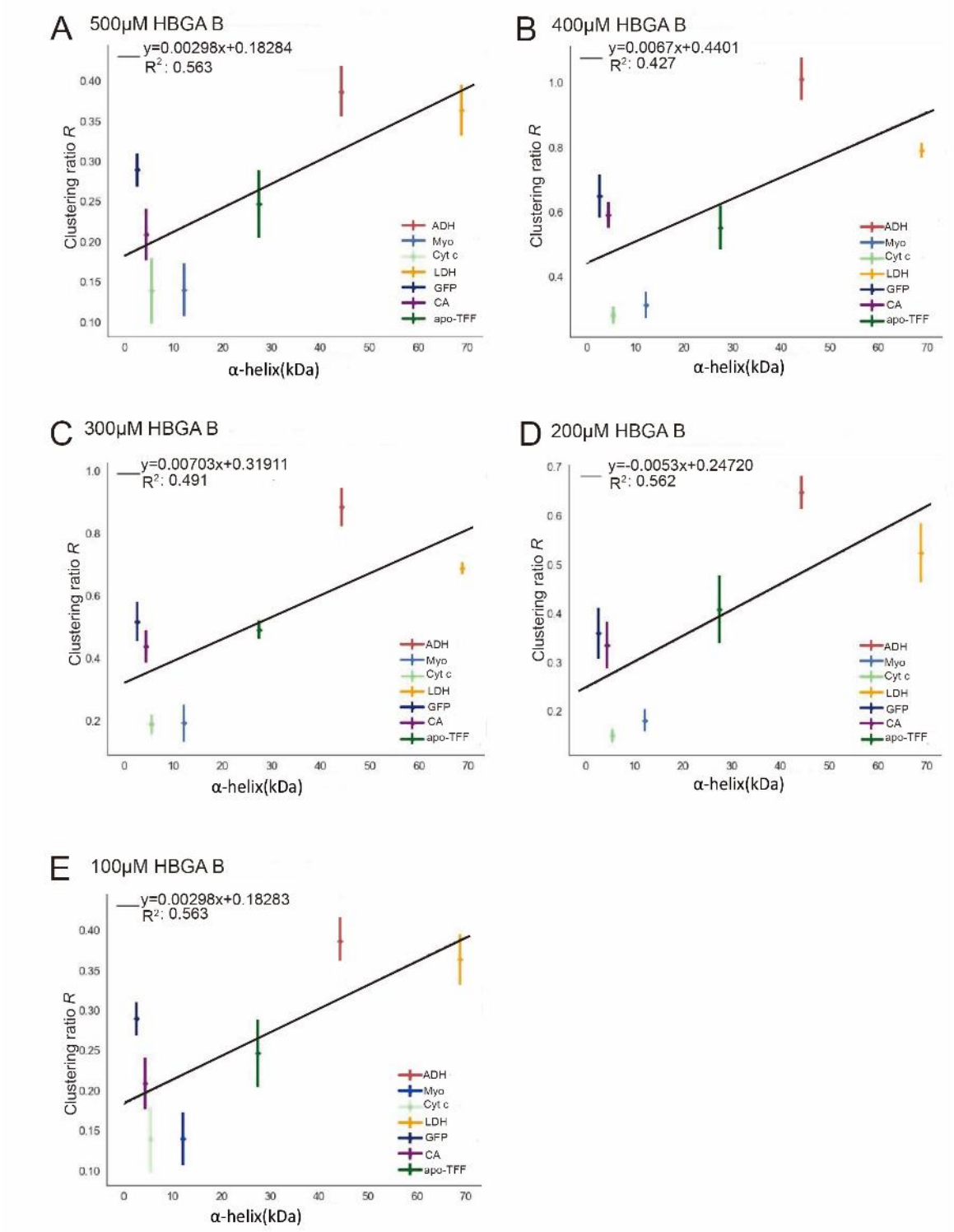


Figure S7 The relationship of the share of α-helix in a protein (kDa) to glycan clustering ratio. The clustering ratio *R* value is calculated from native MS spectra during titration measurement of P dimer and seven listed reference proteins(ADH; CA; Cyt c; GFP; Myo; LDH; apo-TFF), respectively, with HBGA B ligand at five concentrations ranging from 100μM to 500μM in 150

mM ammonium acetate solution in pH7. The linear-fit results at different concentrations are listed as: A) 500µM; B) 400µM; C) 300µM; D) 200µM; E) 100µM. The black line indicates the linear regression result and the corresponding linear fit equation is shown.

S8 Correlating the share of β-sheet in a protein to glycan clustering ratio

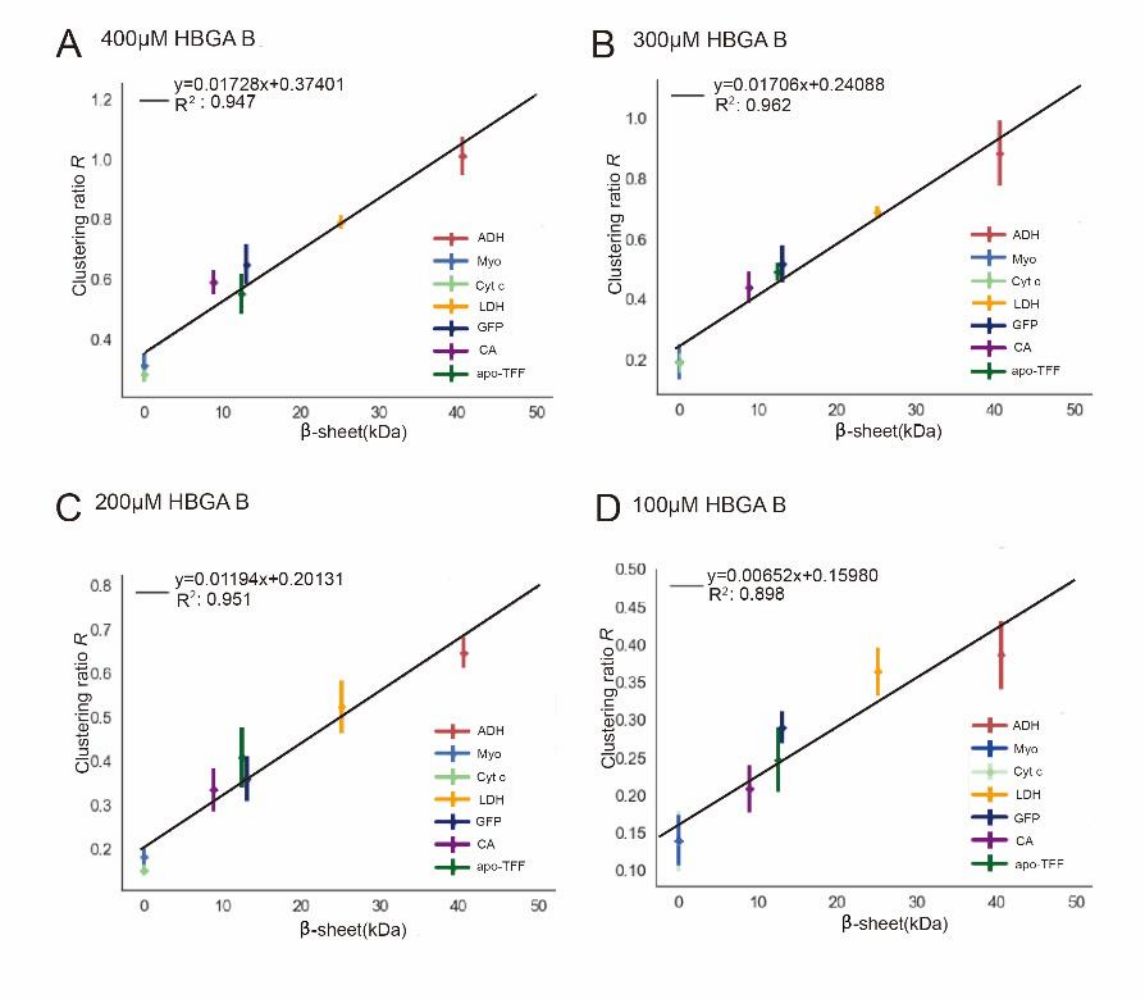


Figure S8 The relationship of the share of β-sheet in a protein (kDa) to glycan clustering ratio. The clustering ratio *R* value is calculated from native MS spectra during titration measurement of P dimer and seven listed reference proteins(ADH; CA; Cyt c; GFP; Myo; LDH; apo-TFF), respectively, with HBGA B ligand at five concentrations ranging from 100µM to 400µM in 150 mM ammonium acetate solution in pH7. The linear-fit results at different concentrations are listed as: A) 400µM; B) 300µM; C) 200µM; D) 100µM. The plot of 500µM concentration is shown in Figure3.5. The black line indicates the linear regression result and the corresponding linear fit equation is shown.

S9 Correlating collision cross section to glycan clustering ratio

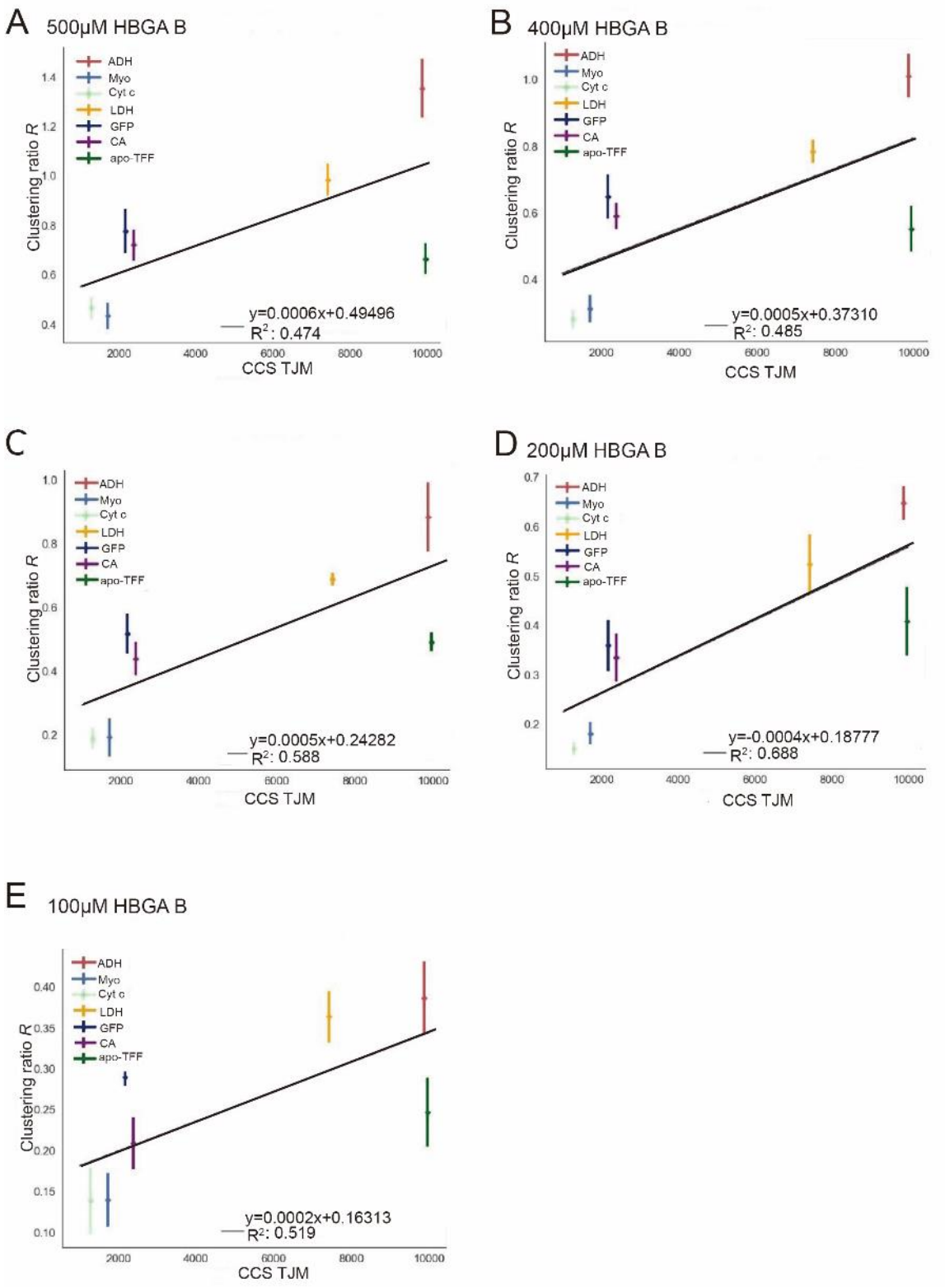


Figure S9 The relationship of collision cross section (CCS TJM) to glycan clustering ratio. The clustering ratio *R* value is calculated from native MS spectra during titration measurement of P dimer and seven listed reference proteins(ADH; CA; Cyt c; GFP; Myo; LDH; apo-TFF), respectively, with HBGA B ligand at five concentrations ranging from 100μM to 500μM in 150 mM ammonium acetate solution in pH7. The linear-fit results at different concentrations are listed as: A) 500μM;

B) 400μM; C) 300μM; D) 200μM; E) 100μM. The black line indicates the linear regression result and the corresponding linear fit equation is shown.

