Glycan-induced protein dynamics in human norovirus P dimers depend on virus strain and deamidation status

Jasmin Dülfer[a], Hao Yan[a], Maxim N Brodmerkel[b], Robert Creutznacher[c], Alvaro Mallagaray[c], Thomas Peters[c], Carl Caleman[d,e], Erik G Marklund[b], Charlotte Uetrecht[a,f]

a. Heinrich Pette Institute, Leibniz Institute for Experimental Virology, Hamburg, Germany
b. Department of Chemistry – BMC, Uppsala University, Uppsala, Sweden
c. Institute of Chemistry and Metabolomics, University of Lübeck, Lübeck, Germany
d. Department of Physics and Astronomy, Uppsala University, Uppsala, Sweden
e. Center for Free-Electron Laser Science, Deutsches Elektronen-Synchrotron, Hamburg, Germany
f. European XFEL GmbH, Schenefeld, Germany

Correspondence: Charlotte Uetrecht, charlotte.uetrecht@xfel.eu

Keywords: glycan interaction, norovirus capsid protein VP1, protruding domain, HDX-MS, native MS, hydrogen/deuterium exchange mass spectrometry

Running title: Glycan-induced protein dynamics in norovirus P dimers

Abbreviations

A_{sas}	solvent accessible area
CSP	chemical shift perturbation
ESI	electrospray ionization
HBGA	histo-blood group antigen
HDX	hydrogen/deuterium exchange
iDiD	fully deamidated (2x isoD373) P dimer
iDN	half deamidated (isoD373, native N373) P dimer
K_d	dissociation constant
MD	molecular dynamics
NN	fully native (2x native N373) P dimer
P domain	protruding domain
S domain	shell domain
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
VLP	virus-like particle
VP1	major capsid protein

Abstract

Noroviruses are the major cause of gastroenteritis and re-emerge worldwide every year, with GII.4 currently being the most frequent human genotype. The norovirus capsid protein VP1 is essential for host immune response. The P domain mediates cell attachment via histo blood-group antigens (HBGAs) in a strain-dependent manner but how these glycan-interactions actually relate to cell entry remains unclear. Here, hydrogen/deuterium exchange mass spectrometry (HDX-MS) is used to investigate glycan-induced protein dynamics in P dimers of different strains, which exhibit high structural similarity but different prevalence in humans. While the almost identical strains GII.4 Saga and GII.4 MI001 share glycan-induced dynamics, the dynamics differ in the emerging GII.17 Kawasaki 308 and rare GII.10 Vietnam 026 strain. We also further examine structural effects of N373 deamidation upon glycan binding in partially deamidated GII.4 P dimers, which are likely present during infection. Such mixed species exhibit increased exposure to solvent in the P dimer upon glycan binding as opposed to pure wildtype. Furthermore, deamidated P dimers display increased flexibility and a monomeric population. Our results indicate that glycan binding induces strain-dependent structural dynamics, which are further altered by N373 deamidation, and hence hint at a role of deamidation in modulating cell attachment and entry in GII.4 strains.

Introduction

Noroviruses are non-enveloped single strand (+)sense RNA viruses of the *Caliciviridae* family that cause an estimated 20 % of gastroenteritis cases worldwide [1]. The virus possesses an icosahedral capsid, consisting of dimers of the major capsid protein VP1. A minor capsid protein VP2 is located inside the icosahedral shell. Based on its VP1 sequence, ten genogroups have been categorized, of which GI, II, IV, VIII and IX can infect humans [2]. Noroviruses of genogroup II (GII), especially genotype GII.4, dominated outbreaks in the last two decades [1]. With the emergence of new strains, e.g. GII.17 in Asia, it is an open question as to whether GII.4 will be displaced [3, 4] or resurgent [5].

VP1 is divided into the inner shell (S) domain and the outward-facing protruding (P) domain [6]. The P domain is further subdivided into P1 and P2 subdomains, with P2 being essential for host immune response and binding to histo blood-group antigens (HBGAs) for cell attachment in a strain-dependent manner [7]. The importance of interactions with HBGAs for host cell attachment has been shown in several studies [8], but how these interactions actually mediate cell entry remains unclear.

In order to shed light on glycan contribution to cell entry several biophysical techniques including nuclear magnetic resonance (NMR), X-ray crystallography, native mass spectrometry (native MS) and hydrogen-deuterium exchange mass spectrometry (HDX-MS) have been applied to characterize binding of P dimers and whole virus-like particles (VLPs) to HBGAs and other glycans [9-17]. These studies revealed that glycan preferences and binding affinities are strongly genotype- and strain-dependent. For instance, crystallization studies showed that two fucose binding pockets on the top of the P dimer are highly conserved among different strains [8, 18], while in the GII.10 Vietnam 026 strain two additional binding sites located in the P2 cleft are occupied at high fucose concentrations [12].

Recently, 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. 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 an Asp at the equivalent position [19]. HDX-MS measurements confirmed the loss of binding of deamidated P dimers to HBGA B trisaccharide and revealed increased flexibility in the P2 domain compared to the wildtype P dimer.

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 [20]. 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 is causing these different profiles.

Therefore, we set out to examine whether glycan binding or deamidation can induce distinct structural dynamics changes in P dimers, thereby modulating infectivity. 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. GII.4 MI001 infects humans and mice [21] 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. The structural aspects of glycan binding to fully deamidated GII.4 P dimers have been investigated before [19]. However, *in vivo,* large fractions of partially deamidated P dimers with potentially altered dynamics are likely to occur. Therefore, we also examined glycan binding to partially deamidated GII.4 Saga and GII.4 MI001 P dimers.

Our targets share a different amount of sequence identity with the GII.4 Saga P dimer, but are highly similar on the structural level (Figure 1), with largest differences in the loop regions of the P2 domain. The results reveal identical glycan binding behavior in GII.4 Saga and GII.4 MI001 strains but distinct glycan induced dynamics in GII.17 Kawasaki 308 and GII.10 Vietnam 026. Furthermore, all strains apart from GII.4 Saga form a second P domain species that is highly protected from HDX. In partially deamidated GII.4 P dimers, fucose binding leads to different structural dynamics than in pure wildtype or fully deamidated samples, hinting at a potential biological function. Moreover, molecular dynamics (MD) simulations with an aggregated simulation length of 14 μs are employed to dissect the origin of observed differences.

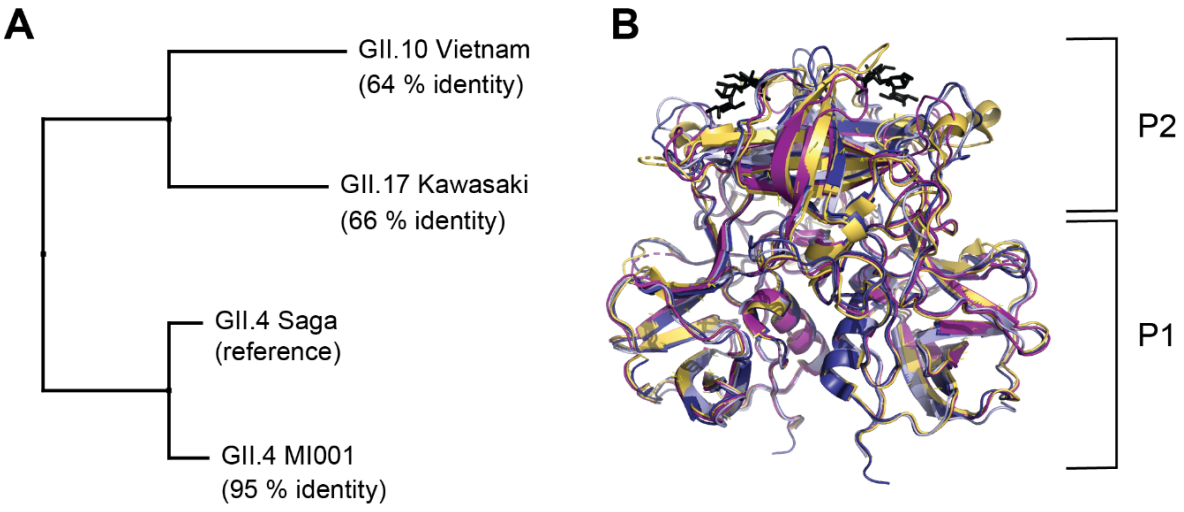


Figure 1: Comparison of human norovirus VP1 sequences (A) and P dimer structures with HBGA B trisaccharide (black) binding to the canonical binding site (B). P dimer sequences of three virus strains were aligned to the already investigated [19] GII.4 Saga strain as reference. For the GII.4 MI001 P dimer, a homology model was created for comparison using the GII.4 Farmington Hills P dimer structure with 94% sequence identity to MI001 as reference. Crystal structures of GII.4 Saga (pdb 4X06, dark blue), GII.10 Vietnam (pdb 3ONY, yellow), GII.17 Kawasaki (pdb 5F4O, purple) and the homology modelled structure of GII.4 MI001 (light blue) were superimposed in PyMOL with the following RMSDs to GII.4 Saga: 7.4 Å (GII.10 Vietnam), 6.5 Å (GII.17 Kawasaki), 1.6 Å (GII.4 MI001). P1 and P2 indicate the respective domains of the P dimer.

Material and Methods

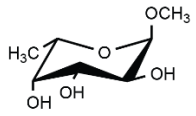
Expression and purification of P dimers

GII.4 Saga 2006 (VP1 residues 225–530), GII.4 MI001 (VP1 residues 225–530), GII.10 Vietnam 026 (VP1 residues 224–538), and GII.17 Kawasaki 308 (VP1 residues 225–530) P domains (see Figure S1 for VP1 sequence alignment), with GenBank accession numbers AB447457, KC631814, AF504671, and LC037415, respectively, were synthesized and purified as described elsewhere [19]. Briefly, *E. coli* BL21(DE3) were transformed with a pMal-c2x expression vector encoding the genes for ampicillin resistance, a fusion protein of maltose-binding protein, two His-tags, an HRV 3C cleavage domain, and the P domain. Due to the cloning strategy, the sequences from GII.4 Saga 2006 and GII.17 Kawasaki 308 2015 P domains contain an extra GPGS sequence preceding K225, whereas GII.10 Vietnam 026 contains a GPG sequence preceding S224. Transformed cells were grown for 3 h at 37 °C. Overexpression was induced with 1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) at an OD_{600} value of 1.5. Incubation was continued at 16 °C for 48 h. Cells were lysed using a high-pressure homogenizer (Thermo). The lysate was clarified by centrifugation, and the fusion protein was purified using a Ni-NTA resin (Qiagen). MBP and the His-tag were cleaved from the P domain using HRV 3C protease (Novagen). Cleaved P domain protein eluted from Ni-NTA resin and was further purified by size-exclusion chromatography using a Superdex 16/600 200 pg column (GE Healthcare) in 20 mM sodium phosphate buffer (pH 7.3). Protein purity and dimer concentration were monitored by SDS-PAGE and ultraviolet absorption. Separation of fully, partially, and non-deamidated (pure N373 wildtype) GII.4 P dimer species was achieved by cation exchange chromatography using a 6 ml Resource S column (GE Healthcare) at 6 °C. After separation protein samples were prepared in 20 mM sodium acetate buffer (pH 4.9) to prevent further spontaneous deamidation and eluted using a linear salt gradient.

