



Structural Basis and Evolution of Glycan Receptor Specificities within the Polyomavirus Family

Luisa J. Ströh,^{a*} Nils H. Rustmeier,^a Bärbel S. Blaum,^a Josephine Botsch,^a Philip Rößler,^a Florian Wedekink,^a W. Ian Lipkin,^b Nischay Mishra,^b  Thilo Stehle^{a,c}

^aInterfaculty Institute of Biochemistry, University of Tübingen, Tübingen, Germany

^bCenter for Infection and Immunity, Columbia University, New York, New York, USA

^cDepartment of Pediatrics, Vanderbilt University School of Medicine, Nashville, Tennessee, USA

ABSTRACT Asymptomatic infections with polyomaviruses in humans are common, but these small viruses can cause severe diseases in immunocompromised hosts. New Jersey polyomavirus (NJPyV) was identified via a muscle biopsy in an organ transplant recipient with systemic vasculitis, myositis, and retinal blindness, and human polyomavirus 12 (HPyV12) was detected in human liver tissue. The evolutionary origins and potential diseases are not well understood for either virus. In order to define their receptor engagement strategies, we first used nuclear magnetic resonance (NMR) spectroscopy to establish that the major capsid proteins (VP1) of both viruses bind to sialic acid in solution. We then solved crystal structures of NJPyV and HPyV12 VP1 alone and in complex with sialylated glycans. NJPyV employs a novel binding site for a α 2,3-linked sialic acid, whereas HPyV12 engages terminal α 2,3- or α 2,6-linked sialic acids in an exposed site similar to that found in *Trichodysplasia spinulosa*-associated polyomavirus (TSPyV). Gangliosides or glycoproteins, featuring in mammals usually terminal sialic acids, are therefore receptor candidates for both viruses. Structural analyses show that the sialic acid-binding site of NJPyV is conserved in chimpanzee polyomavirus (ChPyV) and that the sialic acid-binding site of HPyV12 is widely used across the entire polyomavirus family, including mammalian and avian polyomaviruses. A comparison with other polyomavirus-receptor complex structures shows that their capsids have evolved to generate several physically distinct virus-specific receptor-binding sites that can all specifically engage sialylated glycans through a limited number of contacts. Small changes in each site may have enabled host-switching events during the evolution of polyomaviruses.

IMPORTANCE Virus attachment to cell surface receptors is critical for productive infection. In this study, we have used a structure-based approach to investigate the cell surface recognition event for New Jersey polyomavirus (NJPyV) and human polyomavirus 12 (HPyV12). These viruses belong to the polyomavirus family, whose members target different tissues and hosts, including mammals, birds, fish, and invertebrates. Polyomaviruses are nonenveloped viruses, and the receptor-binding site is located in their capsid protein VP1. The NJPyV capsid features a novel sialic acid-binding site that is shifted in comparison to other structurally characterized polyomaviruses but shared with a closely related simian virus. In contrast, HPyV12 VP1 engages terminal sialic acids in a manner similar to the human *Trichodysplasia spinulosa*-associated polyomavirus. Our structure-based phylogenetic analysis highlights that even distantly related avian polyomaviruses possess the same exposed sialic acid-binding site. These findings complement phylogenetic models of host-virus codivergence and may also reflect past host-switching events.

KEYWORDS polyomavirus, evolution, glycan receptor, host-pathogen interactions, sialic acid, virus tropism

Citation Ströh LJ, Rustmeier NH, Blaum BS, Botsch J, Rößler P, Wedekink F, Lipkin WI, Mishra N, Stehle T. 2020. Structural basis and evolution of glycan receptor specificities within the polyomavirus family. *mBio* 11:e00745-20. <https://doi.org/10.1128/mBio.00745-20>.

Invited Editor Colin R. Parrish, Cornell University

Editor Carolyn B. Coyne, University of Pittsburgh School of Medicine

Copyright © 2020 Ströh et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Thilo Stehle, thilo.stehle@uni-tuebingen.de.

* Present address: Luisa J. Ströh, Institute of Virology, Hannover Medical School, Hannover, Germany.

Received 30 March 2020

Accepted 22 June 2020

Published 28 July 2020

Polyomaviruses are small double-stranded DNA (dsDNA) viruses with a circular genome of approximately 5,000 bp. More than 100 distinct polyomaviruses have been found in samples taken from diverse mammals and birds, as well as fish (1) and invertebrates (2), and the discovery of new polyomaviruses continues (3). Of these identified viruses, 14 have been proposed to be human-tropic, but only a subset has so far been directly associated with diseases in immunocompromised individuals. The JC polyomavirus (JCPyV) is the causative agent of progressive multifocal leukoencephalopathy (PML), a fatal brain disease (4), whereas BK polyomavirus (BKPyV) causes polyomavirus-associated nephropathy (PVAN) and hemorrhagic cystitis in renal transplant recipients (5). Merkel cell polyomavirus (MCPyV) plays a key role in the development of a rare form of skin cancer, Merkel cell carcinoma, and human polyomavirus 7 (HPyV7) and *Trichodysplasia spinulosa*-associated polyomavirus (TSPyV) have been linked to diseases such as thymic and lymphoid cancers and nonmalignant skin dysplasias, respectively (6–8). The recent discovery of new family members has heightened the interest in polyomaviruses and their potential contributions to human disease, in particular cancer (9–11).

To mount an infection, virions must adhere to molecules presented on the surface of the target host cell. Such molecules have been identified for many polyomaviruses, and they include gangliosides (12–15), glycoproteins (16), and glycosaminoglycans (GAGs) (17, 18). However, some polyomaviruses use other, still unknown attachment factors (19, 20). All glycan-binding polyomaviruses analyzed to date engage oligosaccharide (glycan) structures featuring terminal sialic acids that in mammals are usually presented either on gangliosides, which are molecules composed of a glycosphingolipid, i.e., a ceramide and an oligosaccharide, or on glycoproteins, which contain mostly complex and branched oligosaccharides attached to amino acids during co- and posttranslational modification steps. These sialylated glycan structures interact with loop regions at the outer margin of the major capsid protein VP1, at the apical surface of the polyomavirus capsid, as revealed by X-ray crystallography (14, 15, 21–26) and cryoelectron microscopy (27).

Sialic acids comprise a family of more than 50 naturally occurring carbohydrates that are derivatives of the nine-carbon pyranose neuraminic acid (5-amino-3,5-dideoxy-D-glycero-D-galactononulosonic acid) (28). In humans, ganglioside and glycoproteins are commonly capped with 5-*N*-acetyl neuraminic acid (Neu5Ac) in α 2,3, α 2,6, or α 2,8 linkages (28). The Neu5Ac moiety is central for the recognition of the polyomavirus-glycan receptor. The VP1 surface exhibits an astonishing plasticity, providing virus-specific interactions for a range of Neu5Ac linkage types and further specific characteristics of Neu5Ac-containing glycans (29). Satellite residues outside a core binding pocket are often critical for establishing specificities for a single glycan receptor motif (29), Neu5Ac linkage (21), or even different sialic acid modifications (24, 30). For example, a single amino acid mutation in BKPyV VP1 is sufficient to switch the viral receptor specificity to the specificity of the related simian virus 40 (SV40) *in vitro* and in cell culture (14). A small number of mutations in the GM1 binding site of SV40 alters ganglioside usage, resulting in a change in tropism and suggesting that VP1 divergence might be driven primarily by a requirement to interact with certain receptors (31). In contrast, molecular details for interactions of polyomaviruses with the nonsialylated GAGs are still missing. GAGs are long linear polysaccharides consisting of repeating disaccharide units that are usually highly sulfated. For example, MCPyV attaches during early cell binding events to GAGs, followed by secondary interactions with a sialylated entry cofactor (17). JCPyV uses GAGs, which are also present in oligodendrocytes and astrocytes, as alternative entry pathway attachment receptors, and similarly, BKPyV virus-like particles, but not VP1 pentamers, can employ GAGs to attach to target cells (18), suggesting that GAGs bind in regions where neighboring pentamers interact within an assembled virion. A recent cryoelectron microscopy study showed additional electron density after treatment of BK virions with GAGs on top of a patch of positively charged surface in the grooves between VP1 pentamers, but the resolution of the obtained map was not sufficient to identify details of the interaction (27).

Although evidence is emerging that coreceptor or pseudoreceptor molecules as well as downstream events also significantly contribute to virus entry upon carbohydrate receptor binding (32), virus-receptor interactions are major determinants of polyomavirus host and tissue tropism (14).

New Jersey polyomavirus-2013 (NJPyV) and human polyomavirus 12 (HPyV12) are two of the most recently identified human polyomaviruses (33, 34). In both cases, a close phylogenetic relationship to animal polyomaviruses was found, raising questions about the origin and evolution of these two viruses. NJPyV was first discovered through muscle biopsy in an organ transplant recipient with systemic vasculitis, myositis, and retinal blindness (33). Histopathology analysis suggested a tropism for endothelial cells, and thus, NJPyV may have contributed to muscle and ocular damage in the patient. However, it remained unclear if the discovery reflected merely a single human transmission event (33). A recent serological study with serum samples from 706 Italians suggests seroprevalence rates of up to 57.5% in the elderly (35), while very low seroprevalence rates of about 5% in samples from 1,050 Dutch blood donors (36) and of 1.8% in Japan (37) have been detected. NJPyV is closely related by sequence to chimpanzee polyomavirus (ChPyV), and indeed, its existence had been predicted in a serological study as a potential human homolog to ChPyV prior to its discovery (38).

HPyV12 was first described in a study focusing on the identification of new human polyomaviruses in the gastrointestinal tract, spleen, and lymph nodes (34). The virus likely represents an early offshoot of a large and diverse clade comprising polyomaviruses from apes, bats, monkeys, rodents, and humans (MCPyV and TSPyV) in the initial phylogenetic analysis based on the polyomavirus large T antigen (LT) (34). However, other VP1-based analyses suggested close relationships of HPyV12 with sheep polyomavirus (ShPyV) found in commercially available sheep meat (2) and with avian viruses such as budgerigar fledgling disease polyomavirus (BFDPyV) and finch polyomavirus (FiPyV) (3, 39). More recently, new polyomaviruses possessing especially high sequence identities with HPyV12 were detected in shrews and nutria, and thus, HPyV12 may be a variant of a nonhuman polyomavirus that naturally infects these animals (40, 41). In contrast, a recent serological study with peak seroprevalence rates of 97.3% (35) indicates that HPyV12 or a related virus circulates widely in humans, outranging earlier and more recently reported seroprevalence rates of 20% (34) and 4% (36), respectively.

Regardless of their origins, the recent identification of NJPyV, HPyV12, and other polyomaviruses has raised doubts about earlier proposed evolutionary models and has triggered new phylogenetic approaches (2, 3, 42–44). In 2016, criteria for the definition of polyomavirus species based on sequences coding for LT were introduced, leading to a new classification system that has four genera: *Alpha-*, *Beta-*, *Gamma-*, and *Delta-polyomavirus* (45). Many polyomaviruses clustering in phylogenetic trees based on the whole genome or confined parts of the genome are found in distinct but closely related hosts (43), including several “human-simian pairs.” Such pairs include, for example, (i) human JCPyV and simian agent 12 (SA12) (46), (ii) human BKPyV and SV40 (46), (iii) human polyomavirus 9 (HPyV9) and African green monkey-derived lymphotropic polyomavirus (LPyV) (46, 47), and (iv) TSPyV and orangutan polyomavirus (OUPyV) (25).

A recent study suggested that host-switching events had been so far generally underestimated for dsDNA viruses and should be considered in revised evolutionary models for polyomaviruses (48). Indeed, classic reservoirs for zoonotic viruses, including primates, bats, and rodents, are highly populated by polyomaviruses, and zoonotic transmissions have been discussed for NJPyV (33) and HPyV12 (40), as well as for the most recently identified human polyomavirus, Lyon IARC polyomavirus (LIPyV), which is related to the putatively oncogenic raccoon polyomavirus (49). However, to date, the two avipolyomaviruses BFDPyV and goose hemorrhagic polyomavirus (GhPyV) and bat PyV species are the only known examples of polyomaviruses that can infect hosts belonging to multiple species and are thus capable of a host switch (50–52).

To define the structural requirements for receptor binding within the growing polyomavirus family and to connect these with the recently proposed evolutionary models, we established the glycan binding properties of the VP1 proteins of NJPyV,

HPyV12, and related animal viruses. While NJPyV and HPyV12 specifically bind sialylated glycans, their binding sites differ significantly in location and architecture. NJPyV VP1 specifically engages Neu5Ac that is α 2,3-linked to galactose (Gal) with a novel binding site that differs in location from the sites in all other structurally characterized sialic acid-binding polyomaviruses. Structural analyses indicate that the Neu5Ac-binding site of NJPyV is conserved in the closely related ChPyV.

In contrast, the Neu5Ac-binding site of HPyV12 is similar in location and glycan specificity to that of TSPyV (25) and three animal polyomaviruses, ShPyV, GhPyV, and FiPyV. It is therefore clear that HPyV12, TSPyV, ShPyV, GhPyV, and FiPyV (and likely additional members of the polyomavirus family) share a core glycan-binding strategy that is conserved across mammalian and nonmammalian species.

Our results expand the number of physically distinct sialic acid-binding sites in polyomaviruses and show that structure-based phylogenetic analyses may help our understanding of factors that modulate virus-carbohydrate receptor interactions, polyomavirus evolution, and potential host-switching events in the past and in the future.