S10 Correlating isoelectric point to glycan clustering ratio

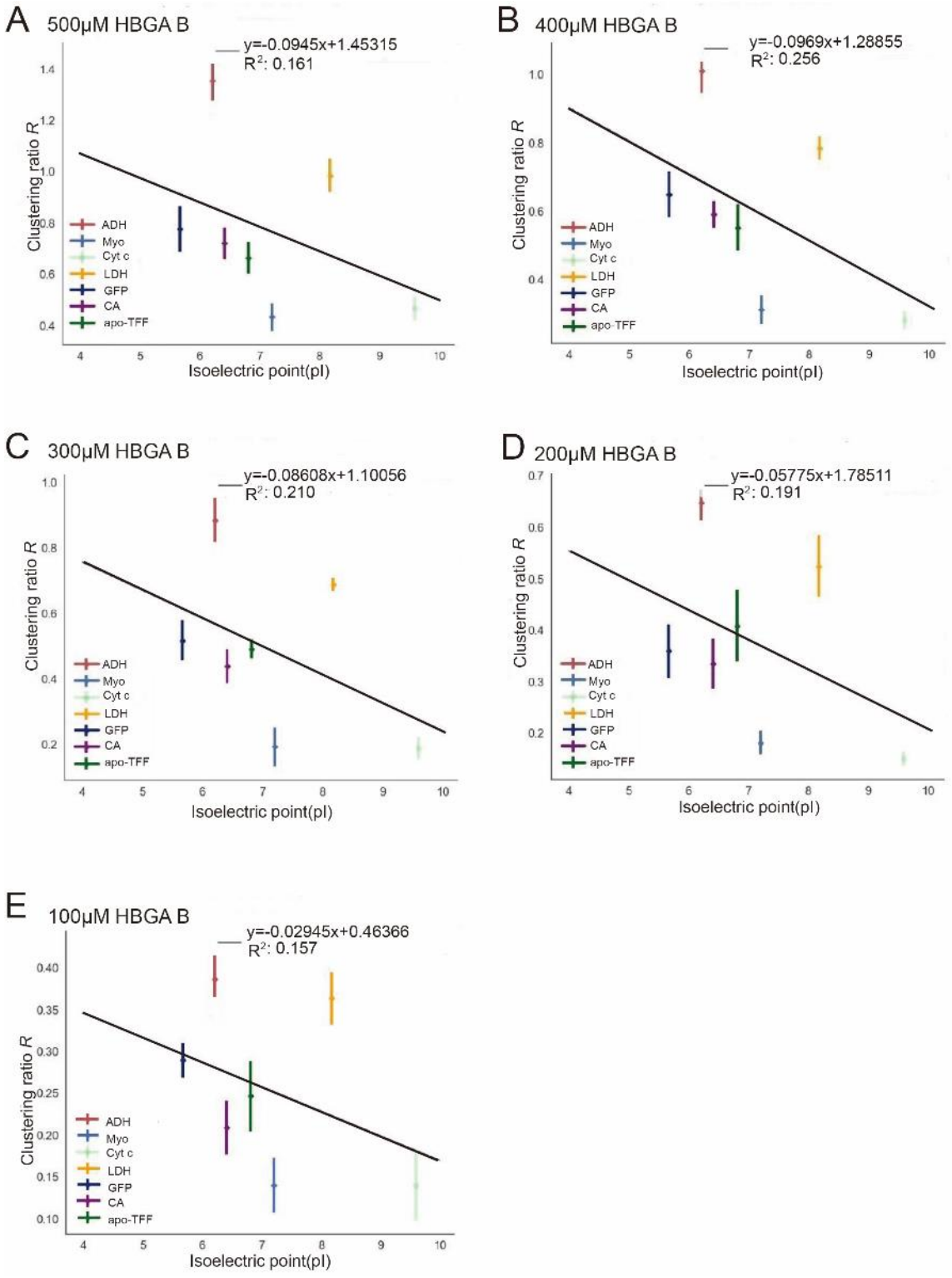


Figure S10 The relationship of isoelectric point to glycan clustering ratio. The clustering ratio *R* value is calculated from native MS spectra during titration measurement of P dimer and seven listed reference proteins(ADH; CA; Cyt c; GFP; Myo; LDH; apo-TFF), respectively, with HBGA B ligand at five concentrations ranging from 100μM to 500μM in 150 mM ammonium acetate solution in pH7. The linear-fit results at different concentrations are listed as: A) 500μM; B) 400μM; C) 300μM; D) 200μM; E) 100μM. The black line indicates the linear regression result and the corresponding linear fit equation is shown.

S11. **Bound glycan number on** Saga **P dimer**

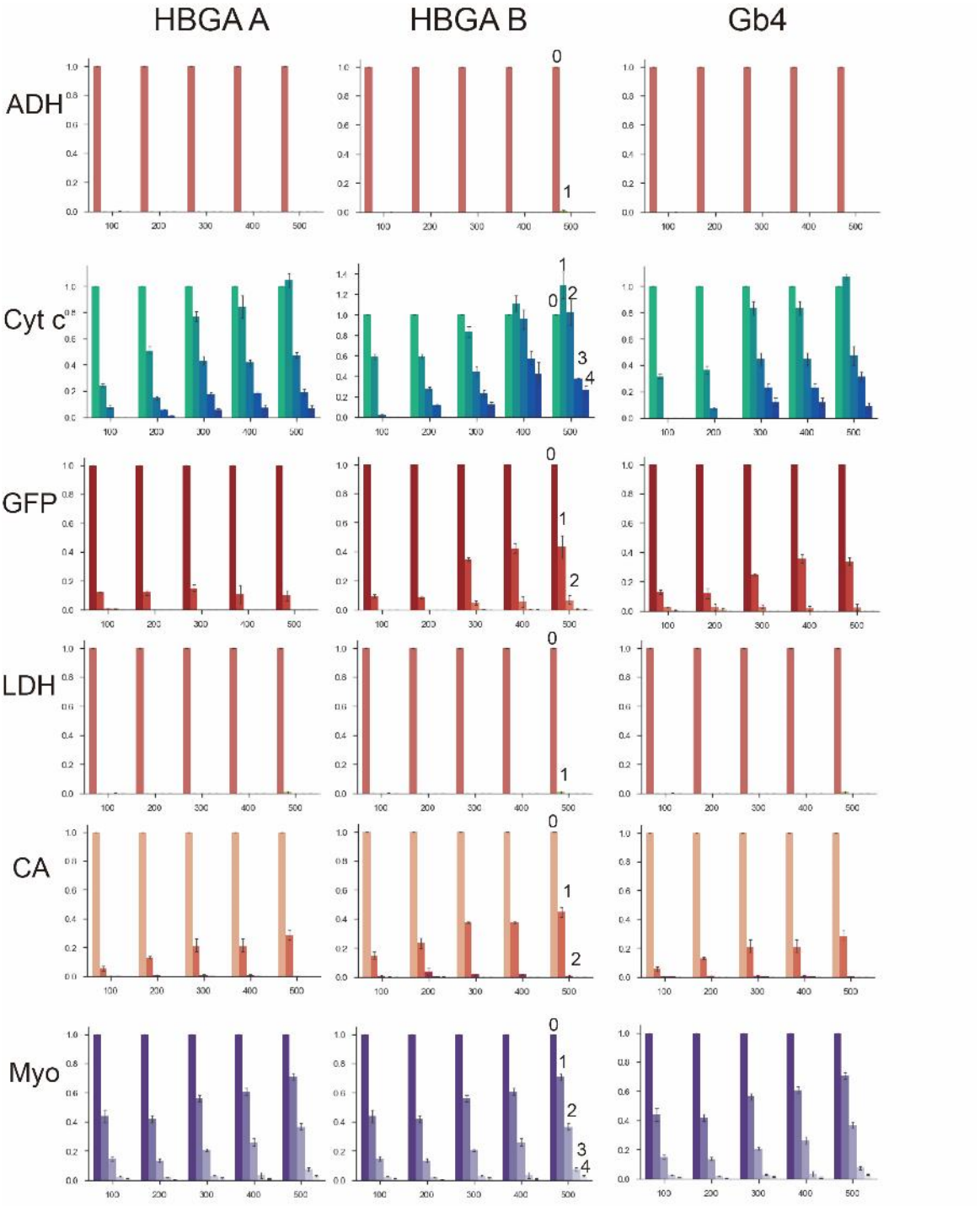


Figure S11 Bound glycan number on P dimer. Titration experiment of three ligands HBGA A, B and Gb4, respectively binding to P dimer is performed ranging from 100μM to 500μM.the bound ligand number(s) is obtained based on native mass spectrum of wildtype Saga P dimer (1 μM) and reference proteins (3 μM), respectively at 150 mM ammonium acetate solution at pH 7. The number 0 to 4 represents the bound number to the P dimer.

S12. Synthetic scheme of precision glycomacromolecules

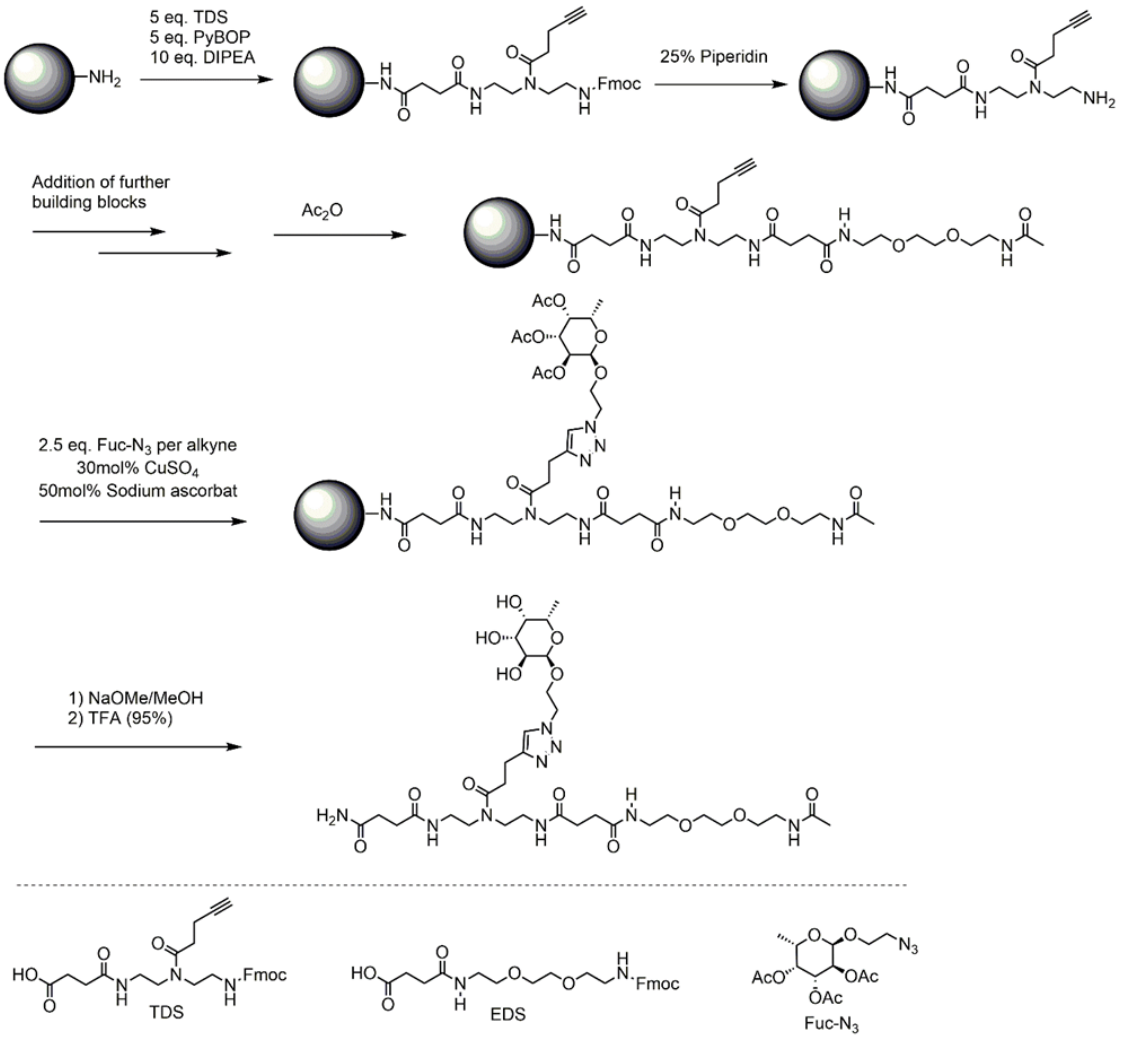


Figure S12. Synthetic scheme for the preparation of precision glycomacromolecules (fucose attached) via solid phase assembly of tailor-made building blocks

S13. native mass spectra of glycomacromolecule 1 binding to GII.4 Saga P dimer

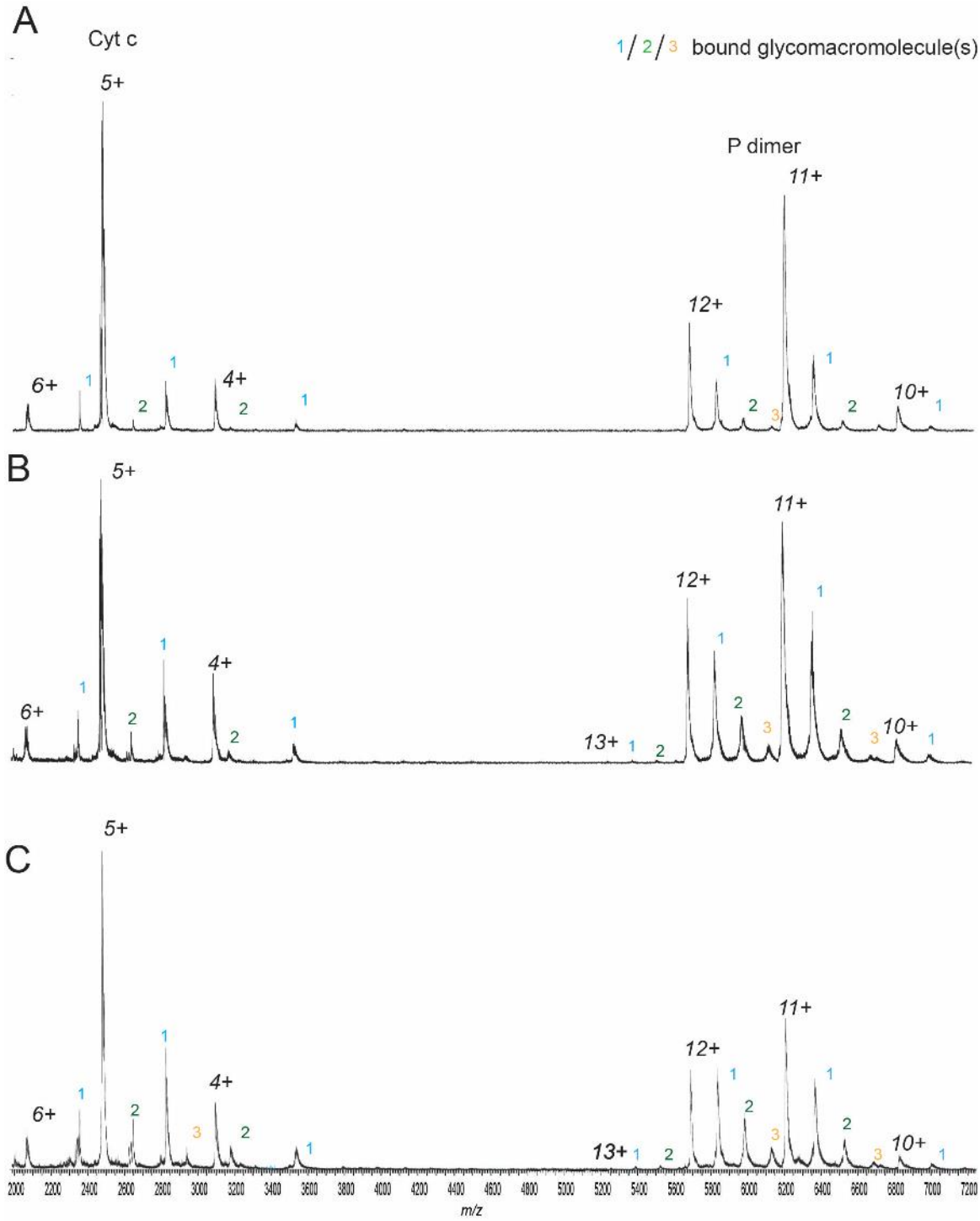


Figure S13 Glycomacromolecule 1 binding to GII.4 Saga P dimer. Mass spectra of GII.4 Saga P dimer and reference protein Cyt c with Glycomacromolecule 1 in three concentrations ranging from 100µM to 200µM at 300 mM ammonium acetate and 20 mM TEAA buffer solution in pH 7. The titration experiments performed in the concentration listed as: A: 100µM(top), B: 150µM(middle) and C: 200µM(bottom).

