

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Virology Papers

Virology, Nebraska Center for

2021

Chlorovirus PBCV-1 Multidomain Protein A111/114R Has Three Glycosyltransferase Functions Involved in the Synthesis of Atypical N-Glycans

Eric Noel

Anna Notaro

Immacolata Speciale

Garry A. Duncan

Cristina De Castro

See next page for additional authors

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



Part of the [Biological Phenomena](#), [Cell Phenomena](#), and [Immunity Commons](#), [Cell and Developmental Biology Commons](#), [Genetics and Genomics Commons](#), [Infectious Disease Commons](#), [Medical Immunology Commons](#), [Medical Pathology Commons](#), and the [Virology Commons](#)






This Article is brought to you for free and open access by the Virology, Nebraska Center for at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Virology Papers by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Authors

Eric Noel, Anna Notaro, Immacolata Speciale, Garry A. Duncan, Cristina De Castro, and James L. Van Etten

Article

Chlorovirus PBCV-1 Multidomain Protein A111/114R Has Three Glycosyltransferase Functions Involved in the Synthesis of Atypical N-Glycans

Eric Noel ^{1,2,†} , Anna Notaro ^{3,4,†} , Immacolata Speciale ^{3,5,†} , Garry A. Duncan ¹, Cristina De Castro ^{5,*}  and James L. Van Etten ^{1,6,*} 

¹ Nebraska Center for Virology, University of Nebraska, Lincoln, NE 68583-0900, USA; enoel@huskers.unl.edu (E.N.); gduncan2@unl.edu (G.A.D.)

² School of Biological Sciences, University of Nebraska, Lincoln, NE 68588-0118, USA

³ Department of Chemical Sciences, University of Napoli Federico II, Via Cintia 4, 80126 Napoli, Italy; anna.notaro@unina.it (A.N.); immacolata.speciale@unina.it (I.S.)

⁴ Information Génomique et Structurale, Centre National de la Recherche Scientifique, Aix-Marseille Université, UMR 7256, IMM FR3479, Institut de Microbiologie (FR3479), CEDEX 9, 13288 Marseille, France

⁵ Department of Agricultural Sciences, University of Napoli Federico II, Via Università 100, 80055 Portici, Italy

⁶ Department of Plant Pathology, University of Nebraska, Lincoln, NE 68583-0722, USA

* Correspondence: decastro@unina.it (C.D.C.); jvanetten1@unl.edu (J.L.V.E.)

† These authors contributed equally to this work.

Abstract: The structures of the four N-linked glycans from the prototype chlorovirus PBCV-1 major capsid protein do not resemble any other glycans in the three domains of life. All known chloroviruses and antigenic variants (or mutants) share a unique conserved central glycan core consisting of five sugars, except for antigenic mutant virus PIL6, which has four of the five sugars. A combination of genetic and structural analyses indicates that the protein coded by PBCV-1 gene *a111/114r*, conserved in all chloroviruses, is a glycosyltransferase with three putative domains of approximately 300 amino acids each. Here, in addition to in silico sequence analysis and protein modeling, we measured the hydrolytic activity of protein A111/114R. The results suggest that domain 1 is a galactosyltransferase, domain 2 is a xylosyltransferase and domain 3 is a fucosyltransferase. Thus, A111/114R is the protein likely responsible for the attachment of three of the five conserved residues of the core region of this complex glycan, and, if biochemically corroborated, it would be the second three-domain protein coded by PBCV-1 that is involved in glycan synthesis. Importantly, these findings provide additional support that the chloroviruses do not use the canonical host endoplasmic reticulum–Golgi glycosylation pathway to glycosylate their glycoproteins; instead, they perform glycosylation independent of cellular organelles using virus-encoded enzymes.

Keywords: glycosyltransferases; multidomain protein; N-glycan; chloroviruses; PBCV-1



Citation: Noel, E.; Notaro, A.; Speciale, I.; Duncan, G.A.; De Castro, C.; Van Etten, J.L. Chlorovirus PBCV-1 Multidomain Protein A111/114R Has Three Glycosyltransferase Functions Involved in the Synthesis of Atypical N-Glycans. *Viruses* **2021**, *13*, 87. <https://doi.org/10.3390/v13010087>

Academic Editors: Jacques Le Pendu and Göran Larson

Received: 9 December 2020

Accepted: 8 January 2021

Published: 10 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Glycosylation is an important post-translational modification that confers a diversity of structures and functions in both cell and virus biology. One of the most common forms of protein modification is N-linked glycosylation, in which a high mannose core is linked to the amide nitrogen of asparagine in the context of the conserved motif Asn–X–Ser/Thr. This attachment occurs early in protein synthesis, followed by a complex process of trimming and remodeling of the oligosaccharide during transit through the endoplasmic reticulum (ER) and Golgi [1] producing glycoproteins with varying oligosaccharide structures. Typically, viruses use host-encoded glycosyltransferases (GT) and glycosidases located in the ER and Golgi to add and remove N-linked sugar residues to/from virus glycoproteins either co-translationally or shortly after translation of the protein. The viral glycoproteins are then often transported to the host plasma-membrane where progeny

viruses bud through the membrane, and so the viruses only become infectious as they leave the cell [2].

However, the large dsDNA chloroviruses in the family *Phycodnaviridae* are interesting because, unlike other viruses, they encode most, if not all, of the machinery to glycosylate their major capsid proteins (MCP) independent of the cellular organelles [3,4]. Furthermore, the process occurs in the cytoplasm, and infectious viruses are formed inside the cell prior to cell lysis. The prototype chlorovirus, *Paramecium bursaria chlorella virus* (PBCV-1), has a 330-kb genome that is predicted to encode as many as 400 proteins, many of which are unusual for a virus [5], including at least 17 related to various aspects of carbohydrate metabolism [6].

The PBCV-1 MCP (also referred to as Vp54) is coded by gene *a430l*, and the protein has four glycosylation sites [7]. The predominant oligosaccharide is a nonasaccharide (Figure 1) with several unique features, including the following: (i) its structure does not resemble any structure previously reported in the three domains of life [8]; (ii) the four glycoforms are generated by the non-stoichiometric presence of two monosaccharides, L-arabinose (Araf) and D-mannose (Man); (iii) the most abundant glycoform consists of nine neutral monosaccharide residues organized in a highly-branched fashion; (iv) none of the N-linked glycans is attached to a typical Asn–X–(Thr/Ser) consensus site in the protein [9]; (v) the glycans are attached to the protein by a β -glucose (Glc) linkage, which is rare in nature and has only been reported in glycoproteins from a few organisms [10–13]; and (vi) the glycoform contains a dimethylated rhamnose (Rha) as the capping residue of the main chain, a hyper-branched fucose (Fuc) residue and two Rha residues with opposite configurations.

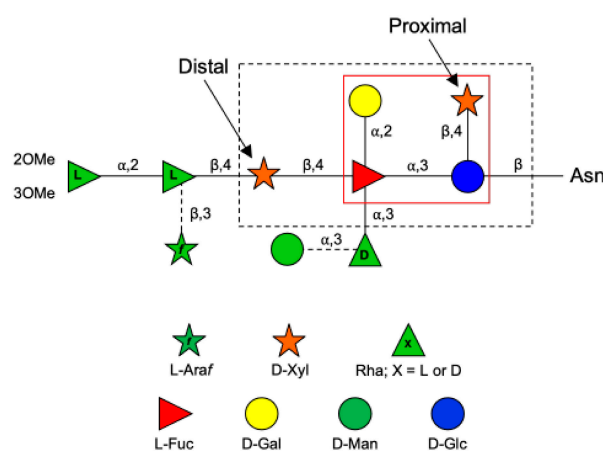


Figure 1. Structure of the N-glycan attached to the chlorovirus PBCV-1 major capsid protein (Vp54). Monosaccharides (Man and Araf) connected by dashed lines are non-stoichiometric substituents. The larger black box encloses the conserved pentasaccharide core structure common to all chloroviruses analyzed to date. The inner red box indicates the tetrasaccharide structure of the N-glycan of the antigenic mutant chlorovirus PIL6 [6].

Interestingly, all the chloroviruses studied to date, including those with different host specificities, have the capsid protein N-glycosylated with other types of oligosaccharides; however, all chloroviruses share the same pentasaccharide core oligosaccharide [14,15] composed of an N-linked Glc, a hyperbranched Fuc, a distal and a proximal xylose (Xyl), and a galactose (Gal) (Figure 1). Additional monosaccharides decorate this core N-glycan, producing a molecular signature for each chlorovirus [16]. This oligosaccharide core was also found in all mutants (or antigenic variants) of PBCV-1 analyzed to date, except for PIL6 (Figure 1 [6]). Mutant PIL6 is a representative of the antigenic class D, characterized by a large genomic deletion spanning genes *a014r* through *a078r* [6,17]. Its N-glycan is a tetrasaccharide and is significantly truncated compared to that of wild-type PBCV-1. It contains all the units of the chloroviruses N-glycans present in the conserved core region

except the distal Xyl attached to Fuc [16]. By analyzing the genomes of all chloroviruses and of the mutants sequenced to date, we noted that the *a111/114r* gene is the only annotated orthologous GT gene found outside of the region of the large deletion mutants that is present in all other chloroviruses. This finding suggests its likely involvement in the synthesis of the initial part of the unusual N-glycan shared by all of these viruses and prompted us to investigate its role in the assemblage of the conserved core oligosaccharide.

To predict the A111/114R protein structure and functions, we used a combination of genetic and structural analyses, together with hydrolytic activity assays and in silico evidence by sequence analysis and protein modeling. The combined results suggest that the A111/114R protein is a multi-domain/multi-functional GT likely involved in the attachment of three of the five monosaccharides in the conserved core region of the N-glycan (Figure 1). Specifically, this large protein of 860 amino acids is comprised of three putative domains (Figure S1; Figure 2), each with a specific role; the N-terminal domain (1–260 aa) is a galactosyltransferase (GalT), the central domain (261–559 aa) is a xylosyltransferase (XylT), and the C-terminal domain (560–860 aa) is a fucosyltransferase (FucT).