RESULTS

Overall structures of NJPyV and HPyV12 VP1. In order to enable an understanding of their structural features and receptor-binding strategies, we expressed and purified VP1 pentamers of NJPyV and HPyV12 and solved X-ray structures of both native and liganded proteins (see Materials and Methods) (Table 1). The two pentamers lack N- and C-terminal residues, which renders them assembly incompetent but does not affect the structure of the VP1 surface loops and binding of sialylated ligands (53). Complexes were prepared by soaking crystals with α 2,3-linked sialyllactose (3'SL) for NJPyV (Fig. 1D) and with α 2,3-linked sialyllactosamine (3'SLN) for HPyV12 (Fig. 2D). Both trisaccharides feature a terminal Neu5Ac moiety and thus represent terminal oligosaccharide structures usually included in more-complex branched glycans on glycoproteins. The 3'SL glycan is, for example, also found linked to a ceramide in the context of various gangliosides. All structures were solved to high resolution, ranging from 2.3 Å to 1.55 Å, with excellent refinement statistics (Table 1). As expected, both proteins feature the conserved jelly-roll fold typical for polyomavirus VP1. This fold includes two antiparallel β -sheets, B, I, D, and G2 and C, H, E, and F (Fig. 1A and Fig. 2A), which are connected by variable surface loops named after the β -strands connected by them (12, 14, 15, 19–25, 53, 54). The outer surface of a polyomavirus is exclusively defined by the BC, DE, EF, and HI loops, and it is therefore not surprising that the two viruses engage glycans via binding sites formed by the four loops. The long BC loop emanates from the top of VP1 in two different directions, which are referred to as the BC1 and BC2 loops.

Structure of NJPyV bound to 3'SL. NJPyV VP1 binds the α 2,3-Neu5Ac-Gal motif of 3'SL in a binding site that is located between the HI and BC loops of one VP1 monomer (Fig. 1C). The DE loop, which is elongated in NJPyV VP1 compared to other structurally known polyomavirus VP1 pentamers, is not involved in any interactions with the glycan. Instead, side chains of residues K296 and N299 from the HI loop and residue H80 of the BC2 loop engage the Neu5Ac carboxylate group. Residue E77 interacts with the Neu5Ac C-4 hydroxyl, and the *N*-acetyl group forms a water-mediated hydrogen bond with the backbone of residue V78. Additionally, hydrophobic interactions between the *N*-acetyl methyl group, residue L90, and the hydrophobic part of the E77 side chain are seen, but the Neu5Ac glycerol chain is facing into solution and is not engaged by VP1. The Gal ring interacts with residue D298 via a hydrogen bond formed to the C-4 hydroxyl and via hydrophobic interactions of its C-6 atom. Like most polyomaviruses, NJPyV VP1 does not undergo any rearrangements upon glycan binding (root mean square deviation [RMSD] of 0.2 Å for the structural superposition of native and liganded NJPyV VP1 C α atoms [55]).

NJPyV VP1 binds specifically to α 2,3-linked sialic acids in solution. In order to confirm the observed interactions and to probe for specificity of binding, we examined

TABLE 1 Crystallographic data collection and refinement statistics for native and complexed NJPyV and HPyV12 VP1 structures

Parameter ^a	Value for indicated structure ^b			
	NJPyV VP1 (6Y5X) ^c	NJPyV VP1-3'SL (6Y5Y)	HPyV12 VP1 (6Y5Z)	HPyV12 VP1-3'SLN (6Y60)
Data collection statistics				
Space group	P2 ₁	P2 ₁	P2 ₁	P2 ₁ ,2 ₁
Unit cell dimensions <i>a</i> , <i>b</i> , <i>c</i> (Å)	86.48, 151.76, 130.91	86.38, 151.08, 130.62	61.22, 136.42, 85.38	83.32, 141.72, 251.30
β (°)	106.85	106.56	109.57	90
Resolution range (Å)	48.30–2.30 (2.36–2.30)	48.20–1.80 (1.91–1.80)	44.10–1.55 (1.64–1.55)	47.40–1.80 (1.91–1.80)
No. of unique reflections	143,046 (10,546)	295,110 (47,350)	181,696 (28,058)	274,251 (42,872)
Total no. of reflections	828,661 (60,401)	1,695,688 (255,694)	689,847 (106,706)	1,930,655 (294,756)
R_{meas} (%)	25.1 (104.5)	15.6 (106.4)	5.6 (45.6)	12.5 (177.2)
$I/\sigma I$	6.7 (1.6)	9.8 (1.6)	16.4 (2.9)	12.2 (1.3)
$CC_{1/2}$ (%)	98.4 (63.8)	99.6 (66.6)	99.9 (85.5)	99.9 (72.5)
Completeness (%)	99.9 (99.9)	99.8 (99.3)	95.0 (90.9)	99.4 (97.0)
Wilson B-factors (Å ²)	29.3	26.0	24.2	33.3
Refinement statistics				
$R_{\text{work}}/R_{\text{free}}$ (%)	18.9/24.4	15.9/19.6	14.3/16.9	20.4/24.3
No. of atoms				
Protein	21,197	21,615	10,318	19,809
Water	839	2,796	1,547	1,829
Glycan		430		287
B-factor (Å ²)				
Protein	29.8	23.3	17.5	33.3
Water	24.4	33.4	30.4	37.5
Glycan		31.9		41.1
RMSD				
Bond length (Å)	0.008	0.010	0.009	0.009
Bond angle (°)	1.578	1.656	1.548	1.629

^a R_{meas} , data redundancy-independent R-factor; $I/\sigma I$, intensity-to-noise ratio; $CC_{1/2}$, half-set correlation coefficient; R_{work} , refinement R-factor of work set; R_{free} , refinement R-factor of test set.

^bValues in parentheses correspond to the highest data resolution shell.

^cPDB accession numbers are given in parentheses after the structure designations.

the interactions of 3'SL and related glycans with NJPyV VP1 pentamers in solution using saturation transfer difference nuclear magnetic resonance (STD NMR) measurements (56, 57). This technique has been used to study similar protein-glycan interactions that are typically characterized by weak binding affinities (57, 58). It is based on the fast cross-relaxation of excited states in large molecules and the intermolecular nuclear Overhauser effect (NOE) between protons of a protein and protons of a ligand within a complex with a fast off-rate. In short, after selective excitation of proton resonances of the protein, ligand protons at distances of up to 5 Å from the protein protons receive saturation transfer from the protein during the lifetime of the complex through intermolecular NOEs and are observed in the STD NMR difference spectrum.

The STD NMR experiment showed that NJPyV VP1 pentamers specifically bind 3'SL but not 6'SL in solution (Fig. 1E to G). This is very much consistent with the crystal structure of the NJPyV-3'SL complex, which elucidated interactions with both Neu5Ac and Gal, suggesting that a defined orientation (or linkage) between the two sugars is required for the interaction. In our crystal structure, Gal forms fewer interactions than Neu5Ac and contributes only about 30% of the total interface area buried upon 3'SL binding. Likewise, in the NMR spectrum, only resonances belonging to the Neu5Ac moiety of 3'SL are observed, and the well-dispersed resonances for protons H1 and H3 of the Gal or protons of the terminal Glc are not visible in the STD NMR difference spectrum. The interactions of α 2,3-linked Gal with VP1 nevertheless appear critical for

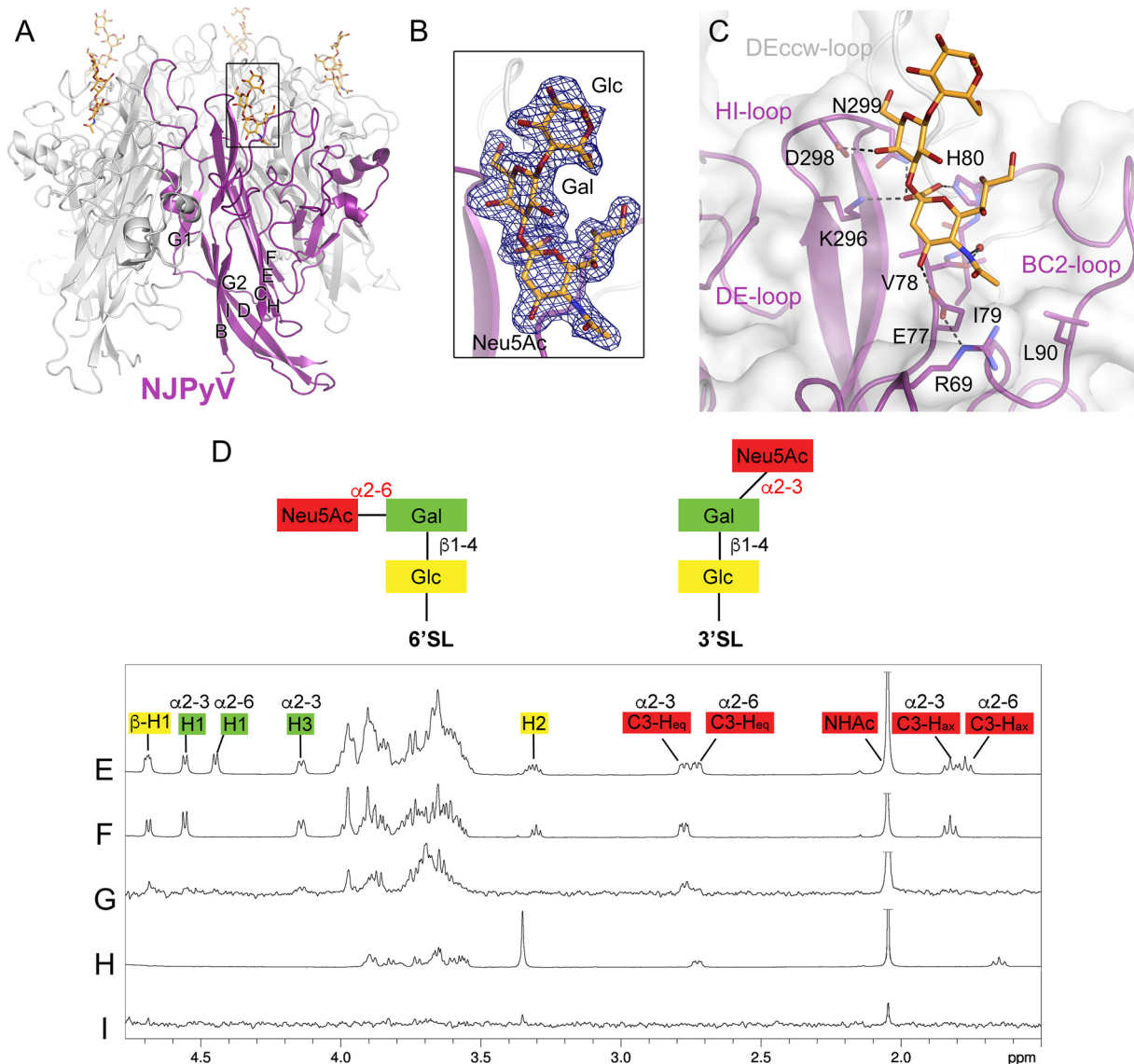


FIG 1 NJPyV VP1 engages the Neu5Ac- α 2,3-Gal motif of 3'SL. (A) Structure of the NJPyV VP1 pentamer structure in complex with the 3'SL glycan. One VP1 monomer is highlighted in purple, the others are colored in gray. Glycans are shown in stick representation with carbons in orange, oxygens in red, and nitrogens in blue. (B) The simulated annealing difference electron density map for the 3'SL is contoured at a σ level of 2.5 and is shown with a radius of 2 Å around the ligand. (C) Interactions of NJPyV with 3'SL. Side chains of VP1 in the binding site are shown as sticks, and water molecules are shown as a red sphere. Direct and water-mediated interactions are indicated with dashed lines. (D) Schematic representation of the glycans, 3'SL and 6'SL, used for NMR experiments. Glc, glucose; Gal, galactose; Neu5Ac, *N*-5-acetyl neuraminic acid. (E) 1H NMR reference spectrum of 27 μ M NJPyV VP1 with 1 mM 6'SL and 1 mM 3'SL. The equatorial H3 resonance of α 2,3-linked-Neu5Ac is slightly shifted with respect to the same resonance of α 2,6-Neu5Ac. (F) 1H NMR reference spectrum of 1 mM 3'SL. (G) STD NMR difference spectrum of the same sample as shown in panel D. (H) 1H NMR reference spectrum of 2 mM 2-*O*-methyl sialic acid and 27 μ M NJPyV VP1. (I) STD NMR difference spectrum of the same sample as shown in panel G. Neu5Ac amide methyl group resonances are truncated.

ligand recognition, as isolated sialic acid, represented by the compound 2-*O*-methyl- α Neu5Ac, clearly did not interact with NJPyV VP1 in solution (Fig. 1H and I).