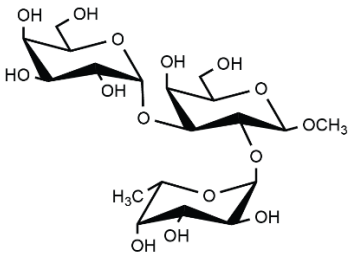
Wildtype P dimer samples were stored at 5 °C in the following buffers until analysis: GII.10 Vietnam: 25 mM TrisHCl, 300 mM NaCl, pH 7.3; GII.17 Kawasaki, GII.4 Saga and GII.4 MI001: 20 mM sodium acetate, 100 mM NaCl pH 4.9 (the last two pure wildtype N373). To create mixed species of wildtype (NN), partially deamidated (iDN) and fully deamidated (iDiD) GII.4 MI001 and GII.4 Saga P dimer [19], pure wildtype (NN) P dimer samples were stored at pH 7.3, which favors spontaneous deamidation, for several months. The storage conditions were 25 mM TrisHCl, 300 mM NaCl, pH 7.3, 4 °C for GII.4 MI001 P dimer and 75 mM sodium phosphate buffer, 100 mM NaCl, pH 7.3, 4 °C for GII.4 Saga P dimer.

Glycan structures

Methyl α-L-fucopyranoside (α-L-Fuc-(1,O)-CH_3) and HBGA B trisaccharide (α-D-Gal-(1,3)-[α-L-Fuc-(1,2)]-β-D-Gal-(1,O)-CH_3) were purchased from Carbosynth.



Methyl α-L-fucopyranoside



HBGA B trisaccharide

Native MS

Native MS measurements were performed using 3 to 4.5 µM purified P dimers. Proteins were subjected to buffer exchange to different concentrations of ammonium acetate (GII.10 Vietnam: 125 mM; GII.17 Kawasaki and GII.4 Saga: 300 mM; GII.4 MI001: 250 mM) at pH 7.5 and 4°C via centrifugal filter units (13000 x *g*, Vivaspin 500, MWCO 10000 (Sartorius) or Micro Bio-Spin 6

columns (Bio-Rad)) according to the manufacturers' protocols. Mass spectra were acquired at room temperature (25 °C) in positive ion mode on an LCT mass spectrometer modified for high mass (Waters, UK and MS Vision, the Netherlands) with a nano-electrospray ionization (ESI) source. Gold-coated electrospray capillaries were produced in house for direct sample infusion. Capillary and sample cone voltages were 1.20 kV to 1.35 kV and 150 to 240 V, respectively. The pusher was set to 100-150 µs. Pressures were 7 mbar in the source region and 6.2 x 10^{-2} to 6.5 x 10^{-2} mbar argon in the hexapole region. A spectrum of a 25 mg/ml cesium iodide solution from the same day was applied for calibration of raw data using the MassLynx software (Waters, UK). OriginPro 2016 (Origin Lab Corporation) software was used for peak integration and calculation of oligomer fractions.

HDX-MS

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 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 and injected onto a cooled (2 °C) HPLC System (Agilent Infinity 1260, Agilent Technologies) equipped with a home packed pepsin column (IDEX guard column with an internal volume of 60 µL, Porozyme immobilized pepsin beads, Thermo Scientific) in a column oven (25 °C), a peptide trap column (OPTI-TRAP for peptides, Optimize Technologies) and a reversed-phase analytical column (PLRP-S for Biomolecules, Agilent Technologies). 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. MS was performed using an Orbitrap Fusion Tribrid in positive ESI MS only mode (Orbitrap resolution 120000, 4 microscans).

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 peptide carry over, 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 carry over and only deuteration differences, which are present in both datasets, are considered real.

Peptide and PTM identification

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 microscan, 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 [22]. 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. For

peptides carrying a deamidation, spectra were checked manually and chromatographic peak areas where calculated in Xcalibur (Thermo Scientific) to obtain a wildtype/deamidated peptide ratio.

HDX data analysis

DeutEx software (obtained from peterslab.org) was used to determine the deuterium uptake via centroid analysis. Excel (Microsoft), GraphPad Prism (GraphPad Software, Inc.), OriginPro 2016 (Origin Lab Corporation) and PyMOL (Schrödinger) software were used for data visualization and statistical analysis. For comparison of triplicate data, a two-sided Student's T-test using deuteration differences 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 [23, 24] of the dataset either for several time points or for the same time point in overlapping peptides. For some peptides deuteration of FD controls was lower than deuteration of the 8 h labeling time point. Therefore, datasets were not normalized to the absolute FD deuterium uptake and only relative differences between states are presented. For comparison of the unbound wildtype and deamidated MI001 P dimer, the ratio of the FD controls from both measurements was used for normalization. Additionally, a higher cut-off of $\Delta D > 0.42$ (99% percentile calculated according to [25]) 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).

Deuterated spectra of peptides in certain protein regions showed bimodal peak distributions that led to lower deuteration values in centroid analysis. To validate the deuteration differences observed in centroid data analysis and to calculate relative intensities of both peak distributions, bimodal spectra of peptides representative for certain regions were analyzed by binomial fitting in

HXExpress [26]. 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. 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 [27] and can be found in the supplement (Figures S12-16).

Experimental Design and Statistical Rationale

The rationale for experimental design and data analysis is based on HDX-MS community-recommendations [28]. In brief, sample quality was assessed with native MS and HDX-MS conditions were optimized for maximum sequence coverage and detection sensitivity. Labeling time points were chosen to cover 3-4 orders of magnitude. Three independent labeling reactions were performed for each time point and the level of back exchange was assessed with a fully deuterated protein control as well as a mix of deuterated model peptides. Details about the peptide identification method, statistical analysis with Student's T-test and color mapping procedure are given in the individual methods sections. Fragmentation spectra for identification of deamidated peptides are given in the supplement (Figures S3-9). All HDX-MS data has been manually inspected and exchange differences in bimodal peak distributions have been validated by binomial fitting. HDX summary tables with detailed information about experimental conditions and statistics as well as deuterium uptake plots for each dataset can be found in the supplement (Table S3 and Figures S24-32).

Structure and sequence alignment

GII.4 Saga, GII.4 MI001, GII.10 Vietnam and GII.17 Kawasaki VP1 protein sequences were aligned with T-Coffee [29] and visualized with Jalview (version 2.11.0) [30]. A phylogenetic tree was created in Jalview with BLOSUM62 and Neighbor joining. GII.4 Saga (pdb 4X06), GII.10 Vietnam (pdb 3ONY) and GII.17 Kawasaki (pdb 5F4O) P dimer crystal structures were superimposed in PyMOL.

Homology modeling of GII.4 MI001 P dimer structure

The SWISS-MODEL template library [31] [32] SMTL version 2019-10-24, PDB release 2019-10-18 was searched with BLAST [33] and HHBlits [34] for evolutionary related structures matching the GII.4 MI001 P dimer target sequence. Based on the search results the GII.4 Farmington Hills P dimer structure (pdb 4OOV, 94 % sequence identity) was used for model building. Models were built based on the target-template alignment using ProMod3. The global and per-residue model quality has been assessed using the QMEAN scoring function [35]. The resulting GMQE (Global Model Quality Estimation) was 0.99 and QMEAN was 0.57 indicating very good accuracy and quality of the model structure (Figure S2).

MD simulations

We performed molecular dynamics (MD) simulations on the following proteins: GII.4 Saga (pdb 4X06), GII.4 Saga containing a deamidated P domain (pdb 6H9V) (iDN), GII.4 MI001 (pdb 4OOV), GII.10 Vietnam (3ONU), GII.17 Kawasaki (5F4O). An additional proteoform was generated comprising fully deamidated P dimers at residue 373 in both peptide chains, based on GII.4 Saga (pdb 6H9V) (iDiD). All pdb-structures were refined by adding missing atoms and residues using the UCSF Chimera tool (version 1.14) [36]. GII.4 Saga P dimers were additionally simulated with α-L-methyl-fucose (F) ligands to explore a potential influence of deamidation on protein dynamics.

Hence, the amount of systems is expanded to include wildtype GII.4 Saga P dimers (pdb 4X7C) (NN) with one (N_FN) and two (N_FN_F) fucose ligands, iDN P dimers with one fucose complexing each individual chain (iD_FN and iDN_F), and two fucoses (iD_FN_F), and further include iDiD P dimers with one (iD_FiD) and two (iD_FiD_F) fucose ligands.

All MD simulations were performed using Gromacs on the Rackham cluster of the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX), and the Kebnekaise cluster at the High Performance Computing Center North (HPC2N) [37]. The amber99sb force field was utilized for all simulations [38], modified to include parameters for iD and F [39]. The MkVsites tool provided virtual sites and dummy-mass constructions for F [40]. Structures were placed in a dodecahedral box under periodic boundary conditions, solvated using the TIP3P water model [41], and neutralized in a 154 mM saline solution by adding NaCl. Protonation states of all systems were based on the sidechains' pKa at pH 7. Each system was minimized using the steepest descent algorithm, followed by a 100 ps simulation with applied position-restrains. Temperature and pressure were maintained at 300 K and 1 bar by the v-rescale thermostat and the Parrinello—Rahman barostat, with coupling constants of 50 fs for both [42-44]. Neighbor lists were updated every 10 steps. The particle mesh Ewald algorithm was used for Coulomb interactions, with a real-space cut-off of 1.0 nm [45, 46]. The systems were allowed to relax for 100 ns with a 5 fs time step, extracting one frame every 10^{th} ns as starting structures for the production runs. Final simulations were performed for ten 100 ns production runs at a 5 fs time step. As such, each of the 14 systems was simulated in 10 replicates from different starting structures, resulting in an aggregated simulation time of 1 μs per system, making 14 μs in total for all systems.