S14 native mass spectra of glycomacromolecule 2 binding to GII.4 Saga P dimer

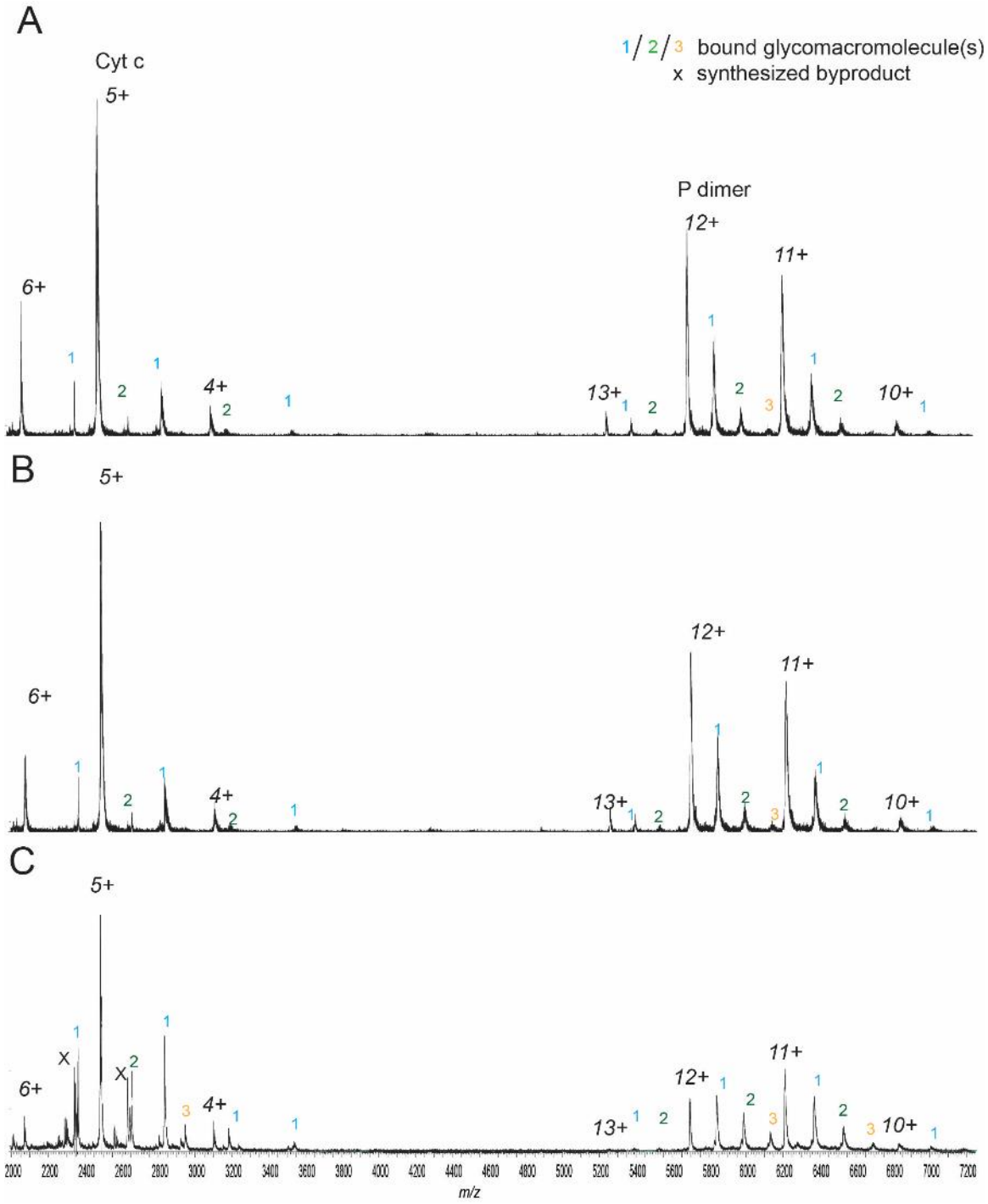


Figure S14 Glycomacromolecule 2 binding to GII.4 Saga P dimer. Mass spectra of GII.4 Saga P dimer and reference protein Cyt c with Glycomacromolecule 2 in three concentrations ranging from 100µM to 200µM at 300 mM ammonium acetate and 20 mM TEAA buffer solution in pH 7. The titration experiments performed in the concentration listed as: A: 100µM(top), B: 150µM(middle) and C: 200µM(bottom).

S15 native mass spectra of glycomacromolecule 3 binding to GII.4 Saga P dimer

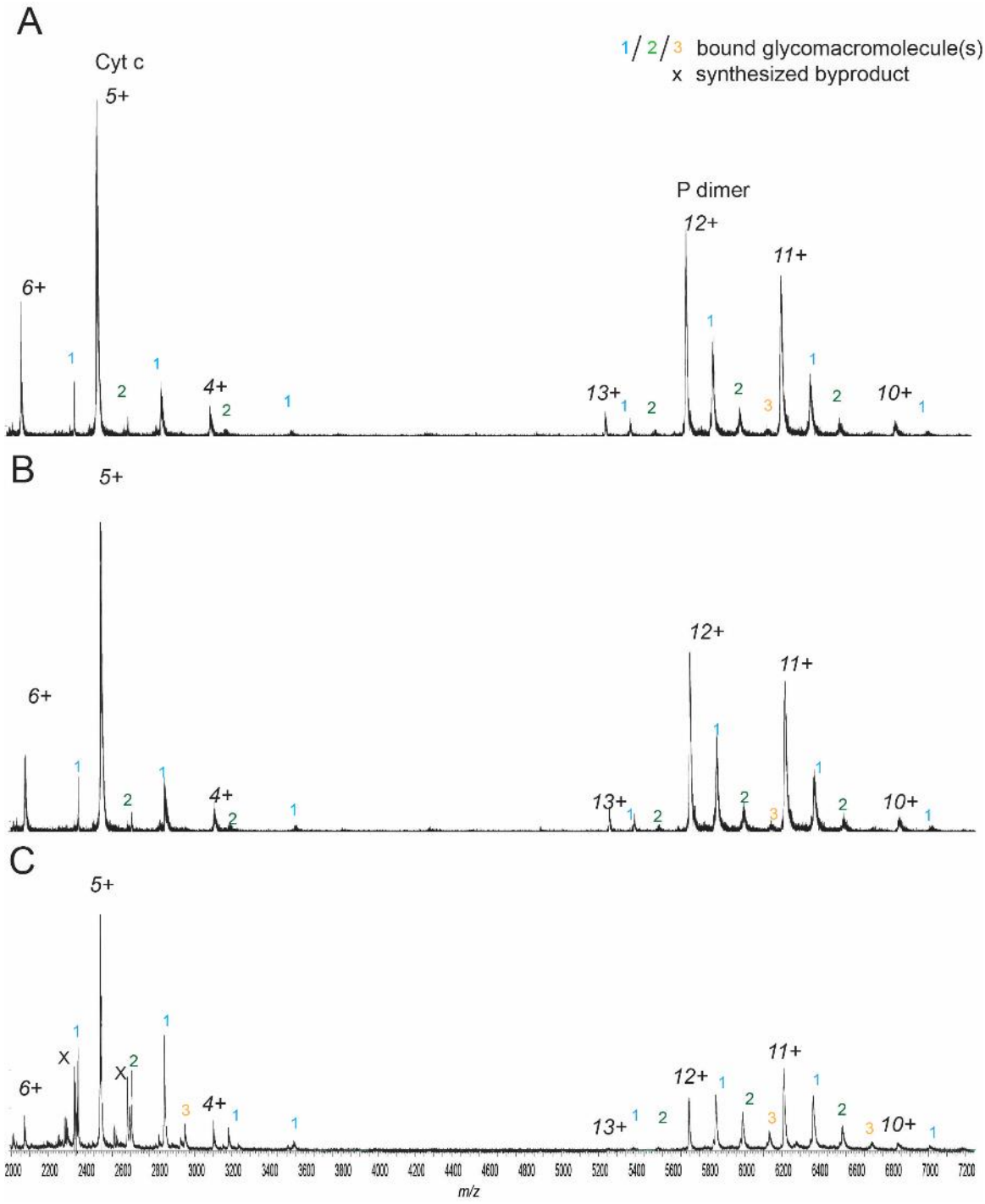


Figure S15 Glycomacromolecule 3 binding to GII.4 Saga P dimer. Mass spectra of GII.4 Saga P dimer and reference protein Cyt c with Glycomacromolecule 3 in three concentrations ranging from 100µM to 200µM at 300 mM ammonium acetate and 20 mM TEAA buffer solution in pH 7. The titration experiments performed in the concentration listed as: A: 100µM(top), B: 150µM(middle) and C: 200µM(bottom).

S16 native mass spectra of glycomacromolecule 4 binding to GII.4 Saga P dimer

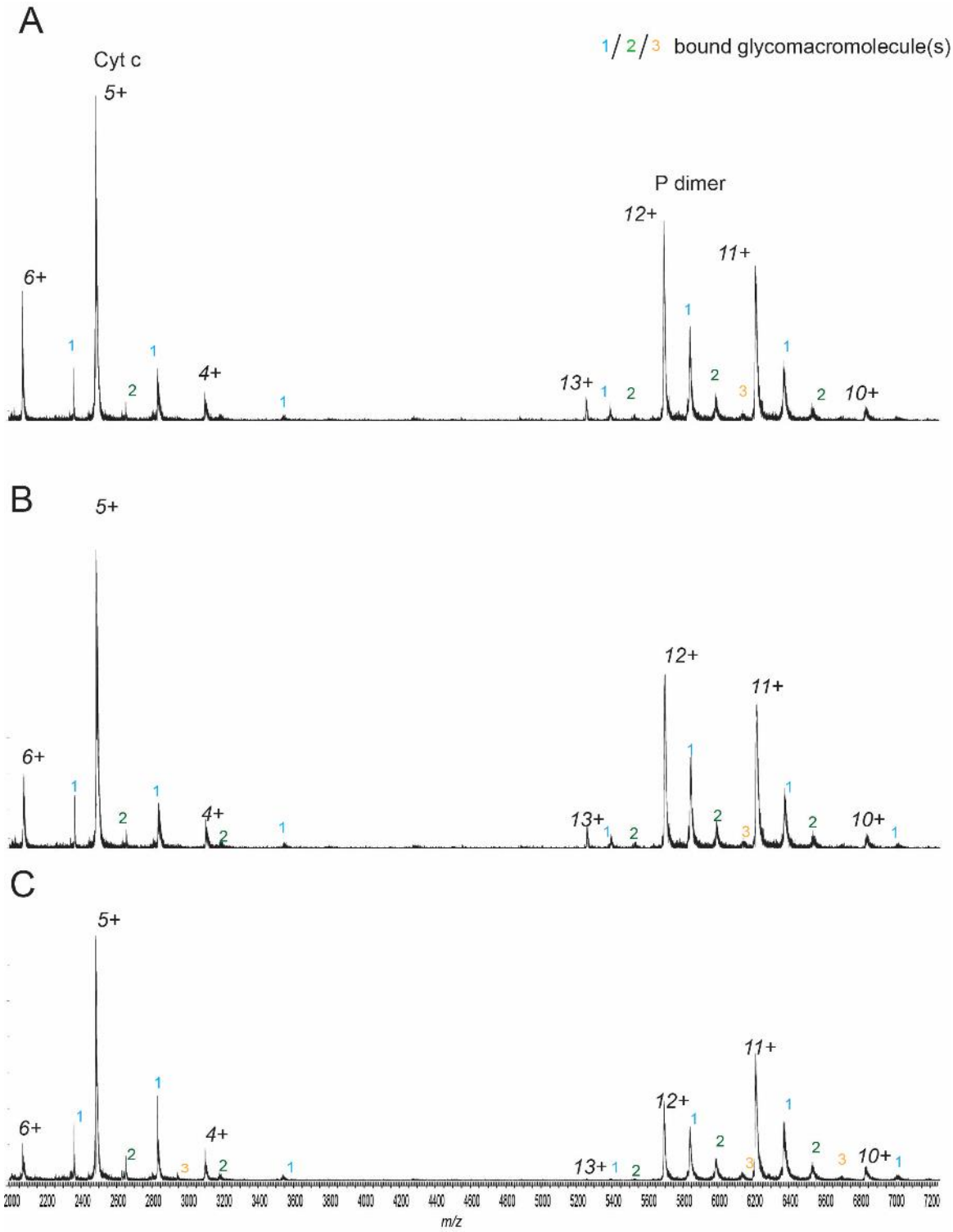


Figure S16 Glycomacromolecule 4 binding to GII.4 Saga P dimer. Mass spectra of GII.4 Saga P dimer and reference protein Cyt c with Glycomacromolecule 4 in three concentrations ranging from 100µM to 200µM at 300 mM ammonium acetate and 20 mM TEAA buffer solution in pH 7.

The titration experiments performed in the concentration listed as: A: 100µM(top), B: 150µM(middle) and C: 200µM(bottom).