Structure of HPyV12 bound to 3'SLN. In the crystal structure of the HPyV12 VP1 pentamer in complex with 3'SLN, the Neu5Ac moiety of the linear trisaccharide is engaged in a binding site mainly formed by the BC2 loop of one VP1 monomer (Fig. 2A to C). The Neu5Ac *N*-acetyl group inserts into a shallow cavity, forming a hydrogen bond between its amide group and residue T69. Additional, nonpolar interactions involve the methyl group and residues P79 and H128 of the counterclockwise (ccw) VP1 monomer. Finally, the *N*-acetyl carbonyl oxygen forms direct and water-mediated hydrogen bonds with the backbone amide groups of R78 and T72, respectively. The

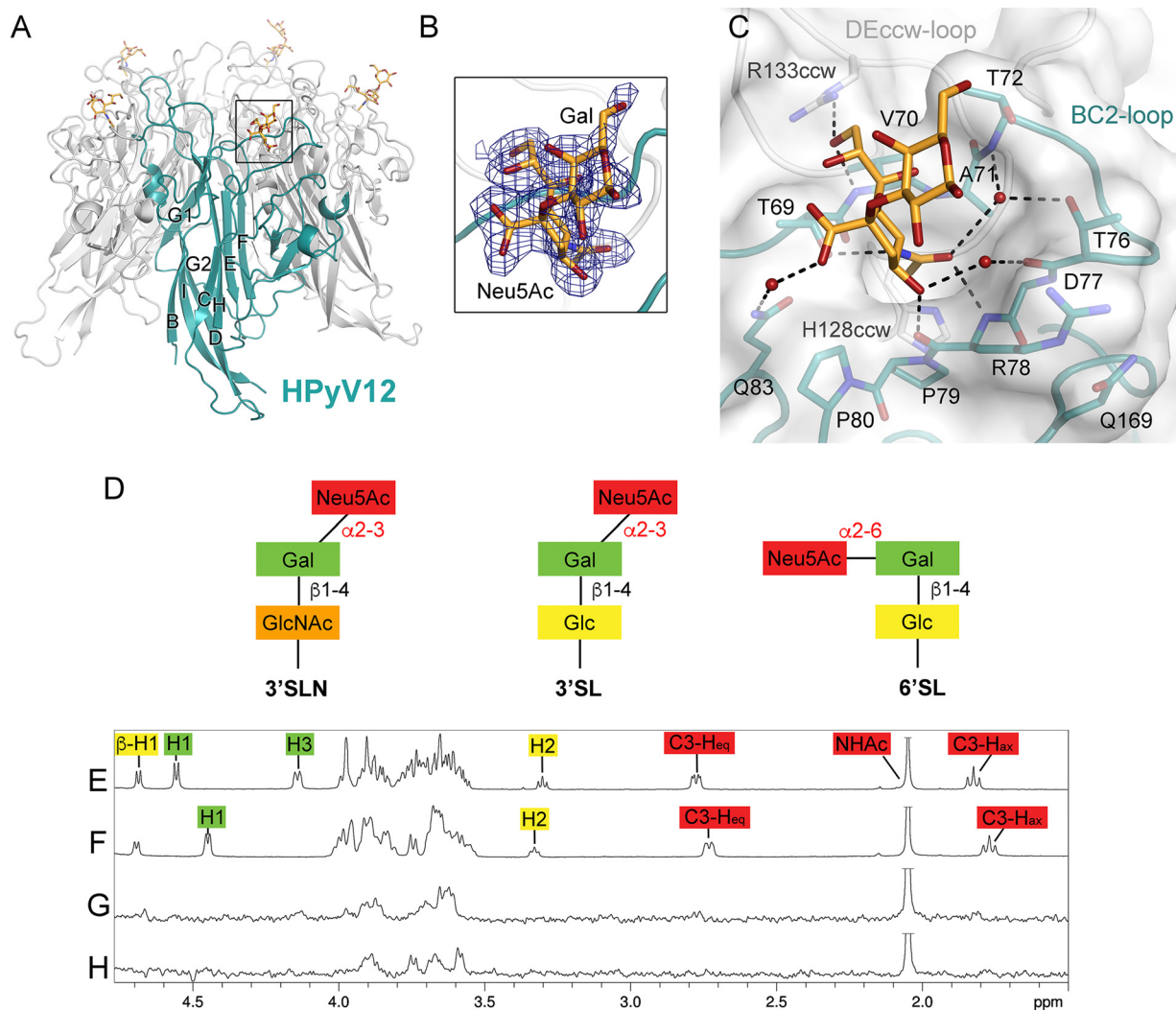


FIG 2 HPyV12 VP1 interacts with α 2,3-linked and α 2,6-linked sialic acids. (A) Structure of HPyV12 VP1 complexed with 3'SLN. One VP1 monomer of the VP1 pentamer is highlighted in deep teal. (B) A simulated annealing difference electron density map contoured at a σ level of 2.5 is displayed with a radius of 2 Å around the Neu5Ac-Gal motif of 3'SLN. (C) Close-up view of the binding site for 3'SLN. (D) Schematic representation of glycans used for X-ray structure determination (3'SLN) and NMR experiments (3'SL and 6'SL). GlcNAc, *N*-acetylglucosamine; Glc, glucose; Gal, galactose; Neu5Ac, *N*-5-acetyl neuraminic acid. (E) ¹H NMR reference spectrum of 50 μ M HPyV12 VP1 with 2 mM 3'SL. (F) ¹H NMR reference spectrum of 50 μ M HPyV12 VP1 with 2 mM 6'SL. (G) STD NMR difference spectrum recorded with the same sample as shown in panel E. (H) STD NMR difference spectrum recorded with the same sample as shown in panel F. Neu5Ac amide methyl group resonances are truncated.

Neu5Ac glycerol chain forms hydrogen bonds with the side chain of R133 from the ccw DE loop and with the backbone amine of residue V70. On the other side of the BC2 loop, the Neu5Ac C-4 hydroxyl group interacts with the protein backbone. The Neu5Ac carboxylate group is not involved in any direct interactions with VP1, only forming a water-mediated hydrogen bond with the side chain of residue Q83. $C\alpha$ atoms of the BC2 loop (residues 67 to 82) from the liganded HPyV12 VP1 structure superpose with the respective residues of the native structure with an RMSD of 0.3 Å (55), indicating that the Neu5Ac docks into a preformed binding pocket within the BC2 loop and that no conformational changes occur upon ligand binding. Well-defined electron density was observed for the Neu5Ac ring in all 10 VP1 binding sites of the asymmetric unit, and in some sites the Gal ring could also be unambiguously built into the electron density. Although the Gal does not interact with VP1 via directed polar interactions, it is likely that conformational restraints of the Neu5Ac-Gal glycosidic bond in the bound state stabilize the Gal ring in a preferred orientation.

HPyV12 VP1 interacts with α 2,3-linked and α 2,6-linked sialic acid in solution.

The STD NMR analysis of HPyV12 VP1 reveals that the protein interacts with α 2,3- and

α 2,6-linked Neu5Ac (represented by 3'SL and 6'SL, respectively) in solution. In both cases, the interactions involve only the terminal Neu5Ac moiety common to both glycans, as no resonances for protons from Gal and Glc are observed (Fig. 2E to H). Again, these observations are in excellent agreement with the HPyV12 crystal structure, which shows that the Neu5Ac moiety of 3'SLN mediates contacts with VP1 while the remaining portion of the glycan faces into solution and does not participate in any interaction with VP1 (Fig. 2C). The binding site could also clearly accommodate α 2,6-linked Neu5Ac, for instance in 6'SL.

Location of NJPyV and HPyV12 binding sites on VP1. In the NJPyV VP1-3'SL complex, the Neu5Ac ring is positioned roughly in between the locations seen for JCPyV VP1 and TSPyV VP1 (Fig. 3A), defining a novel polyomavirus Neu5Ac-binding site. The NJPyV VP1 sialic acid binding mode shares some similarity with MCPyV VP1 binding of sialylated glycans (Fig. 3B and C). Both proteins bind Neu5Ac in an almost identical orientation, although clearly not in the same location. In particular, the *N*-acetyl and carboxylate groups of Neu5Ac maintain multiple directed and nondirected interactions with VP1 of either virus, while the Neu5Ac glycerol chains project away from the respective protein surfaces. Despite the similar orientations, specific interactions are realized by nonhomologous amino acids as the NJPyV sialic acid-binding site is shifted toward the BC2 loop by about 5 Å (Fig. 3B and C). The interactions used by HPyV12 to engage terminal Neu5Ac are strikingly similar to those used by TSPyV (25) (Fig. 3A, D, and E). Superposition of the two complexes based on the VP1 chains brings the bound Neu5Ac moieties into almost perfect congruence (Fig. 3A). Moreover, key residues on both sides of the binding pockets are conserved, and the BC2 loop folds in a very similar manner on top of both VP1 proteins. Therefore, it is not surprising that the direct VP1-Neu5Ac interactions are also conserved between the two viruses (Fig. 3D and E), which share an overall VP1 amino acid sequence identity of 60%.

NJPyV and HPyV12 share basic glycan receptor-binding site characteristics with related nonhuman viruses. Phylogenetic analysis of the polyomavirus family based on the complete viral genome revealed that many human polyomaviruses are closely related to one or more simian polyomaviruses (10). These relationships are also present on the amino acid level of VP1. NJPyV shares amino acid sequence identities of about 84% with ChPyV isolates (ChPyV-Azzi, ChPyV-Bob, and ChPyV-Tanu) and about 73% with two other primate polyomaviruses, vervet monkey polyomavirus 1 (VmPyV1) and *Ptilocolobus rufomitratu*s polyomavirus 1 (PrufPyV1). All of these viruses belong to the *Alphapolyomavirus* genus. In contrast, the phylogenetic classification of HPyV12 is not very clear. A phylogenetic relationship to ShPyV with a VP1 amino acid sequence identity of 60% has been proposed (2), but also VP1 proteins of viruses from the distinct *Gammapolyomavirus* clade comprising the avipolyomaviruses GhPyV and BFDPyV and also FiPyV, crow polyomavirus (CPyV), butcherbird polyomavirus (BbPyV), and canary polyomavirus (CaPyV) exhibit amino acid sequence identities of about 58% with HPyV12 VP1. Moreover, a recent study found full-genome sequence identities of more than 86% between HPyV12 and shrew polyomaviruses and questioned if HPyV12 is a truly human-tropic virus (40). In order to rationalize the impact of VP1 differences on receptor specificity of closely related viruses from different species, we mapped amino acid differences on the surfaces of the NJPyV and HPyV12 VP1 pentamers (Fig. 4). Differences map predominantly to the exposed surface loops on top of VP1. Interestingly, the elongated DE loop, a prominent feature of NJPyV VP1, is unique to this human virus, whereas VP1 residues involved in the recognition of the Neu5Ac-Gal motif are highly conserved in ChPyV, VmPyV1, and PrufPyV1 (Fig. 4A and B). The Neu5Ac-binding region of HPyV12 VP1, formed primarily by the BC2 loop, is as expected mostly conserved, as is the rest of the VP1 surface for the shrew polyomaviruses *Sorex coronatus* polyomavirus 1 (ScorPyV1) and *Sorex araneus* polyomavirus 1 (SaraPyV1) (Fig. 4C and F). However, residues P79 and P80 at the base of the Neu5Ac binding site are not present in the shrew *Sorex minutus* polyomavirus 1 (SminPyV1) isolates (40). Furthermore, residues important for Neu5Ac recognition are conserved not only in

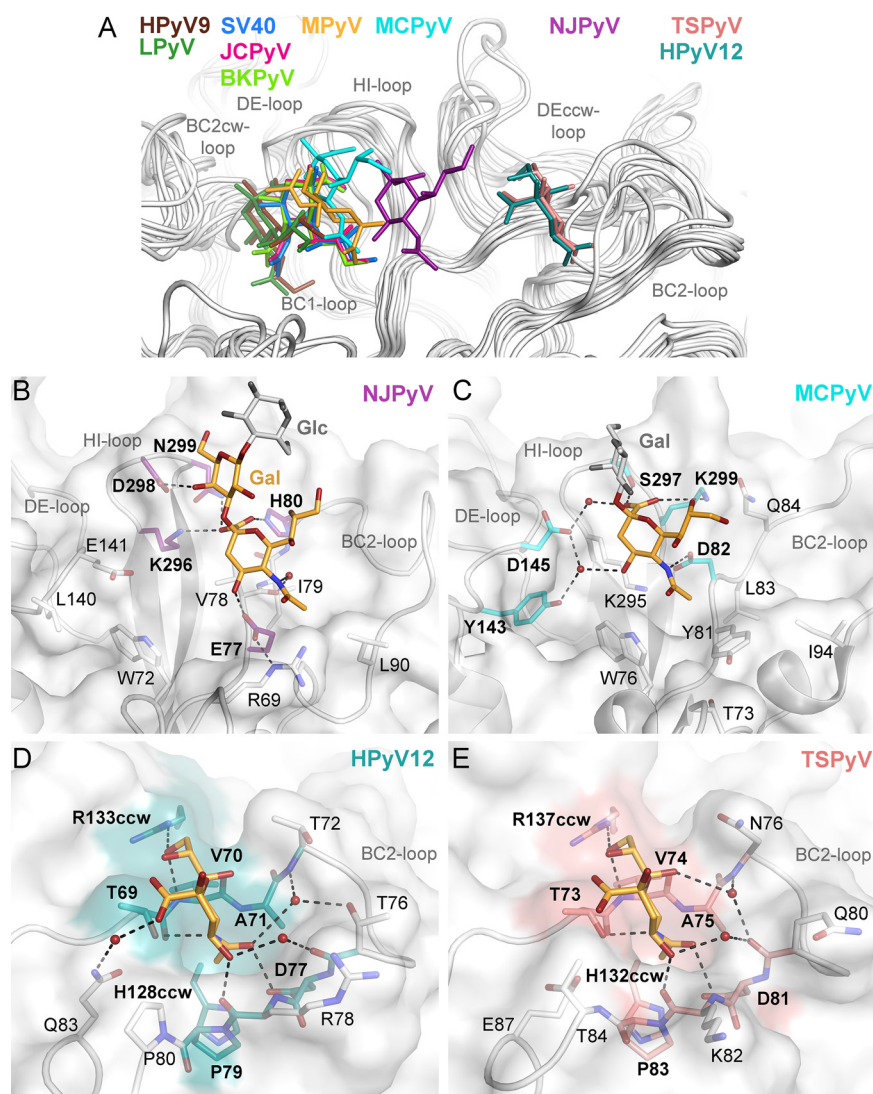


FIG 3 Glycan receptor-binding sites from HPyV12 and NJPyV compared to those from other structurally characterized polyomaviruses. (A) Structural superposition of polyomavirus VP1-glycan complex structures: HPyV12 VP1-3'SLN, NJPyV VP1-3'SL, SV40 VP1-GM1 glycan (PDB 3BWR), BKPyV VP1-GD3 glycan (PDB 4MJ0), JCPyV VP1-LSTc (PDB 3NXD), MCPyV VP1-GD1a glycan (PDB 4FMJ), LPyV VP1-3'SL (PDB 4MBY), MPyV VP1-Neu5Ac- α 2,3-Gal- β 1,3-[α 2,6-Neu5Ac]-GlcNAc- β 1,3-Gal- β 1,4-Glc (PDB 1VPS), HPyV9 VP1-3' Neu5Gc-SL (PDB 4POT), and TSPyV VP1-GM1 glycan (PDB 4U60). In the case of structures with more-complex glycans, only the Neu5Ac residues are shown. (B and C) Comparison of binding sites from NJPyV and MCPyV. Glycan residues not involved in intermolecular interactions are colored in gray. Direct and water-mediated hydrogen bonds between VP1 side chain residues and the terminal Neu5Ac are depicted as dashed black lines. Side chains of VP1 residues interacting with the glycan via hydrogen bonding or charged interactions are colored in purple and cyan for NJPyV and MCPyV, respectively. (D and E) Binding sites of HPyV12 and TSPyV VP1 are shown in the same orientation. VP1 complex structures were aligned using C α atoms and the secondary-structure matching (SSM) tool in Coot (76). Only terminal Neu5Ac residues are shown as orange sticks, with carbons in orange, oxygens in red, and nitrogens in blue. Direct and water-mediated hydrogen bonds between VP1 side chain residues and Neu5Ac are depicted as dashed black lines. Conserved VP1 residues are labeled in bold and are colored on the VP1 surface.

