

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

James Van Etten Publications

Plant Pathology Department

---

1-2016

## Chloroviruses *N*-linked glycans share a new type of conserved core architecture unprecedented in any form of life / [Published as] *N*-Linked Glycans of Chloroviruses Sharing a Core Architecture without Precedent

Cristina De Castro

*University of Napoli*, [decastro@unina.it](mailto:decastro@unina.it)

Immacolata Speciale

*University of Napoli*

Garry Duncan

*Nebraska Wesleyan University*, [gduncan@nebrwesleyan.edu](mailto:gduncan@nebrwesleyan.edu)

David Dunigan

*University of Nebraska - Lincoln*, [ddunigan2@unl.edu](mailto:ddunigan2@unl.edu)

Irina Agarkova

*University of Nebraska-Lincoln*, [iagarkova2@unl.edu](mailto:iagarkova2@unl.edu)

Follow this and additional works at: <https://digitalcommons.unl.edu/vanetten>

See next page for additional authors



Part of the [Genetics and Genomics Commons](#), [Plant Pathology Commons](#), and the [Viruses Commons](#)

---

De Castro, Cristina; Speciale, Immacolata; Duncan, Garry; Dunigan, David; Agarkova, Irina; Lanzetta, Rosa; Sturiale, Luisa; Palmigiano, Angelo; Garozzo, Domenico; Molinaro, Antonio; Tonetti, Michela; and Van Etten, James L., "Chloroviruses *N*-linked glycans share a new type of conserved core architecture unprecedented in any form of life / [Published as] *N*-Linked Glycans of Chloroviruses Sharing a Core Architecture without Precedent" (2016). *James Van Etten Publications*. 25.

<https://digitalcommons.unl.edu/vanetten/25>

This Article is brought to you for free and open access by the Plant Pathology Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in James Van Etten Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

---

**Authors**

Cristina De Castro, Immacolata Speciale, Garry Duncan, David Dunigan, Irina Agarkova, Rosa Lanzetta, Luisa Sturiale, Angelo Palmigiano, Domenico Garozzo, Antonio Molinaro, Michela Tonetti, and James L. Van Etten



Published in final edited form as:

*Angew Chem Int Ed Engl.* 2016 January 11; 55(2): 654–658. doi:10.1002/anie.201509150.

## Chloroviruses N-linked glycans share a new type of conserved core architecture unprecedented in any form of life

Prof. Cristina De Castro<sup>a</sup>, Dr. Immacolata Speciale<sup>b</sup>, Prof. Garry Duncan<sup>c</sup>, Dr. David D. Dunigan<sup>d</sup>, Dr. Irina Agarkova<sup>d</sup>, Prof. Rosa Lanzetta<sup>b</sup>, Dr Luisa Sturiale<sup>e</sup>, Dr. Angelo Palmigiano<sup>e</sup>, Prof. Domenico Garozzo<sup>e</sup>, Prof. Antonio Molinaro<sup>b</sup>, Prof. Michela Tonetti<sup>f</sup>, and Prof. James L. Van Etten<sup>d</sup>

Cristina De Castro: decastro@unina.it

<sup>a</sup>Department of Agricultural Sciences, University of Napoli Via Università 100, 80055 Portici (NA), Italy

<sup>b</sup>Department of Chemical Sciences, University of Napoli Via Cintia 4, 80126 Napoli, Italy

<sup>c</sup>Department of Biology, Nebraska Wesleyan University, Lincoln, NE, USA

<sup>d</sup>Department of Plant Pathology and Nebraska Center for Virology University of Nebraska Lincoln, NE 68583-0900 USA

<sup>e</sup>CNR, Institute for Polymers, Composites and Biomaterials Via P. Gaifami 18, 95126 Catania, Italy

<sup>f</sup>Department of Experimental Medicine and Center of Excellence for Biomedical Research, University of Genova, Viale Benedetto XV/1, 16132 Genova, Italy

### Abstract

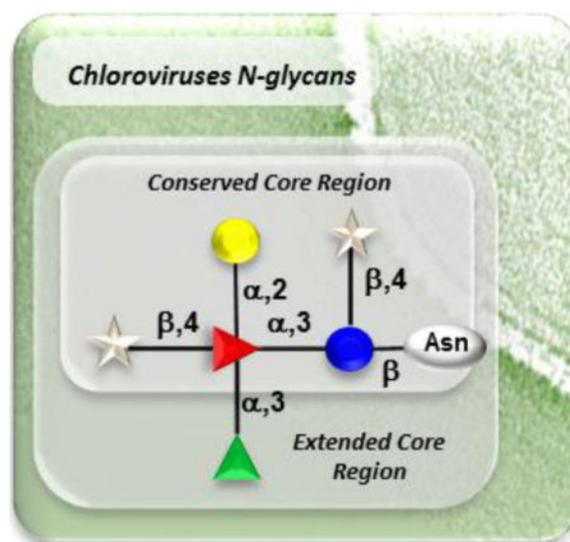
N-glycosylation is a fundamental modification of proteins that exists in the three domains of life and in some viruses, including the chloroviruses, for which a new type of core N-glycan is described. This N-glycan core structure common to all chloroviruses is a pentasaccharide with a  $\beta$ -glucose linked to an asparagine residue that is not located in the typical sequon N-X-T/S. The glucose is linked to a terminal xylose unit and a hyperbranched fucose, in turn substituted with a terminal galactose and a second xylose residue. The third position of the fucose unit is always linked to a rhamnose, which is a semi-conserved element because its absolute configuration is virus-dependent. Additional decorations occur on this core N-glycan and represent a molecular signature for each chlorovirus.

### Graphical abstract

---

Correspondence to: Cristina De Castro, decastro@unina.it.

Supporting information for this article is available on the WWW under <http://>



## Keywords

Glycoproteins; Chloroviruses; *N*-glycans; Natural Products; Structural Biology

Protein *N*-glycosylation is a fundamental post-translational modification that occurs in all domains of life, Eukarya, Archaea, and Bacteria,<sup>[1]</sup> and also in many viruses, such as rhabdoviruses, herpesviruses, poxviruses, and paramyxoviruses.<sup>[2]</sup> These viruses infect eukaryotic organisms and use the host-encoded machinery to add and remove sugars from virus glycoproteins. Consequently, *N*-acetylglucosamine is the sugar directly linked to the Asn residue in these viral glycoproteins and the glycan portion resembles host glycans. Therefore, the only way to alter glycosylation of virus proteins is to either grow the virus in a different host<sup>[3]</sup> or have a mutation that changes the viral protein such that it alters glycosylation.

*Chloroviruses* (family *Phycodnaviridae*) infect eukaryotic algae but differ from this paradigm because they encode most, if not all, of the machinery to glycosylate their major capsid proteins (MCP).<sup>[4]</sup> This genus includes large (190 nm in diameter) icosahedral, plaque-forming viruses with an internal lipid membrane and dsDNA genomes encoding as many as 400 proteins.<sup>[5]</sup> We have recently shown that Vp54, the MCP of *Paramecium bursaria* chlorella virus (PBCV-1) contains four *N*-linked glycans,<sup>[6]</sup> with structures that do not resemble any other reported for Bacteria, Archaea, and Eukarya,<sup>[7]</sup> including unicellular algae.<sup>[8]</sup>

The PBCV-1 Vp54 glycans are attached to the protein by an Asn  $\beta$ -glucose linkage; the Asn is not located in the typical N-X-T/S consensus sequon. The glycans consist of eight to ten neutral monosaccharide residues, for a total of four glycoforms (Figure 1).<sup>[7]</sup> These *N*-glycans have a highly branched architecture with two residues, arabinose and mannose, as non-stoichiometric substituents, along with a hyperbranched fucose, and a dimethylated rhamnose as a capping residue. Prompted by the peculiarity of these motifs, we examined the *N*-glycans of additional chloroviruses, to determine if those described for PBCV-1 were

unique or if their structural features were conserved in the other chloroviruses and represent an unprecedented *N*-glycosylation pattern.

Five chloroviruses with different host specificities were chosen: NY-2A, like PBCV-1, infects *Chlorella variabilis* NC64A (NC64A viruses); ATCV-1 and TN603 infect *Chlorella heliozoae* (SAG viruses) and MT325 and CVM-1 infect *Micractinium conductrix* Pbi (Pbi viruses). The MCP amino acid sequence from PBCV-1 was used to identify the MCPs from the five additional chloroviruses (Figure S1). All of these viruses encode one MCP with a calculated molecular weight lower than that determined experimentally by ESI-MS (Table 1) or SDS-PAGE (Figure 2).

The ESI-MS data indicated a MW increase 5.9–12.5% for all the MCPs, except for MT325, for which clean MS data were not obtained, and the SDS-PAGE positive staining for glycoproteins (Figure 2A) suggested that these MCPs were glycosylated.

Virus NY-2A was an exception: it encodes two closely related (or paralog) MCPs, B585L and B617L (Table S1), both glycosylated as deduced by SDS-PAGE (Figure 2A) and ESI-MS data (Table 1). These two MCPs could not be separated and further analysis was performed on the combined pair.

Glycopeptide isolation from each MCP was accomplished as reported (Supporting Information for details).<sup>[7]</sup> Each virus gave one glycopeptide (<sup>1</sup>H NMR in Figure 3), except NY-2A for which two different fractions were isolated, NY-2A<sub>1</sub> and NY-2A<sub>2</sub>.

Monosaccharide chemical analysis (Table S2)<sup>[7]</sup> identified all the residues (Figure S2), except for xylulose, a labile residue that degrades during the methanolysis reaction. The structure of each glycopeptide was obtained by combined analysis of NMR and MALDI data (see Supporting Information). MALDI MS spectrometry studies supported NMR results and determined the glycosylation sites for CVM-1 and NY-2A.

Virus CVM-1 had the simplest glycan among all of this study (Figure 1), consisting of only six monosaccharides as noted in the HSQC spectrum (Figure S3). Unit **D** (5.01/80.7 ppm) had the anomeric carbon *N*-linked to the protein as disclosed by its carbon chemical shift (80.7 ppm), and it was a glucose (Figure S4–S5, Table S3) substituted as in PBCV-1 with an  $\alpha$ -fucose (**A**) at *O*-3, and a  $\beta$ -xylose **F** at *O*-4. This last xylose is named proximal xylose because closer than the other xylose to the peptide backbone. Fucose (**A**) was fully substituted and it had  $\alpha$ -galactose **B** at *O*-2,  $\alpha$ -rhamnose **C** at *O*-3, and finally the distal  $\beta$ -xylose **E** at *O*-4.

Compared to the PBCV-1 MCP, CVM-1 had some similarities with a few differences (Figure 1): it was not elongated at the distal xylose, and rhamnose **C** was L instead of D, in addition to being acetylated at *O*-2 and methylated at *O*-3. Consistent with the different absolute configuration of **C**, carbon chemical shifts (Table S4) of **A** of CVM-1 diverged from those of the equivalent residue of PBCV-1,<sup>[7]</sup> and such variation fitted with the rules reported by Shashkov et al.<sup>[9]</sup> MALDI MS confirmed CVM-1 *N*-glycan structure and determined the sites of attachment (Figure S6, Table 1): Asn-47, -279, -285, and -293. With the exception of Asn-47, all positions were equivalent to those found glycosylated in PBCV-1 (Figure S1).

Asn-47 is the only one found in a typical sequon (NGS, Table 1) and although it is conserved in all MCPs, its glycosylation only occurs in CMV-1.

The ATCV-1 *N*-glycan (Figure 1, NMR spectra in Figures 3, S7–S8, NMR data in Table S4) had several similarities with that of CVM-1; the main differences occurred in  $\alpha$ -rhamnose **C** and the two xylose units. The anomeric signal of **C** was slightly different because this unit was methylated at *O*-3, but not acetylated at *O*-2, while inspection of the xylose anomeric region (Figure S8C) indicated a high degree of complexity. Such complexity originated from a methyl group which occurred as not stoichiometric substituent at *O*-4 of both the xylose units, generating four different glycopeptides (Figure 1) as confirmed by MALDI spectrometry (Figure S9).

TN603 glycopeptides (NMR spectra in Figures 3, S10–11, NMR data in Table S5) shared with ATCV-1 the complex pattern of xylose residues, which arose from the non-stoichiometric presence of two different types of residues: a methyl group and a xylulose residue. These two appendages were always at *O*-4 of the xylose residues, with the methyl group occurring only at the distal unit while xylulose was at both xylose units. In addition, xylulose was also methylated in a non-stoichiometric fashion at *O*-1, increasing the overall variability of this *N*-glycan as depicted in Figure 1. The great variation in structures observed by NMR was confirmed by MALDI MS analysis (Figure S12 and detailed description is given in Supporting Information).

The NMR spectra of MT325 (Figures 3, S13–14, data in Table S6) resembled those of TN603: divergences related with the non-stoichiometric groups at *O*-4 of the two xylose residues. The methyl group was only at the proximal xylose, while xylulose was at both xylose residues, with that at the distal xylose further decorated with a methyl at *O*-3.

As a result, several complex glycoforms were identified for MT325 glycopeptides and supported by MALDI spectrometry (Figure S15); the most relevant structures are reported in Figure 1.

NY-2A MCPs produced two glycopeptide fractions: interestingly, in NY-2A<sub>1</sub> (NMR spectra in Figures 3, S16–S17) the fucose **A** carbon chemical shift values (Table S7) were similar to those found for the equivalent residue in PBCV-1. This agreement resulted from the nature of **C**, a  $\alpha$ -rhamnose at *O*-3 of **A**, having the **D** absolute configuration as in PBCV-1 and not **L** as for the other viruses. **C** was substituted at both *O*-3 and *O*-4 by  $\alpha$ -L-3OMeRha (**I**) and  $\alpha$ -D-Gal (**L**), in turn further substituted at *O*-2 with a  $\alpha$ -D-3OMeXyl residue (**H**). Taken together, this information defined NY-2A<sub>1</sub> glycopeptide as a highly branched nonasaccharide (Figure 1).

NY-2A<sub>2</sub> NMR spectra (Table S8, NMR in Figures 3, S18–S20) contained all the signals assigned in NY-2A<sub>1</sub> along with a new anomeric signal at 5.04 ppm, labelled **C'**. This new residue was a non-substituted  $\alpha$ -rhamnose linked at *O*-3 of the hyperbranched fucose. Thus, NY-2A<sub>2</sub> was a mixture of two different glycopeptides, a nonasaccharide and its truncated form (Figure 1); MALDI spectrometry confirmed this information (Figure S21) and also identified the glycosylation sites of the MCPs, which occurred at the same positions

(Asn-280, -302, -399 and -406) as PBCV-1, along with two additional ones, Asn-54 and -291 (Table 1, Figure S21).

Glycosylation of NY-2A MCPs is interesting for several reasons. First, Asn-54 is in a region conserved in all MCPs but not glycosylated in PBCV-1 and CVM-1, while Asn-291 is not conserved and only exists in the NY-2A B617L paralog. Second, glycosylation is site-specific: nonasaccharide occurs at Asn-302, -399 and -406, while the hexasaccharide at Asn-54, -280 and -291 (Table 1). Finally, the presence of five (B585L) or six (B617L) glycosylation sites does not fit with the average number of glycans calculated for these MCPs (Table 1), suggesting that not all of the possible sites are always occupied in each paralog, an issue for future studies.

Hence, solving the *N*-glycan structures of five additional chloroviruses addresses the question that prompted this study: this is an unprecedented *N*-glycosylation pattern. All the MCPs from these chloroviruses contain *N*-linked glycans (Figure 1) that have a conserved central core structure new and unique (Figure 4). This motif starts with a *N*-linked glucose and contains a hyperbranched fucose, a distal and a proximal xylose, and a galactose.

This basic core motif is extended with a semi-conserved element: a rhamnose is always at *O*-3 of fucose. However, this unit has the D-configuration in NC64A viruses, while it is L and methylated at *O*-3 in SAG and Pbi viruses (Figure 4). Therefore, its inclusion in the basic core structure generates an extended conserved core motif that is group specific. This extended core structure is further decorated with other appendages and the overall oligosaccharide is the signature of each virus (Figure 1). Other relevant features are the *N*-glycans capping with one or more methyl groups, a modification common to many organisms except mammals,<sup>[11]</sup> while the glycosylation sites determined in this study for some of the MCPs confirmed that they do not occur in the typical sequon N-X-T/S, with the only exception of Asn-47 in CVM-1.

In conclusion, *N*-glycans from chloroviruses have a new structural motif that can be considered as a signature or as a distinctive trait for this class of organisms, opening the way to many intriguing questions. For example, how does biosynthesis occur and what are the genes/proteins involved in the process? How is the glucose-asparagine linkage assembled? How does the L,D rhamnosyl transferase work to equally transfer both L and D rhamnose nucleotides to the growing oligosaccharide?

But also, how much is this motif diffused in Nature and how many non eukaryotic-like *N*-glycosylation patterns were overlooked so far? Are the chloroviruses the only organisms possessing their own *N*-glycans?

Answers to these questions will increase the global knowledge in Glycoscience and will open new avenues into the understanding of such an important matter as the biosynthesis and function of glycoproteins in Nature.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

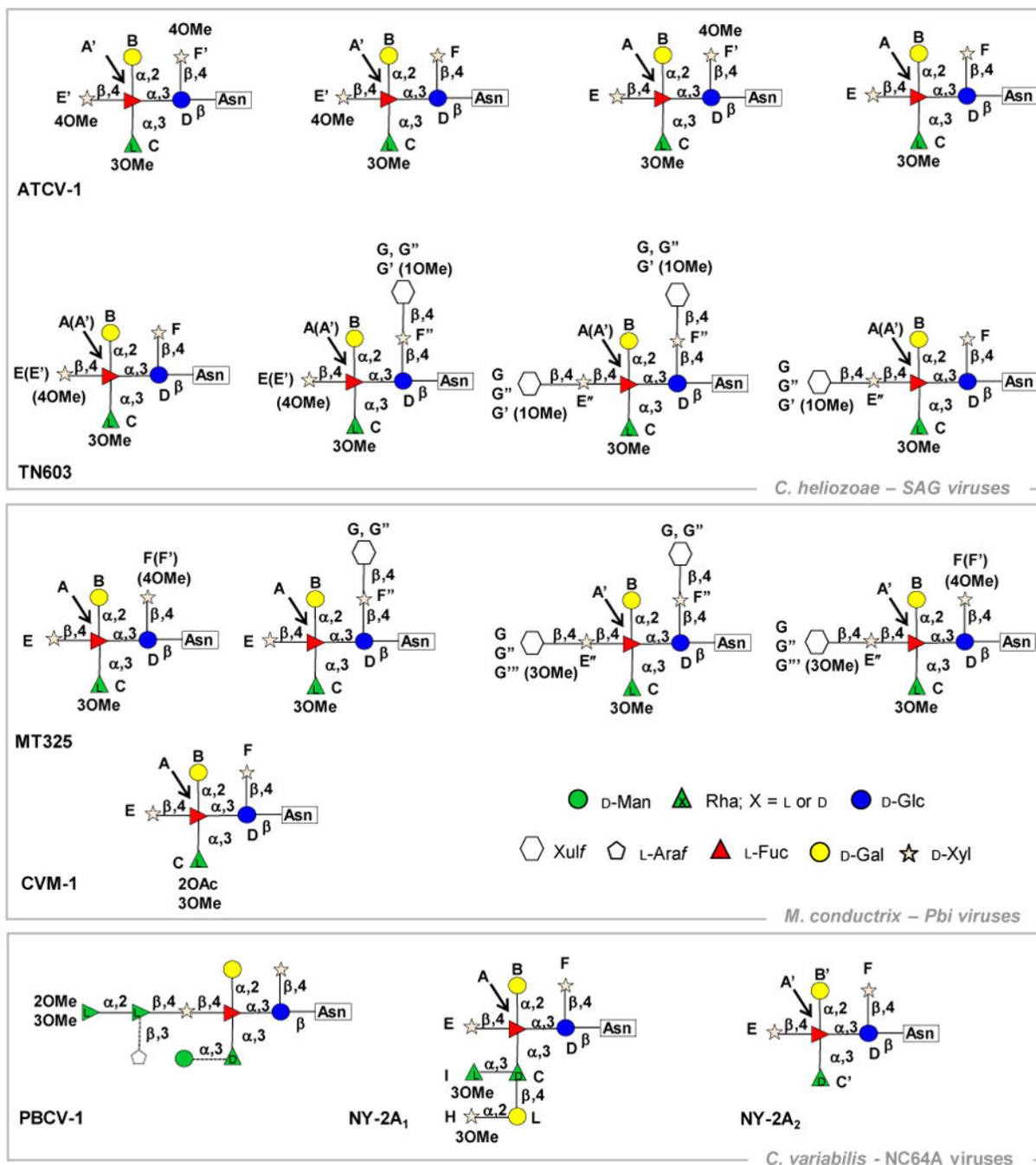
Work in the Van Etten laboratory was partially supported by National Science Foundation grant EPS-1004094, Stanley Medical Research Institute grant 11R-003, and the National Institute of Health grant P20-RR15635 from the Centers of Biomedical Research Excellence Program of the National Center for Research Resources. GD was supported by National Institute for General Medical Science grant P20GM103427.

