

Article

Protein Secondary Structure Affects Glycan Clustering in Native Mass Spectrometry

Hao Yan ^{1,†}, Julia Lockhauserbäumer ^{1,†}, Gergo Peter Szekeres ^{2,3} , Alvaro Mallagaray ⁴ , Robert Creutzmacher ⁴, Stefan Taube ⁵ , Thomas Peters ⁴, Kevin Pagel ^{2,3} and Charlotte Uetrecht ^{1,6,7,*} 

- ¹ Leibniz Institute for Experimental Virology (HPI), 20251 Hamburg, Germany; hao.yan@leibniz-hpi.de (H.Y.); julia.lockhauserbaeumer@leibniz-hpi.de (J.L.)
- ² Organic Chemistry, Free University Berlin, 14195 Berlin, Germany; gergo.peter.szekeres@fhi-berlin.mpg.de (G.P.S.); kevin.pagel@fu-berlin.de (K.P.)
- ³ Fritz-Haber-Institut der Max-Planck-Gesellschaft, 14195 Berlin, Germany
- ⁴ Center of Structural and Cell Biology in Medicine (CSCM), Institute of Chemistry and Metabolomics, University of Lübeck, 23562 Lübeck, Germany; alvaro.mallagaraydebenito@uni-luebeck.de (A.M.); r.creutzmacher@uni-luebeck.de (R.C.); thomas.peters@chemie.uni-luebeck.de (T.P.)
- ⁵ Institute of Virology, University of Lübeck, 23562 Lübeck, Germany; stefan.taube@uni-luebeck.de
- ⁶ European XFEL GmbH, 22869 Schenefeld, Germany
- ⁷ Centre for Structural Systems Biology (CSSB), 22607 Hamburg, Germany
- * Correspondence: charlotte.uetrecht@xfel.eu; Tel.: +49-40-48051-261; Fax: +49-40-48051-252
- † Authors contributed equally to this work.



Citation: Yan, H.; Lockhauserbäumer, J.; Szekeres, G.P.; Mallagaray, A.; Creutzmacher, R.; Taube, S.; Peters, T.; Pagel, K.; Uetrecht, C. Protein Secondary Structure Affects Glycan Clustering in Native Mass Spectrometry. *Life* **2021**, *11*, 554. <https://doi.org/10.3390/life11060554>

Academic Editors: Adam Taylor and Marko Noerenberg

Received: 2 May 2021
Accepted: 8 June 2021
Published: 11 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Infection by the human noroviruses (hNoV), for the vast majority of strains, requires attachment of the viral capsid to histo blood group antigens (HBGAs). The HBGA-binding pocket is formed by dimers of the protruding domain (P dimers) of the capsid protein VP1. Several studies have focused on HBGA binding to P dimers, reporting binding affinities and stoichiometries. However, nuclear magnetic resonance spectroscopy (NMR) and native mass spectrometry (MS) analyses yielded incongruent dissociation constants (K_D) for the binding of HBGAs to P dimers and, in some cases, disagreed on whether glycans bind at all. We hypothesized that glycan clustering during electrospray ionization in native MS critically depends on the physicochemical properties of the protein studied. It follows that the choice of a reference protein is crucial. We analysed carbohydrate clustering using various P dimers and eight non-glycan binding proteins serving as possible references. Data from native and ion mobility MS indicate that the mass fraction of β -sheets has a strong influence on the degree of glycan clustering. Therefore, the determination of specific glycan binding affinities from native MS must be interpreted cautiously.

Keywords: ion mobility; native MS; electrospray ionization; norovirus capsid protein; carbohydrate binding; HBGA

1. Introduction

Human norovirus (hNoV) infection is the most common cause of acute gastroenteritis, leading to an estimated 685 million cases annually worldwide. The elderly, immunocompromised patients and children under 5 years are the most severely affected. hNoV belongs to the family of *Caliciviridae*, non-enveloped viruses of icosahedral shape with the viral genome consisting of positive-sense single-stranded RNA. Histo blood group antigens (HBGAs) serve as attachment factors in viral infection [1,2]. Previous nuclear magnetic resonance spectroscopy (NMR), X-ray crystal structure and mass spectrometry (MS) investigations identified the L-fucose moieties within HBGAs as a minimal binding motif for hNoV attachment. Most studies 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 [3,4]. Because the determination of dissociation constants and binding stoichiometries is key

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 [5], 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 [6–8]. 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 [4,9–12]. Notably, results disagreed [4,11], raising the question of whether the selection of a reference protein influences data interpretation. Moreover, with the exception of glycan mimetics [13], 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_D s for binding of several sialic acid containing carbohydrates hNoV P dimers [9,11,12,14], 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) [3,9,12]. This questions the validity of the results from direct MS measurements.

