# Structural and Functional Characterization of Polyomavirus Capsid Proteins in Interaction with Sialylated and Neutral Glycan Ligands

**Dissertation**

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Nils Rustmeier, M.Sc.

aus Kiel

Tübingen

2023

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Für meine Eltern

# Abstract

Polyomaviruses are non-enveloped, double-stranded DNA viruses infecting animals, including mammals, birds, and fish. In immunocompromised individuals, certain polyomaviruses can inflict severe disease. In most cases, however, infection with a polyomavirus is asymptomatic and remains latent. Polyomavirus particles comprise a capsid containing the genomic DNA, and that capsid is an icosahedral construct of several capsid proteins. The capsid is scaffolded by the major capsid protein VP1, a pentameric protein that protects the genome and enables polyomaviruses to recognize and eventually infect their target cells. Most known natural cellular receptors of polyomaviruses are carbohydrate structures (glycans) that carry a terminal sialic acid. The sialic acid, a nine-carbon comprising monosaccharide, is recognized by VP1 through peripheral protrusions from its core domain. These carbohydrate receptor-engaging protein loops are highly diverse between polyomavirus species and vary in sequence and length. This diversity, often called plasticity, constitutes the platform to recognize the different structural contexts of sialic acids, such as their chemical modifications or their linkage to the next monosaccharide. The individual interactions between a VP1 and its sialic acid receptors are usually weak with millimolar dissociation constants. However, 360 binding sites on the entire capsid can act in concert during cellular attachment, compensating for this deficiency. Polyomavirus receptors beyond the scope of sialylated glycans are known, including glycosaminoglycans and membrane proteins, but detailed structural information about the respective interactions remains elusive.

In this thesis, I show for the first time that several polyomaviruses can interact with non-sialylated, neutral carbohydrates. Therein, I characterize a binding site for the Forssman antigen (FA) glycan in the sheep polyomavirus (ShPyV) VP1. My functional assays reveal that ShPyV VP1 efficiently binds to FA-positive blood cells, surpassing established FA-binding proteins in specificity and affinity. These findings complement our understanding of receptor engagement by polyomaviruses and include the first report of a virus specifically interacting with FA. As an aberrant glycosylation linked to human cancer, the Forssman antigen is a potential biomedical marker. ShPyV VP1 may facilitate its detection in the future.

# Kurzfassung

Polyomaviren sind unbehüllte Viren mit doppelsträngiger DNA, die Tieren wie Säuger, Vögel und Fische befallen. Für immungeschwächte Individuen können gewissen Polyomaviren eine Gefahr darstellen. Jedoch verläuft die Ansteckung in den meisten Fällen ohne Symptome und verbleibt ruhend. Polyomavirenpartikel besitzen ein ikosaedrisches Kapsid, welches die genomische DNA enthält und ein Gebilde aus mehreren Kapsidproteinen darstellt. Das Grundgerüst des Kapsids besteht aus dem *major capsid protein* VP1, einem pentameren Protein, das neben dem Genomschutz auch für das Erkennen und Befallen der Zielzellen verantwortlich ist. Der Großteil der bekannten zellulären Rezeptoren der Polyomaviren sind Kohlenhydratstrukturen (Glykane) mit endständiger Sialinsäure. Die Sialinsäure, ein neun-kohlenstoffhaltiger Einfachzucker, wird von Kohlenhydratrezeptor-bindenden Proteinschleifen erkannt, welche zwischen den Polyomavirusspezies nicht nur in ihrer Sequenz, sondern auch der Länge hochgradig variabel sind. Diese Diversität der VP1 Struktur, oft als Verformbarkeit (Engl. *plasticity*) bezeichnet, bildet den Rahmen zur Erkennung der unterschiedlichen Sialinsäure-Strukturen mit ihren verschiedenen chemischen Modifikationen oder Verkettungen an nachfolgende Zuckereinheiten. Einzelne Wechselwirkungen zwischen VP1 und einer Sialinsäure sind üblicherweise schwach und besitzen millimolare Dissoziationskonstanten. Allerdings wird dieses Defizit durch das Vorhandensein von insgesamt 360 Bindungsstellen auf dem Kapsid kompensiert, von denen viele gleichzeitig an einer Zellbindung beteiligt sein können. Außer sialylierten Glykanen sind Glykosaminoglykane und Membranproteine als Polyomavirusrezeptoren bekannt, jedoch fehlen bisher genaue strukturelle Informationen über die zugrundeliegenden Interaktionen.

In dieser Arbeit beschreibe ich erstmalig die Interaktionen zwischen Polyomaviren und nicht-sialylierten, neutralen Kohlenhydraten. Zu diesen zählt das Glykan des Forssman Antigens (FA), für welche ich im Schafspolyomavirus (ShPyV, von Engl. *sheep polyomavirus*) eine Bindestelle charakterisiert habe. Meine funktionellen Untersuchungen zeigen, dass ShPyV VP1 effizient an FA-positive Blutzellen bindet und dabei bisher bekannte FA-Bindeproteine in Spezifität und Affinität übertrifft. Diese Ergebnisse erweitern unser Wissen um die Rezeptornutzung von Polyomaviren und enthalten den ersten Bericht einer spezifischen Bindung des FA's durch ein Virus. Als eine anomale Glykosylierung in Verbindung mit menschlichem Krebs ist das Forssman Antigen ein potenzieller biomedizinischer Marker, dessen Detektion zukünftig durch das ShPyV VP1 erleichtert werden könnte.

# Contributions

Experiments and analyses described in this thesis were performed by me unless indicated otherwise in the text. This dissertation further contains material with my contribution that was or will be published elsewhere.

In the peer-reviewed research article Ströh, L. J. *et al.* Structural basis and evolution of glycan receptor specificities within the polyomavirus family. *mBio* **11,** 1–21 (2020), I contributed to the structural characterization of the major capsid proteins (VP1s) from sheep polyomavirus, goose hemorrhagic polyomavirus, finch polyomavirus, and chimpanzee polyomavirus, as seen in the article's figures 5a, 6, and 7, supplemental figure 1, and supplemental tables 1 and 2. Although the manuscript is not included as a chapter in this dissertation, its results constitute the basis for many of my successive findings and are referred to throughout the thesis.

The preprint Rustmeier, N. H. *et al.* A novel and highly specific Forssman antigen-binding protein from sheep polyomavirus. *bioRxiv,* 2023.04.10.536218 (2023) is attached in the form of a scientific manuscript with co-authors. I was responsible for most of the study's design and performed all experiments except those of the glycan array screenings (see below) and the NMR spectroscopy, which was conducted by Dr. Vincent Truffault (MPI for Biology, Tübingen) in my presence using reagents provided by me. Further author contributions are minuted within the manuscript.

All glycan array experiments were conducted by the Glycosciences Laboratory (Imperial College London, UK), funded by the Wellcome Trust Biomedical Resource Grant "Carbohydrate Microarray Facility for the New Era of Glycomics" (Nov 2019 - Oct 2023). The glycan array screening results and the respective charts were kindly provided by Dr. Yan Liu and Prof. Ten Feizi for incorporation into this thesis. Analyses and discussions proceeded in collaboration.

# Danksagung

Die Erstellung meiner Doktorarbeit war ein sehr lehrreiches, herausforderndes und sicherlich auch emotionales Unterfangen, für dessen Erfolg ich einer Menge Menschen zu danken habe.

Prof. Dr. Thilo Stehle - ich danke dir, dass du mir die Möglichkeit gegeben hast, in deinem Labor zu wirken und die wunderbare Welt der Polyomaviren kennenzulernen. Du warst immer an meinen Ergebnissen interessiert und hast nie mit wertvollen Tipps gespart. Deine Unterstützung meines Beitrages zu deinem langjährigsten Forschungsprojekt war stets zu spüren. Es war toll, eine Tür zu neuen Erkenntnissen über die Rezeptoren unserer „Hausviren" aufzustoßen.

Prof. Dr. Mario Schelhaas - ich danke dir für die Übernahme der zweiten Berichterstattung. Hoffentlich findest du die Ergebnisse zu den Polyomaviren genau so spannend wie wir.

Dr. Yan Liu and Prof. Ten Feizi - thank you for your awesome glycan arrays and the results we could gather together. I remember how excited I was when I saw that enormous peak indicating the binding of Forssman pentaose by the sheep polyomavirus VP1. The fruitful discussions with you were always helpful for my understanding of the glycosciences world, which I am exceptionally grateful for.

Meiner Vorgängerin Luisa - deine Vorarbeiten haben den Weg für meine eigenen Ergebnisse geebnet. Du hast mir viel beigebracht, auch wenn du zu dieser Zeit nicht mehr im Labor warst. Ich freue mich sehr mit dir und Thilo über Polyomaviren publiziert zu haben.

My colleagues Joana and Jasmin - although not at the same time, each of you had a huge impact on my thriving as a scientist and person. Without you, being in the lab would have been way less fun.
Joana - from the moment I saw you, I knew we would become friends. You were there for me when I needed it. Your smile brightened up my day at any time, and it felt like such a loss when you left the lab after finishing your PhD. I'm sure our friendship will last, no matter where we are. Muito obrigado!

x

Jasmin - wer hätte gedacht, dass ein kleiner Schnitt in meinen Finger zu diesen unzähligen Insidern führen würden, mit denen wir heute kommunizieren. Du warst die beste Büronachbarin. Mit wem sonst hätte ich so viel Blödsinn (und wertvolle Infos über Polyomaviren) austauschen können. Danke dafür!