ShPyV VP1 (Fig. 4D) but also in the avian virus sequences, whereas most of the VP1 top surface differs (Fig. 4E). A broader VP1 sequence alignment reveals that six key residues of the BC2 and DE loop Neu5Ac-binding region (T69, V70, D77, P79, H128, and R133 in the case of HPyV12) are indeed strictly conserved across the avipolyomaviruses, ShPyV, EPyV, and TSPyV, in contrast to the rest of the residues at the VP1 surface (Fig. 4F). In contrast, these key residues are not conserved in a recently discovered animal polyo-

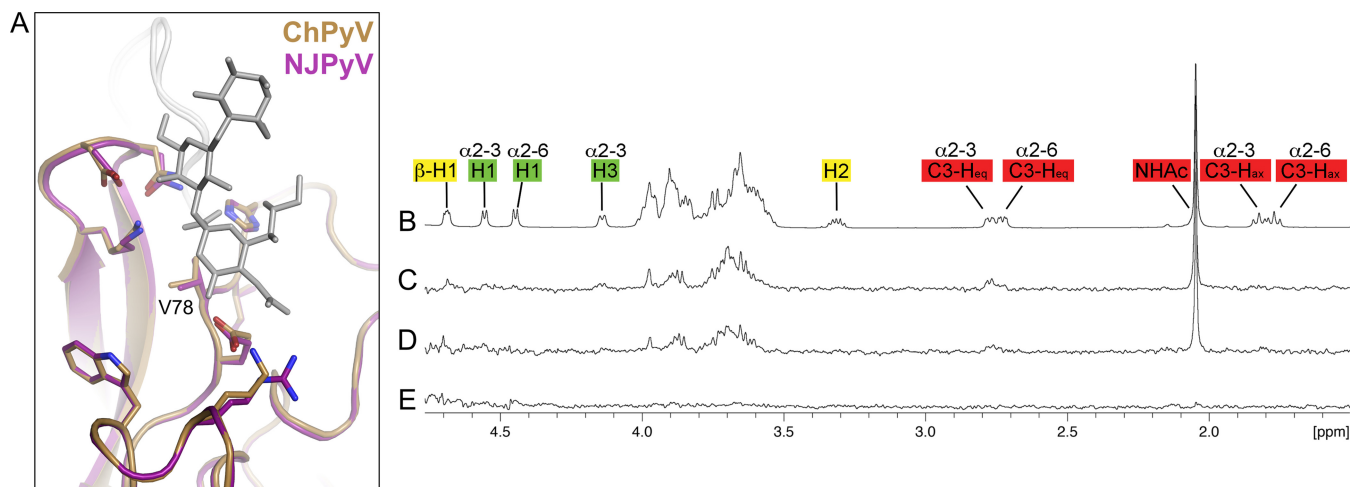


FIG 5 STD NMR spectroscopy of ChPyV VP1 and a putative Neu5Ac-negative ChPyV VP1 mutant. (A) Superposition of NJPyV VP1-3'SL (pink) and native ChPyV VP1 (gold) structures in cartoon representation. NJPyV VP1 side chains responsible for binding of 3'SL and respective ChPyV VP1 residues are shown as sticks, with oxygen and nitrogen atoms colored in red and blue, respectively. (B) ^1H NMR reference spectrum of $27\ \mu\text{M}$ NJPyV VP1 with a mix of $1\ \text{mM}$ 6'SL and $1\ \text{mM}$ 3'SL. (C) STD NMR difference spectrum of the same sample as shown in panel B. (D) STD NMR difference spectrum of $50\ \mu\text{M}$ ChPyV VP1 with the $1\ \text{mM}$ 6'SL- $1\ \text{mM}$ 3'SL mix. (E) STD NMR difference spectrum of $50\ \mu\text{M}$ ChPyV VP1 V78F with the $1\ \text{mM}$ 6'SL- $1\ \text{mM}$ 3'SL mix.

mavirus from nutria (*Myocastor coypus* polyomavirus 1 [McoyPyV1]), which clusters in the phylogenetic tree based on the complete amino acid LTA_g region (41).

Nevertheless, predictions of the surface loop architecture can be rather error-prone, especially with the low sequence conservation in the case of HPyV12, hampering modeling and mapping of glycan binding sites and the prediction of glycan receptor specificities based only on sequence information. Hence, we used two different approaches to test our hypothesis of the existence and conservation of the core Neu5Ac-binding sites of NJPyV and HPyV12 in the respective viruses. In both cases, we expressed and purified VP1 proteins of closely related nonhuman viruses, specifically, ChPyV for the comparison with NJPyV and ShPyV, GhPyV, and FiPyV for the comparison with HPyV12.

Interactions of ChPyV with Neu5Ac. As the Neu5Ac-binding site residues of NJPyV are highly conserved in ChPyV VP1, we hypothesized that this virus is also able to bind 3'SL. In order to define the ligand binding properties of ChPyV, we solved the unliganded structure of its VP1 pentamer at $1.9\ \text{\AA}$ resolution (see Table S2 in the supplemental material). With an RMSD of about $0.5\ \text{\AA}$, the ChPyV VP1 monomer superposes very well with a VP1 monomer of the NJPyV VP1-3'SL complex structure, including the region that binds sialic acid in NJPyV (Fig. 5A). The putative ligand-binding sites are largely blocked in the ChPyV VP1 crystals, preventing soaking with ligands. Cocrystallization with sialylated glycans did not yield well-diffracting crystals. Therefore, we assessed the binding of ChPyV VP1 to 3'SL by STD NMR spectroscopy and introduced a valine-to-phenylalanine mutation (V78F) (equivalent to amino acid residue V78 of NJPyV) in the putative sialic acid-binding site of ChPyV in order to interfere with potential glycan binding (Fig. 5B to E). While the STD NMR difference spectrum of wild-type ChPyV VP1 closely resembles that obtained for NJPyV VP1 (Fig. 5C and D), no saturation transfer was observed for the V78F mutant (Fig. 5E). We therefore conclude that ChPyV binds 3'SL in a manner that is identical to that observed for NJPyV. Hence, from our receptor-focused point of view, the two viruses can indeed be regarded as a closely related "human-simian pair."

Structures of VP1 proteins of the animal viruses ShPyV, GhPyV, and FiPyV. Our structure-based sequence analysis shows that the BC2 loop of HPyV12, which essentially mediates all interactions with Neu5Ac, is especially highly conserved in several animal viruses (Fig. 4). In order to probe whether these viruses are able to bind Neu5Ac in a similar manner, and whether differences in regions surrounding the binding site

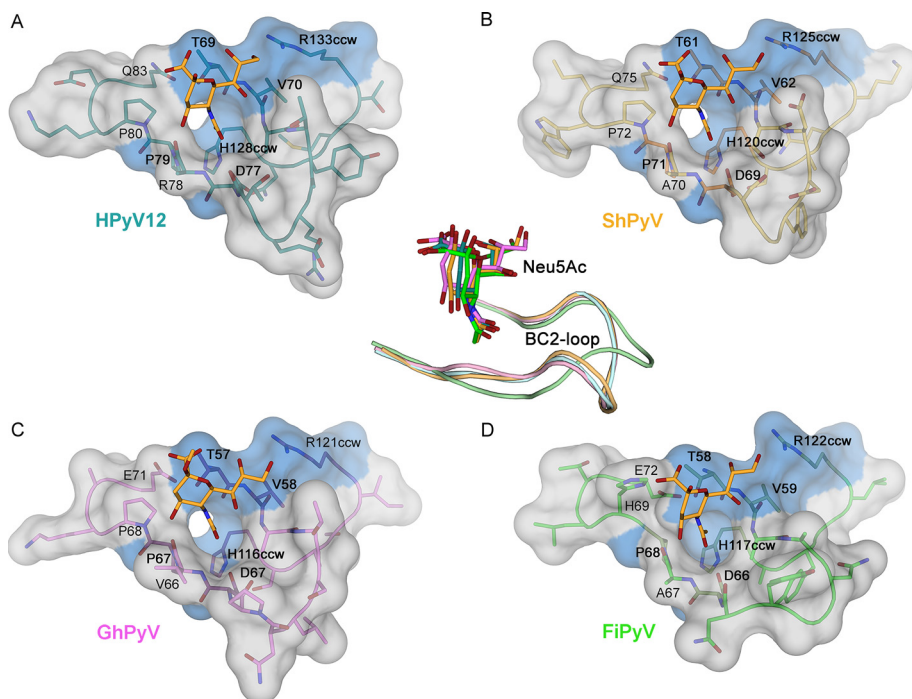


FIG 6 Engagement of Neu5Ac in a highly conserved binding site. Illustration of the BC2 loop binding sites from the complex structures of HPyV12 VP1-3'-SLN (A), ShPyV VP1-6'-SLN (B), GhPyV VP1-2-O-Me-Neu5Ac (C), and FiPyV VP1-2-O-Me-Neu5Ac (D) in the same orientation. Only the terminal Neu5Ac portions of the respective glycan structures are shown. Strictly conserved protein residues are distinguished by a blue surface. The middle panel displays a structural superposition of the four binding modes with the Neu5Ac coloring according to panels A to D.

might contribute to subtle differences in receptor engagement, we investigated the structures and ligand binding properties of three representative animal viruses from this alignment: the mammalian ShPyV, which was found in commercial sheep meat (2), and two avian polyomaviruses infecting geese (GhPyV) and finches FiPyV (39). GhPyV is a bona fide animal pathogen, as it causes hemorrhagic nephritis enteritis in geese (59).

We expressed and purified assembly-incompetent VP1 pentamers from all three viruses by following the protocol established for HPyV12 and solved their crystal structures at high resolution (Table S1). All three proteins fold into the familiar VP1 pentamer structure (53). Superpositions of the VP1 pentamers of ShPyV, GhPyV, and FiPyV with that of HPyV12 confirm that the core VP1 structure is highly conserved (RMSD values of 0.65 Å, 0.83 Å, and 0.73 Å, respectively) (55). There are subtle differences in surface loop regions, however. In HPyV12 VP1, the BC2 loop comprising the Neu5Ac-binding site is formed by amino acids T69 to Q83 (T73 to E87 in TSPyV VP1), whereas the putative binding sites in ShPyV VP1, GhPyV VP1, and FiPyV VP1 include amino acid residues T61 to Q75, T57 to E71, and T58 to E72, respectively (Fig. 6).

Complexes of ShPyV, GhPyV, and FiPyV VP1 with glycans. In order to define the ligand binding properties of the three animal viruses ShPyV, GhPyV, and FiPyV, we conducted crystal soaking experiments with sialylated glycans, which yielded additional electron density for the glycan ligand (Fig. S1) so that we could solve and refine complex structures for ShPyV VP1 with 6'-SLN and 3'-SLN, GhPyV VP1 with 2-O-Me-Neu5Ac, and FiPyV VP1 with 2-O-Me-Neu5Ac to 1.6 Å, 1.65 Å, 1.95 Å, and 2.8 Å resolution, respectively (Table S1).

All three proteins bind Neu5Ac in a shallow, conserved binding site above the BC2 loop (Fig. 6). The location of this site on the protein and the contacts are highly similar to those of HPyV12 VP1 (Fig. 6) and TSPyV. In fact, the site was first identified and shown to be functionally relevant in TSPyV (25), a human virus. Although a range of different

sialyloligosaccharides (such as α 2,3- and α 2,6-linked sialyllactosamine) were used to generate complexes, the complex structures show that only Neu5Ac forms interactions with the four proteins. The residues that mediate interactions with Neu5Ac are highly conserved among the four proteins.

