2013

Structure of N-linked oligosaccharides attached to chlorovirus PBCV-1 major capsid protein reveals unusual class of complex N-glycans

Cristina De Castro
University of Napoli Federico II

Antonio Molinaro
University of Napoli Federico II

Francesco Piacente
University of Genova

James R. Gurnon
University of Nebraska-Lincoln

Luisa Sturiale
Consiglio Nazionale delle Ricerche–Istituto di Chimica e Tecnologia dei Polimeri

See next page for additional authors

Structure of N-linked oligosaccharides attached to chlorovirus PBCV-1 major capsid protein reveals unusual class of complex N-glycans

Cristina De Castro[a], Antonio Molinaro[a], Francesco Piacente[b], James R. Gurnon[c], Luisa Sturiale[d], Angelo Palmigiano[d], Rosa Lanzetta[a], Michelangelo Parrilli[a], Domenico Garozzo[d], Michela G. Tonetti[b], and James L. Van Etten[c,1]

[a]Department of Chemical Sciences, University of Napoli Federico II, 80126 Naples, Italy; [b]Department of Experimental Medicine and Center of Excellence for Biomedical Research, University of Genova, 16132 Genoa, Italy; [c]Department of Plant Pathology and Nebraska Center for Virology, University of Nebraska, Lincoln, NE 68583; and [d]Consiglio Nazionale delle Ricerche–Istituto di Chimica e Tecnologia dei Polimeri, Sezione di Catania, 95126 Catania, Italy

Contributed by James L. Van Etten, July 11, 2013 (sent for review May 15, 2013)

The major capsid protein Vp54 from the prototype chlorovirus *Paramecium bursaria* chlorella virus 1 (PBCV-1) contains four Asn-linked glycans. The structure of the four N-linked oligosaccharides and the type of substitution at each glycosylation site was determined by chemical, spectroscopic, and spectrometric analyses. Vp54 glycosylation is unusual in many ways, including: (*i*) unlike most viruses, PBCV-1 encodes most, if not all, of the machinery to glycosylate its major capsid protein; (*ii*) the glycans are attached to the protein by a β-glucose linkage; (*iii*) the Asn-linked glycans are not located in a typical N-X-(T/S) consensus site; and (*iv*) the process probably occurs in the cytoplasm. The four glycoforms share a common core structure, and the differences are related to the nonstoichiometric presence of two monosaccharides. The most abundant glycoform consists of nine neutral monosaccharide residues, organized in a highly branched fashion. Among the most distinctive features of the glycoforms are (*i*) a dimethylated rhamnose as the capping residue of the main chain, (*ii*) a hyperbranched fucose unit, and (*iii*) two rhamnose residues with opposite absolute configurations. These glycoforms differ from what has been reported so far in the three domains of life. Considering that chloroviruses and other members of the family *Phycodnaviridae* may have a long evolutionary history, we suggest that the chlorovirus glycosylation pathway is ancient, possibly existing before the development of the endoplasmic reticulum and Golgi pathway, and involves still unexplored mechanisms.

virus-encoded glycosylation | glucose/asparagine linkage | cytoplasmic glycosylation | glycopeptide | glycobiology

Many viruses, such as rhabdoviruses, herpesviruses, poxviruses, and paramyxoviruses, have structural proteins that are glycosylated. All viruses studied to date use host-encoded glycosyltransferases and glycosidases located in the endoplasmic reticulum (ER) and Golgi apparatus to add and remove sugar residues from their glycoproteins cotranslationally or shortly after translation of the protein. The virus glycoproteins are subsequently transported to host membranes, and progeny viruses then bud through these virus-specific target membranes, which is often the final step in the assembly of infectious virions (1–3). Thus, virus glycoproteins are host-specific because they are glycosylated by the same mechanism as host glycoproteins. Consequently, the only way to alter the glycan structure of virus glycoproteins is to grow the virus in a different host or have a mutation in the protein that alters its glycosylation site.

In contrast, viruses in the genus *Chlorovirus* (family Phycodnaviridae) differ from this paradigm in that they encode most, if not all, of the machinery to glycosylate their major capsid protein (4). Furthermore, the process probably occurs in the cytoplasm of the infected cell. Chloroviruses are large (190 nm in diameter) icosahedral, plaque-forming, dsDNA-containing viruses with an internal lipid membrane; they have genomes of 290 to 370 kb that contain ∼400 protein-encoding genes (5–7). The prototype chlorovirus, *Paramecium bursaria* chlorella virus 1 (PBCV-1), infects *Chlorella variabilis* (formerly known as *Chlorella* NC64A) that is normally a symbiont in the protozoan *P. bursaria*. The 330 kb PBCV-1 genome encodes 416 protein-encoding genes of 40 codons or larger (8). The PBCV-1 major capsid protein Vp54 comprises ∼40% of total virus protein. The primary translation product of its gene, *a430l*, has a predicted weight of 48,165 Da, but the protein undergoes additional posttranslational modifications so that the mature product has a molecular weight of approximately 54 kDa (9).

Previous studies established that glycosylation of Vp54 is unusual: the protein lacks *N*-acetylglucosamine and *N*-acetylgalactosamine, typical of Asn-linked (N-linked) and many Ser/Thr-linked (O-linked) glycoproteins (10). Instead, Vp54 glycans were reported to contain seven neutral monosaccharides, glucose, fucose, galactose, mannose, xylose, rhamnose, and arabinose (10). Crystallization of Vp54 at 2 Å resolution revealed four N-linked and two O-linked glycans (11). Comparison of the molecular mass of Vp54 with the predicted molecular weight from the amino acid sequence (48,165 Da) indicates that the protein contains ∼35 sugar moieties, of which ∼20 were detected, but not identified, in the crystal density map. At a minimum, single sugars were attached to ^{57}Ser and ^{387}Ser and branched-chain sugar moieties were attached to ^{280}Asn, ^{302}Asn, ^{399}Asn, and ^{406}Asn. Surprisingly, none of these asparagines reside in a N-X-(T/S) sequence, where X is anything but a Pro, commonly recognized by eukaryotic and prokaryotic enzymes involved in N-glycosylation. Instead, ^{302}Asn, ^{399}Asn, and ^{406}Asn occur in the amino acid sequence (A/G)-N-T-X-T and ^{280}Asn occurs in an A-N-I-P-G sequence. These results explain why previous tests for N-glycosylation in Vp54 were negative (12).

These findings suggest that the structure of the Vp54 glycans differ from typical glycans and prompted us to investigate their structures. The current manuscript describes the unusual structure of the PBCV-1 N-linked oligosaccharides.

Results

Isolation of the Two Largest Vp54 Glycans. Vp54 was extensively digested with proteinase K to minimize the length and the heterogeneity of the peptide moiety of the glycopeptides. The crude mixture was chromatographed on size-exclusion columns and four fractions containing glycopeptides were isolated. An ^{1}H

Author contributions: C.D.C., M.G.T., and J.L.V.E. designed research; C.D.C., A.M., F.P., J.R.G., L.S., A.P., R.L., M.P., D.G., and M.G.T. performed research; C.D.C., A.M., F.P., J.R.G., L.S., A.P., R.L., M.P., D.G., M.G.T., and J.L.V.E. analyzed data; and C.D.C., A.M., L.S., M.G.T., and J.L.V.E. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: jvanetten1@unl.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1313005110/-/DCSupplemental.

NMR spectrum of these samples showed the same pattern of the anomeric proton signals (5.7–4.3 ppm; Fig. S1), but differed in the number of signals around 4.5 ppm and for those occurring in the high and low field regions of the spectrum. This information suggested that the samples differed in the peptide portion containing the oligosaccharide, and also explained their different chromatographic retention volumes. With regard to the anomeric profile, not all of the protons occurred with the same intensity, as typically occurs when more glycoforms are present. Therefore, each fraction was investigated by MALDI MS at the same time the full NMR analysis was carried out on fraction 4. These approaches led to the full elucidation of the glycoforms attached to Vp54 (Fig. 1).

NMR Characterization of the Two Most Abundant Vp54 Glycoforms. The complete assignment of all resonances of the glycopeptide mixture in fraction 4 was performed by a combination of homo- and heteronuclear 2D NMR experiments. The heteronuclear single quantum correlation (HSQC) spectrum (Fig. S2; structures and labels in Fig. 1) contained 12 anomeric proton and one ring proton signals (range, 5.7–4.4 ppm), a crowded carbinolic region (4.4–3.0 ppm), two O-Me (3.47 and 3.45 ppm), and, eventually, four methyl signals at approximately 1.3 ppm, typical of 6-deoxy residues (Table S1).

The anomeric proton of **H** at 5.00 ppm was identified as an N-linked residue by virtue of its correlated carbon chemical shift at 80.5 ppm and therefore located at the reducing end of the oligosaccharide (13). In support of this conclusion, the **H** anomeric proton in the heteronuclear multiple bond correlation (HMBC) spectrum correlated with a carbonyl at 174.4 ppm, namely C^{γ} of the Asn residue. The *gluco* stereochemistry of **H** determined the efficient magnetization transfer in the total correlation spectroscopy (TOCSY) spectrum (Fig. S3*A*). Accordingly, the **H** anomeric signal displayed correlations up to H-6s and, integrating TOCSY information with those from the correlation spectroscopy (COSY) spectrum, the chemical shifts of all of the ring protons were assigned. Carbon chemical shift values of **H** were then determined analyzing HSQC (Fig. S2), HMBC (Fig. S4*A*), and HSQC-TOCSY (Fig. S4*B*) spectra. As a result, residue **H** was identified as a glucose β-configured at the anomeric center ($^3J_{H1H2}$, 9.1 Hz) and glycosylated at O-4; this last information was deduced by the low field displacement of the corresponding carbon (74.9 ppm) with respect to the standard values (70.6 ppm) (14). In agreement with this information, the HMBC spectrum correlated H-1 of **N** with C-4 of **H** (Fig. S4*A*) and contained an additional long-range correlation connecting H-3 of **H** to the anomeric proton of residue **A**. These results indicated that **H** was glycosylated at O-4 from **N** and at O-3 from **A**.