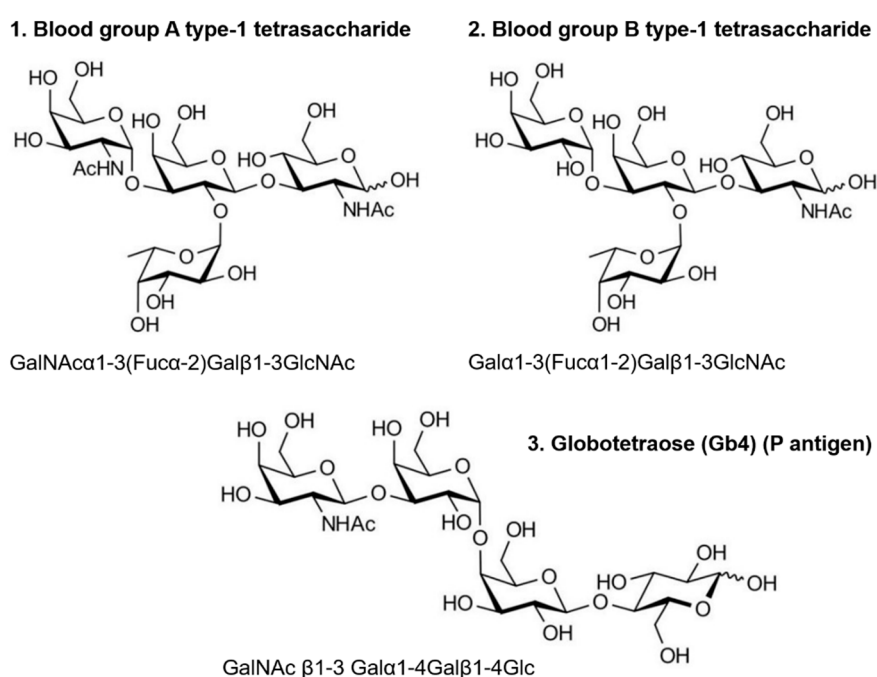
Additionally, the reported K_D values from carbohydrate-binding studies to P dimers are not comparable in STD NMR, native MS and isothermal titration calorimetry (ITC) [11,12,15]. Importantly, K_D s obtained for active pharmaceutical agents based on different biophysical assays such as native MS, ITC, surface plasmon resonance (SPR) and circular dichroism (CD) were equivalent in numerous other cases [8,15,16]. However, these compounds showed higher binding affinities (μ M 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 μ M to mM range. This also holds for hNoV–carbohydrate interactions [9–12,14,15]. In former studies, the origin of the discrepancies between NMR and native MS data for P dimer–glycan interaction was not in focus [12]. 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 [4,13].

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 [18]. 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.

2. Materials and Methods

2.1. Glycans

The following glycans were purchased from Elicityl-Oligotech, dissolved in H₂O for native MS analysis and are shown in Scheme 1. Blood group antigens: (1.) A tetrasaccharide type 1 (>90% NMR) (GalNAc α 1-3(Fuc α -2)Gal β 1-3GlcNAc, product code: GLY035-1-90%), (2.) B tetrasaccharide type 1 (>90% NMR) (Gal α 1-3(Fuc α 1-2)Gal β 1-3GlcNAc, product code: GLY038-1-90%). Globo-series: (3.) P antigen Gb4 (>90% NMR) (GalNAc β 1-3 Gal α 1-4Gal β 1-4Glc, product code: GLY121-90%). The HBGA ligands were chosen as known binders based on their fucose binding moiety, whereas Gb4 was the fucose-free non-glycan-binding reference. HBGAs comprise various oligosaccharides, the tetrasaccharides were chosen here instead of e.g., trisaccharides to have a larger mass increment upon association with the P dimer to allow for gentler conditions in the native MS measurements.



Scheme 1. Structure of used carbohydrates: Blood group A and B type-1 tetrasaccharides (HBGAs) are known to bind to hNoV P dimers; Gb4 is considered a non-binder.

2.2. Proteins

P dimers of hNoV VP1 were used as target proteins. Native MS investigations were performed with P dimers from the hNoV strains GII.4 Saga (Saga 2006 (GenBank ID: AB447457, aa 225–530), GII.4 MI001 (KC631814, aa 225–530) and murine noroviruses (MNV-1, CW1: DQ285629, aa 228–530). *Escherichia coli* was used for overexpression of the P domains as described in [4,18]. Purified P dimers were stored at 4 °C before preparation for native MS experiments. The commercially available reference candidates cytochrome c from equine heart (cyt c, CAS number 9007-43-6), ubiquitin from bovine erythrocytes (Ubq, CAS number 75986-22-4), carbonic anhydrase isozyme II from bovine erythrocytes (CA, CAS number 9001-03-0), alcohol dehydrogenase from *Saccharomyces cerevisiae* (ADH, CAS number 9031-72-5), L-lactic dehydrogenase from rabbit muscle (LDH, CAS number 9001-60-9), myoglobin from equine heart (Myo, CAS number 100684-32-0), human apo-transferrin (apo-TFF, CAS number 11096-37-0) were purchased from Sigma-Aldrich (Merck, Germany) and stored according to manufacturer recommendation. Furthermore, a superfolder green fluorescent protein (GFP) was kindly provided by Henning Tidow (University of Hamburg, Hamburg, Germany) and used as a reference protein candidate [19].

2.3. Native Mass Spectrometry

Prior to MS analysis, purified P dimers were buffer exchanged to 150 mM ammonium acetate (Sigma-Aldrich, reagent grade $\geq 98\%$) at pH 7 via centrifugal filter units at $12,000\times g$ at $4\text{ }^{\circ}\text{C}$ (Vivaspin 500, MWCO 10000, Sartorius). Glycans for native MS analysis were mixed with $1\text{ }\mu\text{M}$ (per monomer) purified GII.4 Saga P dimer and $3\text{ }\mu\text{M}$ of the reference proteins at indicated concentrations. Native mass spectra were measured at room temperature in positive ion mode on a high mass modified LCT Premier mass spectrometer (Waters/Micromass, UK and MS Vision, the Netherlands) with a nano-ESI source [20]. In house-made gold-coated electrospray capillaries were used for direct-sample infusion without any accessory chromatographic separation. The voltages and pressures were optimized for non-covalent protein complexes. The gas pressure used in the source hexapole was set between 6.5 and 6.8×10^{-2} mbar argon optimized for minimal complex dissociation and the backing pressure of the source roughing pump was set between 7.0 and 8.5 mbar. Spectra were recorded with applied voltages for the capillary and cone of 1.20 kV and 240 V, respectively. Note that charge state distributions for P dimers were sometimes slightly shifted as was the degree of glycan clustering per charge state. These changes were highly variable and not quantifiable. We attribute them to spray variations and hence slight variation in activation.