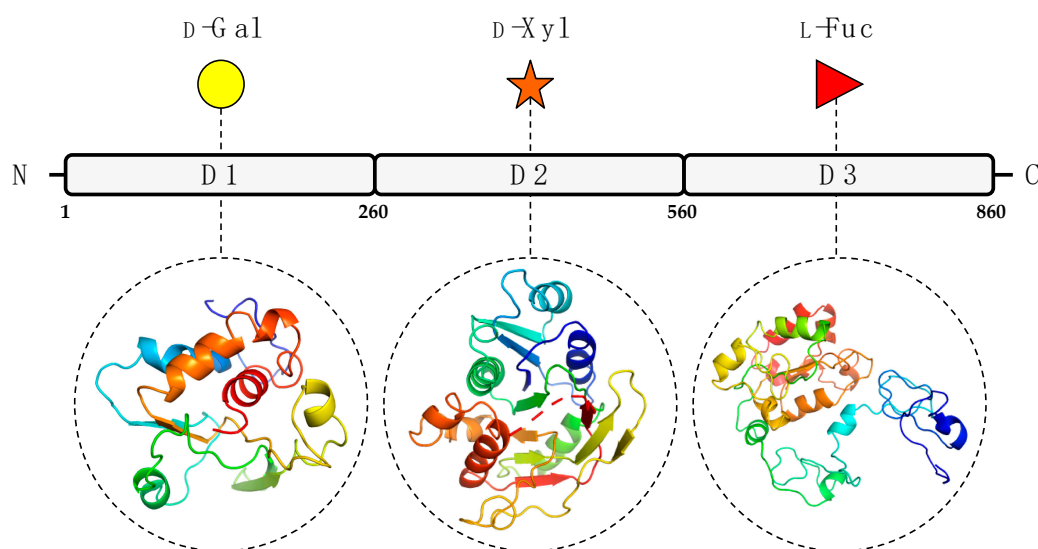


Figure 2. Predicted GT domains of PBCV-1 encoded protein A111/114R. A111/114R domain analysis based on remote homology identified three putative GT domains labeled as D1, D2, and D3, located at the N-terminal (1 to 260 aa), central (261 to 559 aa), and C-terminal (560 to 860 aa) regions, respectively. Below individual domains are the corresponding three-dimensional protein models assigned by Phyre2 [18] based on alignments to known protein structures identified by their PDB entry: 1GA8 Chain A (D1), 2Z86 Chain D (D2), 2NZY Chain A (D3). Protein ribbon models are rendered using rainbow colors from N-terminus (blue) to C-terminus (red). The putative domain, predicted protein model, and sugar substrate are connected by the black dashes.

2. Materials and Methods

2.1. Protein Modeling

The prediction of the different domains of A111/114R was performed by HHpred [19] based on remote homology detection. Then, the 3D model of each domain was built by Phyre2 (Protein Homology/analogy Recognition Engine V 2.0) in a normal mode [18]. Each 3D model was based on an alignment generated by HMM–HMM matching. Final drawings and residue analysis were prepared with the molecular graphics system PyMol Version 1.2r3pre, Schrödinger, LLC.

2.2. Cloning and Expression

PBCV-1 *a111/114r* and domain variants were cloned from PCR-amplified viral DNA using oligonucleotide primers with restriction sites NotI–BamHI. PCR fragments of the expected size were digested and inserted into the restriction sites of the pMAL-c6T expression vector (New England Biolabs, Ipswich, MA, USA). This process produced a

maltose-binding protein (MBP) tag at the N-terminus of the target protein. The resulting plasmid was transformed into *E. coli* strain One Shot TOP10 competent cells (Invitrogen) for maintenance. The *E. coli* cells containing positive cloned plasmids were selected with 100 µg/mL carbenicillin. The cloned structure of each vector was sequence verified. Plasmids were isolated with a QIAprep Spin Miniprep kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions and transformed into NEBExpress competent cells for expression. Viral genes were expressed by growing cells overnight at 37 °C in 10 mL of LB medium (10 g/L tryptone, 5 g/L yeast extract, 5 g/L sodium chloride) containing 100 µg/mL carbenicillin. Then, 5 mL of the over-night culture was sub-cultured into 200 mL LB medium containing 100 µg/mL carbenicillin. The batch culture was grown to an OD₆₀₀ of 0.6 at 37 °C and then induced with 0.1 mM IPTG and incubated at 16 °C overnight. The cells were harvested by centrifugation at 3500 × g, for 5 min at 4 °C, and resuspended in 35 mL of PBS with 2 mM phenylmethylsulfonyl fluoride (PMSF). After incubation on ice for 30 min, cells were disrupted by sonication for 3 min using a Tekmar sonic disruptor at 30% amplitude, in 5 s pulses. Samples were centrifuged at 10,000 rpm for 15 min to separate soluble and insoluble fractions.

2.3. Purification of Recombinant Enzymes

Amylose resin (New England Biolabs) was loaded onto a 5-mL self-packing column with a 45- to 90-µm-pore-size polyethylene filter (frit) (Life Science Products, Chestertown, MD, USA), and the resin was allowed to settle. The column was equilibrated with 5 column volumes of cold wash buffer (50 mM NaH₂PO₄, 150 mM NaCl, 1 mM DTT, 1 mM EDTA, pH 7.2). The soluble bacterial fraction was applied to the column and allowed to drain. The column was washed again with 5 column volumes of cold wash buffer. The recombinant proteins were eluted with the MBP moiety using elution buffer (wash buffer plus 10 mM maltose). The recombinant protein concentrations were determined by a NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). Eluted proteins were resolved by SDS-PAGE (7.5% acrylamide) with Coomassie brilliant blue staining.

2.4. UDP-GloTM and GDP-GloTM GT Assays

Detection of free uridine diphosphate (UDP) after hydrolysis of the sugar nucleotide was performed using the UDP-GloTM GT assay kit (Promega Corporation, Madison, WI, USA), which detects UDP after UDP-sugar hydrolysis or transfer by converting UDP to light (measured in Relative Luminescence Units) in a luciferase-type reaction. Detection of free guanosine diphosphate (GDP) from GDP-sugar hydrolysis was evaluated using the GDP-GloTM GT assay kit (Promega Corporation), which operates by the same principles as above. A standard curve using 0–25 µM of the respective nucleoside diphosphate (NDP) was performed, and the range of measurements was determined to be in the linear range of detection, where the luminescence detected is directly proportional to NDP concentration.

Enzymes were diluted with an optimized GT solution (0.1 M MOPS-NaOH pH 7, 10 mM Mn²⁺) and supplemented with 100 µM of the targeted nucleotide sugar(s). Each sugar-nucleotide hydrolysis reaction was incubated at 16 °C for 16 h. Following the manufacturer's protocol, each reaction was combined in a ratio of 1:1 (25 µL: 25 µL) with the UDP-GloTM detection reagent in independent wells of a white, flat bottom 96-well assay plate and allowed to incubate at ambient temperature. After 1 h of incubation, luminescence was measured in triplicate, which is directly proportional to UDP or GDP concentration based on the standard curves. Luminescence measurements were performed with a VeritasTM microplate luminometer (Turner Biosystems, Sunnyvale, CA, USA) using a 96-well microplate with standard 128 × 86-mm geometry with an integration time of 1.0 s. Luminescence was measured by a low-noise photomultiplier detector through an empty filter position in the emission filter wheel.

3. Results and Discussion

3.1. In Silico Analysis of A111/114R Domain 1

PBCV-1 encoded protein A111/114R was annotated in NCBI (<https://blast.ncbi.nlm.nih.gov/>) as a hypothetical protein (NP_048459) of which only one region was predicted to be a GT (262 to 382 aa). PSI-BLAST analysis [20] revealed that A111/114R was conserved among all chloroviruses with 62–94% sequence identity, reinforcing the notion that it could be involved in the assembly of the conserved oligosaccharide core. Therefore, to elucidate the function of A111/114R, we analyzed the full-protein sequence using HHpred tool [19], which predicted three putative GT domains. N-terminal homology with several known GTs (>98% probability) was identified for residues 1 to 256; hence, we assigned residues 1 to 260 aa as domain 1 (A111/114R-D1). The other two domains are discussed in the following sections.

Protein structure prediction by Phyre2 analysis [18] identified six GT crystal structures (Table S1) to model three-dimensional structures of A111/114R-D1 based on 50–82% protein coverage (>90% confidence and 12–23% sequence identity). Models were ranked according to raw alignment score using the sequence and the secondary structure similarity, inserts, and deletions. Interestingly, the top ranked models were based on two XylTs (PDB:6BSV, 4WMA), one glucosyltransferase (GlcT) (PDB:1LL2), and three GalTs (PDB:1GA8, 5GVV, 6U4B), in agreement with our initial hypothesis, as both Xyl and Gal are part of the conserved region of the core glycan (Figure 1). The highest ranked model was the XylT XXT1 from *Arabidopsis* (PDB: 6BSV) based on 120 residues (19% identity with 96% confidence). However, the homologous region of A111/114R-D1 (9 to 168 aa) that aligned with XXT1 residues 167 to 287 had no homology with the residues in the active site of the XXT1 model, which utilized Lys-382, Asp-317, Asp-318, and Gln-319 to bind the nucleotide sugar for catalytic activity [21]. Similarly, important residues for enzyme activity in the XylT XXYL1 (PDB: 4WMA) from *Mus musculus* [22] (His-262, Trp-265, and Gly-325) did not share homologous positions with A111/114R-D1. Based on this evidence, we deduced that A111/114R-D1 has no XylT activity, and for this reason we examined the other high top ranked models, all reporting well characterized GalTs (PDB: 1GA8, 5GVV, 6U4B), except for 1LL2 which is a GlcT involved in glycogen synthesis [23]. Although GlcT activity cannot be definitely ruled out, A111/114R-D1 lacks the equivalent 1LL2 residue, Tyr-194, involved in Glc addition.