The same spectroscopic approach was applied to the two substituents of **H**, and in turn to all of the other residues. In some cases, TOCSY connectivity did not propagate along all of the protons of the sugar ring because of the stereochemistry of the residue; therefore, information from the HMBC or the transverse rotating-frame overhauser effect spectroscopy (TROESY) spectra were used to complete assignment of the residue (a detailed discussion is provided in *SI Results*). **N** was identified as a terminal β-xylose, whereas **A** was a fully substituted α-fucose carrying a terminal α-galactose (**E**) at O-2, an α-rhamnose (**F**) at O-3, and a β-xylose (**M**) at O-4 (Fig. 1). The α-rhamnose (**F**) was substituted at O-3 with a terminal α-mannose (**G**), whereas the xylose (**M**) had a β-rhamnose (**I**) at O-4. The β-rhamnose (**I**) also had a terminal α-rhamnose (**C**) at O-2 that was methylated at O-2 and O-3. With the addition of residue **C**, the structure of the first glycoform (glycoform **1**; Fig. 1) was completed.

Spectroscopic investigation of the left minor signals started with **L**, which exhibited the same pattern found for **I** (*SI Results*). Therefore, it was identified as a β-rhamnose linked at O-4 of **M**. The carbon chemical shift analysis of **L** revealed that it was glycosylated at O-2 and O-3, and TROESY spectrum (Fig. S3*B*) showed that O-3 was substituted with **D**, whereas O-2 was connected to **B**, a terminal α-rhamnose methylated at O-2 and O-3. The anomeric proton of **D** (Fig. S3*A*) correlated with five other protons in the TOCSY spectrum, and their assignment, together with the information from the HSQC spectrum (Table S1), identified this residue as a pentose monosaccharide in the furanose form because of the deshielded value of its C-4 (83.0 ppm). Comparison of **D** carbon chemical shifts with those of reference compounds (12) identified this unit as a terminal β-arabinofuranose residue. Therefore, the identification of residues **L**, **B**, and **D** allowed a second N-linked oligosaccharide to be constructed (glycoform **2**; Fig. 1), which differed from the major glycoform **1** by the presence of an additional residue.

Analysis of other minor signals in the spectra led to the identification of a threonine residue, whose occurrence was confirmed through MS analysis performed on fractions 3 and 4. Finally, the low intensity of the few remaining signals prevented any further spectroscopical assignment. These additional signals indicated that minute amounts of other glycoforms were present; information regarding these species was obtained by GC-MS and MALDI.

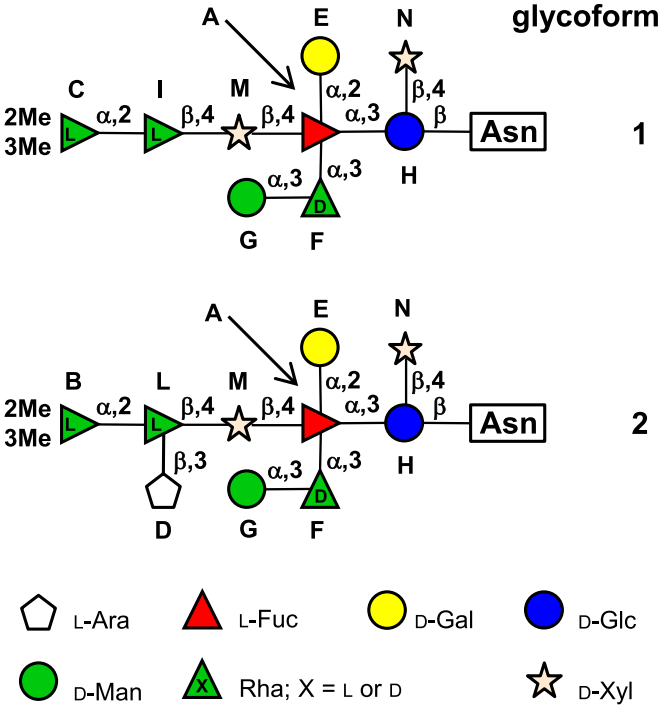

Fig. 1. Structures of glycoforms of the major capsid protein Vp54 from chlorovirus PBCV-1. Residues are labeled with the letter used during the NMR assignment. Glycoform **3** (or **4**) differs from **1** (or **2**) by lacking a mannose unit.

Chemical Composition of the Vp54 Glycans. The determination of the monosaccharide composition and linkage analysis was performed according to standard procedures (15) using fractions 3 and 4. Accordingly, residues identified in the NMR analyses together with their linkage patterns were confirmed. Determination of the absolute configuration of all residues was performed by analyzing the fully acetylated *R*-(-)-2-octylglycosides (15), which revealed the occurrence of L-arabinose, D-xylose, L-fucose, D-mannose, D-galactose, D-glucose, D- and L-rhamnose, and the 2,3-di-OMe-rhamnose derivative (units **B** and **C**; Fig. 1). For this latter residue, it was necessary to synthesize an ad hoc standard, the

4-O-Ac-2,3-di-OMe-rhamnose 2-(-)-octylglycoside. The compound was synthesized by methylating the O-antigen polysaccharide from *Salmonella enterica* ssp. *enterica* (16), which contains a 4-substituted L-rhamnose, the proper precursor of the target compound (Fig. S5).

As for the occurrence of D- and L-rhamnose residues, we surmised that the residues **I** (or **L**) and **F** of the two glycoforms possessed the opposite stereochemistry; accordingly, their configuration was inferred applying the same approach used to deduce the **C** configuration and using the polysaccharide from *Kaistella flava* (17) as a precursor for the standards (Fig. S6). Therefore, **I** possessed the L configuration, whereas **F** was the opposite D. All this information completed the structural data necessary to fully describe the two major glycoforms present in PBCV-1 virus Vp54 capsid protein (Fig. 1). This approach also revealed a small percentage of nonsubstituted D-rhamnose **F**, as later confirmed by MALDI MS.

Determination of the N-Glycosylation Sites and the Site-Specific Glycan Structure Heterogeneity. To identify the specific N-glycosylation sites, all Bio-Gel P10 fractions were analyzed by MALDI MS. The spectrum of fraction 2 (Fig. S7) showed two major peaks at *m/z* 1,916.66 and 1,932.64, and two less intense peaks at *m/z* 1,784.60 and 1,800.59 (*Δm/z* 132). The two main ions were ascribed to the Na^+ (*m/z* 1,916.64) and K^+ (*m/z* 1,932.64) ion adducts of the same glycopeptide composed from the ^{280}NIPG amino acidic sequence linked to glycoform **2**. The pair of minor peaks at *Δm/z* 132 (i.e., the 132 mass shift corresponding to a pentose residue) were assigned to the analogous glycopeptide ions composed of the same ^{280}NIPG sequence plus glycoform **1**. This interpretation was confirmed by the MS/MS fragmentation spectrum of the precursor ion at *m/z* 1,916.66 by using MALDI TOF/TOF MS/MS (Fig. 2; nomenclature in ref. 18).

MALDI spectra of fractions 3 and 4 comprised glycopeptides consistent with glycoform **1** and glycoform **2** linked to a single Asn or to a NT dipeptide. All these species could be located at glycosylation sites ^{302}Asn, ^{399}Asn, or ^{406}Asn; thus, they do not identify the specific glycosylation site.

Conversely, MALDI spectra of fraction 1 showed two major peaks, at *m/z* 1,848.58 and 1,980.62, whose assignments, based on the molecular masses and MS/MS spectra (MS/MS of 1,848.58 in Fig. S8*A*), were consistent with the glycoforms **1** and **2** (both present as Na^+ ions) linked at the ^{399}NTET amino acid sequence. The MS spectrum also contained a pair of less intense peaks at *m/z* 1,686.51 and *m/z* 1,818.55 that differed by a hexose monosaccharide (*Δm/z* 162) from the two main ions. Taking into consideration the compositional data, these peaks were assigned to two different glycoforms, named **3** and **4**, linked at the same tetrapeptide, with the same structure of glycoforms **1** and **2**, but lacking the mannose residue **G**.

To obtain meaningful glycopeptides with amino acid lengths suitable for site identification by MS, we digested Vp54 with a more selective protease. Trypsin was expected to produce two glycopeptides, the first containing ^{280}Asn and ^{302}Asn and the second containing ^{399}Asn and ^{406}Asn, with estimated molecular masses greater than *m/z* 6,300; these sizes are too large to acquire useful MS and MS/MS spectra. We then tried thermolysin because of its predicted specificity. Reverse-phase HPLC of the glycoprotein digestion with thermolysin yielded 25 fractions, mostly consisting of complex mixtures of peptides and glycopeptides; they were all analyzed by MALDI MS. Glycopeptides were identified by their specific spacing patterns (*Δm/z* 132 or 162) between the glycoforms, and by their MALDI TOF/TOF MS/MS fragmentation patterns. This process led to the identity of several glycopeptides composed of glycoforms **1** to **4** linked to ^{302}Asn, ^{399}Asn, or ^{406}Asn. Fig. S8 *B–D* shows the MS/MS spectra of the main glycopeptide ions, containing glycoform **1** as the glycan moiety. In summary, the combined digestion procedure revealed the identity of the oligosaccharide(s) present at a specific glycosylation site together with their proportion. Glycoform **2** is the prominent species at ^{280}Asn whereas glycoform **1** predominates over the other three glycoforms at the other glycosylation sites, even though most of the glycosylation heterogeneity occurs at ^{399}Asn and ^{302}Asn (Table 1).

