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.

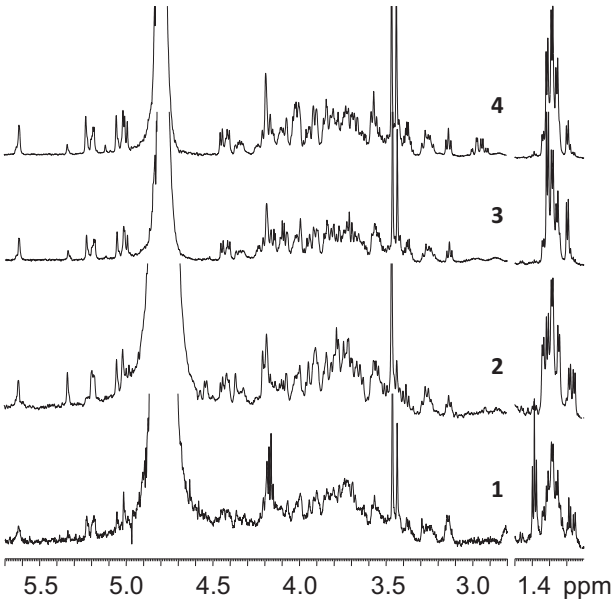

Fig. S1. Expansion of the proton spectra recorded for fractions 1 to 4, isolated from Vp54 protein after proteinase K digestion and Bio-Gel P10 purification. Anomeric region: 5.7 to 4.4 ppm; carbinolic region, 4.4 to 3.0 ppm; O-Me signals at 3.47 and 3.45 ppm; methyl signals of 6-deoxy residues at approximately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

