2017

Structure of the chlorovirus PBCV-1 major capsid glycoprotein determined by combining crystallographic and carbohydrate molecular modeling approaches

Cristina De Castro
University of Napoli, decastro@unina.it

Thomas Klose
Purdue University, tklose@purdue.edu

Immacolata Speciale
University of Napoli

Rosa Lanzetta
University of Napoli Federico II

Antonio Molinaro
University of Napoli Federico II

See next page for additional authors

Structure of the chlorovirus PBCV-1 major capsid glycoprotein determined by combining crystallographic and carbohydrate molecular modeling approaches

Cristina De Castro[a,1], Thomas Klose[b], Immacolata Speciale[c], Rosa Lanzetta[c], Antonio Molinaro[c], James L. Van Etten[d,1], and Michael G. Rossmann[b,1]

[a]Department of Agricultural Sciences, University of Napoli, 80055 Portici, Italy; [b]Department of Biological Sciences, Purdue University, West Lafayette, IN 47907-2032; [c]Department of Chemical Sciences, University of Napoli, 80126, Napoli, Italy; and [d]Department of Plant Pathology and Nebraska Center for Virology, University of Nebraska, Lincoln, NE 68583-0900

Contributed by James L. Van Etten, November 28, 2017 (sent for review August 18, 2016; reviewed by Koichi Kato and Annabelle Varrot)

The glycans of the major capsid protein (Vp54) of *Paramecium bursaria* chlorella virus (PBCV-1) were recently described and found to be unusual. This prompted a reexamination of the previously reported Vp54 X-ray structure. A detailed description of the complete glycoprotein was achieved by combining crystallographic data with molecular modeling. The crystallographic data identified most of the monosaccharides located close to the protein backbone, but failed to detect those further from the glycosylation sites. Molecular modeling complemented this model by adding the missing monosaccharides and examined the conformational preference of the whole molecule, alone or within the crystallographic environment. Thus, combining X-ray crystallography with carbohydrate molecular modeling resulted in determining the complete glycosylated structure of a glycoprotein. In this case, it is the chlorovirus PBCV-1 major capsid protein.

virus PBCV-1 | capsid protein | N-glycans | glycoprotein structure | chloroviruses

Many viruses, such as rhabdoviruses, herpesviruses, poxviruses, and paramyxoviruses, have structural proteins that are glycosylated. Typically, viruses use host-encoded glycosyltransferases and glycosidases located in the endoplasmic reticulum and Golgi apparatus to add and remove *N*-linked sugar residues from virus glycoproteins either cotranslationally or shortly after translation of the protein. Thus, virus glycoproteins are host-specific (1–4).

One group of viruses that differs from this scenario is chloroviruses (family *Phycodnaviridae*) that infect eukaryotic algae (5). The prototype chlorovirus, *Paramecium bursaria* chlorella virus (PBCV-1), infects *Chlorella variabilis* NC64A, which is normally a symbiont in the protozoan *P. bursaria* (6). The PBCV-1 major capsid protein (hereafter named Vp54) has a predicted weight of 48,165 Da, but undergoes additional posttranslational modifications, so that the mature product has a molecular weight of ~53,790 Da (7).

Previous studies established that glycosylation of Vp54 is unusual and that the process probably occurs in the cytoplasm (5, 8). The structures of the four *N*-linked Vp54 glycans were recently determined (ref. 9 and Fig. 1*A*) and confirmed that PBCV-1 glycosylated Vp54 differently than other viruses in several aspects: none of the glycans were located in a typical Asn-X-(Thr/Ser) consensus site, the oligosaccharides were highly branched, β-glucose (Glc) was the site of attachment, there was a fucose (Fuc) substituted at all available positions, and each glycan contained two rhamnose residues with opposite configurations (abbreviated as L-Rha or D-Rha) plus a capped L-Rha with two O-methyl groups (diOMe-L-Rha). Two monosaccharides, arabinose (Ara) and mannose (Man), occurred as nonstoichiometric substituents, which resulted in four glycoforms (Fig. 1*A*), with two, glyco1 and glyco2, being the most abundant. Glyco1 is a nonasaccharide, it has all the monosaccharides except Ara, and it is the predominant form at Asn-302, Asn-399, and Asn-406. Glyco2 is a decasaccharide, it includes Ara, and it is the predominant form linked at Asn-280.

The structure of these *N*-glycans consists of two regions: the core region is located near the protein backbone and is highly conserved among the chloroviruses (10, 11). It is a type of *N*-glycan architecture and consists of the *N*-linked Glc, two xylose (Xyl) units [one located close (proximal unit or Xyl_{prox}) and one far (distal unit or Xyl_{dist}) from the protein backbone], the hyperbranched Fuc, and galactose (Gal) (Fig. 1*A*). The second region extends the conserved core with other monosaccharides, which are specific for each chlorovirus (10, 11). All these *N*-glycans are unique and do not resemble any known eukaryotic (12) or prokaryotic glycan (Bacterial Carbohydrate Structure database, csdb.glycoscience.ru/bacterial/).

To gain insight into the architecture of these glycans, we reexamined the previous X-ray diffraction data of the Vp54 crystals (13). In general, crystallographic analysis of carbohydrates is still a challenging task (14), and during the original study, the glycan structures were unknown; as a consequence, fitting potential glycans into the electron density data was performed using sugars now known to be incorrect. Glucosamine (not Glc) was placed near the sites of glycosylation, and Man was used to fill out the interpretable electron densities. However, reexamination of the Vp54 electron density data produced a structure that was compatible with the composition of the glycan

Significance

The three-dimensional structure of Vp54, the major capsid protein of the chlorovirus *Paramecium bursaria* chlorella virus (PBCV-1), was determined by combining two powerful techniques, X-ray diffraction and carbohydrate molecular modeling. This strategy resolved the limitations posed by each technique alone and increases our understanding of how glycans fold, their flexibility, and the type of interactions they exert within the protein component.

Author contributions: C.D.C., J.L.V.E., and M.G.R. designed research; T.K., I.S., and A.M. performed research; J.L.V.E. and M.G.R. contributed new reagents/analytic tools; C.D.C., T.K., I.S., R.L., and A.M. analyzed data; and C.D.C., T.K., J.L.V.E., and M.G.R. wrote the paper.

Reviewers: K.K., Okazaki Institute for Integrative Bioscience, National Institutes of Natural Sciences; and A.V., University of Grenoble Alpes, CNRS, Centre de Recherches sur les Macromolécules Végétales.

The authors declare no conflict of interest.

Data deposition: The atomic coordinates have been deposited in the Protein Data Bank, www.wwpdb.org (PDB ID codes 5TIP and 5TIQ).

[1]To whom correspondence may be addressed. Email: decastro@unina.it, jvanetten1@unl.edu, or mr@purdue.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1613432115/-/DCSupplemental.

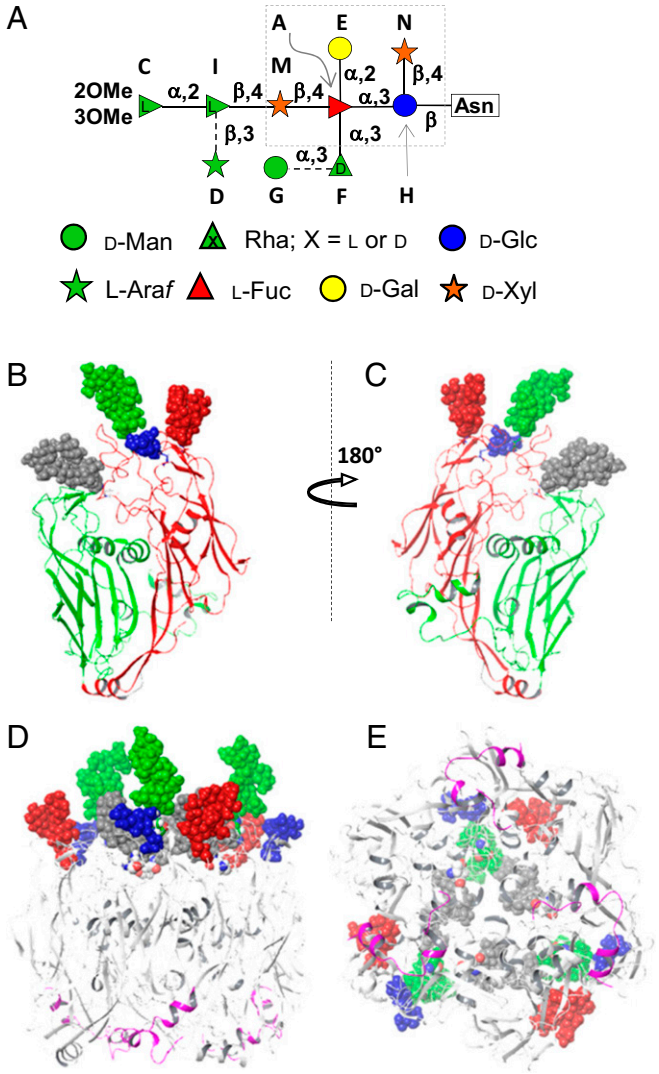

Fig. 1. (*A*) Structures of PBCV-1 Vp54 *N*-glycans (adapted from ref. 9). Each monosaccharide is denoted with the letter used during the NMR experiments, and residues inside the dotted box define the conserved core region. Ara and Man are not stoichiometric substituents and create four different glycoforms; their glycosidic linkages are denoted with a broken line. Glyco1 and glyco2 are the most abundant glycoforms: both have Man, glyco1 lacks Ara that is present in glyco2. (*B* and *C*) Different views of PBCV-1 Vp54 monomer (5TIP). The two jelly-roll domains, D1 and D2, are colored in green and red, respectively. Space-filling representation is used for the glycans, and their atoms are colored according to the residue they are attached to (Asn-280, green; Asn-302, gray; Asn-399, red; Asn-406, blue). Different views of the Vp54 trimer (pseudo hexamer) are given in (*D*) lateral view and (*E*) bottom view. The N-terminal 24 AAs are highlighted in pink.

structures, although not all the residues were detected. A carbohydrate molecular modeling (MM) approach provided the missing information, so that the combined use of X-ray crystallography and MM overcame the limitations presented by each approach alone. Together, these procedures allowed us to obtain a complete 3D description of the chlorovirus PBCV-1 glycosylated Vp54.

Results and Discussion

Reexamination of the Vp54 Diffraction Data. The original structures that were deposited in the protein database (PDB accession numbers 1M3Y and 1J5Q) were revised on the basis of the correct *N*-glycan structures and contained four (5TIP) and two (5TIQ) independent polypeptide chains in the crystallographic asymmetric unit. The data extended to 2.0- and 2.5-Å resolution, respectively. In addition, a cryoEM structure of the whole virus has also been reported (15). Reevaluation of the Vp54 electron density data (statistics in *SI Appendix*, Table S1) indicated that there was no density that supported O-glycosylation at either Ser-57 or Ser-387, contrary to what was previously reported (13). The calculated molecular weight of the revised structure (53,640 Da) is close to the Vp54 molecular weight determined by mass spectrometry (53,790 Da) (7) and confirmed our experimental work, which failed to detect O-linked glycans.

The fold of the protein (Fig. 1 *B* and *C*) consists of two consecutive "jelly-roll" domains, D1 (residues 2–212) and D2 (residues 225–437), linked by an α helix. The refined crystal structures also contained interpretable density for the first 24 amino acids (AAs) of Vp54 for each crystallographic independent polypeptide. No such density was reported in the original structure determination (13). In each crystal structure, the first 24 AA residues form two short α-helices connected by a linker. Three Vp54s form a trimeric capsomer, which has pseudosix-fold symmetry, and these first 24 AAs are located at the internal face of the Vp54 trimer (Fig. 1 *D* and *E*).

A comparison of the revised crystal structures with the averaged cryoEM electron density (15) of a Vp54 trimer suggests these helices are slightly rearranged in the mature virus and may help stabilize the interactions between neighboring capsomers and minor capsid proteins located on the inner surface of Vp54. Indeed, capsomers are further organized into more complex architectures, known as trisymmetrons and pentasymmetrons, containing 66 and 30 copies of the trimer, respectively (13), and contributing to the hycosahedral shape of the virus. Knowledge of the correct chemical structures for the *N*-glycans (9) permitted the adoption of the correct monosaccharide templates, and more residues were detected (*SI Appendix*, Table S2, and electron densities in *SI Appendix*, Fig. S1).

Taking the best-resolved structure into account, 5TIP, and considering chain A as an example, the glycan at Asn-302 contained the complete oligosaccharide; at Asn-280, only Ara could not be modeled (Fig. 1*A*), although some additional density was located in this area. For the Asn-399 glycan, two of the nine monosaccharides were not detected, whereas at Asn-406, only Glc and Fuc were observed (*SI Appendix*, Table S2). Glycans at the other three chains of Vp54 had similar dihedral angle values (*SI Appendix*, Table S2), although more residues were missing, such as diOMe-L-Rha of the glycan at Asn-280 of all of the other chains, or Fuc of the glycan at Asn-406 of chain C.

As for chains A and B of the X-ray structure with lower resolution, 5TIQ (*SI Appendix*, Table S2), the glycan at Asn-280 was missing the Ara residue, the complete nonasaccharide was detected at Asn-302, and only Glc and Fuc were visible for the glycan at Asn-406. As for Asn-399, chain A presented an oligosaccharide missing the two L-Rha residues, whereas at chain B, the oligosaccharide was complete.

Arrangement of the Glycan at the Glycoprotein Surface. To understand the conformation of the glycan, the best-resolved X-ray structure, 5TIP, was analyzed by inspecting one protein chain. During this analysis, we did not consider the hydrogens because they were not used during the refinement of the X-ray structures; therefore, evaluation of hydrogen bonds (HBs) was based on the distances measured between oxygen and nitrogen (or oxygen) atoms. Chain A was considered because it had the major number of monosaccharide residues resolved (*SI Appendix*, Table S2). The four *N*-glycans were located on the same side of the protein, with their terminal parts pointing outward while their inner regions interacted with either the protein backbone or with other *N*-glycans (Fig. 1 *B* and *C*).

Regarding the first glycan, Asn-280 is held in position by three different HBs with Gly-289, Ala-390, and Thr-391 (Fig. 2*A*); Fuc

and the lower face of the *N*-linked Glc face the loop defined from AAs 289–291 (Fig. 2*A*); the other side of Glc along with the Xyl_{dist}, D-Rha, and the units attached to them, extends to regions devoid of other residues (Fig. 2*B*), except Gal, which points to Glc linked at Asn-406.

The glycan at Asn-302 instead is embedded between the two domains of the protein (Fig. 2*C*) facing the loops of the D1 domain defined from AAs 81–83 and 136–140, and some of the AAs of the long segment from 284 to 295 (part of the D2 domain), which has the *N*-glycan at Asn-280 on the other side. The glycan at Asn-302 is held in position by different sets of HBs (Fig. 2*D*), with a key role played by Asp-299, which has two HBs with Asn-302 and one with OH2 of Glc; one additional HB instead involves Gly-287 and OH4 of Gal. As a result, some of the residues of this glycan (Glc, Fuc, Gal, and to a lesser extent Xyl_{dist}, D-Rha, and Man) are inserted between D2–D1 domains of the protein (Fig. 2*C*), whereas Xyl_{prox}, L-Rha and diOMe-Rha are located in a region free of other residues (Fig. 2 *C* and *E*). As for the glycan at Asn-399, the disaccharide diOMe-α-L-Rha-(1→3)-β-L-Rha is not detected; most of the other residues are located in an unoccupied region of the space (Fig. 2*F*). Orientation of the inner part of this glycan is fixed by two HBs occurring between Asn-399 with Ala-395 and Thr-400, and two other HBs involve Glc-OH2 and Gly-398, and Gal-OH3 and Ala-394 (Fig. 2*F*).

As for Asn-406, most of the monosaccharides are not detected except for Glc and Fuc, suggesting most of the glycan is in a region devoid of other residues. The only unit that is fixed in its position is Asn, which has three different HBs with Ile-388, Ala-390, and Thr-409 (Fig. 2*G*).

Analysis of chain A did not explain why the terminal parts of the glycans at Asn-280 and Asn-302 were almost entirely visible, in contrast to those at Asn-399 and Asn-406. Accordingly, we examined the crystallographic data further searching for other elements that could influence the flexibility of the glycans. To this end, we examined the mates in the crystal lattice within 10 Å of the glycans of chain A and found four glycoproteins, referred to as mate-2 to mate-5 (mate-1 is chain A). Mate-2 and mate-3 are symmetric to chain A in 5TIP, and together form the capsomeric unit, whereas mate-4 and mate-5 are symmetric with other chains in the asymmetric unit and approached chain A from the side or from the top (Fig. 3), respectively. These five mates will be referred to as assembly from here on.