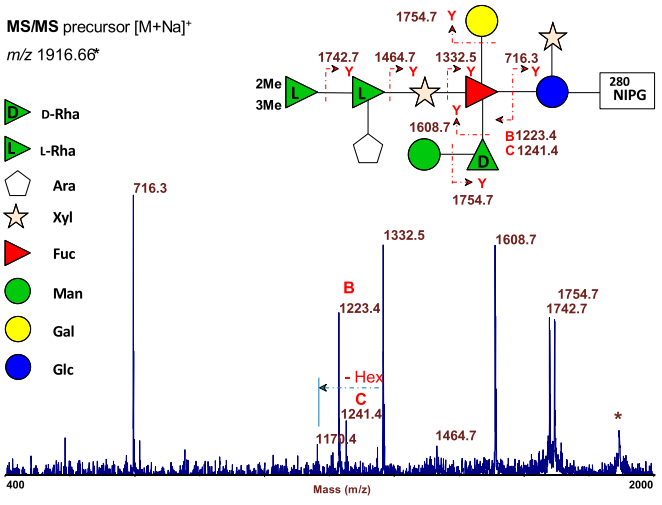

Fig. 2. MALDI TOF/TOF MS/MS spectrum of the precursor ion at *m/z* 1,916.66 of glycoform **2**, isolated from Vp54 protein after proteinase K digestion and Bio-Gel P10 purification. Structure of the glycopeptide (located at ^{280}NIPG) is drawn together with the fragmentation observed in the spectrum. Ion at *m/z* 1,170.4 results from dual fragmentation.

Discussion

A detailed chemical, spectroscopic, and spectrometric analysis allowed us to determine the structure of the four Asn-linked glycoforms attached to the major capsid protein Vp54 from virus PBCV-1 (Fig. 1). As expected, the structure of the Vp54 glycoforms are unique and do not resemble any type of structure reported so far for any organism in the three domains of life.

MALDI MS experiments extended our knowledge on the site-specific glycosylation of Vp54, demonstrating that there is a preference for the type of oligosaccharide at the four glycosylation sites. ^{280}Asn primarily has glycoform **2**, whereas glycoform **1** is the main oligosaccharide at ^{406}Asn. ^{302}Asn and ^{399}Asn are the positions where most of the major glycosylation heterogeneity occurs, even though glycoform **1** is still the main oligosaccharide.

Therefore, PBCV-1 glycosylation opens many questions regarding its biosynthetic pathway. PBCV-1 encodes five putative glycosyltransferases, A064R, A111/114R, A219/222/226R, A473L, and A546L, which are scattered throughout the PBCV-1 genome (4, 6). The A111/114R protein has two glycosyltransferase domains. However, clearly, there are more glycan linkages than virus-encoded glycosyltransferases, which may result from some of the viral-

Table 1. Proportion and type of oligosaccharides found at each glycosylation site in PBCV-1 major capsid protein Vp54

	Glycoform, %			
Position	1	2	3	4
^{280}Asn	5	94	—	4
^{302}Asn	58	10	21	—
^{399}Asn	45	32	12	12
^{406}Asn	71	—	29	—

The structures are reported in Fig. 1.

encoded glycosyltransferases serving more than one function or some host-encoded enzymes are involved in the process. Like Vp54, none of these PBCV-1–encoded glycosyltransferases has an identifiable signal peptide that would target them to the ER of the host. Furthermore, the cellular protein localization program PSORT predicts that all these proteins, with the exception of A473L, are located in the cytoplasm. None of these five PBCV-1–encoded putative glycosyltransferases have been enzymatically characterized until now because potential substrates were unknown. Solving the structure of the final biosynthetic products will allow functional characterization of these enzymes.

The most abundant oligosaccharide contains nine sugar residues (glycoform **1**; Fig. 1): D- and L-rhamnose, 2,3-di-OMe-L-rhamnose, L-fucose, D-glucose, D-mannose, D-galactose, and D-xylose (two units), organized in a highly branched fashion. This glycoform can be dissected into a main chain, consisting of the five residues **C-I-M-A-H** (Fig. 1) further substituted with two monosaccharides, **E** and **N**, and one disaccharide, **G-F**. Glycoform **2** possesses an additional branch carrying a terminal L-arabinofuranose residue, unit **D**. The other two minor glycoforms, **3** and **4**, differ from the first two by the lack of the mannose residue **G**.

The structures of these oligosaccharides have several remarkable features. The glycoprotein β-Glc-Asn linkage is rare in nature and has only been reported in glycoproteins from a few organisms. It was first described in surface glycopeptides and flagellins of the extreme halophilic archaea *Halobacterium halobium* (19, 20) and *Halobacterium volcanii* (21). Later, this glycoside linkage was reported in bacterial adhesion glycoproteins from *Haemophilus influenzae* (22, 23) and *Actinobacillus pleuropneumoniae* (13). In these cases, the Glc-Asn linkage was found in the typical sequon for N-glycosylation, and Glc was elongated with short linear chains of Glc or glucuronic acid. In eukaryotic organisms, a β-Glc-Asn linkage was reported in the B2 chain of rat kidney laminin, a major basement membrane glycoprotein, by using antibody that specifically recognizes this linkage (24). However, the glycoprotein has not been characterized further, leaving this N-glucosylation example as the only exception to the general N-glycosylation pattern in Eukarya.

Thus, the β-Glc-Asn linkage, although rare, has been reported in all three domains of life. However, the virus PBCV-1 glycosylation of Vp54 differs in many aspects from the systems previously mentioned. (*i*) The four N-linked glycans are not located in a typical N-X-(T/S) consensus site. Three of the four glycosylation sites share the sequence NTXT (X represents A, E, or G), whereas the fourth is NIPG; all of these asparagine flanking residues denote a new pattern of N-glycosylation. (*ii*) The PBCV-1 N-glycans are more complicated than those reported previously. In fact, the main Vp54 glycoform consists of nine different neutral sugars arranged in a highly branched fashion with several unusual features. (*iii*) The fucose unit is completely substituted and not terminal like in most N-linked glycans (25). Extending our view to all known carbohydrate structures, regardless of the type of glycosylation and source, Vp54 glycoforms differ from those reported in the literature. As far as we know, a hyperbranched fucose has been reported only in the complex phosphoglycan from *Trypanosoma cruzi* (26), but its substitution pattern does not resemble that of Vp54 glycoforms. (*iv*) The terminal residue at the nonreducing end of the main chain is a rhamnose capped with two methyl groups, which is a rare type of termination and so far only reported for glycolipids produced from *Mycobacterium haemophilum* (27) and *Mycobacterium leprae* (28). (*v*) The Vp54 oligosaccharide contains both D- and L-rhamnose residues, a feature that is rare, occurring only in bacteria and that returned 15 hits for natural oligosaccharides in the Bacterial Carbohydrates Structure Database (http://csdb.glycoscience.ru/bacterial). Interestingly, PBCV-1 encodes a functional "de novo" GDP-L-fucose biosynthetic pathway, which also produces GDP-D-rhamnose as a major product (29). Thus, PBCV-1 encodes a more complex glycosylation machinery, which comprises the glycosyltransferases and some of the enzymes required to produce the nucleotide-sugar substrates, which presumably are present in limited supply or absent in the host cell.

In summary, the structures of four glycans associated with virus PBCV-1 major capsid protein Vp54 are very unusual and provide additional support for the hypothesis that PBCV-1 and probably all of the chloroviruses glycosylate their major capsid proteins differently than all of the typical known viruses, and in a different fashion with respect to the other known glycosylation pathways. Chloroviruses and other members of the family *Phycodnaviridae* are believed to have a long evolutionary history, possibly dating back to the time when eukaryotes arose from prokaryotes (30–33), and we have suggested that the chlorovirus glycosylation pathway may be ancient, possibly existing before the development of the ER and Golgi (9).

In the past few years, increasing evidence has indicated that other giant viruses, in particular those belonging to the *Mimiviridae* family, encode enzymes involved in glycan production (34). Moreover, genome sequencing of several other members of the *Phycodnaviridae* family, which are closely related to the chloroviruses, indicate that they encode putative glycosyltransferases and other genes involved in glycan formation and remodeling (35). Some of these viruses are widely distributed in the oceans, where they are active participants in ecological and evolutionary events (36). Altogether, these data indicate that the presence of glycosylation machinery can no longer be considered as a hallmark solely of cellular organisms, but that viruses also encode unique and complex glycan systems, which are still mostly unknown. Identification of the Vp54 glycan structure represents a starting point to understanding these systems.

Materials and Methods

Glycan Isolation. Procedures for growing and purifying virus PBCV-1 have been described (37–39). The Vp54 capsid protein was purified to near-homogeneity as reported previously (10). Seven milligrams of Vp54 were suspended in 1 mL of 100 mM Tris, 50 mM NaCl, 10 mM $MgCl_2$, pH 7.5, and digested with proteinase K (P6556; Sigma) at 55 °C. To ensure complete digestion, proteinase K was added three times (0.7 mg each time) at 8-h intervals. The mixture was freeze-dried, suspended in water, and purified on a Bio-Gel P-10 chromatographic column (1.5 × 118 cm; 150-4144; Bio-Rad) eluted with distilled water. The chromatographic profile was monitored with an online refractive index detector (K-2310; Knauer), and fractions were pooled accordingly and monitored via ^{1}H NMR. Carbohydrate containing fractions 1 (0.1 mg), 2 (0.26 mg), 3 (0.4 mg), and 4 (0.5 mg) were identified and eluted at 37.0%, 52.9%, 56.7%, and 61.5% of the column volume.

Glycan Analysis. Monosaccharide analysis was performed on fractions 3 and 4 and required the preparation of the corresponding: acetylated alditols, partially methylated and acetylated alditols, acetylated methylglycosides, partially methylated and acetylated octylglycosides, or only acetylated octylglycosides, by using standard protocols (15). All GC-MS analyses were performed with an Agilent instrument (GC instrument Agilent 6850 coupled to an Agilent 5973 MS instrument), equipped with a SPB-5 capillary column (30 m × 0.25 i.d.; flow rate, 0.8 mL·min^{-1}; Supelco) and He as carrier gas. Electron impact mass spectra were recorded at an ionization energy of 70 eV and an ionizing current of 0.2 mA. The temperature program used for the analyses was: 150 °C for 5 min, from 150 to 280 °C at 3 °C/min, 300 °C for 5 min.