Allen anderen Kolleginnen und Kollegen im AK Stehle - in einem Labor zu sein bedeutet Teamarbeit, und mit euch hat das super funktioniert. Besonderer Dank geht an Georg und sein wechselndes Team an Betreuern des Linuxnetzwerks, - ohne euch wären die vielen Stunden in der Workstation nicht möglich gewesen. Danke außerdem Yinglan, Georg, Michael, Elena, Niklas, Nina und Nele für die vielen Fragen, interessanten Diskussionen und lustigen Gespräche. Irmi, du hast mir von Anfang an bis zu deinem Ruhestand gezeigt, wie das Labor funktioniert. Das Büro im alten IFIB mit dir zu teilen war wirklich schön. Christoph, mit dir diesen riesigen Haufen Metall, der sich Röntgenquelle nennt, zu bändigen war ein Abenteuer. Ich danke Joshua Müller und Alexander Herrmann, deren Bachelorarbeiten ich betreuen durfte - euer Ehrgeiz und eure Zuverlässigkeit im Labor suchen ihresgleichen.

Meinen Freunden, ohne die alles anders wäre. Während meiner Schulzeit in Kiel, meines Studiums in Greifswald und meiner Doktorandenzeit in Tübingen habe ich so viele liebe Menschen kennengelernt, die ich heute Freunde nennen darf, was einfach unfassbar ist. Niklas - danke, dass du die letzten Jahre immer am Start warst. Mit dir, Mario, und Michel großzuwerden war so wichtig.

Tim - wir haben nach jeder Abgabe meiner Abschlussarbeiten telefoniert. Danke für deine Unterstützung!

Dariah - es ist schön, dass wir es schaffen uns einmal jährlich sehen. Jannis - mit dir um die Welt zu reisen war die beste Zeit meines Lebens.

Caner - you know you saved me. Starting my PhD with you and Michael was probably the luckiest coincident. Thank you for being my friend.

Meiner Tübinger WG in wechselnder Zusammensetzung, insbesondere Karlotta - ihr habt die Wohnung in der Hechinger Straße zu einem Zuhause gemacht.

Meinen Eltern - Liebe Mama und Papa, euch widme ich diese Arbeit. Mein ganzes Leben lang habt ihr mich in allen meinen Vorhaben unterstützt und wart immer da, wenn ich euch gebraucht habe. Ihr habt mir von Kleinauf so viel ermöglicht, wofür ich nicht dankbarer sein könnte. Ohne euch wäre ich nicht das, was ich heute bin. Wie schön es ist, solch liebe Eltern zu haben.

# Contents

# Contents

# List of Figures

# List of Tables

# Abbreviations

| | |
|---|---|
| 2-$O$-Me-Neu5Ac | 2-$O$-Methyl-$\alpha$-D-$N$-acetylneuraminic acid |
| 3'SL | 3-sialyllactose |
| 3'SLN | 3'-sialyl-$N$-acetyllactosamine |
| 6'SL | 6-sialyllactose |
| 6'SLN | 6'-sialyl-$N$-acetyllactosamine |
| APyV | Avian polyomavirus |
| ASU | Asymmetric unit |
| BKPyV | BK polyomavirus |
| BSA | Bovine serum albumine |
| BFDPyV | Budgerigar fledgling disease polyomavirus |
| ChPyV | Chimpanzee polyomavirus |
| CV | Column volume |
| *E. coli* | *Escherichia coli* |
| FA | Forssman antigen |
| FiPyV | Finch polyomavirus |
| $F_P$ | Forssman pentaose |
| GAG | Glycosaminoglycan |
| Gal4 | Galactotetraose |
| Gb3 | Globotriose |
| GHPyV | Goose hemorrhagic polyomavirus |
| GSL | Glycosphingolipid |
| HPyV6 | Human polyomavirus 6 |
| HPyV7 | Human polyomavirus 7 |
| HPyV9 | Human polyomavirus 9 |
| IMAC | Immobilized metal ion affinity chromatography |
| Kd | Dissociation constant |
| JR | Jelly roll |
| JCPyV | JC polyomavirus |
| LacNAc | $N$-acetyllactosamine |
| LB | Lysogeny broth |
| LIPyV | Lyon IARC polyomavirus |
| LNT | Lacto-$N$-tetraose |
| LPyV | B-lymphotropic polyomavirus |
| LSTc | Lactoseries tetrasaccharide c |

| | |
|---|---|
| LTAg | Large T antigen |
| mPyV | Murine polyomavirus |
| MCPyV | Merkel cell polyomavirus |
| ML | Mother liquor |
| MSA | Multiple sequence alignment |
| NMR | Nuclear magnetic resonance |
| Neu5Ac | *N*-acetylneuraminic acid |
| Neu5Gc | *N*-glycolylneuraminic acid |
| NJPyV | New Jersey Polyomavirus |
| NLS | Nuclear localization sequence |
| PCR | Polymerase chain reaction |
| pLNnH | Para-lacto-*N*-neohexaose |
| PDB | Protein Data Bank |
| PML | Progressive multifocal leukoencephalopathy |
| PVAN | Polyomavirus-associated nephropathy |
| PyV | Polyomavirus |
| RCR | Rolling-circle replicating |
| RCRE | Rolling-circle replicating endonuclease |
| RMSD | Root mean square deviation |
| SDS-PAGE | Sodium dodecyl sulfate-polyacrylamide gel electrophoresis |
| NaBH$_3$CN | Sodium cyanoborohydride |
| SEC | Size exclusion chromatography |
| ShPyV | Sheep polyomavirus |
| SJR-CP | Single jelly roll capsid protein |
| sLeX | 3'-Sialyl Lewis$^X$ |
| SOPyV | Sea otter polyomavirus |
| SF3 | Superfamily 3 |
| SPR | Surface plasmon resonance |
| sTAg | Small T antigen |
| STD | Saturation transfer difference |
| SV40 | Simian virus 40 |
| TEMED | Tetramethylethylenediamine |
| TEV | Tobacco Etch Virus |
| TMAO | Trimethylamine *N*-oxide |
| TSPyV | *Trichodysplasia spinulosa*-associated polyomavirus |
| VP1 | Viral Protein 1 |
| VP2/3 | Viral Protein 2 and 3 |

# Chapter 1: Introduction

## 1.1 The origin of viruses

Viruses are biological entities that rely on hosts for their reproduction. They do not possess an intrinsic metabolism and are not considered life forms. Generally, viruses hijack host cells and use the resident transcription and translation machinery to replicate, often accompanied by host cell destruction. The origin of viruses is considered to coincide or even predate the emergence of the first cellular life forms such as early archaea and bacteria [3, 4]. The virus world theory postulates that self-replicating genetic elements, which have emerged during abiogenesis, used the metabolism of organic clusters that eventually evolved into the initial cellular life [4, 5]. Two other theories of viral provenance include the cell regression hypothesis, in which viruses descended from a now-extinct cellular life form, and the escape hypothesis proposing a polyphyletic virogenesis from genetic elements of cellular organisms [4, 6–8]. Synoptic to these theories is that the evolutionary paths of viral agents and cells have been intertwined since the formation of complex molecular biology on Earth. After billions of years of conjoint evolution, the contemporary virosphere exhibits a vast diversity of viruses that significantly vary in genome type, size, organization, and the composition and shape of their particles [9]. The sizes of viral genomes range from under two kilobases for the porcine circoviruses up to megabases for large DNA viruses (e.g., pandoraviruses) with a predicted number of over 2000 encoded proteins [10–13]. Viruses, in stark contrast to cellular life forms, employ a wide range of nucleic acid manifestations to store their genetic information. These include both DNA and RNA, which can be in single-stranded (ss) or double-stranded (ds) forms and can be linear, circular, or even segmented [9].

Viruses are the most abundant biological entities on the planet [14, 15] and constitute the largest reservoir of genetic information in the biosphere [9]. Due to the vast diversity, achieving a conclusive and unified evolutionary model for the origin of viruses is challenging [13]. Yet, certain viral genes encode proteins that are usually absent in cellular life and were frequently used to detangle viral lineages [3, 4, 6, 16]. These proteins, including superfamily 3 (SF3) helicases, rolling-circle replicating endonucleases (RCREs), and beta-barrel "jelly roll" capsid proteins, allow for the identification of gene transfer events and have significant implications for the evolution of viruses [4]. The acquisition of a capsid that bears and protects the viral genome is often regarded as the "cornerstone in the origin of viruses" (Krupovič & Bamford [17]). Likewise, the structural characterization of viral particles and their capsid proteins significantly impacted the classification

of viral families. Protein structure homology may unveil evolutionary relationships where sequence-based procedures are hindered by low genetic similarity [18–20]. The structural similarity of capsid proteins from viruses infecting hosts across different domains suggests a common capsid protein lineage [18, 21]. The extensive prevalence of structurally homologous capsid proteins without traces of genetic conservation indicates their ancient origin [17].