In this more comprehensive view, the glycan at Asn-280 of chain A is flanked by the glycans of mate-3 and mate-5 (*SI Appendix*, Fig. S2). The best proximity occurs between Xyl_{prox} and the L-Rha disaccharide at Asn-302 of mate-3 (ca. 4 Å), whereas the terminal part extends toward the L-Rha-Xyl_{dist} fragment at Asn-280 of mate-5, with a distance of ∼7 Å. Thus, the terminal residues of the glycan at Asn-280 of chain A expand into a region of space occupied by other molecules, with the effect that their overall flexibility is limited.

The inner part of the glycan at Asn-302 of chain A is inserted within the polypeptide chain, and its terminal part approaches the glycans of mate-2 (*SI Appendix*, Figs. S2 and S3); indeed, it is rather close to that at Asn-302 (∼5 Å), while further away from that at Asn-399 (>11 Å; *SI Appendix*, Figs. S2 and S3).

As for Asn-399 of chain A (*SI Appendix*, Fig. S4), the inner part intercalates among mate-4 and mate-3. Mate-4 is more than 10 Å away, whereas AAs Val-141 and Leu-140 of mate-3 are close to the Gal (∼2.5 Å) and Man (∼7 Å) units of the glycan. Finally, Xyl_{dist} extends into a free region of space, making it likely that the disaccharide diOMe-Rha-L-Rha attached at its

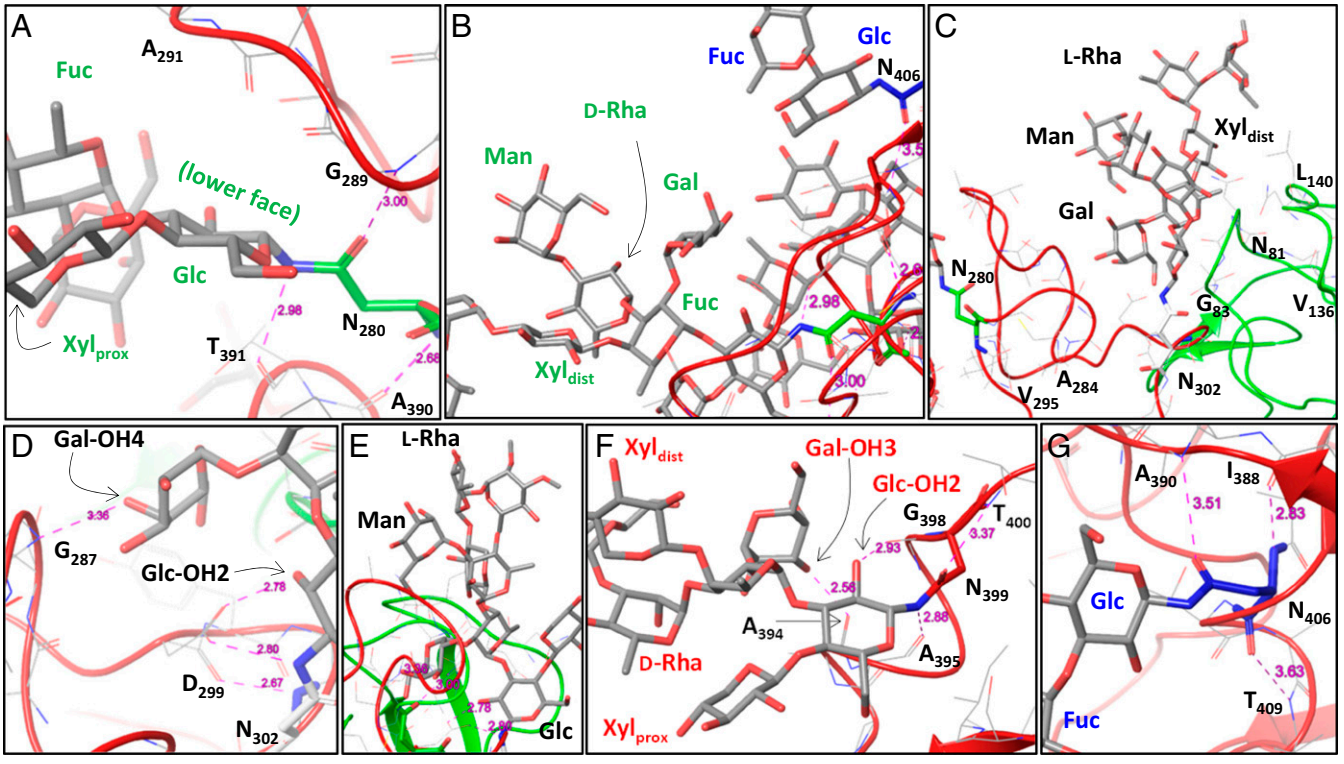

Fig. 2. Views of 5TIP detailing the proximities existing between the *N*-glycans and the protein backbone. To avoid crowding and simplify the vision, all hydrogens and AAs further than 3 Å from any glycan are excluded. (*A* and *B*) Different views of Asn-280 glycan; (*C*–*E*) details of the glycan at Asn-302, Val-295, and Ala-284 are included even if they are further than 3 Å from the glycan because they are important for the discussion. (*F*) Expansion of the inner part of the glycan at Asn-399 and at (*G*) Asn-406. The color used to label the monosaccharides of the *N*-glycans depends on the Asn that is glycosylated: Asn-280, green; Asn-302, black; Asn-399, red; Asn-406, blue. AAs have a black label.

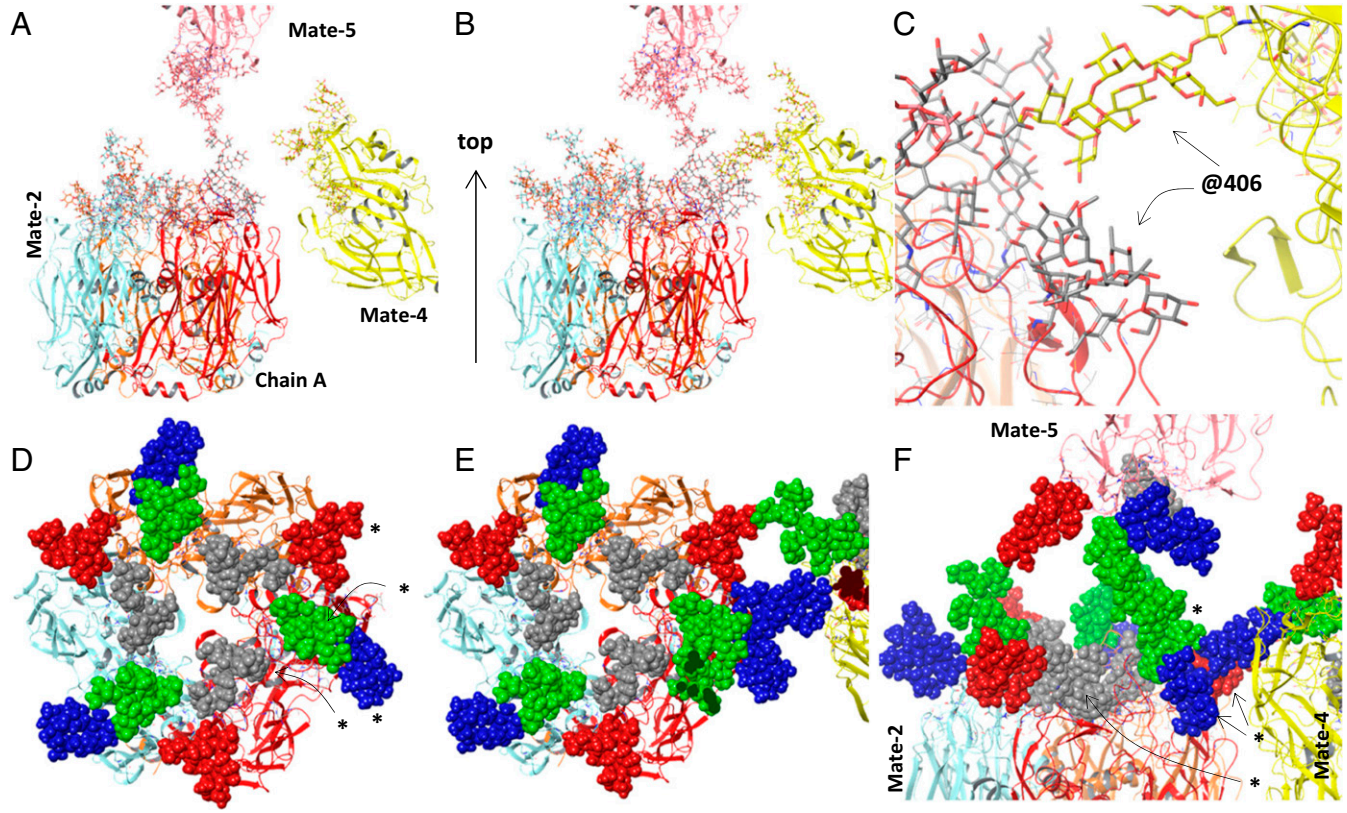

Fig. 3. (*A*) Assembly made from 5TIP by selecting the mates of chain A located within 10 Å of the glycan moieties. (*B–F*) Different views of the conformer at the global energy minimum obtained by the MCMM conformational search. In particular: (*C*) expansion of the region occupied by the glycans at Asn-406 of chain A and mate-4. (*D*) Top view of the capsomer part of the MCMM assembly, mate-4 and mate-5 are omitted to better visualize the glycans. (*E*) As in *D*, but including mate-4 and mate-5. (*F*) Lateral view of the assembly detailing the arrangement of the glycans. In all panels, chain A (mate-1) is depicted in red, mate-2 (cyan) and mate-3 (orange) define the capsomeric unit, and mate-4 (yellow) and mate-5 (pink) are located at the side and at the top of chain A, respectively. In *D–F*, monosaccharides are represented in the space filling mode, and a color code is used to quickly visualize them. Glycan at Asn-280, green; Asn-302, gray; Asn-399, red; Asn-406, blue. Glycans labeled with a star (*) are attached to chain A.

position 4 can move without restrictions and is thus too flexible to be detected in the X-ray structure.

As for the last glycosylated position, Asn-406 (*SI Appendix*, Fig. S5), the small fragment of the glycan, is located in a region of space far from any other crystal mate, supporting the hypothesis that the other monosaccharides have large conformational freedom that results in no reasonable density in the X-ray structure.

In conclusion, the conformational space accessible to the glycans depends on intramolecular and intermolecular factors. Indeed, glycans at Asn-280, Asn-302, and Asn-399 face other regions of the same glycoprotein or other mates, whereas the same situation does not exist for the oligosaccharide at Asn-406, the only one poorly detected and that exists in a void region of space.

Conformation of the *N*-Glycans in the Free Form, Evaluation of the Most Appropriate Force Field. To add the missing monosaccharides to the crystallographic structure and to determine the conformation of all of the glycans, a MM approach was used. Assembling the free oligosaccharides was performed first (*Materials and Methods*), and the molecular mechanic approach was applied to identify the preferred dihedral angles adopted at each glycosidic junction. This approach dissects the entire oligosaccharide into many disaccharide entities and evaluates the orientation for each of the two monosaccharides that is compatible with their relative stereochemistry along with the exo-anomeric effect (16). This initial screening does not take into account possible restrictions that may arise when a disaccharide is part of a more complex oligosaccharide, as in glyco1 or glyco2 (Fig. 1*A*), but it is useful to identify an initial set of dihedral angles (Φ/Ψ) to build the molecule. In essence, two sets of "Ramachandran plots" were constructed by using the force fields (FF), MM3 and Amber, suitable for carbohydrates. For each glycosidic junction, the best Φ/Ψ values were deduced (*SI Appendix*, Table S3, and graphics are reported only for Amber maps in *SI Appendix*, Fig. S6) and used to build the two complex *N*-glycans, glyco1 and glyco2, not bound to the protein and with the Glc *N*-linked at an Asn residue.

To evaluate which FF was more appropriate, the dynamic behavior of each oligosaccharide was inspected by running canonical molecular dynamic (MD) simulations with both FFs and then validating the results with experimental NMR data.

Canonical MD was selected in place of replica exchange MD, a very accurate but computer power-demanding approach (17). Replica exchange MD succeeds in determining the types and proportions of the different conformers, avoiding their trapping in local minima, but the two approaches converge when simulations are performed at high temperatures, with 300 K considered the lower limit (17, 18). Hence, a NMR T-ROESY (transverse rotating-frame Overhauser enhancement spectroscopy) spectrum was recorded at 310 K, so that each MD simulation (310 K, 20 ns) reasonably sampled the space accessible to each oligosaccharide.

The averaged interproton distances derived from the ensemble of conformers was compared with the experimental NMR values deduced from the T-ROESY spectrum (*SI Appendix*, Table S4). In general, both FFs performed well, but simulations performed with Amber gave a better agreement with the experimental data for both glyco1 and glyco2, with a higher number of

predicted distances close to the experimental values (*SI Appendix*, Table S4). Thus, our experiments proceeded by first analyzing the output from the Amber simulations, and then using this information to build the complete glycoprotein.

Analysis of Φ/Ψ plots was performed for all residues, except the Glc-Asn linkage, because Asn was not connected to other AAs in the protein. Φ/Ψ plots (*SI Appendix*, Fig. S7) identified a rather large value distribution for each glycosidic linkage, giving an indication of the space accessible to each residue when the glycan was not attached to the protein. Comparison of these distributions with the values from the 5TIP data (*SI Appendix*, Fig. S7 and Table S2) showed that the crystallographic residues were within the conformational space allowed by MD simulation. On the basis of these results, an optimized conformer for glyco1 and glyco2 was built by using the X-ray Φ/Ψ values that were available and integrating those missing with those of the most populated conformation from the MD simulation (*SI Appendix*, Table S3). The final oligosaccharides were minimized to relieve any steric clash created by the manual setting of some dihedral values.

Montecarlo Multiple Minimum Conformational Search of Glycosylated Vp54. A fully glycosylated Vp54 was built with the aim of investigating whether the *N*-glycan conformations in 5TIP corresponded to the global minimum, or whether the process of packing the molecules in the crystal lattice selected one conformer over others with comparable energy.

The complete glycoprotein was constructed by using the protein backbone of chain A and replacing the *N*-glycans with the optimized glyco1 conformer at Asn-302, Asn-399, and Asn-406, and glyco2 at Asn-280. This structure was used as the starting model for the Montecarlo Multiple Minimum (MCMM) conformation search, which was made by varying all the glycan dihedrals, including those of the Glc-Asn junction. Energy minimization was performed with Amber FF, mimicking water with a dielectric constant ($\varepsilon = 80$), and was restricted to the carbohydrate moieties including the corresponding Asn residues. During the MCMM search, the protein backbone was kept in the conformation of the X-ray structure (see *Material and Methods* for details).

The similarity between chain A and the conformers generated from the search was rated by superposition and subsequent calculation of a rmsd value (where a low rmsd value means a better overlay). Inspection of rmsd and energy plots (*SI Appendix*, Fig. S8) indicated several species with low energy (60 kJ/mol above the global minimum; *SI Appendix*, Fig. S8 *A* and *B*), and other information was obtained from analyzing the rmsd frequency plot (*SI Appendix*, Fig. S8*C*). In general, rmsd varied up to 16 arbitrary units, and distribution of conformers between the different values gave rise to four main groups (*SI Appendix*, Fig. S8*C*). The first group was sparsely populated and included the conformers with the best rmsd values; group B was more populated than the first group, but its rmsd increased. Group C was the most populated and also included the conformer at the global minimum; however, the rmsd range of this group was consistently higher than that of the previous two. Finally, group D enclosed elements that had very high rmsd and that were also higher in energy (*SI Appendix*, Fig. S8*C*).

To understand the differences between these conformers, one representative of each group was analyzed. Conformer A was selected because of its very low rmsd value, conformer B pointed to the maximum of rmsd distribution of its group and had the lowest energy value, and conformer D was the lowest in energy of its group. After initial inspection of group C, some conformers (C, C1, and C2; *SI Appendix*, Fig. S8*B*) were found to be similar, and conformer C (the global minimum of the search) was taken as representative.