NMR Spectroscopy. All NMR experiments were carried out on a Bruker DRX-600 spectrometer equipped with a cryoprobe. Chemical shifts of spectra recorded in D_2O are expressed in δ relative to internal acetone (2.225 and 31.45 ppm). Two-dimensional spectra [double-quantum filtered (DQ)-COSY, TOCSY, TROESY, HSQC, gHMBC, and HSQC-TOCSY] were measured by using standard Bruker software. For all experiments, 512 free induction decays (FIDs) of 2,048 complex data points were collected, 40 scans per FID were acquired for homonuclear spectra, and mixing times of 120 and 250 ms were used for TOCSY and for TROESY spectra acquisition, respectively. The spectral width was set to 10 ppm and the frequency carrier placed at the residual HOD peak. HSQC spectrum was acquired with 50 scans per

FID, and the globally-optimized, alternating phase, rectangular pulses (GARP) sequence was used for ^{13}C decoupling during acquisition. HSQC-TOCSY and gHMBC scans tripled and doubled, respectively, those of HSQC spectrum, and a mixing time of 100 ms was used for HSQC-TOCSY spectrum. Data processing was performed with the standard Bruker Topspin 3 program.

MS. Purified Vp54 was digested using thermolysin (Promega) in 50 mM Tris·HCl, pH 7.8, 0.5 mM $CaCl_2$, at 80 °C for 4 h. The peptides were then separated by reversed-phase HPLC on a Agilent 1200 system by using an Alltech C18 4.6 × 250-mm, 5-μM particle size column, with flow at 1 mL/min. Isocratic mobile phase A [acetonitrile/H_2O (5:95) containing 0.1% TFA] was used for 5 min; then, a linear gradient from 0% to 100% mobile phase B [acetonitrile/H_2O (90:10) containing 0.05% TFA] over 50 min was applied. Detection was performed by UV absorbance at 220 nm. Fractions, collected every 0.5 min, were dried by using a Speed-Vac and then directly used for MS analyses.

MALDI-TOF mass spectra were recorded in a linear mode on a PerSeptive Voyager STR equipped with a pulsed UV laser beam (nitrogen laser, $\lambda = 337$ nm) and in reflectron mode on a 4800 Proteomic Analyzer (Applied Biosystems) supplied with a Nd:yttrium–aluminum garnet laser at a wavelength of 355 nm. Mass spectra were acquired twice using 2,5-dihydroxybenzoic acid (DHB) oralphacyano-4-hydroxycinnamic acid, respectively, as the matrix. DHB has proved to be a better matrix for glycopeptides. MALDI TOF/TOF MS/MS spectra were also recorded on the 4800 system without any collision gas.

ACKNOWLEDGMENTS. Work in the J.L.V.E Etten laboratory was partially supported by National Science Foundation-Experimental Program to Stimulate Competitive Research Grant EPS-1004094, Stanley Medical Research Institute Grant 11R-0001; National Center for Research and Resources Centers of Biomedical Research Excellence (COBRE) Program Grant P20-RR15635; Italian Ministry for University and for Scientific and Technology Research Grants PRIN 2009 and FIRB-MERIT RBNE08HWLZ; Regione Liguria; the National Council of Research; and Programma Operativo Regionale Campania Project Campania Research in Experimental Medicine Fondo Europeo Sviluppo 2007-2013 (to C.D.C. and R.L.).

1. Doms RW, Lamb RA, Rose JK, Helenius A (1993) Folding and assembly of viral membrane proteins. *Virology* 193(2):545–562.
2. Olofsson S, Hansen JES (1998) Host cell glycosylation of viral glycoproteins—a battlefield for host defence and viral resistance. *Scand J Infect Dis* 30(5):435–440.
3. Vigerust DJ, Shepherd VL (2007) Virus glycosylation: Role in virulence and immune interactions. *Trends Microbiol* 15(5):211–218.
4. 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(2): 152–159.
5. Van Etten JL (2003) Unusual life style of giant chlorella viruses. *Annu Rev Genet* 37: 153–195.
6. Dunigan DD, Fitzgerald LA, Van Etten JL (2006) Phycodnaviruses: A peek at genetic diversity. *Virus Res* 117(1):119–132.
7. Van Etten JL, Dunigan DD (2012) Chloroviruses: Not your everyday plant virus. *Trends Plant Sci* 17(1):1–8.
8. Dunigan DD, et al. (2012) *Paramecium bursaria* chlorella virus 1 proteome reveals novel architectural and regulatory features of a giant virus. *J Virol* 86(16):8821–8834.
9. 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(2):332–345.
10. Wang I-N, et al. (1993) Evidence for virus-encoded glycosylation specificity. *Proc Natl Acad Sci USA* 90(9):3840–3844.
11. 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(23):14758–14763.
12. Que Q, et al. (1994) Protein glycosylation and myristylation in Chlorella virus PBCV-1 and its antigenic variants. *Virology* 203(2):320–327.
13. Schwarz F, Fan Y-Y, Schubert M, Aebi M (2011) Cytoplasmic N-glycosyltransferase of *Actinobacillus pleuropneumoniae* is an inverting enzyme and recognizes the NX(S/T) consensus sequence. *J Biol Chem* 286(40):35267–35274.
14. Bock K, Pedersen S (1983) Carbon-13 nuclear magnetic resonance spectroscopy of monosaccharides. *Adv Carbohydr Chem Biochem* 41:27–65.
15. De Castro C, Parrilli M, Holst O, Molinaro A (2010) Microbe-associated molecular patterns in innate immunity: Extraction and chemical analysis of gram-negative bacterial lipopolysaccharides. *Methods Enzymol* 480:89–115.
16. De Castro C, Lanzetta R, Leone S, Parrilli M, Molinaro A (2013) The structural elucidation of the *Salmonella enterica* subsp. *enterica*, reveals that it contains both O-factors 4 and 5 on the LPS antigen. *Carbohydr Res* 370:9–12.
17. Gargiulo V, et al. (2008) Structural elucidation of the capsular polysaccharide isolated from *Kaistella flava*. *Carbohydr Res* 343(14):2401–2405.
18. Domon B, Costello CE (1988) A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconj J* 5:397–409.
19. Wieland F, Heitzer R, Schaefer W (1983) Asparaginylglucose: Novel type of carbohydrate linkage. *Proc Natl Acad Sci USA* 80(18):5470–5474.
20. Wieland F, Paul G, Sumper M (1985) Halobacterial flagellins are sulfated glycoproteins. *J Biol Chem* 260(28):15180–15185.
21. Mengele R, Sumper M (1992) Drastic differences in glycosylation of related S-layer glycoproteins from moderate and extreme halophiles. *J Biol Chem* 267(12): 8182–8185.
22. Gross J, et al. (2008) The *Haemophilus influenzae* HMW1 adhesin is a glycoprotein with an unusual N-linked carbohydrate modification. *J Biol Chem* 283(38):26010–26015.
23. Grass S, Lichti CF, Townsend RR, Gross J, St Geme JW, 3rd (2010) The *Haemophilus influenzae* HMW1C protein is a glycosyltransferase that transfers hexose residues to asparagine sites in the HMW1 adhesin. *PLoS Pathog* 6(5):e1000919.
24. Schreiner R, Schnabel E, Wieland F (1994) Novel N-glycosylation in eukaryotes: Laminin contains the linkage unit beta-glucosylasparagine. *J Cell Biol* 124(6):1071–1081.
25. Stanley P, Schachter H, Taniguchi N (2009) *Essentials of Glycobiology*, eds Varki A, et al. (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY), pp 101–114.
26. Allen S, Richardson JM, Mehlert A, Ferguson MAJ (2013) Structure of a complex phosphoglycan epitope from gp72 of *Trypanosoma cruzi*. *J Biol Chem* 288(16): 11093–11105.
27. Besra GS, et al. (1991) Structural elucidation and antigenicity of a novel phenolic glycolipid antigen from *Mycobacterium haemophilum*. *Biochemistry* 30(31): 7772–7777.
28. Daffé M, Lanéelle MA (1989) Diglycosyl phenol phthiocerol diester of *Mycobacterium leprae*. *Biochim Biophys Acta* 1002(3):333–337.
29. Tonetti M, et al. (2003) *Paramecium bursaria* Chlorella virus 1 encodes two enzymes involved in the biosynthesis of GDP-L-fucose and GDP-D-rhamnose. *J Biol Chem* 278(24):21559–21565.
30. Iyer LM, Aravind L, Koonin EV (2001) Common origin of four diverse families of large eukaryotic DNA viruses. *J Virol* 75(23):11720–11734.
31. Iyer LM, Balaji S, Koonin EV, Aravind L (2006) Evolutionary genomics of nucleocytoplasmic large DNA viruses. *Virus Res* 117(1):156–184.
32. Raoult D, et al. (2004) The 1.2-megabase genome sequence of Mimivirus. *Science* 306(5700):1344–1350.
33. Villarreal LP, DeFilippis VR (2000) A hypothesis for DNA viruses as the origin of eukaryotic replication proteins. *J Virol* 74(15):7079–7084.
34. Piacente F, et al. (2012) Giant DNA virus mimivirus encodes pathway for biosynthesis of unusual sugar 4-amino-4,6-dideoxy-D-glucose (Viosamine). *J Biol Chem* 287(5): 3009–3018.
35. Weynberg KD, Allen MJ, Gilg IC, Scanlan DJ, Wilson WH (2011) Genome sequence of *Ostreococcus tauri* virus OtV-2 throws light on the role of picoeukaryote niche separation in the ocean. *J Virol* 85(9):4520–4529.
36. Suttle CA (2007) Marine viruses—major players in the global ecosystem. *Nat Rev Microbiol* 5(10):801–812.
37. VAN Etten JL, Burbank DE, Kuczmarski D, Meints RH (1983) Virus infection of culturable chlorella-like algae and dlevelopment of a plaque assay. *Science* 219(4587): 994–996.
38. Van Etten JL, Burbank DE, Xia Y, Meints RH (1983) Growth cycle of a virus, PBCV-1, that infects Chlorella-like algae. *Virology* 126(1):117–125.
39. Agarkova IV, Dunigan DD, Van Etten JL (2006) Virion-associated restriction endonucleases of chloroviruses. *J Virol* 80(16):8114–8123.