S17 native mass spectra of glycomacromolecule 6 binding to GII.4 Saga P dimer

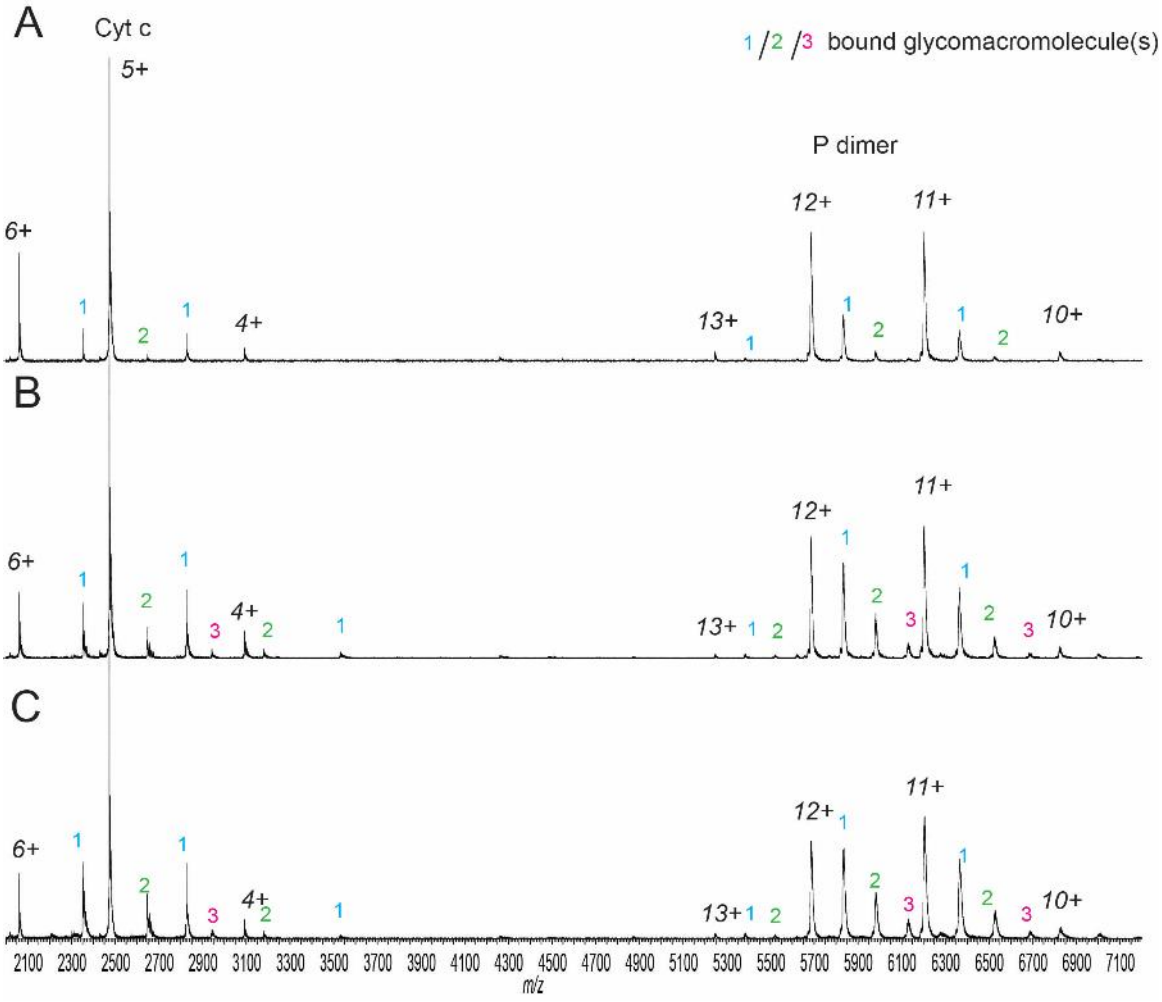


Figure S17 Glycomacromolecule 6 binding to GII.4 Saga P dimer. Mass spectra of GII.4 Saga P dimer and reference protein Cyt c with Glycomacromolecule 6 in three concentrations ranging from 100µM to 200µM at 300 mM ammonium acetate and 20 mM TEAA buffer solution in pH 7. The titration experiments performed in the concentration listed as: A: 100µM(top), B: 150µM(middle) and C: 200µM(bottom).

S18 native mass spectra of glycomacromolecule 7 binding to GII.4 Saga P dimer

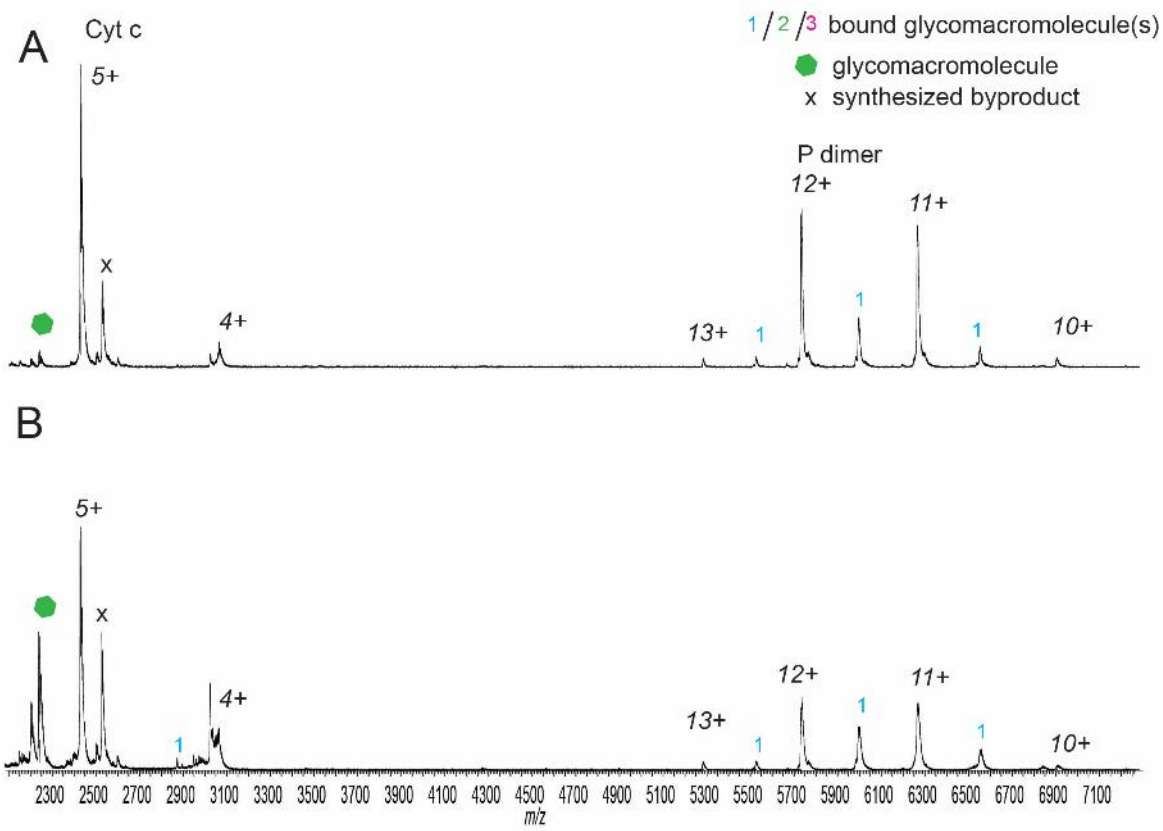


Figure S18 Glycomacromolecule 7 binding to GII.4 Saga P dimer. Mass spectra of GII.4 Saga P dimer and reference protein Cyt c with Glycomacromolecule 7 in two concentrations of 100µM and 200µM at 300 mM ammonium acetate and 20 mM TEAA buffer solution in pH 7. The titration experiments performed in the concentration listed as: A: 100µM(top), B: 200µM(bottom).

S19 native mass spectra of glycomacromolecule 8 binding to GII.4 Saga P dimer

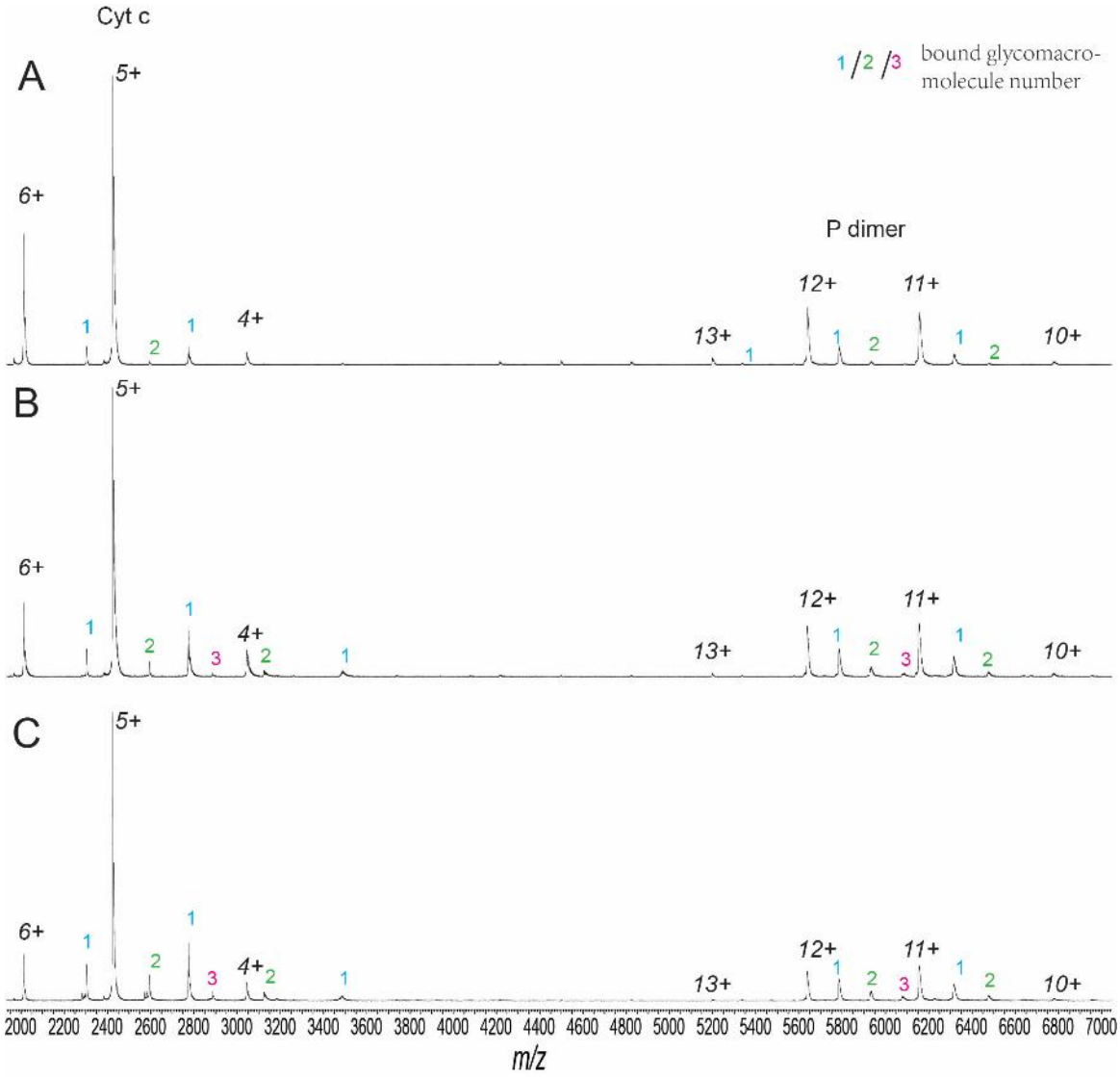


Figure S19 Glycomacromolecule 8 binding to GII.4 Saga P dimer. Mass spectra of GII.4 Saga P dimer and reference protein Cyt c with Glycomacromolecule 8 in three concentrations ranging from 100µM to 200µM at 300 mM ammonium acetate and 20 mM TEAA buffer solution in pH 7. The titration experiments performed in the concentration listed as: A: 100µM(top), B: 150µM(middle) and C: 200µM(bottom).

S20 native mass spectra of glycomacromolecule 9 binding to GII.4 Saga P dimer

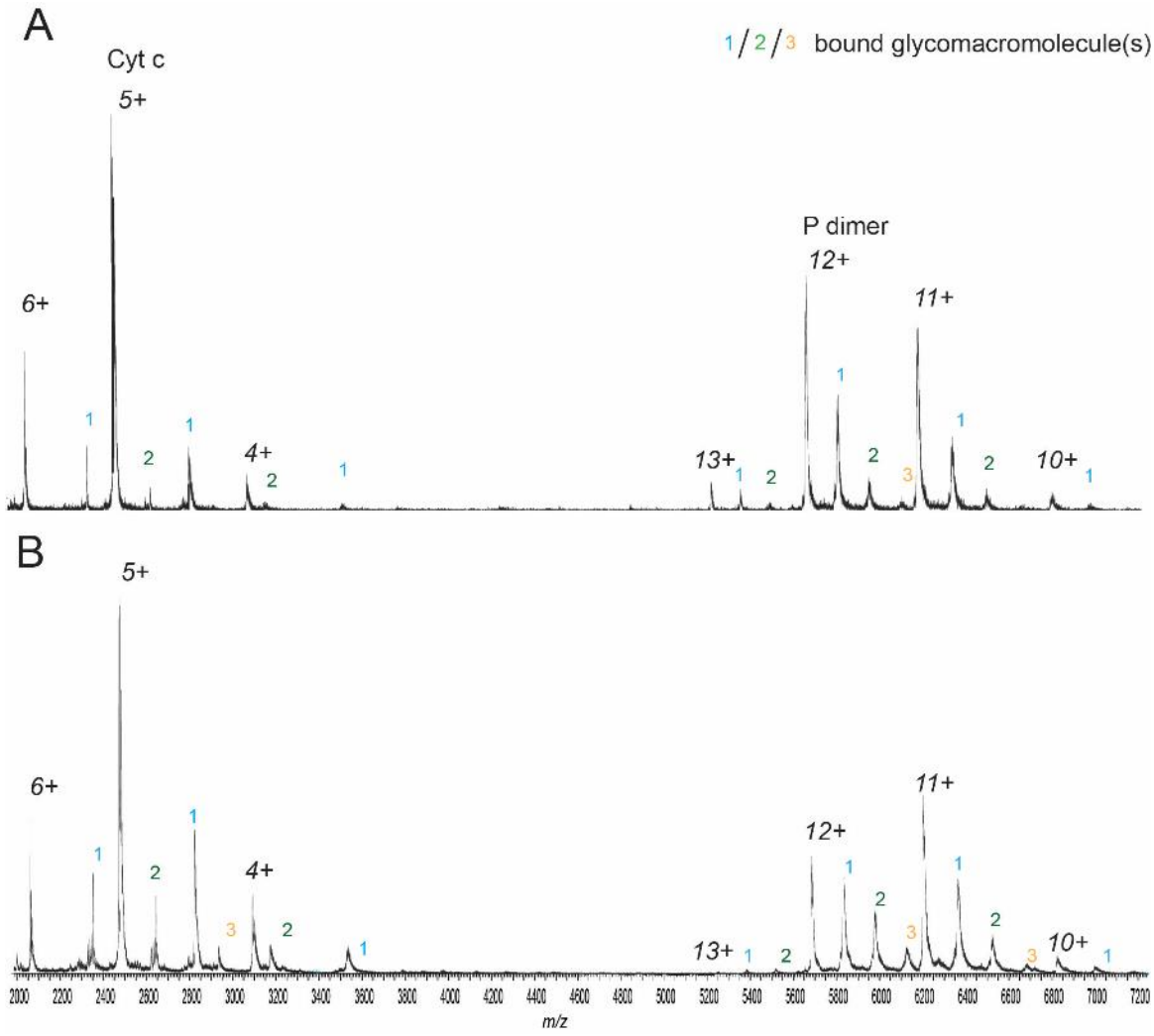


Figure S20 Glycomacromolecule 9 binding to GII.4 Saga P dimer. Mass spectra of GII.4 Saga P dimer and reference protein Cyt c with Glycomacromolecule 9 in two concentrations of 100µM and 200µM at 300 mM ammonium acetate and 20 mM TEAA buffer solution in pH 7. The titration experiments performed in the concentration listed as: A: 100µM(top), B: 200µM(bottom).