In order to investigate possible GalT activity for A111/114R-D1, we chose, as reference protein, LgtC (1GA8), a GalT of *Neisseria meningitidis*, for which the residues involved in the sugar–nucleotide binding and in the catalysis were well characterized. LgtC is a retaining GT that transfers α -D-Gal from UDP-Gal to a terminal lactose. The structure of LgtC was solved in complex with Mn^{2+} and a non-cleavable analog of the donor sugar (UDP-Gal) in which the hydroxyl at the 2' position of the Gal was substituted by a fluorine for stability purposes [23]. The alignment of the protein sequences of LgtC and A111/114R-D1 (Figure 3), along with the structural superimposition of the 3D model of A111/114R-D1 based on LgtC (Table S1) with 1GA8 (Chain A) in complex with Mn^{2+} and an analog of the donor sugar (Figure 4), clearly revealed that all the residues responsible for binding and catalysis were preserved. In detail, A111/114R-D1 shares four sugar binding residues with LgtC, namely Asp-85, Asn-149, Asp-193, and Gln-194, which correspond to Asp-103, Asn-153, Asp-188, and Gln-189 in LgtC (1GA8). Mutagenesis experiments have established that LgtC Gln-189, contained within the invariant D/EQD motif found in all members of GTs in family 8 (GT8), plays a crucial role in binding and probably in catalysis as well [24]. A catalytic mechanism leading to the retention of the anomeric carbon configuration for A111/114R-D1 is consistent with the presence in the oligosaccharide core of an α -D-Gal (Figure 1). It has also been proven that LgtC is a cation-dependent GT [23]. Indeed, LgtC exhibits the typical DXD motif ($^{103}DXD^{105}$), common in a wide range of GTs, both in prokaryotes and eukaryotes [25]. In addition, a crucial role has been attributed to Asp-103, as LgtC D103E and D103N mutants present a dramatic reduction in their activity compared to the wild-type [23].

(a)

D1 1 MTIVGITFGTEQYLGSAAVLRHSALTTFGEFDEFVFEKKDISWLMDTYPNHFENSRGFGFWCWKPELIRNVMNQLPDCGLV
 LgtC 67 -----EDFAGFPLNIRHISITTYARLKI GEYIADCLK

D1 81 VYICDSTMYFERSTIKFYIDHVEHNNPVLCLRGWSDNKNKYRNRWTKKSVFNSMGACNTVAEEIQLNASFOYK-NSP
 LgtC 99 VLYLDIDVLRVDSLTFLWDTDLGDNWLGASID-----LFVERQEGYKIKLQX-----DGEYFFNAGVLIINLKVK

D1 160 ETRAFVDQYLOVCLNLDITNDEGRDGGIFDTRHDQSILSIVASEHPRVTFSRDVSQWGKQDPPCSISQPTGGAIELDALD
 LgtC 165 RREDFIKXSSEVEQYKDVXQYQ-----DQDILNGIFKGG-----

D1 240 ENGVMHNLVNHHRMMKIPKI
 LgtC -----

(b)

D2 261 AVITPTTG-GKFLDAQIKSVQSSITLPNIEHWIIVDGREHEGKVDMLIEKYRHKHPVIRLVLPPNNIGSGGWNHGRVFC SIP
 K4CP 435 SYIPAYNCSKYIVRQVESALNQTITDIEVQICDDGSTD--TLRIIQEHYANHPRVFISQKN---KIGSASNTAVRL

D2 340 WIINADYISYLDNIEPKHFKDLVSSIISTPNAKWSYCLRLIDNNGNLIQYDNCSLGGISHTVAGRGDYIIDTSCY
 K4CP 545 --CRGFYIGQLSDDFLEPDAVELCDEFKRLSLACVYTTNENIDREGNLI SNGYNWPIYSRE----KLTSANICHHFR

D2 420 MIDRDLAISASPAWNARFRDPDGKPEPDEI SKCLLSAAPYGVVRKHSINRYIGSTGLSVSNFFVQGNSIFGYDFEKEYE
 K4CP 619 MFTARAWNLTG-----FNESISNAVYDYMLKLSVGPPEKHINKICYNRVLHD-----

D2 500 DIYVFHFSEKATADFMNARRQYKNRSYALEEWQMTLLKGLDGMNGKYNLLNGFTNFPNI
 K4CP -----

(c)

D3 560 PNKATILVSLCNFGDLPMDFFKERVLDLQRI-VYTLSEPNI---RHQGWNFNLTQHFDVATYFPIIENKSIHTIFT
 hpFucT 63 PNEFSDLVFGNPLGSAR-KILSYQNAKRVFYTGENSEPNFNLFDYATGFDELNDNRYLRVPIYYDLHH-----KAE

D3 635 PNTTHHDLDDPHDATAALLRVNKGVG--KSAGMILERRPHLFHTRDYAINGVHLRLCDYLRFDLVRLGLEDVTVFGINWGE
 hpFucT 135 SVNDTTPYKLDNSLYALKKPSHCFEKHPNCAVVNDESDELKRGFASFASNPNAPIIRNAFYDALNSIEPVTG-CGS

D3 713 VADGKKIKLGHAHRSQDENSSVDLRSKFEVDVVENCDAEGYVSEKLFDSLSAGCVPIYYGNMYDELGLDIPEGDVIFD
 hpFucT 214 VRN-----TLGYNVKNKNEFLSQYKENCFENTQGYGYVTEKIDAYFSHTPIYFGSP--SVAKDFNPKS-IVN

D3 793 LKRNITTGKQLQELDTLS--DERVEGMR
 hpFucT 281 VED--FKNFDEAIDYKYLHTHKNAYLDML

Figure 3. Amino acid sequence alignment of A111/114R domains and known GTs. Invariant and similar residues are highlighted in black and gray, respectively. Sequences from the following organisms were used (PDB code) for three individual domains (D1, D2, and D3), respectively: (a) *N. meningitidis* LgtC GalT (1GA8), (b) *E. coli* K4CP dual GalNAcT and GlcAT (2Z86), and (c) *H. pylori* FucT (2NZW). Homologous residues positioned less than 4 Å from the nucleotide sugar in the three-dimensional model analysis are marked with an asterisk (*). Known residues from annotated enzymes that interact with the nucleotide sugar donor and ion are marked with a dot (•) and underline (—), respectively. Multiple alignment was performed by Pyre2 using structural information and homology extension. File output was compiled by BOXSHADE.

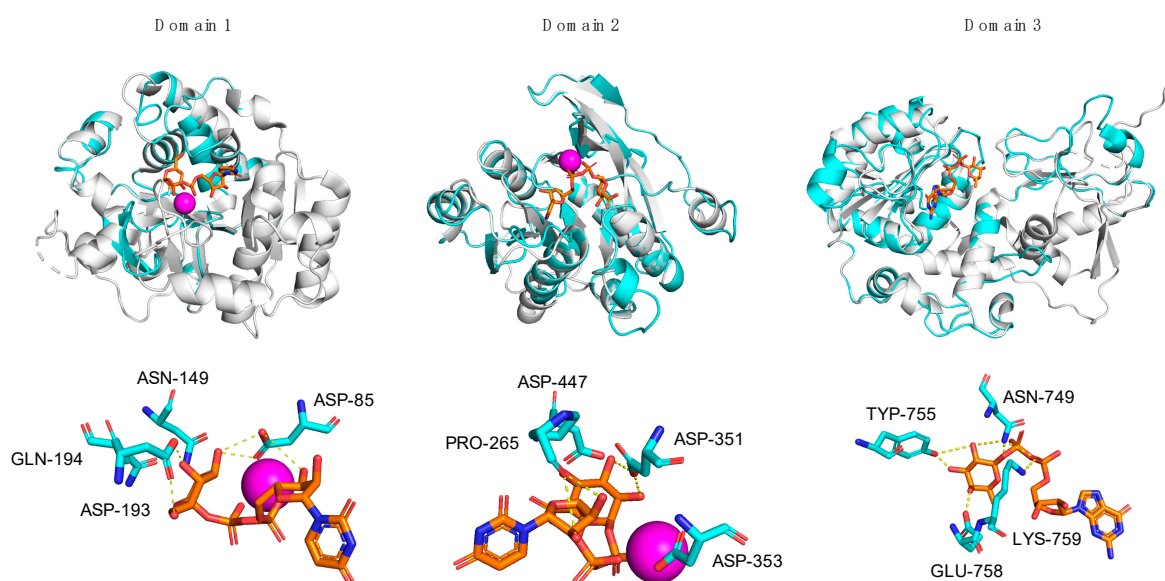


Figure 4. Superpositions of A111/114R-D1, -D2, and -D3 with structural homologs. Individual domains of A111/114R (cyan) are shown independently as ribbon diagrams superimposed with known GTs (gray) bound to respective nucleotide sugars drawn as stick models (orange) and Mn^{2+} ion (magenta): *N. meningitidis* LgtC with UDP-fluorogalactose (PDB: 1GA8), *E. coli* K4CP with UDP-glucuronic acid (PDB: 2Z86), and *H. pylori* FucT with GDP-Fuc (PDB: 2NZY) are superposed with D1, D2, and D3, respectively (left to right). The corresponding active sites of D1, D2, and D3 are shown magnified below the complexed stereoviews with labeled residues proposed to be involved in sugar and ion coordination. Hydrogen bonds are represented as yellow dotted lines.