For calibration of the raw data, a 25 mg/mL cesium iodide spectrum from the same day was used. MassLynx V4.1 (Waters, Manchester, UK) was used to assign peak series to protein species and to determine the mass after minimal smoothing. Furthermore, Origin-Pro 2016 SR2 (OriginLab, Northampton, MA, USA) was used for spectral deconvolution and curve fitting. Correction for nonspecific protein-ligand clustering was performed as described [5]. Corrected peak areas were averaged and normalized to the free-protein signal. Data were based on at least three independent measurements. Python code Panda and Numpy package were used to calculate the specific binding of P dimers according to the glycan clustering ratio obtained from the selected reference protein candidates. Seaborn and matplotlib.pyplot package were imported into a python script to plot the results. The module of sklearn.linear model was applied for linear regression. The matplotlib and heatmap package were used to plot the heatmap.

2.4. Determination of Glycan Binding: Calculation of the Dissociation Constant (K_D)

Data from native ESI MS measurements were translated into dissociation constants (K_D values). The detailed calculations were based on the equations listed below. A protein (P) and a glycan ligand (L) form a complex through non-covalent interactions leading to a reversible association described by (Equation (1)):



when equilibrium is reached, the K_D can be calculated directly from the concentrations via the law of mass action (Equation (2)):

$$K_D = [P_{free}][L_{free}] / [PL] \quad (2)$$

Theoretically, the K_D is calculated by measuring the concentration of the three components (P_{free} , L_{free} , PL) in solution. Usually, these values are only indirectly accessible because they require ligand titration up to binding saturation. However, native MS allows for the directly measurement of P_{free} and PL , from which L_{free} can be calculated using the input concentrations. In the present work, the non-covalent interaction was measured in the gas phase assuming identical ESI and detection efficiency for free and glycan-associated proteins, both for the analyte and the reference protein. The peak areas (A) of the non-bound protein ion (P^{n+}) and the bound protein ion (PL^{n+}) were used to calculate the ratio (R) as these most accurately reflected the total signal and hence were assumed to be comparable to the ratio of a ligand-bound protein complex and an unbound protein

in solution (e.g., equilibrium state). For P dimers, the equations reflect glycan binding to the individual monomers [4] assuming that binding sites on each monomer are equal and independent [21].

Using R described in Equation (3), the K_D calculation can be re-written into Equation (4) with the known initial concentration of glycan ligand $[L_0]$ and protein $[P_0]$, adapted from Sun et al. [5]:

$$R = PL_{equil} / P_{equil} = \sum_n A(PL^{n+}) / n / \sum_n A(P^{n+}) / n \quad (3)$$

$$K_D = \frac{L_0 - \frac{RP_0}{1+R}}{R} \quad (4)$$

2.5. Titration Measurement of Glycan Binding on P Dimers

For the carbohydrate-binding experiments, it was essential to keep the initial protein concentrations and pH value constant during the analysis of the protein-carbohydrate solutions. While acidification occurs in any ESI droplet, the remaining solution in the capillary changes pH over time. Therefore, accumulated spectra were obtained within 10 min after the voltages were applied to minimize the influence of pH changes during the measurements. The LCT mass spectrometer with only a short hexapole prior to the ToF analyser reduces potential activation and hence in-source dissociation (ISD). Three different glycan ligands including ligands of interest (HBGA B and A) and non-binding control (Gb4) were measured in a range of concentrations to determine the K_D of the first glycan binding event to P dimers (Table 1).

2.6. Ion Mobility Mass Spectrometry (IMMS)

IMMS experiments on the GII.4 Saga P dimers were performed in a Synapt G2-S instrument modified with a linear drift-tube ion-mobility cell operating with He drift gas under 2.4 mbar pressure. Samples with 3 μ M P dimer protein and 0, 100, 200, and 500 μ M of the HBGA B tetrasaccharide or Gb4, respectively, were prepared, thoroughly homogenized, then loaded onto a Pd/Pt-coated borosilicate glass capillary. The sample was introduced by nano-ESI with the following instrumental parameters: capillary voltage, 1200 V; cone voltage, 15 V; source offset, 15 V; source gas flow, 0.0 mL/min; backing pressure, 0.1 bar; source temperature, 30 °C. These settings ensured soft ionization conditions that allowed the preservation of the native protein structure. The arrival time distribution curves shown later on in Figure 5 were recorded at a He cell DC of 35 V and a bias of 55 V. Each experiment was performed at least twice to address reproducibility.

3. Results

3.1. 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 [4], 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 [9,10]. 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 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_D s for the wildtype determined by NMR (12 mM for B tetrasaccharide [18]), suggesting an intrinsic issue with the measurement approach. Additionally, the non-binding glycan Gb4 shows a

pattern similar to HBGA B after correction (Figure 1B), but with more glycan attached to the wildtype protein.