The root-mean-square deviation (RMSD) and fluctuation (RMSF), as well as the solvent accessible surface area (A_{sas}), were calculated to analyze the behavior of each system. The RMSD was computed with the first frame of the individual trajectory as reference structure. The trajectories of the ten replicas were combined to a single trajectory, of which the average structure

was calculated with the Gromacs software package. This average structure was then taken as reference for RMSF calculations, as this most accurately represents the standard deviation of the individual atomic positions. To further support the RMSF calculations, we computed the A_{sas} of the protein backbone for the initial conformation (pdb-structure) and the final production simulations, of which latter was combined to an average representation of the area over all ten production replicas. The resulting values were subtracted from each other, to eventually visualize an increase or decrease of the A_{sas} after 100 ns. The RMSF and A_{sas} values for the different P dimer strains were aligned with the sequence to compare each residue between the dimers.

Data availability

Full HDX data tables as well as MS raw data and peptide identification results have been deposited to the ProteomeXchange Consortium [47] via the PRIDE [48] partner repository (dataset identifier PXD019884). Annotated fragment ion spectra of all protein/peptide identifications can be viewed with MS-Viewer using the respective search keys given in the supplement (Table S4).

Results

P dimer quality control by peptic digest and native MS

To verify the deamidation status P dimer samples were subjected to peptic cleavage followed by LC-MS for peptide identification (Table S2). No deamidated peptides were identified for GII.10 Vietnam and GII.17 Kawasaki after 1 year and 4 months (Vietnam) and 1 year (Kawasaki) of storage at 5°C, respectively. For GII.4 MI001 P dimers stored for 1 year at pH 7.3 and 5°C, a fraction of approximately 64 % was deamidated at N373. Based on this the ratio of purely native (NN) to half deamidated (iDN) to fully deamidated (iDiD) P dimers is statistically predicted as 13:46:41 %. Furthermore, a minor fraction was deamidated at N239 and N448, respectively. For GII.4 MI001 P dimers stored at pH 4.9 for 5 months at 5°C, no deamidation of N373 was observed. Only a small fraction (< 10 %) of deamidated N448 was identified. GII.4 Saga P dimers stored for more than 2 years at pH 7.3 at 5°C were approximately 88 % deamidated at N373 leading to a ratio of NN:iDN:iDiD of 1.5:21:77.5 % (Table S1 and Figures S3-9).

Prior to HDX-MS analysis P dimers were subjected to native MS for quality control. Furthermore, ion exchange separated wildtype (NN) and fully deamidated (iDiD) 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 unspecific tetramers formed during the ESI process (Figure S10). 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 S2). As deamidation rates in these strains are identical [49] this suggests that monomers are a result of iDiD P dimer dissociation, while iDN species are still primarily dimeric.

Bimodality of deuterated peak distributions points towards a second, more protected population

Based on glycan binding data for GII.4 Saga P dimers [19, 49], we wanted to expand our HDX MS experiments to P dimers of other human norovirus strains to analyze possible strain specific differences in the structural response to glycan binding. Therefore, we incubated GII.4, GII.10 and GII.17 P dimers with 10 mM HBGA B trisaccharide or 100 mM fucose at pH 7 and measured differences between the unbound and ligand-bound state using HDX-MS. During inspection of the glycan binding data, we observed that deuterated spectra of some peptides had a bimodal character with one low intense, low deuterated and a second high intense, higher deuterated peak distribution (Figure 2A). These bimodal peak distributions can have many causes, e.g. two distinct protein conformations, conformational rearrangements that lead to EX1 exchange kinetics, insufficient ligand saturation or peptide carry over from the analytical or protease column [50, 51]. To rule out effects induced by carry over, an additional dataset with randomized sample order and additional washing of the pepsin column between sample injections was measured for GII.4 MI001 wildtype (wt, N373) P dimers incubated with fucose, which still showed bimodality. Moreover, bimodality is also observed in absence of any ligand, strongly indicating that undersaturated binding sites are not the origin of bimodality. Furthermore, ligand concentrations were chosen to provide high and comparable saturation of binding sites. HBGA B trisaccharide affinities are almost identical for GII.4 Saga and GII.4 MI001 P dimers [49]. Assuming that P dimers of other strains bind glycans with affinities similar to GII.4 P dimers, binding pocket occupancy can be estimated based on K_d values measured by NMR (GII.4 Saga (NN): K_d = 5.5 mM for HBGA B trisaccharide and K_d = 22 mM for fucose) [19]. Binding of HBGA B trisaccharide in our setup would hereby correspond to 95 % binding site occupancy during equilibration with ligand (98 % for fucose) and 65 % during deuterium labeling (82 % for fucose). For GII.4 P dimers, recent chemical shift perturbation titrations demonstrate the presence of two independent HBGA binding sites [19,

49]. The presence of four fucose binding sites shown for GII.10 Vietnam P dimers in the presence of high fucose concentrations appears to be an exception [12].

Bimodality can be seen for P dimers of GII.4 MI001, GII.10 Vietnam and GII.17 Kawasaki in the unbound proteins as well as in ligand bound forms. However, no bimodality is observed for GII.4 Saga P dimers [19]. For all three strains, bimodality is almost exclusively present in the P2 domain and in the lower part of the P1 domain (Figure 2B). Peptides that are affected by glycan binding are often bimodal as well; therefore, it was necessary to manually analyze deuteration differences in these regions again by binomial fitting of the individual peak distributions (an example analysis can be seen in Figure S11). Residues 334-354, for example, are bimodal and also involved in glycan binding in all three strains. However, not all peptides that are protected upon glycan binding are bimodal. Peptides covering the canonical binding site are unimodal in GII.17 Kawasaki and GII.10 Vietnam and only show slight bimodality in GII.4 MI001 in the presence of fucose.

Relative intensities of the individual peak distributions are constant over time (Figure S11) and highly similar for peptides within the same protein, which lead us to the assumption that the P dimer adopts two distinct conformations, a compact and a more flexible one. The relative intensity ratios of the peak distributions vary between experiments, but can still be compared within a certain experiment (Figure 2C). Depending on the strain and the experiment, the relative intensity of the first peak distribution in the unbound P dimer varies between 7 and 17 %. For GII.17 Kawasaki and GII.4 MI001 P dimers, incubation with HBGA B trisaccharide has no significant effect on the relative intensity of the first distribution, while there is a slight increase for GII.10 Vietnam P dimers. Presence of fucose, in contrast, significantly increases the relative intensity of the first distribution in all strains. While there is only a slight increase for GII.17 Kawasaki P dimers, the relative intensity of the first distribution increases by a factor of 2 for GII.10 Vietnam and GII.4 MI001 P dimers.

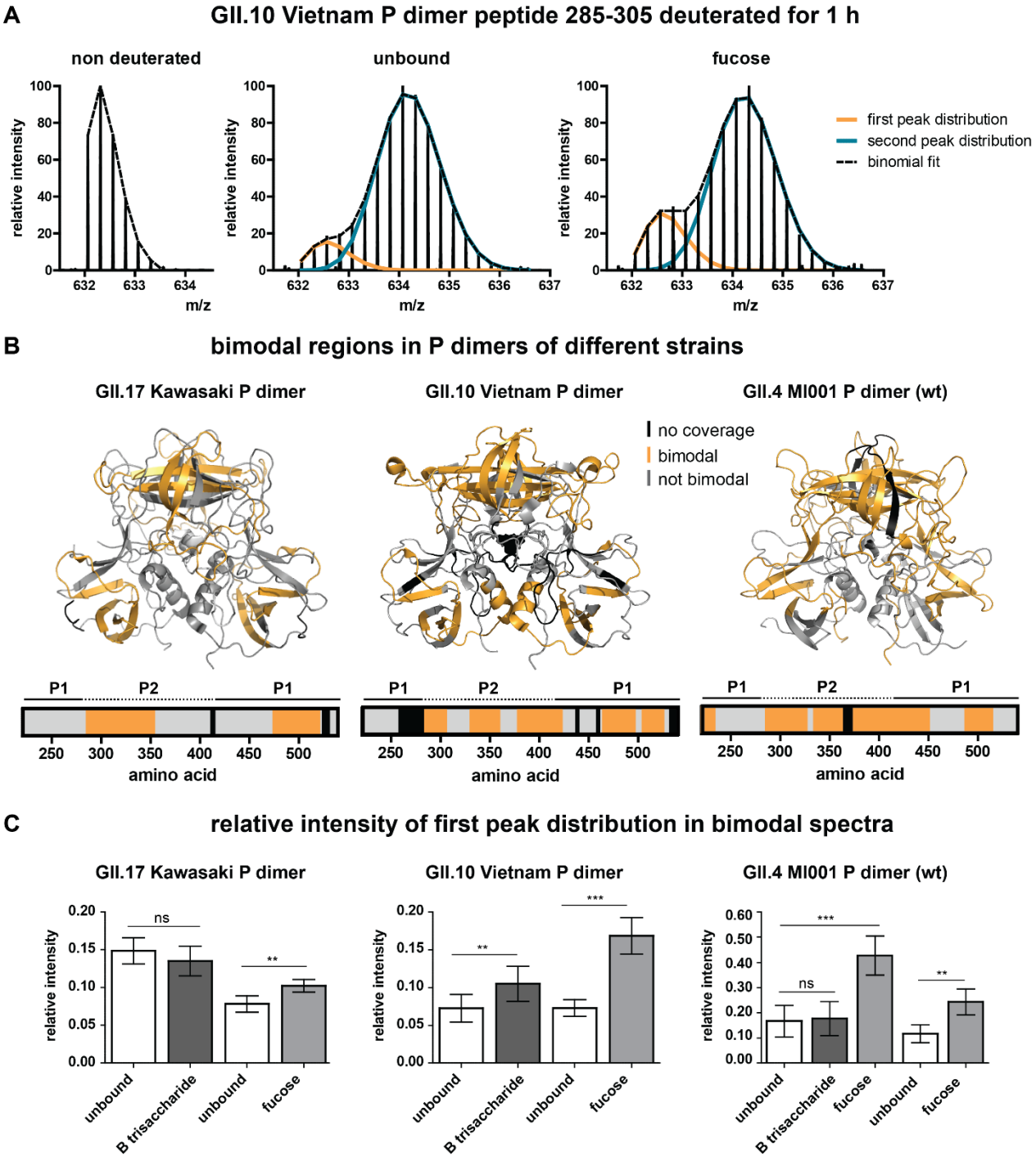


Figure 2: Bimodal peak distributions in some peptides point towards the existence of a second, more protected population. A) Bimodality in deuterated spectra of an exemplary peptide of the GII.10 Vietnam P dimer. Bimodality occurs in both the unbound and the glycan bound state. Single peak distributions can be separated by binomial fitting. B) Regions in P dimers of different strains that show bimodality in deuterated spectra (orange). Bimodality mainly occurs in the P2 domain and the lower part of the P1 domain in all

414 strains. The amino acid numbering is based on the VP1 sequence. C) Relative intensity of the first (lower
415 deuterated) peak distribution for different strains and experiments calculated by binomial fitting. For GII.4
416 MI001 both HDX-MS experiments are shown (triplicate measurement with B trisaccharide and fucose vs.
417 single replicate measurement with fucose and additional wash steps between the injections to eliminate
418 peptide carry over from the pepsin column). Relative intensity of the first peak distributions stays constant
419 over time, so the averages over all bimodal time points for all states were calculated for several peptides of
420 different protein regions and combined into bar graphs. The error bar represents the standard deviation of
421 the average relative intensity calculated from N ≥ 5 peptides. Significant differences between the unbound
422 and the glycan bound state were assessed using a two-sided Student's T-test for each pair in an individual
423 experiment. P values are indicated by asterisks: p<0.001 (***), p<0.01 (**) and not significant (ns).