However, the $^{103}DXD^{105}$ motif of LgtC is not preserved in position with A111/114R-D1, except for LgtC Asp-103, which corresponds to A111/114R-D1 Asp-85. Homologous to residue LgtC Asp-103, Asp-85 of A111/114-D1 is positioned in close proximity to a divalent cation and to the nucleotide sugar, as denoted with the characteristic ligand distance $< 4 \text{ \AA}$ (Table 1). It is likely that Asp-85 is sufficient in providing one side-chain oxygen ligand in coordination with Mn^{2+} , as evident from the structural superimposition (Figure 4). The evidence that A111/114R-D1 possesses all residues involved in binding and catalysis is further supported by the fact that other well-noted GalTs exhibit similar homologies. The GT GlyE from *Streptococcus pneumoniae* TIGR4 (PDB: 5GVV) binds UDP-Gal [26], and it shares the same conserved sugar binding residues as LgtC and A111/114R-D1 (Asp-103, Asn-142, Asp-177, and Gln-178) (Figure S2). Mutation of these key residues in GlyE completely abolished the hydrolytic activity [26]. Additionally, the bifunctional domain polymerase WbbM from *Klebsiella pneumoniae* (PDB: 6U4B) possesses a C-terminal galactopyranosyltransferase [27] with similar homologous residues, Asp-486 and Gln-487, resembling the known GT8 family enzyme signature likely to bind UDP-Gal.

3.2. In Silico Analysis of A111/114R Domain 2

Three-dimensional renderings of the central domain of A111/114R (261 to 559 aa), referred to as A111/114R-D2, were assembled based on protein alignment and secondary structure similarities. Phyre2 analysis assigned protein model predictions based on as high as 97% coverage with 100% confidence (11–18% sequence identity) from multiple N-acetylgalactosaminyltransferases (GalNAcT) (Table S1). The top ranked model was based on the second domain of the chondroitin polymerase K4CP from *E. coli* (PDB: 2Z86), namely 197 residues (66% of the protein sequence) of A111/114R-D2 were modelled with 100% confidence. K4CP is a bi-functional enzyme organized into two GT-A domains (A1 and A2) that catalyzes elongation of the bacterial chondroitin chain [28]. K4CP A1 (1–417) and A2 (418–682), located respectively to the N- and C-terminal, are engaged in the transfer of N-acetylgalactosamine (GalNAc) and glucuronic acid (GlcA) residues alternatively from UDP-GalNAc and UDP-GlcA. The 3D model of A111/114R-D2 (261–474)

is based on the second domain (A2) of K4CP (435 to 631 aa), which binds UDP-GlcA, thus excluding GalNAcT activity. The GlcA, absent in the oligosaccharidic core, has the same stereochemistry of Xyl and differs from this monosaccharide by a carboxyl function attached to carbon 5. This finding suggests that the A111/114R-D2 could be a XylT, and for that reason we used K4CP-A2 as a reference to assess the conservation of the residues implicated in binding and in catalysis. Sequence alignment of A111/114R-D2 with K4CP-A2 validated the conserved residues involved with GlcA binding and divalent cation coordination (Figure 3). In detail, A111/114R-D2 residues involved in the sugar-nucleotide binding are Pro-265, Asp-294, and Asn-323, which correspond to Pro-439, Asp-469, and Asn-496 in K4CP-A2 [28]. Superposition of K4CP Chain D with the predicted A111/114R-D2 structure (Figure 4) showcase these residues aligning in close proximity (<4 Å) to the nucleotide sugar (Table 1), strengthening their participation in sugar-binding and catalysis. K4CP, in analogy with other GT-A fold GTs, has a ⁵¹⁹DSD⁵²¹ motif coordinating the Mn²⁺. This DXD motif is preserved in A111/114R-D2 and corresponds to ³⁵¹DDD³⁵³.

Table 1. Predicted H-bond distances between homologous A111/114R residues and nucleotide–sugar ligand.

	A111/114R Residues	H-Bond Distance of Nucleotide-Sugar (Å)
Domain 1	Tyr-83	<4
	Asp-85	2.6, 3.0, 3.6
	Asn-149	<4
	Ala-150	<4
	Asp-193	2.4, 2.5
	Gln-194	<4
Domain 2	Pro-265	2.8, 3.3
	Asp-294	<4
	Asn-323	<4
	Gly-328	<4
	Asp-351	3.2, 3.3
	Asp-353	<4
	Asp-447	<4
Domain 3	Leu-617	<4
	Val-684	<4
	Arg-693	<4
	Gly-711	<4
	Asn-749	1.6, 3.1, 3.3
	Tyr-755	1.9, 3.4
	Ser-757	<4
	Glu-758	2.2
	Lys-759	2.5
	Asp-762	<4

Together, the corresponding catalytic sites and DXD motif support orthology in enzyme activity. Of further note, two additional DXD motifs (Figure S1) are present downstream of A111/114R-D2 (⁴²²DRD⁴²⁴ and ⁴³⁹DPD⁴⁴¹) that potentially could play an active role in substrate recognition or catalysis, but they do not exhibit homology with K4CP (Figure 3). Should A111/114R-D2 have functional XylT activity, it would be the first XylT sandwiched between two domains specific for different nucleotide sugars. The placement of the proximal Xyl in the core glycan structure, positioned closely to Gal and Fuc, is agreeable with the proposed domain organization of A111/114R.

3.3. In Silico Analysis of A111/114R Domain 3

The C-terminal domain (560 to 860 aa), referred to as A111/114R-D3, was clearly predicted as a putative α-1,3-FucT (Table S1) modelled from *Helicobacter pylori* (PDB: 2NZW) with 100% confidence and 15% sequence identity (86% coverage). The prediction is that Fuc

is linked to the O-3 of a Glc, which, like the monosaccharide contributions of domains D1 and D2, is a component of the overall virus core glycan structure. FucT belongs to the GT B family, in which the protein contains N- and C-terminal domains binding to the acceptor and donor substrates, respectively. Normally, these GTs do not have a recognizable DXD motif responsible for the Mn^{2+}/Mg^{2+} binding; however, A111/114R-D3 has a ⁶⁴²DLD⁶⁴⁴ signature (Figure S1) that must be evaluated for activity.

A sequence alignment of A111/114R-D3 and *H. pylori* FucT (hpFucT) revealed conserved residues involved in binding of the donor substrate, GDP-Fuc (Figure 3); Asn-240, Tyr-246, Glu-249, and Lys-250 correspond to Asn-749, Tyr-755, Glu-758, and Lys-759 in A111/114R-D3, respectively (Table 1). Classified as an inverting GT, the proposed catalytic mechanism of hpFucT incorporates Glu-95 as a general base in catalysis, while Lys-250 and Arg-195 in hpFucT share a key role in the neutralization of the negative charged phosphate groups from GDP-Fuc to facilitate the glycosidic bond cleavage [29]. Glu-249 acts to stabilize in part the positive charge developed in the transition state as well as to form two hydrogen bonds with both the ribose and the Fuc residues of GDP-Fuc. The Glu-95 residue of hpFucT is located in the N-terminal domain, presumably associated with the acceptor substrate, and it has no equivalent in A111/114R-D3. This finding could be related to alternative acceptors in the two FucT enzymes compared here, for example, hpFucT binding GlcNAc or A111/114R-D3 binding Glc.

Comparison of superimposed proteins disclosed that the second half of A111/114R-D3 traces the C-terminal domain of hpFucT Chain A (160 to 320 aa), illustrating similarities in secondary structures. In contrast, structure resemblance decreases towards the N-terminal regions, which could be a result of different acceptor substrates. The open structure of the N-terminal region of the A111/114R-D3 model could be a consequence of a larger acceptor.

To evaluate the sugar-binding and catalytic sites of A111/114R-D3, we superimposed the domain with hpFucT (Chain A) in complex with GDP-Fuc (PDB: 2NZY). This overlap confirmed that all the expected residues involved in the GDP-Fuc binding were preserved and in good orientation (Figure 4), in agreement with the sequence alignment data (Figure 3). A111/114R-D3 residues Asn-749, Tyr-755, Glu-758, Lys-759 appeared <4 Å from the nucleotide sugar in agreement with their equivalent residues in hpFucT (Figure 4). This observation suggests these A111/114R-D3 residues likely participate in GDP-Fuc binding and are critical for A111/114R-D3 activity. Specifically, Asn-749, Tyr-755, and Glu-758 likely stabilize the positive charge on the Fuc moiety like their equivalent residue in hpFucT. Positioned close in proximity to GDP-Fuc is A111/114R-D3 residue Arg-693 that corresponds to Arg-195 in hpFucT, an essential residue for enzyme activity. In fact, Ala mutants of Arg-195 or Lys-250 resulted in no detectable activity, supporting the idea that the two residues provide positive charges to interact with negatively charged GDP-Fuc [29]. The important role of Arg-693 is supported by its conservation amongst different chloroviruses.

3.4. In Vitro Evidence of A111/114R Hydrolytic Activity

In order to test A111/114R for GT activity, we used the bioluminescent UDP/GDP-Glo assays to detect free UDP/GDP released by GT-mediated hydrolysis of the nucleotide sugars. This allowed us to screen for nucleotide donor specificities without their target substrates (acceptor). A time course experiment with the full-length recombinant A111/114R protein in the presence of the core monosaccharides UDP-Gal, UDP-Xyl, GDP-Fuc, and Glc displayed evidence of UDP-hydrolysis (Figure 5). Glc was supplemented to simulate the Glc-Asn acceptor located on the nascent glycoprotein. The release of the UDP increased steadily over time, suggestive of GT activity by A111/114R-mediated hydrolysis.