### 1.1.1 The icosahedral capsid and quasi-equivalence

Most viral families comprise members that employ icosahedral capsids, and these viruses infect hosts from all domains of life [22–24]. The icosahedron is a polyhedron with 20 triangular faces and twelve fivefold vertices, conforming to the symmetry of the icosahedral 532 point group. With the genetic material inside, the icosahedral capsid allows for efficient packing with minimal space requirements. It is constructed by a single or very few capsid proteins, wherein the capsid proteins constitute the tiles of a regular spherical pattern [25]. The first elaborate description of a viral capsid with icosahedral 532 point group symmetry was by Caspar [22], proposing that the protein shell of the tomato bushy stunt virus[1] comprises 60 identical asymmetric units (capsomers), which can be composed of smaller protein subunits called protomers. Later, it was found that many icosahedral viruses deviate from the perfectly symmetrical composition of 60 capsomers (or a multiple thereof), as seen in the 72 capsomers of papilloma- and polyomaviruses particles [27]). To overcome this systematic problem, Caspar & Klug [28] instituted the theory of quasi-equivalence, which allows the construction of icosahedral capsids without an identical environment for all subunits. In other words, a spherical arrangement of subunits with quasi-equivalent (but not identical) interactions will result in a quasi-symmetrical capsid. Caspar and Klug also described the icosahedron as a three-dimensional arrangement of equilateral triangles ("triangulation"), which enabled the classification of the capsid architectures by the triangulation number T (for examples, see Figure 1) [28]. The triangulation number can be calculated from the count of hexagons connecting one vertex with another, in direct (h) or indirect (k) direction using the formula $T = h^2 + hk + k^2$ (see Figure 9). In the simplest case (T=1), each triangular face of the icosahedron comprises three subunits, which results in a total number of 60 subunits for the whole capsid [22, 25]. Higher triangulation numbers exist, and a higher resolved structure of the tomato bushy stunt virus mentioned above revealed that it pos-

---

[1]In the original publication, Caspar referred to the "bushy stunt virus" without specifying the plant host in his article. However, an amino acid analysis from that time showed they were working with the tomato bushy stunt virus (TBSV) [26].

sesses 180 capsid proteins arranged in a T=3 capsid [29]. From $T \geq 7$ onwards, two possible capsid enantiomorphs occur, which are defined by differing k directions and are distinguished by the suffixes d (*dextro*) and l (*laevo*).



| T=1 | T=3 | T=4 | T=7d |
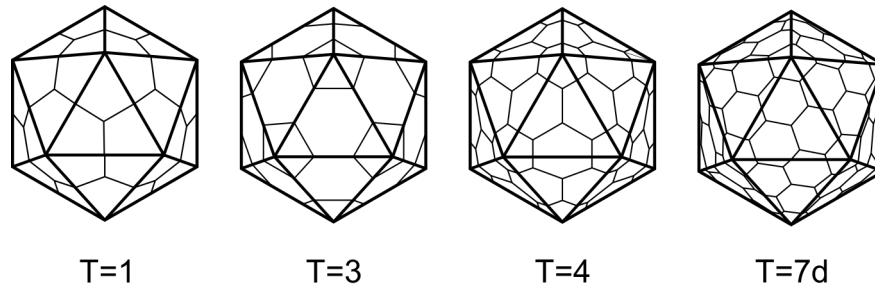
Figure 1: Icosahedral capsid architectures. For the triangulation number (T) calculation, see also Figure 9.

### 1.1.2 Jelly roll proteins

Many icosahedral capsids comprise proteins featuring a specific tertiary structure, first described in the tomato bushy stunt virus and later designated beta-barrel or the "jelly roll" fold [30]. The jelly roll (JR) fold is a structural motif of eight beta strands (historically labeled B – I), which fold into two pairs of antiparallel beta sheets, BIDG and CHEF (see Figure 2d). Sustained mainly by hydrophobic interactions, the JR fold adopts a barrel-like structure [31]. The JR fold is present in the structural proteins of dsRNA and dsDNA viruses, ssRNA viruses of plants and animals, and bacteriophages [31–35]. Initially, single jelly roll capsid proteins (SJR-CPs), which occur in about 30% of viral taxa, were considered virus-exclusive and reflect an ancient acquisition event during viral evolution [8]. However, more recent advances in structural and computational biology have revealed that the provenance of SJR-CPs are different carbohydrate-binding proteins found in cellular life [8, 36]. Computational analyses also signified that the capsid proteins of many ssDNA viruses are evolutionarily linked to the JR proteins of RNA viruses. Consequently, the lineage of these ssDNA viruses was proposed to originate from plasmids that recombined with reversely transcribed capsid protein genes of RNA viruses at various times during evolution [37].

## 1.2 Polyomaviruses

The polyomaviruses (PyVs) constitute a viral family, the *Polyomaviridae*, of currently over 110 officially assigned species that infect mammals, birds, and fish. Polyomaviruses

are non-enveloped double-stranded DNA viruses with a circular genome of ~5000 base pairs. The viral DNA is packed inside a capsid and associated with host cell histones. The genomes organize into a region for the early genes, encoding the regulatory proteins large T antigen (LTAg) and small T antigen (sTAg), and a late region with genes for the major capsid protein, Viral Protein 1 (VP1), minor capsid proteins, Viral Protein 2 and 3 (VP2/3), and, in avian polyomaviruses, a structural VP4 [38]. Some PyVs encode the non-structural agno protein, which participates in processes such as replication, particle maturation, and release [39]. The non-coding control region of the genome, situated between the early and late regions, holds the origin of DNA replication and genetic elements guiding transcription [40, 41]. The infection with a polyomavirus is usually asymptomatic and results in latency. However, in immunocompromised individuals, certain polyomaviruses, such as BK polyomavirus (BKPyV) and JC polyomavirus (JCPyV), can cause severe diseases, including the polyomavirus-associated nephropathy (PVAN) and progressive multifocal leukoencephalopathy (PML). Merkel cell polyomavirus (MCPyV) is causally linked to Merkel cell carcinoma, an aggressive form of skin cancer [42]. Other human polyomaviruses are associated with milder conditions such as dermatoses, namely *Trichodysplasia spinulosa*-associated polyomavirus (TSPyV), human polyomavirus 6 (HPyV6) and human polyomavirus 7 (HPyV7) [43–45]. In mice, the murine polyomavirus causes several tumors, which was eponymous for the virus family. Other severe viral diseases of animals are linked to the budgerigar fledgling disease polyomavirus (BFDPyV) and a raccoon polyomavirus found in brain tumors of raccoons [46–48]. The vacuolating virus, simian virus 40 (SV40), which was discovered as a contamination in early batches of human polio vaccine, is associated with carcinogenesis in rodents [49, 50]. Ever since, the studies of polyomaviruses such as SV40 and their tumor antigens have also led to groundbreaking insights into the molecular mechanism of DNA replication and tumor suppression [51]. SV40 LTAg has also regularly been employed in the establishment of immortalized cell lines [52, 53].

### 1.2.1 Discovery and structure

The first known polyomavirus, murine polyomavirus (mPyV), was described as an infectious viral particle with oncogenic potential in mice and a size of 30-60 nm in diameter as determined by electron microscopy in the late 1950s [54–56]. During the 1970s, the polyomavirus capsid was resolved to comprise 72 capsomers, which is similar to the human wart papillomavirus [27, 57, 58]. In 1982, the crystal structure of mPyV was solved to 22.5 Å. It showed for the first time that all capsomer positions, including the morphological hexamers, were occupied by a pentameric VP1, revealing that the poly-
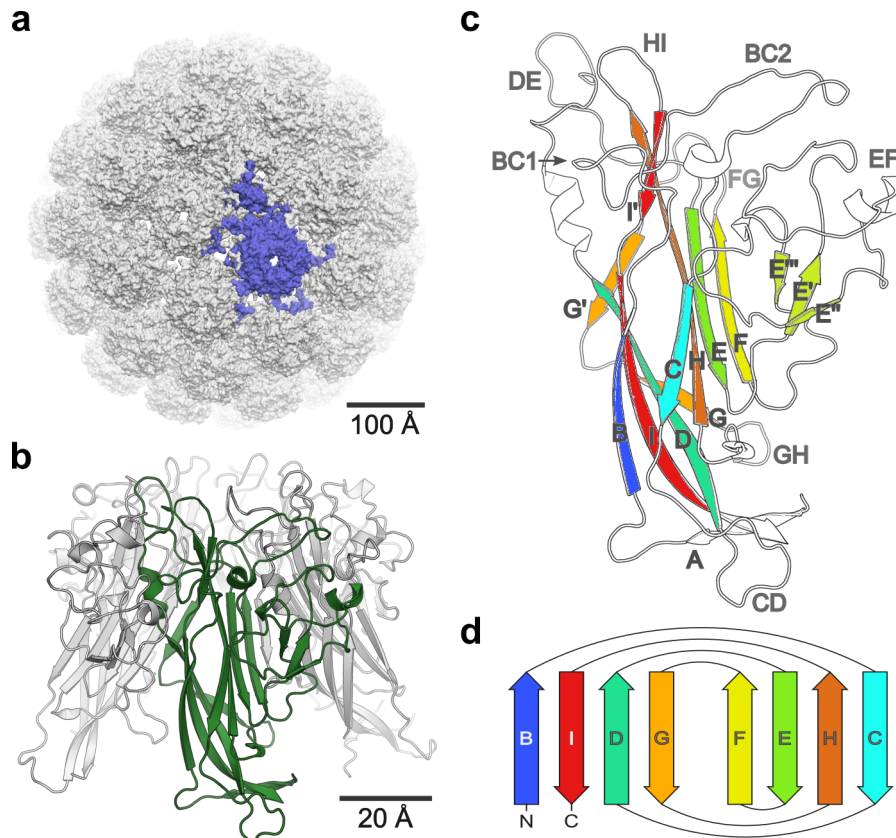
Figure 2: The structure of the outer polyomavirus capsid and VP1. **a)** The capsid structure of SV40 at 3.1 Å resolution, Protein Data Bank (PDB) entry 1SVA. One asymmetric unit of the capsid, comprising six individual VP1 chains, is highlighted in blue. **b)** Structure of the BKPyV VP1 pentamer at 1.7 Å resolution (PDB 4MJ0) with one chain colored green. **c)** The structure of a VP1 chain in isolated representation. Beta strands and loops are labeled and colored according to **d)**, the jelly roll fold topology.

omavirus capsid holds 360 copies of VP1 [23]. The polyomavirus capsid is icosahedral with a T=7d architecture (see Figure 1). The asymmetric unit of the capsid comprises six individual VP1 chains (see Figure 2a), in which every C-terminal arm undergoes distinct interactions with neighboring chains, establishing quasi-symmetry [34]. At that, VP1 is sufficient to form a capsid without the help of posttranslational modifications or minor capsid proteins [59]. Each VP1 chain comprises an N-terminal arm containing positively charged amino acid residues for viral DNA interaction and cysteines that form disulfide bridges to stabilize the capsid [60–63]. The basic amino acids in the N-terminal arms also often serve as a nuclear localization sequence (NLS) [64–66]. The polyomavirus capsomer is a homopentameric VP1 (see Figure 2b). The core domain of VP1 adopts a JR fold with the beta strands linked via extended protein loops. While the BC, DE, EF, and HI loops form the outer surface of the capsid, the FG, GH, and CD loops are situated inside (see Figure 2c). The CD loop participates in the formation of inter-chain disulfide

bonds for capsid stabilization [34, 63, 67, 68]. Among polyomavirus species, VP1 sequences exhibit the most variation in the outer surface loops, serving as the primary site of receptor engagement and antibody recognition [69–71].