Supporting Information

De Castro et al. 10.1073/pnas.1313005110

SI Results

NMR Characterization of Two More Abundant Vp54 Glycans. NMR investigation of fraction 4 was performed at 310 K (Table S1; structures in Fig. 1), and the heteronuclear single quantum correlation (HSQC) spectrum (Fig. S2) showed 12 anomeric and one ring proton in the range of 5.7 to 4.4 ppm, a crowded carbinolic region (4.4–2.93 ppm) containing, *inter alia*, two O-Me signals (3.47 and 3.45 ppm), two diasterotopic geminal protons at 3.00 and 2.93 ppm attributed to the methylene of the Asn residue, and four methyl signals at approximately 1.3 ppm, typical of 6-deoxyresidues.

The anomeric protons were labeled with a capital letter, and NMR analysis started from the anomeric proton of **H** at 5.00 ppm, identified as an N-linked residue by virtue of its carbon chemical shift at 80.5 ppm (1). In support of this information, H-1 of **H** correlated, in the heteronuclear multiple bond correlation (HMBC) spectrum, with a carbonyl at 174.4 ppm, namely C^{γ} of the Asn residue; all the proton and carbon chemical shifts for this amino acid are reported in Table S1.

The *gluco* stereochemistry of **H** was recognized by analyzing the anomeric region of the total correlation spectroscopy (TOCSY) spectrum (Fig. S3*A*). In this experiment, the efficiency of the magnetization transfer is directly proportional to the magnitude of the vicinal coupling constants $^3J_{H,H}$ and hence by the stereochemistry of the monosaccharide investigated. Accordingly, H-1 of **H** showed the complete correlation pattern up to H-6s; integrating TOCSY information with those from correlation spectroscopy (COSY), led to the chemical shift assignment of all of the ring protons. Carbon chemical shift values of **H** were determined analyzing the heteronuclear correlated spectra HSQC, HMBC (Fig. S4*A*), and HSQC-TOCSY (Fig. S4*B*). As result, residue **H** was a glucose β-configured at the anomeric center ($^3J_{H1H2}$, 9.1 Hz), glycosylated at O-4 as suggested from the low field value of the corresponding carbon (74.9 ppm) with respect to the standard values (70.6 ppm) (2). In agreement with this information, HMBC spectrum correlated H-1 of **N** with C-4 of **H** (Fig. S4*A*), and contained an additional long-range signal connecting H-3 of **H** to the anomeric proton of residue **A**. Therefore, **H** was glycosylated at O-4 from **N** and at O-3 from **A**; C-3 of **H** did not display the expected glycosylation shift because it was probably cancelled by the counteracting β-glycosylation effect of the substituent at position O-4. This spectroscopic analysis approach was applied to the two substituents of **H** and to all of the others found thereafter.

H-1 of **N** displayed intense correlations in the TOCSY spectrum (Fig. S3*A*) with five other protons; two of them (4.03 and 3.28 ppm) were located on the same carbon atom, and indeed this monosaccharide was identified as a xylose in the pyranose form (C-5 at 66.1 ppm), β-configured ($^3J_{H1H2}$ 8.0 Hz), and not further substituted as deduced by its carbon chemical shift values.

Residue **A** was α-configured ($^3J_{H1H2}$, 3.3 Hz) fucose: TOCSY spectrum (Fig. S3*A*) displayed three correlations, two almost coincident at 4.18 and 4.21 ppm (H-2 and H-3, respectively) and one at 3.88 ppm. COSY spectrum pointed at the signal at 4.18 ppm as H-2 whereas H-2/H-3 cross peak was not detected because it was merged into the spectrum diagonal, but the cross peak linking the signals at 4.21 and at 3.88 ppm completed the proton assignment up to H-4 proton. C-2 to C-4 chemical shifts were identified analyzing the HSQC and HSQC-TOCSY spectra; identification of H-5/C-5 occurred through the analysis of the HMBC spectrum (Fig. S4*A*), which correlated the **A** anomeric proton with two different carbons, its C-3 and one at 68.3 ppm with the corresponding proton at 4.75 ppm. This proton was H-5 of **A** because it correlated with a methyl carbon at 1.30 ppm and with the previously recognized C-4. Indeed, the complete attribution of **A** was accomplished, and analysis of the carbon chemical shift established its complete glycosylation: actually, C-2, C-3, and C-4 resonated at 71.7, 76.2, and 81.9 ppm, respectively, and each displayed a low field glycosylation shift with respect to the standard values (69.0, 70.6, and 72.9 ppm, respectively) (2).

HMBC spectrum (Fig. S4*A*) disclosed that **A** had **M** at O-4 and **F** at O-3 (Fig. 1) and the substituent at O-2 was determined by analyzing the transverse rotating-frame overhauser effect spectroscopy (TROESY) spectrum (Fig. S3*B*) and was **E**. The spectroscopic pattern of **E** was similar to that of **A**, with the difference that H-5 of **E** had a long-range correlation with a carbon at 62.6 ppm. Therefore, **E** was a galactose unit, α-configured at the anomeric center ($^3J_{H1H2}$, 3.8 Hz), whereas the carbon chemical shift values identified this unit as terminal. The anomeric proton of residue **F** displayed three TOCSY correlations (Fig. S3*A*), with the H-1/H-2 cross peak more intense than the others, as occurs for *manno* configured residues. COSY and TOCSY analyses determined the position of all the ring protons up to H-6s, which resonated at 1.29 ppm. Therefore, **F** was as a rhamnose, glycosylated at O-3 (C-3 at 79.8), and its C-5 value at 70.4 ppm indicated the α-anomeric configuration of the residue (2).

The substituent of **F** was **G** as deduced from the long-range correlation in the HMBC spectrum (Fig. S4*A*). Residue **G** displayed a spectroscopical pattern similar to that of **E**, with the difference that its C-6 resonated at 62.8 ppm; indeed, **G** was a mannose, not further substituted because of the carbon chemical shift values and α-configured on the basis of C-3 and C-5 chemical shifts with respect to the standard values (2).

With regard to **M**, its spectroscopic pattern was similar to that of **N**, so **M** was a β-xylose ($^3J_{H1H2}$, 7.8 Hz), and, differently from **N**, it was glycosylated at O-4 as suggested by the deshielded value of the corresponding carbon (80.9 ppm) with respect to the standard value (70.4 ppm) (2).

HMBC spectrum (Fig. S4*A*) showed that two different residues, **I** and **L**, were connected at O-4 of **M**; these two residues had different intensities and, as determined later, reflected the presence of two glycoforms.

Determination of the sugar chain sequence proceeded analyzing the more intense **I** unit. TOCSY correlation from H-1 stopped at H-2 (Fig. S3*A*), but, starting from H-2, four other correlations were seen, one with a methyl unit at 1.32 ppm. Indeed, residue **I** was a rhamnose and its H-5 resonated at 3.48 ppm, a value typical of a β-configured rhamnose. This information was supported by the TROESY spectrum (Fig. S3*B*), which correlated H-1 with both H-3 and H-5, a pattern resulting from the *cis* diaxial orientation of these three protons, and typical of β-configured residues. **I** was glycosylated at O-2 because of the deshielded value of the corresponding carbon and HMBC spectrum (Fig. S4*A*) identified **C** as substituent.

Similar to **F**, H-1 of **C** displayed one intense H1/H2 correlation in the TOCSY spectrum and three additional ones (Fig. S3*A*); COSY and TOCSY analysis determined the position of all the ring protons up to H-6s, which resonated at 1.26 ppm. This information identified **C** as a rhamnose, α-configured at the anomeric center (C-5 at 69.7 ppm), and substituted at O-2 and O-3 with a methyl group, seen via a long-range correlation in the HMBC spectrum. Indeed, **C** was a terminal α-configured rhamnose residue, methylated at O-2 and O-3. With the attribution of this last residue, the

structure of the first glycoform (Fig. 1) was accomplished, and spectroscopic investigation proceeded with the evaluation of the other minor signals.

L was chosen as the start, and it displayed the same spectroscopical pattern described for **I**; indeed, **L** was a β-configured rhamnose glycosylated at C-2, but, different from **I**, also at C-3. The substituent at C-2 of **L** was detected in the TROESY spectrum (Fig. S3*B*), which had a cross peak correlating H-1 of **B** with H-2 of **L**; **B** was a terminal rhamnose, α-configured at the anomeric center and substituted with two methyl groups at O-2 and O-3, exactly as found for **C**.

Last, the TOCSY spectrum of the anomeric proton of the **D** residue showed five different correlations, and the combined analysis of TOCSY, COSY, and HSQC led to the definition of all the proton and carbon chemical shifts of this unit (Table S1). **D** was a pentose in the furanose form because of the deshielded value of its C-4 (83.0 ppm) together with the **C**-5 value (62.8 ppm). Comparison of **D** carbon chemical shifts with those of reference compounds (2) identified this unit as a β-configured arabinofuranose residue. HMBC did not contain information relevant to an understanding of the **D** location, which was placed at O-3 of **L** on the basis of the TROESY spectrum (Fig. S3*B*); actually, H-1 of **D** was in close proximity with two protons of **L**, H-2 and H-3, and, considering that O-2 of **L** was already glycosylated from **B**, residue **D** was placed at O-3 of **L**, completing the structure determination of the second virus glycoform (Fig. 1).