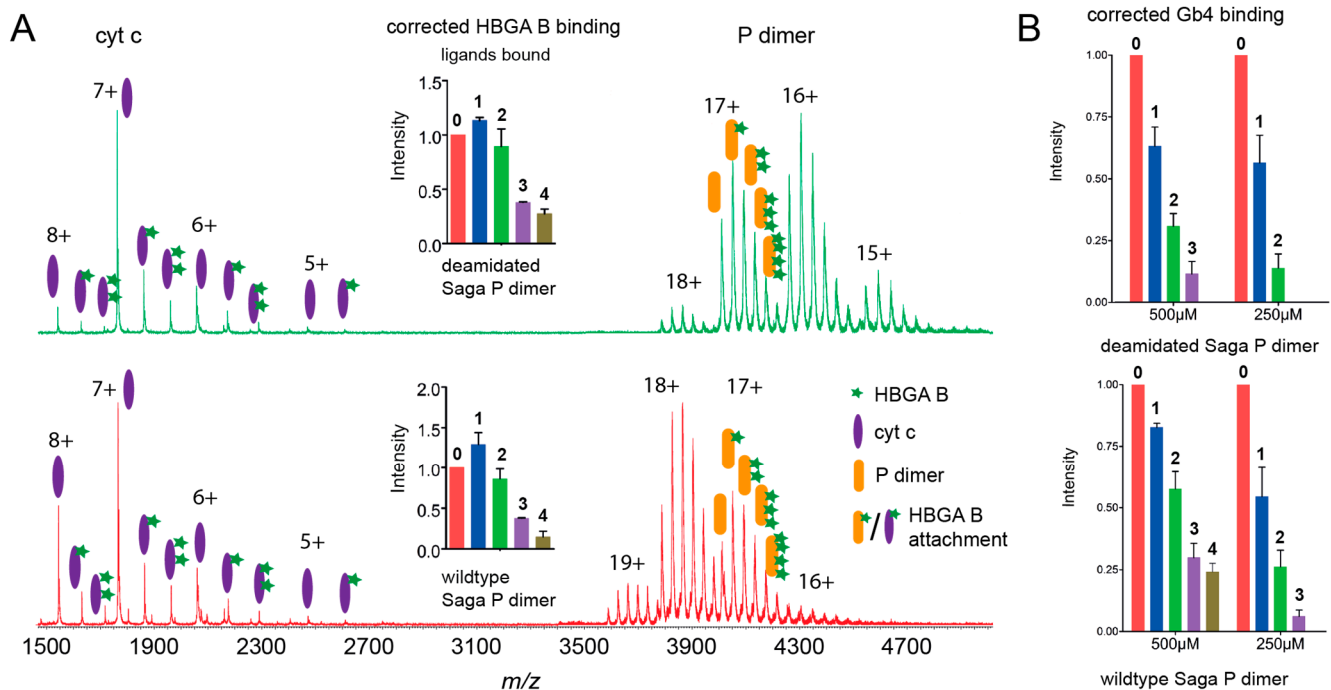


Figure 1. Influence of deamidation on glycan clustering. (A) Native mass spectra of cytochrome c (cyt c) with wildtype (red, bottom) or deamidated (green, top) GII.4 Saga P dimer and 500 μ M HBGA B ligand in 150 mM ammonium acetate solution at pH 7. Signal intensity was normalized to the base peak in 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 μ M and 250 μ M Gb4.

Protein NMR experiments recently demonstrated that MNV P dimers do not bind to HBGAs [18]. Moreover, MNV shows altered dimerization properties and presents a large fraction of P monomers at neutral pH. While correction is not possible due to the insufficient quality of the reference protein signal (see Figure S1), it is evident that similar glycan clustering is observed for the P monomer and dimer. Glycan binding requires a dimeric protein; therefore, these interactions between MNV P monomers [22] and glycans have to be unspecific in line with the study by Creutzmacher et al. [18]. 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 behavior; hence, the P monomer and dimer signals. The corrected data shows equal clustering behavior for HBGA B, A and 3'-sialyllactose (GM3) on the monomer and dimer (Figure S2). 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.

3.2. Glycan Clustering to Various 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 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 2). 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.

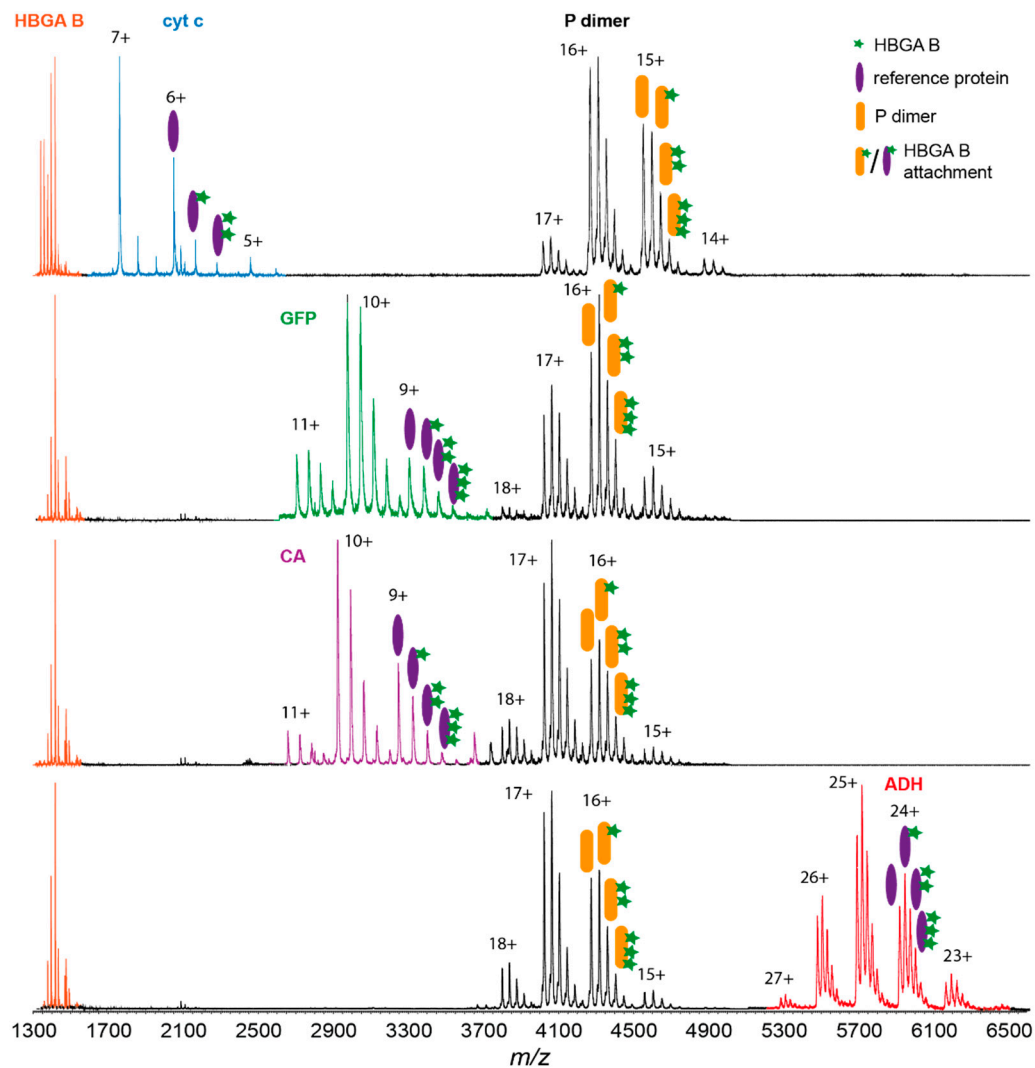


Figure 2. Native mass spectra of four selected reference proteins. 3 μM reference protein and 1 μM wildtype Saga P dimer (all protein concentrations based on the monomer) are measured with 500 μM HBGA B ligand in 150 mM ammonium acetate solution at pH 7. From top to bottom: cyt *c* (blue), GFP (green), carbonic anhydrase II (CA, purple), ADH (red). Signal intensity was normalized to the protein base peak. P dimer: 3800–5000 m/z , ADH: 5200–6500 m/z , GFP: 2700–3500 m/z , CA: 2600–3700 m/z .