## 1.2.2 Evolution and taxonomy

Recent advances in computational biology helped to date back the lineage of polyomaviruses to rolling-circle replicating (RCR) plasmids that held a superfamily 3 (SF3) helicase gene [37]. Via recombination with reversely transcribed single jelly roll capsid protein (SJR-CP) genes of RNA viruses, these plasmids acquired the ability to encapsulate their DNA, resulting in the formation of several small RCR ssDNA viruses [72, 73]. The direct ancestor of polyomaviruses (and papillomaviruses) was a member of the *Parvoviridae*, vertebrate-infecting linear ssDNA viruses featuring icosahedral capsids of pentameric SJR-CPs [37, 74]. The transition from a linear ssDNA to a circular dsDNA genome, as outlined in Figure 3a, was the critical step in polyomavirus evolution. Due to their lineage, PyVs contain rolling-circle replicating endonuclease (RCRE) and SF3 helicase domains in their LTAg. However, the RCR mechanism was lost, and the inactive RCRE now fulfills non-enzymatic tasks [37, 73].

Polyomaviruses have speciated in association with their hosts over approximately 500 M years, following a co-evolutionary pathway [76]. Host switching also occurs, resulting in several closely related PyVs that infect distant hosts such as bats and apes [76, 77]. In some cases, PyVs evolved faster than their hosts, causing intra-host divergence as observed in the human BKPyV and JCPyV [76]. Ancient recombination events between the early and late regions of distantly related PyVs resulted in incongruent phylogenetic trees of LTAg and VP1 [76]. Species assignment of polyomaviruses is based on the LTAg phylogeny, following these criteria: (i) Availability of the complete genome, (ii) family typical genome organization, (iii) availability of natural host information, (iv) genetic distance to the next relative amounts to >15% for the LTAg sequence, and (v) if the distance is <15%, biological properties such as host specificity and distinct pathogenicity may also allow for the assignment of a new species [78].

Today, polyomaviruses fall into six different genera, which are designated *Alpha*-, *Beta*-, *Gamma*-, *Delta*-, *Epsilon*-, and *Zetapolyomavirus* (see Figure 3b) [79].

The alphapolyomaviruses constitute the largest genus with over 50 PyV species. Members of this genus primarily infect mammals, including humans [77, 80]. The human MCPyV and TSPyV are alphapolyomaviruses.

The second largest genus, *Betapolyomavirus*, has over 40 members that infect mammals, including some of the most intensively studied PyV species such as SV40 and the
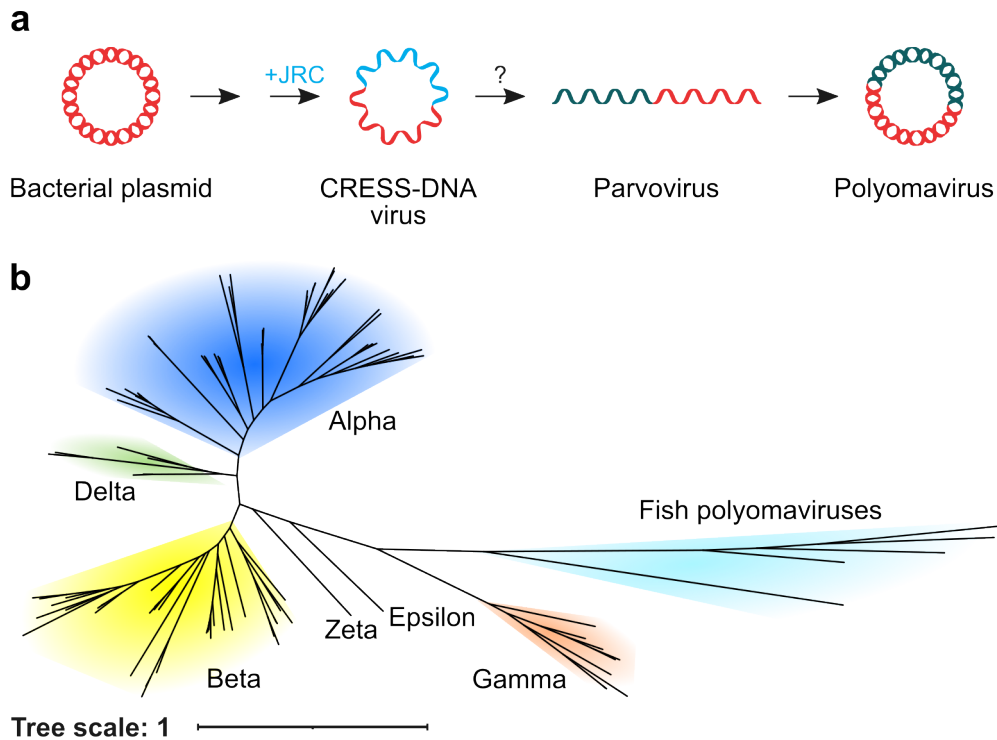
Figure 3: The evolution and taxonomy of polyomaviruses. **a)** The proposed evolutionary route of polyomaviruses' early and late genomic regions. The lineage of plasmid genes/ viral regulatory proteins is indicated in red. Genes for the (non-orthologous) capsid proteins are colored in blue or dark teal. Panel adapted from Kazlauskas *et al.* [37], Figure 6 (distributed under the CC BY 4.0 deed, `https://creativecommons.org/licenses /by/4.0/`). **b)** Unrooted phylogenetic tree of polyomaviruses. Background colors unify branches corresponding to the same genus. The scale indicates the number of substitutions per site. Tree file downloaded from https://ictv.global/ and compiled via iTOL [75].

human pathogenic BKPyV and JCPyV.

The gammapolyomaviruses exclusively infect birds with currently nine species. One member, the BFDPyV, is a severe pathogen of young psittacines.

The *Deltapolyomavirus* genus holds seven species, four of which were found in humans [78]. In contrast to many other human polyomaviruses to date, the deltapolyomaviruses lack structurally characterized receptors [81].

*Epsilonpolyomavirus* comprises three species that infect artiodactyls[2] [79, 82].

*Zetapolyomavirus* holds the single species *Zetapolyomavirus delphini*, isolated from the trachea of a dolphin [83].

Several polyomaviruses isolated from fishes and arthropods have no assigned genus yet.

---

[2]The Artiodactyla order includes even-toed ungulates, such as bovine animals and swine.

## 1.2.3 Receptors and binding of sialic acid moieties

Polyomaviruses use their VP1 proteins to recognize structures on their target cells, enabling them to attach to the outer cell membrane and eventually infect the host. These target structures (receptors), among others, comprise glycolipids, such as gangliosides (see Figure 8), and glycoproteins bearing terminal sialic acids [84]. Sialic acids are derivatives of neuraminic acid (see Figure 4), and in humans, the most common sialic acid is *N*-acetylneuraminic acid (Neu5Ac) [85]. Different polyomaviruses have acquired different specificities for certain sialic acid modifications, linkages, or the overall sequence of the carbohydrate moiety [86].
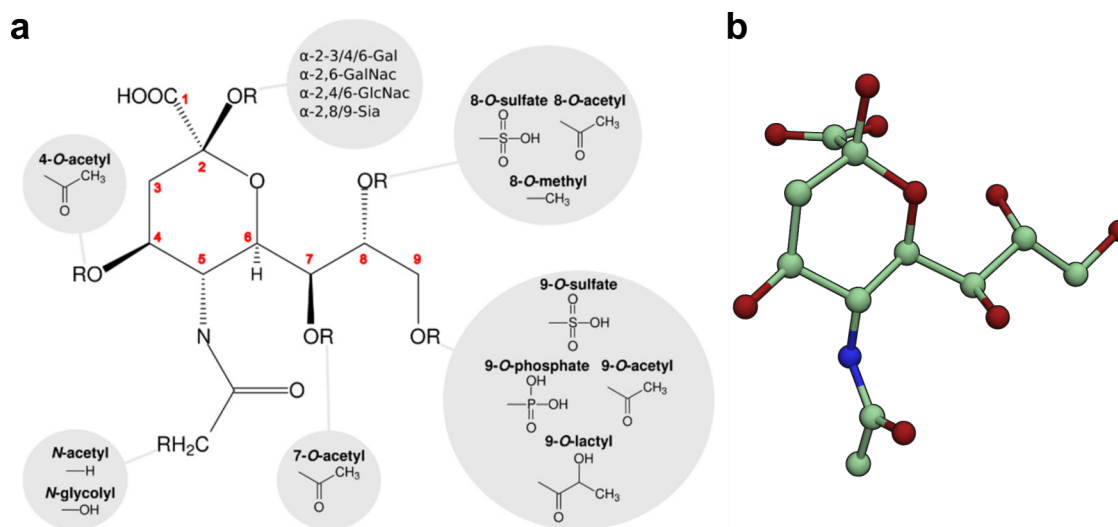


Figure 4: **a)** Outline of the *N*-acetylneuraminic acid (Neu5Ac) scaffold with the possible modifications found in sialic acids. Red numbers indicate carbon positions. **b)** Three-dimensional structure of *N*-acetylneuraminic acid (Neu5Ac) in ball and stick representation, derived from PDB entry 3BWR [87]. Oxygen, nitrogen, and carbon atoms are colored red, blue, and pale green, respectively. Illustration taken from Figure 1 of Rustmeier, N. H. *et al. The symmetry of viral sialic acid binding sites-implications for antiviral strategies* 2019 [88], distributed under the CC BY 4.0 deed (License: https://creativecommons.org/licenses/by/4.0/).