424 Analysis of glycan induced changes in P dimers of different strains by HDX-
425 MS

426 Most of the deuteration changes in presence of HBGA B trisaccharide or fucose are detected in
427 peptides that show bimodality. This commonly causes falsely low deuteration differences in
428 centroid analysis, so individual binomial fitting of the two peak distributions was performed for
429 some representative peptides in these regions to validate the observed deuteration differences in
430 the main (second) peak distribution. As expected, binding of HBGA B trisaccharide and fucose
431 induces changes in P dimers of all three strains, primarily in the P2 domain, indicating occupation
432 of the glycan binding pocket (Figure 3). Protected regions in wildtype GII.4 MI001 P dimers are
433 highly similar to GII.4 Saga P dimers [19] for both glycans (canonical binding site G443, Y444 and
434 residues 283-303). In addition, protection of a β-sheet region in the top cleft of the P2 domain
435 (residues 333-353) can be detected in GII.4 MI001 in presence of HBGA B trisaccharide. Chemical
436 shift perturbations (CSP) in this region could also be seen in NMR experiments with GII.4 Saga
437 P dimers in presence of glycans [19, 49]. Overall, protected regions in GII.4 MI001 match with
438 regions showing CSPs in GII.4 Saga NMR data, suggesting that both strains respond similarly to
439 glycan binding.

Protection of residues 333-353 in the P2 domain can be seen in all three 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 center below the P2 domain, can be detected (Figure 3 A). In contrast to GII.4 MI001 and GII.17 Kawasaki, GII.10 Vietnam P dimers show protection in the P2 domain including the canonical binding site (G451, Y452) and the β-sheet region in the binding cleft, but also in the lower part of the P1 domain (Figure 3 D). All observed differences 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 labeling time was too short to detect deuteration differences in already strongly protected regions.

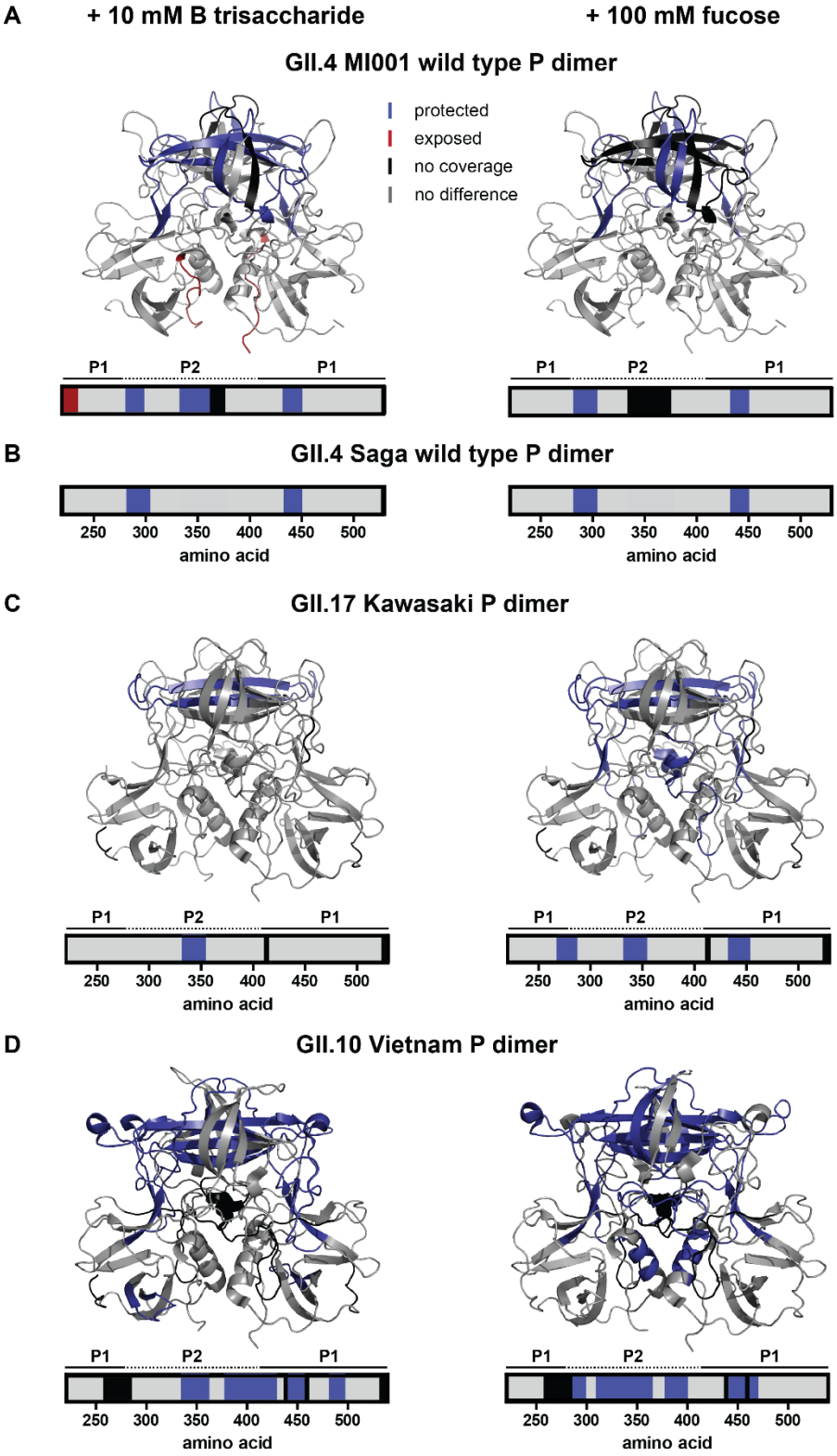


Figure 3: HDX differences upon glycan binding in human GII.4 MI001 (A), GII.17 Kawasaki 308 (C) and GII.10 Vietnam (D) norovirus P dimers. (B) Protected regions in wildtype GII.4 Saga P dimers are shown for comparison [19]. Depicted are protein regions with significant differences in deuterium uptake 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). The deuteration difference in the second peak distribution was manually validated by binomial fitting in case of bimodal spectra. Bar graphs and colored structures indicate regions of P dimers, which get more protected (dark blue) or exposed (red) upon interaction with glycans. Areas colored 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).

Influence of N373 deamidation on dynamics and glycan binding of GII.4 P dimers

To study the influence of partial vs. complete deamidation of N373 on glycan binding [19] (Figure 4C), HDX-MS experiments with partially deamidated GII.4 MI001 and GII.4 Saga P dimer samples in the presence of 100 mM fucose were performed. Strikingly, protection of the canonical fucose binding site (G443, Y444) could be detected in both strains (Figure 4 A/B). This shows that fucose binding is still possible in partially deamidated (iDN) or even fully deamidated (iDiD) P dimers at the given concentration, even though binding is attenuated compared to the N373 wildtype [19]. Occupation of the canonical binding sites has been seen in crystal structures of deamidated GII.4 Saga P dimers at elevated concentrations of 600 mM fucose, but binding interactions were slightly different from wildtype [19]. In contrast to wild type GII.4 MI001 P dimer, no other region was protected from HDX under fucose treatment, but increased deuteration in the main peak distribution was observed in the P2 domain of both GII.4 strains, suggesting a more exposed conformation. Interestingly, residues 335-362, which are protected in the wildtype proteins of all strains, show increased deuteration in partially deamidated GII.4 P dimers. As we have a mixture of wildtype and deamidated P domains in the sample, the mass shift we see in the deuterated spectra reflects the average of all components, unbound and fucose-bound NN, iDN and iDiD P dimers, which cannot be discriminated. However, binding probability calculations can give a hint, which species contribute most to the observed increase in deuteration in presence of fucose.

Considering the fractions of wildtype and deamidated P domains and their different K_ds for fucose binding [19], in GII.4 MI001 only 17 % of binding events occur in pure wildtype NN dimers, 54 % in half-deamidated iDN dimers and 29 % in fully deamidated iDiD dimers. For GII.4 Saga even more binding events take place in fully deamidated P dimers (2:31:67 % for NN, iDN; iDiD), clearly showing that the detected increased deuteration is caused by fucose binding to at least half-deamidated P dimers. This suggests that fucose binding results in different dynamics in the wildtype and partially deamidated protein.

In GII.4 MI001 regions with bimodal peak distributions (Figure 4D), relative intensities of both distributions are similar to the ones in the wildtype protein. However, interaction with fucose in the partially deamidated GII.4 MI001 P dimer does not lead to a significant increase in the relative intensity of the lower deuterated peak distribution, as seen in the wildtype protein (Figure 4D). Slight bimodality is also present in peptides covering the canonical fucose binding site in the fucose bound state but relative intensities are similar to the ones observed for the wildtype protein. A comparison of FD normalized deuteration levels for wildtype and deamidated GII.4 MI001 P dimers without glycans reveals increased deuterium incorporation in large parts of the P2 domain, as well as the P1 domain (Figure 4E). Highest deuteration differences ($\Delta D > 1$ Da) are detected for residues 335-432 located in the P2 β-sheet cleft. This is in line with the increased flexibility in the P2 domain of the deamidated GII.4 Saga P dimer [19] (Figure 4F), however, in GII.4 MI001 this effect is propagated into regions more distant from the glycan binding pocket and deamidation site. The increased dynamics could weaken the dimer interfaces and therefore explain the dissociation into monomers in the deamidated protein. Additionally, monomers will most probably also experience higher HDX because of missing dimer interactions in the P1 domain and the P2 β-sheets [52] that will add to the observed increase in deuteration compared to the exclusively dimeric wildtype protein.