Work in Garozzo laboratory was partially supported from Stellalucente Trust and from the Italian National Council of Research (CNR).

CDC and AM wish to acknowledge financial support from NMR IR-RMN-THC Fr3050 CNRS for the use of 950 MHz NMR spectrometer.

## References

1. Schwarz F, Aebi M. *Curr Opin Struct Biol.* 2011; 21:576–582. [PubMed: 21978957]
2. Flint, SJ.; Enquist, LW.; Racaniello, VR.; Skalka, AM. *Principles of Virology, Molecular Biology, Pathogenesis, and Control of Animal Viruses.* ASM Press; Washington DC: 2004.
3. Vigerust DJ, Shepherd VL. *Trends Microbiol.* 2007; 15:211–218. [PubMed: 17398101]
4. Van Etten JL, Gurnon J, Yanai-Balser G, Dunigan D, Graves MV. *Biochem Biophys Acta.* 2010; 1800:152–159. [PubMed: 19654039]
5. Van Etten JL, Dunigan DD. *Trends Plant Sci.* 2012; 17:1–8. [PubMed: 22100667]
6. Nandhagopal N, Simpson A, Gurnon JR, Yan X, Baker TS, Graves MV, Van Etten JL, Rossmann MG. *Proc Natl Acad Sci.* 2002; 99:14758–14763. [PubMed: 12411581]
7. De Castro C, Molinaro A, Piacente F, Gurnon JR, Sturiale L, Palmigiano A, Lanzetta R, Parrilli M, Garozzo D, Tonetti M, Van Etten LJ. *Proc Natl Acad Sci.* 2013; 110:13956–13960. [PubMed: 23918378]
8. Mathieu-Rivet E, Kiefer-Meyer MC, Vanier G, Ovide C, Burel C, Lerouge P, Bardor M. *Front Plant Sci.* 2014; 5:1–13.
9. Shashkov AS, Lipkind GMY, Knirel YA, Kochetkov NK. *Magn Res Chem.* 1988; 26:735–747.
10. Stanley, P.; Schchter, H.; Taniguchi, N. *Essential of Glycobiology.* 2nd. Varki, A.; Cummings, RD.; Esko, JD.; Freeze, HH.; Stanly, P.; Bertozzi, CR.; Hart, GW.; Etzler, ME., editors. Cold Spring Harbor, NY: 2009. p. 101-114.
11. Staudacher E. *Biol Chem.* 2012; 393:675–685. [PubMed: 22944672]

**Figure 1.**

Structures of *N*-glycans from six chloroviruses. The basic core structure is defined by residues **A–B**, **D–F** and was conserved between all the chloroviruses. Xylose **F** is named the “proximal xylose” because it is closer than xylose **E** to the peptide backbone; xylose **E** is named the “distal xylose”. Inclusion of **C** gives the extended conserved core structure. The *N*-glycan of the prototype chlorovirus PBCV-1 is given for comparison and monosaccharides connected by broken lines are not stoichiometric substituents. Structures from the same virus are arranged along a line and different viruses are grouped according to

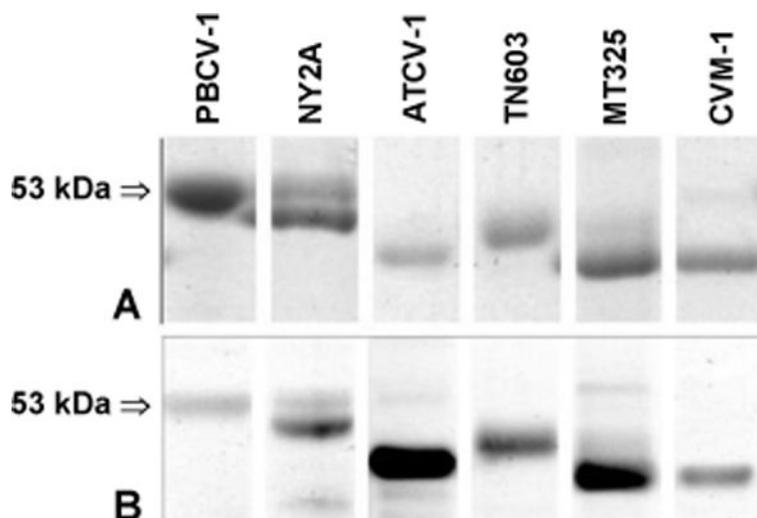
host specificity, indicated at the right-bottom margin of the grouping box. Letter labels are those used for NMR attribution and all sugars are in the pyranose form except where specified

Author Manuscript

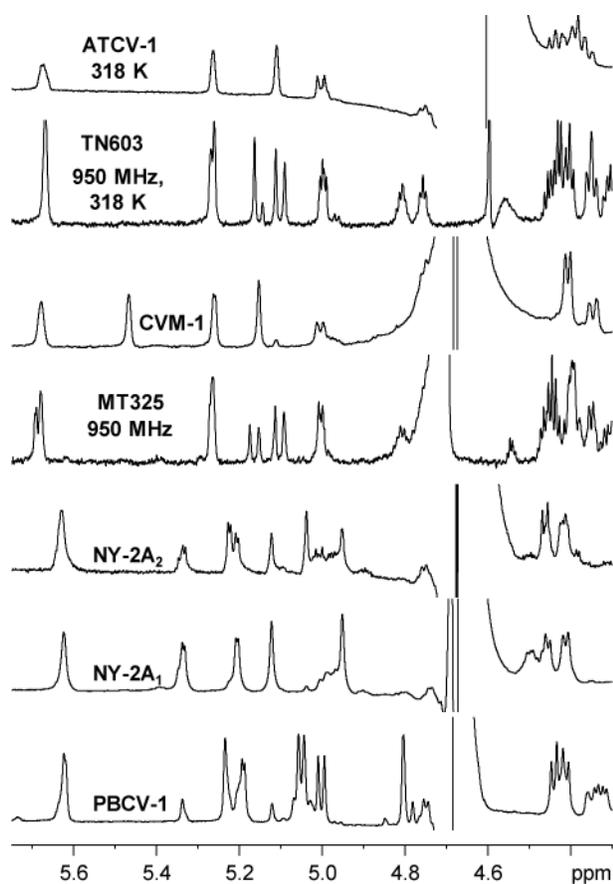
Author Manuscript

Author Manuscript

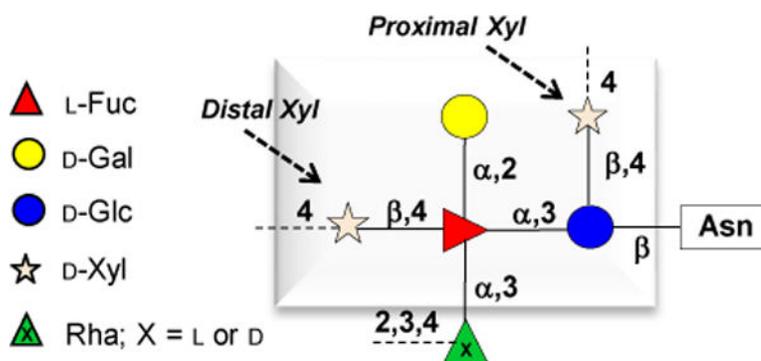
Author Manuscript



**Figure 2.** SDS-PAGE of total proteins of the MCPs from chloroviruses PBCV-1, NY-2A, ATCV-1, TN603, CVM-1 and MT325. A) Glycoprotein staining. B) Silver staining. Note that NY-2A has two bands.



**Figure 3.**  $^1\text{H}$  NMR spectra reporting the anomeric region of the glycopeptides obtained from the capsid protein from the different chloroviruses. Where not specified, spectra were recorded at 600 MHz and at 310 K. The spectrum of PBCV-1 is provided as a reference.



**Figure 4.** Oligosaccharide core structure common to the *N*-glycans of the MCPs from six different chloroviruses. The strictly conserved region is made by the five monosaccharides inside the box. The unit outside is always a rhamnose and its inclusion in the formula gives the extended form of the conserved core structure. The absolute configuration of this rhamnose is D for NC64A viruses (NY-2A and PBCV-1) otherwise, it is L and methylated at *O*-3. Broken lines indicate positions further decorated: the type and number of appendages is virus specific as detailed in Figure 1.

Table 1

Major capsid protein statistics for six chloroviruses: PBCV-1 (A430L), NY2A (B585L and B617L, two nearly identical paralogs), MT325 (M463L), CVM-1 (548L), TN603.4.2 (325L) and ATCV-1 (Z280L). MW% represent the net increase in percentage of molecular weight for each protein, calculated as:  $(MW - MW^o) \times 100 / MW^o$

Virus (gene code)	% AA identity	MW <sup>a</sup> [a]	MW	MW (%)	N-glycan MW[b]	Avg. no. glycans	Asn sequon found
PBCV-1 (A430L)	100	48168	53790[c]	5622 (11.7)	1384	4	<sup>280</sup> NIPG; <sup>302</sup> NTGT; <sup>399</sup> NTET; <sup>406</sup> NTAT
NY-2A (B585L)[d]	96	48359	52641, 51250	4282 (8.9) 2891 (5.9)	1348, 880	4 or 3[e]	<sup>54</sup> NKVS; [f] <sup>280</sup> NIPG; [g] <sup>302</sup> NTGT; [g] <sup>399</sup> NTET; [g] <sup>406</sup> NTAT [g]
NY-2A (B617L) [d]	94	48354	As above	As above	As above	As above	As above and <sup>29</sup> NVAT[h]
MT325 (M463L)	78	47652	ND	ND	1186	ND	ND
CVM-1 (548L)	79	47622	51247	3625 (7.5)	936	4	<sup>47</sup> NGSV; [i] <sup>279</sup> NLTA; [g] <sup>285</sup> NVGY; [g] <sup>293</sup> NTAV [g]
TN603.4.2 (325L)	79	48437	54508 [i]	6071 (12.5)	1186	5	ND
ATCV-1 (Z280L)	78	48421	52827	4406 (9.1)	922	5	ND

[a] MCP molecular weight without glycans.

[b] For NY-2A the molecular weight of the two oligosaccharides are given; for ATCV-1, TN603 and MT325 the MW of the largest oligosaccharide is listed.

[c] taken from: M.V. Graves, C.T. Bernadt, R. Cerny, J.L. Van Etten. *Virology* **2001**, 285, 332–345.

[d] experimental work performed on the mixture of the two NY-2A MCPs, for which two different MWs were detected. Nonsaccharide was at Asn-302, -399, -406; hexasaccharide at Asn-54, -280, -291.

[e] averaged number calculated considering two nonsaccharides and two hexasaccharides (major MW) or one nonsaccharide and two hexasaccharides (minor MW).

[f] Asn glycosylated but previously predicted or conserved in PBCV-1.

[g] conserved with respect to PBCV-1 (see alignment sequence in Figure S1).

[h] site not predicted and present only in NY2A MCP: B617L.

[i] glycan present in this site in in CVM-1 but not in PBCV-1.

[j] MCP highly glycosylated, higher MW is given in the table, but these other masses were obtained after spectrum deconvolution: 54363, 54227, 54079, and 53950 Da.

<b>Glycopeptide isolation and chemical composition</b>	<b>S2</b>
<b>NMR and MALDI structural characterization of the most abundant MCP glycoforms</b>	<b>S3</b>
<i>Virus CVM-1</i>	S3
<i>Virus ATCV-1</i>	S5
<i>Virus TN603</i>	S8
<i>Virus MT325</i>	S9
<i>Virus NY-2A</i>	S11
<b>Experimental section.</b>	<b>S13</b>
<b>Glycan isolation</b>	<b>S13</b>
<b><i>N</i>-glycan chemical analysis</b>	<b>S13</b>
<b>NMR spectroscopy.</b>	<b>S14</b>
<b>MALDI TOF/TOF MS/MS</b>	<b>S14</b>
<b>HPLC-ESI-MS</b>	<b>S15</b>
<b>References</b>	<b>S15</b>
<b>Table S1. Virus NY-2A two major capsid proteins</b>	<b>S17</b>
<b>Table S2 (<i>N</i>-glycan chemical data)</b>	<b>S18</b>
<b>Table S3 (CVM-1 glycopeptide NMR data)</b>	<b>S19</b>
<b>Table S4 (ATCV-1 glycopeptide NMR data)</b>	<b>S20</b>
<b>Table S5 (TN606 glycopeptide NMR data)</b>	<b>S21</b>
<b>Table S6 (MT325 glycopeptide NMR data)</b>	<b>S22</b>
<b>Table S7 (NY-2A<sub>1</sub> glycopeptide NMR data)</b>	<b>S23</b>
<b>Table S8 (NY-2A<sub>2</sub> glycopeptide NMR data)</b>	<b>S24</b>
<b>Figure S1 (MCPs alignment sequence)</b>	<b>S25-S26</b>
<b>Figure S2 (GC-MS chromatograms of CVM-1 and NY2A<sub>1</sub>)</b>	<b>S27</b>
<b>Figure S3 (NMR, CVM-1 HSQC and HMBC)</b>	<b>S28</b>
<b>Figure S4 (NMR, CVM-1 COSY, TOCSY and TROESY)</b>	<b>S29</b>
<b>Figure S5 (NMR, CVM-1 HSQCTOCSY)</b>	<b>S30</b>
<b>Figure S6 (CVM-1 MALDI)</b>	<b>S31</b>
<b>Figure S7 (NMR, ATCV-1 HSQC and HMBC)</b>	<b>S32</b>
<b>Figure S8 (NMR, ATCV-1 COSY, TOCSY and TROESY)</b>	<b>S33</b>
<b>Figure S9 (ATCV-1 MALDI)</b>	<b>S34</b>
<b>Figure S10 (NMR, TN603 HSQC and HMBC)</b>	<b>S35</b>
<b>Figure S11 (NMR, TN603 COSY, TOCSY and NOESY)</b>	<b>S36</b>
<b>Figure S12 (TN603 MALDI)</b>	<b>S37</b>
<b>Figure S13 (NMR, MT325 HSQC)</b>	<b>S38</b>
<b>Figure S14 (NMR, MT325 COSY, TOCSY and TROESY)</b>	<b>S39</b>
<b>Figure S15 (MT325 MALDI)</b>	<b>S40</b>
<b>Figure S16 (NMR, NY-2A<sub>1</sub> HSQC and HMBC)</b>	<b>S41</b>
<b>Figure S17 (NMR, NY-2A<sup>1</sup> HSQC and HMBC)</b>	<b>S42</b>
<b>Figure S18 (NMR, NY-2A<sub>2</sub> COSY and TOCSY)</b>	<b>S43</b>
<b>Figure S19 (NMR, NY-2A<sub>2</sub> HSQC and HMBC)</b>	<b>S44</b>
<b>Figure S20 (NMR, NY-2A<sub>2</sub> HMBC)</b>	<b>S45</b>
<b>Figure S21 (NY-2A MALDI)</b>	<b>S46</b>

## Glycopeptide isolation and chemical composition

Glycopeptides were isolated by size exclusion chromatography after protease treatment of the major capsid protein (MCP). Chromatographic peaks were inspected by proton NMR using as reporter signals those of the  $\alpha$ -anomeric protons of the monosaccharides that occur in a region (5.5-5.0 ppm) devoid of peptide signals. Glycopeptides were eluted at approximately 50% of the column volume, and each virus gave one main glycopeptide fraction ( $^1\text{H}$  NMR in Figure 3), in a 15-20% yield with respect to the starting MCP amount. NY-2A was an exception because two different fractions were isolated, NY-2A<sub>1</sub> and NY-2A<sub>2</sub>, which were studied separately.

Monosaccharide chemical analysis was performed after the full set of NMR spectra was recorded, and the samples transformed into their corresponding acetylated methylglycoside to determine the type of monosaccharide present (Table S2).<sup>[1]</sup> Absolute configuration was determined by preparing the acetylated or methylated 2-(-)-octyl derivatives according to the procedures developed during the study of PBCV-1 *N*-glycans.<sup>[2]</sup> Pbi and SAG viruses contained rhamnose methylated at *O*-3 for which a standard was not available, indeed each sample was fully methylated and converted into the corresponding chiral octylglycosides (Figure S2A), for which the necessary standards were available from the previous studies on PBCV-1, or produced by starting from commercial monosaccharides.

For NY-2A, absolute configuration was performed on fraction NY-2A<sub>1</sub> because it contained the monosaccharide in pure form (Figure S2B). Fully acetylated 2-octylglycosides disclosed the absolute configuration of residues **A-F** and **L**, while **H** (3OMe-Xyl) and **I** (3OMe-Rha) went unassigned because of the lack of a proper standard. Information on their chirality was gained in a successive step, replacing the acetyl with methyl groups by methylation of the octylglycosides mixture and comparing the new chromatogram with that from authentic standards.

Information on the xylulose residue was lost by the two chemical procedures used above, because it is a labile monosaccharide and degrades during the alcoholysis reaction. Thus, supportive information on this residue was gained by preparation of partially methylated and acetylated derivatives of MT325 and TN603 glycopeptides.<sup>[1]</sup> Analysis of these derivatives confirmed the substitution pattern of all the residues of these glycopeptides and established that xylose occurred as a terminal and as a *O*-4 linked residue, thus confirming the location at one of the xylose units. In addition, two additional peaks were discovered because they were generated by the non stereospecific reduction of the keto function of xylulose. Interpretation of their EI-MS spectra fragmentation was consistent with a non substituted ketopentose in the furanose form.

Chemical data are summarized in Table S2 and it is worthy to note that D-rhamnose occurs only in

the NC64A viruses (PBCV-1 and NY-2A) and not in the other two groups of viruses. This result is in agreement with previous studies, which addressed the biosynthetic pathway of this monosaccharide in PBCV-1 as operated by a GDP-D-mannose-4,6-dehydratase (GMD) possessing also reductase activity.<sup>[3]</sup> A homolog to this enzyme was characterized in ATCV-1 virus as well, but it had no reductase activity so that no production of D-rhamnose occurred.<sup>[4]</sup> Hence, D-rhamnose appears to be group-specific and further bioinformatic studies on the distribution of GMD with reductase activity in the different viruses may clarify this issue.

### **NMR and MALDI structural characterization of the most abundant MCP glycoforms**

The key points of the NMR and MALDI interpretation of the glycopeptides is discussed starting from the simplest glycopeptide and proceeding in order of increasing difficulty. For each virus and at the beginning of its section, NMR field strength, the temperature of spectra acquisition and the proton/carbon chemical shift table, are described.

#### *Virus CVM-1 (600 MHz, 310 K, Table S3)*

Inspection of the low field region (5.7-4.3 ppm) of the HSQC spectrum (Figure S3) displayed six anomeric carbons, and **D** (<sup>1</sup>H 5.01 ppm) was *N*-linked to the protein because of its diagnostic carbon chemical shift (80.7 ppm), Thus this residue was located at the reducing end of the oligosaccharide. **D** was a glucose as inferred by the efficient propagation of magnetization in the TOCSY spectrum (Figure S4) and by its <sup>13</sup>C chemical shift values (Table S3), with C-3 and C-4 (77.0 and 74.7 ppm, respectively) shifted at low field by glycosylation, similarly to what was reported for the analogue residue in PBCV-1 glycopeptide.