In the case of ShPyV VP1, the Neu5Ac *N*-acetyl methyl group is situated in a shallow groove in the BC2 loop formed by residues T61, V62, S63, D69, A70, P71, P72, and H120ccw (Fig. 6B). Neu5Ac binding further involves the side chain of residue T61, which forms a hydrogen bond with the *N*-acetyl group amide. The *N*-acetyl carbonyl oxygen interacts with the backbone amide and the Neu5Ac C-4 hydroxyl group interacts with the backbone carbonyl oxygen of residue A70. The ligand's glycerol moiety interacts with the R125ccw side chain, forming a hydrogen bond between the Neu5Ac C-8 hydroxyl group and a guanidinium N of the arginine. In two of the binding sites, an additional interaction between the Neu5Ac C-9 hydroxyl group and the side chain of R131ccw is formed (not shown in Fig. 6B). The Neu5Ac carboxylate forms a hydrogen bond with the side chain oxygen of residue Q75. Additional carbohydrate monomers of the trisaccharide 6'SLN could be built into the electron density at four equivalent sites in other VP1 chains, showing electron density for Gal in three sites and for Gal-GlcNAc (lactosamine) in one site, respectively. These additional moieties do not, however, form any interactions with the protein and are most likely visible due to reduced glycosidic linkage flexibility in the context of crystal contacts. In the complex structure of ShPyV VP1 and 3'SLN, additional electron density allowed modeling of Gal in four of the six binding sites occupied by Neu5Ac; however, these galactose moieties do not contribute to the overall binding of the glycan to the protein. The interactions between the protein and the Neu5Ac part of 3'SLN are identical to the interactions observed in the 6'SLN complex structure.

GhPyV VP1 basically undergoes the same interactions with its glycan receptor, which in this case is represented by 2-*O*-Me-Neu5Ac (Fig. 6C). The shallow groove in the BC2 loop is formed by residues T57, V58, A59, D65, V66, P67, P68, and H116ccw. Again, residue T57 interacts with the Neu5Ac *N*-acetyl amide, and the backbone amide and carbonyl oxygen of residue V66 interact with the *N*-acetyl carbonyl oxygen and the Neu5Ac C-4 hydroxyl group, respectively. The carboxylate interacts with E71 via a hydrogen bond. The glycerol chain C-8 hydroxyl group forms a hydrogen bond with the side chain of R121ccw.

FiPyV VP1, which was also derivatized with 2-*O*-Me-Neu5Ac, forms interactions with Neu5Ac that are very similar to those observed in GhPyV and ShPyV VP1 (Fig. 6D). The methyl group of the *N*-acetyl moiety reaches into a shallow groove formed by residues T58, V59, A60, D66, A67, P68, E72, and H117ccw. The *N*-acetyl amide is recognized by the hydroxyl group of T58. VP1 residue A67 interacts with the Neu5Ac carbonyl oxygen and C-4 hydroxyl via its backbone atoms, as described above for ShPyV VP1. The R122ccw side chain interacts with the Neu5Ac glycerol moiety, contacting the hydroxyl groups of C-8 or C-9, depending on the VP1 chain. Residue H69 of the FiPyV VP1 contacts both the Neu5Ac carboxylate group and the E72 side chain, thereby providing for an additional hydrogen bond between the protein and the putative Neu5Ac-containing receptor.

In order to obtain insights into possible determinants of specificity that extend beyond direct contacts with Neu5Ac, we analyzed the structures and amino acid compositions of the four BC2 loop sequences in more detail. Structurally, the Neu5Ac-binding regions superimpose well. We calculated all-atom RMSD values with PyMOL using the BC2 loop of HPyV12 VP1 as a reference. With a value of 1.82 Å, the BC2 loop of FiPyV VP1 shows the largest deviation, while the BC2 loops of ShPyV and GhPyV VP1 yield lower deviations, of 0.77 Å and 0.84 Å, respectively. However, the observed structural conservation of the BC2 loop translates only partially to a conservation of residues surrounding the Neu5Ac-binding site. While the residues that mediate direct contacts with Neu5Ac (and are thus determinants of specificity for this portion of the receptor) are highly conserved, the remaining residues, even in the direct vicinity of Neu5Ac, display a much lower level of conservation (Fig. 4F).

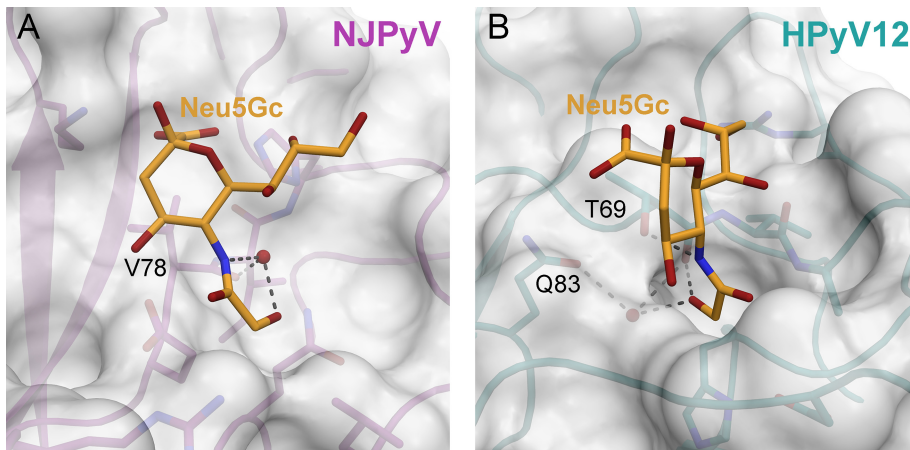


FIG 7 Models of putative Neu5Gc binding in NJPyV and HPyV12 VP1. Neu5Gc models are shown as sticks with carbons colored in orange, oxygens in red, and nitrogens in blue. NJPyV VP1 (A) and HPyV12 VP1 (B) are shown as cartoons with a transparent surface representation around the proteins. Putative hydrogen bonds involving the *N*-glycolyl group of Neu5Gc are depicted as dashed lines.

Putative recognition of Neu5Gc by human and animal VP1. Humans lack *N*-glycolylneuraminic acid (Neu5Gc) in their receptors due to a loss of the CMP-*N*-acetylneuraminic acid hydroxylase (CMAH) gene activity (60). However, Neu5Gc is a common component of cell surface glycans in many mammals, including nonhuman primates. In the case of the simian polyomavirus SV40, it could be demonstrated that ganglioside Neu5Gc-GM1 is a better receptor than its Neu5Ac counterpart (30), whereas the closely related human BK polyomavirus is unable to engage Neu5Gc-containing receptors (24). The human polyomavirus 9 (HPyV9) and the related B-lymphotropic polyomavirus (LPyV) of the African green monkey have counterintuitive binding properties, as the human virus displays a preference for the Neu5Gc receptor variants (22, 24). In order to assess the Neu5Gc binding capacities of the VP1 proteins introduced in this study, we prepared representative models of the Neu5Gc complexes of NJPyV and HPyV12 VP1 by replacing the Neu5Ac moiety with a Neu5Gc model within the complex X-ray structures (Fig. 7). These models demonstrate that the sialic acid-binding sites of both NJPyV and HPyV12 VP1 can accommodate Neu5Gc. In the NJPyV VP1 complex, the additional hydroxyl group of the *N*-glycolyl group would be able to establish a hydrogen bond with the water molecule that bridges the Neu5Gc amide nitrogen and the protein backbone at residue V78 (Fig. 7A). In HPyV12 VP1, the Neu5Ac-binding site possesses a recessed cavity with a hydrophilic bottom side that potentially prefers the extra hydroxyl group of Neu5Gc. Here, the hydroxyl could establish a hydrogen bond with a water molecule held by the side chain of Q83 and the backbone of T69 (Fig. 7A). The animal PyV species of chimpanzee and sheep, goose, and finch possess highly conserved binding pockets (Fig. 5A and 6, respectively) compared with their related human VP1 counterparts, and they therefore should be able to engage Neu5Gc as well. Mutations within the core binding sites, which would reflect evolutionary changes to preferentially accommodate Neu5Gc over Neu5Ac in the case of the animal PyVs, were not detected. However, Neu5Gc-favoring amino acid changes with rather long-range effects cannot be ruled out.

DISCUSSION

We solved the crystal structures of the VP1 proteins of the recently discovered polyomaviruses NJPyV and HPyV12 and established that the two viruses use significantly different binding site architectures to bind sialylated glycans. The NJPyV structure defines a new sialic acid-binding site in polyomaviruses, while HPyV12 uses a site that was previously observed in the human TSPyV. In each case, however, the respective binding mode is conserved in closely related animal polyomaviruses. For NJPyV,

this is demonstrated in solution by STD NMR spectroscopy of 3'SL binding to ChPyV VP1 and by the structure determination of unliganded ChPyV VP1. For HPyV12, we have solved crystal structures of three animal polyomavirus VP1 proteins (ShPyV, GhPyV, and FiPyV) that show essentially the same interactions and specificity for a terminal sialic acid. These results advance an understanding of virus-glycan specificities, and they also provide insight into the evolution of polyomaviruses.

Although many polyomaviruses engage sialylated receptors, extensive structure-function analyses have shown that there is remarkable plasticity in the recognition of sialic acids. The previously reported structures of polyomavirus complexes identified no less than five different binding modes of sialic acid. One is exemplified by JCPyV (21), BKPyV (14), and SV40 (15), the second by murine polyomavirus (26), a third by MCPyV (23), a fourth by LPyV (22) and HPyV9 (24), and a fifth by TSPyV (25). The five binding sites differ not only in their location on VP1 but also in the orientation of the bound sialic acid and the individual interactions that confer glycan specificity. The structure of NJPyV in complex with 3'SL extends this level of plasticity even further with yet another sialic acid-binding site that lies in between the sites observed in other polyomaviruses and that shares some similarities with respect to individual contacts with MCPyV. A common feature of the Neu5Ac binding sites of SV40, JCPyV, BKPyV, LPyV, and HPyV9 is a hydrophobic pocket, which is formed by residues of HI and BC1 loops from one VP1 monomer and BC2 loop residues of the clockwise (cw) VP1 monomer (14, 15, 21, 22, 24). The location of this hydrophobic pocket is not conserved in NJPyV VP1. Instead, BC2 loop residues E77 and L90 provide a similar hydrophobic environment for the methyl group of Neu5Ac on NJPyV VP1 in a different location.

The plasticity of polyomavirus-sialic acid interactions outlined in this and previous reports is, at least to our knowledge, unprecedented. While many viruses recognize sialylated glycans, they typically do so with a single binding site that is conserved across strains. A well-documented example is the influenza virus hemagglutinin (HA)-sialic acid interaction (61). While human and avian HAs bind to sialylated glycans that differ markedly in linkage type and overall conformation, all HA proteins essentially recognize the terminal sialic acid of the sialoglycan receptors in the same region and bind Neu5Ac in the same orientation (62). The variety of sialic acid-binding sites seen in polyomaviruses is particularly striking, as even identical glycans (such as 3'SLN) are bound in different locations by different polyomaviruses. In these cases, the viruses have evolved differing strategies to bind not only the Neu5Ac itself but also the glycosidic linker region, thus achieving specificity for one ligand with entirely different contacts. It is possible that this is a consequence of the actual physiological glycan receptors having more-complex structures that are not reflected by 3'SLN, perhaps branched glycans that necessitate binding in different orientations to either virus. Another possibility is that fine-tuning of affinity to a specific ligand can be realized only in a certain location.

The structural analyses of the HPyV12, ShPyV, GhPyV, and FiPyV VP1 proteins establish the existence of essentially the same sialic acid-binding site in viruses that infect different hosts. The same site also exists in the human-tropic TSPyV VP1, where it was shown to be required for cell attachment and infection (25). In all five cases, interactions with Neu5Ac are limited to a small set of highly conserved residues, and these residues are clearly sufficient to confer specificity for Neu5Ac. In contrast, the residues surrounding the Neu5Ac-binding site display a much lower level of conservation. One possible explanation is that there is no evolutionary pressure on the virus capsids to maintain conservation that goes beyond direct interactions with Neu5Ac because this pyranose is the sole binding determinant. However, given the narrow receptor specificities observed for other polyomaviruses, this seems unlikely: all other known polyomavirus-sialyloligosaccharide complexes show interactions that extend beyond the Neu5Ac ring, and it is precisely these additional interactions that determine (in some cases extraordinary) specificity for a particular sialylated glycan or a small range of structurally related glycans (12, 14, 15, 21–24). Another explanation is that TSPyV, HPyV12, ShPyV, GhPyV, and FiPyV utilize sialic acid-based receptors that contain sialic acid variants other than Neu5Ac. Interestingly, the BC2 loop binding cavity

features a hydrophilic bottom and would potentially prefer the extra hydroxyl group of Neu5Gc. However, Neu5Gc-favorable mutations within the core binding sites, which would suggest preferential binding of Neu5Gc over Neu5Ac in case of the animal PyVs, were not detected.

Certainly, it is conceivable that the residues surrounding the sialic acid-binding site contribute to the host tropism, perhaps by providing contacts with sialic acid variants or linkages that are preferentially expressed in a certain host and/or a certain tissue. In addition, the hemagglutinating activity of polyomaviruses has been attributed primarily to the binding of VP1 to sialylated glycans on the surface of human red blood cells (63–65). Conversely, the absence of hemagglutination activity for a given virus does not preclude glycan receptor binding, as shown for SV40 and its specific sialylated receptor, the GM1 ganglioside (13, 15). In the case of avian viruses BFDPyV, CaPyV, and GhPyV, hemagglutination activity has been reported with chicken and human erythrocytes, functionally supporting the sialylated glycan receptor binding properties we report for GhPyV (66, 67). However, FiPyV did not hemagglutinate chicken or human erythrocytes (66), suggesting an alternative receptor specificity compared to that of the related viruses, likely influenced by residues surrounding the conserved BC2 loop Neu5Ac-binding site.

Additionally, unknown nonsialylated coreceptors may likewise impact the host tropism.