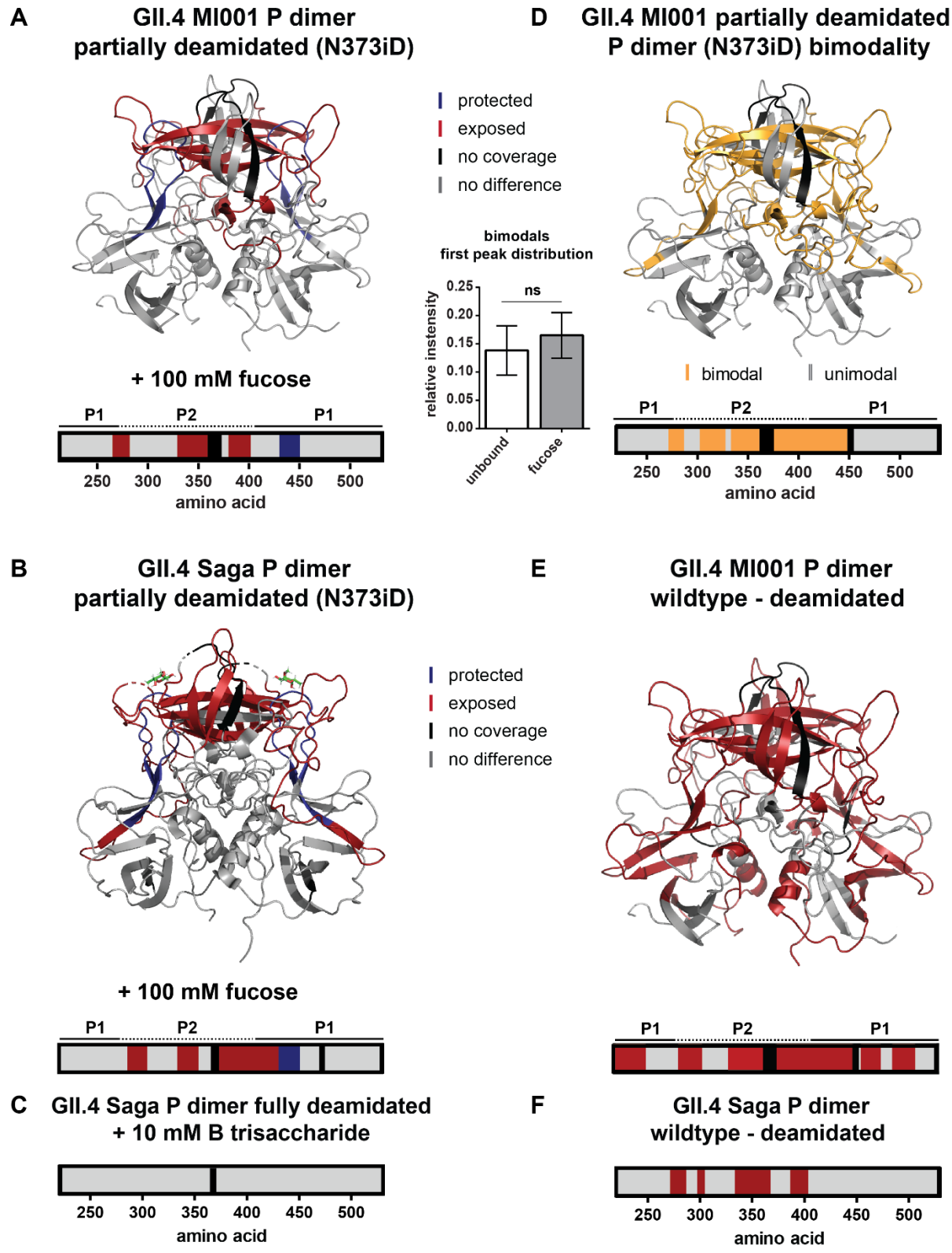


Figure 4: Significant HDX differences in partially deamidated (N373iD) GII.4 MI001 P dimers. A/B) Fucose can still bind to the canonical glycan binding site in partially deamidated GII.4 MI001 (A) and GII.4 Saga (B)

P dimers (pdb 6H9V). In contrast to the wildtype N373 P dimer, parts of the P2 domain get more exposed upon interaction with 100 mM fucose. (C) HDX differences in fully deamidated GII.4 Saga P dimers in presence of 10 mM HBGA B trisaccharide are shown for comparison [19]. (D) Bimodality occurs in similar regions as for the wildtype P dimer, implying that the more protected population is also present in the partially deamidated sample. In contrast to the native P dimer, the relative intensity of the first peak distribution does not significantly increase under fucose treatment (for statistics refer to description of Figure 3). (E) The partially deamidated GII.4 MI001 N373iD P dimer shows a higher deuterium uptake in large parts of the structure, which points towards higher flexibility, like in the fully deamidated GII.4 Saga P dimer [19] (F).

MD simulations

MD simulations were utilized to further investigate the norovirus P dimer strains. The RMSD relative to the starting structure post-equilibration were calculated in order to estimate protein dynamics during MD simulations. RMSF calculations were employed to examine fluctuations throughout the simulated time frame, and highlight alterations in flexible regions of the different protein chains. As support for the RMSF data, we calculated the A_{sas} of the P dimers during the simulation with respect to their crystal structure, providing an understanding of an increase or decrease of the surface area of each individual residue.

The RMSD for the four norovirus P dimer strains are reported in Figure S18, in which the simulated GII.4 Saga and MI001 P dimers reached a value of 1.5 Å after 100 ns. GII.10 Vietnam P dimers show a maximum deviation around 90 ns at 2 Å, decreasing to 1.8 Å after 100 ns. GII.17 Kawasaki P dimers demonstrate a still slightly increasing trend at the end of the simulation, indicating that this system has not yet fully adapted to the solution environment. The RMSFs and A_{sas} relative to the respective crystal structures reveal differences in protein chain flexibility of the norovirus strains, as depicted in Figure 5 and S17. The sequences were aligned for a better comparison. Hence, resulting gaps in the individual RMSF graphs are due to missing residues at that specific position. Least stability is introduced for the GII.4 Saga strain, as the RMSF values suggest only a limited increase in fluctuation during the 100 ns of simulation. GII.4 MI001 and GII.17 Kawasaki

follow a similar trend. In contrast, GII.10 Vietnam P dimers show overall higher flexibility compared to the other strains, in particular a peak around residue 350 in the P2 domain. Similar trends can be observed for the A_{sas} graphs depicted in Figure S17. GII.10 Vietnam demonstrates the highest area values, which support the peaks observed in the RMSF graph in Figure 5A. The various P dimer structures were overlaid in the PyMOL software, where areas of interest were imaged in order to further explore differences of the protein crystal structures (Figure S19).

Investigating GII.4 Saga P dimers complexed by fucose, and the potential role of deamidation, the data revealed minimal difference between the RMSD values of said systems. RMSDs for the NN, iDN and iDiD G.II 4 Saga dimers show a similar trend, reaching a value between 1.6 and 1.75 Å after 100 ns of simulation (Figure S20). RMSF and A_{sas} calculations to investigate the role of deamidation in the GII.4 Saga strains show that the individual graphs follow a similar trend, suggesting only limited influence of the deamidation on the overall P dimer structure (Figure S21-23) which is in line with previous crystallography data [19].

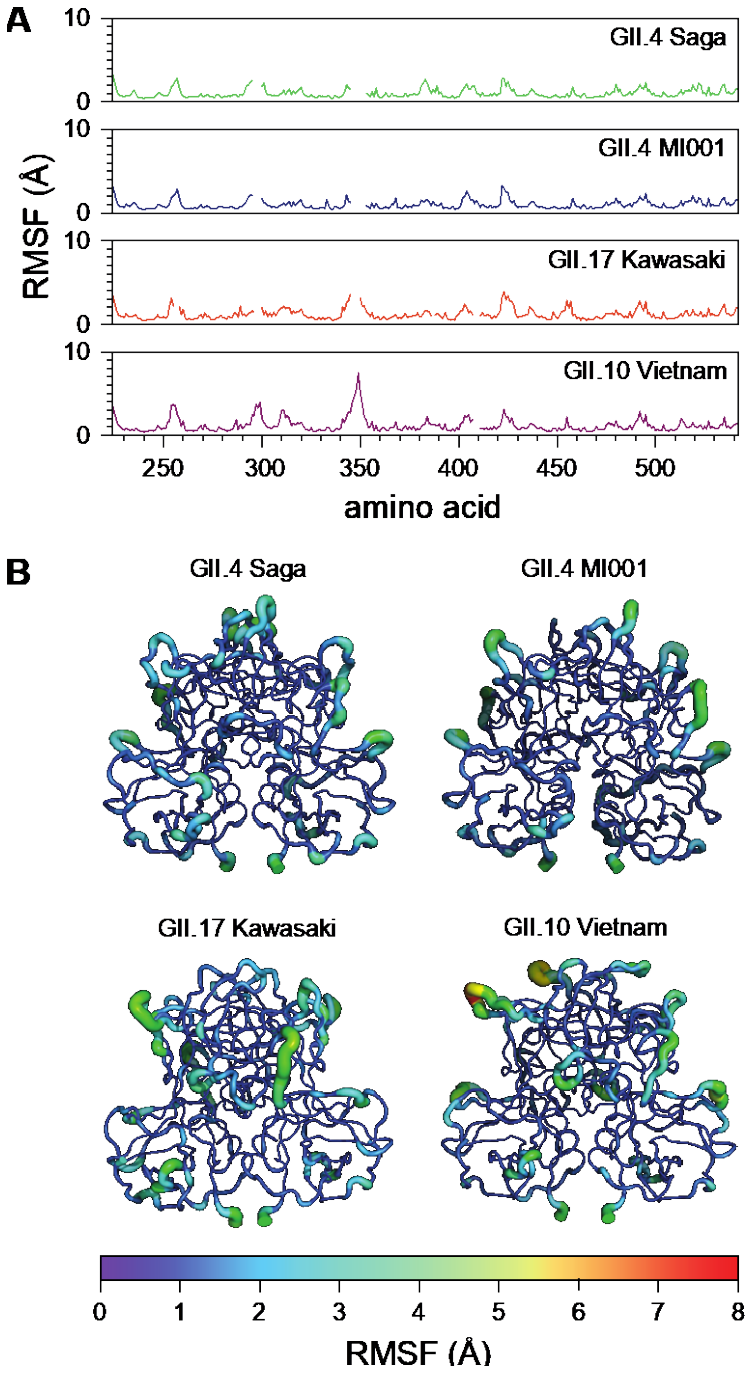


Figure 5: Difference in fluctuations between the P dimer strains propose increased flexibility in the protein chains in absence of a ligand. A) RMSF data of GII.4 Saga, GII.4 MI001, GII.17 Kawasaki and GII.10 Vietnam, simulated for a total of 1 μs each, reveal different protein chain dynamics between the strains, with most prominent peak around residue 350. Gaps in the data originate from alignment of the norovirus P dimer sequences. B) RMSF values of GII.4 Saga, GII.4 MI001, GII.17 Kawasaki and GII.10 Vietnam visualized in the structures, highlighting residues with increased fluctuations during the simulations.