Then each conformer was compared with 5TIP; in general, most of the glycans occupied roughly the same location as in 5TIP (Fig. 4), with variations resulting from some tilting and/or rotation with respect to the analog residues in the reference structure. At Asn-280 (Fig. 4*A*), the Fuc of conformers C and D adopted an orientation different from the X-ray structure, which shifted the moiety comprising Xyl_{dist} up to Ara in a different area of the space (Fig. 4*A*). At Asn-302 instead (Fig. 4*B*), the Glc-Asn linkage of conformer D differed from that of the others and shifted the whole glycan in a different space region. As for Asn-399, the orientation of Fuc in conformer C differed from that of 5TIP; thus, Fuc and all the other sugars attached orientated differently from the X-ray structure (Fig. 4*C*). As for Asn-406 (Fig. 4*D*), all the conformers maintained the same Glc-Asn and Fuc-Glc orientation, but no other differences or similarities could be deduced, as this glycan was poorly detected in the X-ray structure. Thus, conformers A and B were very similar to chain A, and the two others had significant differences. The finding of different possible orientations with comparable energy posed two questions: Can they interconvert? And are they compatible with the crystal packing?

MCMM Conformer Interconversion and Compatibility with 5TIP Structure. To answer the first question, we ran MD simulations using each conformer, A, or C or D, as the starting structure. Conformer B was not considered because it was similar to conformer A. MD focused on the glycan part of the glycoprotein by setting the appropriate constrains on the protein backbone and leaving the four oligosaccharides and the glycosylated Asn free to move. The conformational behavior of the glycans of the three molecules was compared with that of the free *N*-glycans by inspecting the distribution of the dihedrals of each glycosidic connection.

As for conformer A, the Φ/Ψ distribution of all four *N*-glycans was within the boundaries defined by the oligosaccharide in the free form, but did not cover all the space originally accessible (*SI Appendix*, Fig. S9). This is especially evident for the fully substituted Fuc, for the terminal Gal, and to a lesser extent for the other residues. The best overlap between the two distributions occurred for the residues far from the protein backbone, indicating that their flexibility was similar to that of the unbound *N*-glycan. The Glc-Asn linkage is defined by four different dihedrals, but variation occurs for three of them, as the amide linkage ($\Psi = C_1NC_\gamma C_\beta$) prefers the *trans* geometry. Indeed, representation of a Glc-Asn linkage has only three variables, and its graph (*SI Appendix*, Fig. S10) denotes a unique distribution of dihedrals at each glycosylated site. The presence of two distributions for Asn-280–302 and Asn-406 is only apparent because some of their dihedrals (e.g., W and X for Asn-280 and Asn-302 in conformer A, respectively), are centered about 180°, and therefore, fluctuations include also negative values that appear in a different area of the graph.

As for conformer C, Glc-Asn junctions matched that of conformer A (*SI Appendix*, Fig. S10), whereas the same did not happen for the other glycosidic connections (*SI Appendix*, Fig. S11). At Asn-280, the dihedral distribution of Fuc was within the boundaries defined from the free glycan, but shifted to another region. More interestingly, at Asn-399, Fuc dihedrals were utterly different and located outside the region defined from the free glycan, but still at the edge of the boundaries identified in the flexible map of the Fuc-Glc disaccharide (*SI Appendix*, Fig. S6). Gal orientation at the Asn-406 glycan was similar to that at Asn-280, and both occupied a state less populated in the free *N*-glycan. No significant variations were detected in all the other residues.

As for conformer D, the most striking difference was in Glc-Asn at Asn-302 (*SI Appendix*, Fig. S10), which oriented the whole glycan in a different region of space during the entire time of the simulation. However, this difference did not substantially affect the *N*-glycan conformation because the dihedral values of the other residues (*SI Appendix*, Fig. S12) were similar to those

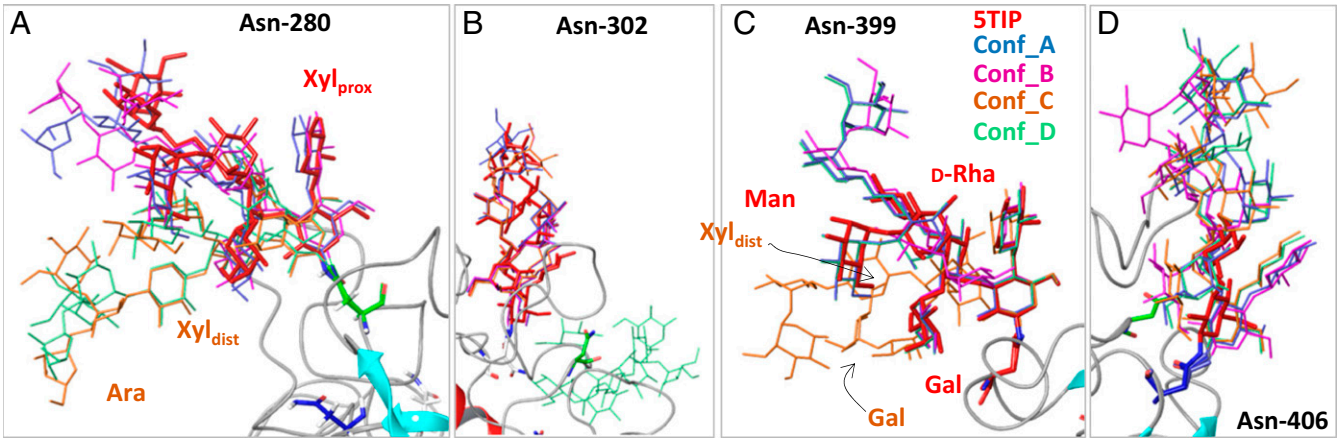

Fig. 4. Superposition of Vp54 conformers A–D found during the MCMM search. Chain A of 5TIP is used as reference and it is shown as a red thick tube; conformer A (Conf_A) is blue, conformer B (Conf_B) is pink, conformer C (Conf_C) is brown, and conformer D (Conf_D) is green. Some of the monosaccharides are labeled and the color used depends on the conformer considered. Each panel focuses on one glycosylated position, and to avoid crowding, other parts of the molecule are not visualized. (*A*) Glycans at Asn-280. (*B*) Glycans at Asn-302. (*C*) Glycans at Asn-399. (*D*) Glycans at Asn-406.

of the unbound glycan, with only some reduction in their overall flexibility. Of note, Gal at Asn-280 could explore different conformational states, as visible by the dual distribution of its dihedrals (*SI Appendix*, Fig. S12).

Thus, our first conclusion is that glycans at the surface of monomeric Vp54 can exist in different conformations, which do not interconvert because they are probably trapped in local minima under the simulation condition used.

Next, we determined whether these conformers were compatible with the crystal packing: We built three different assemblies, named assembly-A, assembly-C, and assembly-D, by replacing the five chains of the 10-Å assembly (Fig. 3*A*) with the corresponding MCMM conformers. As a result, the arrangement of the glycans on the three assemblies differed (*SI Appendix*, Fig. S13). The glycans of assembly-A (*SI Appendix*, Fig. S13*B*) occupied almost the same space as in crystallographic assembly (*SI Appendix*, Fig. S13 *A* or *A′*). The only obvious difference is that assembly-A has the four complete oligosaccharide chains. As for assembly-C and assembly-D (*SI Appendix*, Fig. S13 *C* and *D*), the glycans appear as flattened on the surface of the capsomeric unit, as evidenced by the void region located under mate-5. In addition, the glycans of assembly-C and assembly-D heavily collapsed with other parts of the assembly, whereas for assembly-A, clashes were minimal. Therefore, to relieve the clashes, minimization restricted to the glycans and to the connecting Asns were performed and produced conformers devoid of steric clashes (*SI Appendix*, Fig. S13 *C′* and *D′*), but still different from the crystallographic assembly (*SI Appendix*, Fig. S13 *A* and *A′*). Last, the three minimized assemblies were ranked according to their energies calculated over all of the atoms, and assembly-A (E = 58,461 kJ/mol), the one more similar to the crystallographic structure, had a stability intermediate to that of assembly-C (E = 58,226 kJ/mol) and assembly-D (E = 58,490 kJ/mol).

Thus, our second conclusion is that different conformations of the glycans are compatible with the crystal lattice, even though the crystallization process selects one over the many conformers close to the energy minimum.

Arrangement of *N*-Glycans at the Vp54 Surface. The final objective was to define the best position of the monosaccharides that escaped the X-ray detection. This task was performed by analyzing chain A together with its four mates; the choice of such a large system was necessary to create around the glycans of chain A the same space constrictions of the crystallographic unit cell. This assembly was generated by adjusting the glycan dihedrals of assembly-A to those of the crystallographic structure (*SI Appendix*, Table S2) and complementing the values not available with those of the MD of the free glycans (*SI Appendix*, Table S3).

This conformer presented few clashes between the residues undetected in the X-ray structure, which were relieved by a minimization restricted to them; namely, the two L-Rha units at Asn-399; all the residues at Asn-406, except Glc and Fuc; and both Ara and diOMe-L-Rha at Asn-280. This last residue was included in the minimization because of the poor density surrounding it in 5TIP; it appeared only in chain A and not in the others. A MCMM conformational search was applied on this minimized assembly; this approach was preferred to MD simulation to avoid the possibility that the system could be trapped in a local minimum, as observed for conformers A, C, and D.

The MCMM search considered the glycans of the five mates and focused on the glycosidic linkages of the monosaccharides not detected in the X-ray structure, the same selected for the minimization. The analysis of the output focused on the glycans of chain A and on the structures within 10 kcal/mol from the global minimum to pinpoint the features of the most stable conformations (*SI Appendix*, Figs. S14–S16).

At the Asn-280 glycan, Ara more than diOMe-L-Rha can adopt different orientations (*SI Appendix*, Fig. S14*A*), with the most representative reported in *SI Appendix*, Fig. S14 *B–H* (the Ara and diOMe-Rha position is denoted with a green and turquoise circle, respectively); it is worth noting that when seen in the context of the whole assembly, the glycan at Asn-280 extends toward the homolog chain of mate-5 and close proximities occur between diOMe-L-Rha and Ara of each unit (*SI Appendix*, Fig. S14 *B′–H′*). Indeed, both Ara and diOMe-L-Rha suffer some restriction in their flexibility, in agreement with the finding that this area of the crystallographic structure contains partial densities for diOMe-L-Rha (chain A), along with other fragmented densities that could not be attributed to Ara with confidence.

Regarding the glycan at Asn-399, the terminal disaccharide diOMe-Rha-(1→2)-L-Rha can assume many different orientations, as summarized in the fuzzy superimposition of all the low-energy conformers (*SI Appendix*, Fig. S15*A*). For clarity, some representative conformers are reported alone (*SI Appendix*, Fig. S15 *B–I*, the diOMe-Rha position is denoted with a green circle), and their visualization, together with the other mates, shows that depending on the conformation adopted, this terminal disaccharide can be less than 5 Å from the Ara unit of mate-4 (*SI Appendix*, Fig. S15 *B′–I′*).

The glycan at Asn-406, the one with more dihedrals included in the search, can exist in different conformations (*SI Appendix*, Fig. S16 *A* or *A′*), with the most relevant reported in *SI Appendix*, Fig. S16 *B–I* or *B′–I′*. First, the Xyl_{prox} has one main orientation (*SI Appendix*, Fig. S16*A*), and it appears at less than 5 Å from the glycan at Asn-280 of chain A and that at Asn-406 of mate 4. The two other terminal segments, Man-(1→3)-D-Rha and diOMe-Rha-(1→2)-L-Rha-(1→4)-Xyl, instead are more flexible. The orientation of Man-(1→3)-D-Rha is intermediate to that of the Xyl_{prox} and to the other fragment; the main variation is limited to some tilting and rotation of the two units (*SI Appendix*, Fig. S16 *B–I*, Man position is denoted with a turquoise circle). In contrast, the trisaccharide diOMe-Rha-(1→2)-L-Rha-(1→4)-Xyl explores large areas of the space, as appears following the position of its terminal moiety, diOMe-L-Rha (*SI Appendix*, Fig. S16 *B–I*, the diOMe-L-Rha position is denoted with a green circle). Overall, apart from the few loose proximities found between the Xyl_{prox} and other oligosaccharides, the rest of this glycan can exist in many different conformations, with no impediment from residues from chain A or its mates, explaining why they are not detected in the X-ray structure.

Finally, the assembly with the lowest energy of the MCMM search was taken as representative of the most likely conformation of the glycans (Fig. 3*B*), and attention was paid mostly to chain A to appreciate further details. Indeed, compared with the crystallographic assembly (Fig. 3*A*), the region between chain A and mate-4 (Fig. 3*B*) was not void but filled with the glycans at Asn-406 of chain A and mate-4 (Fig. 3*C*). This arrangement is also evident changing the perspective and watching the complex from the top, where it appears that the glycans at Asn-399 and Asn-406 protrude radially outward (Fig. 3*D*), approaching those at Asn-280 and Asn-406, respectively, of mate-4 (Fig. 3*E*). Similarly, the glycan at Asn-280 extends outside along the vertical axis of the capsomer, where it engages with the analog glycan of mate-5 (Fig. 3*F*). In contrast, the glycan at Asn-302 extends toward the center of the capsomer (Fig. 3*E*), where it appears deeply buried (Fig. 3*F*).

Thus, analysis of the assembly included information on the capsomer, the unit at the base of more complex elements (penta- and trisymmetrons), which in turn make up the whole capsid of PBCV-1 (15). Indeed, this information will pave the way to further studies aimed to understand whether and how the glycans stabilize the intra- and intercapsomer interactions necessary for the structural integrity of the capsid and, finally, which role, if any, the glycans play during the host–virus interaction process.

Conclusions

The 3-dimensional structure of Vp54, the major capsid protein of chlorovirus PBCV-1, was refined by applying two different approaches: X-ray crystallography and carbohydrate MM. Reexamining the original X-ray data fixed several aspects missing in a previous investigation (13). First, the 24 AAs at the N-terminal of the polypeptide backbone, minus Met, formed two α-helices that interconnect in the capsomer, a trimer of Vp54, presumably stabilizing its overall architecture (Fig. 1). The absence of the N-terminal AA, Met, confirmed previous results indicating that it is removed by a posttranslational event (5). Second, no density related to O-linked glycans was detected at Ser-57 or Ser-387. This result agrees with our recent unsuccessful attempts to detect O-linked glycans at these two Ser residues. The calculated MW (53,640) of Vp54 agrees reasonably well with the experimental value, [53,790 Da (7)]. Our calculated MW considers the occurrence of minor *N*-glycans without Man, and the discrepancy with the experimental value (150 Da) might be related to the formation of adducts with sodium or potassium ions during the MS measurement. Third, in light of the recently reported *N*-glycan structures (9), the choice of the right monosaccharide templates allowed modeling of most of the residues in the X-ray density maps. It was possible to detect the full nonasaccharide at Asn-302, nine of the 10 residues of the glycan at Asn-280, seven of the nine residues of the glycan at Asn-399, and only two residues of the nonasaccharide at Asn-406.

Analysis of the crystallographic environment of a single Vp54 molecule disclosed that the glycans at Asn-280, Asn-302, and Asn-399 are well resolved because they have some conformational restrictions that arise from contact with the same molecule (Fig. 2) and/or with the nearby crystal symmetry mates (*SI Appendix*, Figs. S2–S6).

Integration of the crystallographic data with a MM approach revealed other features. First, glycans can exist in multiple conformations (Fig. 4), all in principle compatible with the crystal packing (*SI Appendix*, Fig. S13), even though only one is selected during the crystallization process. Second, the space accessible to the residues undetected in the X-ray analysis is ample, and these residues can adopt many different conformations because they are not constricted by other components of the crystal lattice, thus escaping detection (*SI Appendix*, Figs. S14–S16).

Thus, the fully glycosylated structure of PBCV-1 Vp54 was determined by merging the power of two complimentary approaches, X-ray crystallography and MM, which showed aspects inaccessible for each of them separately. Here, the computation protocols and the analysis of the simulation data focus on the carbohydrate moiety, with the result that a complete view of the molecule is obtained, enabling us to appreciate features that would have otherwise been lost.

PBCV-1 glycosylation is important, as suggested by the biology of its spontaneous mutants, also termed antigenic variants (5, 8). These viruses have no mutations in the Vp54 gene, but the glycans are truncated at different positions (5, 8), and depending on the level of truncation, some variants are less stable during purification. Indeed, *N*-glycans seem to have a role in the capsid properties and assemblage, and the information about the Vp54 structure will enable us to investigate this aspect of the biology of this virus at the molecular level.