Table 1. Glycan clustering ratio R and K_D of HBGA B interaction with P dimers analysed by native MS. Data correction is based on the reference protein method. n.a.—not available.

Ref. Protein	Mass /kDa	PDB ID	Secondary Structure Composition		Glycan-Clustering Ratio R at 500 μ M			K_D for HBGA B /mM
			% β -sheet	% α -helix	HBGA B	Gb4	HBGA A	
Ubq	8.9	1UBQ	24.7	20.8	0.25 ± 0.00	0.09 ± 0.03	0.10 ± 0.05	0.20 ± 0.13
Cyt c	13.2	2N3B	0	53.8	0.35 ± 0.03	0.45 ± 0.02	0.41 ± 0.09	0.16 ± 0.05
Myo	17.0	1AZI	0	79.2	0.36 ± 0.05	0.37 ± 0.06	0.34 ± 0.03	0.25 ± 0.10
GFP	26.8	1BFP	55.0	18.5	0.67 ± 0.02	0.80 ± 0.03	0.86 ± 0.04	0.90 ± 0.35
CA	29.1	1V9E	35.8	15.8	0.75 ± 0.07	0.75 ± 0.03	0.74 ± 0.07	0.40 ± 0.10
apo-TFF	79.6	2HAU	15.7	34.5	0.82 ± 0.03	0.80 ± 0.05	0.79 ± 0.03	n.a.
LDH	146.8	3H3F	17.1	46.9	1.02 ± 0.04	0.91 ± 0.06	n.a.	n.a.
ADH	147.0	4W6Z	34.8	28.4	1.45 ± 0.11	1.26 ± 0.03	1.32 ± 0.10	24.00 ± 16.00
SAGA P dimers	68.1	4X7C	32.5	6.6	1.55 ± 0.05	1.40 ± 0.21	1.44 ± 0.15	n.a.

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 μ M 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] but is also barely detectable (Figure S3). Differences in R and K_D are also consistent when titrating five glycan concentrations (Table S1). When the clustering ratio R is plotted over the glycan concentration (Figure 3), 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 [5,11]. 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, grey line). Therefore, Ubq is excluded from the following analysis.

3.3. Biophysical Properties Influence the 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 4, Figures S4–S6). 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 tryptophans 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 4), whereby 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.

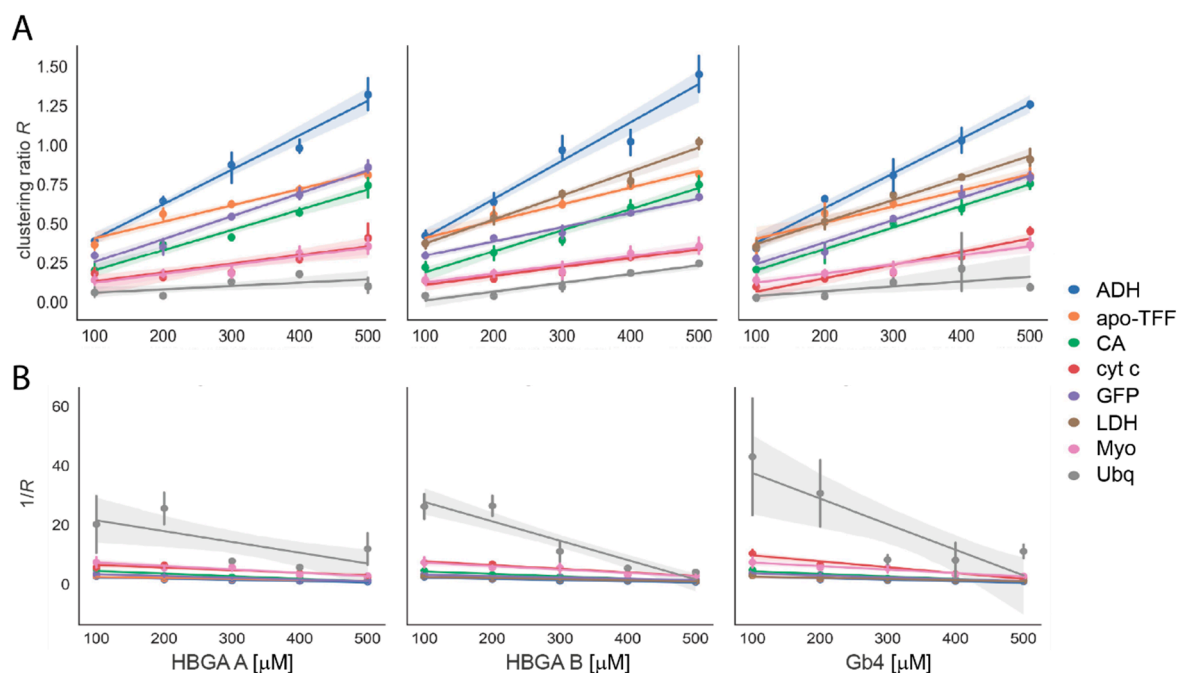


Figure 3. Correlating glycan concentrations to glycan clustering ratios. (A) R value calculated based on the peak area of the reference proteins (ADH; CA, cyt c, GFP; Myo, Ubq, LDH, apo-TFF). (B) Corresponding reciprocal value. Titration experiments were performed with P dimers (1 μ M, monomer) and reference proteins (3 μ M, monomer) in 150 mM ammonium acetate solution pH 7 with ligand concentrations ranging from 100 μ M to 500 μ M for HBGA A, B and Gb4. The shaded areas represent the standard error of the slope for the linear fit.