Discussion

In this study, we address 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 pandemic GII.4 genotype and has been shown to infect mice as well [21].

Bimodal peak distributions could originate from P particle formation

In our glycan-binding data, we observe the presence of bimodal peak distributions in a large variety of peptides located in the P2 domain and the lower part of the P1 domain in all strains but the previously analyzed GII.4 Saga. The intensity of the lower deuterated peak distribution was between 7% and 17 % and relative intensities of both distributions remained constant over time. This observation points towards two distinct protein populations [26] that experience a different level of HDX over the whole exchange period. The low deuteration of the first peak distribution suggests the presence of a compact conformation that is shielded from HDX. This could be in principle true for protein aggregates, however, a protection of only the top and bottom part of the protein appears very distinct. In case of protein aggregation, we would expect bimodality all over the protein surface.

So, if the lowly deuterated population is no artifact, what could it be instead? The P domain can form larger oligomers of different stoichiometry, up to whole 24-mer P particles, depending on the protein concentration [53, 54]. P oligomers form contacts through interactions in the lower part of the P1 domain of each P dimer [53], which could explain the reduced deuteration in this area. Closer inspection of the cryo EM structure [53] also suggests more contacts between the P2 domains compared to free P dimer. Importantly, the absence of bimodality in GII.4 Saga P dimers implies that this strain has a different ability to form P oligomers than the closely related GII.4 MI001 strain.

P particles can bind HBGAs and are even suspected to interact with them in the same way as VLPs [53, 55]. However, in the presence of glycans no significant deuteration difference could be detected in low deuterated peak distributions. Nevertheless, this does not explicitly mean that there is no binding in these areas. Interpretation of no significant deuteration difference on a certain time scale as the absence of any structural change should be treated with caution for several reasons [28]. First, the intensity and thus the signal-to-noise ratio in the lower deuterated peak distribution is low, which makes it difficult to detect statistically significant changes in deuteration. Secondly, HDX is reduced in the lower deuterated population compared to P dimers and would therefore need longer exchange times to reflect potential differences. In addition, peak distributions in peptides covering the canonical glycan binding site are unimodal and show protection in all strains, meaning this interaction can be found in the entire protein population.

In our data there is a clear increase of the potential P particle population in presence of 100 mM fucose, which could mean that interaction with glycans supports the formation of P particles that is otherwise less pronounced [55]. It has to be noted that we did not observe P particle oligomers in native MS of a 4.5 µM P dimer solution and our protein constructs lack the C-terminal arginine cluster that has been shown to be important for P particle formation [56, 57]. However, the < 20% of monomers, assumingly assembled into 24-mer P particles in absence of glycans, would amount to around 1% of total signal intensity split up into many charge states in native MS, which likely drop below detection limit. In contrast, fractions of structural variants of less than 5 % can be detected in a properly conducted HDX-MS experiment [58].

Glycan binding in different strains

Apart from the presence of two distinct protein species in the sample, we could also detect differences in glycan-induced protein dynamics in different strains. For GII.4 MI001, protected regions were almost identical to the earlier investigated GII.4 Saga P dimer (canonical binding site G443, Y444 and residues 283-303) [19], 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 [19]. Furthermore, a recent NMR study suggests identical glycan binding behavior of both GII.4 strains [49]. 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 [11, 59]. 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 [60-62].

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 [10, 12]. 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. Due

to close proximity of interacting amino acids in fucose binding sites 1/2 and 3/4 we cannot distinguish these binding sites in HDX data at peptide resolution, but occupation of all four binding sites is likely at the given concentration [12]. Interestingly, we see a protection of several residue stretches that cannot be explained by known glycan interactions. Residues 285-298 are protected in presence of fucose, similar to GII.4 strains. Residues 311-336 belong to an unstructured region below the P2 binding cleft and could link the protection observed in the cleft to the one in residues 285-298. 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 [12]. HBGA B trisaccharide binding is mainly mediated by the fucose residue, with an additional interaction of galactose with G451 and some water mediated interactions [10]. 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.

Table 1 Comparison of protected residues in HDX with known glycan interactions in crystal structures. Binding sites 1/2 are conserved for many strains and glycans, binding sites 3/4 have so far only been detected in GII.10 Vietnam. For GII.17 Kawasaki only crystal structures with fucose and A trisaccharide are available, so protected residues for B trisaccharide are compared to binding sites seen for A trisaccharide. P dimer chain annotations are given for fucose binding sites 1 and 3. No crystal structure is available for GII.4 MI001 P dimers, so binding sites are marked as unknown (NA). Protected residues for wildtype GII.4 Saga P dimers [19] are shown for comparison.

Protected residues in HDX	**Fucose binding site 1/2**	**Fucose binding site 3/4**
GII.10 Vietnam + 100 mM fucose [12]		
285-298	-	-

311-336	-	-
337-364	N355 (chain A) R356 (chain A)	E359 (chain A)
379-399	D385 (chain A)	W381 (chain A)
442-458	G451 (chain B) Y452 (chain B)	L449 (chain A)
GII.10 Vietnam + 10 mM HBGA B trisaccharide [10]		
336-361	N355 (chain A) R356 (chain A)	-
379-428	D385 (chain A)	-
440-458	G451 (chain B) Y452 (chain B)	-
483-496	-	-
GII.17 Kawasaki + 100 mM fucose [11, 59]		
269-286	-	-
333-353	T348 (chain A) R349 (chain A)	-
434-452	G443 (chain B) Y444 (chain B)	-
-	D378 (chain A)	-
GII.17 Kawasaki + 10 mM HBGA B trisaccharide [11]		
333-353	T348 (chain A) R349 (chain A)	-
-	G443 (chain B) Y444 (chain B)	-
GII.4 MI001 + 100 mM fucose		
283-303	NA	NA
434-449	NA	NA

GII.4 MI001 + 10 mM HBGA B trisaccharide		
283-298	NA	NA
333-353	NA	NA
434-450	NA	NA
GII.4 Saga + 100 mM fucose **GII.4 Saga + 10 mM HBGA B trisaccharide [18, 19]**		
283-303	-	-
no coverage	D374 (chain A)	
434-449	G443 (chain B) Y444 (chain B)	-

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. 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 and GII.10 P dimers show protection of residues 285-298, which is also absent in GII.17 P dimers. Interestingly, P dimers of the more prevalent strains GII.4 and GII.17 [3, 4] show less changes in HDX upon glycan binding compared to GII.10 Vietnam, which is rarely detected in patients [10].

The RMSD plots for the four investigated P dimer strains without ligand reveal minimal differences between all systems (Figure S22). Whilst GII.4 Saga and MI001 trends reach a plateau, one can observe a still increasing trend GII.17 Kawasaki. This indicates that this system has not yet reached a stable conformation. The GII.10 Vietnam shows a decrease towards the end of the

simulation, indicating that this system just adapted to the environment and obtained a stable structure.

For the MD simulations, we were interested in the dynamics in absence of fucose, and we observed an increase in flexibility throughout the peptide chain, accompanied with changes of the A_{sas}, which we present in Figure 5. The most prominent difference between the strains is a high peak around the glycan binding site near residue 350 exclusively in GII.10 Vietnam P dimers. GII.10 Vietnam has a longer loop around residue 350, which could explain the higher flexibility in absence of fucose (Figure 5). On further investigation however, when complexed by the ligand in the crystal structure, this loop adopts a short helical structure (Figure 6A), forming a pocket shielding nearby residues from deuterium exchange, as observed in the HDX-MS experiment (Figure 3B). The crystal structures of the other strains have more unstructured loops. In our simulations of ligand-free GII.10 Vietnam P dimers, the loop becomes flexible and unstructured, as seen by the high RMSF values of up to 7.5 Å, and evident from snapshots taken from the MD trajectory (Figure 6B), which would explain the protection provided by bound ligands. This is further supported by the increase of area accessible by the solvent (Figure 5B). As such, for GII.10 Vietnam, the binding of fucose or other ligands promotes a structural rearrangement of the glycan binding site near residue 350 (Figure 6A) not seen in the other strains.

Around residue 250 and 300 the GII.10 Vietnam strain presents increased flexibility compared to GII.4 Saga, GII.4 MI001 and GII.17 Kawasaki. Near residue 424, a smaller peak can be observed for all four strains (Figure 5).This could be a result of the individual chain orientations, and the different sequence alignment one can find in these areas (Figure S19A-C). The residues that are part of this area of interest seem to form smaller loops on the surface of the protein, which is likely the reason of the recorded high flexibility.