While not all polyomaviruses investigated to date bind sialylated ligands [81, 89], their VP1 sialic acid binding sites share a common location on the peripheral surface of the pentameric protein, constituted by loops protruding from the core domain. To date, six different binding orientations of sialic acid moieties have been identified in association with VP1 proteins (see Figure 5).

(i) mPyV engages the glycans of gangliosides GT1a, GD1a, and GT1b for infectious entry, all of which contain a Neu5Acα2-3Gal binding motif, in a site constituted by residues of the BC and HI loops [67, 90].
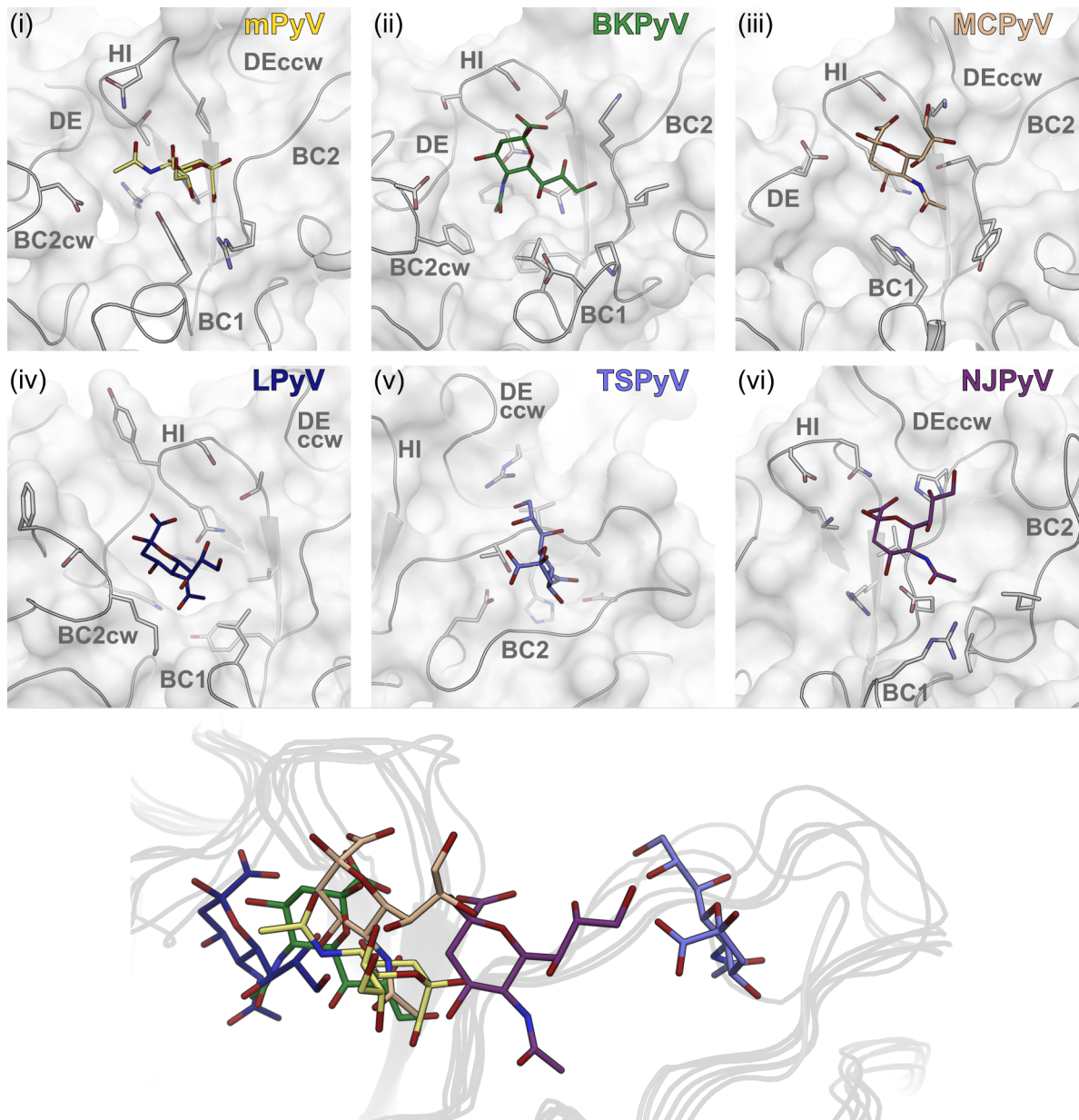
Figure 5: Sialic acid moiety binding orientations in the VP1s of (i) murine polyomavirus (mPyV), (ii) BK polyomavirus (BKPyV), (iii) Merkel cell polyomavirus (MCPyV), (iv) B-lymphotropic polyomavirus (LPyV), (v) *Trichodysplasia spinulosa*-associated polyomavirus (TSPyV), and (vi) New Jersey Polyomavirus (NJPyV). In panels (i)-(vi), all proteins are shown as transparent surfaces with an underlying cartoon representation. Loops are labeled in bold gray. The side chains of residues involved in sialyl binding are shown as sticks. The bottom panel displays the six orientations in superposition, colored according to (i)-(vi). All oxygen and nitrogen atoms are colored red and blue, respectively.

(ii) The simian and human polyomavirus species SV40, JCPyV, and BKPyV preferably recognize sialic acid moieties in an α2-3, α2-6, and α2-8 linked manner, respectively [87, 91, 92]. Some JCPyV strains also bind the α2-3 and α2-8 linked sialyl glycans of gangliosides [93]. While the human viruses engage receptors carrying a terminal Neu5Ac, the simian SV40 preferably binds to *N*-glycolylneuraminic acid (Neu5Gc), which carries an additional hydroxylation at the acetyl group (see again Figure 4a) and is not present in humans [94]. Their mutual binding site features a highly conserved hydrophobic pocket between the BC, DE, and HI loops, and the surrounding amino acid residues determine the differences in receptor specificity [87, 91, 92].

(iii) MCPyV engages the glycan of GT1b ganglioside with a Neu5Acα2-3Gal recognition motif in a binding site constituted by the HI and BC loops, yet differently to mPyV: While in the mPyV complex the amide and carboxyl groups point in parallel direction of the pentamer axis, in MCPyV these groups are perpendicular to it [95].

(iv) The simian LPyV and its counterpart, human polyomavirus 9 (HPyV9), feature a binding site resembling that of SV40 and BKPyV, yet narrower and more recessed. The binding motif corresponds to the linear α2-3 linked 3-sialyllactose (3'SL) or 3'-sialyl-*N*-acetyllactosamine (3'SLN); however, no information on the engagement of specific receptors is available [96].

(v) The fifth sialic acid binding orientation was characterized in the alphapolyomavirus TSPyV. It was later also found in the betapolyomavirus from sheep, sheep polyomavirus (ShPyV), and the gammapolyomaviruses of birds [1, 97]. This binding site is formed by the anterior BC loop (BC2 loop) with just one additional contact with a DE loop residue. Glycolipids mediate the attachment of TSPyV to cells, and the binding site accepts both α2-3 and α2-6 linked sialic acids.

(vi) The latest sialic acid binding orientation was found in NJPyV, the 13th human polyomavirus [1]. The site is located towards that of TSPyV (see Figure 5) and comprises residues of the BC and HI loops, which again differ from mPyV and MCPyV. NJPyV VP1 was complexed with linear 3'SL, suggesting the engagement of glycolipids.

For a few polyomaviruses, receptors not belonging to the classes of sialylated glycolipids and glycoproteins were characterized: MCPyV is known to use glycosaminoglycans (GAGs) for initial cellular attachment [98], which is also discussed for BKPyV and JCPyV [99, 100]. JCPyV further requires serotonin receptors for entry [101–103]. However, no detailed structural information on these sialic acid-independent interactions is available.

## 1.2.4 Endocytosis and the polyomavirus life cycle

After attaching to their host cells, polyomaviral particles (virions) enter the cell via endocytosis. Packed into vesicles, the virions are routed towards the endoplasmic reticulum (ER) via retrograde transport, translocate into the ER lumen, and eventually enter the nucleus, the site of viral replication. Most commonly, the endocytosis of polyomaviruses proceeds via clathrin-, caveolae-, or lipid raft-dependent pathways [104–106]. While host factors, such as clathrin or caveolin-1, aid the vesiculation, virions interacting with their receptors alone can invaginate membranes due to the spherical distribution of the binding sites on their capsid [107]. Recently, polyomaviruses such as JCPyV and BKPyV were also shown to invade host cells via extracellular vesicles [108, 109].