S3 Supplementary Tables

S1 calculated Glycan clustering ratio from reference protein candidates

Table S1. Glycan clustering ratio via the interaction of Gb4 and HBGA A, respectively interaction to Saga P dimer analysed with native MS. Data correction is based on the reference protein method.

Ref . Protein	glycan clustering ratio									
	100µM		200µM		300µM		400µM		500µM	
	Gb4	HBGA A	Gb4	HBGA A	Gb4	HBGA A	Gb4	HBGA A	Gb4	HBGA A
Myo	0.14±0.02	0.13±0.03	0.17±0.03	0.16±0.03	0.20±0.03	0.22±0.03	0.30±0.01	0.29±0.04	0.37±0.06	0.34±0.03
Cyt c	0.10±0.01	0.18±0.03	0.15±0.00	0.16±0.02	0.19±0.03	0.19±0.03	0.29±0.04	0.27±0.02	0.45±0.02	0.41±0.09
Ubq	0.03±0.01	0.06±0.03	0.04±0.01	0.04±0.01	0.12±0.02	0.13±0.01	0.21±0.20	0.18±0.01	0.09±0.03	0.10±0.05
GFP	0.27±0.02	0.30±0.02	0.32±0.07	0.35±0.05	0.53±0.01	0.54±0.01	0.69±0.04	0.68±0.04	0.80±0.03	0.86±0.04
CA	0.21±0.01	0.20±0.05	0.32±0.07	0.37±0.01	0.50±0.00	0.42±0.02	0.60±0.05	0.57±0.02	0.75±0.03	0.74±0.07
ADH	0.34±0.04	0.39±0.02	0.66±0.03	0.65±0.02	0.81±0.11	0.88±0.10	1.03±0.08	0.98±0.05	1.26±0.03	1.32±0.10
apo-TFF	0.36±0.05	0.39±0.02	0.57±0.06	0.57±0.05	0.63±0.02	0.65±0.02	0.69±0.02	0.71±0.03	0.80±0.05	0.79±0.03
LDH	0.35±0.04	n.a.	0.51±0.08	n.a.	0.68±0.02	n.a.	0.80±0.01	n.a.	0.91±0.06	n.a.
P dimer	0.35±0.06	0.37±0.13	0.60±0.09	0.66±0.10	1.05±0.09	0.98±0.08	1.21±0.15	1.34±0.26	1.40±0.21	1.44±0.15

S2 Bound number of carbohydrates structure on P dimer determined by native MS

Table S2. Bound number of carbohydrates structure on P dimer analysed with native MS. Data correction is based on the reference protein method.

Ref. Protein	Bound number of HBGA B on Saga P dimer					Bound number of HBGA A on Saga P dimer					Bound number of Gb4 on Saga P dimer				
	100	200	300	400	500	100	200	300	400	500	100	200	300	400	500
Myo	2	3	3	3	4	2	3	3	3	4	2	3	3	3	4
Cyt c	2	3	4	4	4	2	3	4	4	4	1	2	3	3	4
Ubq	2	3	4	4	4	2	3	4	4	4	2	3	4	4	4
GFP	1	1	2	2	2	1	1	1	1	1	2	2	2	2	2
CA	1	2	2	2	2	1	1	1	1	1	1	1	2	2	2
ADH	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
LDH	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1

S3 Information of reference protein candidate physiochemical properties

Table S3. The information of the analysed protein physiochemical properties during the study. Data is obtained from Protein Data Bank in Europe (https://www.ebi.ac.uk/pdbe-srv/pdbechem/)

Ref. Protein	Analysed protein physiochemical feature									
	The share of β-sheet	The share of α-helix	mass	CCS TJM	positive charge residues number	negative charge residues number	pI	charge residues number	solvent accessible area	m/z
ADH	40.60	44.21	147.00	9876.81	128.00	144	6.21	272.00	31928.42	5933
apo-TFF	12.51	27.46	79.60	9959.74	85.00	87	6.81	172.00	15595.88	4469
CA	8.91	4.40	29.10	2378.38	27.00	30	6.41	57.00	6984.81	2985
Cyt c	0.00	5.44	13.20	1286.00	21.00	12	9.59	33.00	3618.71	1864
GFP	13.08	2.61	29.60	2164.24	26.00	34	5.67	60.00	5608.96	3037
LDH	25.07	68.81	146.10	7421.70	148.00	140	8.17	288.00	25045.90	6120
Myo	0.00	12.14	17.00	1714.99	21.00	21	7.20	42.00	4980.50	2283

S4 Monomer/dimer ratio of P dimer species measured by native MS

Table S4. P dimer species oligomerization measured via native MS. The ratio of the three types of oligomers was calculated based on the raw native mass spectrum obtained at 300 mM ammonium acetate solution at pH 7. non-deamidated/wild type and deamidated GII.4 Saga, non-deamidated/wild type and deamidated GII.4 MI001, GII.17 Kawasaki and GII.10 Vietnam P dimers mainly formed as dimers, monomers was presented in deamidated Saga and MI001 species. The formed tetramers of the P dimers were might due to ESI process.

P dimer species	percentage of oligomers		
	monomer	Dimer	Tetramer
Vietnam	4,2%	95,8%	0,0%
non-deamidated Saga	0,0%	99,5%	0,5%
Deamidated Sage	32,1%	60,4%	7,5%
non-deamidated MI001	0,0%	90,9%	9,1%
Deamidated MI001	16,3%	79,9%	3,8%
Kawasaki	0,0%	99,9%	0,1%

S4 List of Figures

S5 List of Tables

S6 Abbreviations

ACAT	Acyltransferase
ADH	Alcohol Dehydrogenase
apo-TFF	Apo-Transferrin
CA	Carbonic Anhydrase Isozyme II
CCS	Collision Cross Section
CEM	Chain Ejection Model
CID	Collision-Induced Dissociation
Cyt c	Cyt ochrome C
DDA	Data-Dependent Analysis
ESI	Electrospray Ionization
FCV	Calicivirus
FWHM	Full Width At Half Maximum
FUT	Fucosyltransferase
Galβ1-3GlcNAc	O-ß-D-Galactopyranosyl-(1-3)-N-Acetylglucosamine
GFP	Green Fluorescent Protein
HBGA	Histo Blood Group Antigen
HDX	Hydrogen Deuterium Exchange
UPLC	High Performance Liquid Chromatography
Hrs	Hours
hNoV	Human Norovirus
IEM	Ion Evaporation Mechanism
ITC	Isothermal Titration Calorimetry
K_D	Dissociation Constants
Le^b	Lewis B Antigen
LDH	L-Lactic Dehydrogenase
LDLR	Low-Density Lipoprotein Receptor
MALDI	Matrix-Assisted Laser Desorption/Ionization
MD	Molecular Dynamics
MS	Mass Spectrometry
MST	Microscale Thermophoresis
Myo	Myoglobin
NMR	Nuclear Magnetic Resonance
P	Protruding
PTM	Post-Translational Modification
QToF	Quadrupole Time-Of-Flight
SEC	Size-Exclusion Chromatography

S	Shell
SPR	Surface Plasmon Resonance
ToF	Time-Of-Flight
Ubq	Ubiquitin
VLP	Virus-Like Particle

S7 Reference

Ahmed, S. M., Hall, A. J., Robinson, A. E., Verhoef, L., Premkumar, P., Parashar, U. D., Koopmans, M., Lopman, B. A. Global prevalence of norovirus in cases of gastroenteritis: a systematic review and meta-analysis. *The Lancet infectious diseases,* 2014;*Volum*: 725-730.

Aliyari, E., Konermann, L. Formation of Gaseous Proteins via the Ion Evaporation Model (IEM) in Electrospray Mass Spectrometry. *Analytical Chemistry,* 2020;*Volum*: 10807-10814.

Arora, J., Hickey, J. M., Majumdar, R., Esfandiary, R., Bishop, S. M., Samra, H. S., Middaugh, C. R., Weis, D. D., Volkin, D. B. (2015). *Hydrogen exchange mass spectrometry reveals protein interfaces and distant dynamic coupling effects during the reversible self-association of an IgG1 monoclonal antibody.* Paper presented at the MAbs.

Báez Bolivar, E. G., Bui, D. T., Kitova, E. N., Han, L., Zheng, R. B., Luber, E. J., Sayed, S. Y., Mahal, L. K., Klassen, J. S. Submicron Emitters Enable Reliable Quantification of Weak Protein–Glycan Interactions by ESI-MS. *Analytical Chemistry,* 2021;*Volum*: 4231-4239.

Bai, Y., Milne, J. S., Mayne, L., Englander, S. W. Primary structure effects on peptide group hydrogen exchange. *Proteins: Structure, Function, and Bioinformatics,* 1993;*Volum*: 75-86.

Baker, K., Morris, J., McCarthy, N., Saldana, L., Lowther, J., Collinson, A., Young, M. An outbreak of norovirus infection linked to oyster consumption at a UK restaurant, February 2010. *Journal of public health,* 2010;*Volum*: 205-211.

Bakhtiar, R., Nelson, R. W. Electrospray ionization and matrix-assisted laser desorption ionization mass spectrometry: Emerging technologies in biomedical sciences. *Biochemical Pharmacology,* 2000;*Volum*: 891-905.

Bartsch, S. M., Lopman, B. A., Ozawa, S., Hall, A. J., Lee, B. Y. Global economic burden of norovirus gastroenteritis. *PloS One,* 2016;*Volum*: e0151219.

Bertolotti-Ciarlet, A., White, L. J., Chen, R., Prasad, B. V., Estes, M. K. Structural requirements for the assembly of Norwalk virus-like particles. *Journal of Virology,* 2002;*Volum*: 4044-4055.

Brandenburg, B., Lee, L. Y., Lakadamyali, M., Rust, M. J., Zhuang, X., Hogle, J. M. Imaging poliovirus entry in live cells. *PLoS Biology,* 2007;*Volum*: e183.

Braun, P., Nägele, B., Wittmann, V., Drescher, M. Mechanism of multivalent carbohydrate–protein interactions studied by EPR spectroscopy. *Angewandte Chemie International Edition,* 2011;*Volum*: 8428-8431.

Brigitte, Fiege, Dr., Christoph, Rademacher, Jonathan, Cartmell, Dr., Pavel, I. Molecular Details of the Recognition of Blood Group Antigens by a Human Norovirus as Determined by STD NMR Spectroscopy. *Angewandte Chemie International Edition,* 2012;*Volum*.

Bücher, K. S., Yan, H., Creutznacher, R., Ruoff, K., Mallagaray, A., Grafmüller, A., Dirks, J. S., Kilic, T., Weickert, S., Rubailo, A. Fucose-functionalized precision glycomacromolecules targeting human norovirus capsid protein. *Biomacromolecules,* 2018;*Volum*: 3714-3724.

Bull, R. A., Tanaka, M. M., White, P. A. Norovirus recombination. *Journal of General Virology,* 2007;*Volum*: 3347-3359.

Bull, R. A., Tu, E. T., McIver, C. J., Rawlinson, W. D., White, P. A. Emergence of a new norovirus genotype II. 4 variant associated with global outbreaks of gastroenteritis. *Journal of Clinical Microbiology,* 2006;*Volum*: 327-333.

Cannon, J. L., Papafragkou, E., Park, G. W., Osborne, J., Jaykus, L.-A., Vinjé, J. Surrogates for the study of norovirus stability and inactivation in the environment: a comparison of murine norovirus and feline calicivirus. *Journal of Food Protection,* 2006;*Volum*: 2761-2765.

Carlsson, B., Kindberg, E., Buesa, J., Rydell, G. E., Lidon, M. F., Montava, R., Mallouh, R. A., Grahn, A., Rodríguez-Díaz, J., Bellido, J. The G428A nonsense mutation in FUT2 provides strong but not absolute protection against symptomatic GII. 4 Norovirus infection. *PloS One,* 2009;*Volum*: e5593.

Cecioni, S., Imberty, A., Vidal, S. Glycomimetics versus multivalent glycoconjugates for the design of high affinity lectin ligands. *Chemical Reviews,* 2015;*Volum*: 525-561.

Chan, M. C., Hu, Y., Chen, H., Podkolzin, A. T., Zaytseva, E. V., Komano, J., Sakon, N., Poovorawan, Y., Vongpunsawad, S., Thanusuwannasak, T. Global spread of norovirus GII. 17 Kawasaki 308, 2014–2016. *Emerging Infectious Diseases,* 2017;*Volum*: 1350.

Changotra, H., Jia, Y., Moore, T. N., Liu, G., Kahan, S. M., Sosnovtsev, S. V., Karst, S. M. Type I and type II interferons inhibit the translation of murine norovirus proteins. *Journal of Virology,* 2009;*Volum*: 5683-5692.

Chhabra, P., de Graaf, M., Parra, G. I., Chan, M. C.-W., Green, K., Martella, V., Wang, Q., White, P. A., Katayama, K., Vennema, H. Updated classification of norovirus genogroups and genotypes. *The Journal of general virology,* 2019;*Volum*: 1393.

Clemmer, D. E., Hudgins, R. R., Jarrold, M. F. Naked Protein Conformations: Cytochrome c in the Gas Phase. *Journal of the American Chemical Society,* 1995;*Volum*: 10141-10142.

Conley, M. J., McElwee, M., Azmi, L., Gabrielsen, M., Byron, O., Goodfellow, I. G., Bhella, D. Calicivirus VP2 forms a portal-like assembly following receptor engagement. *Nature,* 2019;*Volum*: 377.

Consta, S., Malevanets, A. Manifestations of charge induced instability in droplets effected by charged macromolecules. *Physical Review Letters,* 2012;*Volum*: 148301.

Costantini, V. P., Cooper, E. M., Hardaker, H. L., Lee, L. E., Bierhoff, M., Biggs, C., Cieslak, P. R., Hall, A. J., Vinjé, J. Epidemiologic, virologic, and host genetic factors of norovirus outbreaks in long-term care facilities. *Clinical Infectious Diseases,* 2015;*Volum*: 1-10.

Creutznacher, , T., Dülfer, J., Feldmann, C., Hartmann, V., Knickmann, J., Westermann, L. T., Smith, T. J., Uetrecht, C., Mallagaray, A. Murine norovirus capsid plasticity–Glycochenodeoxycholic acid stabilizes P-domain dimers and triggers escape from antibody recognition. *bioRxiv,* 2021;*Volum*.

Creutznacher, R., Maass, T., Ogrissek, P., Wallmann, G., Feldmann, C., Peters, H., Lingemann, M., Taube, S., Peters, T., Mallagaray, A. NMR Experiments Shed New Light on Glycan Recognition by Human and Murine Norovirus Capsid Proteins. *Viruses,* 2021;*Volum*: 416.

Dabelsteen, E. ABO blood group antigens in oral mucosa. What is new? *Journal of Oral Pathology and Medicine,* 2002;*Volum*: 65-70.