Using a similar approach, **A** was a fucose unit: starting from the anomeric proton (5.68 ppm) its pattern in the TOCSY spectrum (Figure S4) displayed only three correlations due to the small H3-3/H-4 (<sup>3</sup>J<sub>H3,H4</sub> = 2.8 Hz) H-4/H-5 (below resolution limits) coupling constant value that hindered further propagation of the magnetization to the other protons of the residue, as typically occurs for *galacto* configured residue. H-1/C-5 long-range correlation (Figure S5) determined the position of H-5 that in turn correlated with a methyl at 1.33 ppm. Analysis of chemical shift values indicated that both C-3 and C-4 were shifted at low field, indeed glycosylated, while substitution at C-2 was not apparent by the chemical shift of this carbon (69.3 ppm) but it was deduced by the long-range correlation connecting H-2 to C-1 of **B** (Figure S3);  $\alpha$  configuration at the anomeric center was inferred on the basis of the anomeric proton shape (broad singlet) and of its proton chemical shift value (5.68 ppm). The pattern of the TOCSY spectrum of H-1 of **B** was similar to that of **A**, and

identification of its H-5 (3.99 ppm) was indicated by the NOE effect with H-3 (3.84 ppm), while the cross-peak H-5/H-6 in the COSY spectrum completed the attribution and identified this unit as a galactose. Configuration at the anomeric carbon was  $\alpha$  as inferred by the  $^3J_{H1,H2}$  value (3.0 Hz), further supported by comparison of carbon chemical shift values of this restudies with those from reference methylglycoside.<sup>[5]</sup>

Regarding **C**, TOCSY spectrum (Figure S4) displayed one main correlation from the anomeric proton, while the connections up to the methyl group (H-6) were visible from H-2 due to the favorable proton-proton coupling constants ( $^3J_{H2,H3} = 2.6$  Hz,  $^3J_{H3,H4} = ^3J_{H4,H5} = 9.6$  Hz,  $^3J_{H5,H6} = 6.3$  Hz). This residue was a terminal rhamnose,  $\alpha$  configured at the anomeric center on the basis of the agreement between its C-5 value (70.3 ppm) and the value reported for the  $\alpha$ -methylrhamnoside (69.4 ppm for  $\alpha$  anomer, 73.6 for  $\beta$  anomer).<sup>[5]</sup> This unit was acetylated at *O*-2 as suggested by H-2 low-field chemical shift (5.47 ppm), and methylated at *O*-3 due to the C-3 value (80.3 ppm) along with a long range correlation between its H-3 and methyl ( $^1H/^{13}C$  3.44/58.4 ppm). The anomeric signals of **E** and **F** coincided at 4.41 ppm and their  $^3J_{H1,H2}$  (7.1 Hz) was indicative of the  $\beta$  configuration at each anomeric center, in agreement with the chemical shift values associated to the anomeric carbons. Discrimination of the two different spin systems was achieved by analyzing the correlations in the TOCSY spectrum associated with isolated signals from each of them: H-2 for **F** (3.13 ppm) and H-5<sub>eq</sub> for **E** (3.18 ppm). The efficient propagation of the magnetization in the TOCSY spectrum suggested that all the ring protons (other than H-5<sub>eq</sub>) were axial oriented, as occurs for *gluco* configured residues. Indeed, H-2 (or H-5<sub>eq</sub>) of **F** and H-5<sub>eq</sub> of **E** had distinct correlations (Figure S4) each compatible with a five carbon atoms monosaccharide, a xylose. In agreement with this attribution were also the few proton-proton coupling constants that could be deduced from the proton spectrum (**E**:  $^3J_{H4,H5eq} = ^3J_{H5ax,H5eq} = 11.2$  Hz; **F**  $^3J_{H2,H3} = 8.0$  Hz,  $^3J_{H4,H5eq} = ^3J_{H5ax,H5eq} = 11.2$  Hz)

Accordingly, **E** and **F** were two terminal  $\beta$ -xylose units, and integration with heteronuclear spectra information established that **E** was linked at *O*-4 of **A**, while **F** at *O*-4 of **D** (Figures S3 and S5). Xylose **E** was named distal while **F** proximal because of the different distance from the *N*-glycosylation site.

Completion of the CVM-1 glycopeptide structure was inferred by analyzing the HMBC spectrum (Figure S3 and S5), which disclosed that **A** was linked at *O*-3 of **D** and in turn had **B** at *O*-2 and **C** at *O*-3. Taken together, the structure of CVM-1 was established as reported in Figure 1, and confirmed by MALDI MS and MS/MS (Figure S6). The interpretation of glycopeptide MS/MS spectra was particularly clear-cut as CVM-1 MCP had only one *N*-glycan structure. Furthermore,

the presence of fragments at  $\sim m/z$  976, corresponding to C ions retaining the Asn amino group, revealed the molecular weight of each associated peptide moiety. Breaking the glycopeptide linkage revealed easily recognizable C ions since, unlike other possible C-type ions, they have even  $m/z$  values. The most significant MS/MS spectra, showing the distinctive CVM-1 hexasaccharide glycoform linked to four different glycosylation sites at Asn-47, -279, -285, -293, are reported in Figure S6.

Of note and compared to NMR data available for the PBCV-1 *N*-glycans, chemical shift of **A** was different even though it was substituted in a similar fashion (Figure 1). This change was revealed by examining the absolute configuration of all the sugars, which was the same for all the residues but rhamnose **C**, which was L instead of D.  $^{13}\text{C}$  chemical shift displacement with respect to standard values<sup>[5]</sup> in oligosaccharides depends on stereochemical factors that rule the amplitude of the  $\alpha$ - and  $\beta$ -glycosylation effects as a function of the relative stereochemistry of the sugars.<sup>[6]</sup> The rigorous application of these rules to the  $\alpha$ -L-Rha-(1 $\rightarrow$ 3)-L-Fuc motif was complicated because fucose was substituted also at *O*-2 and *O*-4, so that  $^{13}\text{C}$  values at C-3 of fucose experienced the counteracting  $\beta$ -glycosylation shifts of the neighboring substituents. Indeed, focusing on C-1 of rhamnose, rules predicted a large  $\alpha$ -glycosylation effect ( $+8.0 \pm 0.4$  ppm) for the  $\alpha$ -D-Rha-(1 $\rightarrow$ 3)-L-Fuc entity, which agreed well with the experimental value in the PBCV-1 oligosaccharide (103.5 ppm). For the  $\alpha$ -L-Rha-(1 $\rightarrow$ 3)-L-Fuc fragment of the CVM-1 oligosaccharide, rules predicted a small shift ( $+2.9 \pm 0.8$  ppm) and the value measured (94.5 ppm) underwent a shift of -0.5 ppm, probably due to the acetyl group at *O*-2 of the monosaccharide. The correct predicted shift was found for the other viruses having a non *O*-2 acetylated L-rhamnose (as ATCV-1).

#### *Virus ATCV-1 (600 MHz, 318 K, Table S4)*

The proton spectrum of ATCV-1 glycopeptide had several similarities with that of CVM-1 (Figure 3): anomeric protons of **A**, **B** and **D** had the same chemical shift, **C** was not acetylated at *O*-2 and its anomeric proton was mildly shifted at highfield, while *O*-3 was methylated. In addition, proton and carbon chemical shifts of **A**, **B** and **D** were very similar to those found for CVM-1, in agreement with their attribution as  $\alpha$ -fucose,  $\alpha$ -galactose and  $\beta$ -glucose, respectively. The main differences were in the xylose units: the distal and the proximal units were discriminated in the HSQC spectrum (Figure S7) because of the distinct H-1/C-1 chemical shifts, with **E** having the most deshielded anomeric carbon value (105.7 ppm). Inspection of the COSY spectrum (Figure S8C) disclosed a high degree of complexity in the sample: four different H-1/H-2 cross peaks at ca. 4.45/3.41 ppm

(Figure S8C) indicated the presence of a group of distal xylose units, collectively named as group E. Similarly for the proximal xylose units (group F), the H-1/H-2 cross peaks formed three different groups (Figure S8C). Such complexity prevented any reading of the  $^3J_{H1,H2}$  coupling constant, and the  $\beta$ -configuration at the anomeric center of the two groups of signals, was attributed on the basis of the anomeric carbon chemical shift values. Attribution of the different spin systems was achieved for two out of the four distal xylose units (**E** and **E'**), and for the two proximal residues (**F** and **F'**). Analysis started from the H-5<sub>ax</sub> protons, for which overlap was not as severe as for the anomeric protons (Figure S8B). First, the COSY cross peak 4.39/3.20 was not a H-1/H-2 correlation but connected two diastereotopic methylene protons, H-5<sub>eq</sub> and H-5<sub>ax</sub> of **F'** (Figure S8A). The corresponding carbon was shifted at highfield (64.0 ppm, Figure S7) with respect to standard value (66.3 ppm)<sup>[5]</sup> because O-4 was methylated (OMe  $^1H/^{13}C$  3.48/59.7 ppm), accordingly C-4 was shifted at lowfield (80.4 versus 70.4 ppm as standard value) and it was identified by the long-range correlation connecting H-5<sub>ax</sub>/C-4 (Figure S7). COSY spectrum connected H-4 to H-3 and in turn to the other ring protons (Figure S8B), so that H-1 was identified, and found coincident with its own H-5<sub>eq</sub>; the *inter*-residue NOEs of this residue (Figure S8C) confirmed that it was linked at O-4 of **D**. The correlation at 4.41/3.14 ppm (Figure S8A) was given to H-1/H-2 of **F**, H-2 identified H-3, while H-4 was discriminated from the almost coincident H-4 of **E'** by the H-2/H-4 cross peak in the TROESY spectrum. HSQC analysis determined that **F** was a terminal xylose, linked at O-4 of **D**. Identification of **F** allowed us to assign the correlations left to the other xylose unit, **E'**, which had the same H-1 and H-5<sub>eq,ax</sub> chemical shifts as **F**, but diverged for the values of the other ring protons. Inspection of homonuclear spectra disclosed that H-4 and H-2 were coincident; the carbon signal at 80.0 ppm was given to C-4 assuming that this position was O-methylated, a hypothesis supported by the modest highfield displacement observed for C-3 and C-5. Accordingly, C-2 was unaffected by the methylation at O-4 and its chemical shift was equal to that of **E**. Indeed, **E'** was a  $\beta$ -xylose terminal and methylated at O-4 as **F'**, but linked at O-4 of the fucose unit. Attribution of **E** was inferred by COSY and TOCSY analyses, H-5<sub>ax</sub> correlated with H-4 which in turn identified H-3, which was proximal to H-2, therefore not identified by the direct H-2/H-3 COSY cross peak that instead merged in the diagonal of the spectrum. Analysis of the HSQC spectrum established that **E** was a terminal and a non methylated  $\beta$ -xylose, linked at O-4 of the fucose unit, as found for **E'**.

These attributions left the remaining two xylose units of the group E unassigned, the anomeric proton of the two residues coincided at 4.43 ppm but their H-2 were slightly different and found at 3.41 and 3.42 ppm, similar to those of **E** and **E'**, respectively. Inspection of the TOCSY spectrum at these anomeric protons, displayed correlations similar to those reported for both **E** and **E'** (starred

densities in Figure S8A), suggesting that one of these unit was methylated at *O*-4 as **E'** while the other was not.

A more careful inspection of the TOCSY spectrum indicated that the fucose unit was heterogeneous and it was split in **A** and **A'** residues, differing mainly by the position of the H-3 and H-4 protons (detailed in Figure S7). Interestingly, the TROESY spectrum placed the anomeric proton of **E'** with H-4 of **A'** while that of **E** with H-4 of **A** (Figure S8C), so that the difference between **A** and **A'** was explained as depending on the occurrence of a methyl at *O*-4 of the xylose unit present at *O*-4 of the fucose residue (Figure 1).

Taken together, the NMR data supported the structure shown in Figure 1, but it did not explain the elevated number of xylose residues of group E, which instead was understood by analyzing the MALDI spectra (Figure S9). The reflectron positive MALDI spectrum (Figure S9A) contained several ion peaks. We focused our attention on the base peak at  $m/z$  1063.29, consistent with the sodium form of the species  $\text{dhex}_2\text{hex}_2\text{Me}_2\text{pent}_2\text{Asn}$ , a hexasaccharide with two *O*-methyl groups. Based on the NMR results, the first was located on the deoxyhexose residue (rhamnose **C**), consequently the second should belong to xylose **E'** or **F'**. Indeed, to pinpoint *O*-methylation, MS/MS analysis was performed on the base peak. The resulting spectrum (Figure S9B) contained two couples of diagnostic ions:  $m/z$  449.1, 655.2 and  $m/z$  463.1, 641.2. These fragments, originated by the rupture of the linkage connecting the fucose to the glucose unit, were assigned to Y-type ions ( $m/z$  463.1 and 449.1) whose composition were consistent with a hexpent Asn moiety whether bearing ( $m/z$  463.13) or not bearing ( $m/z$  449.10) a methyl group, thus indicating that the proximal xylose was not always methylated at *O*-4, as found for **F** and **F'**. The second pair of ions at  $m/z$  641.2 and 655.2 were assigned to C-type fragments complementary to the Y ions at  $m/z$  463.1 and 449.1 respectively, suggesting that once the proximal xylose is methylated, the distal xylose residue was not substituted or vice versa. Accordingly, the dominant ion of the MS glycopeptide profile (Figure S9A,  $m/z$  1063.29) accounted for two isobaric species (as depicted in Figure S9B) depending on the type of xylose methylated. Moreover, two additional MS peaks, differing at higher and lower mass values for 14 u ( $m/z$  1077.31 and  $m/z$  1049.28) were assigned to two glycopeptide structures, one having both xylose units methylated and the other had no methyl at any xylose residue, respectively. Collectively these data implied that ATCV-1 had a mixture of four glycopeptides (Figure 1), sharing the same oligosaccharide scaffold and methylated in a non stoichiometric fashion at *O*-4 of both proximal and distal xylose units.

*Virus TN603 (homonuclear spectra at 950 MHz, heteronuclear spectra at 600 MHz, 318 K, Table S5)*

The proton spectrum of TN603 glycopeptide (Figure 3) was very similar to those of the previous samples, and detailed NMR analysis (HSQC in Figure S10, NMR attribution in Table S5) supported from the NMR data of the previous glycopeptides, confirmed the presence of a hyperbranched fucose (residue **A**), a terminal galactose (**B**), a terminal 3-OMe-rhamnose (**C**) and glucose (**D**) *N*-linked to an asparagine. Similar to ATCV-1, distal and proximal xylose units (groups **E** and **F**, respectively) were heterogeneous, in that three anomeric protons were counted for group **E** and two for **F** group (see COSY spectrum in Figure S11A). As for ATCV-1, attribution was possible by analyzing the correlations in the region at ca. 3.2 ppm of the COSY and TOCSY spectra, where the axial orientated H-5' were found (Figure S11B). By this analysis, **E'** was a terminal xylose methylated at *O*-4, linked at *O*-4 of the fucose residue, as occurring for the other two xylose units of the same group, **E** and **E''** (TROESY spectrum in Figure S11C). The two residues, **E** and **E''** were not equivalent: C-3 of **E''** was upfield shifted (75.8 ppm) with respect to C-3 of **E** (76.8 ppm) while the opposite was true for the C-4 values (71.3 vs 70.6 ppm), and a similar trend was observed for **F** and **F''**, as well. These modest chemical shift variations were compatible with the occurrence of a ketose residue located at *O*-4 of **E''** and **F''**.

This hypothesis was confirmed and a group of ketofuranose residues was identified and named **G**. Attribution started from H-5/C-5 signal in the HSQC spectrum (Figure S10) at <sup>1</sup>H/<sup>13</sup>C 4.32-3.66/71.6 ppm; H-5 at 4.32 ppm correlated with H-4 (ca. 4.46 ppm) which then with H-3. Appreciation of the number of components in the **G** group came from inspection of TOCSY spectrum, which discerned three **G**, **G'** and **G''** residues (Figure S11A) on the basis of their H-3/C-3 values (Figure S10, Table S5), C-3 from **G** and **G'** in turn had a long range correlation with their own C-1, so that H-1/C-1 of **G''** (<sup>1</sup>H/<sup>13</sup>C 3.51/62.3 ppm) was given by difference. Identification of C-2 chemical shift was possible by the long-range correlation with H-5' at 3.66 ppm (Figure S10). Of note, most of the C-2 values were at 106.2 ppm with the exception of one, attributed to **G'**, which was upfield shifted at 105.5 ppm. Attribution of H-1/C-1 for **G** and **G'** was determined by the corresponding H-1/C-3 long range correlation, while H-1/C-1 of **G''** (<sup>1</sup>H/<sup>13</sup>C 3.51/62.3 ppm) was given by exclusion, but supported by the corresponding long range correlation, which connected its H-1 to an anomeric carbon at 106.2 ppm (Figure S10).

Indeed, comparison of carbon chemical shift with those reported in literature,<sup>[7]</sup> identified **G**, **G'** and **G''** as terminal β-xylulofuranose residues, with **G'** further methylated at *O*-1, as indicated by the value of its C-1 (73.8 ppm) and by the long range connection to a methyl group. Location of **G** (or

**G''**) at *O*-4 of **F''** was inferred by long-range correlation in the HMBC spectrum, corroborating the assumption that **E''** was also glycosylated with a ketose.

Collectively, this information indicated that TN603 had a complex mixture of oligosaccharides, differing by the substitution type of both distal and proximal xylose residues. These residues could either be methylated at *O*-4 (as **E'**) or attached to a xylulose unit (as **E''** and **F''**) or be terminal (as **E** and **F**). In addition, xylulose could be methylated (**G'**), increasing the level of structural complexity of the oligosaccharide. Indeed, this great variation in structures detected by NMR analysis was confirmed by MALDI MS analysis.

MALDI mass spectra of two glycopeptide mixtures obtained by proteinase K digestion of the TN603 capsid protein are shown in Figures S12A and S12B. These last spectra are very dense with  $[M+Na]^+$  ion signals, accounting for the extreme heterogeneity in their *N*-glycan moieties and also for the variability of the (small) peptide portions. To obtain in-depth information on the structures of the associated glycan moieties, MS/MS analyses were performed on many glycopeptide ions (those underlined in Figures S12A and S12B). Here we report a representative set of the fragmentation spectra ( $m/z$  1150.37, 1280.42, 1252.49 and 1266.50) as shown in Figures S12C-F, respectively. All the MS/MS analyses, support the NMR findings that indicate that a terminal xylulose residue can be attached to either the proximal or distal xylose unit.

#### *Virus MT325 (950 MHz, 310 K, Table S6)*

NMR spectra of MT325 (Figures S13 and S14, Table S6) had several similarities with TN603, along with some differences. First, the low field region of the proton spectrum (Figure 3) contained the same anomeric signals but the proportion of the C group signals (5.18-5.08 ppm, 3-OMe-rhamnose) differed from those of TN603. Second, a new methyl group occurred at 3.53 ppm while that at 3.41 ppm, attribute to *O*-1 of Xulf (residue **G'**) in TN603, was absent.

Inspection of 2D homonuclear spectra (Figure S14) detected different types of xylulofuranose units (group G) along with a complicated pattern for both distal and proximal xylose residues (groups E and F), for which attribution of the main forms was possible inspecting the region of the axial orientated H-5 protons (Figure S14B). Accordingly, two different types of distal xylose residues were found, one terminal (**E**) and one with a xylulofuranose residue at *O*-4 (**E''**). None of the distal xyloses were methylated at *O*-4, as found in the TN603 glycopeptides. In addition, three different proximal xylose units were detected, one terminal (**F**), one having a methyl at *O*-4 (**F'**) as in ATCV-1 but not TN603, and one having a xulf residue at *O*-4 (**F''**).

Regarding the xylulofuranose residues, inspection of the H-4/C-4 region (ca. 4.50/75.0 ppm) in the HSQC spectrum (Figure S13) detected three different units, and their full attribution confirmed that two of them were terminal and an analogue to **G** and **G'** previously found in TN603. The third residue, **G''**, differed from the other two especially for the H-3/C-3 chemical shifts (3.67/87.9 ppm): the low field carbon chemical shift displacement suggested *O*-methylation at this position, as confirmed by the NOE effect between the methyl at 3.53 ppm and H-3. In addition, H-4 (4.54 ppm) of **G''** did not overlap with any other proton and had a weak cross peak with H-1 of distal xylose residues, supporting its location at O-4 of **E''**.

Similar to TN603, MT325 produced a complex mixture of glycopeptides (Figure 1), which shared the same internal oligosaccharide structure (residues **A-D**) but differed in the substitution pattern of the proximal and distal xylose units. These could be terminal (**F** and **E**) or substituted with a xylulofuranose residue (**F'** and **E'**) and harbored other small variations: methylation was restricted at *O*-4 of the proximal xylose (**F'**) or at *O*-3 of Xulf at the distal Xyl unit. All these variations generated a complex number of glycopeptides for which the residues of the inner part of the molecules appeared at distinct NMR chemical shift, so four different types of terminal 3-OMeRha and two hyperbranched fucose were identified, even though it was not possible to connect any of them to a specific glycoform.

A comprehensive picture of some of the possible glycoforms is given in Figure 1 and is supported by MALDI analysis. MALDI mass spectra of two glycopeptide mixtures obtained by proteinase K digestion of MT325 capsid protein are shown in Figures S15A and S15B. Also in this case, several peaks were present due to the heterogeneity of the *N*-glycoforms attached, that give rise, in some cases, to ions corresponding to multiple glycopeptide structures, as revealed by the MS/MS experiments. Again, MS/MS spectra were performed on a large number of peaks (those underlined) and more informative experiments are reported in Figure S15C-F. The fragmentation analysis of some significant ions (*m/z* 1209.24, 1252.25, 1063.19 and 1162.23), disclosed the particularly complex glycosylation pattern in MT325 and its minor differences with TN603 glycosylation. As an example, the MS/MS spectrum of the ion at 1252.25 (Figure S15D), also present in the TN603 glycopeptide mixture (MS/MS in Figure S12E) and associated in both cases with a couple of structures, showed a rather different fragmentation pattern, due to a slight variation in one of the two glycopeptide components. These spectra, which supported the NMR conclusions, demonstrated that, in some cases, the proximal xylose could be missing; these glycoforms have low abundance and were outside the NMR detection limits.