Materials and Methods

X-Ray Diffraction Data. The coordinates and structure factors for PBCV-1 Vp54 were retrieved from the PDB and evaluated by visual inspection in the program COOT. Sugars in these structures were removed and the chemically identified oligosaccharides placed into positive density of a difference map obtained from the refined crystal structure without any glycans. After refinement of this initial structure, which contained about 80% of the final glycans, monosaccharides were added into positive density of a calculated difference map and refined again until convergence. Restrains for unknown monosaccharides were generated with ACEDRG from the CCP4 suite. All coordinates were refined using phenix.refine from the PHENIX software suite, using torsion-angle restrains as well as custom dictionaries to define glycosidic bonds involving D-Xyl, whose nonstandard atom descriptions (e.g., C1B instead of C1) do not conform to predefined glycosidic bond definitions in the monomer library. The refinement also accounted for the twinning present in 5TIP, and TLS refinement was used with one TLS group per chain, excluding the glycans. The refined structures were checked with PRIVATEER (19, 20) and remodeled using COOT with the same custom dictionary. The process was repeated until all glycans conformed to the expected low-energy conformations. The final Rwork/Rfree factors were 0.18/0.21 and 0.18/0.23 for the structures, with PDB accession codes 1M3Y and 1J5Q, respectively. These rerefined structures have been deposited with the PDB with the accession code 5TIP (1M3Y replacement) and 5TIQ (1J5Q replacement), respectively. The refinement statistics are summarized in *SI Appendix*, Table S1.

Evaluation of Optimal Φ/Ψ Values for Free *N*-Glycan Glycosidic Linkages by Molecular Mechanics. MMs were performed using the MM3* and Amber FFs in the MacroModel program in the Schrödinger suite 2015. The Molecular Mechanic approach was used to evaluate the optimal dihedral angles of each glycosidic junction of the *N*-glycans, and the calculations were performed with a dielectric constant $\varepsilon = 80$, as an approximation for bulk water. Energy maps were calculated employing the Coordinate Scan utility

(modulated with the DEBG option 150, which imposes on the program to start each incremental minimization reading the initial input structure file). In more detail, both Φ and Ψ were varied incrementally, using a grid step of 18°; each (Φ, Ψ) point of the map was optimized using 10,000 P.R. (Polak-Ribiere) conjugate gradients. Φ is defined as O_n-C_1-O-C_n and Ψ as C_1-O-C_n-O_{n+1}. The plotting and the analysis of the relaxed maps were performed with the Plot Coordinate Scan facility built in the MacroModel package (*SI Appendix*, Fig. S6).

MD of the Free Oligosaccharides. On the basis of the values of the minimum energy obtained from the MM approach, the two glycans, glyco1 and glyco2, were built, minimized again with MM3 or Amber FF, approximating water solvent with a dielectric constant ε = 80. MD simulation (either with MM3 or Amber FF) was started for each optimized structure at 310 K for 20 ns, setting an equilibration time of 250 ps, using a dynamic time step of 2.0 fs together, and the SHAKE protocol to the HBs was applied and coordinates were saved regularly, leading to the collection of 10,000 frames. For each frame, interproton distances or Φ/Ψ values were extracted with the tools offered from the Maestro program in the Schrödinger suite. Φ/Ψ scattered maps (*SI Appendix*, Fig. S7), and calculation of distance-averaged values were made with the MS Excel program (*SI Appendix*, Table S4). For Amber simulation, the frequency count of Φ and Ψ was performed with the Origin program to appreciate which value had the highest probability. Some of these values (*SI Appendix*, Table S3) were used to build the complete *N*-glycan used in the MCMM searches.

NMR Spectra Acquisition and Evaluation of Experimental Distances. Determination of experimental distances was performed measuring the T-ROESY spectrum on the *N*-glycopeptide mixture isolated in our previous study (9). The T-ROESY experiment was carried out in a Bruker DRX-600 instrument equipped with a cryo-probe, the spectral width was set to 10 ppm, and the frequency carrier was placed at the residual peak of the solvent signal: 512 FIDs (free induction decay) of 2,048 complex data points and 32 scans per FID were acquired with 250 ms as mixing time. Data processing and T-ROESY cross peak integrations were performed with the standard Bruker Topspin 3 program, elaboration of cross peak volumes into distances was performed with the MS Excel program, and a distance to volume 10^{-6} dependency was used.

Construction of the Complete PBCV-1 Vp54 Protein and MCMM Search. The 5TIP structure was imported into MacroModel, and all molecules (water, Hg, *N*-glycans and B-D chains), except chain A were deleted. This polypeptide chain was used to build the fully glycosylated glycoprotein by connecting the oligosaccharides at the appropriate Asn positions (glyco1 at Asn-302, Asn-399, and Asn-406; glyco2 at Asn-280) and maintaining the orientation of each Glc-Asn connection as in 5TIP chain A (*SI Appendix*, Table S2). Glyco1 or glyco2 were elaborated using the dihedrals of the residues as in the X-ray structure (*SI Appendix*, Table S2), whereas those missing were set up by using the dihedral values extracted by frequency count from the MD simulation (*SI Appendix*, Table S3).

Exploring all the possible conformations was conducted by varying the Φ/Ψ values of the residues over all the possible values, including the Glc-Asn junction, except its amidic linkage, which was left free to minimize but not varied. In the MCMM set up, minimization (Amber, ε = 80) was allowed for the *N*-glycans (including Asn), two constraint shells were applied to the protein: atoms in the first shell (3 Å) were fixed (force constant 200.0 kJ/mol $Å^2$), atoms in the second shell (further 6 Å) were frozen, and the remaining atoms were ignored. The maximum number of steps was set to 15,000, steps per rotatable bond were 100, and the energy window for saving structures was 200 kJ/mol.

The MCMM output generated 11,416 different structures, ordered by increasing potential energy; the whole ensemble was scanned with the Superposition option of Maestro to find those conformers that better matched the monosaccharide residues of the crystallographic data.

Superposition procedure in MCMM was performed as follows: the protein backbone of the conformers was aligned to that of 5TIP chain A, and glycans residues matched "in place"; viz, without reorienting the conformers to find the best fit with the reference. For each monosaccharide, superimposition was performed matching the coordinates of the ring atoms: five carbons and one oxygen. This selection criterion expresses the similarity between two structures calculating the rmsd among the coordinates of two species, where a low rmsd value means a better overlay. Structures were ordered by increasing rmsd values and reported in a graph along with their energy (*SI Appendix*, Fig. S8).

MD of Vp54 Conformers A, C, and D. Each MD simulation focused on the glycan part of the glycoprotein by using the same constraints applied at the polypeptide backbone during the MCMM conformational search: the four oligosaccharides and the corresponding Asn were free to move, atoms within 3 Å were fixed with a force constant of 200.0 kJ/mol $Å^2$, atoms in the second layer (6 Å) were kept frozen, and those left were ignored. MD simulation was performed with Amber FF approximating the bulk with the dielectric constant ε = 80. The initial conformer was equilibrated for 250 ps and successively kept in a thermal bath at 310 K for 100 ns, using a dynamic time step of 2.0 fs and collecting 20,000 frames. Extraction of Φ/Ψ values was performed with Maestro, and graphs were generated with MS Excel or Origin.

Energy Minimization of Assembly-A, Assembly-C, and Assembly-D and Energy Calculation. Minimization focused on the glycan part of the assembly: the 20 oligosaccharides and the corresponding Asn were free to move, atoms within 3 Å were fixed with a force constant of 200.0 kJ/mol $Å^2$, atoms in the second layer (4 Å) were kept frozen, and those left were ignored. Minimization was performed with Amber FF approximating the bulk with the dielectric constant ε = 80. Energy calculation was performed with the Current Energy option of the Macromodel program, with Amber FF for the minimization procedure, but it was extended to all the atoms of the assembly.

Glycosylated Assembly and MCMM Calculation. A fully glycosylated assembly, joining crystallographic and modeling information, was prepared to modify assembly-A as follows: the glycan dihedral values were set to those available from the crystallographic structure (*SI Appendix*, Table S2), and the values missing were taken from the MD of the free glycans (*SI Appendix*, Table S3). This initial conformer presented few clashes among some of the residues absent in the X-ray, and thus, this initial structure was optimized by a minimization restricted to these units; namely, the two L-Rha units at Asn-399, all of the residues at Asn-406 except Glc and Fuc, and both Ara and diOMe-L-Rha at Asn-280. DiOMe-L-Rha was included in the minimization because it was not detected in three of the chains in 5TIP.

The MCMM conformational search was conducted over the five molecules of the assembly, and by varying the Φ/Ψ values of the monosaccharides not detected in the X-ray structure, minimization was limited to these residues and was performed with Amber FF setting ε = 80, and applying two constraint shells: atoms in the first shell (12 Å) from the residues of search were fixed (force constant 200.0 kJ/mol $Å^2$), atoms in the second shell (further 5 Å) were frozen, and the remaining atoms were ignored. The maximum number of steps was set to 10,000, steps per rotatable bond were 100, and the energy window for saving structures was 100 kJ/mol.

The number of steps was less than that used for MCMM of monomeric Vp54 (i.e., 15,000 steps); 10,000 steps were a reasonable compromise between the accuracy of information to retrieve and the computer time requested, a choice supported from the redundancy of the system that contains five copies of the same glycoprotein and also from the total number of dihedrals varied, which is comparable to that of the monomeric Vp54; viz, 110 (22 for each mate) versus 78, respectively. The MCMM output was ordered by increasing potential energy, and about 50% of the structures were within the energy window; analysis focused on the conformers within 40 kJ/mol (10 kcal/mol) from the global minimum to pinpoint the structural elements occurring in conformers with similar energy.

ACKNOWLEDGMENTS. This project was partially supported by National Science Foundation Grant EPS-1004049 (to J.L.V.E.), the National Institute of Health Grant P20-RR15635 from the Centers of Biomedical Research Excellence Program of the National Center for Research Resources (to J.L.V.E.) and the National Institute of Health Grant AI011219 (to M.G.R.).

1. Doms RW, Lamb RA, Rose JK, Helenius A (1993) Folding and assembly of viral membrane proteins. *Virology* 193:545–562.
2. Knipe DM (1996) Virus-host cell interactions. *Fields Virology*, ed Fields BN (Lippincott-Raven Publ., Philadelphia), 3rd Ed, pp 273–299.
3. Olofsson S, Hansen JES (1998) Host cell glycosylation of viral glycoproteins–A battlefield for host defence and viral resistance. *Scand J Infect Dis* 30:435–440.
4. Vigerust DJ, Shepherd VL (2007) Virus glycosylation: Role in virulence and immune interactions. *Trends Microbiol* 15:211–218.
5. Van Etten JL, Gurnon JR, Yanai-Balser GM, Dunigan DD, Graves MV (2010) Chlorella viruses encode most, if not all, of the machinery to glycosylate their glycoproteins independent of the endoplasmic reticulum and Golgi. *Biochim Biophys Acta* 1800:152–159.
6. Van Etten JL, Dunigan DD (2012) Chloroviruses: Not your everyday plant virus. *Trends Plant Sci* 17:1–8.
7. Graves MV, Bernadt CT, Cerny R, Van Etten JL (2001) Molecular and genetic evidence for a virus-encoded glycosyltransferase involved in protein glycosylation. *Virology* 285:332–345.

8. Wang I-N, et al. (1993) Evidence for virus-encoded glycosylation specificity. *Proc Natl Acad Sci USA* 90:3840–3844.
9. De Castro C, et al. (2013) Structure of N-linked oligosaccharides attached to chlorovirus PBCV-1 major capsid protein reveals unusual class of complex N-glycans. *Proc Natl Acad Sci USA* 110:13956–13960.
10. De Castro C, et al. (2016) N-linked glycans of chloroviruses sharing a core architecture without precedent. *Angew Chem Int Ed Engl* 55:654–658.
11. Van Etten JL, et al. (2017) Chloroviruses have a sweet tooth. *Viruses* 9:88–110.
12. Stanley P, Schechter H, Taniguchi N (2009) N-glycans. *Essentials of Glycobiology*, eds Varki A, et al. (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY), 2nd Ed, pp 101–114.
13. Nandhagopal N, et al. (2002) The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus. *Proc Natl Acad Sci USA* 99:14758–14763.
14. Agirre J, Davies GJ, Wilson KS, Cowtan KD (2017) Carbohydrate structure: The rocky road to automation. *Curr Opin Struct Biol* 44:39–47.
15. Zhang X, et al. (2011) Three-dimensional structure and function of the *Paramecium bursaria* chlorella virus capsid. *Proc Natl Acad Sci USA* 108:14837–14842.
16. Wormald MR, et al. (2002) Conformational studies of oligosaccharides and glycopeptides: Complementarity of NMR, X-ray crystallography, and molecular modelling. *Chem Rev* 102:371–386.
17. Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett* 314:141–151.
18. Yamaguchi T, et al. (2014) Exploration of conformational spaces of high-mannose-type oligosaccharides by an NMR-validated simulation. *Angew Chem Int Ed Engl* 53: 10941–10944.
19. Agirre J, et al. (2015) Privateer: Software for the conformational validation of carbohydrate structures. *Nat Struct Mol Biol* 22:833–834.
20. Agirre J, Davies G, Wilson K, Cowtan K (2015) Carbohydrate anomalies in the PDB. *Nat Chem Biol* 11:303.

Supporting Information for

Structure of the Chlorovirus PBCV-1 Major Capsid Glycoprotein Determined by Combining Crystallographic and Carbohydrate Molecular Modeling Approaches.

Running title:
Glycan structure of Vp54

Cristina De Castro[a,1], Thomas Klose[b], Immacolata Speciale[c], Rosa Lanzetta[c], Antonio Molinaro[c], James L. Van Etten[d,1], Michael G. Rossmann[b,1]

[a]Department of Agricultural Sciences, University of Napoli, Via Università 100, 80055 Portici (NA), Italy, [b]Department of Biological Sciences, Purdue University, 240 South Martin Jischke Drive, West Lafayette, IN 47907-2032, [c]Department of Chemical Sciences, University of Napoli, Via Cintia 4, 80126, Napoli, Italy, [d]Department of Plant Pathology and Nebraska Center for Virology, University of Nebraska, Lincoln, NE 68583-0900.

[1]To whom correspondence should be addressed. E-mail: decastro@unina.it**,** jvanetten1@unl.edu**,** mr@purdue.edu

This pdf file includes:

- Figures S1 to S16
- Tables S1 to S4

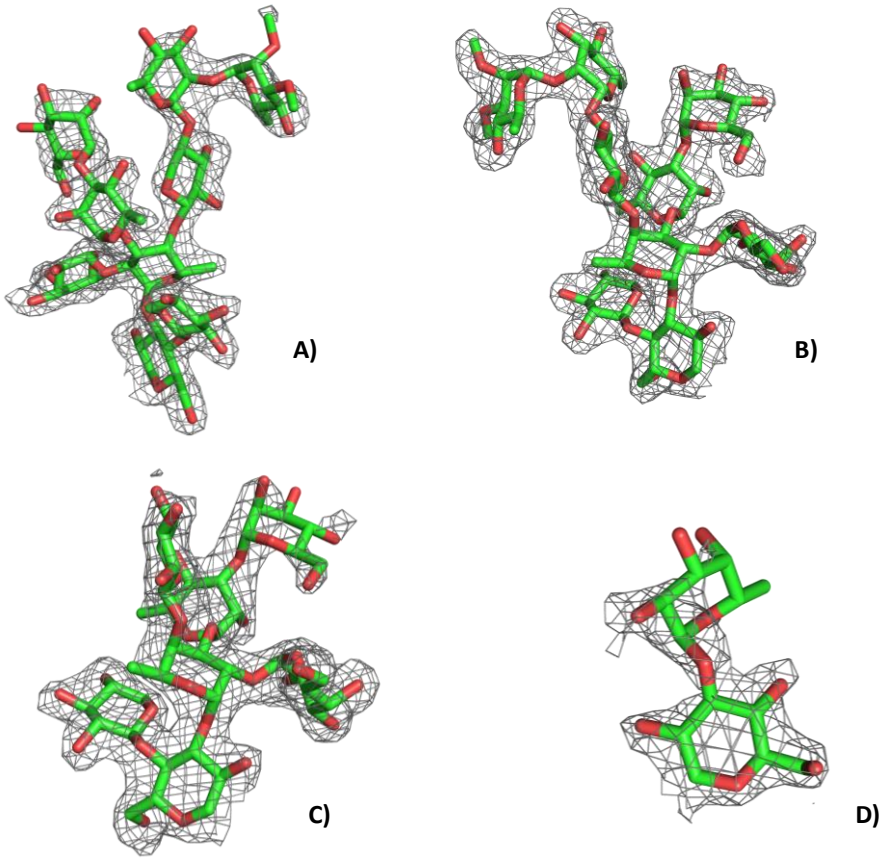

Fig. S1. The electron density surrounding each glycan is shown as a mesh based on a 2Fo-Fc map contoured at 1.2 sigma. A), B), C) and D) show enlarged versions for the glycans located at Asn280, Asn302, Asn399 and Asn406, respectively.