Interestingly, the occurrence of the **D** residue at C-3 of the β-rhamnose of the main glycan chain induced magnetic differentiation of a residue not directly involved in the linkage; namely, depending on the presence of **D**, the terminal methylated rhamnose appeared as a **B** or as a **C** unit. Analysis of other minor signals in the spectra led to the identification of a threonine residue, whose occurrence was confirmed by MS analysis performed on fractions 2 and 3. Finally, the low intensity of the few remaining signals prevented any further spectroscopical assignment. These other signals reflected the occurrence of minute amounts of other glycoforms, and information regarding these species was obtained by GC-MS and MALDI.

Determination of the Absolute Configuration of the Monosaccharide Residues of Vp54 Glycans. Determination of the absolute configuration of all residues was performed by analyzing the fully acetylated *R*-(-)-2-octylglycosides (3), which revealed the occurrence of L-arabinose, D-xylose, L-fucose, D-mannose, D-galactose and D-glucose, and both D- and L-forms of rhamnose, together with one species related to the 2,3-diMe rhamnose derivative (units **B** and **C**; Fig. 1). For the double-methylated rhamnose 2-(-)-octylglycoside, the standard was prepared by derivatizing the O-antigen polysaccharide from *Salmonella enterica* ssp. *enterica* (4), which contains a 4-linked L-rhamnose as the only 6-deoxyhexose residue. The sample was completely methylated (3) and split in two aliquots, the first was treated with racemic 2-octanol and the second with pure *R*-(-)-2-octanol; finally, each sample was acetylated and analyzed via GC-MS. Because of the chemical manipulation, the 4-linked rhamnose was transformed in the corresponding octylglycoside bearing at position 4 an acetyl group and methyl groups at positions 2 and 3, identical to the rhamnose derivative present in the octylglycosides from fraction 4.

The rhamnose derivative with racemic 2-octanol was identified in the GC-MS chromatogram, applying the fragmentation rules reported for methylglycosides (5); it gave two peaks (Fig. S5*A*) at 18.88 and 19.80 min, sharing the same electron impact (EI)-MS spectrum (Fig. S5*B*) and characterized by the fragment at *m*/*z* 217 diagnostic of the oxonium ion, and from the ion at *m*/*z* 88, related to the occurrence of two vicinal methyl groups. Comparing this chromatogram with that acquired for the derivative with the enantiomeric pure octanol, it was possible to determine the retention time of two diastereoisomers, so that 4-Ac-2,3-di-OMe-2-(-)-octyl-L-rhamnoside eluted first, and 4-Ac-2,3-diOMe-2-(+)-octyl-L-rhamnoside [equivalent to its enantiomer 4-Ac-2,3-di-OMe-2-(-)-octyl-D-rhamnoside] eluted later. This information was applied to the 4-Ac-2,3-diMe-2-(-)-octylrhamnoside present in fraction 4 and its L absolute configuration was determined.

Discrimination of the D- and L-rhamnose present in the two glycoforms was solved by methylating fraction 3 and preparing the corresponding chiral partially methylated and acetylated octylglycosides (3). This approach allowed the differentiation of the four types of rhamnose residues occurring in this glycopeptide (Fig. 1): unit **B** (or **C**) was detected as fully methylated 2-(-)-octyl-rhamnoside, unit **I** as 2-O-acetyl-3,4-diOMe-2-(-)-octyl-rhamnoside, unit **F** as 3-O-acetyl-2,4-diOMe-2-(-)-octyl-rhamnoside, and unit **L** as 2,3-diOacetyl-4-OMe-2-(-)-octyl-rhamnoside.

The standard for the fully methylated rhamnoside was prepared by treating L-rhamnose with racemic or *R*-(-)-2-octanol, methylating the final glycoside mixture (3), and confirming the identity of residue **C** (or **B**) as L-Rha. During this analysis, a small amount of D-Rha was detected, suggesting the existence of a minor glycoform with a rhamnose **F** that had no additional substitutions.

Synthesis of the partially methylated and acetylated 2-octylrhamnosides that were used as standards was performed by methylating the polysaccharide from *Kaistella flava* (6) and transforming it into the corresponding acetylated octylglycosides (Fig. S6), as done for the O-antigen from *Salmonella enterica* ssp. *enterica*. The polysaccharide from *Kaistella* was selected because it contains L-rhamnose in three different forms: 2-linked (as residue **I**), 3-linked (as residue **F**), and 2,3-linked (as residue **L**); therefore, it produces the same type of derivatives expected for the two glycoforms found in Vp54 protein, in addition to 3-linked fucosamine and terminal glucosamine.

Identification of the derivatives from the 2-linked rhamnose at 16.15 and 16.27 min (Fig. S6*A*) was possible by analyzing the EI-MS spectrum (Fig. S6*B*), which displayed the oxonium ion at *m*/*z* 217 and an intense ion at *m*/*z* 88, diagnostic of the occurrence of two vicinal methylated hydroxyl group; 3-linked rhamnose derivatives (Fig. S6*C*) were individuated at 18.01 and 18.32 min, and displayed the same oxonium ion, but not the fragment at *m*/*z* 88, because of the different substitution pattern of the hydroxyl functions. The 2,3-linked rhamnose derivatives were identified at 20.69 and 20.89 min on the basis of the expected oxonium ion (Fig. S6*D*). Comparing this chromatogram with that acquired for the derivative with the enantiomeric pure octanol, it was possible to determine the retention time of each couple of diastereoisomers so that determination of the absolute configuration of the rhamnose derivatives present in fraction 3 was possible. Indeed, residue **I** and **L** were L-configured, whereas **F** was D-configured, completing the structural information necessary to fully describe the glycoforms present in Vp54 envelope protein from *Paramecium bursaria* chlorella virus 1 (Fig. 1).

1. Schwarz F, Fan Y-Y, Schubert M, Aebi M (2011) Cytoplasmic N-glycosyltransferase of *Actinobacillus pleuropneumoniae* is an inverting enzyme and recognizes the NX(S/T) consensus sequence. *J Biol Chem* 286(40):35267–35274.
2. Bock K, Pedersen S (1983) Carbon-13 nuclear magnetic resonance spectroscopy of monosaccharides. *Adv Carbohydr Chem Biochem* 41:27–65.
3. De Castro C, Parrilli M, Holst O, Molinaro A (2010) Microbe-associated molecular patterns in innate immunity: Extraction and chemical analysis of gram-negative bacterial lipopolysaccharides. *Methods Enzymol* 480:89–115.
4. De Castro C, Lanzetta R, Leone S, Parrilli M, Molinaro A (2013) The structural elucidation of the *Salmonella enterica* subsp. *enterica*, reveals that it contains both O-factors 4 and 5 on the LPS antigen. *Carbohydr Res* 370:9–12.
5. Lönngren J, Svensson S (1974) Mass spectrometry in structural analysis of natural carbohydrates. *Adv Carbohydr Chem Biochem* 29:41–106.
6. Gargiulo V, et al. (2008) Structural elucidation of the capsular polysaccharide isolated from *Kaistella flava*. *Carbohydr Res* 343(14):2401–2405.

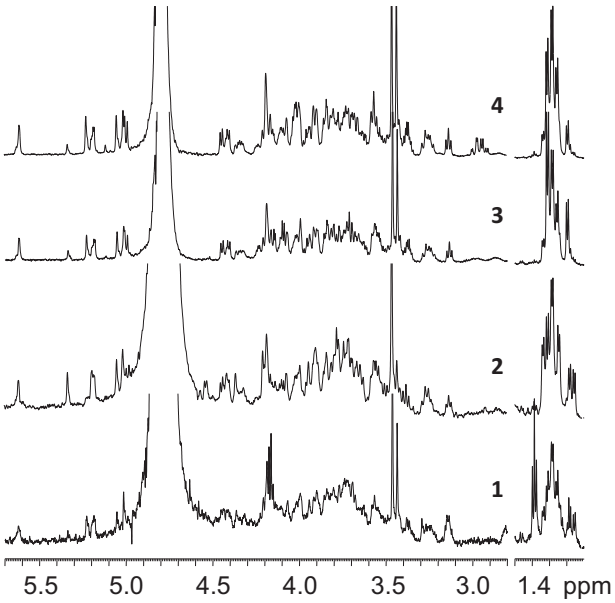

Fig. S1. Expansion of the proton spectra recorded for fractions 1 to 4, isolated from Vp54 protein after proteinase K digestion and Bio-Gel P10 purification. Anomeric region: 5.7 to 4.4 ppm; carbinolic region, 4.4 to 3.0 ppm; O-Me signals at 3.47 and 3.45 ppm; methyl signals of 6-deoxy residues at approximately 1.3 ppm.