*Virus NY-2A (600 MHz, 310 K, Table S7 and S8)*

NY-2A produced two different glycopeptide fractions. Compared to NY-2A<sub>2</sub>, NY-2A<sub>1</sub> displayed a minor number of anomeric signals, all in a 1:1 ratio (Figure 3), and so it was studied first (Table S7). Analysis of the HSQC spectrum (range 5.7-4.3 ppm, Figure S16) showed nine anomeric signals along with two ring proton signals, a crowded carbinolic region (range 4.4-3.01 ppm) two O-Me signals (3.62 and 3.47 ppm) and three methyl groups at ca. 1.3 ppm, typical of deoxyhexoses. Full attribution of the complete set of spectra disclosed that NY-2A<sub>1</sub> shared many traits with the other glycopeptides: the hyperbranched fucose (**A**), the terminal  $\alpha$ -galactose (**B**,  $^3J_{H1,H2} = 3.6$  Hz), the *N*-linked glucose (**D**), and the proximal and distal  $\beta$ -xylose units (**F** and **E**,  $^3J_{H1,H2} = 7.3$  and  $7.0$  Hz, respectively). All these residues were linked as in the other glycopeptides. Of note and within these conserved residues, proximal and distal xylose units were not heterogeneous as occurred for the glycopeptides from ATCV-1, TN603 and MT325. The anomeric proton signal of **D** appeared rather broad and unresolved; however, this pattern was related to the heterogeneity of the peptide portion containing the Asn residue to which this monosaccharide was linked, and not to its neighbouring sugar residues. Regarding the hyperbranched fucose **A**, its carbon chemical shifts values (Table S7) were similar to those published for the equivalent unit in PBCV-1. This agreement resulted from the nature of the **C** residue, an  $\alpha$ -rhamnose on the basis of its H-5/C-5 chemical shifts and linked at O-3 of **A**, for which chemical analysis disclosed that its absolute configuration was D and not L like the glycopeptides of the other viruses (see discussion in CVM-1 section). In addition, **C** was substituted at both O-3 and O-4 and examination of HMBC spectrum (Figure S17) identified in **I** and **L** its substituents, two residues not contained in any of the previous glycopeptides.

The HSQCTOCSY propagation pattern from **L** anomeric proton contained only three densities (Figure S17), diagnostic of the *galacto* ring stereochemistry of the residue and integration of this information with those from the other spectra defined this unit as a  $\beta$ -galactose (C-5 at 76.6 ppm) and glycosylated at O-2 because of the lowfield shift of its C-2 (75.4 ppm) with respect to the standard value (71.7 ppm).<sup>[5]</sup> HSQCTOCSY (Figure S17) of **I** contain only the correlation with C-2, characteristic of *manno* configured residues; analysis of the spectra starting from H-2 determined the position of all the other protons and carbons (Table S7). As a result, this residue is a terminal  $\alpha$ -rhamnose because of the similarity of its C-5 value (69.2 ppm) with that of the reference methylglycoside ( $\alpha$ -anomer is 69.4 ppm while for  $\beta$ -anomer is 73.6 ppm);<sup>[5]</sup> moreover its C-3 value (80.4 ppm) indicated that the hydroxyl function was capped with a methyl group ( $^1\text{H}/^{13}\text{C}$  3.47/57.0 ppm).

The last **H** was also methylated as prompted by the downfield shift of its C-3 ( $^1\text{H}/^{13}\text{C}$  3.62/61.2 ppm Figure S13), but differently from **I**, its pattern in the TOCSY spectrum showed the complete magnetization propagation from H-1 up to the two protons H-5, identifying this unit as a terminal  $\alpha$ -xylose ( $^3J_{\text{H1,H2}} = 3.8$  ppm) methylated at O-3. This unit was located at O-2 of **L** unit on the basis of the NOE effect detected in the TROESY spectrum, completing the structure of the glycopeptide, a highly branched nonasaccharide (Figure 1). Analysis of 2-octylglycosides determined the absolute configuration of all the residues and in particular showed that the two rhamnose units did not possess the same absolute stereochemistry, being **C** a D-rhamnose and **I** L-configured.

Full attribution of the glycopeptide in the NY-2A<sub>1</sub> fraction facilitated the interpretation of the NMR spectra of the second chromatographic fraction (Table S8), which contained all the signals assigned previously to the nonasaccharide and for which the same letters were maintained through the attribution. Indeed, attention focused on the new signals and in particular on the anomeric proton at 5.04 ppm: TOCSY spectrum (Figure S18) from this signal, labelled **C'**, had one intense correlation with H-2 and a weak one, attributed to H-3 by integrating COSY information. This pattern indicated a *manno* configured residue and attribution of the other proton resonances from homonuclear spectra (Figure S18) along with carbon chemical shifts from HSQC spectrum (Figure S19) indicated that **C'** was an unsubstituted rhamnose. Inspection of the HMBC spectrum (Figure S20) detected a long range correlation connecting H-1 of **D'** to a carbon at 77.2 ppm, attributed to C-3 of a fucose, named **A'** having the anomeric proton almost coincident with that indicated as **A**.

Thus, NY-2A produced two different glycopeptides, a nonasaccharide and its truncated form (Figure 1), which differed from the longer form because of the lack of the substituents at the rhamnose **C** at the hyperbranched fucose **A**. The truncated form is a hexasaccharide and the lack of substituents on this unit generated a magnetic environment different from that of the nonasaccharide, so that the chemical shifts of the other sugars were affected. In the spectra of the mixture of the two glycopeptides, this is exemplified by the occurrence of two types of hyperbranched fucose units, **A** and **A'**, and two types of terminal galactose units, **B** and **B'**. The correct placement of all the residues on the correct structure was possible by using the data previously evaluated for fraction NY-2A<sub>1</sub>.

## **Experimental section.**

### **Glycan isolation.**

Procedures for growing and purifying the chloroviruses have been described.<sup>[8a,b,c]</sup> The MCPs were purified to near homogeneity as previously reported.<sup>[9]</sup> Briefly, purified viruses were suspended in 1 mL of 100 mM Tris, 50 mM NaCl, 10 mM MgCl<sub>2</sub> at pH 7.5 and heated at 70°C for 20 min. A first supernatant was obtained by centrifugation (4°C, 15 min. 10000 g) and the pellet was again extracted with the same buffer. The supernatants from the first and the second extraction were treated separately with 5 volumes of cold acetone and left at -20°C overnight. MCPs were recovered as precipitate by centrifugation (4°C, 15 min. 10000 g) and inspected by SDS-PAGE, which disclosed for each of them a homogeneous band at approx. 50 kDa MW, except for NY-2A, which presented two close bands in the same molecular weight range. MCPs were present in both extracts, which were pooled, yielding 0.5-2 mg of MCP for each virus.

Ultra-purified whole virus proteins were separated by 8% SDS-PAGE as described previously.<sup>[10]</sup> Approximately 100 µg of protein were loaded into each well for glycan staining and 10 µg of protein for silver staining. The gel was stained with Pierce glycoprotein staining kit (Thermo Scientific) or silver stain.<sup>[11,12]</sup>

Proteolytic cleavage of MCPs was performed with proteinase K (Sigma P6556) at 37°C. To ensure complete digestion, proteinase K was added two times (0.2 mg each time) at 8 hr intervals. The mixture was freeze-dried, suspended in water and purified on a Bio-Gel P-10 (Bio-Rad 150-4144) chromatographic column. The peptides were monitored with an online refractive index detector (Knauer K-2310) and fractions were pooled accordingly and monitored via <sup>1</sup>H NMR.

### **N-glycan chemical analysis**

The monosaccharide composition was performed by analyzing the acetylated methylglycoside derivatives according to the procedures reported.<sup>[1]</sup> Absolute configuration was determined by preparing the acetylated or methylated 2-(-)-octyl derivatives according to procedures developed during the study of the PBCV-1 N-glycans.<sup>[2]</sup>

All GC-MS analyses were performed with an Agilent instrument (GC instrument Agilent 6850 coupled to a MS Agilent 5973), equipped with a SPB-5 capillary column (Supelco, 30 m x 0.25 i.d., flow rate, 0.8 mL/min) and He as carrier gas. Electron impact mass spectra were recorded at an ionization energy of 70 eV and ionizing current of 0.2 mA. The temperature program used for the analyses was 150°C, for 5 min, 150 → 280 C at 3°C/min 300°C for 5 min.

## **NMR spectroscopy.**

NMR experiments were carried out on a Bruker DRX-600 instrument equipped with a cryo-probe except in those cases where a Bruker 950 spectrometer at the facility in Grenoble was used. The set of two-dimensional spectra (DQ-COSY, TOCSY, TROESY, gHSQC, and gHMBC) were measured for glycopeptides from viruses NY-2A, CVM-1, TN603 and ATCV-1. NY2A<sub>1</sub> fraction and CVM-1 were more abundant and for them it was possible to also acquire HMBC and HSQCTOCSY spectra. For the TN603 glycopeptide, the complexity of the spectrum required the use of the 950 spectrometer to resolve all the signals in the homonuclear spectra (COSY, TOCSY and NOESY), while the glycopeptide from MT325 was measured on the same instrument because of the low amount of the sample. Chemical shifts in spectra recorded in D<sub>2</sub>O are expressed in  $\delta$  relative to internal acetone (2.225 and 31.45 ppm) and the temperature was adjusted for each sample to minimize the overlap of glycopeptide signals with that of the residual water peak; temperatures are indicated for each spectrum. Standard Bruker software was used for all the experiments, the spectral width was set to 10 ppm and the frequency carrier was placed at the residual peak of the solvent signal, 512 FIDs of 2048 complex data points were collected, while scan number was adjusted to reach an optimal signal to noise ratio for each sample and depended on its abundance. Typically, 24-40 scans per FID were acquired for homonuclear spectra and mixing times of 100, 250 and 200 ms were used for TOCSY, TROESY, and NOESY spectra acquisition, respectively. gHSQC spectrum was acquired with 40-100 scans per FID (depending on sample abundance), and the GARP sequence was used for <sup>13</sup>C decoupling during acquisition. When acquired, gHSQCTOCSY and gHMBC scans tripled or doubled, respectively, those of gHSQC spectrum, and a mixing time of 100 ms was used for HSQCTOCSY spectrum mixing time. Data processing was performed with the standard Bruker Topspin 3 program.

## **MALDI TOF/TOF MS/MS**

MALDI-TOF mass spectra were recorded in linear mode on a Perseptive (Framingham, MA, USA) 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:YAG laser at the wavelength of 355 nm. Mass spectra were acquired twice using DHB or alphacyano -4-hydroxycinnamic acid as MALDI matrix, with the classical dried drop sample preparation. It was proved that DHB gave better results for glycopeptides. MALDI TOF/TOF MS/MS spectra were also recorded on the 4800 system without any collision gas. DHB was used as the solid matrix.

Glycopeptides were always present as sodium  $[M+Na]^+$  and potassium  $[M+K]^+$  adducts (pseudo-molecular ions). This characteristic was sometimes used to recognize glycopeptide from peptide

pseudo-molecular ions. Since both *N*-glycan structures and glycosylation sites are extremely peculiar, classical bioinformatics tools were not useful, with the exception of FindPep tool (expasy.org), which was sometimes used for glycosylation site assignments. MS/MS spectra showed several glycosidic linkage cleavages, mainly corresponding to B, C and Y ions,<sup>[13]</sup> allowing the assignments of the branching of the oligosaccharide moieties. Where a peptide portion was large enough, peptide fragments could also be detected.

## HPLC-ESI-MS

The molecular mass of proteins was detected using an Agilent 1100 HPLC system coupled to a MSD Ion Trap XCT mass spectrometer, equipped with an electrospray ion source (HPLC-ESI-MS) (Agilent Technologies, Palo Alto, CA, USA). Separations were performed on a Symmetry C<sub>18</sub> column 1×150 mm with 3-μm particle size (Waters Corporation, Milford, MA, USA). Eluents used were water and acetonitrile added with 0.1% formic acid. The gradient employed was: 10% acetonitrile for 10 min, then linear to 95% in 50 min. The flow rate was set to 30 μL/min and the column temperature was set at 25°C. Injection volume was 8 μL. Ions were detected in an ion charged control with a target ion value of 65000 and an accumulation time of 300 ms, using the following operation parameters: capillary voltage: 2600V; nebulizer pressure: 12 psi; drying gas: 8 L/min; dry temperature: 325°C; rolling averages 3, averages 8. Mass spectra were acquired in the positive ion mode in the 400-2000 *m/z* mass range. Raw spectra were deconvoluted using the LC/MSD Trap Software revision 5.3.

## References

- [1] C. De Castro, M. Parrilli, O. Holst, A. Molinaro. *Methods Enzymol.* **2010**, *480*, 89-115.
- [2] C. De Castro, A. Molinaro, F. Piacente, J.R. Gurnon, L. Sturiale, A. Palmigiano, R. Lanzetta, M. Parrilli, D. Garozzo, M. Tonetti, J.L. Van Etten. *Proc. Natl. Acad. Sci.* **2013**, *110*, 13956-13960.
- [3] M. Tonetti, D. Zanardi, J.R. Gurnon, F. Fruscione, A. Armirotti, G. Damonte, L. Sturla, A. De Flora, J.L. Van Etten. *J. Biol. Chem.* **2003**, *278*, 21559-21565.
- [4] F. Fruscione, L. Sturla, G. Duncan, J.L. Van Etten, P. Valbuzzi, A. De Flora, E. Di Zanni, M. Tonetti. *J. Biol. Chem.* **2008**, *283*, 184-193.
- [5] K. Bock, S. Pedersen. *Adv. Carbohydr. Chem. Biochem.* **1983**, *41*, 27-65.

- [6] A.S. Shashkov, G.M. Lipkind, Y. A. Knirel, N.K. Kochetkov. *Magn. Res. Chem.* **1988**, *26*, 735-747.
- [7] R. P. Gorshova, V. V. Isakov, E. N. Kalmykova, Y. S. Ovodov. *Carbohydr. Res.* **1995**, *268*, 249-255.
- [8] a) J.L. Van Etten, D. E. Burbank, D. Kuczmariski, R.H. Meints. *Science*, **1983**, *219*, 994-996.  
b) J.L. Van Etten, D.E. Burbank, K., Y. Xia, R.H. Meints. *Virology*, **1983**, *126*, 117-125. c)  
I.V. Agarkova, D.D. Dunigan, J.L. Van Etten. *J. Virol.* **2006**, *80*, 8114-8123.
- [9] I.-N. Wang, Y. Li, Q. Que., M. Bhattacharya, L.C. Lane, W.G. Chaney, J.L. Van Etten. *Proc. Natl. Acad. Sci.* **1994**, *90*, 3840-3844.
- [10] M.P.Skrdla, D.E. Burbank, Y. Xia, R.H. Meints, J.L. Van Etten. *Virology*. **1984**, *135*, 308-315.
- [11] J.H. Morrissey. *Analytical Biochem.* **1981**, *17*, 307-31.
- [12] T. Rabilloud. *Analytical Chem.* **2000**, *72*, 48a-55a.
- [13] B. Domon, C.E. Costello. *Glycoconj J.* **1988**, *5*, 397-409.
- [14] D.D. Dunigan, R.L. Cerny, A.T. Bauman, J.C. Roach, L.C. Lane, I.V. Agarkova, K. Wulser, G.M. Yanai-Balser, J.R. Gurnon, J.C. Vitek, B.J. Kronschnabel, A. Jeanniard, G. Blanc, C. Upton, G.A. Duncan, O.W. McClung, F. Ma, J.L. Van Etten. *J. Virol.* **2012**, *86*, 8821-8834.

Table S1. Proteomic analyses of *Chlorovirus* NY-2A MCP paralogs. *NC64A-Chlorovirus* NY-2A twice gradient purified virions were protease-polished, and the proteins were separated by SDS-PAGE. Proteins in the gel were marked with Sypro-Orange, and the gel was comprehensively evaluated by mass spectrometry essentially as described.<sup>[14]</sup> The experiment was repeated 3 times with independent virus preparations and the data were combined.

<b>Virus</b>	<b>Accession number</b>	<b>Gene_tag</b>	<b>Mowse Score</b>	<b>Queries matched</b>	<b>Sequence coverage (%)</b>
NY-2A	GenBank: ABT14984.1	B585L	45497	642	60
	NCBI Reference Sequence: YP_001497813.1	B617L	44652	624	60

Table S2: Chemical data summarizing the features of the glycopeptides from the different viruses. Data from PBCV-1 are included for comparison. NS = not stoichiometric substituent. Bold letters on the side of sugar type refer to the labels used during NMR attribution.

	<i>C. variabilis</i> – NC64A viruses		<i>M. conductrix</i> – Pbi viruses		<i>C. heliozoe</i> – SAG viruses	
	PBCV-1	NY-2A	MT325	CVM-1	ATCV-1	TN603
L-fucose <b>A</b>	+	+	+	+	+	+
D-galactose <b>B</b>	+	+	+	+	+	+
D/L-rhamnose <b>C</b>	+, D-Rha	+, D-Rha	+, L-Rha, 3OMe	+, L-Rha, 2OAc, 3OMe	+, L-Rha, 3OMe	+, L-Rha, 3OMe
D-glucose <b>D</b>	+	+	+	+	+	+
D-xylose <b>E</b>	+	+	+	+	+	+
D-xylose <b>F</b>	+	+	+	+	+	+
Xylulose <b>G</b>	-	-	+, 3OMe (NS)	-	-	+, 1OMe (NS)
D-3OMe- Xylose <b>H</b>	-	+	-	-	-	-
L-3OMe- rhamnose <b>I</b>	-	+	-	-	-	-
D-galactose <b>L</b>	-	+	-	-	-	-

**Table S3:** ( $^1\text{H}$  600 MHz,  $^{13}\text{C}$  150 MHz, 310 K,  $\text{D}_2\text{O}$ ) proton and carbon chemical shifts deduced for CVM-1 glycopeptide. **C** is acetylated at *O*-2 ( $^1\text{H}/^{13}\text{C}$  2.19/21.6 ppm) and methylated at *O*-3 ( $^1\text{H}/^{13}\text{C}$  3.44/58.4 ppm). The structure is presented in Figure 1.

		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5 (5<sub>eq</sub>; 5<sub>ax</sub>)</b>	<b>6; 6'</b>
<b>A</b>	$^1\text{H}$	5.68	4.24	4.34	4.20	4.76	1.33
<b>2,3,4-<math>\alpha</math>-L-Fuc</b>	$^{13}\text{C}$	98.4	69.3	73.4	76.1	67.7	16.3
<b>B</b>	$^1\text{H}$	5.26	3.87	3.84	4.03	3.99	3.76; 3.70
<b><i>t</i>-<math>\alpha</math>-D-Gal</b>	$^{13}\text{C}$	99.7	69.4	70.9	70.5	73.0	62.5
<b>C</b>	$^1\text{H}$	5.15	5.47	3.54	3.50	4.23	1.35
<b><i>t</i>-<math>\alpha</math>-L-2OAc3OMeRha</b>	$^{13}\text{C}$	94.5	69.3	80.3	72.5	70.3	18.1
<b>D</b>	$^1\text{H}$	5.01	3.62	3.97	3.71	3.65	3.93; 3.83
<b>3,4-<math>\beta</math>-D-Glc</b>	$^{13}\text{C}$	80.7	75.0	77.0	74.7	78.1	60.7
<b>E</b>	$^1\text{H}$	4.41	3.41	3.42	3.65	3.86; 3.18	--
<b><i>t</i>-<math>\beta</math>-D-Xyl</b>	$^{13}\text{C}$	105.6	74.6	76.94	70.4	66.3	--
<b>F</b>	$^1\text{H}$	4.41	3.13	3.45	3.45	4.07; 3.26	--
<b><i>t</i>-<math>\beta</math>-D-Xyl</b>	$^{13}\text{C}$	103.9	75.08	76.8	70.9	66.5	--

**Table S4:** ( $^1\text{H}$  600 MHz,  $^{13}\text{C}$  150 MHz, 310 K,  $\text{D}_2\text{O}$ ) proton and carbon chemical shifts deduced for ATCV-1 glycopeptide. **C** is methylated at *O*-3 ( $^1\text{H}/^{13}\text{C}$  3.45/ and 3.46 /57.4 ppm) and **E'** and **F'** at *O*-4 ( $^1\text{H}/^{13}\text{C}$  3.49/59.5 and 3.48/59.7 ppm, respectively). The structures are presented in Figure 1.