Daniel, J. M., Friess, S. D., Rajagopalan, S., Wendt, S., Zenobi, R. Quantitative determination of noncovalent binding interactions using soft ionization mass spectrometry. *International Journal of Mass Spectrometry,* 2002;*Volum*: 1-27.

de Graaf, M., van Beek, J., Vennema, H., Podkolzin, A., Hewitt, J., Bucardo, F., Templeton, K., Mans, J., Nordgren, J., Reuter, G. Emergence of a novel GII. 17 norovirus–End of the GII. 4 era? *Eurosurveillance,* 2015;*Volum*: 21178.

Deng, L., Kitova, E. N., Klassen, J. S. Dissociation Kinetics of the Streptavidin–Biotin Interaction Measured Using Direct Electrospray Ionization Mass Spectrometry Analysis. *Journal of the American Society for Mass Spectrometry,* 2013;*Volum*: 49-56.

Di Martino, B., Di Profio, F., Ceci, C., Di Felice, E., Green, K. Y., Bok, K., De Grazia, S., Giammanco, G. M., Massirio, I., Lorusso, E. Seroprevalence of norovirus genogroup IV antibodies among humans, Italy, 2010–2011. *Emerging Infectious Diseases,* 2014;*Volum*: 1828.

Dominguez-Medina, S., Fostner, S., Defoort, M., Sansa, M., Stark, A.-K., Halim, M. A., Vernhes, E., Gely, M., Jourdan, G., Alava, T. Neutral mass spectrometry of virus capsids above 100 megadaltons with nanomechanical resonators. *Science,* 2018;*Volum*: 918-922.

Duizer, E., Schwab, K. J., Neill, F. H., Atmar, R. L., Koopmans, M. P., Estes, M. K. Laboratory efforts to cultivate noroviruses. *Journal of General Virology,* 2004;*Volum*: 79-87.

Dülfer, J., Yan, H., Brodmerkel, M. N., Creutznacher, R., Mallagaray, A., Peters, T., Caleman, C., Marklund, E. G., Uetrecht, C. Glycan-induced protein dynamics in human norovirus P dimers depend on virus strain and deamidation status. *Molecules,* 2021;*Volum*: 2125.

Dyachenko, A., Gruber, R., Shimon, L., Horovitz, A., Sharon, M. Allosteric mechanisms can be distinguished using structural mass spectrometry. *Proceedings of the National Academy of Ences of the United States of America,* 2013;*Volum*: 7235-7239.

Ebbesen, M. F., Itskalov, D., Baier, M., Hartmann, L. Cu elimination from Cu-coordinating macromolecules. *ACS Macro Letters,* 2017;*Volum*: 399-403.

El-Aneed, A., Cohen, A., Banoub, J. Mass spectrometry, review of the basics: electrospray, MALDI, and commonly used mass analyzers. *Applied Spectroscopy Reviews,* 2009;*Volum*: 210-230.

El-Hawiet, A., Kitova, E. N., Klassen, J. S. Quantifying carbohydrate-protein interactions by electrospray ionization mass spectrometry analysis. *Biochemistry,* 2012;*Volum*: 4244-4253.

El-Hawiet, A., Kitova, E. N., Klassen, J. S. Quantifying carbohydrate–protein interactions by electrospray ionization mass spectrometry analysis. *Biochemistry,* 2012;*Volum*: 4244-4253.

Ettayebi, K., Crawford, S. E., Murakami, K., Broughman, J. R., Karandikar, U., Tenge, V. R., Neill, F. H., Blutt, S. E., Zeng, X.-L., Qu, L. Replication of human noroviruses in stem cell–derived human enteroids. *Science,* 2016;*Volum*: 1387-1393.

Ettayebi, K., Tenge, V. R., Cortes-Penfield, N. W., Crawford, S. E., Neill, F. H., Zeng, X.-L., Yu, X., Ayyar, B. V., Burrin, D., Ramani, S. New insights and enhanced human norovirus cultivation in human intestinal enteroids. *Msphere,* 2021;*Volum*: e01136-01120.

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science,* 1989;*Volum*: 64-71.

Fiege, B., Leuthold, M., Parra, F., Dalton, K. P., Meloncelli, P. J., Lowary, T. L., Peters, T. Epitope mapping of histo blood group antigens bound to norovirus VLPs using STD NMR experiments reveals fine details of molecular recognition. *Glycoconjugate Journal,* 2017;*Volum*: 679-689.

Fiege, B., Rademacher, C., Cartmell, J., Kitov, P. I., Parra, F., Peters, T. Molecular details of the recognition of blood group antigens by a human norovirus as determined by STD NMR spectroscopy. *Angewandte Chemie International Edition,* 2012;*Volum*: 928-932.

Fitzgerald, M. C., Chernushevich, I., Standing, K. G., Whitman, C. P., Kent, S. Probing the oligomeric structure of an enzyme by electrospray ionization time-of-flight mass spectrometry. *Proceedings of the National Academy of Sciences,* 1996;*Volum*: 6851-6856.

Frankevich, V., Barylyuk, K., Chingin, K., Nieckarz, R., Zenobi, R. Native Biomolecules in the Gas Phase? The Case of Green Fluorescent Protein. *Chemphyschem,* 2013;*Volum*.

Frenck, R., Bernstein, D. I., Xia, M., Huang, P., Zhong, W., Parker, S., Dickey, M., McNeal, M., Jiang, X. Predicting susceptibility to norovirus GII. 4 by use of a challenge model involving humans. *The Journal of infectious diseases,* 2012;*Volum*: 1386-1393.

Garcia-Alai, M. M., Heidemann, J., Skruzny, M., Gieras, A., Mertens, H. D., Svergun, D. I., Kaksonen, M., Uetrecht, C., Meijers, R. Epsin and Sla2 form assemblies through phospholipid interfaces. *Nature communications,* 2018;*Volum*: 1-13.

Geng, J., Mantovani, G., Tao, L., Nicolas, J., Chen, G., Wallis, R., Mitchell, D. A., Johnson, B. R., Evans, S. D., Haddleton, D. M. Site-directed conjugation of “clicked” glycopolymers to

form glycoprotein mimics: binding to mammalian lectin and induction of immunological function. *Journal of the American Chemical Society,* 2007;*Volum*: 15156-15163.

Gerke, C., Ebbesen, M. F., Jansen, D., Boden, S., Freichel, T., Hartmann, L. Sequence-controlled glycopolymers via step-growth polymerization of precision glycomacromolecules for lectin receptor clustering. *Biomacromolecules,* 2017;*Volum*: 787-796.

Glass, R. I., Parashar, U. D., Estes, M. K. Norovirus gastroenteritis. *New England Journal of Medicine,* 2009;*Volum*: 1776-1785.

Golden, J. W., Schiff, L. A. Neutrophil elastase, an acid-independent serine protease, facilitates reovirus uncoating and infection in U937 promonocyte cells. *Virology Journal,* 2005;*Volum*: 48.

Grandori, R. Origin of the conformation dependence of protein charge-state distributions in electrospray ionization mass spectrometry. *Journal of Mass Spectrometry,* 2010;*Volum*: -.

Grayson, M. A. (2002). *Measuring mass: from positive rays to proteins*: Chemical Heritage Foundation.

Guiard, J., Fiege, B., Kitov, P. I., Peters, T., Bundle, D. R. “Double‐Click” Protocol for Synthesis of Heterobifunctional Multivalent Ligands: Toward a Focused Library of Specific Norovirus Inhibitors. *Chemistry–A European Journal,* 2011;*Volum*: 7438-7441.

Gulbakan, B., Barylyuk, K., Schneider, P., Pillong, M., Schneider, G., Zenobi, R. Native Electrospray Ionization Mass Spectrometry Reveals Multiple Facets of Aptamer-Ligand Interactions: From Mechanism to Binding Constants. *Journal of the American Chemical Society,* 2018;*Volum*: jacs.7b13044.

Gülbakan, B., Barylyuk, K., Zenobi, R. Determination of thermodynamic and kinetic properties of biomolecules by mass spectrometry. *Current Opinion in Biotechnology,* 2015;*Volum*: 65-72.

Haga, K., Fujimoto, A., Takai-Todaka, R., Miki, M., Doan, Y. H., Murakami, K., Yokoyama, M., Murata, K., Nakanishi, A., Katayama, K. Functional receptor molecules CD300lf and CD300ld within the CD300 family enable murine noroviruses to infect cells. *Proceedings of the National Academy of Sciences,* 2016;*Volum*: E6248-E6255.

Hageman, T. S., Weis, D. D. Reliable identification of significant differences in differential hydrogen exchange-mass spectrometry measurements using a hybrid significance testing approach. *Analytical Chemistry,* 2019;*Volum*: 8008-8016.

Han, Kitova, E. N., Tan, M., Jiang, X., Pluvinage, B., Boraston, A. B., Klassen, J. S. Affinities of human histo-blood group antigens for norovirus capsid protein complexes. *Glycobiology,* 2014;*Volum*: 170-180.

Han, Tan, M., Xia, M., Kitova, E. N., Jiang, X., Klassen, J. S. Gangliosides are ligands for human noroviruses. *Journal of the American Chemical Society,* 2014;*Volum*: 12631-12637.

Han, L., Kitov, P. I., Kitova, E. N., Tan, M., Wang, L., Xia, M., Jiang, X., Klassen, J. S. Affinities of recombinant norovirus P dimers for human blood group antigens. *Glycobiology,* 2012;*Volum*: 276-285.

Han, L., Kitov, P. I., Kitova, E. N., Tan, M., Wang, L., Xia, M., Jiang, X., Klassen, J. S. Affinities of recombinant norovirus P dimers for human blood group antigens. *Glycobiology,* 2013;*Volum*: 276-285.

Han, L., Kitova, E. N., Tan, M., Jiang, X., Klassen, J. S. Identifying carbohydrate ligands of a norovirus P particle using a catch and release electrospray ionization mass spectrometry assay. *Journal of the American Society for Mass Spectrometry,* 2014;*Volum*: 111-119.

Han, L., Tan, M., Xia, M., Kitova, E. N., Jiang, X., Klassen, J. S. Gangliosides are Ligands for Human Noroviruses. *Journal of the American Chemical Society,* 2014;*Volum*: 12631.

Han, L., Zheng, R., Richards, M. R., Tan, M., Kitova, E. N., Jiang, X., Klassen, J. S. Quantifying the binding stoichiometry and affinity of histo-blood group antigen oligosaccharides for human noroviruses. *Glycobiology,* 2018;*Volum*: 488-498.

Hansman, Biertümpfel, C., Georgiev, I., McLellan, J. S., Chen, L., Zhou, T., Katayama, K., Kwong, P. D. Crystal structures of GII. 10 and GII. 12 norovirus protruding domains in complex with histo-blood group antigens reveal details for a potential site of vulnerability. *Journal of Virology,* 2011;*Volum*: 6687-6701.

Hansman, G. S., Biertümpfel, C., Georgiev, I., McLellan, J. S., Chen, L., Zhou, T., Katayama, K., Kwong, P. D. Crystal structures of GII. 10 and GII. 12 norovirus protruding domains in complex with histo-blood group antigens reveal details for a potential site of vulnerability. *Journal of Virology,* 2011;*Volum*: 6687-6701.

Harrington, P. R., Lindesmith, L., Yount, B., Moe, C. L., Baric, R. S. Binding of Norwalk virus-like particles to ABH histo-blood group antigens is blocked by antisera from infected human volunteers or experimentally vaccinated mice. *Journal of Virology,* 2002;*Volum*: 12335-12343.

Hoshino, Y., Imamura, K., Yue, M., Inoue, G., Miura, Y. Reversible absorption of CO2 triggered by phase transition of amine-containing micro-and nanogel particles. *Journal of the American Chemical Society,* 2012;*Volum*: 18177-18180.

Houde, D., Arndt, J., Domeier, W., Berkowitz, S., Engen, J. R. Characterization of IgG1 conformation and conformational dynamics by hydrogen/deuterium exchange mass spectrometry. *Analytical Chemistry,* 2009;*Volum*: 2644-2651.

Houde, D., Berkowitz, S. A., Engen, J. R. The utility of hydrogen/deuterium exchange mass spectrometry in biopharmaceutical comparability studies. *Journal of Pharmaceutical Sciences,* 2011;*Volum*: 2071-2086.

Huang, P., Farkas, T., Marionneau, S., Zhong, W., Ruvoën-Clouet, N., Morrow, A. L., Altaye, M., Pickering, L. K., Newburg, D. S., LePendu, J. Noroviruses bind to human ABO, Lewis, and secretor histo-blood group antigens: identification of 4 distinct strain-specific patterns. *The Journal of infectious diseases,* 2003;*Volum*: 19-31.

Hutson, A. M., Atmar, R. L., Graham, D. Y., Estes, M. K. Norwalk virus infection and disease is associated with ABO histo-blood group type. *The Journal of infectious diseases,* 2002;*Volum*: 1335-1337.

Hvidt, A., Linderstrøm-Lang, K. Exchange of hydrogen atoms in insulin with deuterium atoms in aqueous solutions. *Biochimica et Biophysica Acta,* 1954;*Volum*: 574.

Igde, S., Röblitz, S., Müller, A., Kolbe, K., Boden, S., Fessele, C., Lindhorst, T. K., Weber, M., Hartmann, L. Linear precision glycomacromolecules with varying interligand spacing and linker functionalities binding to Concanavalin A and the bacterial lectin FimH. *Macromolecular Bioscience,* 2017;*Volum*: 1700198.

Jecklin, Schauer, S., Dumelin, C. E., Zenobi, R. Label-free determination of protein-ligand binding constants using mass spectrometry and validation using surface plasmon resonance and isothermal titration calorimetry. *Journal of Molecular Recognition,* 2010;*Volum*: 319-329.

Jecklin, Touboul, D., Bovet, C., Wortmann, A., Zenobi, R. Which electrospray-based ionization method best reflects protein-ligand interactions found in solution? A comparison of ESI, nanoESI, and ESSI for the determination of dissociation constants with mass spectrometry. *Journal of the American Society for Mass Spectrometry,* 2008;*Volum*: 332-343.

Jiang, X., Wang, M., Graham, D. Y., Estes, M. K. Expression, self-assembly, and antigenicity of the Norwalk virus capsid protein. *Journal of Virology,* 1992;*Volum*: 6527-6532.

Jin, M., He, Y., Li, H., Huang, P., Zhong, W., Yang, H., Zhang, H., Tan, M., Duan, Z.-j. Two gastroenteritis outbreaks caused by GII Noroviruses: host susceptibility and HBGA phenotypes. *PloS One,* 2013;*Volum*: e58605.

Johnson, R. S., Walsh, K. A. Mass spectrometric measurement of protein amide hydrogen exchange rates of apo - and holo - myoglobin. *Protein Science,* 1994;*Volum*: 2411-2418.

Jung, J., Grant, T., Thomas, D. R., Diehnelt, C. W., Grigorieff, N., Joshua-Tor, L. High-resolution cryo-EM structures of outbreak strain human norovirus shells reveal size variations. *Proceedings of the National Academy of Sciences,* 2019;*Volum*: 12828-12832.