The identification of shrew polyomaviruses (40) and McoyPyV1 in nutria (41), which are closely related in sequence to HPyV12, and studies reporting low seroprevalence rates and seroactivity for HPyV12, NJPyV, and LIPyV questioned the tropism of recently discovered human polyomaviruses (36). Further analyses are needed, but the presence of a functional core BC2 loop sialic acid-binding site in TSPyV nonetheless demonstrates the conservation of this site in a human-tropic virus (25).

Past studies suggest coevolution of polyomaviruses with their hosts (2, 68) and exclude host-switching events between members of distantly related species. Nevertheless, ancient or sporadic viral transmission events between members of more closely related species or subspecies have to be considered. However, virus-receptor interactions and their impact on transmissibility and zoonotic events often cannot be broken down to a single parameter, as shown for the prominent example influenza A virus and its glycan receptor (69, 70). The interplay of HA-glycan receptor specificity, affinity, and avidity as well as conformational glycan flexibility and HA glycosylation may lead to pandemics if these are sufficient to allow barrier crossovers such as animal-to-human transmission, virus-cell interaction, and finally human-to-human transmission (69).

In conclusion, our data provide clear evidence for close relationships between human and animal polyomaviruses. Such relationships have been suggested by sequence analyses (2, 38, 39, 47), but as the polyomavirus VP1 and LTA_g proteins generally display high levels of sequence conservation, analyses based on sequence alone can be misleading. The structures reported here extend the phylogenetic analyses, demonstrating that key properties of such human-animal pairs, most importantly the engagement of sialic acid receptors, are highly conserved even between distantly related viruses. We expect that this detailed analysis will provide a foundation for assessing receptor-based cell tropism changes and cross-species transmission events, and it may also help to advance an understanding of the evolutionary history of polyomaviruses and perhaps even other viruses.

MATERIALS AND METHODS

Cloning, expression, and protein purification. The NJPyV VP1 DNA was provided by Nischay Mishra (Columbia University), and the pET15b-derived expression plasmid for NJPyV VP1 (amino acids 36 to 323; NCBI accession no. [YP_009030020](#)) was obtained via the InFusion HD cloning strategy according to the manufacturer's protocol (Clontech). Therefore, the vector was linearized via XhoI and the DNA was amplified using the following primers: fw, 5'-CAGCCATATGCTCGACGGAGGAGTTGAAGTTTAAATA-3'; rev, 5'-CAGCCGGATCCTCGATTAATTTTTACAGCCCTTTTCTC-3'.

A synthetic *Escherichia coli* codon-optimized gene coding for amino acids 27 to 299 of HPyV12 VP1 (NCBI accession no. [YP_007684355](#)) was amplified by PCR and cloned into the pET15b vector (Novagen) via NdeI and BamHI restriction enzymes in-frame with an N-terminal hexahistidine tag (His tag) and a

thrombin cleavage site. The following primers were used: fw, 5'-GCATCATATGGGCGGTATTGAGGTTCTG GATGTGAAAAC-3'; and rev, 5'-GACTGGATCCTAGTTACGAACGGCAGCTTTGCGC-3'.

The DNA for the *E. coli* codon-optimized expression constructs for an equivalent N- and C-terminally truncated VP1 of ChPyV (NCBI accession no. [YP_004046682.1](#)), including a mutation of the conserved cysteine in the CD loop (C154S), was purchased from Eurofins Genomics, amplified by PCR, and cloned into the NdeI- and BamHI-linearized pET15b vector by use of InFusion HD cloning. The following primers were used: ChPyV VP1 fw, 5'-CGCGCGGCAGCCATATGGGTGGCGTTGAGGTC-3'; and ChPyV VP1 rev, 5'-G TTAGCAGCCGATCTTAATCTTCACGGCAGCTTTGCG-3'. Site-directed mutagenesis of ChPyV for the STD NMR experiment was carried out using primers as follows (mismatched nucleotides are indicated by boldface): ChPyV V78F fw, 5'-GTACGGTTATAGCGAATTCATTACCATGCCGATGGGTATG-3'; ChPyV VP1 V78F rev, 5'-CATACCCATCGGCATGTGAATGAATTCGCTATAACCGTAC-3'.

The *E. coli* codon-optimized VP1 expression constructs in pET15b (cloning sites NdeI and BamHI) for GhPyV (amino acids 21 to 286; NCBI accession no. [AEC12236.1](#)) and ShPyV (amino acids 20 to 285, including a C95S mutation; NCBI accession no. [AKC98332.1](#)) were purchased from GeneScript (NJ, USA). For FiPyV VP1, the *E. coli* codon-optimized construct (amino acids 22 to 291; NCBI accession no. [YP_529833](#)) was purchased from General Biosystems, Inc. Due to aggregation and oxidation problems during purification trials of an earlier construct (not shown here), an additional C78S mutation was introduced next to the canonical C92S. Furthermore, the construct contained a TEV protease site following the His tag. It was cloned from NcoI to BamHI into the pET-15b vector. The amino acid numbering of all VP1 proteins excludes the N-terminal methionine.

Expression of NJPyV VP1 in *E. coli* Rosetta 2 (DE3) and of HPyV12, ChPyV, ChPyV-V78F, ShPyV, GhPyV, and FiPyV VP1 in *E. coli* BL21(DE3) was induced by addition of 0.4 mM IPTG (isopropyl- β -D-thiogalactopyranoside) and carried out at 20°C for about 18 h. The truncated VP1 cannot assemble into a capsid but forms pentamers or cysteine-mediated dimers of pentamers in solution via the conserved cysteine in the CD loop if present in the expression construct. After cell lysis, VP1 proteins were first purified by nickel affinity chromatography and then by size exclusion chromatography on a Superdex-200 column. The His tag was then removed using thrombin (NJPyV, HPyV12, ChPyV, ShPyV, and GhPyV VP1) or TEV protease (FiPyV VP1) prior to a second nickel affinity chromatography and a final size exclusion chromatography step. A nonnative amino acid sequence, GSHM (or GSMLD in the case of NJPyV VP1), remains at the N terminus after protease cleavage. Buffers were supplemented with 1 M urea for FiPyV VP1 during the nickel affinity chromatography and TEV protease digestion. After size exclusion chromatography, HPyV12 VP1 pentamers were kept in 20 mM HEPES (pH 7.5), 150 mM NaCl, 20 mM dithiothreitol (DTT). Dimers of NJPyV VP1 pentamers and ChPyV, GhPyV, ShPyV, and FiPyV VP1 pentamers were stored in an equivalent buffer without DTT prior to crystallization.

Crystallization and data collection. HPyV12 VP1 was concentrated to 7 mg/ml and crystallized at 20°C by hanging-drop vapor diffusion against a reservoir solution containing 4% (vol/vol) Tacsimate (pH 7.0) and 16% (wt/vol) polyethylene glycol (PEG) 3350. These crystals were not suitable for crystal soaking experiments, because all five binding sites of VP1 were blocked in the crystal by symmetry-related copies of VP1. For complex formation, crystals of HPyV12 VP1 were grown by hanging-drop vapor diffusion at 20°C using a protein concentration of 6 mg/ml and a reservoir solution containing 0.2 M ammonium acetate, 0.1 M bis-Tris (pH 7.0), and 45% (vol/vol) 2-methyl-2,4-pentanediol (MPD). These crystals were soaked for 24 h in the reservoir solution supplemented with 20 mM 3'SLN (Carbosynth, UK).

Crystals of NJPyV VP1 obtained at 4°C using the sitting-drop vapor diffusion method, from a reservoir solution containing 0.1 M succinic acid (pH 7.0) and 15% (wt/vol) PEG 3350, and a protein solution with a concentration of 7 mg/ml yielded the native structure of the protein. Crystals grown using the sitting-drop vapor diffusion method in 0.1 M MES (pH 6.5), 12% (wt/vol) PEG 20000 at 20°C were derivatized by incubation for 10 min in a reservoir solution supplemented with 50 mM 3'SL (Carbosynth, UK). For the structure solution of unliganded ChPyV VP1, crystals grew from a protein concentration of 2.8 mg/ml via the sitting-drop vapor diffusion method at 20°C. The crystallization condition included 0.1 M sodium acetate (pH 7.0), 12% (wt/vol) PEG 3350, and 30% (vol/vol) 1,5-diaminopentane dihydrochloride.

The native GhPyV VP1 structure was solved with a crystal grown at 20°C using the sitting-drop vapor diffusion method and a Morpheus screen purchased from Molecular Dimensions, UK (0.9 M nitrate phosphate sulfate [mix of 0.3 M NaNO₃, 0.3 M Na₂HPO₄, 0.3 M (NH₄)₂SO₄, 0.1 M MES/imidazole (pH 6.5), 10% (wt/vol) PEG 8000, 20% (vol/vol) ethylene glycol]. The protein concentration used for crystallization of GhPyV was 3.7 mg/ml. A GhPyV VP1 crystal obtained from the same screen but with a reservoir containing 0.1 M carboxylic acids (0.2 M Na formate, 0.2 M ammonium acetate, 0.2 M Na citrate tribasic, 0.2 M Na oxamate, 0.2 M Na K tartrate), 0.1 M MES/imidazole (pH 6.5), 10% (wt/vol) PEG 8000, and 20% (vol/vol) ethylene glycol was incubated in a drop of reservoir solution supplemented with 20 mM 2-O-Me-Neu5Ac for 17 h for complex formation.

ShPyV VP1 was concentrated to 4.1 mg/ml and crystallized at 20°C by the sitting-drop vapor diffusion method. The native structure was obtained from a crystallization setup with a reservoir containing 0.2 M KSCN and 20% (wt/vol) PEG 3350. ShPyV VP1 (3.7 mg/ml) was further crystallized via the hanging-drop vapor diffusion method at 20°C, with a reservoir solution of 0.15 M KSCN and 20% (wt/vol) PEG 3350. These crystals were derivatized in a separate drop supplemented with 20 mM 3'SLN for 30 min. A ShPyV VP1 crystal grown with a reservoir solution containing 0.2 M magnesium chloride, 0.1 M Tris (pH 8.5), and 20% (wt/vol) PEG 800 was transferred into the respective reservoir solution supplemented with 20 mM 6'SLN for 48 h.

FiPyV VP1 was concentrated to 3.7 mg/ml, and the native structure was obtained from a crystal grown at 20°C using the hanging-drop vapor diffusion method. The reservoir solution contained 0.1 M

HEPES (pH 7.5), 10% (wt/vol) PEG 8000, and 8% (vol/vol) ethylene glycol. The structure of FiPyV VP1 in complex with 2-O-Me-Neu5Ac was solved using a crystal grown at 4°C in a hanging-drop setup with a reservoir containing 0.2 M MgCl₂, 0.1 M Tris (pH 7), and 10% (wt/vol) PEG 8000. The crystal was soaked in reservoir solution supplemented with 25 mM 2-O-Me-Neu5Ac at 4°C for 30 min.

Drops were set up with 1 μl protein solution and 1 μl reservoir solution for HPyV12, ShPyV (3'SLN derivative), and FiPyV VP1 using the hanging-drop method, or 0.4 μl of the protein solution and 0.4 μl of the reservoir solution were used for the crystallization of NJPyV, GhPyV, and ShPyV (apo and 6'SLN derivative) VP1 using the sitting-drop vapor diffusion method. For ChPyV VP1, 0.2 μl protein solution and 0.2 μl reservoir solution were used for crystallization by the sitting-drop vapor diffusion method. All crystals were transferred for about 2 s into the respective reservoir or soaking solution supplemented with 30% (vol/vol) glycerol or 25% (vol/vol) MPD for FiPyV VP1-2-O-Me-Neu5Ac and ShPyV VP1-3'SLN prior to freezing in liquid nitrogen. Data sets for HPyV12, ShPyV, GhPyV, FiPyV, and ChPyV VP1 were collected using beamline X06DA (X06SA for native FiPyV VP1) of the Swiss Light Source (SLS), while the data sets for NJPyV VP1 were taken using beamline 14.1 of the Berliner Elektronenspeicherring Gesellschaft für Synchrotronstrahlung (BESSY).

Structure determination and structural analysis. Diffraction data sets were processed with XDS (71). All VP1 structures were solved by molecular replacement (MR) using Phaser MR (72) included in the CCP4 program suite (73). The native MCPyV VP1 structure (PDB 4FMG), modified by CHAINSAW (74), was used as a search model in molecular replacement to solve the HPyV12 and NJPyV VP1 structures. The refined NJPyV VP1 structure was used as search model for the structure determination of ChPyV VP1. The refined HPyV12 VP1 pentamer structure was used as a search model to solve the ShPyV VP1 structure, which then served as the search model for the structure determination procedure of GhPyV VP1. The FiPyV VP1 structure was solved using a model of GhPyV VP1. The search models were obtained by truncating the respective structure to the last common atom by CHAINSAW (74). Rigid-body and simulated annealing refinement was carried out with Phenix (75). Subsequently, alternating rounds of model building in Coot (76) and refinement with Refmac5 (77), including 5-fold noncrystallographic symmetry (NCS) restraints, the translation-libration-screw (TLS) method (78), CCP4 library restraints, and user-defined restraints for the α_{2,3}-glycosidic bond in cases where 3'SL or 3'SLN were included as ligands, were performed. In order to investigate patterns of amino acid conservations on the VP1 surface, an alignment was carried out with the multiple sequence alignment tool Clustal Omega (79). JalView (80) was used to assign values for the conservation of the chemical acid amino acid properties: 11 (conserved; red) to 0 (nonconserved; gray). Values of 1 to 0 are colored in gray. For the calculations of RMSD values, C_α atoms of VP1 complex structures were aligned using the secondary-structure matching (SSM) tool in Coot (76). Buried surface and contact areas were calculated with PISA (81).