		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5 (5<sub>eq</sub>; 5<sub>ax</sub>)</b>	<b>6; 6'</b>
<b>A</b>	$^1\text{H}$	5.67	4.22	4.41	4.25	4.74	1.33
<b>2,3,4-<math>\alpha</math>-L-Fuc</b>	$^{13}\text{C}$	98.4	69.5	73.1	76.3	67.6	16.4.3
<b>A'</b>	$^1\text{H}$	5.66	4.22	4.36	4.23	4.74	1.33
<b>2,3,4-<math>\alpha</math>-L-Fuc</b>	$^{13}\text{C}$	98.6	69.5	73.1	76.3	67.6	16.4
<b>B</b>	$^1\text{H}$	5.26	3.85	3.85	4.04	4.98	3.77; 3.70
<b><i>t</i>-<math>\alpha</math>-D-Gal</b>	$^{13}\text{C}$	99.6	69.6	70.5	70.5	72.9	62.5
<b>C</b>	$^1\text{H}$	5.11	4.19	3.38	3.50	4.11	1.31
<b><i>t</i>-<math>\alpha</math>-L-3OMeRha</b>	$^{13}\text{C}$	96.7	67.2	81.4	72.3	69.9	18.2
<b>D</b>	$^1\text{H}$	5.00	3.64	3.96	3.71	3.65	3.93; 3.81
<b>3,4-<math>\beta</math>-D-Glc</b>	$^{13}\text{C}$	80.7	75.0	77.0	74.7	78.1	60.7
<b>E</b>	$^1\text{H}$	4.45	3.41	3.46	3.67	3.94; 3.29	-
<b><i>t</i>-<math>\beta</math>-D-Xyl</b>	$^{13}\text{C}$	105.6	74.6	76.8	70.6	66.3	-
<b>E'</b>	$^1\text{H}$	4.42	3.42	3.51	3.42	4.15; 3.26	-
<b><i>t</i>-<math>\beta</math>-D-4OMeXyl</b>	$^{13}\text{C}$	105.6	74.7	75.8	80.0	64.0	-
<b>F</b>	$^1\text{H}$	4.41	3.14	3.45	3.49	4.14; 3.24	-
<b><i>t</i>-<math>\beta</math>-D-Xyl</b>	$^{13}\text{C}$	103.8	75.1	76.8	72.4	66.3	-
<b>F'</b>	$^1\text{H}$	4.39	3.15	3.49	3.18	4.39; 3.20	-
<b><i>t</i>-<math>\beta</math>-D-4OMeXyl</b>	$^{13}\text{C}$	103.8	75.1	75.7	80.4	64.0	-

**Table S5:** ( $^1\text{H}$  950 MHz,  $^{13}\text{C}$  237.5 MHz, 318 K,  $\text{D}_2\text{O}$ ) proton and carbon chemical shifts deduced for TN603 glycopeptide. C group is methylated at *O*-3 ( $\text{O-CH}_3$  at  $^1\text{H}/^{13}\text{C}$  3.45/57.3 ppm), **E''** at *O*-4 ( $\text{O-CH}_3$  at  $^1\text{H}/^{13}\text{C}$  3.50/59.5), **G'** at *O*-1 ( $\text{O-CH}_3$  at 3.41/60.1). The structures are presented in Figure 1.

		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5 (<math>\delta_{\text{eq}}</math>; <math>\delta_{\text{ax}}</math>)</b>	<b>6; 6'</b>
<b>A</b>	$^1\text{H}$	5.67	4.22	4.36	4.22	4.81	1.33
<b>2,3,4-<math>\alpha</math>-L-Fuc</b>	$^{13}\text{C}$	98.4	69.5	73.2	76.4	67.7	16.4
<b>A'</b>	$^1\text{H}$	5.67	4.22	4.35	4.18	4.76	1.33
<b>2,3,4-<math>\alpha</math>-L-Fuc</b>	$^{13}\text{C}$	98.4	69.5	72.9	76.4	67.7	16.4
<b>B</b>	$^1\text{H}$	5.26	3.85	3.85	4.04	3.98	3.76; 3.71
<b><i>t</i>-<math>\alpha</math>-D-Gal</b>	$^{13}\text{C}$	99.6	69.6	70.9	70.5	73.0	62.5
<b>C</b>	$^1\text{H}$	5.16	4.17	3.38	3.50	4.09	1.31
<b><i>t</i>-<math>\alpha</math>-L-3OMeRha</b>	$^{13}\text{C}$	96.3	67.2	81.3	72.3	69.8	18.2
<b>C'</b>	$^1\text{H}$	5.14	4.18	3.38	3.51	4.09	1.31
<b><i>t</i>-<math>\alpha</math>-L-3OMeRha</b>	$^{13}\text{C}$	96.3	67.2	81.3	72.3	69.8	18.2
<b>C''</b>	$^1\text{H}$	5.11	4.19	3.38	3.51	4.11	1.31
<b><i>t</i>-<math>\alpha</math>-L-3OMeRha</b>	$^{13}\text{C}$	96.8	67.1	81.3	72.3	69.7	18.2
<b>C'''</b>	$^1\text{H}$	5.09	4.20	3.39	3.51	4.11	1.31
<b><i>t</i>-<math>\alpha</math>-L-3OMeRha</b>	$^{13}\text{C}$	96.7	67.1	81.3	72.3	69.7	18.2
<b>D</b>	$^1\text{H}$	5.00	3.63	3.96	3.71	3.65	3.93; 3.82
<b>3,4-<math>\beta</math>-D-Glc</b>	$^{13}\text{C}$	80.7	75.0	77.0	74.8	78.2	60.8
<b>E</b>	$^1\text{H}$	4.43	3.40	3.45	3.67	3.94; 3.29	-
<b><i>t</i>-<math>\beta</math>-D-Xyl</b>	$^{13}\text{C}$	105.5	74.7	76.8	70.6	66.3	-
<b>E'</b>	$^1\text{H}$	4.42	3.42	3.50	3.42	4.17; 3.24	-
<b><math>\beta</math>-D-4OMeXyl</b>	$^{13}\text{C}$	105.6	74.7	75.8	80.0	64.0	-
<b>E''</b>	$^1\text{H}$	4.43	3.43	3.52	3.86	3.88; 3.32	-
<b>4-<math>\beta</math>-D-Xyl</b>	$^{13}\text{C}$	105.5	74.8	75.8	71.3	66.0	-
<b>F</b>	$^1\text{H}$	4.41	3.14	3.44	3.47	4.14; 3.25	-
<b><i>t</i>-<math>\beta</math>-D-Xyl</b>	$^{13}\text{C}$	104.8	75.3	76.7	70.9	66.3	-
<b>F''</b>	$^1\text{H}$	4.40	3.15	3.50	3.63	4.05; 3.27	-
<b>4-<math>\beta</math>-D-Xyl</b>	$^{13}\text{C}$	104.8	75.2	75.7	71.0	66.0	-
<b>G</b>	$^1\text{H}$	3.62 (x2)	-	4.10	4.46	4.32; 3.66	
<b><i>t</i>-<math>\beta</math>-Xulf</b>	$^{13}\text{C}$	63.0	106.2	78.9	75.5	71.6	
<b>G'</b>	$^1\text{H}$	3.54(x2)	-	4.06	4.44	4.32; 3.66	
<b><i>t</i>-<math>\beta</math>-1OMeXulf</b>	$^{13}\text{C}$	73.8	105.5	79.7	75.5	71.6	
<b>G''</b>	$^1\text{H}$	3.70; 3.51	-	4.14	4.54	4.32; 3.66	
<b><i>t</i>-<math>\beta</math>-Xulf</b>	$^{13}\text{C}$	62.3	106.2	78.3	75.5	71.6	

**Table S6:** ( $^1\text{H}$  950 MHz,  $^{13}\text{C}$  237.5 MHz, 310 K,  $\text{D}_2\text{O}$ ) proton and carbon chemical shifts deduced for MT325 glycopeptide. C group is methylated at *O*-3 (*O*-CH<sub>3</sub> at  $^1\text{H}/^{13}\text{C}$  3.45/57.1 ppm), F' at *O*-4 (*O*-CH<sub>3</sub> at  $^1\text{H}/^{13}\text{C}$  3.48/59.5), G''' at *O*-3 (*O*-CH<sub>3</sub> at 3.53/60.0). The structures are presented in Figure 1.

		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5 (5<sub>eq</sub>; 5<sub>ax</sub>)</b>	<b>6; 6'</b>
<b>A</b>	$^1\text{H}$	5.69	4.22	4.39	4.27	4.75	1.33
<b>2,3,4-<math>\alpha</math>-L-Fuc</b>	$^{13}\text{C}$	98.3	69.3	73.1	76.2	67.6	16.1
<b>A'</b>	$^1\text{H}$	5.68	4.22	4.35	4.20	4.81	1.34
<b>2,3,4-<math>\alpha</math>-L-Fuc</b>	$^{13}\text{C}$	98.3	69.3	72.8	76.2	67.5	16.1
<b>B</b>	$^1\text{H}$	5.26	3.85	3.84	4.03	3.98	3.76; 3.70
<b><i>t</i>-<math>\alpha</math>-D-Gal</b>	$^{13}\text{C}$	99.6	69.4	70.7	70.4	73.0	62.6
<b>C</b>	$^1\text{H}$	5.17	4.18	3.37	3.51	4.10	1.31
<b><i>t</i>-<math>\alpha</math>-L-3OMeRha</b>	$^{13}\text{C}$	96.3	67.1	81.3	72.2	69.8	18.0
<b>C'</b>	$^1\text{H}$	5.15	4.18	3.38	3.54	4.10	1.31
<b><i>t</i>-<math>\alpha</math>-L-3OMeRha</b>	$^{13}\text{C}$	96.3	67.1	81.3	72.2	69.8	18.0
<b>C''</b>	$^1\text{H}$	5.11	4.20	3.38	3.49	4.12	1.31
<b><i>t</i>-<math>\alpha</math>-L-3OMeRha</b>	$^{13}\text{C}$	96.8	67.1	81.3	72.2	70.0	18.0
<b>C'''</b>	$^1\text{H}$	5.09	4.21	3.38	3.50	4.12	1.31
<b><i>t</i>-<math>\alpha</math>-L-3OMeRha</b>	$^{13}\text{C}$	96.8	67.1	81.3	72.2	70.0	18.0
<b>D</b>	$^1\text{H}$	5.00	3.62	3.97	3.70	3.65	3.92; 3.82
<b>3,4-<math>\beta</math>-D-Glc</b>	$^{13}\text{C}$	80.6	74.1	76.8	74.7	78.0	60.6
<b>E</b>	$^1\text{H}$	4.45	3.40	3.45	3.68	3.94; 3.29	-
<b><i>t</i>-<math>\beta</math>-D-Xyl</b>	$^{13}\text{C}$	105.6	74.6	76.8	70.6	66.2	-
<b>E''</b>	$^1\text{H}$	4.44	3.42	3.49	3.88	3.88; 3.31	-
<b>4-<math>\beta</math>-D-Xyl</b>	$^{13}\text{C}$	105.6	74.8	75.6	71.0	65.9	-
<b>F</b>	$^1\text{H}$	4.41	3.14	3.45	3.47	4.14; 3.25	-
<b><i>t</i>-<math>\beta</math>-D-Xyl</b>	$^{13}\text{C}$	103.8	75.1	76.8	70.9	66.1	-
<b>F'</b>	$^1\text{H}$	4.40	3.15	3.49	3.18	4.38; 3.19	-
<b><math>\beta</math>-D-4OMeXyl</b>	$^{13}\text{C}$	103.8	75.1	75.6	80.3	63.9	-
<b>F''</b>	$^1\text{H}$	4.40	3.15	3.50	3.60	4.05; 3.27	-
<b>4-<math>\beta</math>-D-Xyl</b>	$^{13}\text{C}$	103.8	75.1	75.6	70.9	66.0	-
<b>G</b>	$^1\text{H}$	3.63; 3.59*	-	4.10	4.47	4.32; 3.67	-
<b><i>t</i>-<math>\beta</math>-Xulf</b>	$^{13}\text{C}$	62.8	ND	78.7	75.7	71.5	-
<b>G''</b>	$^1\text{H}$	3.70; 3.50**	-	4.14	4.46	4.30; 3.68	-
<b><i>t</i>-<math>\beta</math>-Xulf</b>	$^{13}\text{C}$	62.3	ND	78.0	75.3	71.5	-
<b>G'''</b>	$^1\text{H}$	***	-	3.67	4.54	4.35; 4.04	-
<b><i>t</i>-<math>\beta</math>-3OMeXulf</b>	$^{13}\text{C}$	***	ND	87.9	75.0	71.7	-

\*,\*\* values determined on the basis of those found for TN603 glycopeptides

\* the two protons of this hydroxymethylene group are distinguishable at 237.5 MHz, but not at 150 MHz, Table S3.

\*\* H-1/C-1 density is visible at 237.5 MHz, but not at 150 MHz, (Table S3)

\*\*\* H-1/C-1 could not be assigned unequivocally,

**Table S7:** ( $^1\text{H}$  600 MHz,  $^{13}\text{C}$  150 MHz, 310 K,  $\text{D}_2\text{O}$ ) proton and carbon chemical shifts deduced for NY-2A<sub>1</sub> glycopeptide. Both **H** and **I** are methylated at *O*-3 ( $^1\text{H}/^{13}\text{C}$  in ppm are 3.62/61.2 and 3.47/57.0, respectively). The structure is presented in Figure 1

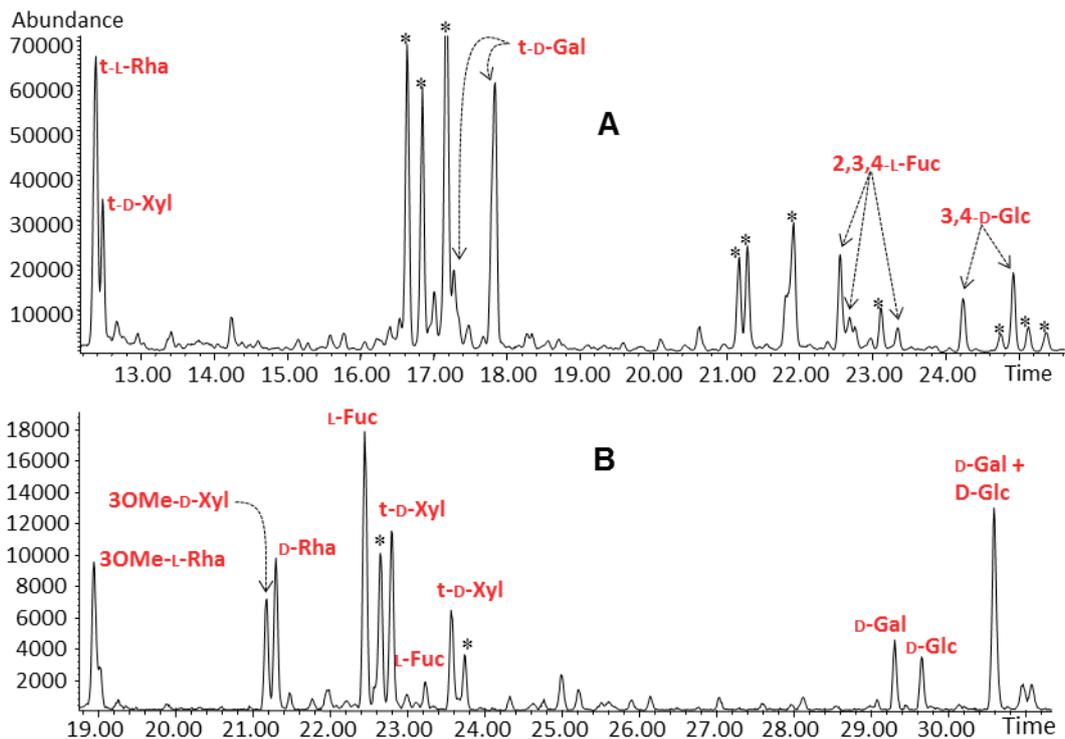
		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5 (5<sub>eq</sub>; 5<sub>ax</sub>)</b>	<b>6; 6'</b>
<b>A</b>	$^1\text{H}$	5.62	4.20	4.20	3.86	4.74	1.30
<b>2,3,4-<math>\alpha</math>-L-Fuc</b>	$^{13}\text{C}$	98.5	72.2	76.2	82.5	68.5	16.1
<b>B</b>	$^1\text{H}$	5.21	3.82	3.86	4.03	4.04	3.73; 3.69
<b><i>t</i>-<math>\alpha</math>-D-Gal</b>	$^{13}\text{C}$	100.1	69.5	70.9	70.4	72.6	62.4
<b>C</b>	$^1\text{H}$	5.12	4.16	3.87	3.85	4.06	1.43
<b>3,4-<math>\alpha</math>-D-Rha</b>	$^{13}\text{C}$	103.2	68.7	74.2	77.4	70.3	18.3
<b>D</b>	$^1\text{H}$	4.98	3.57	3.96	3.70	3.65	3.94; 3.84
<b>3,4-<math>\beta</math>-D-Glc</b>	$^{13}\text{C}$	80.7	74.9	76.9	74.9	78.1	60.6
<b>E</b>	$^1\text{H}$	4.46	3.40	3.45	3.69	4.04; 3.20	--
<b><i>t</i>-<math>\beta</math>-D-Xyl</b>	$^{13}\text{C}$	105.6	74.5	76.8	70.4	66.5	--
<b>F</b>	$^1\text{H}$	4.42	3.16	3.46	3.59	4.05; 3.28	--
<b><i>t</i>-<math>\beta</math>-D-Xyl</b>	$^{13}\text{C}$	104.1	74.8	77.1	70.9	66.1	--
<b>H</b>	$^1\text{H}$	5.34	3.58	3.49	3.66	3.94; 3.58	--
<b><i>t</i>-<math>\alpha</math>-D-3OMeXyl</b>	$^{13}\text{C}$	98.5	72.3	84.1	70.1	62.7	--
<b>I</b>	$^1\text{H}$	4.95	4.24	3.69	3.42	4.50	1.30
<b><i>t</i>-<math>\alpha</math>-L-3OMeRha</b>	$^{13}\text{C}$	97.8	67.2	80.4	73.0	69.3	18.2
<b>L</b>	$^1\text{H}$	4.71	3.63	3.73	3.91	3.70	3.91; 3.78
<b>2-<math>\beta</math>-D-Gal</b>	$^{13}\text{C}$	104.5	75.4	72.8	70.9	76.6	62.8

**Table S8:** ( $^1\text{H}$  600 MHz,  $^{13}\text{C}$  150 MHz, 310 K,  $\text{D}_2\text{O}$ ) proton and carbon chemical shifts deduced for NY-2A<sub>2</sub> glycopeptide. Both **B** and **F** are methylated at *O*-3 ( $^1\text{H}/^{13}\text{C}$  in ppm are 3.62/61.2 and 3.47/57.0, respectively). The structure is presented in Figure 1