Following the internalization, the exact mechanisms that govern the transport towards the ER differ between polyomaviruses, involving different cellular proteins and stages of endosomes [84, 110–114]. For example, JCPyV utilizes a clathrin-mediated endocytic pathway and localizes in early but not in late endosomes or endolysosomes [115], whereas SV40 associates with both early and late endosomes [105]. Despite these differences, the trafficking towards the ER is active and progresses along microtubules as shown for the PyV species mPyV, MCPyV, and BKPyV [116–119].

After the virions enter the ER lumen, isomerases break up intermolecular disulfide bonds and destabilize the capsid [120–123]. As a result, the PyV virions (partially) uncoat, which mediates the exposure of the minor capsid proteins. Then, the virions make contact with the ER membrane and its transmembrane proteins to initiate the egress into the cytosol (see Figure 6) [120, 124, 125].

VP1 but also the minor capsid proteins VP2 and VP3 of SV40 comprise NLSs needed for the translocation of the uncoated virions into the nucleus [126–128]. At that, the signal sequences interact with importins, which then mediate the transport through nucleoporins [129–131]. While BKPyV infection requires the presence of VP2/3 NLS, the NLS of JCPyV VP1 alone is sufficient for nuclear translocation [66, 132]. For mPyV, the VP1 NLS suffices, but VP2/3 also contribute to translocation [133]. Reports about direct translocation from the ER into the nucleus are disputed [134] and, conclusively, the processes that guide viral DNA uptake remain not fully understood [135]. Within the nucleus, cellular RNA polymerases initiate transcription of the early genes, leading to the expression of T antigens. After migrating into the nucleus, the LTAg helicase activity induces viral DNA replication [51]. Eventually, the T antigens activate the late genes of PyVs, priming the expression of capsid proteins. All viral capsid components translocate into the nucleus, where progeny virions assemble. Upon their release from the cell, the viral life cycle is completed.
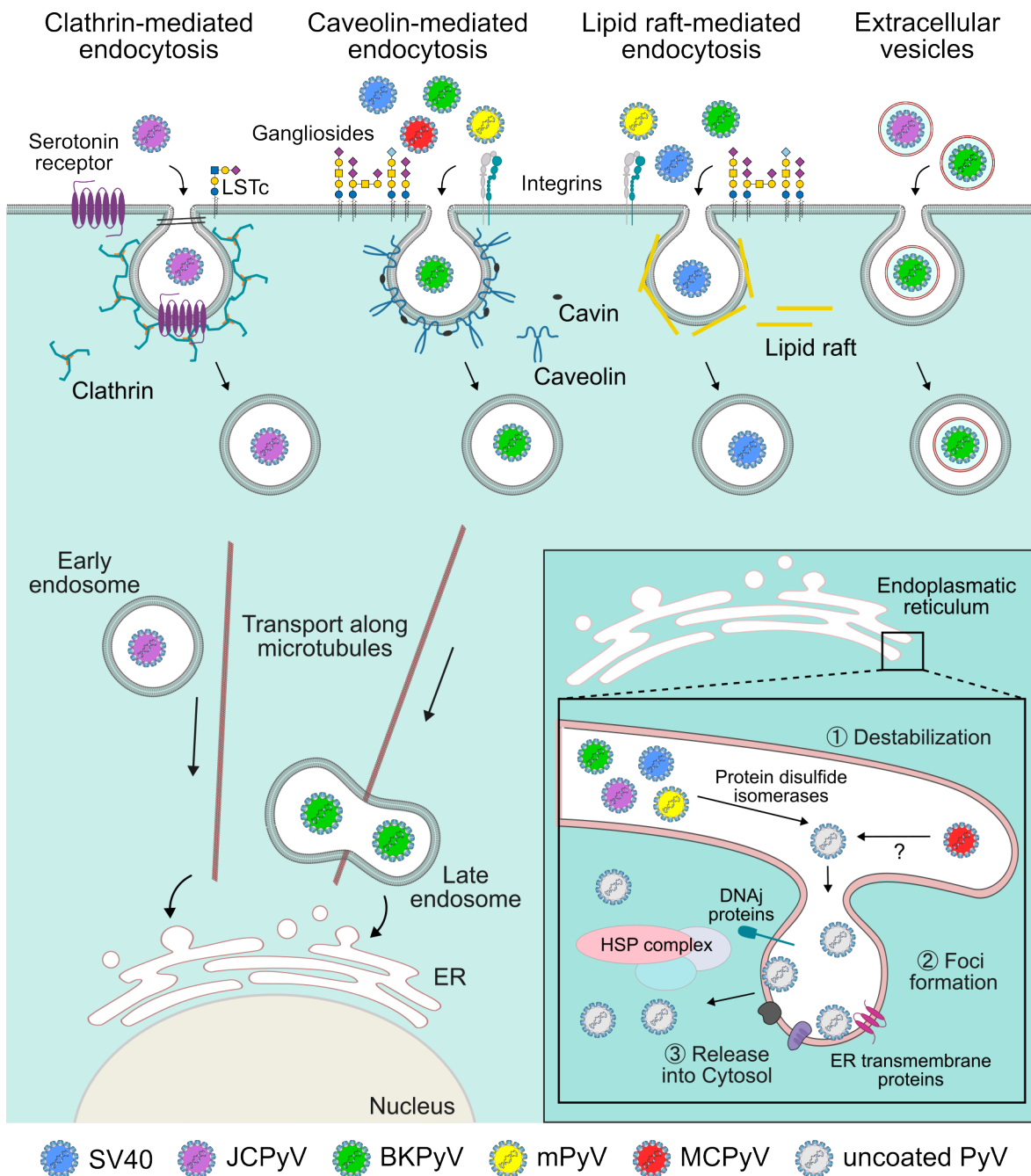
Figure 6: Endocytosis and intracellular trafficking of polyomaviruses. Illustration based on Figures 2, 3, and 4 of Mayberry, C. L. & Maginnis, M. S. *Taking the scenic route: Polyomaviruses utilize multiple pathways to reach the same destination* 2020 [135], published under the CC BY 4.0 deed (`https://creativecommons.org/licenses/by/4.0/`). Legend: ER - endoplasmic reticulum, HSP - heat shock protein, LSTc - lactoseries tetrasaccharide c.

## 1.2.5 Polyomavirus species in this work

An overview of the genomes of the viruses named below is provided in Figure 7.

### 1.2.5.1 Sheep polyomavirus

The sequence of a sheep-associated polyomavirus, designated sheep polyomavirus (ShPyV), was isolated from a lamb meat sample in 2016 [76]. Another study investigating the viral load in sheep feces, however, did not find the sheep polyomavirus [79]. In phylogenetic analyses of LTAg, ShPyV manifests as a betapolyomavirus and clusters with several bat polyomaviruses, horse polyomavirus, mouse polyomavirus 2 (aka murine pneumotropic virus, MPtV), multimammate mouse polyomavirus, and lion polyomavirus. It is distantly related to the lineage of SV40, BKPyV, and JCPyV, which diverged earlier during betapolyomavirus evolution [76, 77]. In Ströh *et al.* [1], we have shown that ShPyV features a sialic acid moiety binding site on its VP1 surface that was previously characterized in the alphapolyomavirus TSPyV and is also present in the VP1s of the avian gammapolyomaviruses. In Rustmeier *et al.* [2] attached to this thesis, we introduced a novel binding site for the Forssman antigen, which is a non-sialylated glycosphingolipid (GSL) and a potential tumor marker in humans [136]. This thesis will provide complementary results about ShPyV VP1.

### 1.2.5.2 Chimpanzee polyomavirus

The DNA sequence of chimpanzee polyomavirus (ChPyV) was isolated from the feces of a captive animal (a chimpanzee named Bob) and published in 2005 [137]. A subsequent screening based on the initial ChPyV VP1 sequence by Deuzing *et al.* [138] further identified two distinct genotypes (1 and 2), the second of which was divided into the subclusters 2A and 2B. The tropism of this ChPyV was suggested to be for skin and spleen [138]. Phylogenetically, the ChPyV variants cluster with alphapolyomaviruses such as LPyV, found in the African green monkey, Orangutan PyVs, human TSPyV and MCPyV, and two rodent PyVs [77, 138]. Additionally, ChPyV is a close relative of the subsequently discovered human NJPyV, with which it forms a "human-simian pair" [139]. In Ströh *et al.* [1], we proposed that ChPyV VP1 binds sialylated ligands in the same manner we described for NJPyV VP1. In this thesis, I describe the complex structures of ChPyV VP1 with 3'-sialyl Lewis$^{X}$ (sLeX) pentasaccharide and two non-sialylated oligosaccharides that bind to a novel galactosyl binding site.

### 1.2.5.3 Lyon IARC polyomavirus

Lyon IARC polyomavirus (LIPyV), initially described as the 14th human polyomavirus [140], was later found in fecal samples of diarrheic cats in Canada and China [141, 142]. A seroreactivity study involving 152 human individuals did not detect antibodies against LIPyV, suggesting LIPyV is a feline rather than a human polyomavirus [143]. LIPyV is closely related to a puma polyomavirus and the tumor-associated raccoon polyomavirus (RacPyV) with overall LTAg sequence identities of ~72% and ~65%, respectively [140]. The alphapolyomavirus also clusters with human oncogenic MCPyV, primate, and bat polyomaviruses [77]. Its alleged human provenance and LIPyV's relationship to at least two oncogenic PyVs prompted its investigation. Using glycan array analysis and X-ray crystallography, I show that LIPyV VP1 does not bind sialylated ligands but neutral neolacto-series oligosaccharides in a recessed binding site.