STD NMR. Saturation transfer difference nuclear magnetic resonance (STD NMR) spectra were recorded on a Bruker AVIII-600 MHz spectrometer with a room temperature probe head at temperatures between 285 K and 288 K. Samples were measured in tubes with an inside diameter of 3 mm and contained 27 μM (NJPyV) or 50 μM (HPyV12 and ChPyV mutant and wild type) VP1 and 2 mM of the respective oligosaccharide(s) in 20 mM K₂HPO₄/KH₂PO₄ (pH 7.4), 150 mM NaCl, 99% D₂O. Oligosaccharides were purchased from Carbosynth, UK, and resuspended in D₂O. Data were processed with TOPSPIN 3.0 (Bruker). Off- and on-resonance frequencies of −30 ppm and 7.3 ppm, respectively, were used. The irradiation power and length of the selective pulse train were 57 Hz and 2 s, respectively. A strength of 3.2 kHz was employed to suppress residual protein resonances in a continuous-wave spin-lock pulse. A total of 5,000 scans were recorded, and the relaxation delay was 3 s. Prior to Fourier transformation, spectra were multiplied with a Gaussian window function.

Data availability. All coordinates and structure factor amplitudes were deposited into the RCSB Protein Data Bank (<https://www.rcsb.org>) under the accession numbers shown in parentheses, as follows: NJPyV VP1 (6Y5X), NJPyV VP1-3'SL (6Y5Y), HPyV12 VP1 (6Y5Z), HPyV12 VP1-3'SLN (6Y60), ShPyV VP1 (6Y61), ShPyV VP1-3'SLN (6Y63), ShPyV VP1-6'SLN (6Y64), GhPyV VP1 (6Y65), GhPyV VP1-2-O-Me-Neu5Ac (6Y66), FiPyV VP1 (6Y67), FiPyV VP1-2-O-Me-Neu5Ac (6Y6A), and ChPyV VP1 (6Y9I). Structure figures were prepared with PyMOL (The PyMOL Molecular Graphics System, version 1.8.0.3; Schrödinger, LLC).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, TIF file, 2.5 MB.

TABLE S1, DOCX file, 0.02 MB.

TABLE S2, DOCX file, 0.01 MB.

ACKNOWLEDGMENTS

We are grateful to the SLS (Villigen, Switzerland) and BESSY (Berlin, Germany) for beam time and to the staff at beamlines X06DA (SLS) and 14.1 (BESSY) for assistance during data collection. We thank Anna-Lisa Fuchs (University of Regensburg, Regensburg, Germany) for help with the purification of HPyV12 VP1 and Remco Sprangers (University of Regensburg, Regensburg, Germany) and Vincent Truffault (Max Planck Institute for Developmental Biology, Tübingen, Germany) for assistance in recording the NMR data.

This work was supported by DFG Research Unit FOR2327 (Virocarb) (grant no. STE

1463/7-1 to T.S.; grant no. BL 1294/3-1 to B.S.B.) and by a contract research grant from the Baden-Württemberg Stiftung (to T.S.).

REFERENCES

- Peretti A, FitzGerald PC, Bliskovsky V, Pastrana DV, Buck CB. 2015. Genome sequence of a fish-associated polyomavirus, black sea bass (*Centropristis striata*) polyomavirus 1. *Genome Announc* 3:e01476-14. <https://doi.org/10.1128/genomeA.01476-14>.
- Buck CB, Van Doorslaer K, Peretti A, Geoghegan EM, Tisza MJ, An P, Katz JP, Pipas JM, McBride AA, Camus AC, McDermott AJ, Dill JA, Delwart E, Ng TF, Farkas K, Austin C, Kraberger S, Davison W, Pastrana DV, Varsani A. 2016. The ancient evolutionary history of polyomaviruses. *PLoS Pathog* 12:e1005574. <https://doi.org/10.1371/journal.ppat.1005574>.
- Ehlers B, Anoh AE, Ben SN, Broll S, Couacy-Hymann E, Fischer D, Gedvilaite A, Ingenhutt N, Liebmann S, Martin M, Mossoun A, Mugisha L, Muyembe-Tamfum JJ, Pauly M, Perez de Val B, Preugschas H, Richter D, Schubert G, Szentiks CA, Teichmann T, Walter C, Ulrich RG, Wiersma L, Leendertz FH, Calvignac-Spencer S. 2019. Novel polyomaviruses in mammals from multiple orders and reassessment of polyomavirus evolution and taxonomy. *Viruses* 11:930. <https://doi.org/10.3390/v11100930>.
- Padgett BL, Walker DL, Zuerlein GM, Eckroade RJ, Dessel BH. 1971. Cultivation of papova-like virus from human brain with progressive multifocal leucoencephalopathy. *Lancet* 1:1257–1260. [https://doi.org/10.1016/s0140-6736\(71\)91777-6](https://doi.org/10.1016/s0140-6736(71)91777-6).
- Gardner SD, Field AM, Coleman DV, Hulme B. 1971. New human papovavirus (B.K.) isolated from urine after renal transplantation. *Lancet* 1:1253–1257. [https://doi.org/10.1016/s0140-6736\(71\)91776-4](https://doi.org/10.1016/s0140-6736(71)91776-4).
- Ho J, Jedrych JJ, Feng H, Natalie AA, Grandinetti L, Mirvish E, Crespo MM, Yadav D, Fasanella KE, Prokell S, Kuan SF, Pastrana DV, Buck CB, Shuda Y, Moore PS, Chang Y. 2015. Human polyomavirus 7-associated pruritic rash and viremia in transplant recipients. *J Infect Dis* 211:1560–1565. <https://doi.org/10.1093/infdis/jiu524>.
- Feng H, Shuda M, Chang Y, Moore PS. 2008. Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science* 319:1096–1100. <https://doi.org/10.1126/science.1152586>.
- van der Meijden E, Janssens RW, Lauber C, Bouwes Bavinck JN, Gorbalenya AE, Feltkamp MC. 2010. Discovery of a new human polyomavirus associated with trichodysplasia spinulosa in an immunocompromised [sic] patient. *PLoS Pathog* 6:e1001024. <https://doi.org/10.1371/journal.ppat.1001024>.
- Dalianis T, Hirsch HH. 2013. Human polyomaviruses in disease and cancer. *Virology* 437:63–72. <https://doi.org/10.1016/j.virol.2012.12.015>.
- DeCaprio JA, Garcea RL. 2013. A cornucopia of human polyomaviruses. *Nat Rev Microbiol* 11:264–276. <https://doi.org/10.1038/nrmicro2992>.
- Moens U, Krumbholz A, Ehlers B, Zell R, John R, Calvignac-Spencer S, Lauber C. 2017. Biology, evolution, and medical importance of polyomaviruses: an update. *Infect Genet Evol* 54:18–38. <https://doi.org/10.1016/j.meegid.2017.06.011>.
- Buch MH, Liaci AM, O'Hara SD, Garcea RL, Neu U, Stehle T. 2015. Structural and functional analysis of murine polyomavirus capsid proteins establish the determinants of ligand recognition and pathogenicity. *PLoS Pathog* 11:e1005104. <https://doi.org/10.1371/journal.ppat.1005104>.
- Tsai B, Gilbert JM, Stehle T, Lencer W, Benjamin TL, Rapoport TA. 2003. Gangliosides are receptors for murine polyoma virus and SV40. *EMBO J* 22:4346–4355. <https://doi.org/10.1093/emboj/cdg439>.
- Neu U, Allen SA, Blaum BS, Liu Y, Frank M, Palma AS, Ströh LJ, Feizi T, Peters T, Atwood WJ, Stehle T. 2013. A structure-guided mutation in the major capsid protein retargets BK polyomavirus. *PLoS Pathog* 9:e1003688. <https://doi.org/10.1371/journal.ppat.1003688>.
- Neu U, Woellner K, Gauglitz G, Stehle T. 2008. Structural basis of GM1 ganglioside recognition by simian virus 40. *Proc Natl Acad Sci U S A* 105:5219–5224. <https://doi.org/10.1073/pnas.0710301105>.
- Dugan AS, Gasparovic ML, Atwood WJ. 2008. Direct correlation between sialic acid binding and infection of cells by two human polyomaviruses (JC virus and BK virus). *J Virol* 82:2560–2564. <https://doi.org/10.1128/JVI.02123-07>.
- Schowalter RM, Pastrana DV, Buck CB. 2011. Glycosaminoglycans and sialylated glycans sequentially facilitate Merkel cell polyomavirus infectious entry. *PLoS Pathog* 7:e1002161. <https://doi.org/10.1371/journal.ppat.1002161>.
- Geoghegan EM, Pastrana DV, Schowalter RM, Ray U, Gao W, Ho M, Pauly GT, Sigano DM, Kaynor C, Cahir-McFarland E, Combaluzier B, Grimm J, Buck CB. 2017. Infectious entry and neutralization of pathogenic JC polyomaviruses. *Cell Rep* 21:1169–1179. <https://doi.org/10.1016/j.celrep.2017.10.027>.
- Ströh LJ, Neu U, Blaum BS, Buch MHC, Garcea RL, Stehle T. 2014. Structure analysis of the major capsid proteins of human polyomaviruses 6 and 7 reveals an obstructed sialic acid binding site. *J Virol* 88:10831–10839. <https://doi.org/10.1128/JVI.01084-14>.
- Neu U, Wang J, Macejak D, Garcea RL, Stehle T. 2011. Structures of the major capsid proteins of the human Karolinska Institutet and Washington University polyomaviruses. *J Virol* 85:7384–7392. <https://doi.org/10.1128/JVI.00382-11>.
- Neu U, Maginnis MS, Palma AS, Ströh LJ, Nelson CD, Feizi T, Atwood WJ, Stehle T. 2010. Structure-function analysis of the human JC polyomavirus establishes the LSTc pentasaccharide as a functional receptor motif. *Cell Host Microbe* 8:309–319. <https://doi.org/10.1016/j.chom.2010.09.004>.
- Neu U, Khan ZM, Schuch B, Palma AS, Liu Y, Pawlita M, Feizi T, Stehle T. 2013. Structures of B-lymphotropic polyomavirus VP1 in complex with oligosaccharide ligands. *PLoS Pathog* 9:e1003714. <https://doi.org/10.1371/journal.ppat.1003714>.
- Neu U, Hengel H, Blaum BS, Schowalter RM, Macejak D, Gilbert M, Wakarchuk WW, Imamura A, Ando H, Kiso M, Arnberg N, Garcea RL, Peters T, Buck CB, Stehle T. 2012. Structures of Merkel cell polyomavirus VP1 complexes define a sialic acid binding site required for infection. *PLoS Pathog* 8:e1002738. <https://doi.org/10.1371/journal.ppat.1002738>.
- Khan ZM, Liu Y, Neu U, Gilbert M, Ehlers B, Feizi T, Stehle T. 2014. Crystallographic and glycan microarray analysis of human polyomavirus 9 VP1 identifies N-glycolylneuraminic acid as a receptor candidate. *J Virol* 88:6100–6111. <https://doi.org/10.1128/JVI.03455-13>.
- Ströh LJ, Gee GV, Blaum BS, Dugan AS, Feltkamp MC, Atwood WJ, Stehle T. 2015. Trichodysplasia spinulosa-associated polyomavirus uses a displaced binding site on VP1 to engage sialylated glycolipids. *PLoS Pathog* 11:e1005112. <https://doi.org/10.1371/journal.ppat.1005112>.
- Stehle T, Yan Y, Benjamin TL, Harrison SC. 1994. Structure of murine polyomavirus complexed with an oligosaccharide receptor fragment. *Nature* 369:160–163. <https://doi.org/10.1038/369160a0>.
- Hurdiss DL, Frank M, Snowden JS, Macdonald A, Ranson NA. 2018. The structure of an infectious human polyomavirus and its interactions with cellular receptors. *Structure* 26:839–847.e3. <https://doi.org/10.1016/j.str.2018.03.019>.
- Varki A, Schauer R. 2009. Sialic acids. In Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME (ed), *Essentials of glycobiology*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Ströh LJ, Stehle T. 2014. Glycan engagement by viruses: receptor switches and specificity. *Annu Rev Virol* 1:285–306. <https://doi.org/10.1146/annurev-virology-031413-085417>.
- Campanero-Rhodes MA, Smith A, Chai WG, Sonnino S, Mauri L, Childs RA, Zhang YB, Ewers H, Helenius A, Imberty A, Feizi T. 2007. N-glycolyl GM1 ganglioside as a receptor for simian virus 40. *J Virol* 81:12846–12858. <https://doi.org/10.1128/JVI.01311-07>.
- Magaldi TG, Buch MH, Murata H, Erickson KD, Neu U, Garcea RL, Peden K, Stehle T, Dimaio D. 2012. Mutations in the GM1 binding site of simian virus 40 VP1 alter receptor usage and cell tropism. *J Virol* 86:7028–7042. <https://doi.org/10.1128/JVI.00371-12>.
- O'Hara SD, Stehle T, Garcea R. 2014. Glycan receptors of the Polyomaviridae: structure, function, and pathogenesis. *Curr Opin Virol* 7:73–78. <https://doi.org/10.1016/j.coviro.2014.05.004>.
- Mishra N, Pereira M, Rhodes RH, An P, Pipas JM, Jain K, Kapoor A, Briese T, Faust PL, Lipkin WI. 2014. Identification of a novel polyomavirus in a pancreatic transplant recipient with retinal blindness and vasculitic myopathy. *J Infect Dis* 210:1595–1599. <https://doi.org/10.1093/infdis/jiu250>.
- Korup S, Rietscher J, Calvignac-Spencer S, Trusch F, Hofmann J, Moens U, Sauer I, Voigt S, Schmuck R, Ehlers B. 2013. Identification of a novel human polyomavirus in organs of the gastrointestinal tract. *PLoS One* 8:e58021. <https://doi.org/10.1371/journal.pone.0058021>.