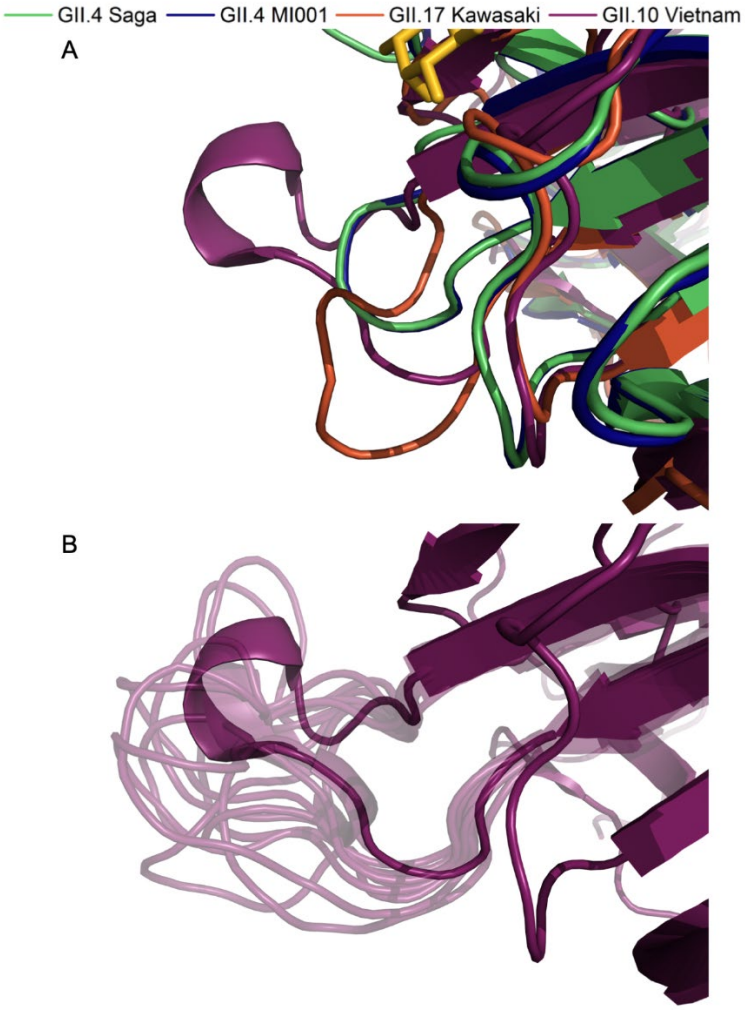


Figure 5: Structural rearrangement of the glycan binding site in absence of fucose. A) Zoomed in image of the overlaid P domain structures around residue 350, complexed with fucose. The presence of fucose stabilizes a small helix, forming an ordered pocket in the GII.10 Vietnam strain (purple). B) Absence of fucose results in a high flexibility of the GII.10 Vietnam pocket and a loss of the short helix, as shown in the snapshots of different conformations throughout the simulation (light purple). These snapshots were taken from every 10th frame of a single trajectory.

The role of N373 deamidation

For GII.4 Saga P dimers it was observed earlier that spontaneous transformation of N373 into iso-aspartate (iD) attenuates glycan binding in fully deamidated iDiD P dimers. Deamidation is site-specific and happens over a timescale of 1-2 days at pH 7.3 and 37°C correlating with the length

of the infection cycle [19]. The high specificity of this deamidation under infection conditions suggested that this process could occur as well *in vivo*, however, the biological relevance for infection remained unclear.

To test if this effect can also be found in the closely related GII.4 MI001 strain, we performed HDX-MS on a spontaneously deamidated P dimer sample, resulting in mixed populations of NN, iDN and iDiD dimers, which is more likely to be found in a natural infection context. For comparison, we also measured fucose binding to a partially deamidated GII.4 Saga P dimer, which contained an even higher fraction of fully deamidated iDiD P dimers. Strikingly, protection of the canonical glycan binding site could still be detected in both isolates under fucose treatment, mainly corresponding to binding to iDN and iDiD P dimers. Additionally, the P2 cleft was more exposed in the partially deamidated sample under fucose treatment, in contrast to the protection observed in the wildtype NN P dimer. This indicates that under natural deamidation conditions, glycan binding at the canonical binding site still happens, but induces different dynamics than in the purely wildtype P dimer. The exposure of the P2 cleft suggests that after glycan binding this area gets more flexible, which could be required to interact with other factors or the until now unknown receptor. As such an increase in deuteration is not present in wildtype NN P dimers, this effect must be caused either by direct binding to deamidated P domains in iDN or iDiD dimers or by binding to wildtype P domains in iDN dimers, whose overall dynamics are altered by the influence of the neighboring deamidated monomer. RMSD, RMSF and A_{sas} calculation for the GII.4 NN, iDN and iDiD Saga P dimers show no striking differences when compared to each other and follow a similar trend (Figure S20-23). The fact that our MD simulations were unable to detect differences between deamidated and non-deamidated P dimers suggests that any differences in dynamics are manifested on timescales longer than a few 100 ns. Neither protection of binding sites nor increased deuteration in P2 has been seen in previous HDX-MS measurements of fully deamidated GII.4 Saga P dimer with 10 mM HBGA B trisaccharide under nearly identical conditions [19]. A possible explanation could be that HBGA B trisaccharide concentration was too

low to induce the observed effects because of decreased binding affinity. Notably, NMR measurements of fully deamidated GII.4 Saga P dimers with HBGA B trisaccharide and fucose show large chemical shift perturbations around residues 370-380 [19], a region where we observe increased deuteration in presence of fucose.

We hypothesize that N373 deamidation serves as a pH and temperature dependent mechanism to control infectivity of the virus. P dimers and VLPs have been shown to be stable under low pH conditions and temperature [63], where deamidation rate is low [19]. After entering the human host via contaminated food and reaching the intestine, the rise in pH and temperature facilitates conversion of pure wildtype to partially deamidated P dimers that are still able to attach to glycans and perform the structural change potentially required for interaction with the receptor and infection of the target cell. This theory is supported by the observation of increased flexibility in the P2 cleft of iDiD GII.4 Saga P dimers compared to the wildtype [19], which is also present in GII.4 MI001 P dimers. In summary, this could mean that deamidation creates the required flexibility for host cell attachment and subsequent receptor binding. Attenuation of glycan binding in the deamidated P dimer could be counteracted by avidity due to high glycan presentation on cell surfaces *in vivo*. Native MS measurements of deamidated GII.4 Saga and MI001 P dimers also show that with increasing deamidation, dissociation into monomers occurs, whereas in NN P dimers no monomers are present (Figure S10). This could as well be linked to the increased flexibility of iDiD P dimers that weaken the dimer interface and shift the monomer-dimer equilibrium. It would be interesting to investigate whether increased flexibility is limited to the P domain or whether it is propagated into the S domain in VLPs as well, which as a result could destabilize the particle and prepare for uncoating.

The question remains, which advantage the evolutionary conserved N373 deamidation site provides for the most prevalent GII.4 strains over other strains. One possibility is that higher flexibility induced by deamidation indeed enables better interactions with host receptors; another possibility is that it is part of an immune escape mechanism. N373 is located in the

753 immunodominant antibody epitope A and even minor changes in the epitope sequence during
754 viral evolution have resulted in the loss of monoclonal antibody response [64]. From all residues
755 in this specific epitope, N373 seems to be highly conserved over time.
756 On the other hand, prevalence of GII.17 strains increased over the last years [4], and based on its
757 D377 sequence the GII.17 Kawasaki P dimer is not able to deamidate at this position. Interestingly,
758 a N373D mutated GII.4 Saga P dimer shows similar affinities to glycans as the N373 wildtype [49]
759 clearly illustrating that iso-aspartate is required at this position to induce the observed changes.
760 The absence of iso-aspartate formation resulting from spontaneous deamidation could also
761 increase stability under a wide range of pH conditions, as dissociation into monomers is less likely
762 to occur. Increased stability under alkaline conditions has been seen for GII.17 Kawasaki VLPs,
763 however, other strains without potential deamidation sites were less stable at alkaline pH [65, 66].
764 GII.10 Vietnam carries a glutamine at the equivalent position 384, which in theory can deamidate
765 but deamidation is much slower and has not been observed after one year and four months of
766 storage at 5°C and pH 7 [19]. Nevertheless, GII.10 Vietnam clearly displays gain of structure upon
767 glycan binding, which may cause the observed long-range effects. The larger structural dynamics
768 could therefore be linked to cellular uptake.
769 Further research is required to clarify the role of N373 deamidation in the norovirus infection
770 process. Therefore, research focus should be shifted from wildtype P dimers alone to the more
771 likely occurring mixture of wildtype and partially deamidated P dimers to elucidate the potentially
772 important role of deamidation in the infection process. Furthermore, glycan binding studies with
773 wildtype and partially deamidated VLPs will give further information about the propagation of
774 structural changes throughout the capsid.
775

Acknowledgements

J.D. and C.U. would like to thank Prof. H. Schlüter (UKE, University of Hamburg, Germany) for access to high resolution mass spectrometers and Dr. Daniel Kavan, Dr. Petr Man and Dr. Alan Kádek for providing the DeutEx software. J.D. and C.U. acknowledge funding from FOR2327 ViroCarb. C.U. acknowledges funding from the Leibniz Association through SAW-2014-HPI-4 grant. The Heinrich-Pette-Institute, Leibniz Institute for Experimental Virology is supported by the Free and Hanseatic City of Hamburg and the Federal Ministry of Health. We thank Dr. Grant Hansman (University of Heidelberg, Germany) for providing us with the plasmids of the P domains of GII.4 Saga, GII.10 Vietnam 026 and GII.17 Kawasaki 308. Prof. Dr. Stefan Taube (University of Lübeck, Germany) is thanked for providing the plasmid of the P domain of GII.4 MI001. Computational resources were supplied through the Swedish National Infrastructure for Computing (SNIC) through the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) and the High Performance Computing Center North (HPC2N) under project SNIC 2019/4-554 and 2020/5-100. C.C. acknowledges the Helmholtz Association through the CFEL at DESY and the Swedish Research Council. This work was supported by MS SPIDOC within the European Union's Horizon 2020 research and innovation program under grant agreement No. 801406.