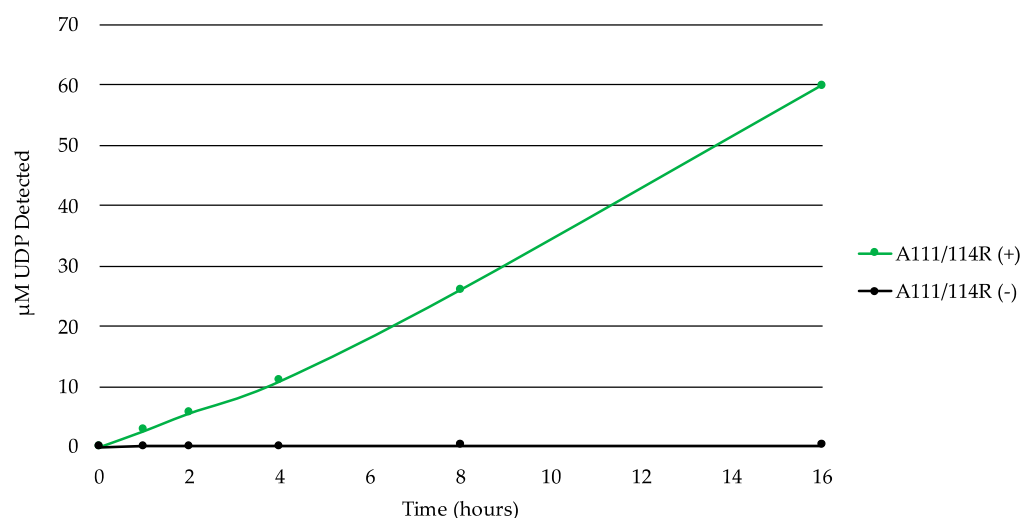


Figure 5. A111/114R-catalyzed hydrolysis of UDP-sugars. Time-course experiment with the full-length recombinant A111/114R protein (6 μ g) in the optimized buffer consisting of 0.1 M MOPS-NaOH (pH 7.0), 10 mM $MgCl_2$, and 100 μ M each of UDP-Gal, UDP-Xyl, GDP-Fuc, and Glc for 16 h at 16 $^{\circ}$ C. The release of UDP was detected by the UDP-GloTM assay. Data are representative from three independent replicates, and error bars represent standard deviation.

We recombinantly expressed the full-length A111/114R protein (1 to 860 aa) and three variants with omitted regions (Figure 6a,b). A111/114R (1–397 aa) contains the complete domain 1 and approximately the first half of domain 2, A111/114R (266–860 aa) contains the complete domains 2 and 3, and A111/114R (391–860 aa) contains the second half of domain 2 and the complete domain 3. Constructs containing only individual domains were originally designed; however, their proteins did not exhibit activity. This may reflect inaccurate residue boundaries or an incorrect folding of the single domains in the absence of adjacent ones. Representative data from each assay are shown in Figure 6c,d and are represented as a ratio of the UDP or GDP measured from reactions containing the indicated GTs relative to the full-length protein-catalyzed hydrolysis, and negative controls in which no enzyme was added.

Analysis of UDP/GDP-hydrolysis by the A111/114R constructs reported in Figure 6b were especially revealing in regard to A111/114R domain assignments. Indeed, starting with the A111/114R (1–397) construct, the UDP molecules were detected only when UDP-Gal was used (Figure 6c). Given A111/114R-D2 is half omitted in A111/114R (1–397), hydrolysis of UDP-Gal implies A111/114R-D1 is a GalT. In agreement, no UDP was produced from UDP-Gal containing reactions involving A111/114R constructs devoid of the first domain, A111/114R (266–860) or A111/114R (391–860). This finding is in agreement with the bioinformatic data.

Hydrolysis of UDP-Xyl by A111/114R (266–860) suggests either A111/114R-D2 or A111/114R-D3 has XylT activity. However, UDP-Xyl was not detected in reactions involving A111/114R (1–397) or A111/114R (391–860), both of which are devoid of a complete D2 (Figure 6b). This suggests that A111/114R-D2 (260–560) harbors the XylT activity. GDP-Fuc hydrolysis was detected in reactions involving A111/114R (266–860) and A111/114R (391–860) exclusively (Figure 6d). These results suggest A111/114R-D3 is a FucT. Notably, GDP-Fuc reactions supplemented with UDP-Gal and UDP-Xyl showed elevated levels of liberated GDP, especially in the presence of A111/114R (266–860). This could be a result of improved protein folding allowed by the extension of residues and co-presence of the nucleotide sugars.

Finally, to evaluate the residues of A111/114R involved in hydrolytic activity, we constructed Ala mutants by site-directed mutagenesis (SDM) (GenScript) to target amino acids from each domain predicted to be involved in nucleotide-sugar or metal-ion binding. Three mutants were expressed, each containing two Ala substitutions inside separate

domains. A111/114R-D1, -D2, and -D3 SDM constructs contained Ala mutants of Asp-85 and Gln-194, Asp-351 and Asp-353, and Arg-693 and Lys-759, respectively. In the presence of UDP-Xyl, UDP-Gal, GDP-Fuc, and Glc, the SDM of D1, D2, and D3 resulted in a significant reduction in UDP-sugar hydrolysis, lowering the activity by 95%, 90%, and 80%, respectively (Figure S3). Likewise, in the presence of the same nucleotide-sugars, GDP-Fuc hydrolysis was reduced by 90%, 70%, and 96%, respectively. This dramatic reduction in detectable nucleotide-sugar hydrolysis supports the idea that these residues are critical for enzyme activity, and that A111/114R functions best when all three domains are active.

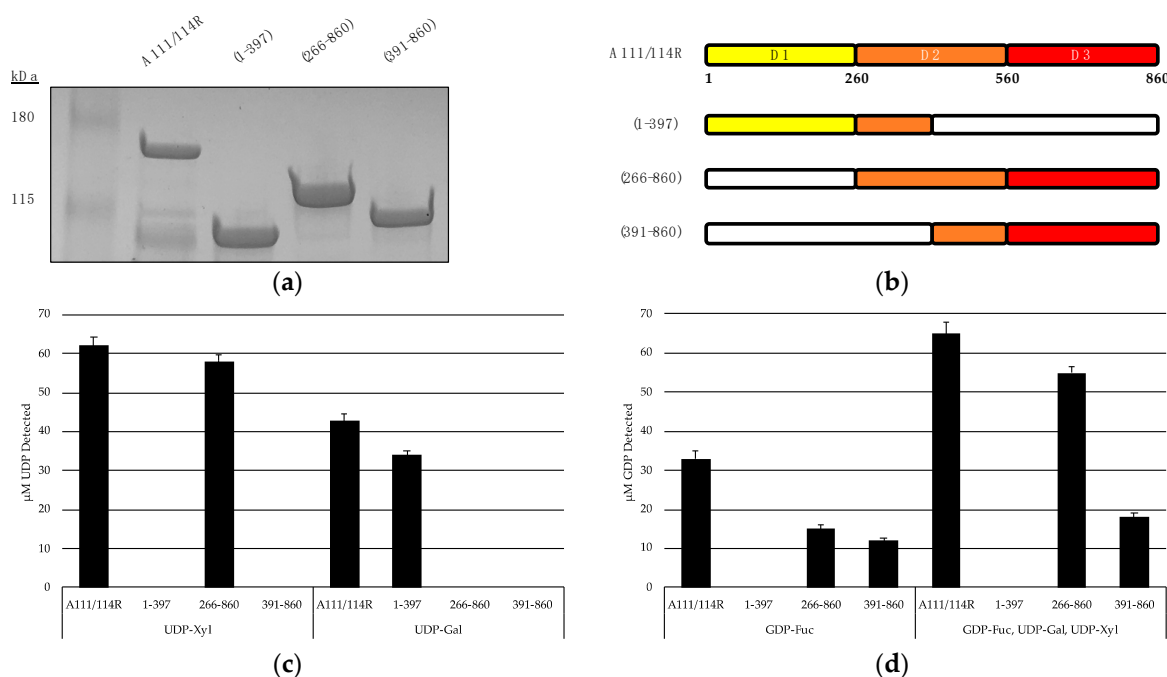


Figure 6. Hydrolysis of UDP- and GDP-sugars by recombinant A111/114R and truncated constructs. **(a)** SDS-PAGE analysis of the expressed proteins: full length recombinant MBP-A111/114R (144 kDa), and truncated variants MBP-A111/114R (1–397) (87 kDa), MBP-A111/114R (266–860) (110 kDa), and MBP-A111/114R (391–860) (96 kDa) were eluted from an amylose column and resolved by SDS/PAGE with Coomassie blue staining. BenchMark™ pre-stained protein ladder and recombinant proteins were separated on a 4–20% tris-glycine gel. **(b)** Cartoon renderings of the full-length A111/114R protein and truncated versions are color coordinated by three putative domains (D1, D2, D3), each corresponding to a different nucleotide sugar donor. White regions denote omitted sections of A111/114R. Shortened constructs are defined by their residues in the left column. **(c,d)** Representative data from hydrolysis assays shown as a ratio of the UDP **(c)** or GDP **(d)** measured from reactions containing the indicated GTs relative to the negative controls where no enzyme was added. The release of UDP and GDP was detected by the UDP-Glo™ assay and GDP-Glo™ assay, respectively. GDP-hydrolysis from GDP-Fuc was significantly elevated in the presence of UDP-Gal and UDP-Xyl with A111/114R (266–860) and A111/114R (391–860). Data are representative from three replicates, and error bars represent standard deviation.

4. Conclusions

The giant chloroviruses continue to challenge our understanding of canonical metabolic pathways in host–virus interplay. The identification of the atypical N-glycan structure attached to the PBCV-1 Vp54 has led to the characterization of virus-encoded enzymes involved in glycosylation independent of host-derived ER and Golgi GTs. The N-glycan's core pentasaccharide conserved among the chloroviruses is especially noteworthy and prompted us to investigate candidate virus-encoded GTs involved in the assemblage of part or all the conserved core oligosaccharide. Results in this study establish that PBCV-1 encoded protein A111/114R has three GT domains of approximately 300 aa each. Evidence from a combination of amino acid alignments, three dimensional renderings, and GT assays indicate that the N-terminal (1 to 260 aa), central (261 to 559 aa), and the C-terminal

(560 to 860 aa) regions resemble a GalT, a XylT, and a FucT, respectively. The three-dimensional protein models built by Phyre2 are predictions and were used only for the purpose of identifying potential individual domains. As with all methods for protein modeling prediction, caution should be exercised when evaluating structural elements of new enzymes that lack homology to currently deposited structures in the Protein Data Bank archives. In fact, since the percent identity between the various domains and the structures on which they were modeled was low, it was not possible to model the various domains accurately. Eventually individual protein domains must be solved by biochemical and crystallographic methods to fully reveal their catalytic mechanism.