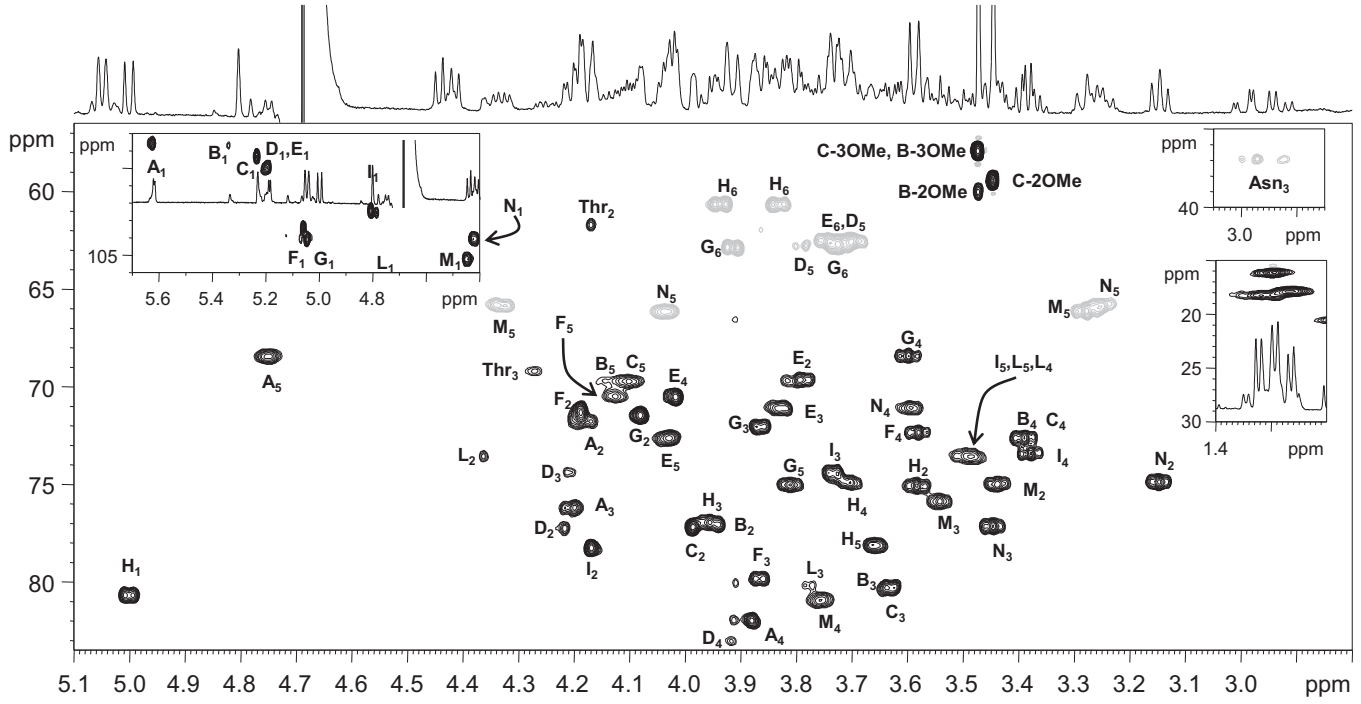

Fig. S2. Expansion of the ^{1}H-^{13}C heterocorrelated HSQC spectrum together with the corresponding ^{1}H spectrum measured for fraction 4, isolated from Vp54 protein after proteinase K digestion and Bio-Gel P10 purification. Correlations attribution follows the letter system of Table S1 and Fig. 1; those appearing in gray have the opposite sign with respect to the other and represent carbons bearing two hydrogen atoms.

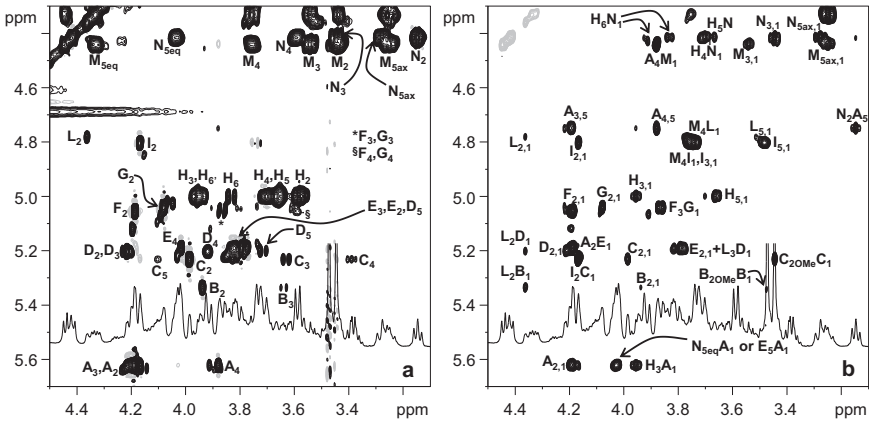

Fig. S3. Expansion of the (*A*) TOCSY and (*B*) TROESY spectra measured for fraction 4. Correlations' attribution follows the letter system of Table S1 and Fig. 1.

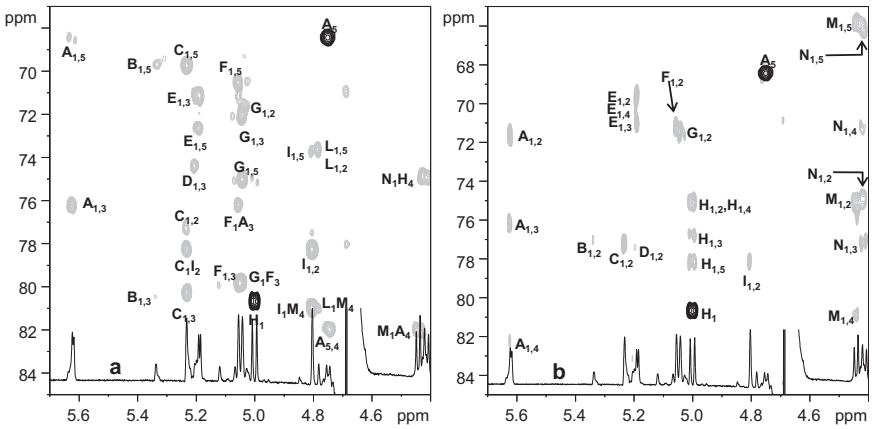

Fig. S4. Expansion of the HSQC spectrum (black) measured for fraction 4 at 310 K, superimposed with (*A*) the HMBC spectrum (gray) or (*B*) the HSQC-TOCSY spectrum (gray). Correlations attribution follows the letter system of Table S1 and Fig. 1.

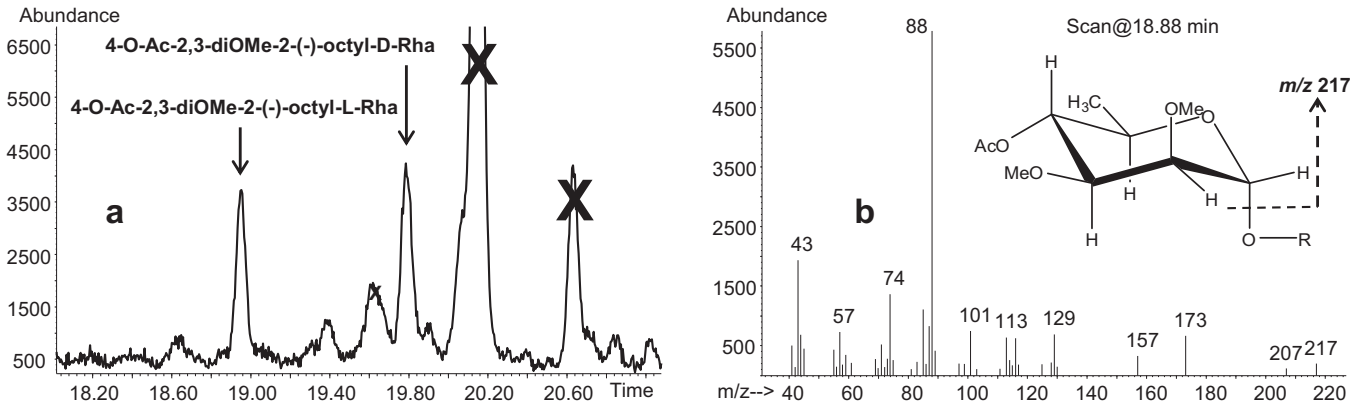

Fig. S5. (*A*) Expansion of the GC-MS chromatogram of the 4-Ac-2,3-diMe-2-(±)-octyl-L-rhamnoside (structure, *B*, *Inset*). (*B*) EI-MS spectrum of the methylated and acetylated rhamnose derivative (*A*, *Inset*). R, 2-octanol.

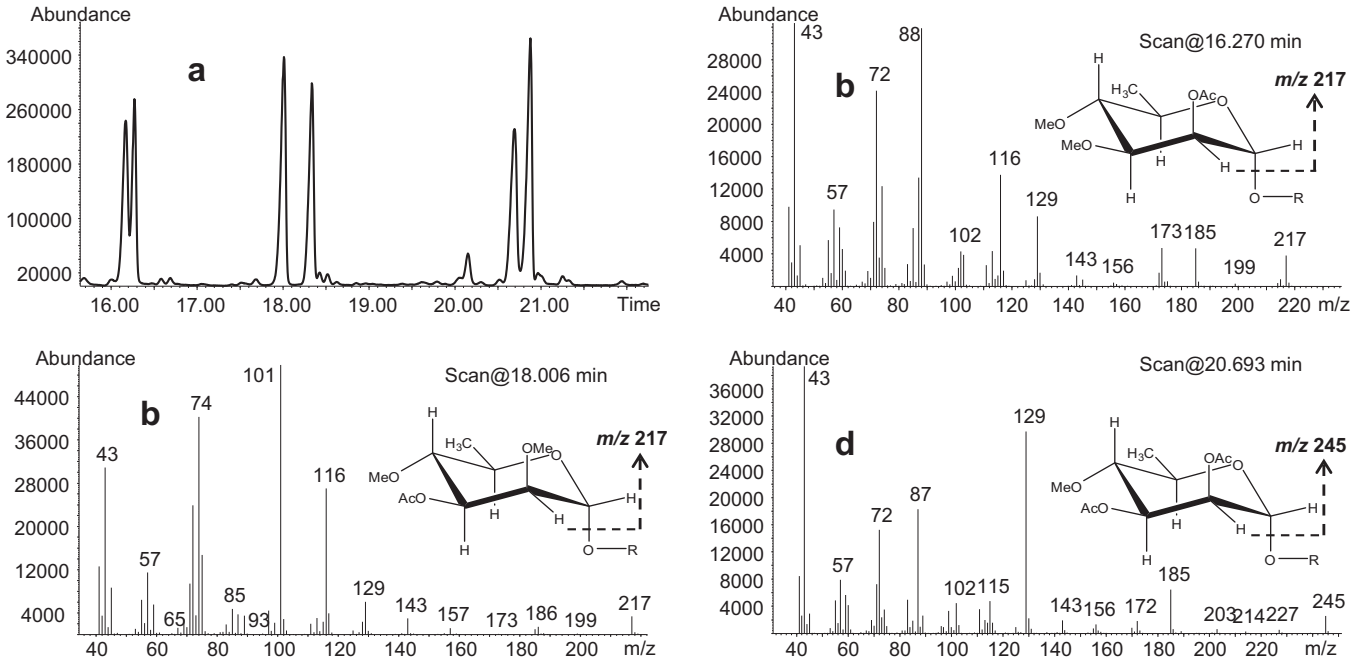

Fig. S6. (*A*) Expansion of the GC-MS chromatogram of the 2-linked, 3-linked, and 2,3-linked rhamnose derivatives after methylation, octanolysis, and acetylation. EI-MS spectra of (*B*) 2-OAc-3,4-diOMe-2-(±)-octyl-L-rhamnoside, (*C*) 3-OAc-2,4-diOMe-2-(±)-octyl-L-rhamnoside, and (*D*) 2,3-OAc-4-OMe-2-(±)-octyl-L-rhamnoside. (*B–D*) (*Insets*) Structures. R, 2-(±)-octanol.