3.4. Ion Mobility on P Dimer Glycan Interactions

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 [9]. 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 [4]. 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 [9]. 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 5, the arrival-time distributions are shown for the wildtype GII.4 Saga P dimer together with its complexes with HBGA B (Figure 5A) and Gb4 (Figure 5B) for all detectable stoichiometries, as well as the respective single complexes of 1 P dimer (0–3 HBGA B (Figure 5C)) and 1 P dimer: (0–3 Gb4 1 molecules (Figure 5D)). 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 5, 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 $3928 \pm 14 \text{ \AA}^2$, while in B and D, the CCS is $3867 \pm 12 \text{ \AA}^2$. The difference between these two CCS values is $\sim 1.6\%$, 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 stoichiometries suggest 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 5A,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 5C,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.

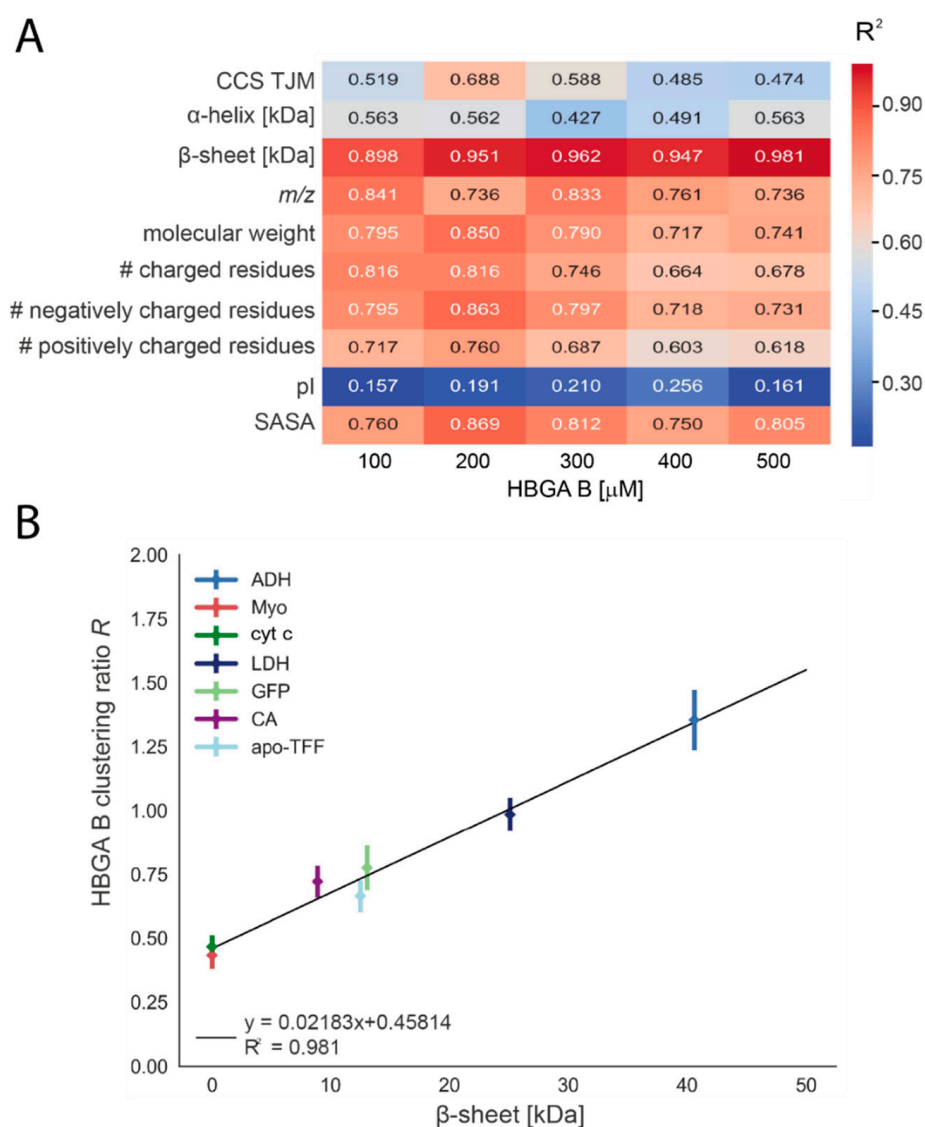


Figure 4. Correlation of clustering ratios R to different protein properties. (A) Heatmap of the R^2 values obtained from the linear regression at the indicated HBGA B glycan concentration for the listed protein properties. (B) The correlation of β -sheet amount in kDa to unspecific glycan clustering 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.