References

1. Ahmed, S.M., et al., *Global prevalence of norovirus in cases of gastroenteritis: a systematic review and meta-analysis.* The Lancet infectious diseases, 2014. **14**(8): p. 725-730.
2. Chhabra, P., et al., *Updated classification of norovirus genogroups and genotypes.* J Gen Virol, 2019. **100**(10): p. 1393-1406.
3. Chan, M.C.W., et al., *Global Spread of Norovirus GII.17 Kawasaki 308, 2014-2016.* Emerg Infect Dis, 2017. **23**(8): p. 1359-1354.
4. de Graaf, M., et al., *Emergence of a novel GII.17 norovirus - End of the GII.4 era?* Eurosurveillance, 2015. **20**(26): p. 8-15.
5. Xue, L., et al., *The resurgence of the norovirus GII.4 variant associated with sporadic gastroenteritis in the post-GII.17 period in South China, 2015 to 2017.* BMC Infect Dis, 2019. **19**(1): p. 696.
6. Prasad, B.V., et al., *X-ray crystallographic structure of the Norwalk virus capsid.* Science, 1999. **286**(5438): p. 287-290.
7. Cao, S., et al., *Structural Basis for the Recognition of Blood Group Trisaccharides by Norovirus.* Journal of Virology, 2007. **81**(11): p. 5949-5957.
8. Taube, S., A. Mallagaray, and T. Peters, *Norovirus, glycans and attachment.* Curr Opin Virol, 2018. **31**: p. 33-42.
9. Bucher, K.S., et al., *Fucose-Functionalized Precision Glycomacromolecules Targeting Human Norovirus Capsid Protein.* Biomacromolecules, 2018. **19**(9): p. 3714-3724.
10. Hansman, G.S., et al., *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.* J Virol, 2011. **85**(13): p. 6687-701.
11. Koromyslova, A., et al., *Human norovirus inhibition by a human milk oligosaccharide.* Virology, 2017. **508**: p. 81-89.
12. Koromyslova, A.D., et al., *The sweet quartet: Binding of fucose to the norovirus capsid.* Virology, 2015. **483**: p. 203-8.
13. Mallagaray, A., et al., *Attachment of Norovirus to Histo Blood Group Antigens: A Cooperative Multistep Process.* Angewandte Chemie-International Edition, 2015. **54**(41): p. 12014-12019.
14. Mallagaray, A., et al., *Saturation transfer difference nuclear magnetic resonance titrations reveal complex multistep-binding of l-fucose to norovirus particles.* Glycobiology, 2016.
15. Wegener, H., et al., *Human norovirus GII.4(MI001) P dimer binds fucosylated and sialylated carbohydrates.* Glycobiology, 2017. **27**(11): p. 1027-1037.
16. Han, L., et al., *Affinities of recombinant norovirus P dimers for human blood group antigens.* Glycobiology, 2012. **23**(3): p. 276-285.
17. Han, L., et al., *Quantifying the binding stoichiometry and affinity of histo-blood group antigen oligosaccharides for human noroviruses.* Glycobiology, 2018. **28**(7): p. 488-498.
18. Singh, B.K., M.M. Leuthold, and G.S. Hansman, *Human noroviruses' fondness for histo-blood group antigens.* J Virol, 2015. **89**(4): p. 2024-40.
19. Mallagaray, A., et al., *A post-translational modification of human Norovirus capsid protein attenuates glycan binding.* Nat Commun, 2019. **10**(1): p. 1320.
20. Dülfer, J., et al., *Structural mass spectrometry goes viral* in *Virus Structure and Function*, F.A. Rey, Editor. 2019, Elsevier.
21. Taube, S., et al., *A mouse model for human norovirus.* mBio, 2013. **4**(4).
22. Tyanova, S., T. Temu, and J. Cox, *The MaxQuant computational platform for mass spectrometry-based shotgun proteomics.* Nat Protoc, 2016. **11**(12): p. 2301-2319.

23. Hageman, T.S. and D.D. Weis, *Reliable Identification of Significant Differences in Differential Hydrogen Exchange-Mass Spectrometry Measurements Using a Hybrid Significance Testing Approach.* Anal Chem, 2019. **91**(13): p. 8008-8016.
24. Houde, D., S.A. Berkowitz, and J.R. Engen, *The utility of hydrogen/deuterium exchange mass spectrometry in biopharmaceutical comparability studies.* J Pharm Sci, 2011. **100**(6): p. 2071-86.
25. Arora, J., et al., *Hydrogen exchange mass spectrometry reveals protein interfaces and distant dynamic coupling effects during the reversible self-association of an IgG1 monoclonal antibody.* MAbs, 2015. **7**(3): p. 525-39.
26. Guttman, M., et al., *Analysis of overlapped and noisy hydrogen/deuterium exchange mass spectra.* J Am Soc Mass Spectrom, 2013. **24**(12): p. 1906-12.
27. Kavan, D. and P. Man, *MSTools—Web based application for visualization and presentation of HXMS data.* International Journal of Mass Spectrometry, 2011. **302**(1): p. 53-58.
28. Masson, G.R., et al., *Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments.* Nat Methods, 2019. **16**(7): p. 595-602.
29. Tommaso, P., et al., *T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension.* Nucleic Acids Research, 2011. **39**: p. W13-W17.
30. Waterhouse, A.M., et al., *Jalview Version 2--a multiple sequence alignment editor and analysis workbench.* Bioinformatics, 2009. **25**(9): p. 1189-91.
31. Waterhouse, A., et al., *SWISS-MODEL: homology modelling of protein structures and complexes.* Nucleic Acids Res, 2018. **46**(W1): p. W296-W303.
32. Bertoni, M., et al., *Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology.* Sci Rep, 2017. **7**(1): p. 10480.
33. Camacho, C., et al., *BLAST+: architecture and applications.* BMC Bioinformatics, 2009. **10**: p. 421.
34. Remmert, M., et al., *HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment.* Nat Methods, 2011. **9**(2): p. 173-5.
35. Benkert, P., M. Biasini, and T. Schwede, *Toward the estimation of the absolute quality of individual protein structure models.* Bioinformatics, 2011. **27**(3): p. 343-50.
36. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-12.
37. Lindahl, E., et al. *GROMACS 2019.1 Manual (Version 2019.1)*. 2019. DOI: http://doi.org/10.5281/zenodo.2564761.
38. Hornak, V., et al., *Comparison of multiple Amber force fields and development of improved protein backbone parameters.* Proteins, 2006. **65**(3): p. 712-25.
39. Paissoni, C., et al., *A critical assessment of force field accuracy against NMR data for cyclic peptides containing beta-amino acids.* Phys Chem Chem Phys, 2018. **20**(23): p. 15807-15816.
40. Larsson, P., R.C. Kneiszl, and E.G. Marklund, *MkVsites: A tool for creating GROMACS virtual sites parameters to increase performance in all-atom molecular dynamics simulations.* J Comput Chem, 2020. **41**(16): p. 1564-1569.
41. Mahoney, M.W. and W.L. Jorgensen, *A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions.* Journal of Chemical Physics, 2000. **112**(20): p. 8910-8922.
42. Bussi, G., D. Donadio, and M. Parrinello, *Canonical sampling through velocity rescaling.* Journal of Chemical Physics, 2007. **126**(1).
43. Nose, S. and M.L. Klein, *Constant Pressure Molecular-Dynamics for Molecular-Systems.* Molecular Physics, 1983. **50**(5): p. 1055-1076.
44. Parrinello, M. and A. Rahman, *Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method.* Journal of Applied Physics, 1981. **52**(12): p. 7182-7190.

45. Allen, M. and D. Tildesley, *Computer Simulations of Liquids*. 1987, Oxford, U.K.: Oxford Science Publications.
46. Darden, T., D. York, and L. Pedersen, *Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems.* Journal of Chemical Physics, 1993. **98**(12): p. 10089-10092.
47. Deutsch, E.W., et al., *The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition.* Nucleic Acids Res, 2017. **45**(D1): p. D1100-D1106.
48. Vizcaino, J.A., et al., *2016 update of the PRIDE database and its related tools.* Nucleic Acids Res, 2016. **44**(D1): p. D447-56.
49. Creutznacher, R., et al., *NMR experiments shed new light on glycan recognition by human and murine norovirus capsid proteins* submitted.
50. Hodge, E.A., M.A. Benhaim, and K.K. Lee, *Bridging protein structure, dynamics, and function using hydrogen/deuterium-exchange mass spectrometry.* Protein Sci, 2019.
51. Fang, J., et al., *False EX1 signatures caused by sample carryover during HX MS analyses.* Int J Mass Spectrom, 2011. **302**(1-3): p. 19-25.
52. Chen, M.W., et al., *Chikungunya virus nsP4 RNA-dependent RNA polymerase core domain displays detergent-sensitive primer extension and terminal adenylyltransferase activities.* Antiviral Res, 2017. **143**: p. 38-47.
53. Tan, M., et al., *Noroviral P particle: structure, function and applications in virus-host interaction.* Virology, 2008. **382**(1): p. 115-23.
54. Bereszczak, J.Z., et al., *Structure, stability and dynamics of norovirus P domain derived protein complexes studied by native mass spectrometry.* J Struct Biol, 2012. **177**(2): p. 273-82.
55. Tan, M. and X. Jiang, *The p domain of norovirus capsid protein forms a subviral particle that binds to histo-blood group antigen receptors.* J Virol, 2005. **79**(22): p. 14017-30.
56. Tan, M., J. Meller, and X. Jiang, *C-terminal arginine cluster is essential for receptor binding of norovirus capsid protein*. Journal of Virology, 2006. **80**(15): p. 7322-7331.
57. Tan, M., et al., *Terminal modifications of norovirus P domain resulted in a new type of subviral particles, the small P particles.* Virology, 2011. **410**(2): p. 345-52.
58. Hageman, T.S. and D.D. Weis, *A Structural Variant Approach for Establishing a Detection Limit in Differential Hydrogen Exchange-Mass Spectrometry Measurements.* Anal Chem, 2019. **91**(13): p. 8017-8024.
59. Singh, B.K., et al., *Structural Evolution of the Emerging 2014-2015 GII.17 Noroviruses.* J Virol, 2015. **90**(5): p. 2710-5.
60. Jung, J., et al., *High-resolution cryo-EM structures of outbreak strain human norovirus shells reveal size variations.* Proc Natl Acad Sci U S A, 2019. **116**(26): p. 12828-12832.
61. Smith, H.Q. and T.J. Smith, *The Dynamic Capsid Structures of the Noroviruses.* Viruses, 2019. **11**(3).
62. Tubiana, T., Y. Boulard, and S. Bressanelli, *Dynamics and asymmetry in the dimer of the norovirus major capsid protein.* PLoS One, 2017. **12**(7): p. e0182056.
63. Pogan, R., J. Dulfer, and C. Uetrecht, *Norovirus assembly and stability.* Curr Opin Virol, 2018. **31**: p. 59-65.
64. Mallory, M.L., et al., *GII.4 Human Norovirus: Surveying the Antigenic Landscape.* Viruses, 2019. **11**(2).
65. Pogan, R., et al., *Norovirus-like VP1 particles exhibit isolate dependent stability profiles.* J Phys Condens Matter, 2018. **30**(6): p. 064006.
66. Shoemaker, G.K., et al., *Norwalk virus assembly and stability monitored by mass spectrometry.* Mol Cell Proteomics, 2010. **9**(8): p. 1742-51.