Preliminary evidence suggests that A111/114R-D2 has XylT activity and, if biochemically confirmed, presents a new structural class of XylTs based on its limited resemblance to known XylTs. Moreover, A111/114R would be the second three-domain protein encoded by PBCV-1 that is involved in glycan synthesis. Indeed, recent studies showed that PBCV-1 protein A064R (638 aa) has three functional domains; the first two are GTs (β -L-rhamnosyltransferase and α -L-rhamnosyltransferase, respectively) and the third is a methyltransferase that methylates O-2 of the terminal α -L-Rha residue [30].

It is known that many of the chlorovirus genes encode enzymes involved in various aspects of carbohydrate metabolism. However, it remains unclear how the virus-encoded proteins are involved in the synthesis and/or assembly of the Vp54 glycan. For example, are the sugars added to Vp54 sequentially or are they synthesized independently of Vp54, possibly on a lipid carrier, and then attached to the protein en bloc? A slight variation of these two possibilities is to synthesize a core glycan(s) independently of the protein and attach it to Vp54. Additional experiments will be required to address this issue.

Importantly, the results described herein provide support that the synthesis of the PBCV-1 core glycan structure, or at least part of it, is accomplished with a multidomain enzyme encoded by the virus itself. This finding is in line with the finding that another GT of PBCV-1, the protein A064R, is able to elongate the viral glycan with two units of Rha and to methylate the ultimate unit at O-2. Taking these findings together, the dogma that all viruses use host enzymes to glycosylate their proteins is further subverted.

Supplementary Materials: The following are available online at <https://www.mdpi.com/1999-4915/13/1/87/s1>, Figure S1: Amino acid sequence of chlorovirus PBCV-1 encoded protein A111/114R, Figure S2: Amino acid sequence alignment of A111/114R-D1 and annotated GalTs, Figure S3: Hydrolysis of UDP- and GDP-sugars by recombinant A111/114R and SDM constructs, Table S1: Glycosyltransferase homologs of A111/114R domains obtained using Phyre2 analysis.

Author Contributions: Conceptualization, E.N., G.A.D., C.D.C., and J.L.V.E.; methodology, E.N., A.N., and I.S.; software, E.N., A.N., and I.S.; validation, E.N.; formal analysis, E.N.; investigation, E.N., A.N., and I.S.; data curation, E.N., A.N., and I.S.; writing—original draft preparation, E.N.; writing—review and editing, E.N., A.N., I.S., G.A.D., C.D.C., and J.L.V.E.; supervision, project administration, resources, funding acquisition, C.D.C. and J.L.V.E. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded in part by the National Science Foundation under grant no. 1736030 (JVE), the National Science Foundation Graduate Research Fellowship Program under grant no. 2505060195001 (EN), and the Mizutani Foundation under grant no. 180047.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article or supplementary material.

Acknowledgments: We thank Fatima Maitham Al-Sammak for her technical support with experiments.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Bieberich, E. Synthesis, Processing, and Function of N-Glycans in N-Glycoproteins. In *Glycobiology of the Nervous System*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 47–70.
2. Vigerust, D.J.; Shepherd, V.L. Virus Glycosylation: Role in Virulence and Immune Interactions. *Trends Microbiol.* **2007**, *15*, 211–218. [[CrossRef](#)] [[PubMed](#)]
3. Van Etten, J.L.; Gurnon, J.R.; Yanai-Balser, G.M.; Dunigan, D.D.; Graves, M.V. Chlorella Viruses Encode Most, If Not All, of the Machinery to Glycosylate Their Glycoproteins Independent of the Endoplasmic Reticulum and Golgi. *Biochim. Biophys. Acta (BBA) Gen. Subj.* **2010**, *1800*, 152–159. [[CrossRef](#)] [[PubMed](#)]
4. Van Etten, J.L.; Agarkova, I.; Dunigan, D.D.; Tonetti, M.; De Castro, C.; Duncan, G.A. Chloroviruses Have a Sweet Tooth. *Viruses* **2017**, *9*, 88. [[CrossRef](#)] [[PubMed](#)]
5. Van Etten, J.L.; Agarkova, I.V.; Dunigan, D.D. Chloroviruses. *Viruses* **2019**, *12*, 20. [[CrossRef](#)] [[PubMed](#)]
6. Speciale, I.; Duncan, G.A.; Unione, L.; Agarkova, I.V.; Garozzo, D.; Jimenez-Barbero, J.; Lin, S.; Lowary, T.L.; Molinaro, A.; Noel, E.; et al. The N-Glycan Structures of the Antigenic Variants of Chlorovirus PBCV-1 Major Capsid Protein Help to Identify the Virus-Encoded Glycosyltransferases. *J. Biol. Chem.* **2019**, *294*, 5688–5699. [[CrossRef](#)] [[PubMed](#)]
7. Nandhagopal, N.; Simpson, A.A.; Gurnon, J.R.; Yan, X.; Baker, T.S.; Graves, M.V.; Van Etten, J.L.; Rossmann, M.G. The Structure and Evolution of the Major Capsid Protein of a Large, Lipid-Containing DNA Virus. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14758–14763. [[CrossRef](#)] [[PubMed](#)]
8. De Castro, C.; Klose, T.; Speciale, I.; Lanzetta, R.; Molinaro, A.; Van Etten, J.L.; Rossmann, M.G. Structure of the Chlorovirus PBCV-1 Major Capsid Glycoprotein Determined by Combining Crystallographic and Carbohydrate Molecular Modeling Approaches. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E44–E52. [[CrossRef](#)] [[PubMed](#)]
9. De Castro, C.; Molinaro, A.; Piacente, F.; Gurnon, J.R.; Sturiale, L.; Palmigiano, A.; Lanzetta, R.; Parrilli, M.; Garozzo, D.; Tonetti, M.G.; et al. Structure of N-Linked Oligosaccharides Attached to Chlorovirus PBCV-1 Major Capsid Protein Reveals Unusual Class of Complex N-Glycans. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 13956–13960. [[CrossRef](#)]
10. Wieland, F.; Heitzer, R.; Schaefer, W. Asparaginyglucose: Novel Type of Carbohydrate Linkage. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 5470–5474. [[CrossRef](#)]
11. Mengele, R.; Sumper, M. Drastic Differences in Glycosylation of Related S-Layer Glycoproteins from Moderate and Extreme Halophiles. *J. Biol. Chem.* **1992**, *267*, 8182–8185. [[CrossRef](#)]
12. Schreiner, R.; Schnabel, E.; Wieland, F. Novel N-Glycosylation in Eukaryotes: Laminin Contains the Linkage Unit Beta-Glucosylasparagine. *J. Cell Biol.* **1994**, *124*, 1071–1081. [[CrossRef](#)]
13. Gross, J.; Grass, S.; Davis, A.E.; Gilmore-Erdmann, P.; Townsend, R.R.; Geme, J.W.S. The *Haemophilus Influenzae* HMW1 Adhesin Is a Glycoprotein with an Unusual N-Linked Carbohydrate Modification. *J. Biol. Chem.* **2008**, *283*, 26010–26015. [[CrossRef](#)] [[PubMed](#)]
14. Quispe, C.F.; Esmail, A.; Sonderman, O.; McQuinn, M.; Agarkova, I.; Battah, M.; Duncan, G.A.; Dunigan, D.D.; Smith, T.P.; De Castro, C. Characterization of a New Chlorovirus Type with Permissive and Non-Permissive Features on Phylogenetically Related Algal Strains. *Virology* **2017**, *500*, 103–113. [[CrossRef](#)] [[PubMed](#)]
15. Speciale, I.; Agarkova, I.; Duncan, G.A.; Van Etten, J.L.; De Castro, C. Structure of the N-Glycans from the Chlorovirus Ne-Jv-1. *Antonie Van Leeuwenhoek* **2017**, *110*, 1391–1399. [[CrossRef](#)] [[PubMed](#)]
16. De Castro, C.; Speciale, I.; Duncan, G.; Dunigan, D.D.; Agarkova, I.; Lanzetta, R.; Sturiale, L.; Palmigiano, A.; Garozzo, D.; Molinaro, A.; et al. N-Linked Glycans of Chloroviruses Sharing a Core Architecture without Precedent. *Angew. Chem.* **2016**, *55*, 654–658. [[CrossRef](#)] [[PubMed](#)]
17. Landstein, D.; Burbank, D.E.; Nietfeldt, J.W.; Van Etten, J.L. Large Deletions in Antigenic Variants of the Chlorella Virus PBCV-1. *Virology* **1995**, *214*, 413–420. [[CrossRef](#)] [[PubMed](#)]
18. Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N.; Sternberg, M.J.E. The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis. *Nat. Protoc.* **2015**, *10*, 845–858. [[CrossRef](#)] [[PubMed](#)]
19. Söding, J.; Biegert, A.; Lupas, A.N. The Hhpred Interactive Server for Protein Homology Detection and Structure Prediction. *Nucl. Acids Res.* **2005**, *33* (Suppl. 2), W244–W248.
20. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped Blast and Psi-Blast: A New Generation of Protein Database Search Programs. *Nucl. Acids Res.* **1997**, *25*, 3389–3402. [[CrossRef](#)]
21. Culbertson, A.T.; Ehrlich, J.J.; Choe, J.-Y.; Honzatko, R.B.; Zabolina, O.A. Structure of Xyloglucan Xylosyltransferase 1 Reveals Simple Steric Rules That Define Biological Patterns of Xyloglucan Polymers. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 6064–6069. [[CrossRef](#)]
22. Haltiwanger, R.S.; Yu, H.; Takeuchi, M.; LeBarron, J.; Kantharia, J.; London, E.; Bakker, H.; Li, H.; Takeuchi, H. Regulation of Notch Signaling by O-Glcucosylation: Notch-Modifying Xylosyltransferase-Substrate Complexes Support an Sni-Like Retaining Mechanism. *FASEB* **2016**, *30* (Suppl. 1), 624.3.
23. Gibbons, B.J.; Roach, P.J.; Hurley, T.D. Crystal Structure of the Autocatalytic Initiator of Glycogen Biosynthesis, Glycogenin. *J. Mol. Biol.* **2002**, *319*, 463–477. [[CrossRef](#)]
24. Persson, K.; Ly, H.D.; Dieckelmann, M.; Wakarchuk, W.W.; Withers, S.G.; Strynadka, N.C. Crystal Structure of the Retaining Galactosyltransferase Lgtc from *Neisseria Meningitidis* in Complex with Donor and Acceptor Sugar Analogs. *Nat. Struct. Biol.* **2001**, *8*, 166–175. [[CrossRef](#)] [[PubMed](#)]