### 1.2.5.4 Sea otter polyomavirus

The sea otter polyomavirus (SOPyV) was discovered in the bodies of deceased southern sea otters (*Enhydra lutris nereis*) by deep sequencing nucleic acids in the tissues of these animals. The virus was found in 51% of the 69 examined sea otter samples. At that, it was primarily associated with spleen tissue [144]. By LTAg sequence, SOPyV is a betapolyomavirus and related to the primate and human polyomaviruses SV40 and BKPyV. The structural investigations of SOPyV VP1 were prompted by the conservation of TSPyV binding site residues in its sequence, coinciding with the presence of the BKPyV binding site. I show that the sialyl binding site of BKPyV is conserved and active in SOPyV. In contrast, the TSPyV site is mostly conserved in the SOPyV VP1 sequence but does not engage sialyl ligands *in crystallo*.

### 1.2.5.5 Avian polyomavirus

The avian polyomavirus (APyV), species name *Gammapolyomavirus avis*, is a pathogen of psittacine birds such as parrots and budgerigars [46, 47, 145]. It can distinctively infect hosts from various species, inducing inflammatory diseases, and is the causative agent of the French molt, which is severe in young birds and gave rise to the common virus name budgerigar fledgling disease polyomavirus (BFDPyV) [38]. It must not be confused with the circovirus infection causing psittacine beak and feather disease (PBFD). The structural investigation of APyV VP1 complements the studies of SOPyV VP1. Characterizing the glycan engagement may offer insights into the virus's vast host range. Here, I describe the structure of APyV VP1 in terms of a prediction model.
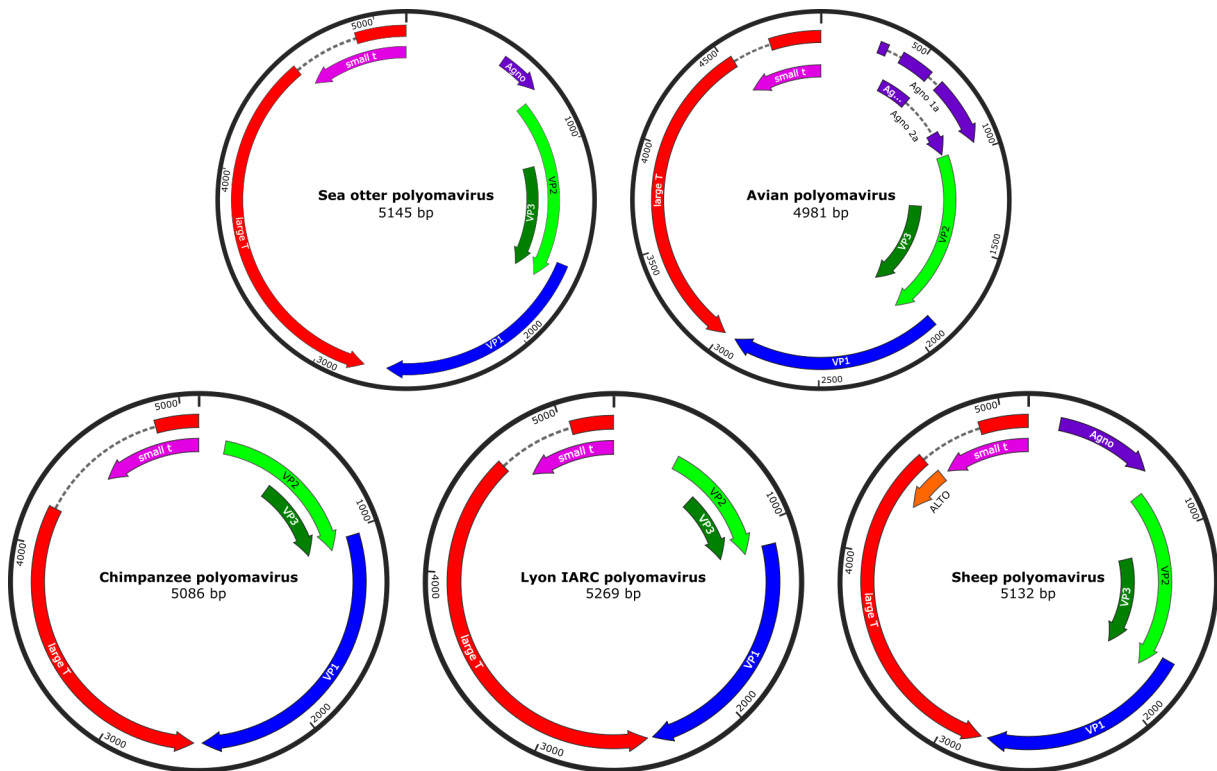
Figure 7: Genomic organization of the polyomaviruses in this work. Colored arrows indicate the direction of gene expression. Introns are drawn as dashed lines. Genomes correspond to the National Center for Biotechnology Information (NCBI) reference sequences NC_025259.1 (SOPyV), NC_004764.2 (APyV), NC_014743.1 (ChPyV), NC_034253.1 (LIPyV), and NC_026942.1 (ShPyV).

### 1.2.5.6 Objectives

In this thesis, I will describe the work that led to the glycan array screening of five VP1s from polyomaviruses found in the sea otter, chimpanzee, cat, sheep, and humans. Aims include the characterization of viral glycan engagement, facilitated by X-ray crystallography, to unveil host tropism factors and gain insights into viral evolution. This work will expand the catalog of established glycan receptor candidates of polyomaviruses by non-sialylated carbohydrates. For the sheep polyomavirus, I will describe new means to detect the Forssman antigen, a potential biomedical marker for cancer, at a cellular level.
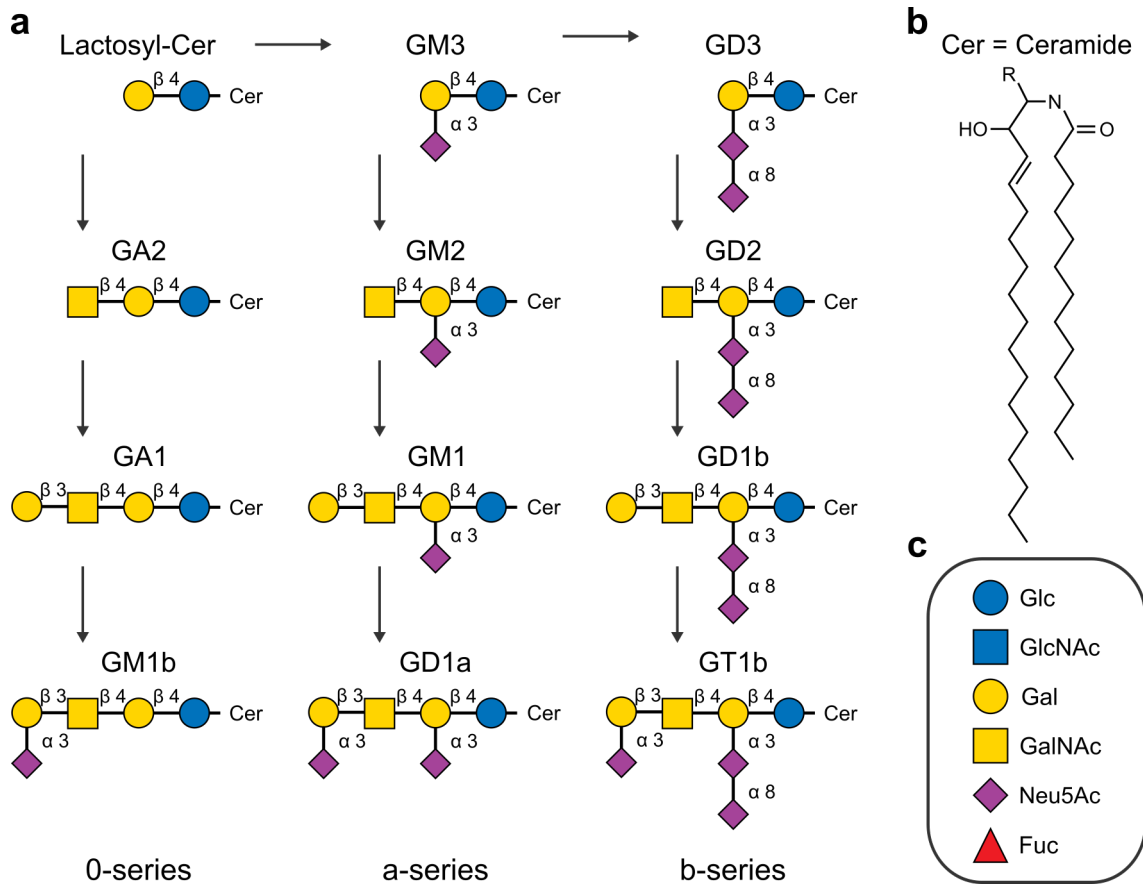
Figure 8: Overview of common cell surface gangliosides. **a)** Glycans drawn in symbol nomenclature. **b)** Structural formula of ceramide, the lipidic anchor of gangliosides. **c)** Legend of the Symbol Nomenclature for Glycans [146, 147].
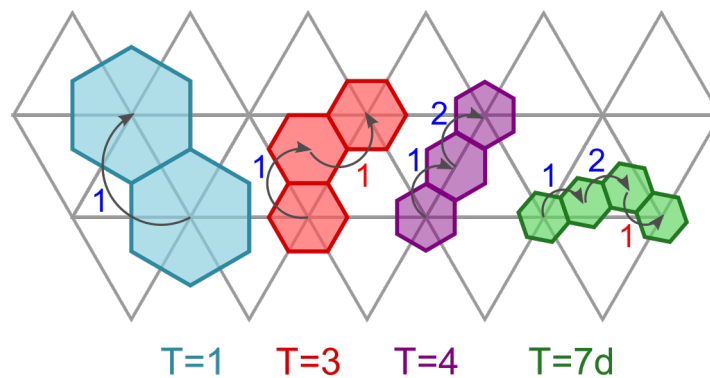


Figure 9: Depiction of the h (blue) and k (red) values for the calculation of the triangulation number of the icosahedral capsids in Figure 1 by the formula $T = h^2 + hk + k^2$.