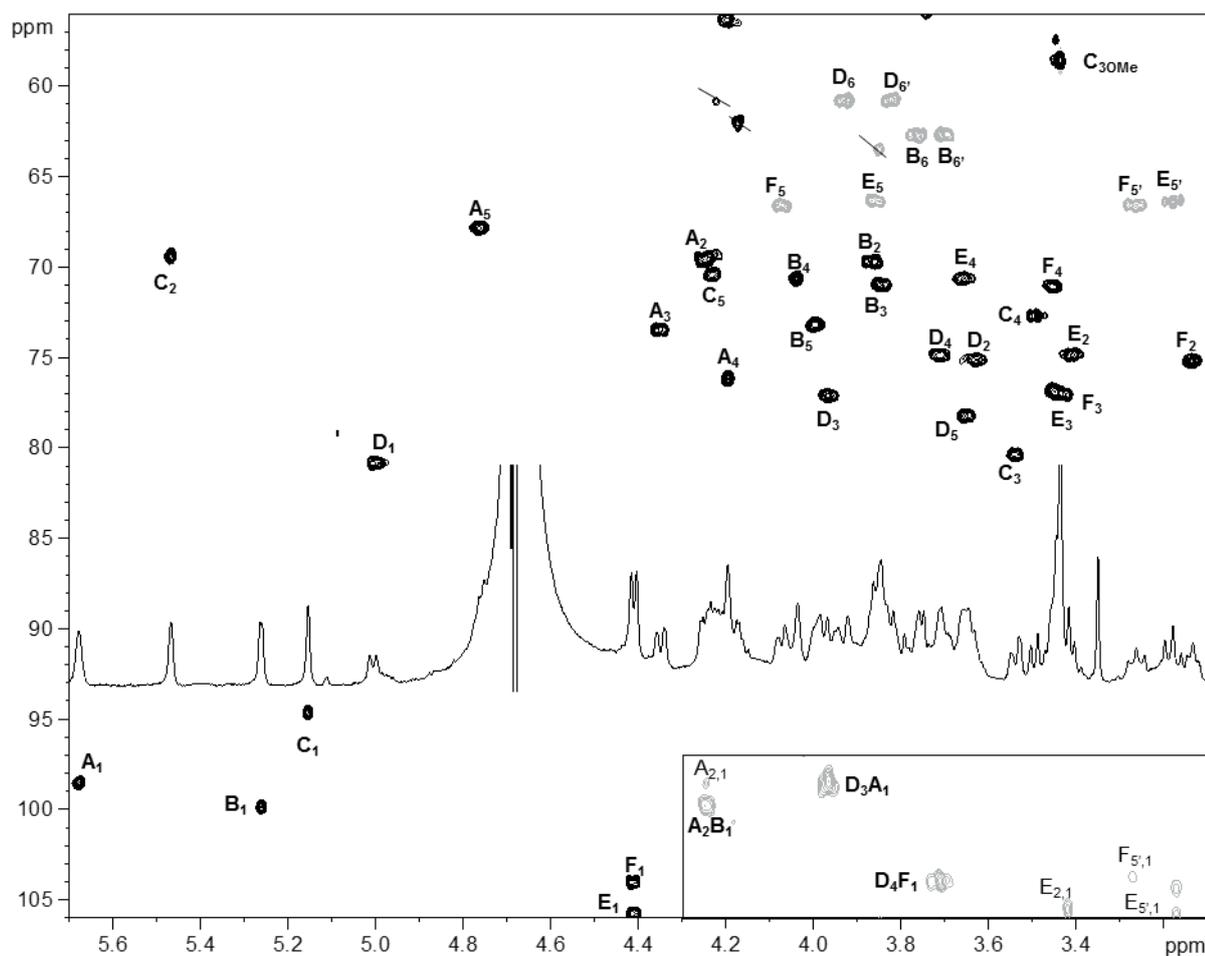
		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5 (5<sub>eq</sub>; 5<sub>ax</sub>)</b>	<b>6; 6'</b>
<b>A</b>	$^1\text{H}$	5.62	4.18	4.20	3.87	4.75	1.30
<b>2,3,4-<math>\alpha</math>-L-Fuc</b>	$^{13}\text{C}$	98.5	72.3	76.2	82.7	68.5	16.3
<b>A'</b>	$^1\text{H}$	5.63	4.16	4.21	3.86	4.75	1.30
<b>2,3,4-<math>\alpha</math>-L-Fuc</b>	$^{13}\text{C}$	98.6	72.1	77.2	82.7	68.5	16.3
<b>B</b>	$^1\text{H}$	5.21	3.81	3.85	4.03	4.03	3.73; 3.70
<b><i>t</i>-<math>\alpha</math>-D-Gal</b>	$^{13}\text{C}$	100.2	69.6	71.0	70.6	72.7	62.4
<b>B'</b>	$^1\text{H}$	5.22	3.82	3.85	4.03	4.03	3.73; 3.70
<b><i>t</i>-<math>\alpha</math>-D-Gal</b>	$^{13}\text{C}$	100.2	69.6	71.0	70.6	72.7	62.4
<b>C</b>	$^1\text{H}$	5.12	4.16	3.87	3.86	4.06	1.42
<b>3,4-<math>\alpha</math>-D-Rha</b>	$^{13}\text{C}$	103.4	68.7	74.3	77.5	70.3	18.2
<b>C'</b>	$^1\text{H}$	5.04	4.03	3.75	3.50	4.10	1.28
<b><i>t</i>-<math>\alpha</math>-D-Rha</b>	$^{13}\text{C}$	104.0	71.4	71.6	73.3	70.2	18.1
<b>D</b>	$^1\text{H}$	4.99	3.58	3.95	3.71	3.66	3.94; 3.83
<b>3,4-<math>\beta</math>-D-Glc</b>	$^{13}\text{C}$	80.7	75.1	77.0	74.9	78.3	60.6
<b>E</b>	$^1\text{H}$	4.46	3.40	3.45	3.67	4.03; 3.22	--
<b><i>t</i>-<math>\beta</math>-D-Xyl</b>	$^{13}\text{C}$	105.6	74.8	77.2	70.7	66.2	--
<b>F</b>	$^1\text{H}$	4.42	3.14	3.45	3.60	4.04; 3.28	--
<b><i>t</i>-<math>\beta</math>-D-Xyl</b>	$^{13}\text{C}$	104.1	74.9	77.2	71.1	66.2	--
<b>H</b>	$^1\text{H}$	5.34	3.58	3.49	3.66	3.94; 3.59	--
<b><i>t</i>-<math>\alpha</math>-D-3OMeXyl</b>	$^{13}\text{C}$	98.5	72.3	84.2	70.5	62.7	--
<b>I</b>	$^1\text{H}$	4.95	4.24	3.70	3.42	4.50	1.29
<b><i>t</i>-<math>\alpha</math>-D-3OMeRha</b>	$^{13}\text{C}$	97.9	67.3	80.5	73.1	69.2	16.2
<b>L</b>	$^1\text{H}$	4.71	3.64	3.74	3.91	3.71	3.90; 3.78
<b>2-<math>\beta</math>-D-Gal</b>	$^{13}\text{C}$	104.5	75.4	72.8	70.9	76.6	62.8



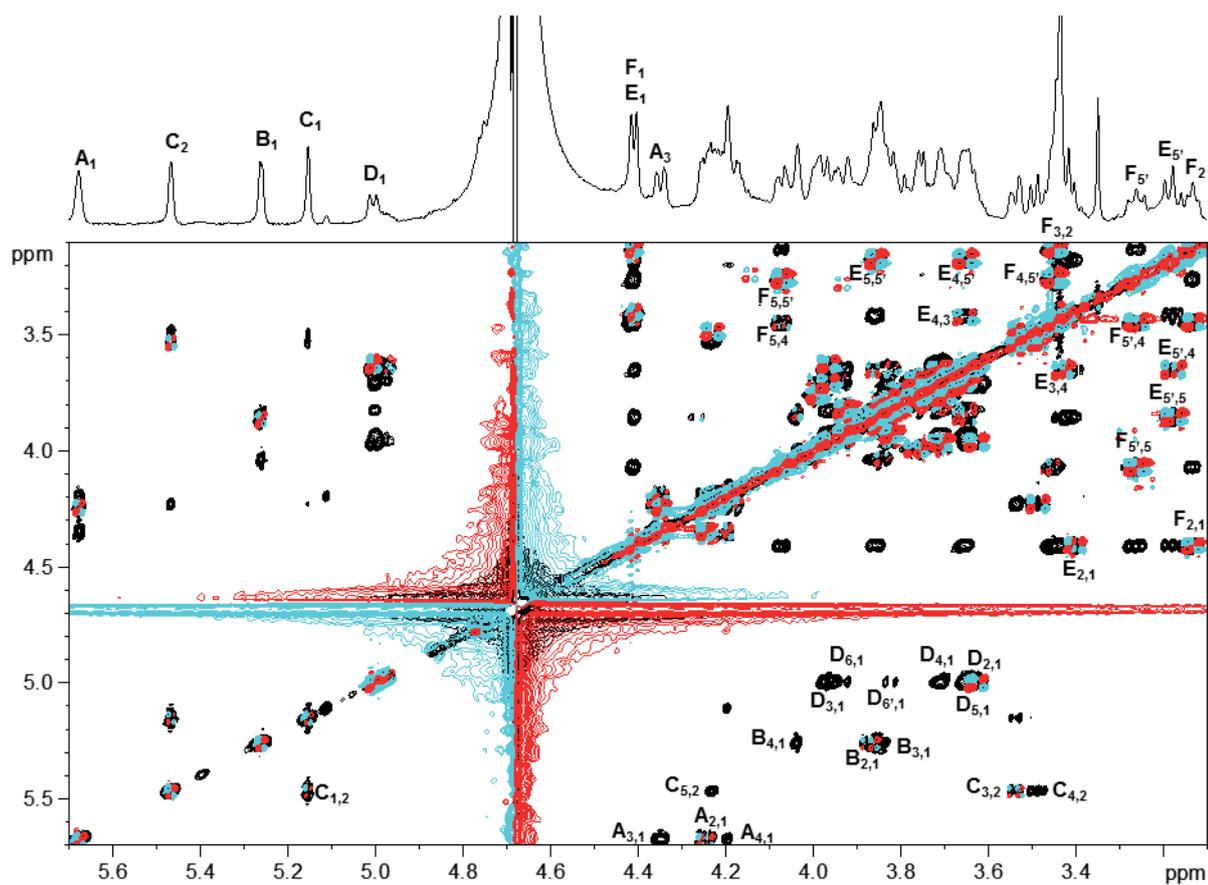
**Figure S1.** Alignment of the MCP sequences for the five chloroviruses included in this study. The MCP of PBCV-1 is included as a reference. The sequences were aligned with ClustalW2 (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>). Each label contains the virus name, gene name and accession number. In PBCV-1, N-linked glycans are at amino acid positions 280, 302, 399 and 406. Since NY-2A encodes two nearly identical paralogs, B585L and B617L, both were included in the alignment. The asparagines highlighted in yellow are *N*-linked glycosylation sites in the virus PBCV-1, along with the putative *N*-linked glycosylation sites in the other viruses. For the virus CVM-1, the asparagines highlighted in turquoise are glycosylated despite the good homology with PBCV-1 MCP where glycosylation does not occur. In NY-2A, glycosylation at Asn in green was not foreseen but found experimentally.



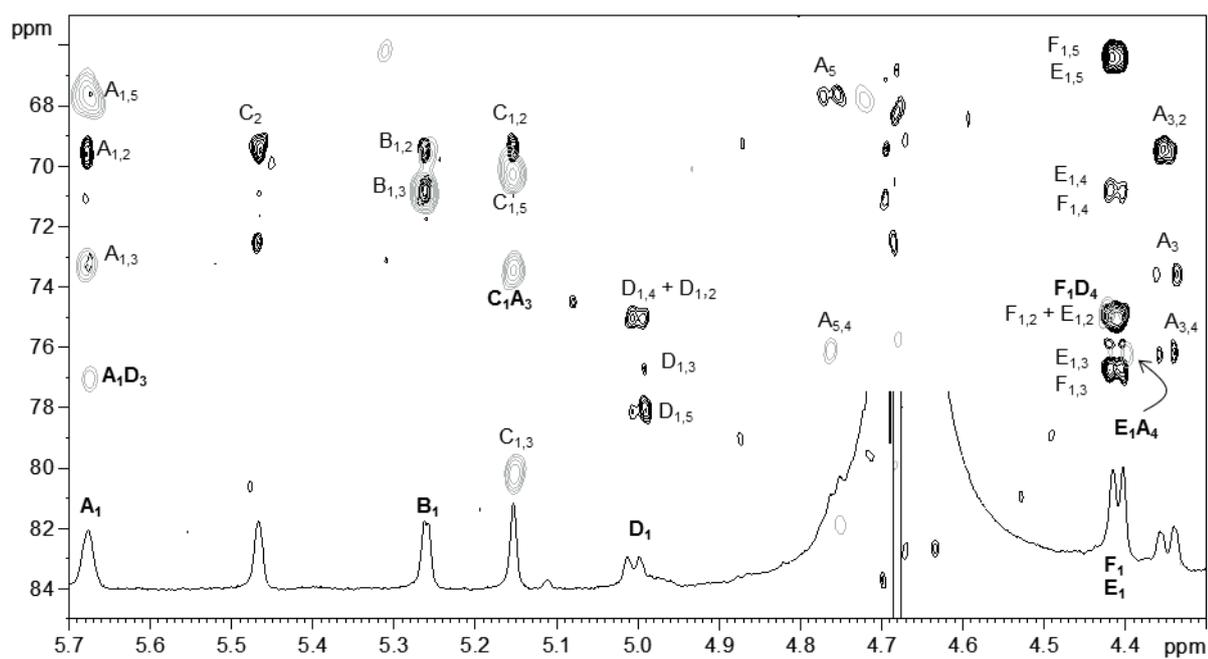
**Figure S2.** GC-MS chromatograms of two representative Chloroviruses. A) permethylated 2-(-)-octylglycosides from CVM-1. For ACTV-1, TN603 and MT325 similar profiles were obtained. B) Acetylated 2-(-)-octylglycosides from NY2A<sub>1</sub> oligosaccharide. \* impurities.



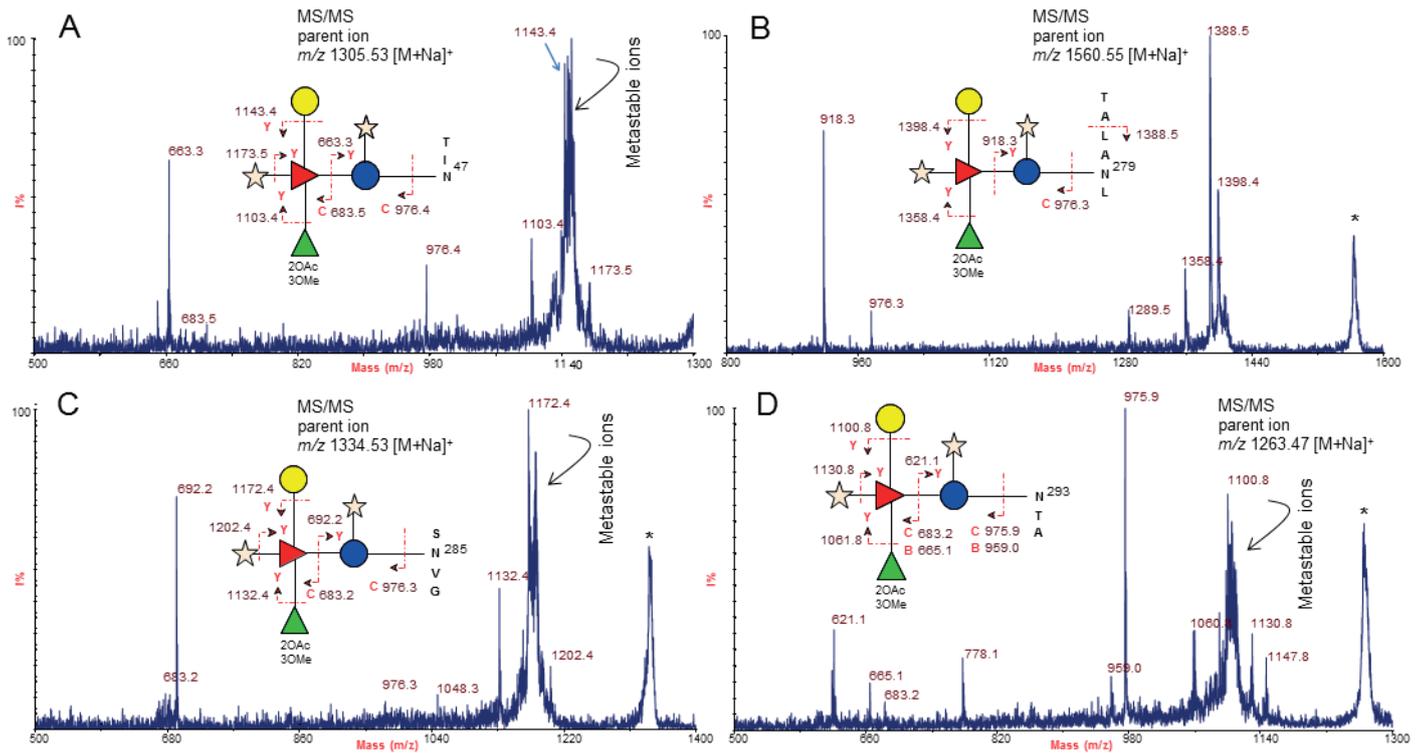
**Figure S3.** Expansion of CVM-1 glycopeptide HSQC spectrum measured at 600 MHz, 310 K, HMBC spectrum is reported in the inset at the right-bottom corner. Correlations attribution follows the letter system of Table S3 and Figure 1; those appearing in gray in the HSQC spectrum have the opposite sign with respect to the other and represent carbons bearing two hydrogen atoms. Signals crossed are related to the peptide moiety of the glycopeptide or to impurities.



**Figure S4.** Superimposition of TOCSY (black) and COSY (turquoise/red) spectra of CVM-1 glycopeptide measured at 600 MHz, 310 K. Correlations attribution follows the letter system of Table S3 and Figure 1.



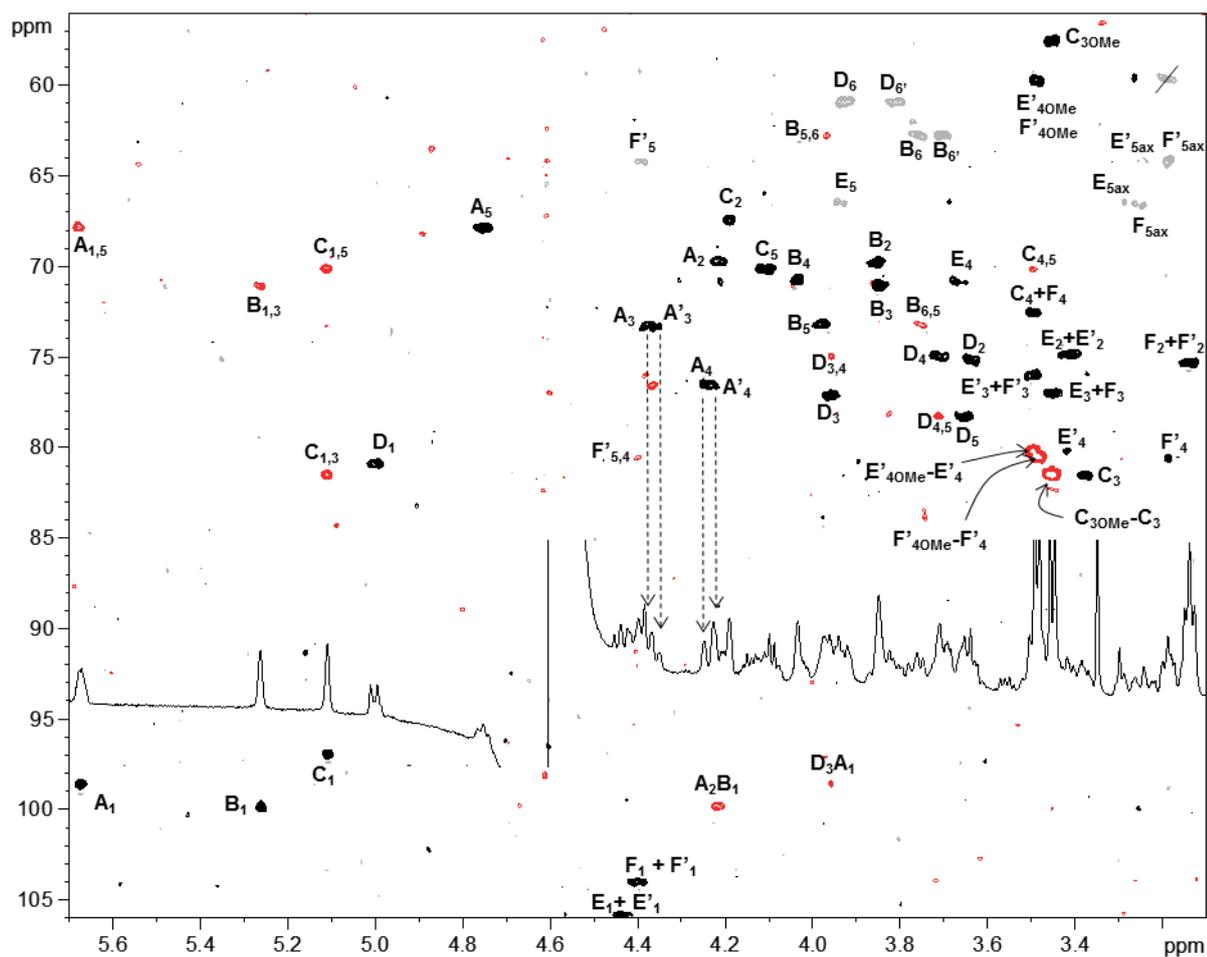
**Figure S5.** Superimposition of CVM-1 glycopeptide HSQCTOCSY (black) and HMBC (grey) spectra measured at 600 MHz, 310 K. Correlations attribution follows the letter system of Table S3 and Figure 1.