25. Campbell, J.A.; Davies, G.J.; Bulone, V.; Henrissat, B. A Classification of Nucleotide-Diphospho-Sugar Glycosyltransferases Based on Amino Acid Sequence Similarities. *Biochem. J.* **1997**, *326*, 929–939. [[CrossRef](#)] [[PubMed](#)]
26. Jiang, Y.-L.; Jin, H.; Yang, H.-B.; Zhao, R.-L.; Wang, S.; Chen, Y.; Zhou, C.-Z. Defining the Enzymatic Pathway for Polymorphic O-Glycosylation of the Pneumococcal Serine-Rich Repeat Protein Psrp. *J. Biol. Chem.* **2017**, *292*, 6213–6224. [[CrossRef](#)]
27. Clarke, B.R.; Ovchinnikova, O.G.; Sweeney, R.P.; Kamski-Hennekam, E.R.; Gitalis, R.; Mallette, E.; Kelly, S.D.; Lowary, T.L.; Kimber, M.S.; Whitfield, C. A Bifunctional O-Antigen Polymerase Structure Reveals a New Glycosyltransferase Family. *Nat. Chem. Biol.* **2020**, *16*, 450–457. [[CrossRef](#)]
28. Ninomiya, T.; Sugiura, N.; Tawada, A.; Sugimoto, K.; Watanabe, H.; Kimata, K. Molecular Cloning and Characterization of Chondroitin Polymerase from Escherichia Coli Strain K4. *J. Biol. Chem.* **2002**, *277*, 21567–21575. [[CrossRef](#)]
29. Sun, H.-Y.; Lin, S.-W.; Ko, T.-P.; Pan, J.-F.; Liu, C.-L.; Lin, C.-N.; Wang, A.H.-J.; Lin, C.-H. Structure and Mechanism of Helicobacter Pylori Fucosyltransferase a Basis for Lipopolysaccharide Variation and Inhibitor Design. *J. Biol. Chem.* **2007**, *282*, 9973–9982. [[CrossRef](#)]
30. Speciale, I.; Laugieri, M.E.; Noel, E.; Lin, S.; Lowary, T.L.; Molinaro, A.; Duncan, G.A.; Agarkova, I.V.; Garozzo, D.; Tonetti, M.G. Chlorovirus PBCV-1 Protein A064R Has Three of the Transferase Activities Necessary to Synthesize Its Capsid Protein N-Linked Glycans. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 28735–28742. [[CrossRef](#)]


```

1  MTIVGITFGT EQYLGSAAVL RHSALTTGEF DEFVFEKGD ISWLMDTYPN HFENSRGFGF
61  WCWKPFLLRN VMNQLPDGDV VVYCDSTMYF ERSIKPYIDH VEHNNPVLC RLGGWSDNKN
121 DYNRRRWTKK SVFNMSGAGN TVAEEIQLNA SFQVYKNSPE TRAFVDQYLQ YCLNLDIIND
181 EGRDGGIFDT RHDQSILSIM ASEHPRVTF S RDVSQWGKQD PPCSISQPTG GAIELDALDE
241 NGVMHNLVNH HRRMMKIPKI AVITPTTGGK FLDACIKSVQ SSTLPNIEHW IIVDGREHEG
301 KVDMMILEKYR HKHPVIKLV PNNIGSGGWN GHRVFGSIPW IINADYISYL DDDNIVEPKH
361 FKDLVSSIIS TPNAKWSYCL RKLIDNNGNL IGYDNCESLG GISHTVAGRG DYLDITSCYM
421 IDRDLAISAS PAWNARFRDP DGKPEPDREL SKCLLSAAPY GVVRKHSLNY RIGSTGLSVS
481 NNFFVQGNSI FGYDFEKYED IYVFHFSEKA TADFMNARRQ YKNRSYALEE WQMTLLKGLD
541 GMNGGKYNLL NGFTNFPNIP NKATILVSLC NPGLPMDFF KERVDLQRIV YTLESPNIRH
601 QGQWNFNWLT QHFDVALTYF KPIIENKSIH TIFTPHNTHH GDLDDPHDAT ALLRVNKGVG
661 KSAGMILERR PHLFHTRDYA INGVHLRCLD YLREDLVRGL EDVTVFGINW GEVADGKKIK
721 LGHAKHRSQD ENSSVDLKSK FVFDIVVENC DAEGYVSEKL FDSL SAGCVP LYYG NMYDEL
781 GDLIPEGDVY FDLKKRNITT GKQLQELLDT LSDERVEGMR KNVIDYREKV LRFAGTKMFA
841 KKVEEAIELS KTTKKNVELV

```

Figure S1. Amino acid sequence of chlorovirus PBCV-1 encoded protein A111/114R. The A111/114R protein (NP_048459.2) has three putative GT domains: domain 1 (red, residues 1 to 260), domain 2 (black, residues 261 to 559), and domain 3 (blue, residues 560 to 860). Domain 1 is predicted to encode a GalT, domain 2 a XylT and domain 3 a FucT, based on homology modelling with other proteins. The signature GT motif DXD is underlined.

D1 1 MTIVGIT---FGTQYLSAA LRHSAITTTGEFDEFRV--FEKKDISWLMDTPNHENSRGFGFWQW-----
 GlyE 1 MRNTKRAVVFAGDYAYTRQ-----
 WbbM 1 MNSKIYTCCHHKPSAFLNASIK---PHVCKANTYNDIGCEGED-SGDNISFKNPYCELTAHYVWVKNEPLADYVGFM
 LgtC 1 MDIVF-----AADDNYAYY-----

D1 65 -----
 GlyE 20 -----
 WbbM 77 HYRRHLNFAEQQNHDPEDNWGVVNYPLINAEYESQFGLSDESISTCVDGYDLLLPKKWSVTSAGSKNNLDHYAKGEFLHIK
 LgtC 15 -----

D1 65 -----
 GlyE 20 -H-----
 WbbM 157 DYQSALDVVEELYPEYKDAIQFNNATDGYITNMFMVRKDMFTDYSEWLFALLSNLEDRISMNNYNAQEKRVIGHIAERL
 LgtC 15 -H-----

D1 65 -----
 GlyE 21 -----ETAKKSC-RH-N--SHLKIYLLNQ
 WbbM 237 FNIYIIKSQQDKQLKIKELQRTFVTAETFNGLNPIFDESVPVVISFDNNYALSGGALINSIV-RHSDANRNYDILVLEN
 LgtC 16 -----CVLAKEAEAAHPD--TEIRFHVLLA

D1 65 -----EFLIRNM-----NQIPDGDVVVYCDSTMYFERSKEY
 GlyE 42 DIPQEWFSQIRIYLQEMGG--DIDCKLIGSQFQMN-SNKIPHINMTIARYFIPDEFVTE-DKVTYLDSDILVTGDLIDL
 WbbM 316 KMSHLNQRLIK-LVAGHNNISLRFFD-VNSFTEMSDVHTRAFISASTYARLFIPQLFREYKKVVFIDSDTVVKSDLATL
 LgtC 39 GISSEANAAVAANLRGGG--NIRFID-VNPEDFAGEELNIRHISITTYARLKIGEYIADCKVLYLIDIVVRDSLTPL

D1 98 IDHVEHNNPIVLRLGGTSDNK-----NDYRNRRWTKKSVFNSMGAGN
 GlyE 119 FEIDLGNMLAAR-SCEGAG-----VGFNAGVLLINKKWGSETIRCK-----
 WbbM 394 LDVEIGTNLVAAR-DVMEGFVKFGTMSSEDDGIMPAGEYLLKKTIGMTNPDEYFQAGTIVFNVEQMVKENTFAQ-----
 LgtC 116 WIDLGNMLGAGT-DDEVERQ-----EGYKQKIGMADGYYFNAGVLLINLKKWRHDIKFM-----

D1 141 TVAEELQNASFQVYKNSPETRAFVDQY--HOYCLNLDIINDEGRDGGI-----FD--IRHD-
 GlyE 162 -----LID--LTEKEHENEE-GDQSILNMLEKDYSSLEDQYNEQIG-VDYGAATFK--HQFT-
 WbbM 468 -----LMS--ALKARK--YWE-LDQDILNKKFFGRVKFPLEWNVYHG-NENTDDFFPNLKFSTYM-
 LgtC 173 -----SCE--WVEQYKDVQY-QDQDILNGLFKGGVCYANSRENFEMPTNYAFEMANRFA--SRHTD

D1 194 -----QSILSIMASEH--PRVTFSRDV-----SQWGKQ
 GlyE 215 ---FDIPLEPLP-LILHYLSQDKPWNQFSVGRLEVVWWEYSLMDWSV-LNEWFSKSVKYPSKSQIFKLQCVNITNSWCVE
 WbbM 523 ---RFEARRNP-KMIHYAGENKPNNTKVDYDFD-L--EN-INTPWEKE
 LgtC 228 PLYRDRNTVMPTAVSHYCGGPAKPWHRDC-----TAWGAE

D1 220 DPP-----CSISQPT-GG-----
 GlyE 291 KIDYLAELPEVHFHIVAYTNMANELLALTRFENVTVYPNSLPMLLEQIVTASDIYLDLNDHRKLEDAYEVLKYKTPM
 WbbM 567 ILY-----RQLEVAT-VVPNQHTLQQTVLITQIK-----FAL
 LgtC 263 RET-----ELAGSIT-TVPEEWRG-----

D1 232 -----ATELDAIDENGVMHNVNHHRRMM-----K--PKI-
 GlyE 370 IADNTCSENLSEISYEGTYPSSTPKMMVAAIRSYM-----R-----
 WbbM 600 MPVVKY-----APVGSPPRNKLTQY-YYVRRSILG-----
 LgtC 281 -----KLAVPHRMFSTKRMLQW-RRKTSARFIRKIY

Figure S2. Amino acid sequence alignment of A111/114R-D1 and annotated GalTs. Invariant and similar residues are highlighted in black and gray, respectively. Sequences of GalTs from the following organisms were used (PDB code): *S. pneumoniae* GlyE (5GVV), *K. pneumoniae* WbbM (6U4B), and *Neisseria meningitidis* LgtC (1GA8). Multiple alignment was performed by T-Coffee [31] using structural information and homology extension. File output was compiled by BOXSHADE.