Kaltashov, I. A., Bobst, C. E., Abzalimov, R. R. Hydrogen/Deuterium Exchange Mass Spectrometry (HDX MS) in the studies of architecture, dynamics, and interactions of biopharmaceutical products. *Mass Spectrometry Handbook. Hoboken, NJ: John Wiley & Sons, Inc,* 2012;*Volum*: 227-241.

Kapikian, A. Z., Wyatt, R. G., Dolin, R., Thornhill, T. S., Kalica, A. R., Chanock, R. M. Visualization by immune electron microscopy of a 27-nm particle associated with acute infectious nonbacterial gastroenteritis. *Journal of Virology,* 1972;*Volum*: 1075-1081.

Karst, S. M., Wobus, C. E., Lay, M., Davidson, J., Virgin, H. W. STAT1-dependent innate immunity to a Norwalk-like virus. *Science,* 2003;*Volum*: 1575-1578.

Kavan, D., Man, P. MSTools—Web based application for visualization and presentation of HXMS data. *International Journal of Mass Spectrometry,* 2011;*Volum*: 53-58.

Keifer, D. Z., Motwani, T., Teschke, C. M., Jarrold, M. F. Measurement of the accurate mass of a 50 MDa infectious virus. *Rapid Communications in Mass Spectrometry,* 2016;*Volum*: 1957-1962.

Kitajima, M., Oka, T., Haramoto, E., Takeda, N., Katayama, K., Katayama, H. Seasonal distribution and genetic diversity of genogroups I, II, and IV noroviruses in the Tamagawa River, Japan. *Environmental Science & Technology,* 2010;*Volum*: 7116-7122.

Kitova, E. N., El-Hawiet, A., Schnier, P. D., Klassen, J. S. Reliable determinations of protein-ligand interactions by direct ESI-MS measurements. Are we there yet? *Journal of the American Society for Mass Spectrometry,* 2012;*Volum*: 431-441.

Klement, E., Halperin, T. Reply to Chan et al., Tan and Jiang, and St. Clair and Patel. *The Journal of infectious diseases,* 2008;*Volum*: 942-943.

Konermann, L., Pan, J., Liu, Y.-H. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chemical Society Reviews,* 2011;*Volum*: 1224-1234.

Konermann, L., Tong, X., Pan, Y. Protein structure and dynamics studied by mass spectrometry: H/D exchange, hydroxyl radical labeling, and related approaches. *Journal of Mass Spectrometry,* 2008;*Volum*: 1021-1036.

Konstantin, B., Shelimov, David, E., Clemmer, Robert, R., Hudgins, Martin. Protein Structure in Vacuo: Gas-Phase Conformations of BPTI and Cytochrome c. *Journal of the American Chemical Society,* 1997;*Volum*.

Koromyslova, Leuthold, M. M., Bowler, M. W., Hansman, G. S. The sweet quartet: binding of fucose to the norovirus capsid. *Virology,* 2015;*Volum*: 203-208.

Koromyslova, Tripathi, S., Morozov, V., Schroten, H., Hansman, G. S. Human norovirus inhibition by a human milk oligosaccharide. *Virology,* 2017;*Volum*: 81-89.

Koromyslova, A. D., Leuthold, M. M., Bowler, M. W., Hansman, G. S. The sweet quartet: binding of fucose to the norovirus capsid. *Virology,* 2015;*Volum*: 203-208.

Kreutz, L. C., Seal, B., Mengeling, W. Early interaction of feline calicivirus with cells in culture. *Archives of Virology,* 1994;*Volum*: 19-34.

Krutchinsky, A. N., Ayed, A., Donald, L. J., Ens, W., Duckworth, H. W., Standing, K. G. (2000). Studies of Noncovalent Complexes in an Electrospray Ionization/Time-of-Flight Mass Spectrometer *Mass spectrometry of proteins and peptides* (pp. 239-249): Springer.

Lay, M. K., Atmar, R. L., Guix, S., Bharadwaj, U., He, H., Neill, F. H., Sastry, K. J., Yao, Q., Estes, M. K. Norwalk virus does not replicate in human macrophages or dendritic cells derived from the peripheral blood of susceptible humans. *Virology,* 2010;*Volum*: 1-11.

Lindesmith, L., Moe, C., Marionneau, S., Ruvoen, N., Jiang, X., Lindblad, L., Stewart, P., LePendu, J., Baric, R. Human susceptibility and resistance to Norwalk virus infection. *Nature Medicine,* 2003;*Volum*: 548.

Ling, H., Kitova, E. N., Ming, T., Xi, J., Benjamin, P., Boraston, A. B., Klassen, J. S. Affinities of human histo-blood group antigens for norovirus capsid protein complexes. *Glycobiology,* 2015;*Volum*: 170.

Loo, J. A. Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrometry Reviews,* 1997;*Volum*: 1-23.

Loo, J. A. Electrospray ionization mass spectrometry: a technology for studying noncovalent macromolecular complexes. *International Journal of Mass Spectrometry,* 2000;*Volum*: 175-186.

Lorenzen, K., Versluis, C., van Duijn, E., van den Heuvel, R. H., Heck, A. J. Optimizing macromolecular tandem mass spectrometry of large non-covalent complexes using heavy collision gases. *International Journal of Mass Spectrometry,* 2007;*Volum*: 198-206.

Lu, J., Sun, L., Fang, L., Yang, F., Mo, Y., Lao, J., Zheng, H., Tan, X., Lin, H., Rutherford, S. Gastroenteritis outbreaks caused by norovirus GII. 17, Guangdong Province, China, 2014–2015. *Emerging Infectious Diseases,* 2015;*Volum*: 1240.

Makino, A., Shimojima, M., Miyazawa, T., Kato, K., Tohya, Y., Akashi, H. Junctional adhesion molecule 1 is a functional receptor for feline calicivirus. *Journal of Virology,* 2006;*Volum*: 4482-4490.

Mallagaray, Creutznacher, R., Dülfer, J., Mayer, P. H., Grimm, L. L., Orduña, J. M., Trabjerg, E., Stehle, T., Rand, K. D., Blaum, B. S. A post-translational modification of human Norovirus capsid protein attenuates glycan binding. *Nature communications,* 2019;*Volum*: 1-14.

Mallagaray, A., Creutznacher, R., Dulfer, J., Mayer, P. H. O., Grimm, L. L., Orduna, J. M., Trabjerg, E., Stehle, T., Rand, K. D., Blaum, B. S., Uetrecht, C., Peters, T. A post-translational modification of human Norovirus capsid protein attenuates glycan binding. *Nat Commun,* 2019;*Volum*: 1320.

Mallagaray, A., Lockhauserbäumer, J., Hansman, G., Uetrecht, C., Peters, T. Attachment of norovirus to histo blood group antigens: a cooperative multistep process. *Angewandte Chemie International Edition,* 2015;*Volum*: 12014-12019.

Marciano, D. P., Dharmarajan, V., Griffin, P. R. HDX-MS guided drug discovery: small molecules and biopharmaceuticals. *Current Opinion in Structural Biology,* 2014;*Volum*: 105-111.

Marionneau, S., Ruvoën, N., Le Moullac–Vaidye, B., Clement, M., Cailleau–Thomas, A., Ruiz–Palacois, G., Huang, P., Jiang, X., Le Pendu, J. Norwalk virus binds to histo-blood group antigens present on gastroduodenal epithelial cells of secretor individuals. *Gastroenterology,* 2002;*Volum*: 1967-1977.

Masson, G. R., Burke, J. E., Ahn, N. G., Anand, G. S., Borchers, C., Brier, S., Bou-Assaf, G. M., Engen, J. R., Englander, S. W., Faber, J. Recommendations for performing, interpreting

and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. *Nature Methods,* 2019;*Volum*: 595-602.

Maunula, L., Miettinen, I. T., Von Bonsdorff, C.-H. Norovirus outbreaks from drinking water. *Emerging Infectious Diseases,* 2005;*Volum*: 1716.

Metwally, H., Duez, Q., Konermann, L. Chain ejection model for electrospray ionization of unfolded proteins: evidence from atomistic simulations and ion mobility spectrometry. *Analytical Chemistry,* 2018;*Volum*: 10069-10077.

Motomura, K., Oka, T., Yokoyama, M., Nakamura, H., Mori, H., Ode, H., Hansman, G. S., Katayama, K., Kanda, T., Tanaka, T. Identification of monomorphic and divergent haplotypes in the 2006-2007 norovirus GII/4 epidemic population by genomewide tracing of evolutionary history. *Journal of Virology,* 2008;*Volum*: 11247-11262.

Mumphrey, S. M., Changotra, H., Moore, T. N., Heimann-Nichols, E. R., Wobus, C. E., Reilly, M. J., Moghadamfalahi, M., Shukla, D., Karst, S. M. Murine norovirus 1 infection is associated with histopathological changes in immunocompetent hosts, but clinical disease is prevented by STAT1-dependent interferon responses. *Journal of Virology,* 2007;*Volum*: 3251-3263.

Nayak, M. K., Balasubramanian, G., Sahoo, G. C., Bhattacharya, R., Vinje, J., Kobayashi, N., Sarkar, M. C., Bhattacharya, M. K., Krishnan, T. Detection of a novel intergenogroup recombinant Norovirus from Kolkata, India. *Virology,* 2008;*Volum*: 117-123.

Neubert, T. A., Walsh, K. A., Hurley, J. B., Johnson, R. S. Monitoring calcium - induced conformational changes in recoverin by electrospray mass spectrometry. *Protein Science,* 1997;*Volum*: 843-850.

Nilsson, M., Hedlund, K.-O., Thorhagen, M., Larson, G., Johansen, K., Ekspong, A., Svensson, L. Evolution of human calicivirus RNA in vivo: accumulation of mutations in the protruding P2 domain of the capsid leads to structural changes and possibly a new phenotype. *Journal of Virology,* 2003;*Volum*: 13117-13124.

Nordgren, J., Kindberg, E., Lindgren, P.-E., Matussek, A., Svensson, L. Norovirus gastroenteritis outbreak with a secretor-independent susceptibility pattern, Sweden. *Emerging Infectious Diseases,* 2010;*Volum*: 81.

Nordgren, J., Nitiema, L. W., Ouermi, D., Simpore, J., Svensson, L. Host genetic factors affect susceptibility to norovirus infections in Burkina Faso. *PloS One,* 2013;*Volum*: e69557.

Nordgren, J., Svensson, L. Genetic susceptibility to human norovirus infection: an update. *Viruses,* 2019;*Volum*: 226.

Ohguro, H., Palczewski, K., Walsh, K. A., Johnson, R. S. Topographic study of arrestin using differential chemical modifications and hydrogen/deuterium exchange. *Protein Science,* 1994;*Volum*: 2428-2434.

Pacholarz, K. J., Garlish, R. A., Taylor, R. J., Barran, P. E. Mass spectrometry based tools to investigate protein–ligand interactions for drug discovery. *Chemical Society Reviews,* 2012;*Volum*: 4335-4355.

Pérez, L., Carrasco, L. Entry of poliovirus into cells does not require a low-pH step. *Journal of Virology,* 1993;*Volum*: 4543-4548.

Perry, J. W., Taube, S., Wobus, C. E. Murine norovirus-1 entry into permissive macrophages and dendritic cells is pH-independent. *Virus Research,* 2009;*Volum*: 125-129.

Pogan, Reimer, R., Hansman, G., Uetrecht, C. Norovirus-like VP1 particles exhibit isolate dependent stability profiles. *Journal of Physics: Condensed Matter,* 2018;*Volum*: 064006.

Ponader, D., Maffre, P., Aretz, J., Pussak, D., Ninnemann, N. M., Schmidt, S., Seeberger, P. H., Rademacher, C., Nienhaus, G. U., Hartmann, L. Carbohydrate-lectin recognition of sequence-defined heteromultivalent glycooligomers. *Journal of the American Chemical Society,* 2014;*Volum*: 2008-2016.

Ponader, D., Wojcik, F., Beceren-Braun, F., Dernedde, J., Hartmann, L. Sequence-defined glycopolymer segments presenting mannose: synthesis and lectin binding affinity. *Biomacromolecules,* 2012;*Volum*: 1845-1852.

Prasad, B. V., Hardy, M. E., Dokland, T., Bella, J., Rossmann, M. G., Estes, M. K. X-ray crystallographic structure of the Norwalk virus capsid. *Science,* 1999;*Volum*: 287-290.

Przybylski, M., Glocker, M. O. Electrospray mass spectrometry of biomacromolecular complexes with noncovalent interactions—new analytical perspectives for supramolecular chemistry and molecular recognition processes. *Angewandte Chemie International Edition in English,* 1996;*Volum*: 806-826.

Rademacher, C., Guiard, J., Kitov, P. I., Fiege, B., Dalton, K. P., Parra, F., Bundle, D. R., Peters, T. Targeting norovirus infection—multivalent entry inhibitor design based on NMR experiments. *Chemistry–A European Journal,* 2011;*Volum*: 7442-7453.

Rayleigh, L. XX. On the equilibrium of liquid conducting masses charged with electricity. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science,* 1882;*Volum*: 184-186.

Regan, A. D., Shraybman, R., Cohen, R. D., Whittaker, G. R. Differential role for low pH and cathepsin-mediated cleavage of the viral spike protein during entry of serotype II feline coronaviruses. *Veterinary Microbiology,* 2008;*Volum*: 235-248.

Robinson, C. V., Groß, M., Eyles, S. J., Ewbank, J. J., Mayhew, M., Hartl, F. U., Dobson, C. M., Radford, S. E. Conformation of GroEL-bound α-lactalbumin probed by mass spectrometry. *Nature,* 1994;*Volum*: 646.

Robinson, C. V., Radford, S. E. Weighing the evidence for structure: electrospray ionization mass spectrometry of proteins. *Structure,* 1995;*Volum*: 861-865.

Rostom, A. A., Robinson, C. V. Detection of the intact GroEL chaperonin assembly by mass spectrometry. *Journal of the American Chemical Society,* 1999;*Volum*: 4718-4719.

Samalikova, M., Grandori, R. Protein charge-state distributions in electrospray-ionization mass spectrometry do not appear to be limited by the surface tension of the solvent. *Journal of the American Chemical Society,* 2003;*Volum*: 13352-13353.

Sánchez-San Martín, C., López, T., Arias, C. F., López, S. Characterization of rotavirus cell entry. *Journal of Virology,* 2004;*Volum*: 2310-2318.

Seo, J., Hoffmann, W., Warnke, S., Bowers, M. T., Pagel, K., von Helden, G. Retention of native protein structures in the absence of solvent: a coupled ion mobility and spectroscopic study. *Angewandte Chemie International Edition,* 2016;*Volum*: 14173-14176.