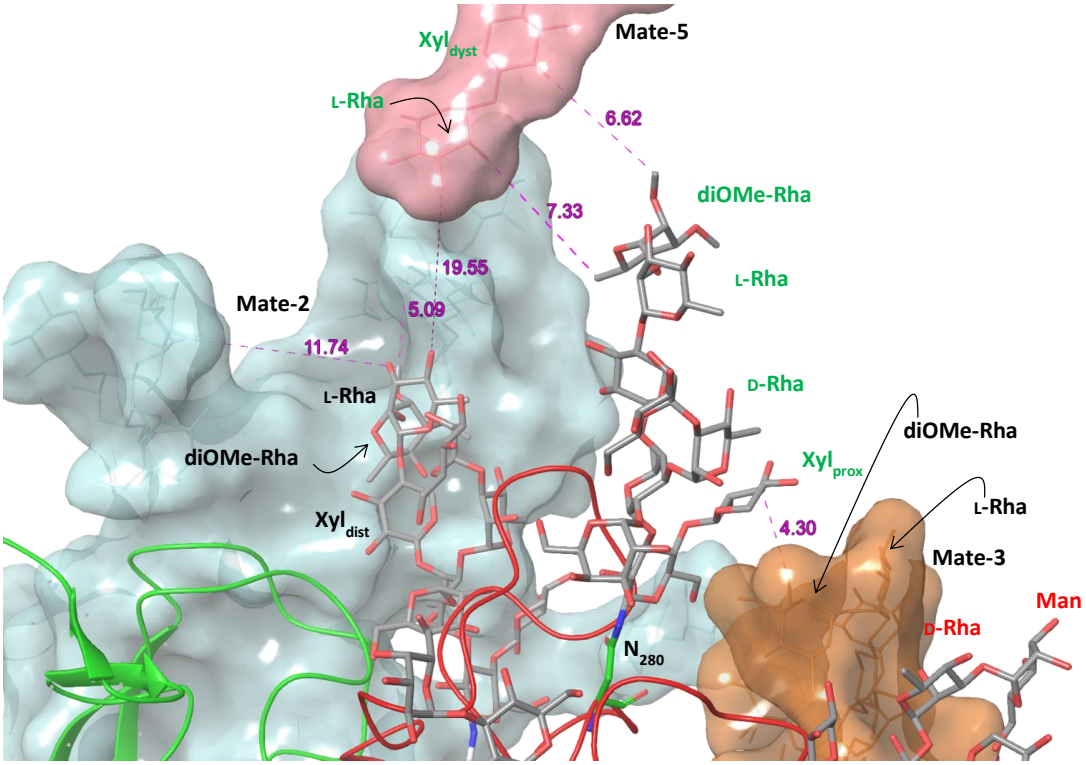

Fig. S2. Expansion of the region detailing the glycan at Asn-280 of Vp54 Chain A; surfaces enclose all the atoms of the mates at less than 10 Å from Chain A and the atoms displayed within are those of the monosaccharide residues. However, amino acids are not displayed unless they are important to the discussion. The color used to label the monosaccharides depends on which Asn is glycosylated: Asn-280: green, Asn-302: black, Asn-399: red, Asn-406: blue. Amino acids have a black label. The glycans of Chain A are represented in tube style. With regard to the crystal mates, only the atom within 10 Å from the glycans of Chain A are included to avoid crowding and the molecular surface of each mate is drawn with a different semitransparent color. The different mates are color-coded: mate-2, cyan; mate-3, orange; mate-4, yellow; mate-5, pink.

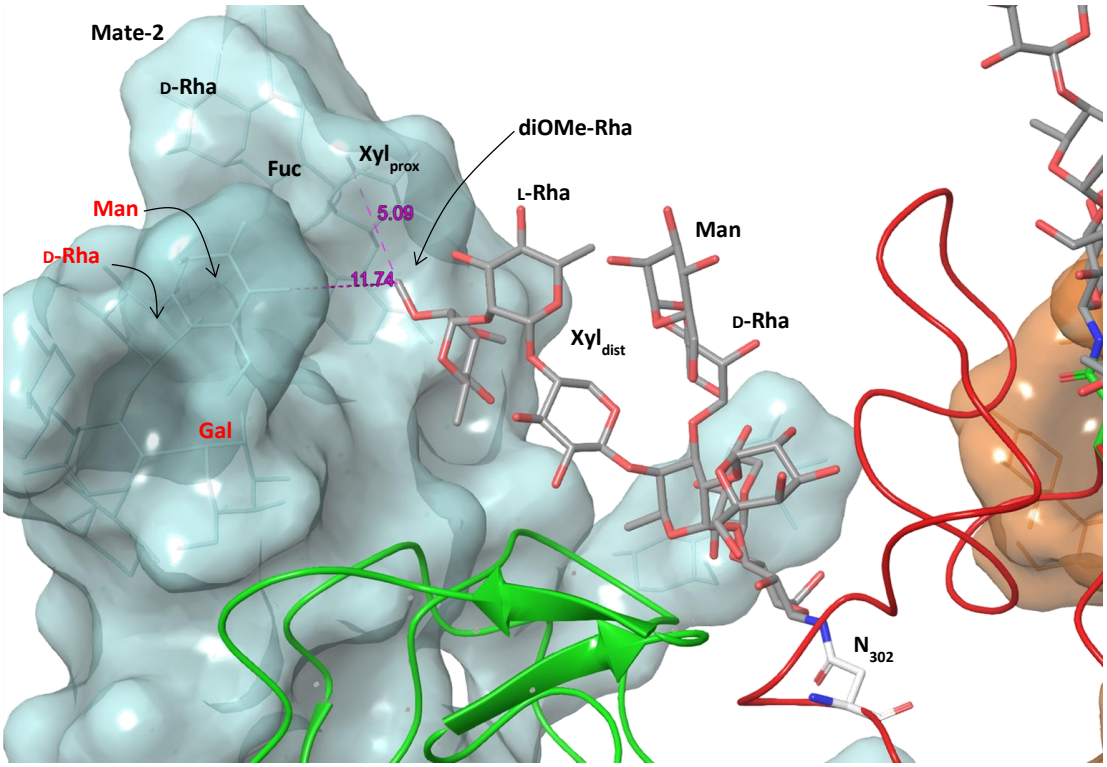

Fig. S3. Expansion of the region detailing the glycan at Asn-302 of Vp54 Chain A; surfaces enclose all the atoms of the mates at less than 10 Å from Chain A and the atoms displayed within are those of the monosaccharide residues. However, amino acids are not displayed unless they are important to the discussion. The colors used to label the monosaccharides depend on which Asn is glycosylated: Asn-280: green, Asn-302: black, Asn-399: red, Asn-406: blue. Amino acids have a black label. The glycans of Chain A are represented in tube style. With regard to the crystal mates, only the atom within 10 Å from the glycans of Chain A are included to avoid crowding and the molecular surface of each mate is drawn with a different semitransparent color. The different mates are color-coded: mate-2, cyan; mate-3, orange; mate-4, yellow; mate-5, pink.

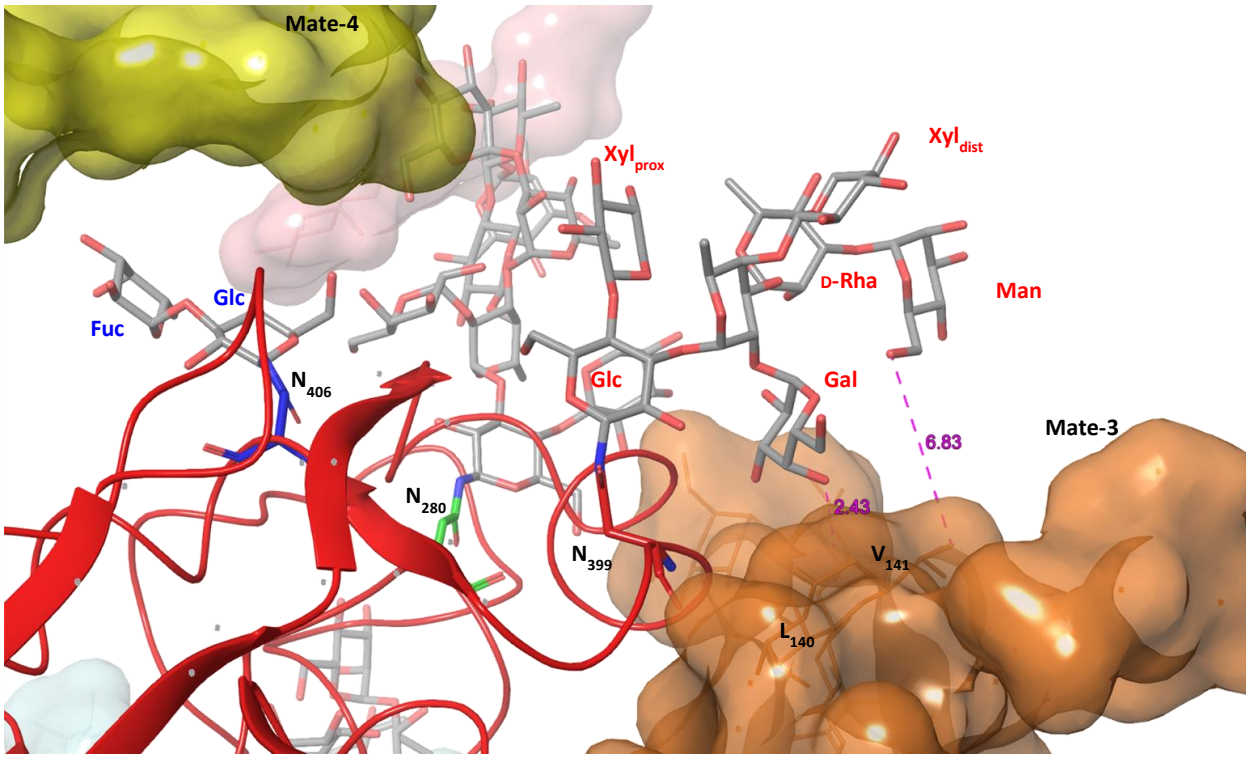

Fig. S4. Expansion of the region detailing the glycan at Asn-399 of Vp54 Chain A; surfaces enclose all the atoms of the mates at less than 10 Å from Chain A and the atoms displayed within are those of the monosaccharide residues. However, amino acids are not displayed unless they are important to the discussion. The color used to label the monosaccharides depends on which Asn is glycosylated: Asn-280: green, Asn-302: black, Asn-399: red, Asn-406: blue. Amino acids have a black label. The glycans of Chain A are represented in tube style. With regard to the crystal mates, only the atom within 10 Å from the glycans of Chain A are included to avoid crowding and the molecular surface of each mate is drawn with a different semitransparent color. The different mates are color-coded: mate-2, cyan; mate-3, orange; mate-4, yellow; mate-5, pink.

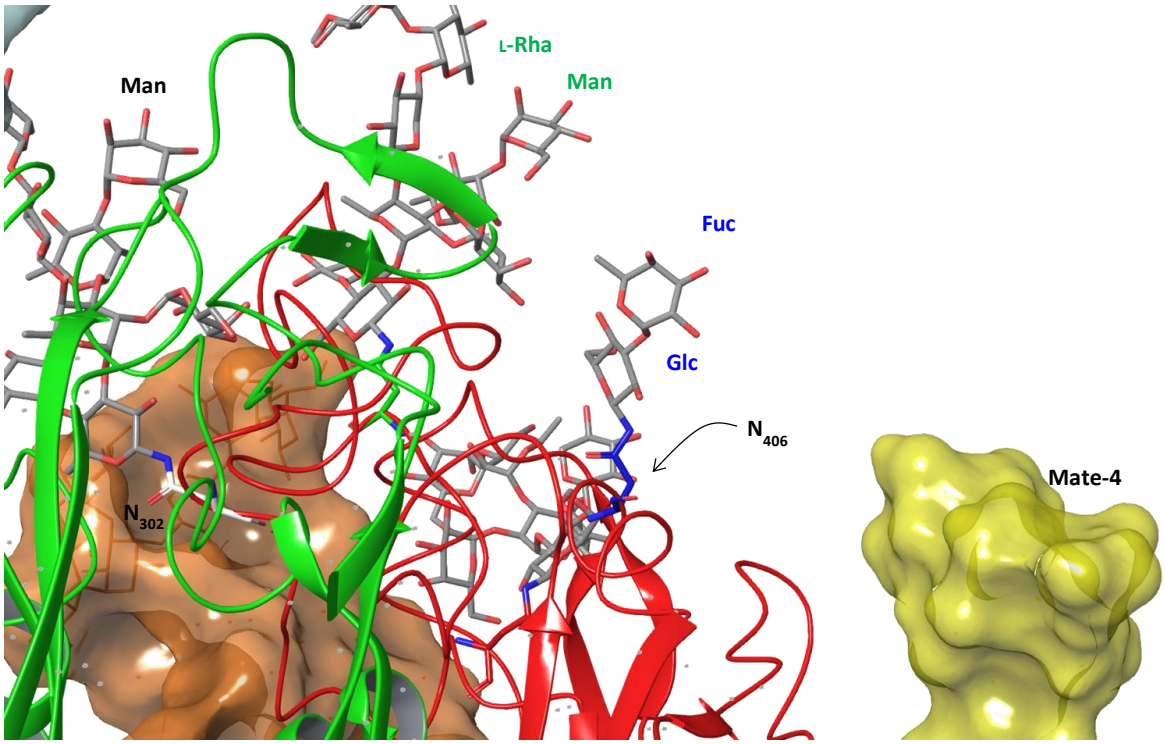

Fig. S5. Expansion of the region detailing the glycan at Asn-406 of Vp54 Chain A; surfaces enclose all the atoms of the mates at less than 10 Å from Chain A and the atoms displayed within are those of the monosaccharide residues. However, amino acids are not displayed unless they are important to the discussion. The color used to label the monosaccharides depends on which Asn is glycosylated: Asn-280: green, Asn-302: black, Asn-399: red, Asn-406: blue. Amino acids have a black label. The glycans of Chain A are represented in tube style. With regard to the crystal mates, only the atom within 10 Å from the glycans of Chain A are included to avoid crowding and the molecular surface of each mate is drawn with a different semitransparent color. The different mates are color-coded: mate-2, cyan; mate-3, orange; mate-4, yellow; mate-5, pink.

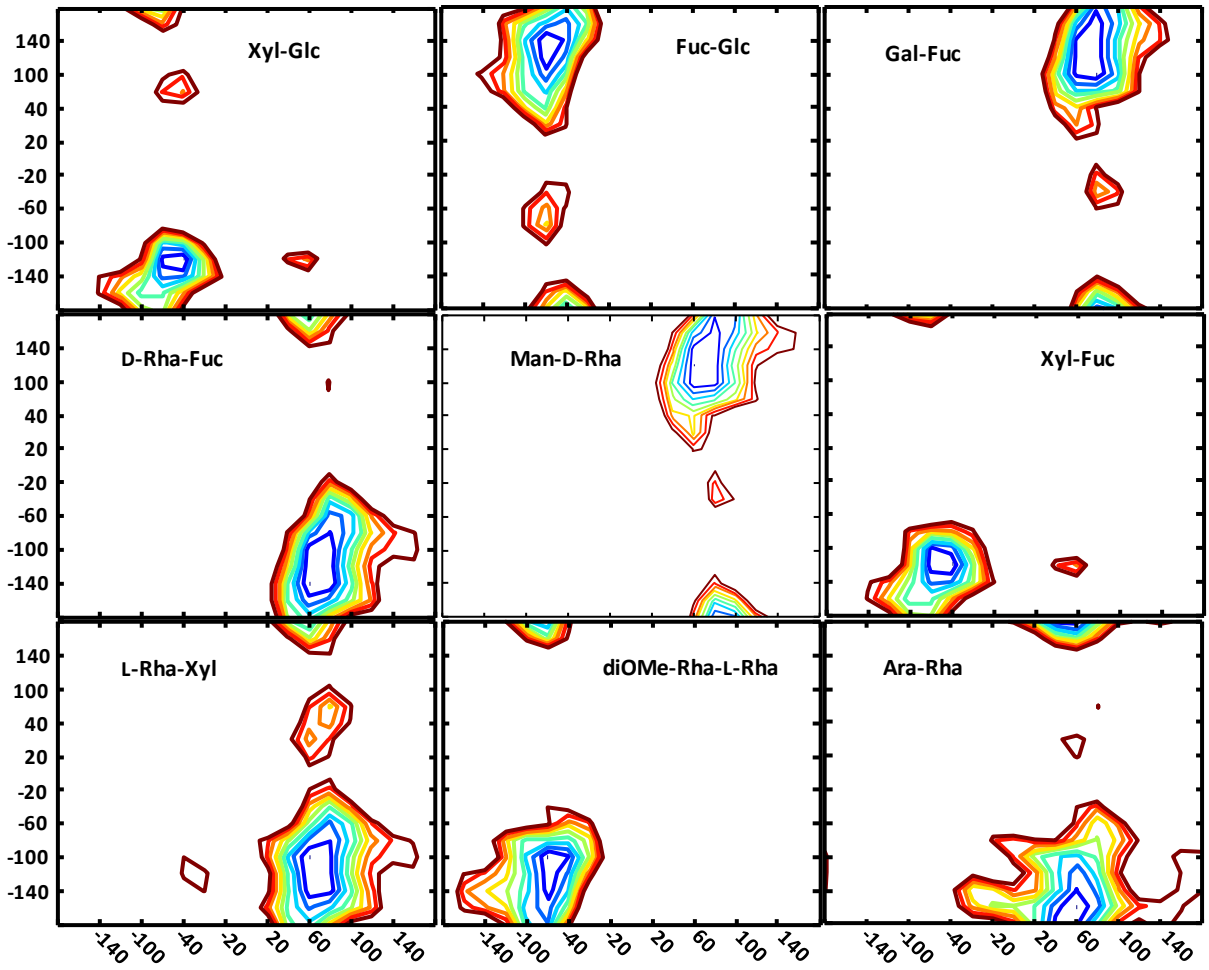

Fig. S6. Φ (abscissa) and Ψ (ordinate) flexible maps calculated for each type of glycosidic linkage present in the Vp54 glycans. Contours are drawn at increments of 0.2 kJ/mol above the absolute minimum found.