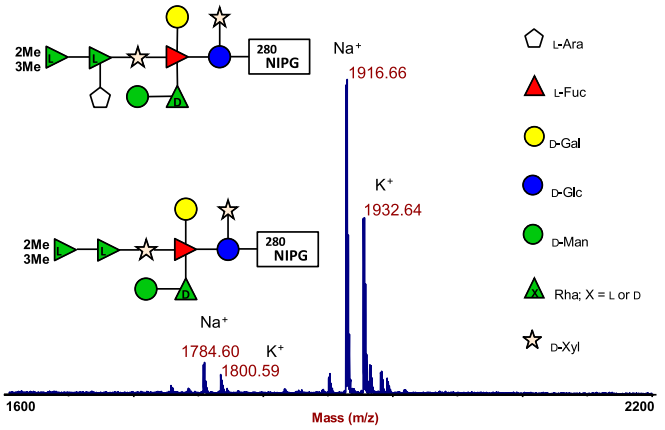

Fig. S7. MALDI MS spectrum of fraction 2 showing the ^{280}NIPG glycopeptide ions of glycoform **2** as the predominant species.

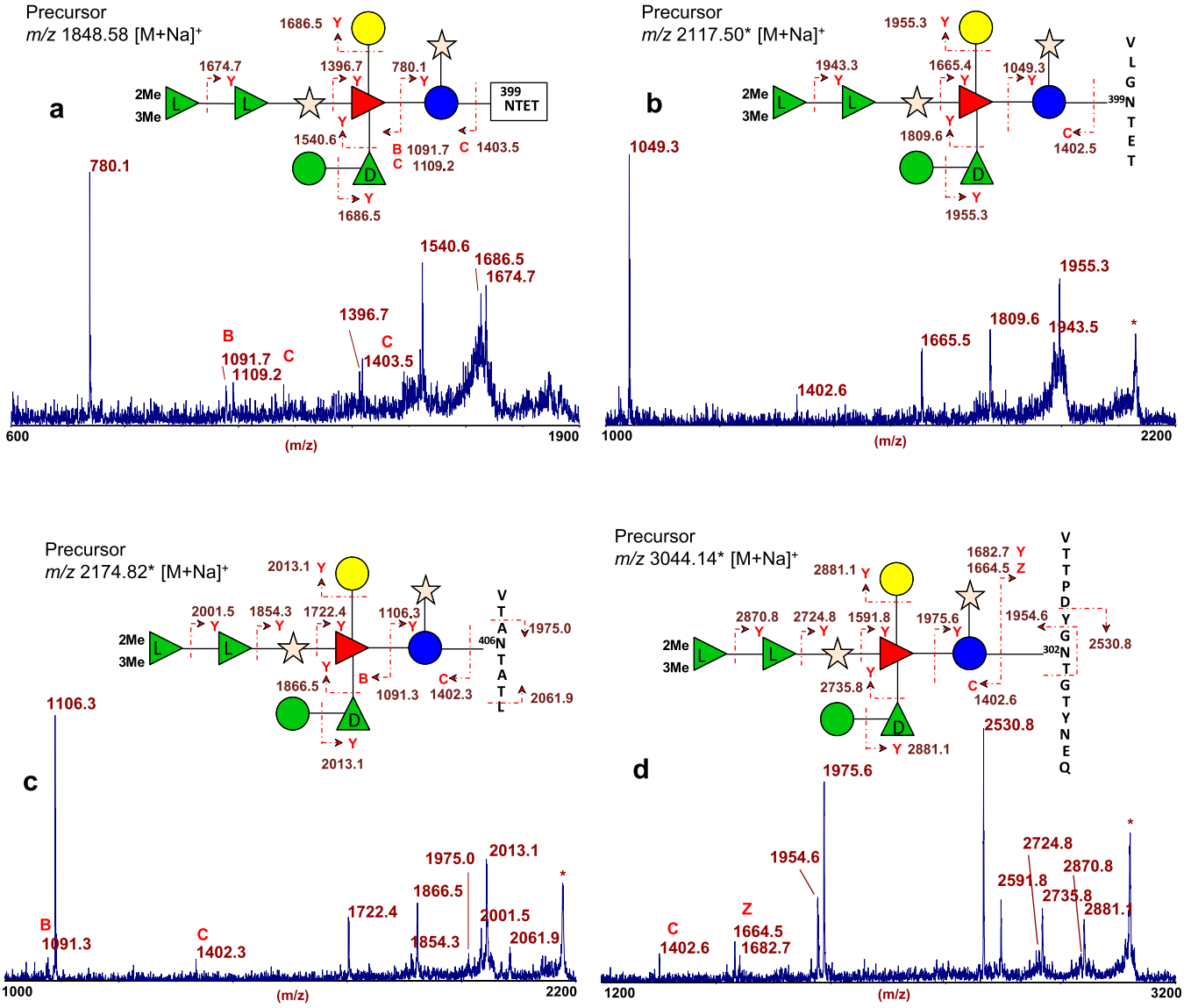

Fig. S8. MALDI MS/MS spectra of (*A*) glycopeptide precursor ions at *m/z* 1,848.58 found in fraction 1 and composed of the ^{399}NTET sequence linked to glycoform **1**; (*B*) glycopeptide precursor ions at *m/z* 2,117.50, (*C*) at 2,174.82, and (*D*) at 3,044.14. Glycopeptides in *B*–*D* were obtained by thermolysin digestion; they share the same glycan moiety (glycoforms **1**) and are linked to ^{399}Asn, ^{406}Asn, and ^{302}Asn, respectively.

Table S1. NMR (^{1}H 600 MHz, ^{13}C 150 MHz) chemical shift registered in deuterated water at 310 K for fraction 4

Residue/nucleus	1	2	3	4	5	6
A: 2,3,4-α-L-Fuc						
^{1}H	5.62	4.18	4.21	3.88	4.75	1.30
^{13}C	98.6	71.7	76.2	81.9	68.3	16.2
B: α-L-Rha-2,3-di-OMe						
^{1}H	5.34	3.94	3.64	3.40	4.14	1.26
^{13}C	98.7	77.0	80.3	72.6	69.7	17.8
C: α-L-Rha-2,3-di-OMe						
^{1}H	5.23	3.99	3.63	3.39	4.10	1.26
^{13}C	99.3	77.2	80.3	72.6	69.7	17.8
D: β-L-Ara*f*						
^{1}H	5.20	4.22	4.21	3.92	3.71; 3.79	—
^{13}C	99.9	77.2	74.3	83.0	62.8	—
E: α-D-Gal						
^{1}H	5.19	3.78	3.82	4.02	4.03	3.73; 3.69
^{13}C	99.9	69.6	71.1	70.5	72.6	62.6
F: 3-α-D-Rha						
^{1}H	5.06	4.19	3.86	3.58	4.12	1.29
^{13}C	103.5	71.3	79.8	72.4	70.4	18.2
G: α-D-Man						
^{1}H	5.04	4.08	3.84	3.60	3.81	3.91; 3.72
^{13}C	104.1	71.4	72.0	68.4	75.0	62.8
H: 3,4-β-D-Glc						
^{1}H	5.00	3.58	3.96	3.70	3.66	3.83; 3.93
^{13}C	80.5	75.0	76.7	74.9	78.1	60.6
I: 2-β-L-Rha						
^{1}H	4.80	4.17	3.73	3.38	3.48	1.32
^{13}C	102.4	78.2	74.4	73.4	73.5	18.2
L: 2,3-β-L-Rha						
^{1}H	4.78	4.36	3.78	3.51	3.52	1.35
^{13}C	102.5	73.6	80.2	73.4	73.4	18.2
M: 4-β-D-Xyl						
^{1}H	4.44	3.44	3.54	3.76	4.33; 3.25	—
^{13}C	105.1	75.0	75.8	80.9	65.8	—
N: β-D-Xyl						
^{1}H	4.42	3.15	3.45	3.59	4.03; 3.28	—
^{13}C	104.0	74.8	77.1	71.0	66.1	—
Asn						
^{1}H	—	4.01	3.00; 2.93	—	—	—
^{13}C	ND	52.4	36.8	174.4	—	—
Thr						
^{1}H	—	4.17	4.27	1.20	—	—
^{13}C	177.8	61.2	69.1	20.5	—	—

Where not explicitly mentioned, all residues are in the pyranose form and the number(s) preceding the name indicates the position glycosylated; B and C are substituted at both O-2 and O-3 with a methyl group (chemical shifts ^{1}H/^{13}C: B-2OMe: 3.48/60.0; B-3OMe: 3.48/57.9; C-2OMe: 3.44/59.4; C-3OMe: 3.48/57.9). Spectra are calibrated on internal acetone (^{1}H: 2.225 ppm, ^{13}C: 31.45 ppm). ND, not detected.