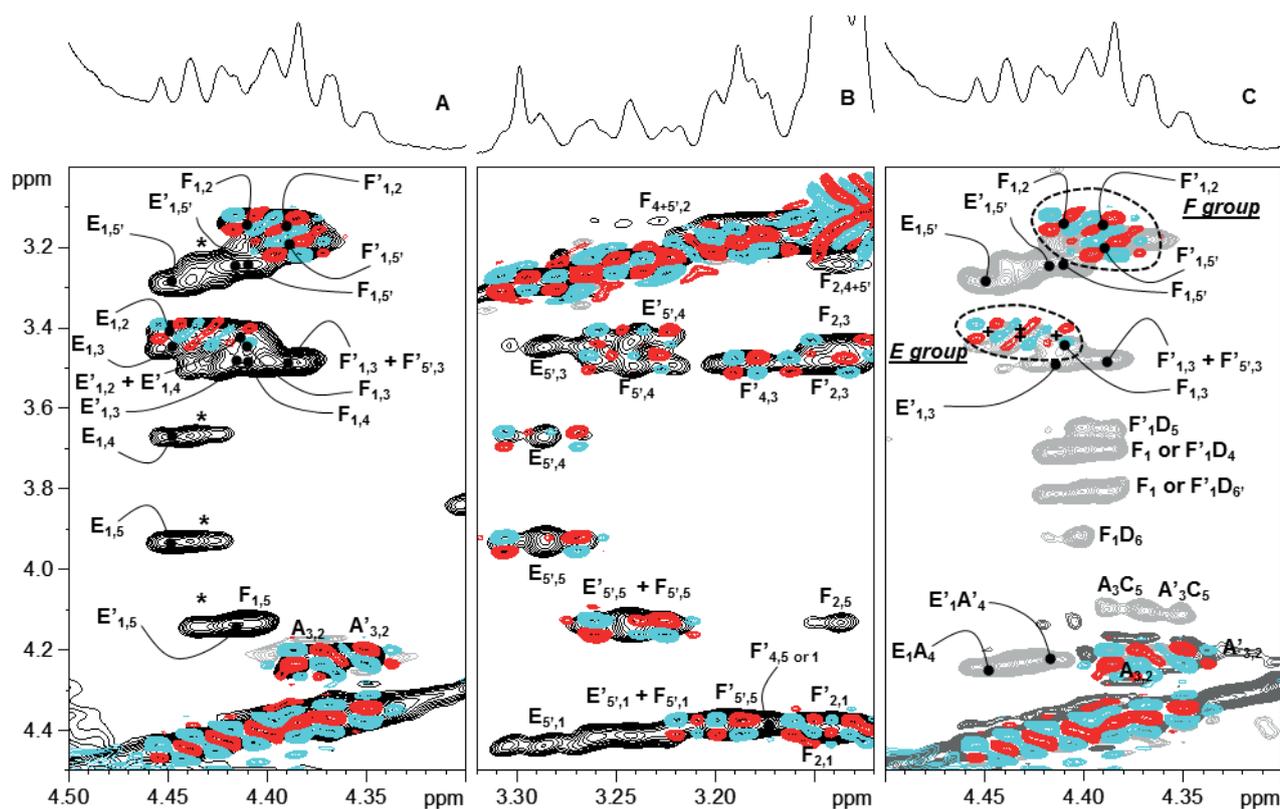
**Figure S6. MALDI MS analysis of CVM-1 MCP glycopeptide fraction.**

MALDI-TOF/TOF MS/MS analysis (A-D) of the most significant *pseudo*-molecular ions from CVM-1 MCP glycopeptide fraction, showing four possible glycosylation sites at Asn-47, -279, -285, -293, all occupied by a unique hexasaccharide glycoform.

Asterisks denote parent ions.

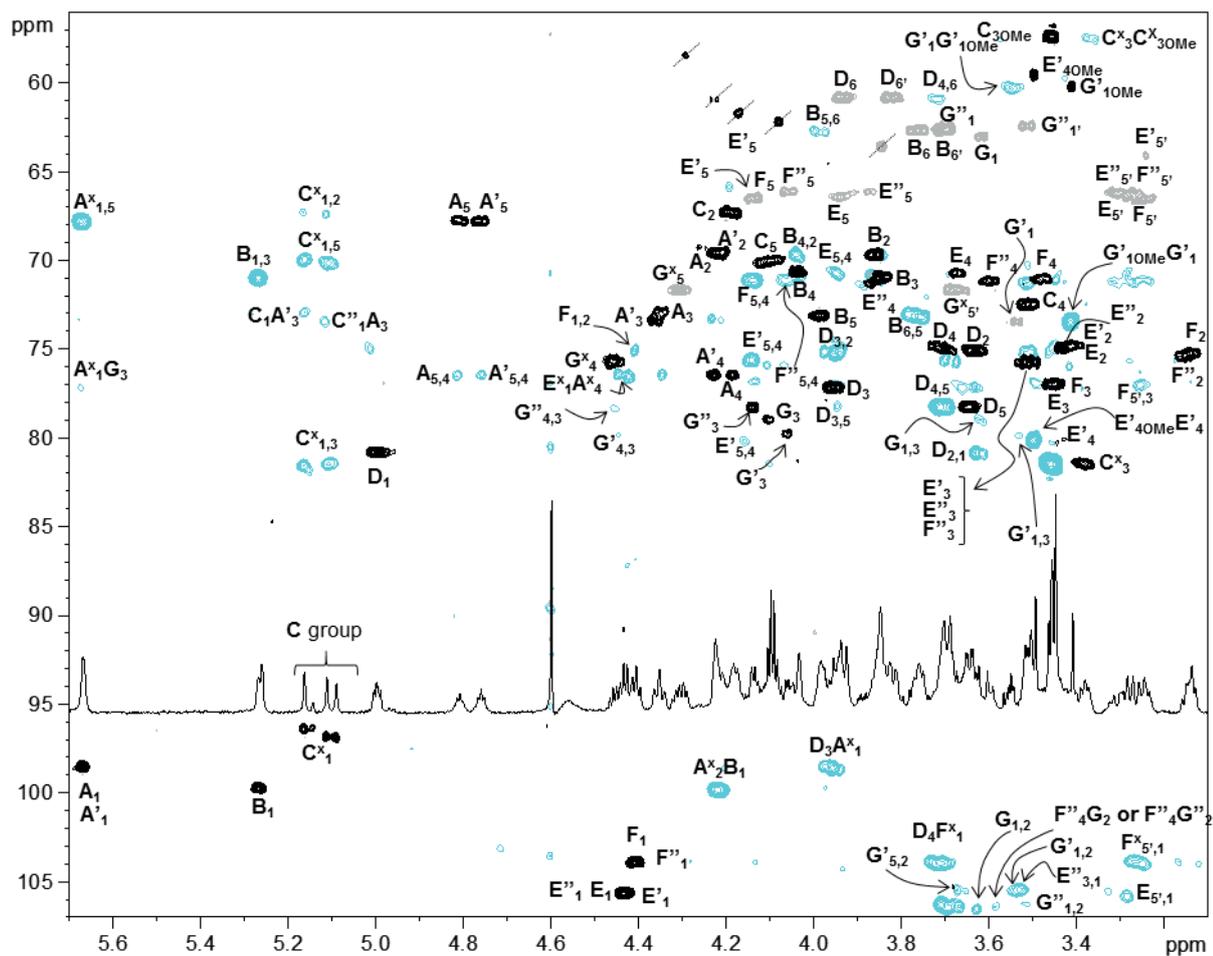


**Figure S7.** Superimposition of ATCV-1 glycopeptide HSQC (black/grey) and HMBC (red) spectra measured at 600 MHz, 318 K. Correlations attribution follows the letter system of Table S4 and Figure 1; those appearing in gray in the HSQC spectrum have the opposite sign with respect to the other and represent carbons bearing two hydrogen atoms. Signals crossed are related to the peptide moiety of the glycopeptide or to impurities.

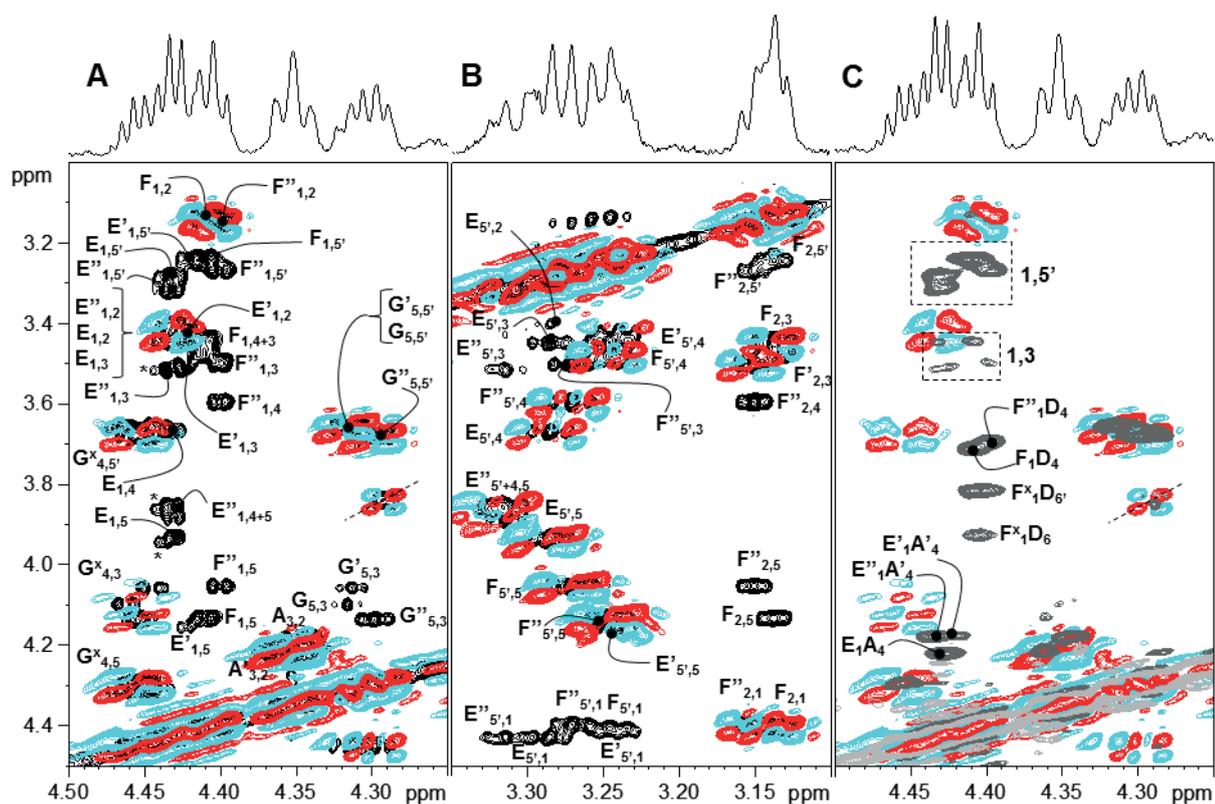


**Figure S8.** Homonuclear spectra measured for ATCV-1 glycopeptides at 600 MHz and 318 K. A) Superimposition of TOCSY (black) and COSY (turquoise/red) spectra detailing the anomeric region of the xylose residues. B) As A) but detailing the region of H-5' signals of distal and proximal xylose units. C) Superimposition of TROESY (grey) and COSY (turquoise/red) spectra detailing the anomeric region of xylose residues. Correlations attribution follows the letter system of Table S4 and Figure 1.

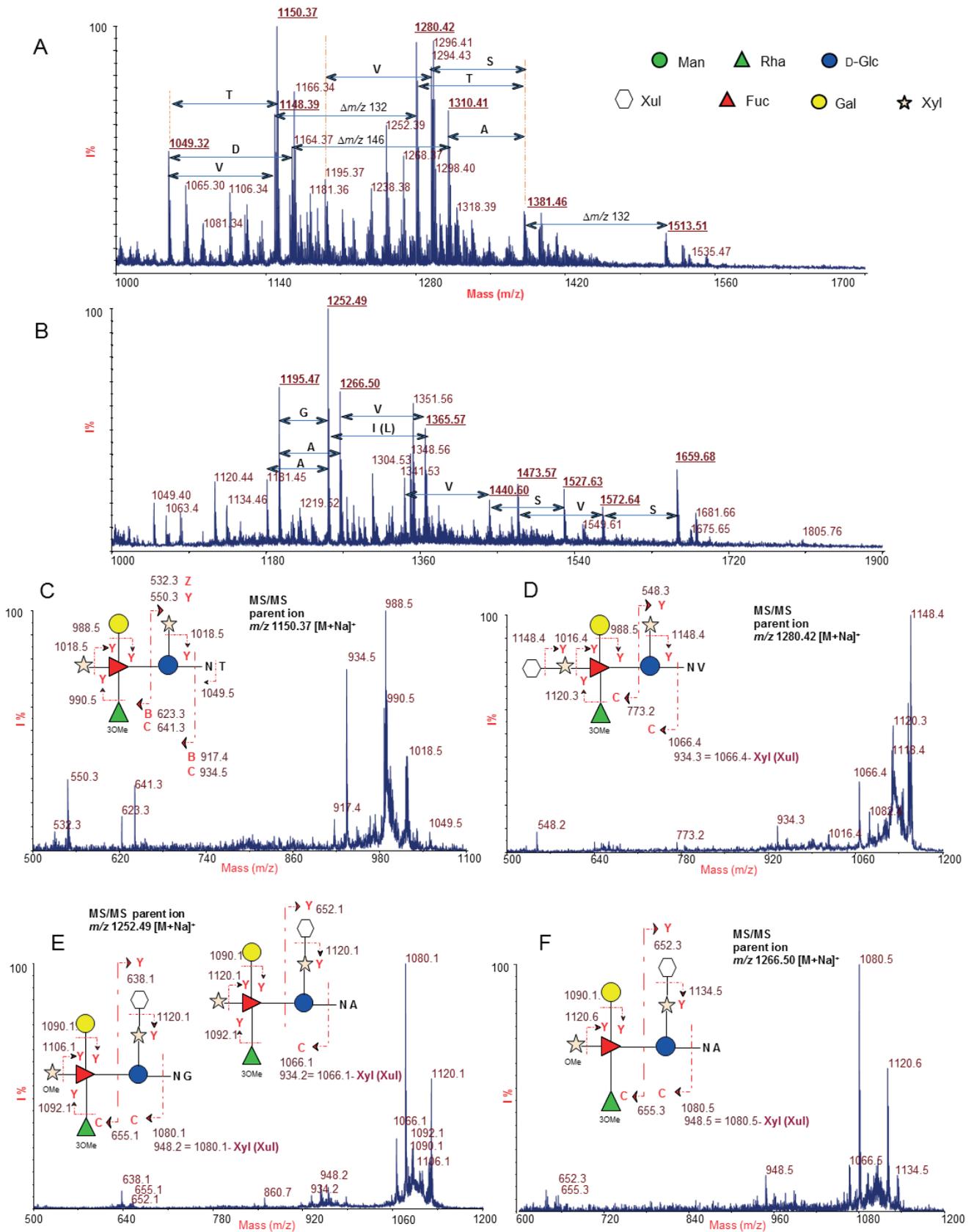




**Figure S10.** Superimposition of TN603 glycopeptide HSQC (black/grey) and HMBC (turquoise) spectra measured at 600 MHz, 318 K. Correlations attribution follows the letter system of Table S5 and Figure 1; those appearing in gray in the HSQC spectrum have the opposite sign with respect to the other and represent carbons bearing two hydrogen atoms. The upscript “x” applied to a label, as C, implies that that density belongs to one of the residues of that group without specifying which one. Signals crossed are related to the peptide moiety of the glycopeptide or to impurities.

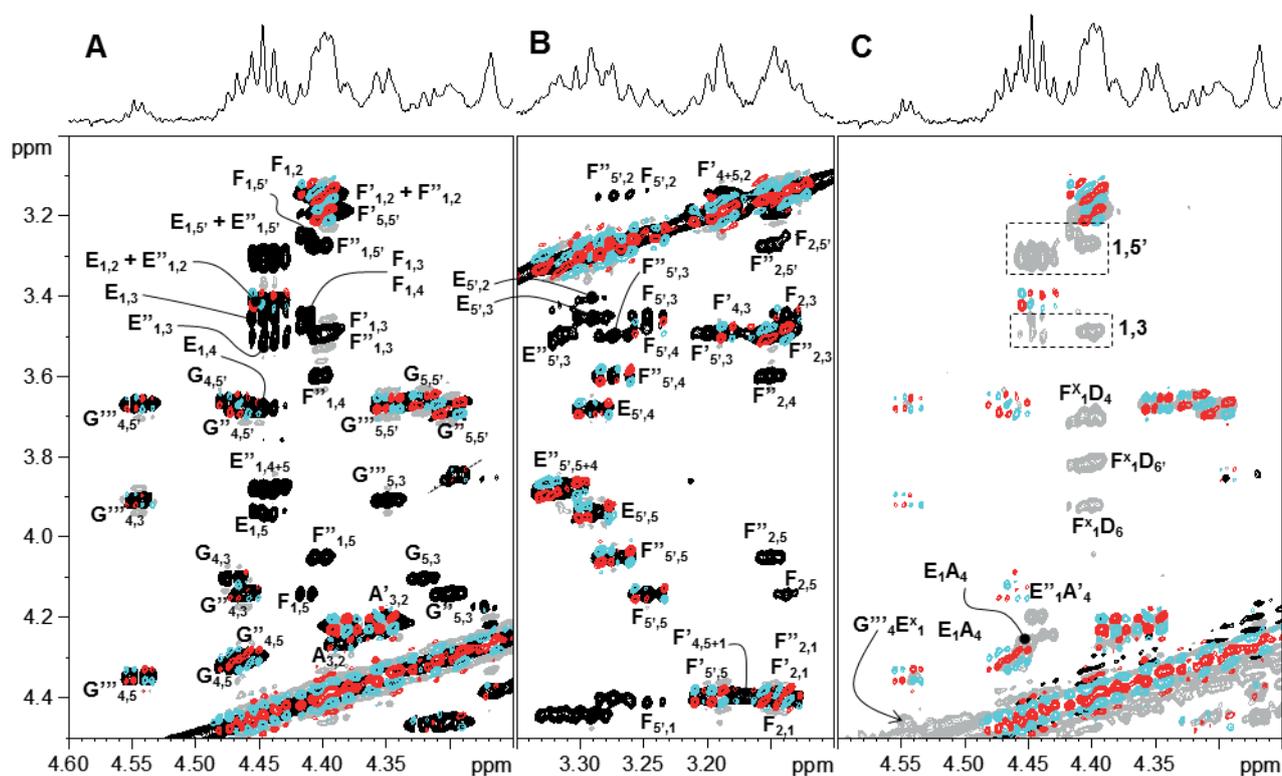


**Figure S11.** Homonuclear spectra measured for TN603 glycopeptides at 950 MHz and 318 K. A) Superimposition of TOCSY (black) and COSY (turquoise/red) spectra detailing the anomeric region of the xylose residues. B) As in A but detailing the region of H-5' signals of distal and proximal xylose units. C) Superimposition of NOESY (grey) and COSY (turquoise/red) spectra detailing the anomeric region of xylose residues. Correlations attribution follows the letter system of Table S5 and Figure 1.