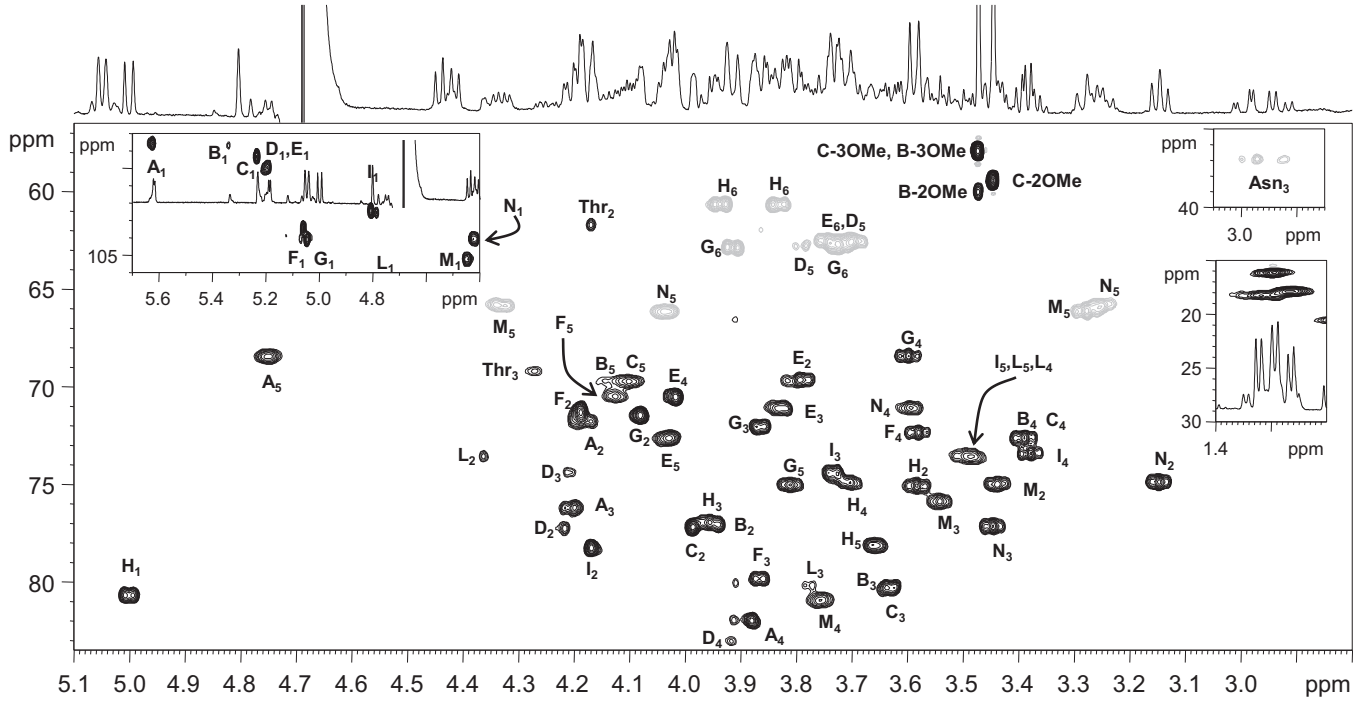

Fig. S2. Expansion of the ^{1}H-^{13}C heterocorrelated HSQC spectrum together with the corresponding ^{1}H spectrum measured for fraction 4, isolated from Vp54 protein after proteinase K digestion and Bio-Gel P10 purification. Correlations attribution follows the letter system of Table S1 and Fig. 1;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

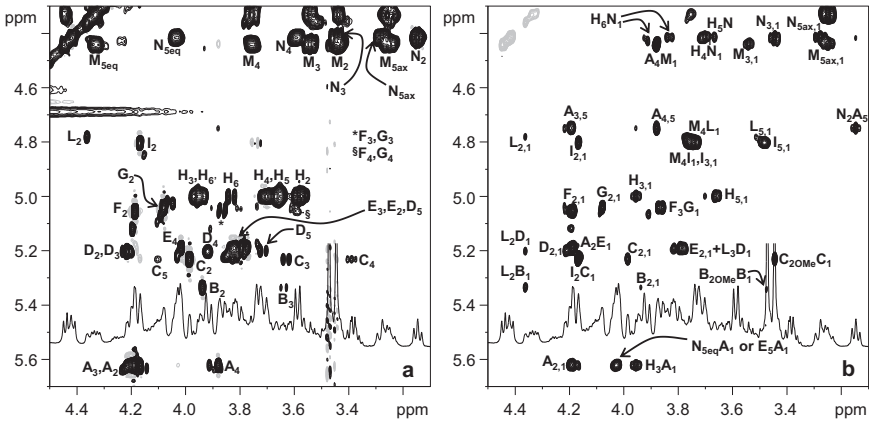

Fig. S3. Expansion of the (*A*) TOCSY and (*B*) TROESY spectra measured for fraction 4. Correlations' attribution follows the letter system of Table S1 and Fig. 1.

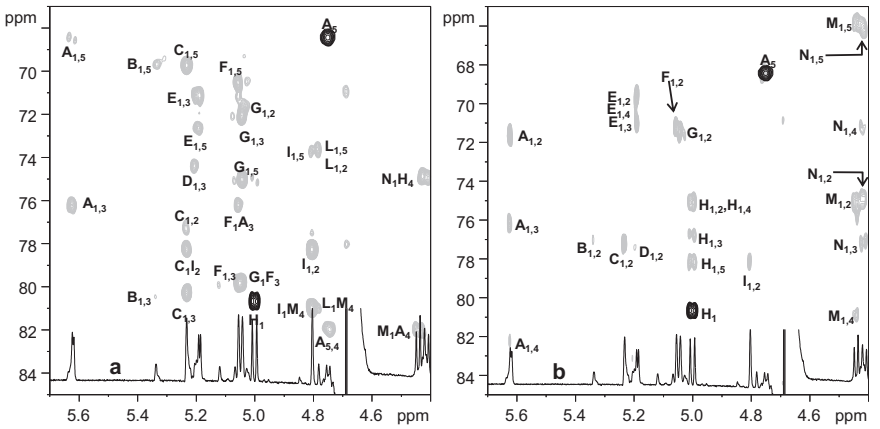

Fig. S4. Expansion of the HSQC spectrum (black) measured for fraction 4 at 310 K, superimposed with (*A*) the HMBC spectrum (gray) or (*B*) the HSQC-TOCSY spectrum (gray). Correlations attribution follows the letter system of Table S1 and Fig. 1.