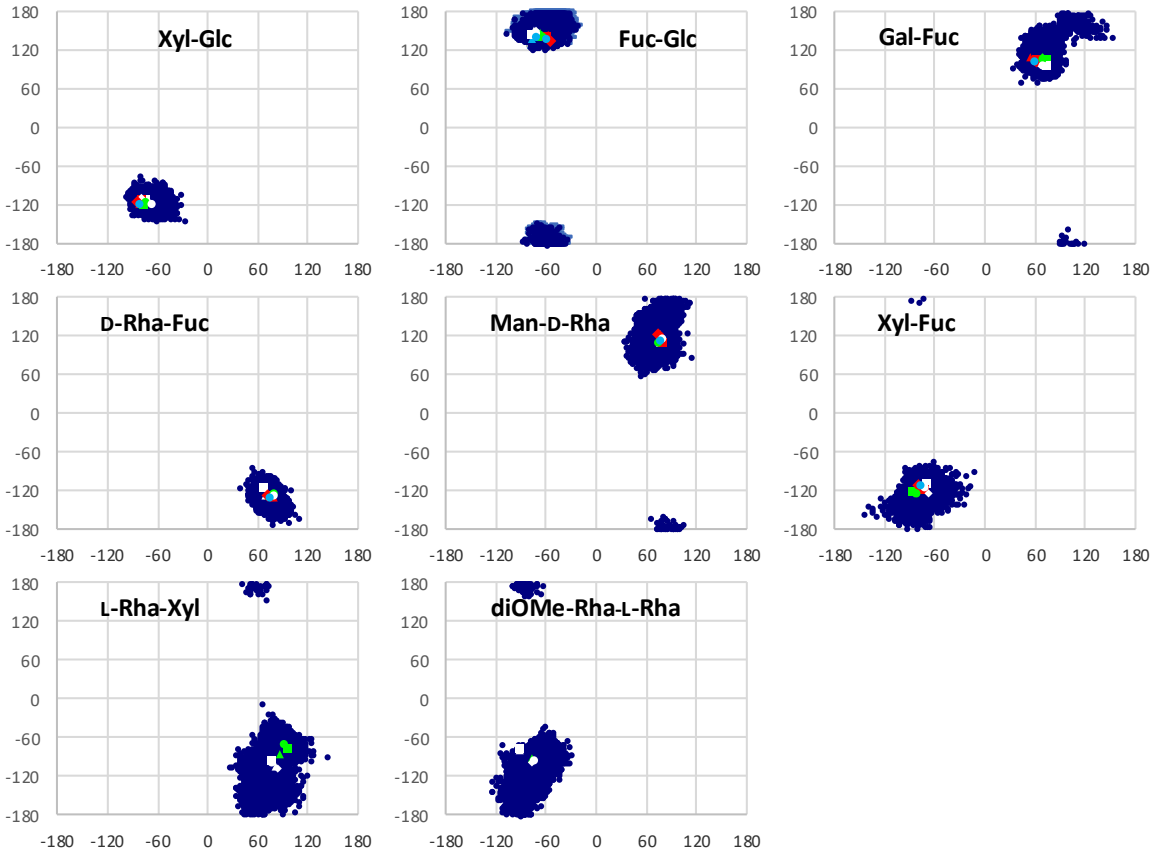

Fig. S7. Φ /Ψ (abscissa/ordinate) plots obtained for glyco2 for MD 20 ns simulation with AMBER FF at 310 K (glyco1 gave the same results). Values from X-ray data are represented with symbols: colors encode for oligosaccharides at different glycosylation sites: Asn-280: green, Asn-302: white, Asn-399: red, Asn-406: cyan while the four Vp54 chains of 5TIP structure are encoded with different symbols: chain A: triangle, chain B: square, chain C: diamond, chain D: circle.

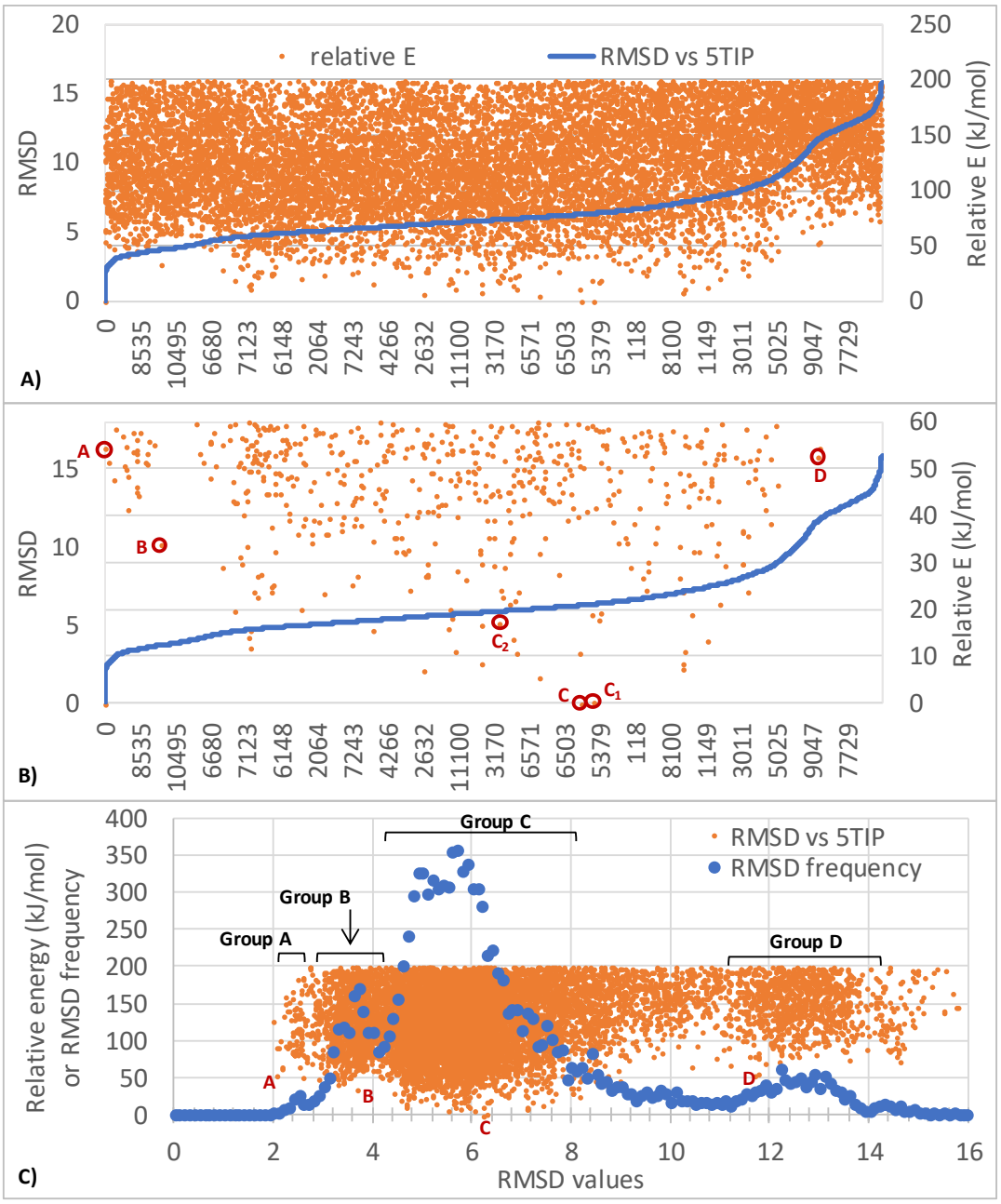

Fig. S8. Results from superposition between MCMM simulation of $Vp54_{hyb}$ and 5TIP structure. A) Profile of RMSD sorted in ascending order and energy values of the corresponding conformers (named with a number on the abscissa). B) As in A) but energy scale set at 60 kJ/mol. Dots enclosed in a circle represents the conformer selected for analysis and the name (as letter) associated with it. C) Frequency plot profile of the RMSD values, superimposed with the corresponding energy values. Conformers selected for further analysis are indicated.

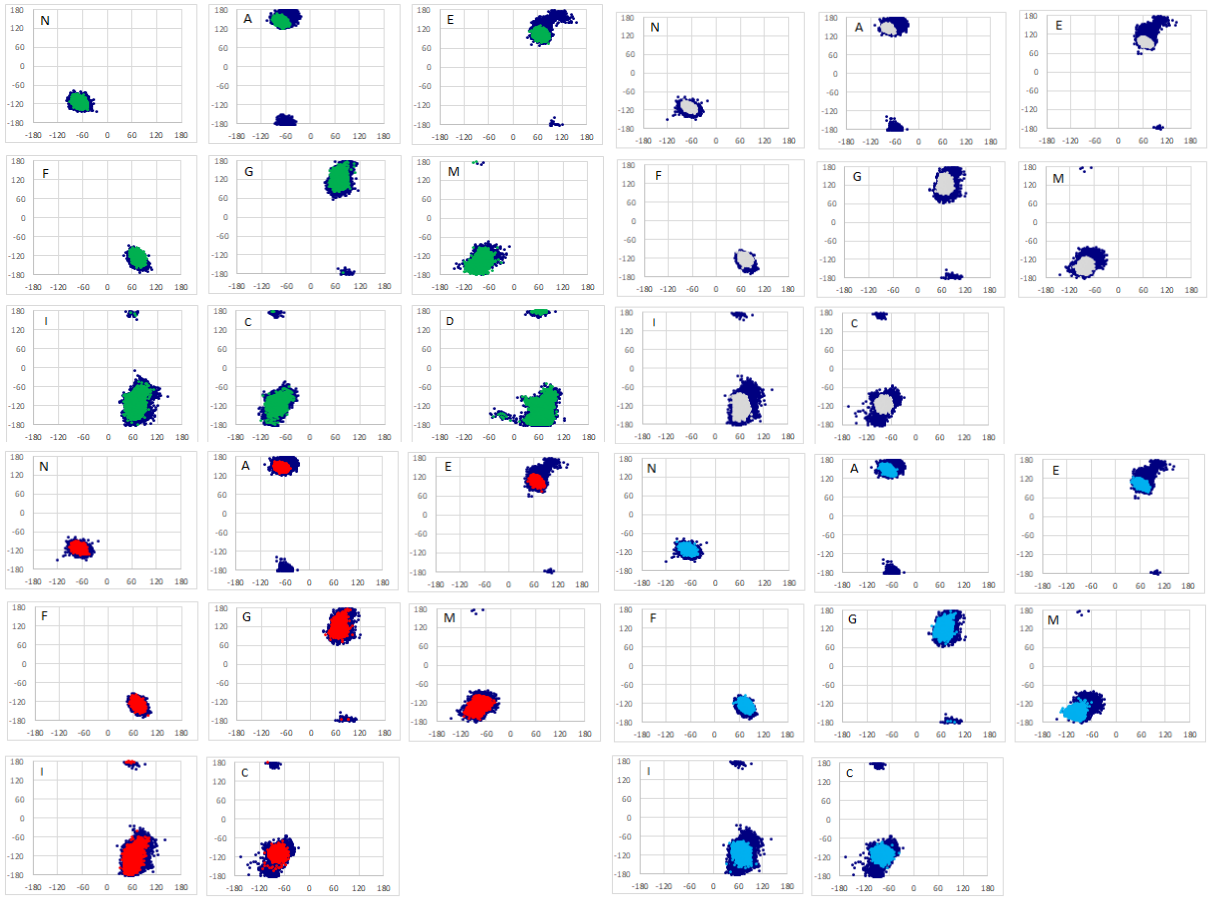

Fig. S9. Superimposition of Φ /Ψ (abscissa/ordinate) plots from two different MD simulations. Blue dots: free N-glycans. Colored dots: N-glycans linked at PBCV-1 Vp54 conformer A after MD simulation of 100 ns with AMBER FF at 310K. Colors code for oligosaccharides at different glycosylation sites: Asn-280: green, Asn-302: grey, Asn-399: red, Asn-406: cyan. Each Φ/Ψ couple is denoted with the letter originally used for the residue during NMR attribution (see Fig. 1A). Values found for the glycan at Asn-280 are superimposed to MD simulation of glyco2, while the others are superimposed to that of glyco1.

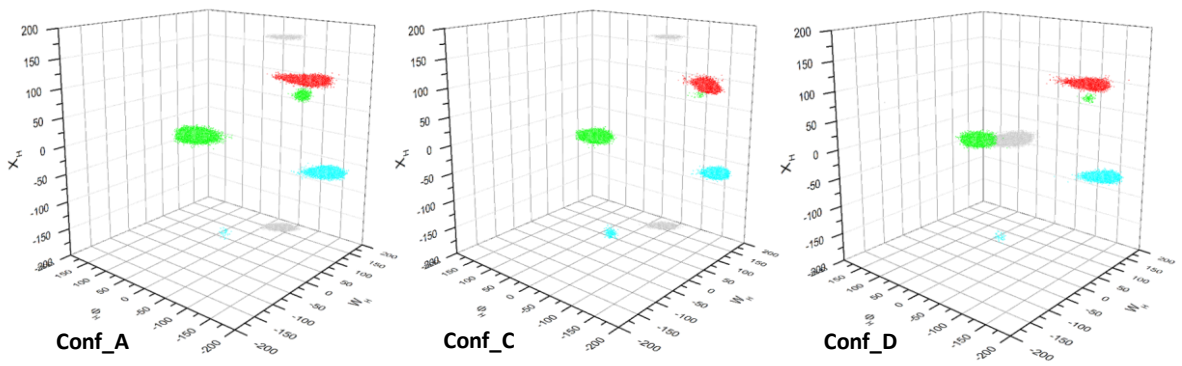

Fig. S10. 3D graph displaying Glc-Asn Φ, W and X dihedrals of conformers A, C and D from MD simulations of 100 ns with Amber FF at 310 K. Colors code for the different glycosylation sites: Asn-280: green, Asn-302: grey, Asn-399: red, Asn-406: cyan. The amide linkage of the Glc-Asn junction (Ψ) adopts the *trans* conformation during the whole simulation, and measures about 180° on average.