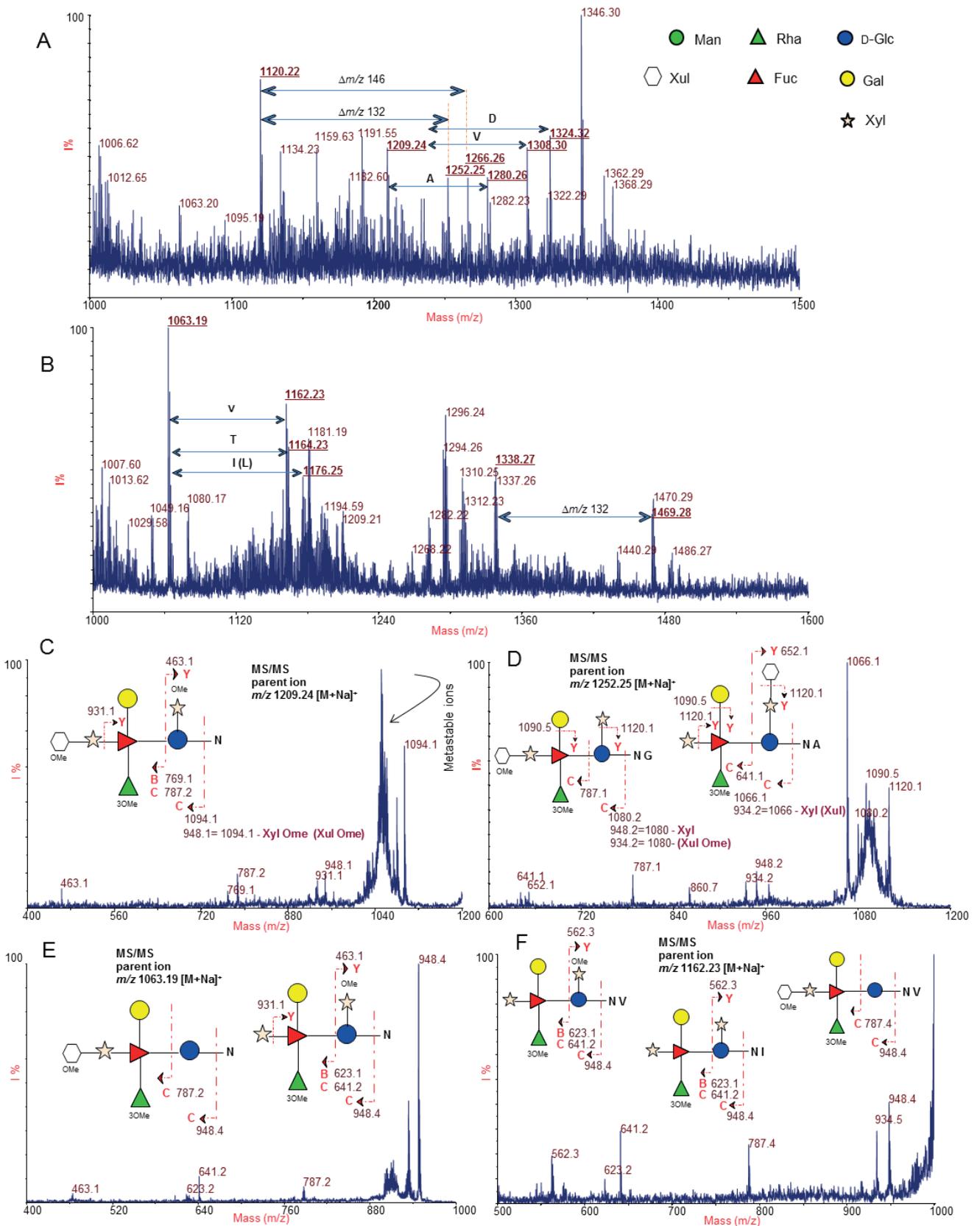


**Figure S12. MALDI MS and MS/MS analysis of TN603 MCP glycopeptide fractions.**  
 (A-B) MALDI mass spectra of the glycopeptide fractions from TN603 MCP. (C-F) MS/MS analysis of selected ions at  $m/z$  1150.37, 1280.42 1252.49 and 1266.50, respectively.

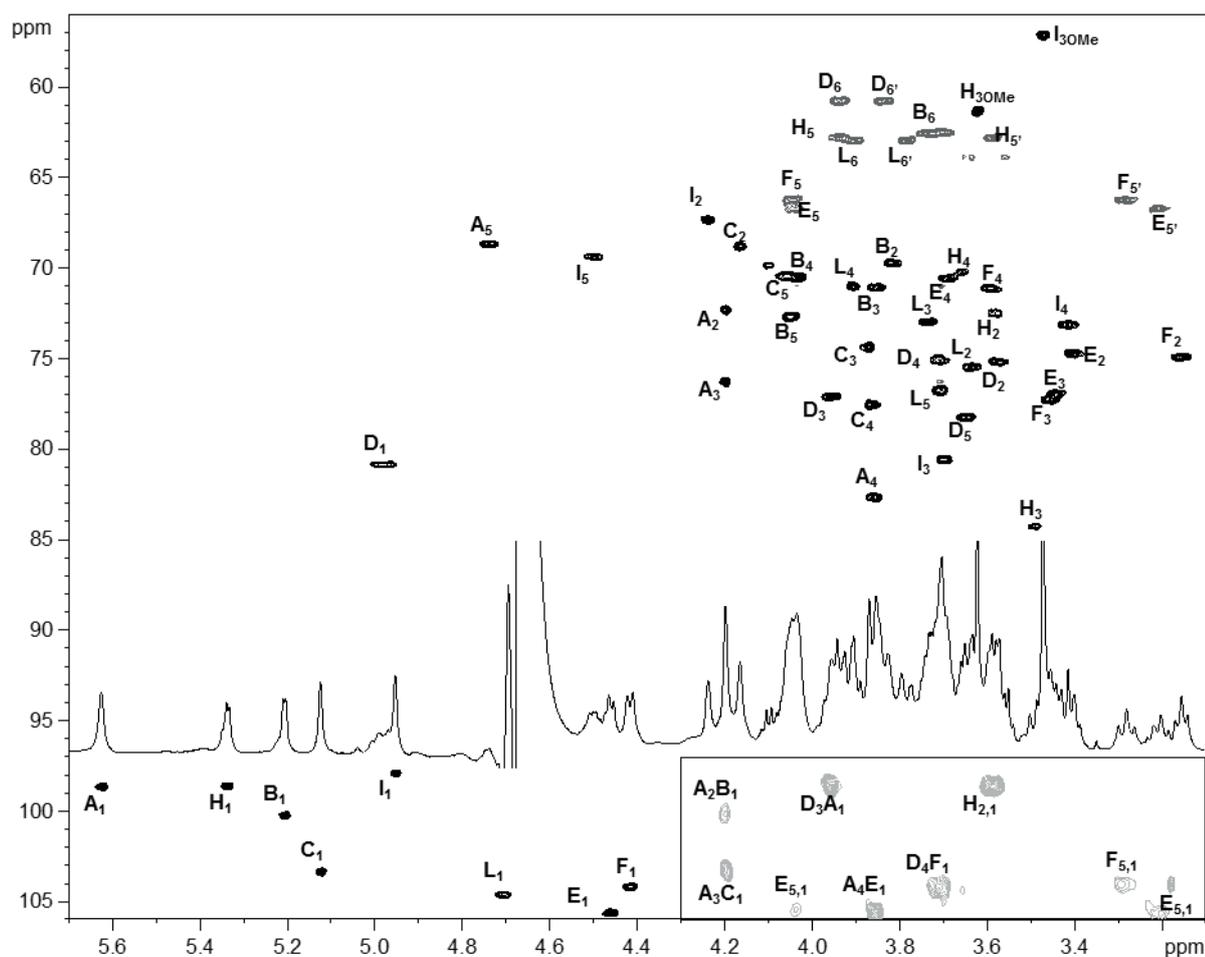




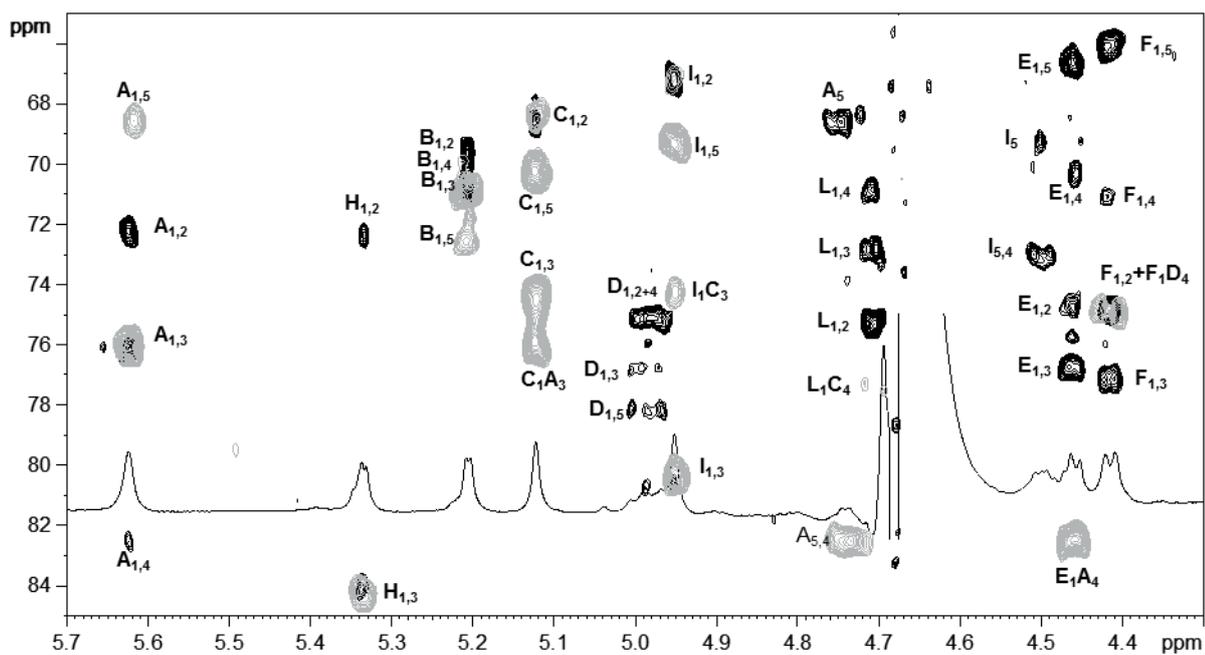
**Figure S14.** Homonuclear spectra measured for MT325 glycopeptides at 950 MHz and 310 K. A) Superimposition of TOCSY (black) and COSY (turquoise/red) spectra detailing the anomeric region of the xylose residues. B) As in A but detailing the region of H-5' signals of distal and proximal xylose units. C) Superimposition of TROESY (grey) and COSY (turquoise/red) spectra detailing the anomeric region of xylose residues. Correlations attribution follows the letter system of Table S6 and Figure 1.



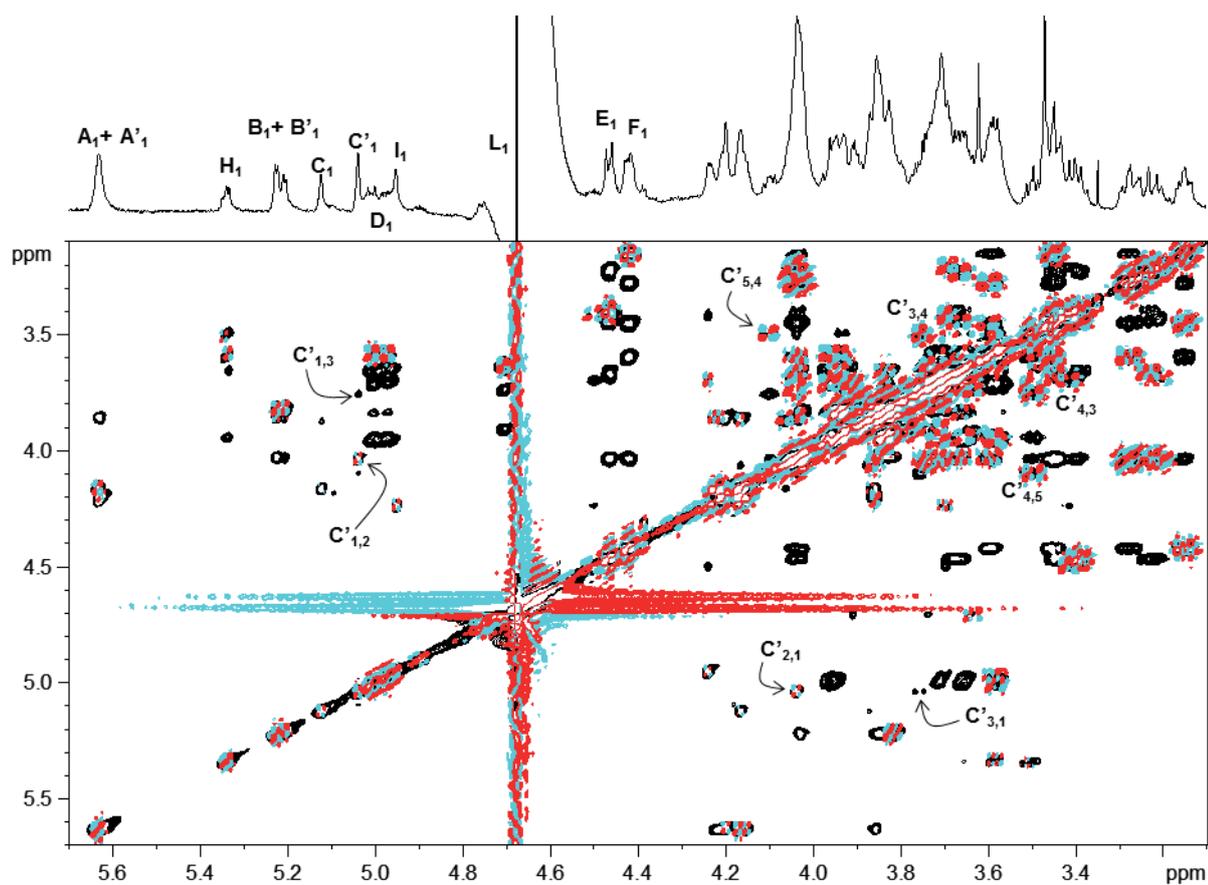
**Figure S15. MALDI MS and MS/MS analysis of MT325 MCP glycopeptide fractions.** (A-B) MALDI mass spectra of the glycopeptide fractions from MT325 MCP. (C-F) MS/MS analysis of selected ions at  $m/z$  1209.24, 1252.25, 1063.19 and 1266.50, respectively.



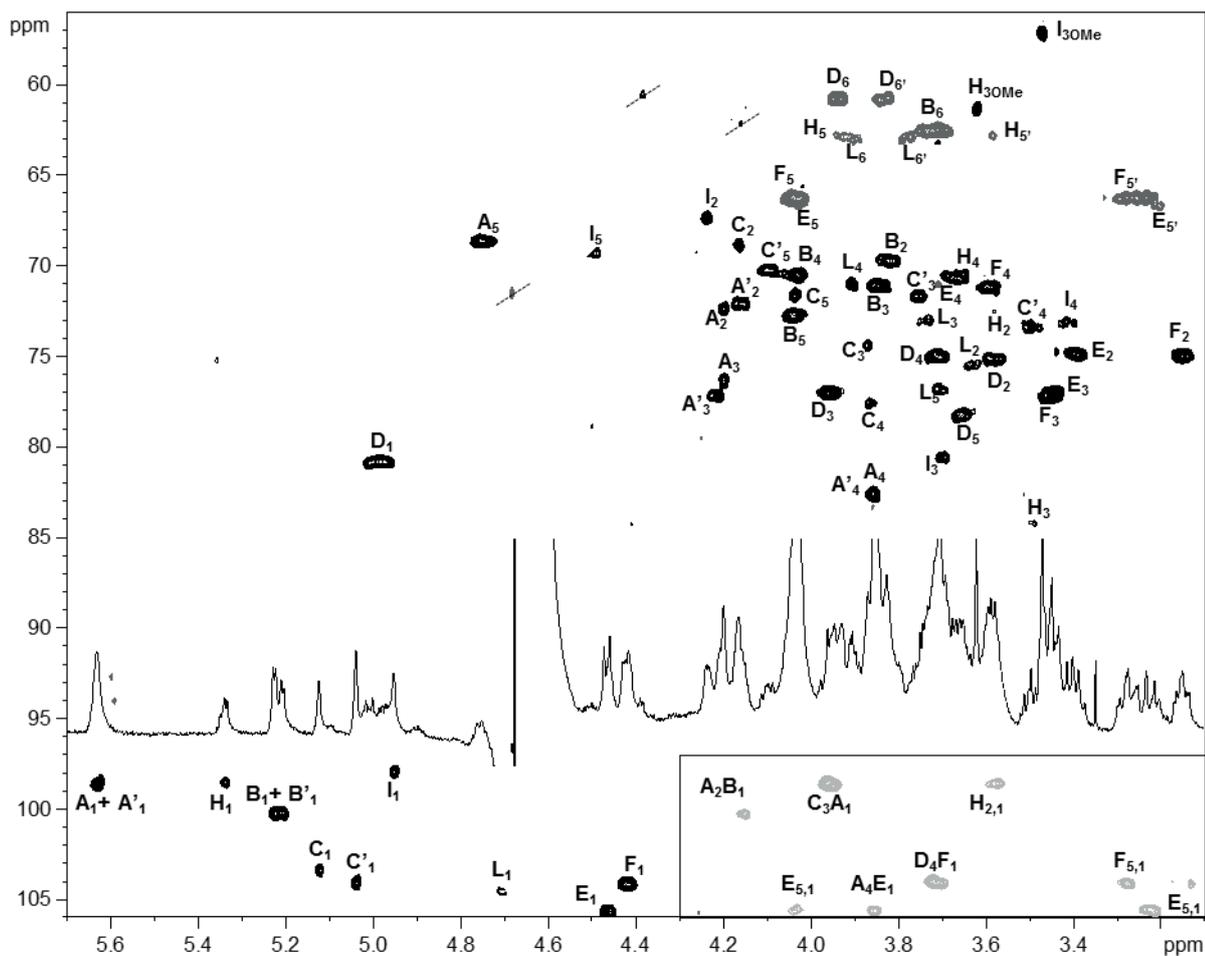
**Figure S16.** Expansion of NY-2A<sub>1</sub> glycopeptide HSQC spectrum measured at 600 MHz, 310 K, HMBC spectrum is reported in the inset at the right-bottom corner. Correlations attribution follows the letter system of Table S7 and Figure 1; those appearing in gray in the HSQC spectrum have the opposite sign with respect to the other and represent carbons bearing two hydrogen atoms. Signals crossed are related to the peptide moiety of the glycopeptide or to impurities.



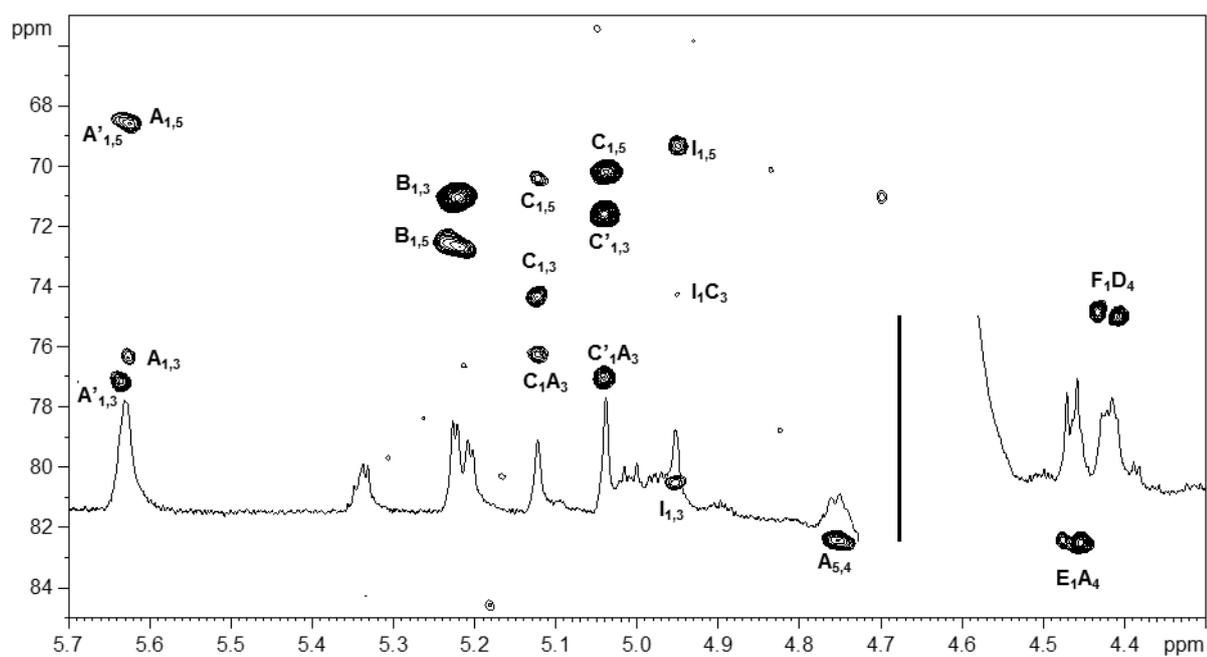
**Figure S17.** Superimposition of NY-2A<sub>1</sub> glycopeptide HSQC (black) and HMBC (grey) spectra measured at 600 MHz, 310 K. Correlations attribution follows the letter system of Table S7 and Figure 1.



**Figure S18.** Superimposition of TOCSY (black) and COSY (turquoise/red) spectra of NY-2A<sub>2</sub> glycopeptide measured at 600 MHz, 310 K. Correlations attribution follows the letter system of Table S8 and Figure 1.



**Figure S19.** Expansion of NY-2A<sub>2</sub> glycopeptide HSQC spectrum measured at 600 MHz, 310 K, HMBC spectrum is reported in the inset at the right-bottom corner. Correlations attribution follows the letter system of Table S8 and Figure 1; those appearing in gray in the HSQC spectrum have the opposite sign with respect to the other and represent carbons bearing two hydrogen atoms. Signals crossed are related to the peptide moiety of the glycopeptide or to impurities.



**Figure S20.** HMBC spectrum of NY-2A<sub>2</sub> glycopeptide measured at 600 MHz, 310 K. Correlations attribution follows the letter system of Table S8 and Figure 1.