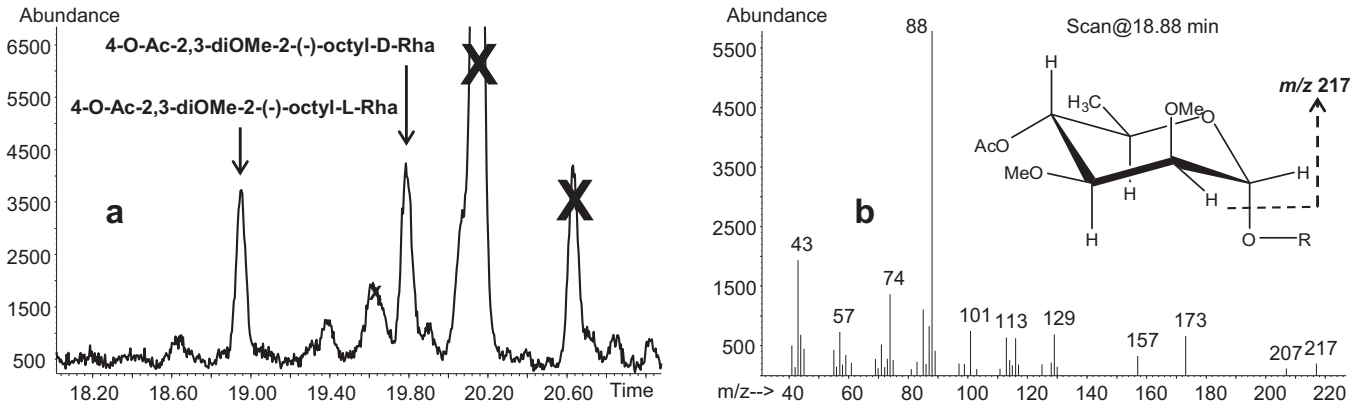

Fig. S5. (*A*) Expansion of the GC-MS chromatogram of the 4-Ac-2,3-diMe-2-(±)-octyl-L-rhamnoside (structure, *B*, *Inset*). (*B*) EI-MS spectrum of the methylated and acetylated rhamnose derivative (*A*, *Inset*). R, 2-octanol.