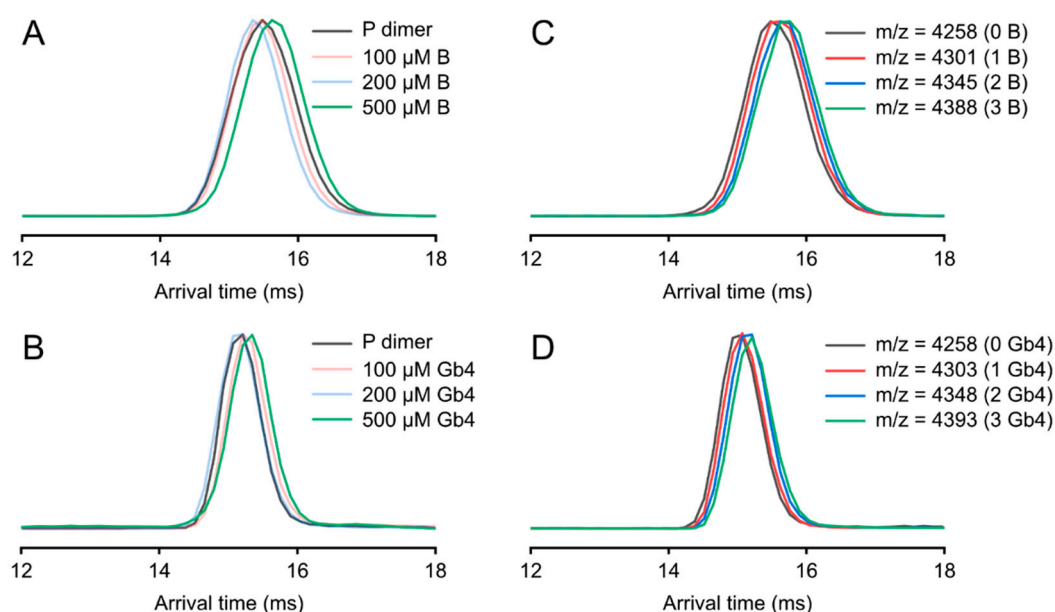


Figure 5. Glycan clustering to wildtype GII.4 P dimers in IMMS. Changes in the arrival time distribution upon increasing HBGA B (A) and Gb4 (B) concentration are depicted for the 16+ ion. Next to these are the extracted arrival-time distributions for free P dimer and P dimer plus 1–3 ligands from a single measurement with HBGA B (C) and Gb4 (D) revealing an increasing trend with increasing number of ligands.

4. Discussion

Glycan clustering was previously observed in native MS experiments [5,9–11]. 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 [12–14,23] 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 [18]. The results indicate that these problems are not limited to neutral glycans but also occur 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) [24], 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 ISD or varying ESI efficiencies. In contrast to Han et al., 2013 and 2018 [9,11], these results indicate an unexpectedly different carbohydrate-clustering propensity, resulting in K_D s 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 [5,9]. 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 occurs during ionization when glycans attach to the protein surface. Furthermore, β -sheets are more labile during the ESI process [25], 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 K_D to other methods was observed [5]. It has also proven invaluable for other ligand types [8,23,26]. 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 [4]. This contrasts with a previous report [11] where the experimental and data acquisition parameters differed. Conformational changes were reported for glycan charge states 17+ and 18+ up to 800 μ M, and we have also checked HBGA B and the 17+ charge state at 800 μ M 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 overactivation 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 overactivation 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 K_D discrepancies in the literature. We had previously used cyt c, both small and mostly α -helical, which overestimated the binding affinity [11,12]. Others used a small-sized monoclonal antibody single-chain fragment (scFv 26 kDa, [5,9,11]), 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 [4,27].

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 [13]. 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 μ M concentration (compare to the stoichiometry information (Table S2 and Figure S5)). 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) [4,18]. Hence, an appropriate reference candidate is crucial for the K_D determination for low-affinity glycans.

5. Conclusions

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. Therefore, 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 [28].

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/life11060554/s1>, Figure S1: glycan clustering on murine norovirus MNV-1 (CW1) P domain; Figure S2: glycan clustering on mutated hNoV GII.4 MI001 strain P domain; Figure S3: Correlation of clustering ratios to multiple protein property parameters; Figure S4: Correlation plotted as R^2 of clustering ratios R to different protein properties over employed HBGA A concentration; Figure S5: Specific binding of different carbohydrates on the P dimer after correction of the unspecific glycan clustering; Figure S6: The relation between the amount of beta-sheet in the reference protein and the unspecific glycan clustering ratio R at different ligand concentrations (400 μ M (A), 300 μ M (B), 200 μ M (C) and 100 μ M (D)). Table S1: Glycan clustering ratio of HBGA B interaction with P dimer analysed with native MS. Data correction is based on the reference protein method; Table S2: Bound number of carbohydrates on P dimer analysed with native MS; Table S3: Protein physicochemical properties used for plots.

Author Contributions: Conceptualization, C.U.; methodology, C.U., H.Y., J.L., G.P.S. and K.P.; software, H.Y.; investigation, H.Y., J.L., G.P.S.; resources, C.U., K.P., T.P., R.C., A.M. and S.T.; data curation, H.Y., J.L., G.P.S.; writing—C.U., J.L., H.Y., G.P.S.; writing—review and editing, all; supervision, C.U. and K.P.; project administration, C.U.; funding acquisition, C.U., T.P. and K.P. All authors have read and agreed to the published version of the manuscript.

Funding: H.Y. is recipient of Hamburg University international doctoral student scholarship; J.L. was funded by the DELIGRAH graduate school funded by the federal state of Hamburg; and C.U. is supported by the Leibniz Association through SAW-2014-HPI-4 grant and acknowledge funding from FOR2327 ViroCarb (UE 183/1-2). T.P. and S.T. also acknowledges funding within ViroCarb FOR2327 (DFG Pe494/12-2). K.P. and G.P.S. are grateful for the funding through the EU Horizon 2020 under grant number 899687 (HS-SEQ). The Leibniz Institute for Experimental Virology (HPI) is supported by the Free and Hanseatic City of Hamburg and the Federal Ministry of Health.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data is available from the corresponding author upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cao, S.; Lou, Z.; Tan, M.; Chen, Y.; Liu, Y.; Zhang, Z.; Zhang, X.C.; Jiang, X.; Li, X.; Rao, Z. Structural Basis for the Recognition of Blood Group Trisaccharides by Norovirus. *J. Virol.* **2007**, *81*, 5949–5957. [[CrossRef](#)]
2. 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. *J. Infect. Dis.* **2003**, *188*, 19–31. [[CrossRef](#)]
3. 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. *Angew. Chem. Int. Ed.* **2012**, *124*, 952–956. [[CrossRef](#)]
4. Mallagaray, A.; Creutzmacher, 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. *Nat. Commun.* **2019**, *10*, 1–14. [[CrossRef](#)]
5. Sun, J.; Kitova, E.N.; Wang, W.; Klassen, J.S. Method for distinguishing specific from nonspecific protein–ligand complexes in nanoelectrospray ionization mass spectrometry. *Anal. Chem.* **2006**, *78*, 3010–3018. [[CrossRef](#)]

