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

Cristina De Castro: decastro@unina.it

^aDepartment of Agricultural Sciences, University of Napoli Via Università 100, 80055 Portici (NA), Italy

^bDepartment of Chemical Sciences, University of Napoli Via Cintia 4, 80126 Napoli, Italy

^cDepartment of Biology, Nebraska Wesleyan University, Lincoln, NE, USA

^dDepartment of Plant Pathology and Nebraska Center for Virology University of Nebraska Lincoln, NE 68583-0900 USA

^eCNR, Institute for Polymers, Composites and Biomaterials Via P. Gaifami 18, 95126 Catania, Italy

^fDepartment 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 β -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,^[1] and also in many viruses, such as rhabdoviruses, herpesviruses, poxviruses, and paramyxoviruses.^[2] 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^[3] 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).^[4] 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.^[5] We have recently shown that Vp54, the MCP of *Paramecium bursaria* chlorella virus (PBCV-1) contains four *N*-linked glycans,^[6] with structures that do not resemble any other reported for Bacteria, Archaea, and Eukarya,^[7] including unicellular algae.^[8]

The PBCV-1 Vp54 glycans are attached to the protein by an Asn β -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).^[7] 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).^[7] Each virus gave one glycopeptide (¹H NMR in Figure 3), except NY-2A for which two different fractions were isolated, NY-2A₁ and NY-2A₂.

Monosaccharide chemical analysis (Table S2)^[7] 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 α -fucose (**A**) at *O*-3, and a β -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 α -galactose **B** at *O*-2, α -rhamnose **C** at *O*-3, and finally the distal β -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,^[7] and such variation fitted with the rules reported by Shashkov et al.^[9] 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 α -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-2A₁ (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 α -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 α -L-3OMeRha (**I**) and α -D-Gal (**L**), in turn further substituted at *O*-2 with a α -D-3OMeXyl residue (**H**). Taken together, this information defined NY-2A₁ glycopeptide as a highly branched nonasaccharide (Figure 1).

NY-2A₂ NMR spectra (Table S8, NMR in Figures 3, S18–S20) contained all the signals assigned in NY-2A₁ along with a new anomeric signal at 5.04 ppm, labelled **C'**. This new residue was a non-substituted α -rhamnose linked at *O*-3 of the hyperbranched fucose. Thus, NY-2A₂ 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,^[11] 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.; Schachter, 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.

^1H 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.

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 ^a [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	²⁸⁰ NIPG; ³⁰² NTGT; ³⁹⁹ NTET; ⁴⁰⁶ NTAT
NY-2A (B585L)[d]	96	48359	52641, 51250	4282 (8.9) 2891 (5.9)	1348, 880	4 or 3[e]	⁵⁴ NKVS; ¹⁷¹ 280NIPG; ¹⁸¹ 302NTGT; ¹⁹¹ 399NTET; ¹⁹¹ 406NTAT[g]
NY-2A (B617L) [d]	94	48354	As above	As above	As above	As above	As above and ²⁹¹ NVAT[h]
MT325 (M463L)	78	47652	ND	ND	1186	ND	ND
CVM-1 (548L)	79	47622	51247	3625 (7.5)	936	4	⁴⁷ NGSV; ¹⁷¹ 279NLTA; ¹⁸¹ 285NVGY; ¹⁹¹ 293NTAV [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 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.