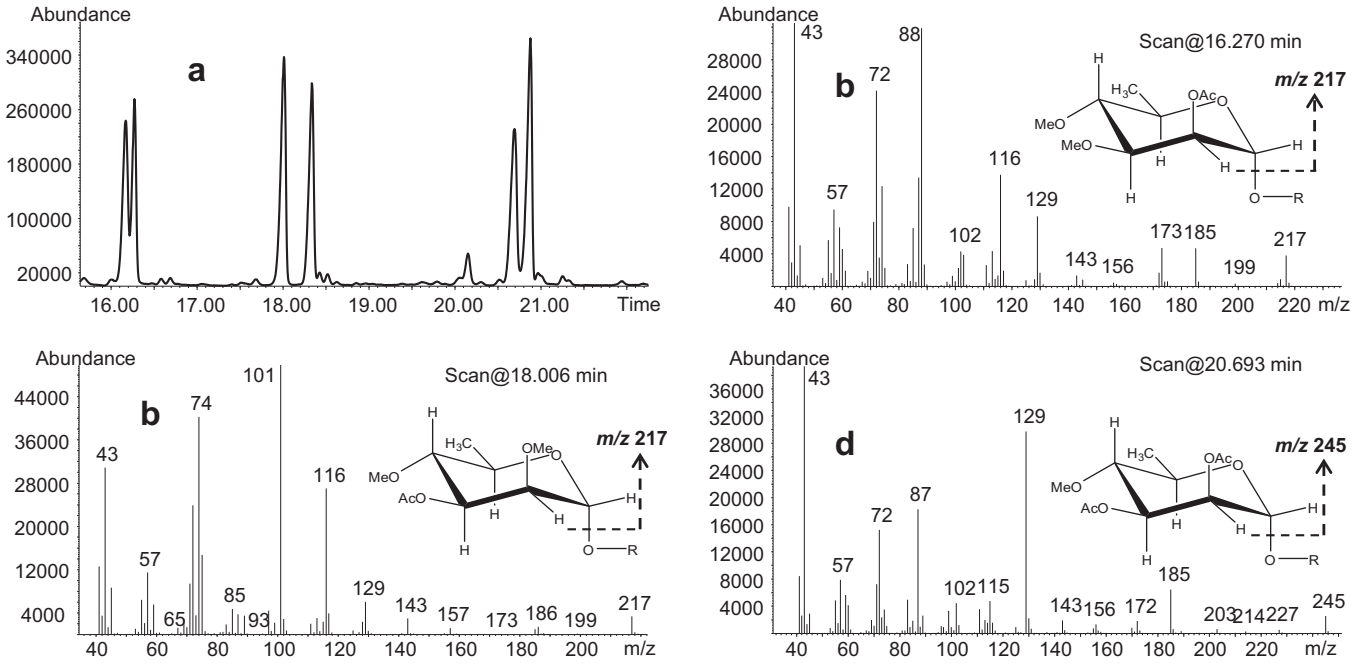

Fig. S6. (*A*) Expansion of the GC-MS chromatogram of the 2-linked, 3-linked, and 2,3-linked rhamnose derivatives after methylation, octanolysis, and acetylation. EI-MS spectra of (*B*) 2-OAc-3,4-diOMe-2-(±)-octyl-L-rhamnoside, (*C*) 3-OAc-2,4-diOMe-2-(±)-octyl-L-rhamnoside, and (*D*) 2,3-OAc-4-OMe-2-(±)-octyl-L-rhamnoside.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

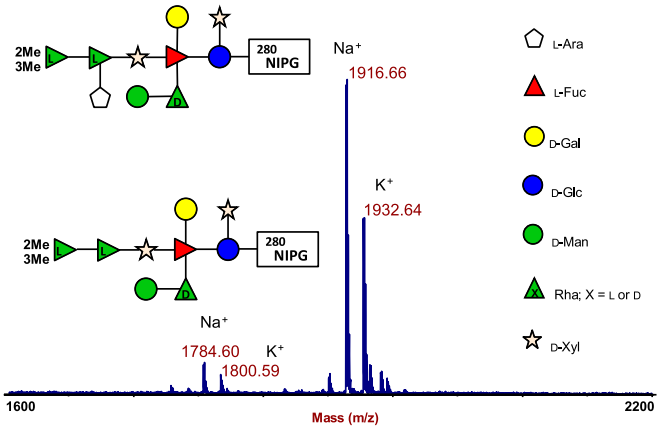

Fig. S7. MALDI MS spectrum of fraction 2 showing the ^{280}NIPG glycopeptide ions of glycoform **2** as the predominant species.