35. Gaboriaud P, Ferté M, Arnold F, Leblond V, Nicol J, Debare H, Le Meur M, Martini F, Tognon M, Touzé A. 2018. Age-specific seroprevalence of human polyomavirus 12 and Saint Louis and New Jersey polyomaviruses. *Emerg Microbes Infect* 7:22. <https://doi.org/10.1038/s41426-018-0026-0>.
36. Kamminga S, van der Meijden E, Feltkamp MCW, Zaaijer HL. 2018. Seroprevalence of fourteen human polyomaviruses determined in blood donors. *PLoS One* 13:e0206273. <https://doi.org/10.1371/journal.pone.0206273>.
37. Zhou X, Bai H, Kataoka M, Ito M, Muramatsu M, Suzuki T, Li TC. 2019. Characterization of the self-assembly of New Jersey polyomavirus VP1 into virus-like particles and the virus seroprevalence in Japan. *Sci Rep* 9:13085. <https://doi.org/10.1038/s41598-019-49541-y>.
38. Scuda N, Madinda NF, Akoua-Koffi C, Adjogoua EV, Wevers D, Hofmann J, Cameron KN, Leendertz SAJ, Couacy-Hymann E, Robbins M, Boesch C, Jarvis MA, Moens U, Mugisha L, Calvignac-Spencer S, Leendertz FH, Ehlers B. 2013. Novel polyomaviruses of nonhuman primates: genetic and serological predictors for the existence of multiple unknown polyomaviruses within the human population. *PLoS Pathog* 9:e1003429. <https://doi.org/10.1371/journal.ppat.1003429>.
39. Rigatti LH, Toptan T, Newsome JT, Moore PS, Chang Y. 2016. Identification and characterization of novel rat polyomavirus 2 in a colony of X-SCID rats by P-PIT assay. *mSphere* 1:e00334-16. <https://doi.org/10.1128/mSphere.00334-16>.
40. Gedvilaite A, Tryland M, Ulrich RG, Schneider J, Kurmauskaitė V, Moens U, Preugschas H, Calvignac-Spencer S, Ehlers B. 2017. Novel polyomaviruses in shrews (Soricidae) with close similarity to human polyomavirus 12. *J Gen Virol* 98:3060–3067. <https://doi.org/10.1099/jgv.0.000948>.
41. da Silva MS, Cibulski SP, Alves C, Weber MN, Budaszewski RF, Silveira S, Mosenca ACS, Mayer FQ, Goltz LV, Campos R, Canal CW. 2018. New polyomavirus species identified in nutria, *Myocastor coypus* polyomavirus 1. *Arch Virol* 163:3203–3206. <https://doi.org/10.1007/s00705-018-3985-5>.
42. Tao Y, Shi M, Conrardy C, Kuzmin IV, Recuenco S, Agwanda B, Alvarez DA, Ellison JA, Gilbert AT, Moran D, Niezgodna M, Lindblade KA, Holmes EC, Breiman RF, Rupprecht CE, Tong S. 2013. Discovery of diverse polyomaviruses in bats and the evolutionary history of the Polyomaviridae. *J Gen Virol* 94:738–748. <https://doi.org/10.1099/vir.0.047928-0>.
43. Hill SC, Murphy AA, Cotten M, Palsler AL, Benson P, Lesellier S, Gormley E, Richomme C, Grierson S, Bhuachalla DN, Chambers M, Kellam P, Boschiroli ML, Ehlers B, Jarvis MA, Pybus OG. 2015. Discovery of a polyomavirus in European badgers (*Meles meles*) and the evolution of host range in the family Polyomaviridae. *J Gen Virol* 96:1411–1422. <https://doi.org/10.1099/vir.0.000071>.
44. Carter JJ, Daugherty MD, Qi X, Bheda-Malge A, Wipf GC, Robinson K, Roman A, Malik HS, Galloway DA. 2013. Identification of an overprinting gene in Merkel cell polyomavirus provides evolutionary insight into the birth of viral genes. *Proc Natl Acad Sci U S A* 110:12744–12749. <https://doi.org/10.1073/pnas.1303526110>.
45. Calvignac-Spencer S, Feltkamp MC, Daugherty MD, Moens U, Ramqvist T, John R, Ehlers B, Polyomaviridae Study Group of the International Committee on Taxonomy of Viruses. 2016. A taxonomy update for the family Polyomaviridae. *Arch Virol* 161:1739–1750. <https://doi.org/10.1007/s00705-016-2794-y>.
46. Sauvage V, Foulongne V, Cheval J, Ar Gouilh M, Pariente K, Dereure O, Manuguerra JC, Richardson J, Lecuit M, Burguiere A, Caro V, Eloit M. 2011. Human polyomavirus related to African green monkey lymphotropic polyomavirus. *Emerg Infect Dis* 17:1364–1370. <https://doi.org/10.3201/eid1708.110278>.
47. Scuda N, Hofmann J, Calvignac-Spencer S, Rupprecht K, Liman P, Kuhn J, Hengel H, Ehlers B. 2011. A novel human polyomavirus closely related to the African green monkey-derived lymphotropic polyomavirus. *J Virol* 85:4586–4590. <https://doi.org/10.1128/JVI.02602-10>.
48. Geoghegan JL, Duchene S, Holmes EC. 2017. Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. *PLoS Pathog* 13:e1006215. <https://doi.org/10.1371/journal.ppat.1006215>.
49. Gheit T, Dutta S, Oliver J, Robitaille A, Hampras S, Combes JD, McKay-Chopin S, Le Calvez-Kelm F, Fenske N, Cherpelis B, Giuliano AR, Franceschi S, McKay J, Rollison DE, Tommasino M. 2017. Isolation and characterization of a novel putative human polyomavirus. *Virology* 506:45–54. <https://doi.org/10.1016/j.virol.2017.03.007>.
50. John R, Muller H. 1998. Avian polyomavirus in wild birds: genome analysis of isolates from Falconiformes and Psittaciformes. *Arch Virol* 143:1501–1512. <https://doi.org/10.1007/s007050050393>.
51. Corrand L, Gelfi J, Albaric O, Etievant M, Pingret JL, Guerin JL. 2011. Pathological and epidemiological significance of goose haemorrhagic polyomavirus infection in ducks. *Avian Pathol* 40:355–360. <https://doi.org/10.1080/03079457.2011.582481>.
52. Carr M, Gonzalez G, Sasaki M, Dool SE, Ito K, Ishii A, Hang'ombe BM, Mweene AS, Teeling EC, Hall WW, Orba Y, Sawa H. 2017. Identification of the same polyomavirus species in different African horseshoe bat species is indicative of short-range host-switching events. *J Gen Virol* 98:2771–2785. <https://doi.org/10.1099/jgv.0.000935>.
53. Stehle T, Harrison SC. 1997. High-resolution structure of a polyomavirus VP1-oligosaccharide complex: implications for assembly and receptor binding. *EMBO J* 16:5139–5148. <https://doi.org/10.1093/emboj/16.16.5139>.
54. Liddington RC, Yan Y, Moulai J, Sahli R, Benjamin TL, Harrison SC. 1991. Structure of simian virus 40 at 3.8-Å resolution. *Nature* 354:278–284. <https://doi.org/10.1038/354278a0>.
55. Krissinel E, Henrick K. 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* 60:2256–2268. <https://doi.org/10.1107/S0907444904026460>.
56. Haselhorst T, Blanchard H, Frank M, Kraschnefski MJ, Kiefel MJ, Szczyzew AJ, Dyason JC, Fleming F, Holloway G, Coulson BS, von Itzstein M. 2007. STD NMR spectroscopy and molecular modeling investigation of the binding of N-acetylneuraminic acid derivatives to rhesus rotavirus VP8* core. *Glycobiology* 17:68–81. <https://doi.org/10.1093/glycob/cwl051>.
57. Mayer M, Meyer B. 1999. Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angew Chem Int Ed* 38:1784–1788. [https://doi.org/10.1002/\(SICI\)1521-3773\(19990614\)38:12<1784::AID-ANIE1784>3.0.CO;2-Q](https://doi.org/10.1002/(SICI)1521-3773(19990614)38:12<1784::AID-ANIE1784>3.0.CO;2-Q).
58. Blaum BS, Neu U, Peters T, Stehle T. 2018. Spin ballet for sweet encounters: saturation-transfer difference NMR and X-ray crystallography complement each other in the elucidation of protein-glycan interactions. *Acta Crystallogr F Struct Biol Commun* 74:451–462. <https://doi.org/10.1107/S2053230X18006581>.
59. Guerin JL, Gelfi J, Dubois L, Vuillaume A, Boucraut-Baralon C, Pingret JL. 2000. A novel polyomavirus (goose hemorrhagic polyomavirus) is the agent of hemorrhagic nephritis enteritis of geese. *J Virol* 74:4523–4529. <https://doi.org/10.1128/jvi.74.10.4523-4529.2000>.
60. Varki A. 2001. Loss of N-glycolylneuraminic acid in humans: mechanisms, consequences, and implications for hominid evolution. *Am J Phys Anthropol* 33(Suppl):54–69. <https://doi.org/10.1002/ajpa.10018.abs>.
61. Gamblin SJ, Skehel JJ. 2010. Influenza hemagglutinin and neuraminidase membrane glycoproteins. *J Biol Chem* 285:28403–28409. <https://doi.org/10.1074/jbc.R110.129809>.
62. Shi Y, Wu Y, Zhang W, Qi J, Gao GF. 2014. Enabling the 'host jump': structural determinants of receptor-binding specificity in influenza A viruses. *Nat Rev Microbiol* 12:822–831. <https://doi.org/10.1038/nrmicro3362>.
63. Cahan LD, Singh R, Paulson JC. 1983. Sialyloligosaccharide receptors of binding variants of polyoma virus. *Virology* 130:281–289. [https://doi.org/10.1016/0042-6822\(83\)90083-1](https://doi.org/10.1016/0042-6822(83)90083-1).
64. Maginnis MS, Stroh LJ, Gee GV, O'Hara BA, Derdowski A, Stehle T, Atwood WJ. 2013. Progressive multifocal leukoencephalopathy-associated mutations in the JC polyomavirus capsid disrupt lactoseries tetrasaccharide C binding. *mBio* 4:e00247-13. <https://doi.org/10.1128/mBio.00247-13>.
65. Liu CK, Wei G, Atwood WJ. 1998. Infection of glial cells by the human polyomavirus JC is mediated by an N-linked glycoprotein containing terminal alpha(2-6)-linked sialic acids. *J Virol* 72:4643–4649. <https://doi.org/10.1128/JVI.72.6.4643-4649.1998>.
66. Zielonka A, Gedvilaite A, Reetz J, Rosler U, Muller H, John R. 2012. Serological cross-reactions between four polyomaviruses of birds using virus-like particles expressed in yeast. *J Gen Virol* 93:2658–2667. <https://doi.org/10.1099/vir.0.044917-0>.
67. Zielonka A, Gedvilaite A, Ulrich R, Luschow D, Sasnauskas K, Muller H, John R. 2006. Generation of virus-like particles consisting of the major capsid protein VP1 of goose hemorrhagic polyomavirus and their application in serological tests. *Virus Res* 120:128–137. <https://doi.org/10.1016/j.virusres.2006.02.010>.
68. Madinda NF, Ehlers B, Wertheim JO, Akoua-Koffi C, Bergl RA, Boesch C, Akonkwa DB, Eckardt W, Fruth B, Gillespie TR, Gray M, Hohmann G, Karhemere S, Kujirakwinja D, Langergraber K, Muyembe JJ, Nishuli R, Pauly M, Petzelkova KJ, Robbins MM, Todd A, Schubert G, Stoinski TS,

- Wittig RM, Zuberbuhler K, Peeters M, Leendertz FH, Calvignac-Spencer S. 2016. Assessing host-virus codivergence for close relatives of Merkel cell polyomavirus infecting African great apes. *J Virol* 90:8531–8541. <https://doi.org/10.1128/JVI.00247-16>.
69. Ji Y, White YJ, Hadden JA, Grant OC, Woods RJ. 2017. New insights into influenza A specificity: an evolution of paradigms. *Curr Opin Struct Biol* 44:219–231. <https://doi.org/10.1016/j.sbi.2017.06.001>.
70. Reperant LA, Kuiken T, Osterhaus AD. 2012. Adaptive pathways of zoonotic influenza viruses: from exposure to establishment in humans. *Vaccine* 30:4419–4434. <https://doi.org/10.1016/j.vaccine.2012.04.049>.
71. Kabsch W. 2010. XDS. *Acta Crystallogr D Biol Crystallogr* 66:125–132. <https://doi.org/10.1107/S0907444909047337>.
72. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. *J Appl Crystallogr* 40:658–674. <https://doi.org/10.1107/S0021889807021206>.
73. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. 2011. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67:235–242. <https://doi.org/10.1107/S0907444910045749>.
74. Stein N. 2008. CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J Appl Crystallogr* 41:641–643. <https://doi.org/10.1107/S0021889808006985>.
75. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66:213–221. <https://doi.org/10.1107/S0907444909052925>.
76. Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66:486–501. <https://doi.org/10.1107/S0907444910007493>.
77. Murshudov GN, Vagin AA, Dodson EJ. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53:240–255. <https://doi.org/10.1107/S0907444996012255>.
78. Painter J, Merritt EA. 2006. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* 62:439–450. <https://doi.org/10.1107/S0907444906005270>.
79. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539. <https://doi.org/10.1038/msb.2011.75>.
80. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. 2009. Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25:1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>.
81. Krissinel E, Henrick K. 2007. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372:774–797. <https://doi.org/10.1016/j.jmb.2007.05.022>.