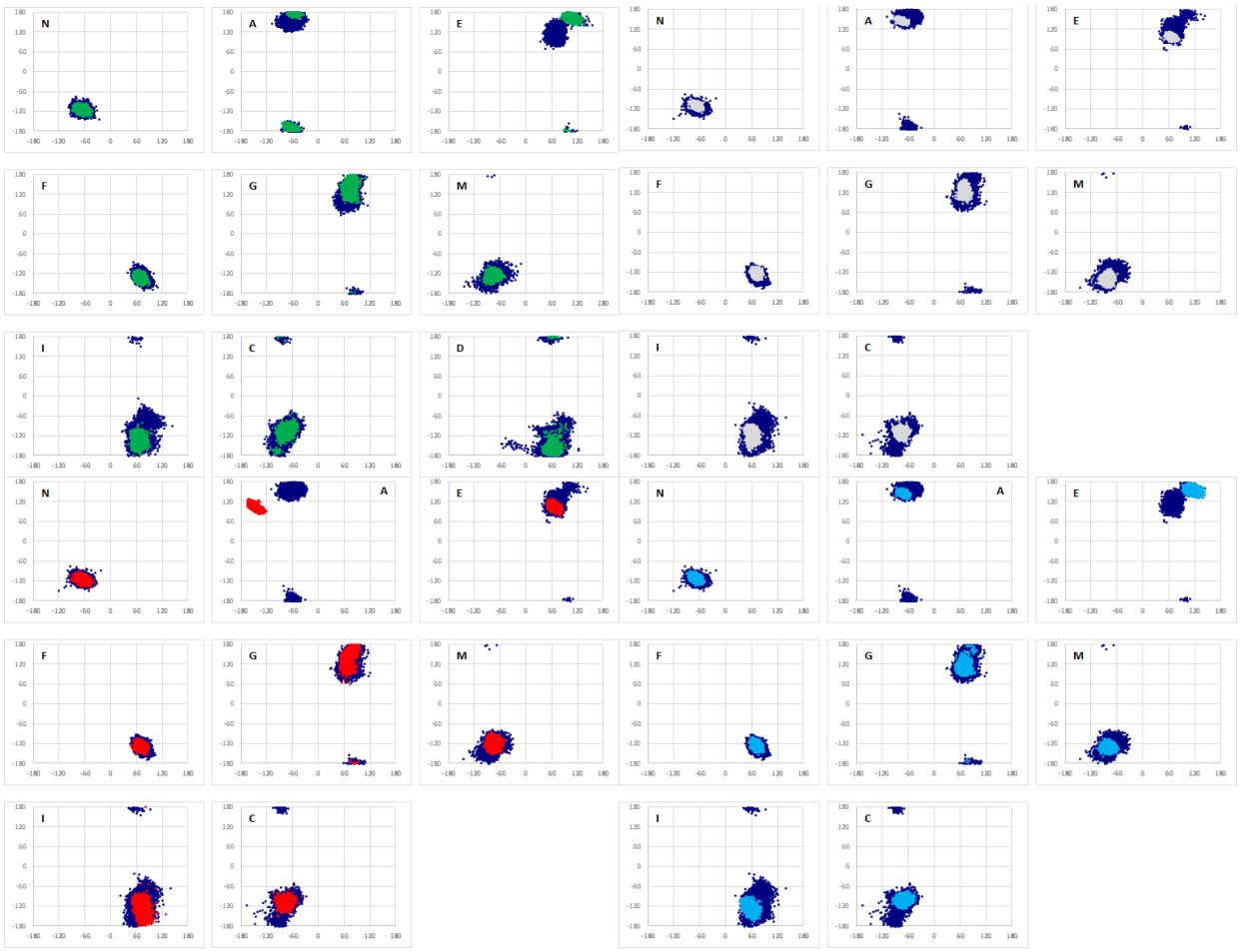

Fig. S11. Superimposition of Φ /Ψ (abscissa/ordinate) plots from two different MD simulations. Blue dots: free N-glycans (as in Fig. S7). Colored dots: N-glycans linked at PBCV-1 Vp54 conformer C after MD simulation of 100 ns with AMBER FF at 310K. Colors code for oligosaccharides at different glycosylation sites: Asn-280: green, Asn-302: grey, Asn-399: red, Asn-406: cyan. Each Φ/Ψ couple is denoted with the letter originally used for the residue during NMR attribution (see Fig. 1A). Values found for glycan at Asn-280 are superimposed to MD simulation of glyco2, while the others to that of glyco1.

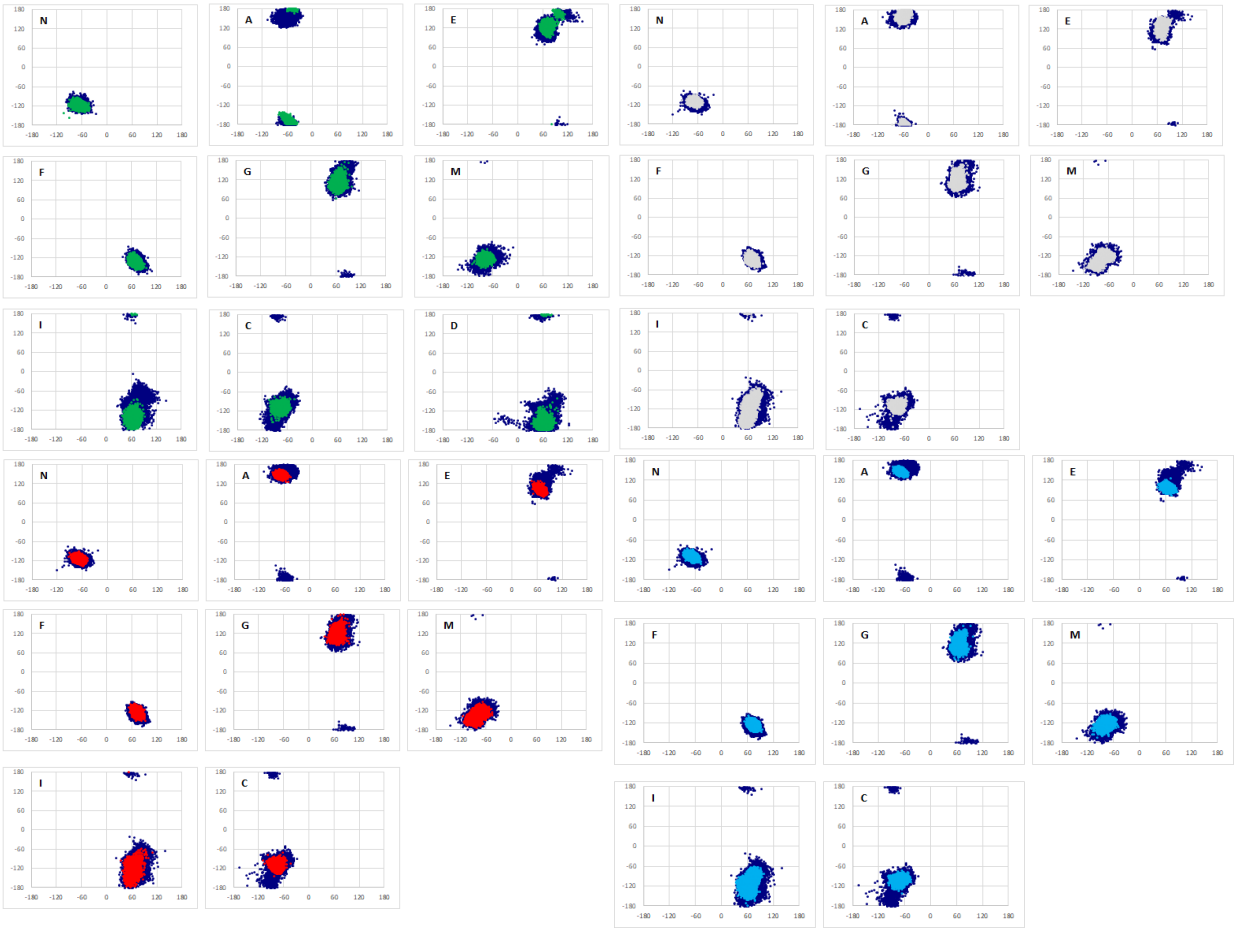

Fig. S12. Superimposition of Φ /Ψ (abscissa/ordinate) plots from two different MD simulations. Blue dots: free N-glycans (as in Fig. S7). Colored dots: N-glycans linked at PBCV-1 Vp54 conformer D after MD simulation of 100 ns with AMBER FF at 310K. Colors code for oligosaccharides at different glycosylation sites: Asn-280: green, Asn-302: grey, Asn-399: red, Asn-406: cyan. Each Φ/Ψ couple is denoted with the letter originally used for the residue during NMR attribution (see Fig. 1A). Values found for the glycan at Asn-280 are superimposed to MD simulation of glyco2, while the others to that of glyco1.

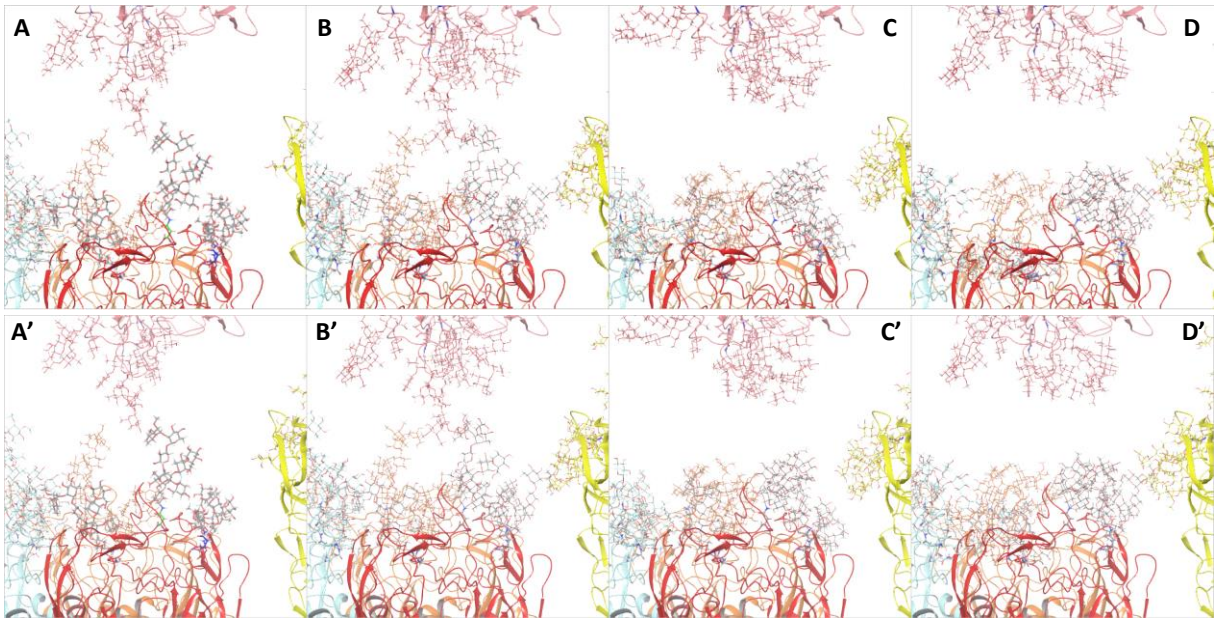

Fig. S13. A and A') Details of the glycan area of the mates of Chain A as in Fig. 3. B) Assembly-A, namely the assembly obtained replacing each Vp54 molecule of the crystallographic structure with the Vp54_A conformer from MCMM conformational search. B') Same as B) but after glycan minimization. C and C') Details of assembly-C structure before and after glycans minimization. D and D') Details of assembly-D structure before and after glycans minimization. The different mates are color coded: Chain A (mate-1), red; mate-2, cyan; mate-3, orange; mate-4, yellow; mate-5, pink.

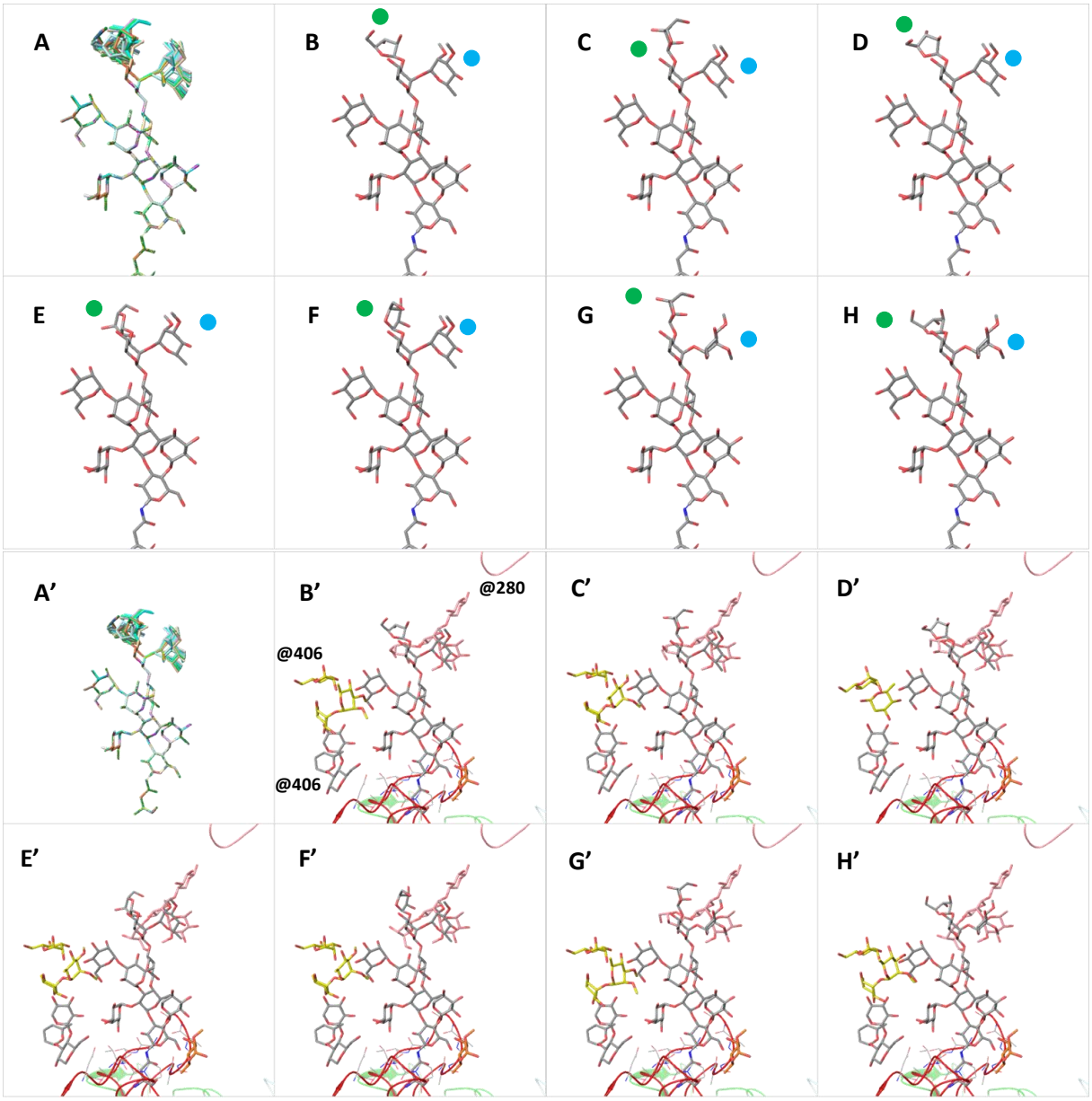

Fig. S14. Analysis of the MCMM search performed on the modeled assembly by varying the dihedrals of the glycan residues not visible in the X-ray structure. Possible conformations of the diOMe-L-Rha and Ara units of the glycan of Asn-280 of Chain A, at less than 10 kcal/mol from the global minimum. A) Superimposition of all the conformers of the Asn-280 glycan found in the energy range inspected; to avoid crowding, the other parts of the assembly are omitted. B-H) View of the most representative conformers, diOMe-Rha and Ara have a blue or a green circle next to facilitate their identification. A') Same as A. B'-H') Views of the same conformers as in B-H, but in the assembly context. To avoid crowding and to simplify the vision, the ribbons of the protein backbone of mates 2-5 are drawn with different colors and only atoms within 5 Å from the glycans are included. The different mates are color-coded: mate-2, cyan; mate-3, orange; mate-4, yellow; mate-5, pink. The diction @xxx indicates the position linked to the glycan (or fragment) appearing in panel B'. Hydrogen atoms are omitted for clarity.