6. Pacholarz, K.J.; Garlish, R.A.; Taylor, R.J.; Barran, P.E. Mass spectrometry based tools to investigate protein-ligand interactions for drug discovery. *Chem. Soc. Rev.* **2012**, *41*, 4335–4355. [[CrossRef](#)]
7. Jecklin, M.C.; 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. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 332–343. [[CrossRef](#)]
8. 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. *J. Am. Chem. Soc.* **2018**, *140*, 7486–7497. [[CrossRef](#)] [[PubMed](#)]
9. 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**, *23*, 276–285. [[CrossRef](#)] [[PubMed](#)]
10. Han, L.; Tan, M.; Xia, M.; Kitova, E.N.; Jiang, X.; Klassen, J.S. Gangliosides are ligands for human noroviruses. *J. Am. Chem. Soc.* **2014**, *136*, 12631–12637. [[CrossRef](#)] [[PubMed](#)]
11. 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**, *28*, 488–498. [[CrossRef](#)]
12. Mallagaray, A.; Lockhauerbäumer, J.; Hansman, G.; Uetrecht, C.; Peters, T. Attachment of norovirus to histo blood group antigens: A cooperative multistep process. *Angew. Chem. Int. Ed.* **2015**, *54*, 12014–12019. [[CrossRef](#)] [[PubMed](#)]
13. Bücher, K.S.; Yan, H.; Creutzmacher, R.; Ruoff, K.; Mallagaray, A.; Grafmuller, A.; Dirks, J.S.; Kilic, T.; Weickert, S.; Rubailo, A.; et al. Fucose-Functionalized Precision Glycomacromolecules Targeting Human Norovirus Capsid Protein. *Biomacromolecules* **2018**, *19*, 3714–3724. [[CrossRef](#)]
14. Wegener, H.; Mallagaray, A.; Schone, T.; Peters, T.; Lockhauerbaumer, J.; Yan, H.; Uetrecht, C.; Hansman, G.S.; Taube, S. Human norovirus GII.4(MI001) P dimer binds fucosylated and sialylated carbohydrates. *Glycobiology* **2017**, *27*, 1027–1037. [[CrossRef](#)] [[PubMed](#)]
15. Han, L.; 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* **2015**, *25*, 170–180. [[CrossRef](#)] [[PubMed](#)]
16. Jecklin, M.C.; 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. *J. Mol. Recognit.* **2010**, *22*, 319–329. [[CrossRef](#)]
17. Deng, L.; Kitova, E.N.; Klassen, J.S. Dissociation Kinetics of the Streptavidin–Biotin Interaction Measured Using Direct Electrospray Ionization Mass Spectrometry Analysis. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 49–56. [[CrossRef](#)] [[PubMed](#)]
18. Creutzmacher, 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**, *13*, 416. [[CrossRef](#)]
19. Kesgin-Schaefer, S.; Heidemann, J.; Puchert, A.; Koelbel, K.; Yorke, B.A.; Huse, N.; Pearson, A.R.; Uetrecht, C.; Tidow, H. Crystal structure of a domain-swapped photoactivatable sfGFP variant provides evidence for GFP folding pathway. *FEBS J.* **2019**, *286*, 2329–2340. [[CrossRef](#)]
20. Van den Heuvel, R.H.; Heck, A.J. Native protein mass spectrometry: From intact oligomers to functional machineries. *Curr. Opin. Chem. Biol.* **2004**, *8*, 519–526. [[CrossRef](#)]
21. Creutzmacher, R.; Schulze, E.; Wallmann, G.; Peters, T.; Stein, M.; Mallagaray, A. Chemical-Shift Perturbations Reflect Bile Acid Binding to Norovirus Coat Protein: Recognition Comes in Different Flavors. *ChemBiochem* **2020**, *21*, 1007–1021. [[CrossRef](#)]
22. Creutzmacher, R.; Maaß, T.; Dülfer, J.; Feldmann, C.; Hartmann, V.; Knickmann, J.; Westermann, L.T.; Smith, T.J.; Uetrecht, C.; Mallagaray, A.; et al. Murine norovirus capsid plasticity–Glycochenodeoxycholic acid stabilizes P-domain dimers and triggers escape from antibody recognition. *bioRxiv* **2021**. [[CrossRef](#)]
23. Garcia-Alai, M.M.; Heidemann, J.; Skruzny, M.; Gieras, A.; Mertens, H.D.T.; Svergun, D.I.; Kaksonen, M.; Uetrecht, C.; Meijers, R. Epsin and Sla2 form assemblies through phospholipid interfaces. *Nat. Commun.* **2018**, *9*, 328. [[CrossRef](#)] [[PubMed](#)]
24. Aliyari, E.; Konermann, L. Formation of Gaseous Proteins via the Ion Evaporation Model (IEM) in Electrospray Mass Spectrometry. *Anal. Chem.* **2020**, *92*, 10807–10814. [[CrossRef](#)]
25. 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. *Angew. Chem. Int. Ed. Engl.* **2016**, *55*, 14173–14176. [[CrossRef](#)] [[PubMed](#)]
26. 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 Disord. Proteins* **2013**, *1*, e25068. [[CrossRef](#)] [[PubMed](#)]
27. Dülfer, J.; Yan, H.; Brodmerkel, M.N.; Creutzmacher, R.; Mallagaray, A.; Peters, T.; Coleman, C.; Marklund, E.G.; Uetrecht, C. Glycan-Induced Protein Dynamics in Human Norovirus P Dimers Depend on Virus Strain and Deamidation Status. *Molecules* **2021**, *26*, 2125. [[CrossRef](#)] [[PubMed](#)]
28. Baez 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. *Anal. Chem.* **2021**, *93*, 4231–4239. [[CrossRef](#)] [[PubMed](#)]