Siebenga, J. J., Vennema, H., Renckens, B., de Bruin, E., van der Veer, B., Siezen, R. J., Koopmans, M. Epochal evolution of GGII. 4 norovirus capsid proteins from 1995 to 2006. *Journal of Virology,* 2007;*Volum*: 9932-9941.

Siuzdak, G., Bothner, B., Yeager, M., Brugidou, C., Change, C. M. Mass spectrometry and viral analysis. *Chemistry and Biology,* 1996;*Volum*: 45-48.

Smith, H. Q., Smith, T. J. The dynamic capsid structures of the noroviruses. *Viruses,* 2019;*Volum*: 235.

Snijder, J., Rose, R. J., Veesler, D., Johnson, J. E., Heck, A. J. Studying 18 MDa virus assemblies with native mass spectrometry. *Angewandte Chemie International Edition,* 2013;*Volum*: 4020-4023.

Snijder, J., van de Waterbeemd, M., Damoc, E., Denisov, E., Grinfeld, D., Bennett, A., Agbandje-McKenna, M., Makarov, A., Heck, A. J. Defining the stoichiometry and cargo load of viral and bacterial nanoparticles by Orbitrap mass spectrometry. *Journal of the American Chemical Society,* 2014;*Volum*: 7295-7299.

Sobott, F., Robinson, C. V. Characterising electrosprayed biomolecules using tandem-MS—the noncovalent GroEL chaperonin assembly. *International Journal of Mass Spectrometry,* 2004;*Volum*: 25-32.

St. Clair, K., Patel, S. Intital descriptive and analytical data on an outbreak of norovirus infection at marine corps recruit depot Parris Island, South Carolina. *The Journal of infectious diseases,* 2008;*Volum*: 941-942.

Stuart, A. D., Brown, T. D. K. α2, 6-Linked sialic acid acts as a receptor for feline calicivirus. *Journal of General Virology,* 2007;*Volum*: 177-186.

Sun, Kitova, E. N., Wang, W., Klassen, J. S. Method for Distinguishing Specific from Nonspecific Protein?Ligand Complexes in Nanoelectrospray Ionization Mass Spectrometry. *Analytical Chemistry,* 2006;*Volum*: 3010-3018.

Sun, E., He, J., Zhuang, X. Live cell imaging of viral entry. *Current Opinion in Virology,* 2013;*Volum*: 34-43.

Sun, J., Kitova, E. N., Sun, N., Klassen, J. S. Method for Identifying Nonspecific Protein?Protein Interactions in Nanoelectrospray Ionization Mass Spectrometry. *Analytical Chemistry*: 8301-8311.

Taflin, D. C., Ward, T. L., Davis, E. J. Electrified droplet fission and the Rayleigh limit. *Langmuir,* 1989;*Volum*: 376-384.

Tamminen, K., Huhti, L., Koho, T., Lappalainen, S., Hytönen, V. P., Vesikari, T., Blazevic, V. A comparison of immunogenicity of norovirus GII‐4 virus‐like particles and P‐particles. *Immunology,* 2012;*Volum*: 89-99.

Tan, M., Hegde, R. S., Jiang, X. The P domain of norovirus capsid protein forms dimer and binds to histo-blood group antigen receptors. *Journal of Virology,* 2004;*Volum*: 6233-6242.

Tan, M., Jiang, X. Norovirus and its histo-blood group antigen receptors: an answer to a historical puzzle. *Trends in Microbiology,* 2005a;*Volum*: 285-293.

Tan, M., Jiang, X. The p domain of norovirus capsid protein forms a subviral particle that binds to histo-blood group antigen receptors. *Journal of Virology,* 2005b;*Volum*: 14017-14030.

Tan, M., Jiang, X. Norovirus gastroenteritis, carbohydrate receptors, and animal models. *PLoS Pathogens,* 2010;*Volum*: e1000983.

Tan, M., Jin, M., Xie, H., Duan, Z., Jiang, X., Fang, Z. Outbreak studies of a GII‐3 and a GII‐4 norovirus revealed an association between HBGA phenotypes and viral infection. *Journal of Medical Virology,* 2008;*Volum*: 1296-1301.

Taube, S., Kolawole, A. O., Höhne, M., Wilkinson, J. E., Handley, S. A., Perry, J. W., Thackray, L. B., Akkina, R., Wobus, C. E. A mouse model for human norovirus. *MBio,* 2013;*Volum*: e00450-00413.

Taube, S., Perry, J. W., Yetming, K., Patel, S. P., Auble, H., Shu, L., Nawar, H. F., Lee, C. H., Connell, T. D., Shayman, J. A. Ganglioside-linked terminal sialic acid moieties on murine macrophages function as attachment receptors for murine noroviruses. *Journal of Virology,* 2009;*Volum*: 4092-4101.

Testa, L., Brocca, S., Santambrogio, C., D'Urzo, A., Habchi, J., Longhi, S., Uversky, V. N., Grandori, R. Extracting structural information from charge-state distributions of intrinsically disordered proteins by non-denaturing electrospray-ionization mass spectrometry. *Intrinsically disordered proteins,* 2013;*Volum*: e25068.

Thorven, M., Grahn, A., Hedlund, K.-O., Johansson, H., Wahlfrid, C., Larson, G., Svensson, L. A homozygous nonsense mutation (428G → A) in the human secretor (FUT2) gene provides resistance to symptomatic norovirus (GGII) infections. *Journal of Virology,* 2005;*Volum*: 15351-15355.

Trcka, F., Durech, M., Vankova, P., Chmelik, J., Martinkova, V., Hausner, J., Kadek, A., Marcoux, J., Klumpler, T., Vojtesek, B. Human Stress-inducible Hsp70 Has a High Propensity to Form ATP-dependent Antiparallel Dimers That Are Differentially Regulated by Cochaperone Binding*[S]. *Molecular & Cellular Proteomics,* 2019;*Volum*: 320-337.

Tu, L.-T., Liu, F.-P., Huang, Y.-C., Huang, C.-G., Yang, S., Tsao, K.-C., Lai, M.-W., Chen, C.-J. Genetic susceptibility to norovirus GII. 4 Sydney strain infections in Taiwanese children. *The Pediatric infectious disease journal,* 2017;*Volum*: 353-357.

Tubiana, T., Boulard, Y., Bressanelli, S. Dynamics and asymmetry in the dimer of the norovirus major capsid protein. *PloS One,* 2017;*Volum*: e0182056.

Tyanova, S., Temu, T., Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature Protocols,* 2016;*Volum*: 2301-2319.

Uetrecht, C., Heck, A. J. Modern biomolecular mass spectrometry and its role in studying virus structure, dynamics, and assembly. *Angewandte Chemie International Edition,* 2011;*Volum*: 8248-8262.

Van Beek, J., Ambert-Balay, K., Botteldoorn, N., Eden, J., Fonager, J., Hewitt, J., Iritani, N., Kroneman, A., Vennema, H., Vinje, J. Indications for worldwide increased norovirus activity associated with emergence of a new variant of genotype II. 4, late 2012. *Eurosurveillance,* 2013;*Volum*: 20345.

van den Heuvel, R. H., Heck, A. J. Native protein mass spectrometry: from intact oligomers to functional machineries. *Current Opinion in Chemical Biology,* 2004;*Volum*: 519-526.

Vanoni, M., Curti, B. Glutamate synthase: a complex iron-sulfur flavoprotein. *Cellular and Molecular Life Sciences CMLS,* 1999;*Volum*: 617-638.

Veenstra, T. D. Electrospray ionization mass spectrometry in the study of biomolecular non-covalent interactions. *Biophysical Chemistry,* 1999a;*Volum*: 63-79.

Veenstra, T. D. Electrospray ionization mass spectrometry: a promising new technique in the study of protein/DNA noncovalent complexes. *Biochemical and Biophysical Research Communications,* 1999b;*Volum*: 1-5.

Wales, T. E., Engen, J. R. Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrometry Reviews,* 2006;*Volum*: 158-170.

Wang, W., Kitova, E. N., Klassen, J. S. Nonspecific protein− carbohydrate complexes produced by nanoelectrospray ionization. Factors influencing their formation and stability. *Analytical Chemistry,* 2005;*Volum*: 3060-3071.

Wegener. Human norovirus GII.4(MI001) P dimer binds fucosylated and sialylated carbohydrates. *Glycobiology,* 2017;*Volum*.

Wegener, Mallagaray, Á., Schöne, T., Peters, T., Lockhauserbäumer, J., Yan, H., Uetrecht, C., Hansman, G. S., Taube, S. Human norovirus GII. 4 (MI001) P dimer binds fucosylated and sialylated carbohydrates. *Glycobiology,* 2017;*Volum*: 1027-1037.

Wei, H., Mo, J., Tao, L., Russell, R. J., Tymiak, A. A., Chen, G., Iacob, R. E., Engen, J. R. Hydrogen/deuterium exchange mass spectrometry for probing higher order structure of protein therapeutics: methodology and applications. *Drug discovery today,* 2014;*Volum*: 95-102.

Weis, D. D. (2016). *Hydrogen exchange mass spectrometry of proteins: fundamentals, methods, and applications*: John Wiley & Sons.

Weis, D. D., Wales, T. E., Engen, J. R., Hotchko, M., Ten Eyck, L. F. Identification and characterization of EX1 kinetics in H/D exchange mass spectrometry by peak width analysis. *Journal of the American Society for Mass Spectrometry,* 2006;*Volum*: 1498-1509.

Williams, A. N., Sherman, M. B., Smith, H. Q., Taube, S., Pettitt, B. M., Wobus, C. E., Smith, T. J. A norovirus uses bile salts to escape antibody recognition while enhancing receptor binding. *Journal of Virology,* 2021;*Volum*: JVI. 00176-00121.

Wilm, M., Mann, M. Analytical properties of the nanoelectrospray ion source. *Analytical Chemistry,* 1996;*Volum*: 1-8.

Winston, R. L., Fitzgerald, M. C. Mass spectrometry as a readout of protein structure and function. *Mass Spectrometry Reviews,* 1997;*Volum*: 165-179.

Wojcik, F., Mosca, S., Hartmann, L. Solid-phase synthesis of asymmetrically branched sequence-defined poly/oligo (amidoamines). *The Journal of organic chemistry,* 2012;*Volum*: 4226-4234.

Wu, Y., Engen, J. R., Hobbins, W. B. Ultra performance liquid chromatography (UPLC) further improves hydrogen/deuterium exchange mass spectrometry. *Journal of the American Society for Mass Spectrometry,* 2006;*Volum*: 163-167.

Yan, H., Lockhauserbäumer, J., Szekeres, G. P., Mallagaray, A., Creutznacher, R., Taube, S., Peters, T., Pagel, K., Uetrecht, C. Protein secondary structure affects glycan clustering in native mass spectrometry. *Life,* 2021;*Volum*: 554.

Zakhour, M., Ruvoën-Clouet, N., Charpilienne, A., Langpap, B., Poncet, D., Peters, T., Bovin, N., Le Pendu, J. The αGal epitope of the histo-blood group antigen family is a ligand for bovine norovirus Newbury2 expected to prevent cross-species transmission. *PLoS Pathogens,* 2009;*Volum*: e1000504.

Zhang, X.-F., Huang, Q., Long, Y., Jiang, X., Zhang, T., Tan, M., Zhang, Q.-L., Huang, Z.-Y., Li, Y.-H., Ding, Y.-Q. An outbreak caused by GII. 17 norovirus with a wide spectrum of HBGA-associated susceptibility. *Scientific Reports,* 2015;*Volum*: 17687.

S8 Acknowledgement

I would to thank the people who encourage me for the native MS study during PhD program and thesis preparation.

Foremost, I would like to thank my supervisor Dr. Charlotte Uetrecht on the supervision of my thesis and the whole PhD program. As the first an international PhD student in the group, I got a detailed guidance on the PhD study and life in Germany. Lucky I am, her patience and immense knowledge motivated me to fulfil the goal on the research work. Although I have entered into industrial area, the skill and knowledge I learned from my supervisor and her group assist me to face my current work on drug discovery platform. I truly appreciated the research projects and leadership skill training in my PhD program.

Here, I would like to thank Prof. Dr.Julia Kehr as the co-supervisor for my PhD program.

As a recipient of Hamburg University international doctoral student scholarship, I want to thank to the scholarship committee to give me the chance to study abroad.

During the preparation of the thesis, I worked with many people from different groups, I learned new skills and be influenced by them personally. It was great pleasure to work with our group members, Julia and Jasmin and Ronja. We worked together for the Virocarb project. Julia and I figured out the glycan binding pattern on different norovirus strains. Ronja's work inspired us to extend the investigation of blood group antigens to some other groups of carbohydrate structures. Jasmin taught me the methodology of HDX, it instructed me to explore the new filed in MS.

I also got valuable input from the persons at external scientific groups. Henrik Wegener from Prof. Dr. Stefan Taube group generated important protein to support our project and my thesis. Dr. Alvaro Mallagaray and Robert Creutznacher from Prof. Thomas Peters group also provide me the norovirus P dimer for the native MS analysis; I also learned the basic theory about NMR.

In the chapter 4, Dr. Katharina Susanne Bücher designed the structure of the glycomacromolecules. Without these 'compounds', we cannot investigate whether this structure can compete with the natural glycans or not. She also taught me a bit skill on organic chemistry synthesis; this is quite interesting for me, a Rookie in this area.

I would like to also thank to Prof. Wouter Roos and his group members, the stay at Groningen University extended my outlook on understand the biophysical assays. It helped me to dig into my own project to study the ligand-protein interaction.

Doctoral program is interesting and also tough, therefore, you especially need some colleagues. They will encourage and stay with you to finally go to the finish line in this part of Journey. Three years in the Panda cave trained my body to tolerate high temperature in summer. The room is quite hot compared to the outside, although this might be not useful in the rest of life. Thanks to the people at 'cousin' room, Julia, Ronja, Johannes and especially Boris. Thanks Bobo, your attitude towards Science cheered me. Especially thanks for all the tips you told to me for living in Harburg. This is great treasure for an international student easily adapting himself in the new environment and culture. Because it is tricky, I would not put into my thesis. Wait, I would like to thank 'Piggie' cave people, Jasmin, Lara, Janine, Alan and Jürgen. Piggie cave is always one of the cosy offices in the building. Thanks for the AG77 group members, Knut, Yinfei, Jürgen Efe, Khon, Marc. I am enjoyable to work in this group.

I would like to thank to my friends, Chris and Frank and Ina. Although my German is still poor, but at least I had some improvement in English. Thank Ina, the dinners you cooked is always delicious. Thanks Frank, Chris,

At the end, I would like to my parents and girlfriend, you always stands up in my side and cheer me up to encounter the difficulty. I am sorry, mom and dad, in the past five years, I did not often come back home. Yuemin, you are the best listener, thank you to let me walk into your journey of life.

Declaration of Authorship

I declare here that I have written the dissertation by my own, the content of this thesis has not been presented previously at any other examination board. The submitted file is related to the document on the storage medium.

17.02.2022 Hao Yan