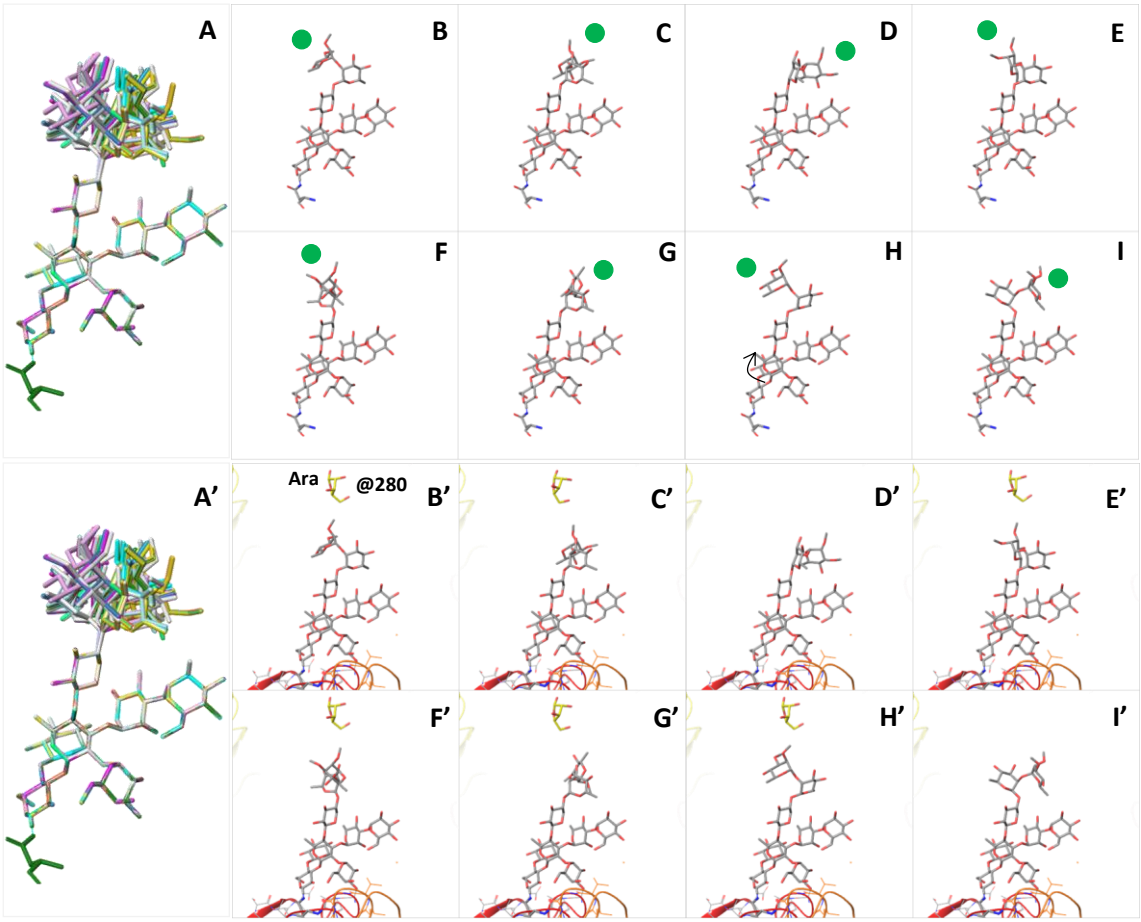

Fig. S15. Analysis of the MCMM search performed on the modeled assembly by varying the dihedrals of the glycan residues not visible in the X-ray structure. Possible conformations of the diOMe-Rha-(1→2)-L-Rha disaccharide of the glycan of Asn-399 of Chain A, at less than 10 kcal/mol from the global minimum. The diOMe-Rha has a green circle next to facilitate its identification. A, A') Superimposition of all conformers of Asn-399 glycan found in the energy range inspected, to avoid crowding, the other parts of the assembly are omitted; B-I) View of the most representative conformers. B'-I') Views of the same conformers as in B-I, but in the assembly context. To avoid crowding and simplify the vision, the ribbons of the protein backbone of mates 2-5 are drawn with different colors and only atoms within 5 Å from the glycans are included. The different mates are color coded: chain A, red; chain A, red; mate-2, cyan; mate-3, orange; mate-4, yellow; mate-5, pink. The diction @xxx indicates the position linked to the glycan (or fragment) appearing in panel B'. Hydrogen atoms are omitted for clarity.

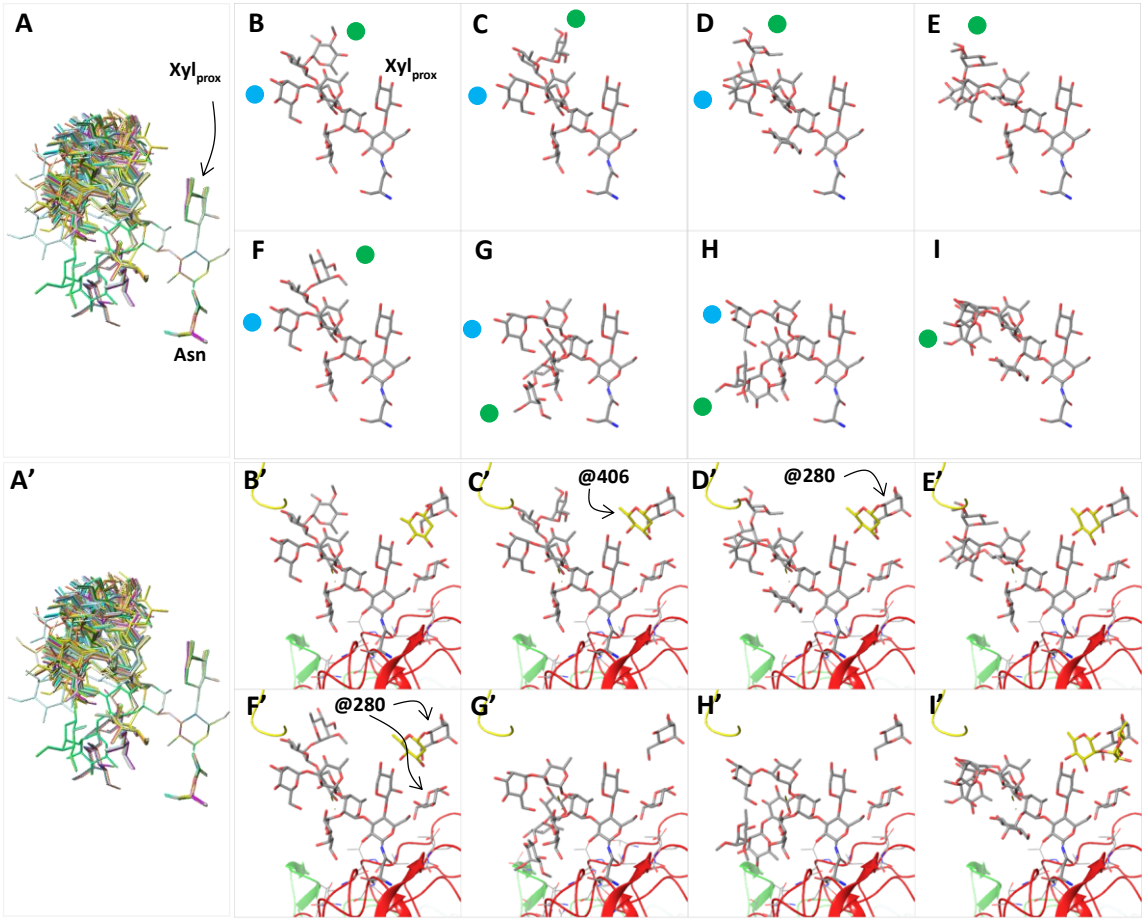

Fig. S16. Analysis of the MCMM search performed on the modeled assembly by varying the dihedrals of the glycan residues not visible in the X-ray structure. Possible conformations of the glycan of Asn-406 of Chain A, at less than 10 kcal/mol from the global minimum. A, A') Superimposition of all conformers of Asn-406 glycan found in the energy range inspected, to avoid crowding, the other parts of the assembly are omitted. B-I) View of the most representative conformers, diOMe-Rha and Man have a green and blue circle, respectively, next to them to facilitate their identification. B'-I') Views of the same conformers as in B-I, but in the assembly context. To avoid crowding and simplify the vision, the ribbons of the protein backbone of mates 2-5 are drawn with different colors and only atoms within 4 Å from the glycans are included. The different mates are color coded: chain A, red; mate-2, cyan; mate-3, orange; mate-4, yellow; mate-5, pink. The diction @xxx indicates the Asn linked to the glycan (or fragment) appearing in panels C', D' and F'. Hydrogen atoms are omitted for clarity.

Table S1. Crystallographic statistics for 5TIP and 5TIQ structures obtained by refining the original structures deposited with the codes 1M3Y and 1J5Q (in brackets), respectively.

	PBCV-1 Vp54	
Data		
PDB ID	5TIP (1M3Y)	5TIQ (1J5Q)
Space group	$P2_13$	$P4_132$
Resolution	84.43-2.00	84.42-2.54
Refinement		
No. of reflections	135050	317487
Rwork/Rfree	0.18/0.21	0.18/0.22
Average B-factor	35.9	36.78
R.m.s deviations		
Bond length (Å)	0.003	0.002
Bond angles (deg.)	0.5	0.5
Ramachandran plot (%)		
Favored	97.2	96.3
Allowed	2.9	3.7
Outliers	0	0

Table S2: Dihedral values of the N-glycans measured for the four Vp54 chains of 5TIP, and for the two chains of 5TIQ. Barred cells mean that no residue was visible. Dihedrals are defined as follows: Glc-Asn junction: $\Phi = O_5C_1NC_\gamma$, $\Psi = C_1NC_\gamma C_\beta$, $W = NC_\gamma C_\beta C_\alpha$, $X = C_\gamma C_\beta C_\alpha C_{C=O}$, for the other glycosidic bonds: $\Phi = O_5C_1OC_n$, $\Psi = C_1OC_nC_{n+1}$. Monosaccharides are labelled with the letter originally used during NMR attribution, see Figure 1.

		5TIP																5TIQ							
		Chain A				Chain B				Chain C				Chain D				Chain A				Chain B			
		280	302	399	406	280	302	399	406	280	302	399	406	280	302	399	406	280	302	399	406	280	302	399	406
H	Φ	-70	-81	-123	-97	-77	-84	-109	-98	-74	-87	-118	-110	-74	-85	-119	-108	-94	-78	-134	-87	-99	-75	-131	-87
	Ψ	168	-173	-162	-173	170	-171	-170	180	168	-174	-162	174	168	-172	-160	-173	177	-173	-157	156	-179	-174	-155	156
	W	177	128	103	142	-179	128	89	150	178	126	96	148	178	127	99	150	-168	124	111	136	-164	122	108	137
	X	70	-177	114	-52	70	-177	115	-48	68	-178	119	-54	70	-177	113	-54	65	-176	112	-57	64	-176	114	-58
N	Φ	-79	-70	-82		-78	-74	-80		-76	-79	-85		-74	-67	-82		-88	-70	-84		-88	-76	-90	
	Ψ	-113	-115	-118		-117	-111	-112		-117	-110	-115		-115	-117	-119		-104	-116	-105		-110	-112	-105	
A	Φ	-68	-74	-57	-76	-68	-76	-59	-77	-62	-72	-55		-67	-72	-59	-72	-56	-74	-49	-76	-56	-75	-45	-81
	Ψ	148	151	145	137	142	145	-142	135	140	152	134		143	146	137	142	138	151	131	130	133	156	127	134
E	Φ	67	65	56		75	75	62		70	65	63		70	70	61		71	65	66		62	65	57	
	Ψ	107	102	109		106	96	105		108	99	103		106	96	103		98	102	84		102	102	96	
F	Φ	74	69	78		75	67	77		77	73	72		78	78	75		72	69	65		68	69	67	
	Ψ	-128	-128	-132		-127	-115	-128		-129	-127	-128		-126	-128	-132		-128	-128	-116		-124	-130	-120	
G	Φ	79	76	79			76	80		78	79	75		75	80	76		79	76	78		76	74	74	
	Ψ	116	121	120			109	110		112	115	121		110	115	114		116	121	120		118	119	125	
M	Φ	-78	-72	-76		-86	-69	-71		-84	-66	-79		-81	-72	-77		-74	-72	-72		-78	-70	-68	
	Ψ	-116	-113	-119		-121	-110	-117		-124	-125	-111		-124	-114	-113		-126	-113	-126		-118	-113	-126	
I	Φ	86	82			96	77				84			91	76			94	82			96	85	103	
	Ψ	-88	-106			-76	-95				-108			-71	-98			-89	-106			-92	-104	-114	
C	Φ	-78	-91				-92				-78				-74			-65	-90			-72	-92	-62	
	Ψ	-94	-81				-78				-99				-97			-101	-81			-93	-87	-101	

Table S3. Values found for the glycosidic linkages of glyco2 by using a molecular modeling approach (Molecular Mechanic and Molecular Dynamic approach). Values found for glyco1 are very similar and omitted. Angles are defined as follows: $\Phi = O_5\text{-}C_1\text{-}O\text{-}C_n$, $\Psi = C_1\text{-}O\text{-}C_n\text{-}C_{n+1}$, $W = N\text{-}C_\gamma\text{-}C_\beta\text{-}C_\alpha$, $X = C_\gamma\text{-}C_\beta\text{-}C_\alpha\text{-}C_{C=O}$. σ = standard deviation. Monosaccharides are labelled with the letter originally used during NMR attribution, as reported in Figure 1A.

		M. Mechanic *Free N-glycan*		**MD (Amber)** *Free N-Glycan*
		AMBER	**MM3**	**(σ)**
H	Φ			
β-Glc	Ψ			
	W			
	X			
N	Φ	-70	-63	-68 (8.7)
β-Xyl_{prox}	Ψ	-116	-119	-112 (8.5)
A	Φ	-76	-63	-60 (10.2)
α-Fuc	Ψ	136	157	158 (12.9)
E	Φ	70	72	70 (10.0)
α-Gal	Ψ	106	156	118 (15.4)
F	Φ	68	58	70 (7.7)
α-D-Rha	Ψ	-134	-155	-132 (10.2)
G	Φ	69	64	70 (9.8)
α-Man	Ψ	126	104	112 (20.1)
M	Φ	-69	-59	-73 (12.2)
β-Xyl_{dist}	Ψ	-116	-115	-118 (14.0)
I	Φ	68	-56	68 (13.0)
β-L-Rha	Ψ	-130	--93	-118 (27.8)
C	Φ	-75	-64	-76 (11.9)
α-L-Me_2Rha	Ψ	-104	-85	-99 (19.6)
D	Φ	69	69	66 (20.4)
β-Ara*f*	Ψ	-172	-154	-161 (22.1)

Table S4. Interproton distances (Å) calculated experimentally from T-ROESY spectrum, recorded at 310 K on a 600 MHz spectrometer, or simulated for glyco1 and glyco2 oligosaccharides, for 20 ns and two different force fields. During NMR attribution, in glyco2 the β-L-Rha and diOMe-α-L-Rha residues were annotated with **B** and **L**, respectively, while in glyco1, **C** and **I** were used. These monosaccharides occupy the same position in the oligosaccharide architecture and differ for the presence of β-Ara*f* that is linked at O3 of **I** in glyco2 and absent in glyco1. Simulated distances in bold have a closer agreement with the experimental values, i.e. fall within 10% of the distance measured.

Proton pair	Experimental Dist.	Glyco1		Glyco2	
	(T-ROESY)	Amber	MM3	Amber	MM3
A_1E_5	2,53	**2,71**	2,18	**2,57**	**2,46**
A_1H_3	2,65	**2,80**	3,21	**2,88**	3,06
B_1L_2	2,23	/	/	2,48	2,81
B_1B_{2OMe}	2,84	/	**/**	3,43	3,15
C_1I_2	2,12	2,70	2,88	/	/
C_1C_{2OMe}	2,51	**2,36**	3,43	/	/
D_1L_2	2,40	/	/	**2,56**	**2,43**
I_1M_4	2,32	**2,38**	**2,38**	/	/
L_1M_4	2,32	/	/	**2,45**	2,58
A_5N_2	2,62	**2,68**	3,65	2,94	3,04
M_1A_4	2,13	**2,32**	2,85	2,44	2,59
$N_1H_{6'}$	2,65	**2,42**	3,19	3,46	3,79
E_5H_2	2,83	**3,01**	3,24	**3,14**	3,27
C_2C_{2OMe}	2,83	**2,83**	**2,79**	/	/
B_2B_{2OMe}	2,83	/	/	**2,88**	**2,87**
H_6N_1	2,96	2,52	2,34	**2,82**	2,54
H_4N_1*	2,55	**2,32**	**2,54**	**2,47**	**2,56**
H_5N_1*	2,91	3,97	3,97	4,02	3,87
G_5F_2	3,55	3,14	2,22	2,57	2,47
$C_{3OMe}C_2$	3,24	2,52	3,80	/	/
$B_{3Ome}B_2$	3,24	/	/	**3,38**	**3,28**
# similar values		**10**	**3**	**9**	**5**

* partially overlapped NOEs