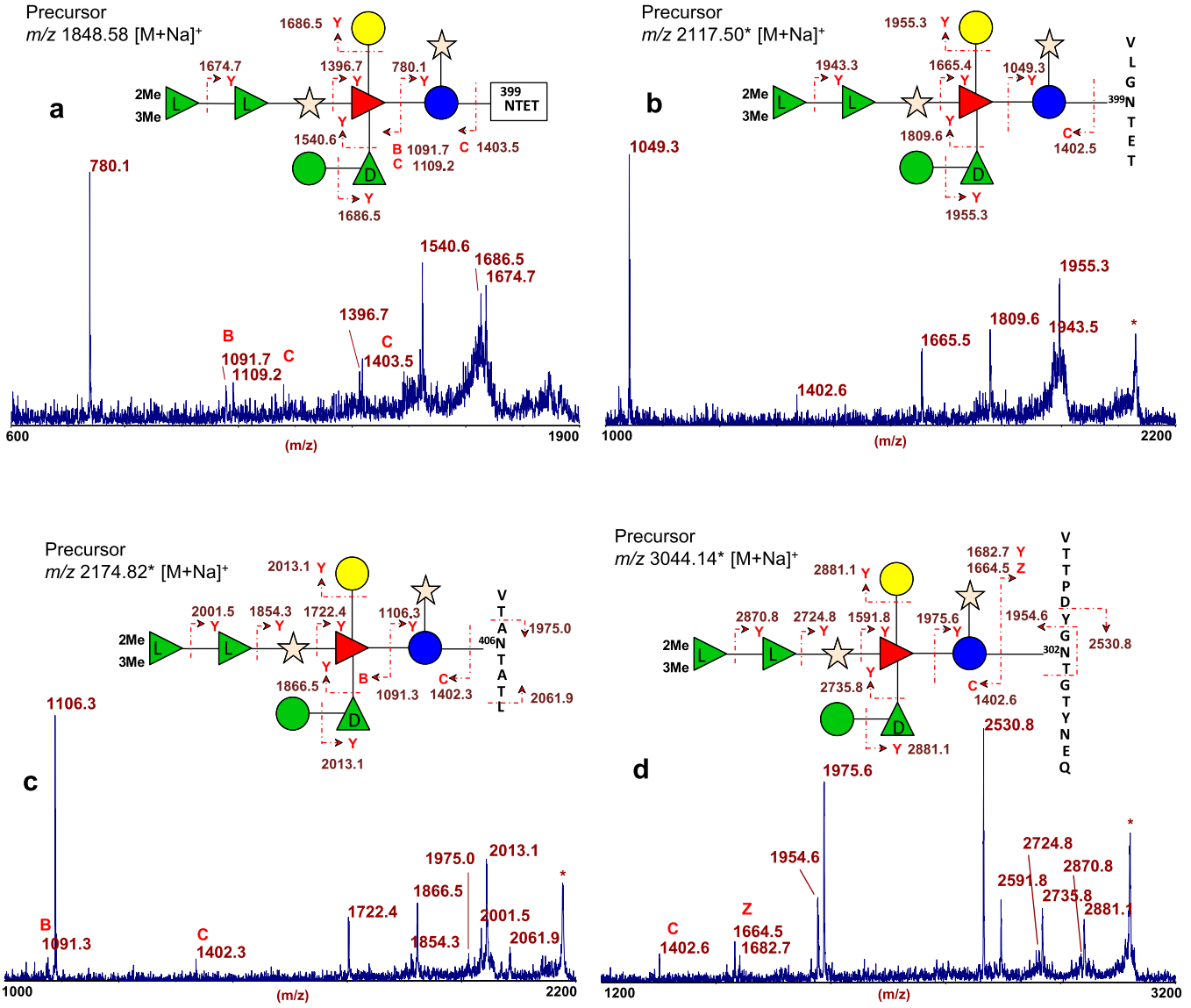

Fig. S8. MALDI MS/MS spectra of (*A*) glycopeptide precursor ions at *m/z* 1,848.58 found in fraction 1 and composed of the ^{399}NTET sequence linked to glycoform **1**; (*B*) glycopeptide precursor ions at *m/z* 2,117.50, (*C*) at 2,174.82, and (*D*) at 3,044.14. Glycopeptides in *B*–*D* were obtained 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.