D3	1	PNKATILVSLCNPGLPMDFFKERVDLRIVYTLESPNIRHQGQWNF-NWLTQHFDVALTYPKPIIDNKSIHTLFTPHNT
NY2A	1	PKATVILNLCNPGDLPMEFFKERVDLRIVYTLESPNIRHQGQWNS-NWLMQHFDVLTYPKPIIDNENISTVFTPHNT
MT325	1	PEGSTVLVSLCNPGLPLEFLKERRDLNRIVYTLESPNIRHAGQWNIPKFLQEHFDVALTYPFLLTS-DFPTVYASHNC
FR483	1	PEGSTVLVSLCNPGLPLEFLKERRDLNRIVYTLESPNIRHAGQWNIPKFLQEHFDVALTYPFLLTS-DFPTVYASHNC
ATCV1	1	PENATVLVSMCNPGELPLEFLKKRQDLRLIVYTLESPNIRHQGQWSL-GFTKEHFDVALTYPKPLMDK--MPTVFAAHNT
NE-JV-3	1	PENATVLVSMCNPGELPLEFLKKRQDLRLIVYTLESPNIRHQGQWSL-GFTKEHFDVALTYPKPLMDK--MPTVFAAHNT
Syngen	1	PKATVILNLCNPGDLPMEFFKERVDLRIVYTLESPNIRHQGQWNL-NWLVQHFDVAMTYFKPIIDNKGLHTVFTPHNT
*		
D3	80	HHGDLDDPHDATAALLRVNKGVGKSAGMILERRPELFHTRDYAINGVHLRCLDYLREDLVRGLEDVTVFGINWGEVADGKK
NY2A	80	HHGDLDDPLDATSLLRPNKGIGKSAGMVLERRPYLFETRDYSINGVHMRCLDYLREDLVRGLEDVTVFGINWGEVADGKK
MT325	80	HHGTLDDPLDRAVLLRDNKGSGRSVGLVLERRPELFNHKEYSVLCCQLQCLDYLREDLVKGOKDVTVFGMNWSEVADGVN
FR483	80	HHGTLDDPLDRAVLLRDNKGSGRSVGLVLERRPELFNHKEYSVLCCQLQCLDYLREDLVKGOKDVTVFGMNWSEVADGVN
ATCV1	78	HQGDRLDPLDRAALLRENAGSDRSVCLVLERRPELFAHNLVSVNGVTLLKCLDELREKLVVGONNITAFGMNWGAVADGKR
NE-JV-3	78	HQGDRLDPLDRAALLRENAGSDRSVCLVLERRPELFAHNLVSVNGVTLLKCLDELREKLVVGONNITAFGMNWGAVADGKH
Syngen	80	HHGNLDDPLDRTMLLRANKGVCKSVGMVLERRPELFHTRDYAINGEHMRCLDYLREDLVRGLEDVTVFGINWNEVADGKK
D3	160	IKLGHAKHRSQDENSSVDLKSKEFVFDIVVENCDAEGYVSEKLFDSLSAGCVPLYYGNYDELGLDPIEGD--VFYDLKKR
NY2A	160	IKLGHAKHRNHDENSSIDLKSNFVFDIVVENCDAEWYVSEKIFYDALGAGCIPLYYGNYDELGLDPIEGE--VFYDLKKR
MT325	160	IKLGHGNHRSRDNKSSVDHKQNFVFDIVVENCDAEGYVSEKIFYDALGAGCIPLYYGNYVFDKLSDLPIEGPDGAEFIDLKKR
FR483	160	IKLGHGNHRSRDNKSSVDHKQNFVFDIVVENCDAEGYVSEKIFYDALGAGCIPLYYGNYVFDKLSDLPIEGPDGAEFIDLKKR
ATCV1	158	IKLGHNNHRSADSLASIDHKQNFVFDIVVENCDAAGYVSEKIFYDALGAGAIPLYYGNYMFDELSALPIEGPGGAYFDLKKR
NE-JV-3	158	VKLGHNNHRSADSLASIDHKQNFVFDIVVENCDAAGYVSEKIFYDALGAGAIPLYYGNYMFDELSALPIEGPGGAYFDLKKR
Syngen	160	IKLGHAKHRSQDENSSIDLKSNFVFDIVVENCDAEWYVSEKIFYDALGAGCVPLYYGNYDELGLDPIEGD--VFYDLKKR
D3	238	NITGKQLQELIDTILSDERVEGMR-----
NY2A	238	NITGKQLQELIDMLSDERVEEMRKNVIDYREKVLRLVGTGMFAKKVEEATQLVKKTKKDIRLL
MT325	240	GIDTGAKLQELIDSLSDQLTDMRQNVAKYREAVLEFVGTKAFKAVEDATLST-----
FR483	240	GIDTGAKLQELIDSLSDQLTGMRQNVAKYREAVLEFVGTKAFKAVEDATLST-----
ATCV1	238	DIVSGRGVQELIDSLTDEDIATMRENVKKVREOVLEFAGSTQFAKSVERATELCETK-----
NE-JV-3	238	DIVSGRGVQELIDSLTDEDIATMRENVKKVREOVLEFAGSTQFAKSVERATELCETK-----
Syngen	238	GIEGELQELIDSLSDERVEEMRTNVNKKYREKVLRLVGTHTFAKKVEEAEVSKKVKKIVELM

Figure S3. Amino acid sequence alignment of A111/114R-D3 and other chlorovirus orthologs. Invariant and similar residues are highlighted in black and gray, respectively. NC64A-chloroviruses PBCV-1 (A111/114R-D3) and NY2A (B159R); Pbi-chloroviruses MT325 (M467R) and FR483 (N472R); SAG-chloroviruses ATCV1 (Z120R) and NE-JV-3 (AGE56873.1:551-845); and Syngen-virus 5 (YP_009325562.1:560-860). The predicted residue essential for enzyme activity, Arg-693, is denoted with an asterisk. Multiple alignment was performed by T-Coffee [31] using structural information and homology extension. File output was compiled by BOXSHADE.

Table S1. Glycosyltransferase homologs of A111/114R domains obtained using Phyre2 analysis.

	Ranking ¹	PDB code	Chain	Protein Name	Organism	Size ²	Confidence ³	% Coverage	% Identity	CAZy family ⁴
Domain 1	1	6BSV	B	Xyloglucan α -1,6-xylosyltransferase (XXT1)	<i>Arabidopsis thaliana</i>	460	96.2	61	19	GT34
	2	5GVV	F	Galactosyltransferase (GlyE)	<i>Streptococcus pneumoniae</i>	406	91.7	50	23	GT8
	3	6U4B	A	α -1,3-galactopyranosyltransferase / β -1,3-galactofuranosyltransferase (WbbM;RfbC)	<i>Klebsiella pneumoniae</i>	630	91.6	50	14	GT8/GT111
	4	1LL2	A	Glycogenin (Gyg)	<i>Oryctolagus cuniculus</i>	333	90.7	77	20	GT8
	5	1GA8	A	α -1,4-galactosyltransferase (LgtC)	<i>Neisseria meningitidis</i>	311	90.3	59	14	GT8
	6	4WMA	A	Xyloside xylosyltransferase 1 (MGC)	<i>Mus musculus</i>	392	90.2	82	12	GT8
Domain 2	1	2Z86	D	Chondroitin polymerase (K4CP)	<i>Escherichia coli</i>	686	100	71	18	GT2
	2	5TZ8	C	Teichoic acid β -N-acetylglucosaminyltransferase	<i>Staphylococcus aureus</i>	573	100	97	15	GT2
	3	1XHB	A	α -N-acetylgalactosaminyltransferase (T1)	<i>Mus musculus</i>	608	100	89	13	GT27
	4	6H4M	A	ss-1,3-N-acetylglucosaminyltransferase (SA1808)	<i>Staphylococcus aureus</i>	327	100	84	16	GT2
	5	6E4R	B	N-acetylgalactosaminyltransferase 9	<i>Drosophila melanogaster</i>	647	100	89	16	GT27
	6	6S24	A	N-acetylgalactosaminyltransferase 3	<i>Taeniopygia guttata</i>	574	100	89	14	GT27
Domain 3	1	2NZW	A	α -1,3-fucosyltransferase	<i>Helicobacter pylori</i>	405	100	86	15	GT10
	2	4W6Q	C	Glucosyltransferase	<i>Streptococcus agalactiae</i>	401	93.1	43	15	GT8
	3	3RHZ	B	Glucosyltransferase 3	<i>Streptococcus parasanguinis</i>	449	90.6	29	16	GTnc ⁵

¹Models are ranked according to raw alignment score accounting sequence and secondary structure similarities

²Number of amino acids

³Confidence is the probability that the A111/114R-D1 sequence and the template are homologous

⁴CAZy Carbohydrate Active Enzyme website: www.cazy.org

⁵GTnc: glycosyltransferase family “not classified”