# Chapter 2: Materials & Methods

## 2.1 Materials

### 2.1.1 Instruments

Table 1: An overview of the instruments used in this work.

| Instrument | Purpose | Manufacturer |
|---|---|---|
| 5415 D | Centrifuge | Eppendorf |
| 5418 | Centrifuge | Eppendorf |
| ÄKTA Basic | FPLC system | Cytiva |
| ÄKTAprime plus | LC system | Cytiva |
| ÄKTApurifier | LC system | Cytiva |
| BP11 | pH meter | Sartorius |
| Digital Sonifier | Bacterial lysis | Branson |
| Ettan LC | FPLC system | Cytiva |
| FastGene® TC01 | PCR cycler | Nippon Genetics |
| Freedom EVO® 150 | Crystallization robot | Tecan Life Sciences |
| iCycler | PCR cycler | Bio-Rad |
| Multifuge 1 L-R | Centrifuge | Heraeus |
| Multitron | Incubation shaker | Infors HT |
| Leica MZ 16 | Microscope | Meyer Instruments |
| Milli-Q® Biocel | Water purification | Merck |
| Nanodrop ND-1000 | Photometer | Thermo Scientific |
| Neoblock 1 | Heating block | NeoLab |
| PowerPac$^{TM}$ Basic | Power supply | Bio-Rad |
| SmartSpec$^{TM}$ Plus | Spectrophotometer | Bio-Rad |
| Sorvall RC 6 Plus | Centrifuge | Thermo Scientific |
|    SLC4000 | Rotor | Thermo Scientific |
|    SS34 | Rotor | Thermo Scientific |
| Superdex$^{TM}$ 200 HiLoad 16/60 | Preparative LC column | Cytiva |
| Superdex$^{TM}$ 200Increase 3.2/300 | Analytical LC column | Cytiva |
| Thermomixer comfort | Incubation shaker | Eppendorf |

Manufacturers are listed as the current brand owners.

## 2.1.2 Software

Table 2: An overview of the software used in this work.

| Software | Distributor/ Author |
| --- | --- |
| BLAST | NCBI [148] |
| Clustal Omega | European Bioinformatics Institute |
| Coot | P. Emsley [149] |
| Expasy translate | Swiss Institute of Bioinformatics |
| IQ-TREE 2 | IQ-TREE [150] |
| iTOL | EMBL[75] |
| Jalview | University of Dundee |
| Lalign | EBI[151] |
| xdlMAPMAN | G.J. Kleywegt & T.A. Jones [152] |
| MUSCLE | European Bioinformatics Institute[151] |
| Phaser | CCP4 [153] |
| Phenix | Phenix development group [154, 155] |
| Protcalc | sourceforge.net |
| PyMOL | Schrödinger |
| Refmac5 | CCP4 [156] |
| SeaView 5 | PRABI-Doua [157] |
| UCSF ChimeraX | RBVI [158] |
| XDS | W. Kabsch [159] |

## 2.1.3 Consumables

Table 3: An overview of the consumables used in this work.

| Consumable | Manufacturer |
| --- | --- |
| Amicon® Ultra-15 centrifugal filter | Merck |
| CELLSTAR® 24-well plate | Greiner Bio-One |
| Centrifugal filters (0.22 µm) | Corning |
| Crystal screens 1 & 2 | Hampton Research |
| Intelli-Plate 96-2 Original | Art Robbins |
| MF-Millipore™ Membrane Filter, 0.22 µm pore size | Merck |
| MicroAmp™ PCR reaction tubes | Applied Biosystems |

Table 3: An overview of the consumables used in this work (cont.).

| Consumable | Manufacturer |
| --- | --- |
| Morpheus crystal screen | Molecular Dimensions |
| PEG/Ion Screens 1 & 2 | Hampton Research |
| Pipet tips | Nerbe Plus |
| Reaction tubes | Greiner Bio-One |
| Syringe filters | VWR |
| Syringes | Greiner Bio-One |
| VIEWSEAL 96-well plate seals | Greiner Bio-One |
| Wizard® Plus SV Minipreps kit | Promega |
| Wizard I-IV crystal screens | Jena Bioscience |

## 2.1.4 Chemicals

Table 4: An overview of the chemicals used in this work.

| Chemical | Manufacturer |
| --- | --- |
| 2-*O*-Methyl-$\alpha$-D-*N*-acetylneuraminic acid | Sigma-Aldrich |
| Acetic acid, glacial | Roth |
| Acrylamide/Bis solution (Rotiphorese® Gel 30) | Roth |
| Ammonium chloride | Roth |
| Ammonium persulfate | Roth |
| Boric acid | Sigma-Aldrich |
| cOmplete™ protease inhibitor cocktail | Roche |
| $D_2O$ | Cortecnet |
| dNTP (10 mM mix) | Sigma-Aldrich |
| Glycerol | Roth |
| HEPES | Roth |
| Imidazole | Roth |
| InstantBlue® | Abcam |
| Isopropanol | Sigma-Aldrich |
| Lactose | Sigma-Aldrich |
| LB broth (Miller) | Sigma-Aldrich |
| Polyethylene glycol 3350 | Sigma-Aldrich |
| Polyethylene glycol 8000 | Sigma-Aldrich |

Table 4: An overview of the chemicals used in this work (cont.).

| Chemical | Manufacturer |
| --- | --- |
| Potassium phosphate dibasic | Sigma-Aldrich |
| Potassium phosphate monobasic | Sigma-Aldrich |
| ReadyBlue™ | Sigma-Aldrich |
| Running buffer (Rotiphorese® 10x SDS-PAGE) | Sigma-Aldrich |
| Sodium acetate | Sigma-Aldrich |
| Sodium cyanoborohydride | Sigma-Aldrich |
| Sodium dodecyl sulfate (SDS) | Roth |
| Sodium chloride | Sigma-Aldrich |
| Sodium sulfate | Roth |
| Sodium tetraborate decahydrate | Sigma-Aldrich |
| Tetramethylethylenediamine (TEMED) | Roth |
| Trizma (base or hydrochloride) | Sigma-Aldrich |

## 2.1.5 Biochemicals and biological products

Table 5: An overview of the biochemicals and biological products used in this work.

| Biochemical or biological product | Manufacturer/Vendor |
| --- | --- |
| Bovine serum albumin (BSA) | Sigma-Aldrich |
| TEV protease | *Lab-made* |
| Thrombin | Cytiva |
| BamHI | NEB |
| NcoI | NEB |
| NdeI | NEB |
| Bovine blood (defibrinated) | bioTRADING Benelux B.V. |
| Sheep blood (defibrinated) | bioTRADING Benelux B.V. |
| Sheep blood (defibrinated) | TCS Biosciences |
| Human reference cells A1A2BO | Immucor |

The certified vendors of biological products guaranteed animal well-being.

The human blood was donated for scientific purposes.

## 2.1.6 Oligosaccharides

Table 6: An overview of the oligosaccharides used in this work.

| Oligosaccharide | Manufacturer |
| --- | --- |
| 3'-Sialyl Lewis X pentasaccharide | Biosynth |
| Blood group H antigen triaose type 2 | Elicityl |
| Forssman pentaose | Biosynth |
| Forssman triaose | Elicityl |
| GD3 oligosaccharide | Sigma-Aldrich |
| Globo-*N*-tetraose | Biosynth |
| Globotriose | Biosynth |
| Galactotetraose (Gal4) | Dextra UK |
| *N*-acetyllactosamine | Elicityl |
| Para-lacto-N-neohexaose | Biosynth |

## 2.1.7 Buffers

Table 7: An overview of the buffers used in this work.

| Buffer | Composition | |
| --- | --- | --- |
| $A_{his}$ | Tris pH 7.5 | 50 mM |
| | NaCl | 250 mM |
| | Imidazole | 20 mM |
| | Glycerol | 5% (w/v) |
| | | |
| $B_{his}$ | Tris pH 7.5 | 50 mM |
| | NaCl | 250 mM |
| | Imidazole | 500 mM |
| | Glycerol | 5% (w/v) |
| | | |
| Dialysis buffer | Tris pH 7.5 | 20 mM |
| | NaCl | 150 mM |
| | Glycerol | 5% |

Table 7: An overview of the biochemicals used in this work (cont.).

| Buffer | Composition | |
| --- | --- | --- |
| Lysis buffer | $A_{his}$ | supplemented with |
| | Urea | 500 mM |
| | Triton-X100 | 1% |
| | cOmplete™ protease inhibitor | 1 tablet (*optional*) |
| | | |
| SEC buffer | HEPES pH 7.5 | 20 mM |
| | NaCl | 150 mM |
| | | |
| Acetate buffer pH 4 (10 mM) | Sodium acetate | 2.27 mM |
| | Acetic acid | 7.73 mM |
| | | |
| Borate buffer pH 8.5 (400 mM) | Sodium tetraborate ($Na_2B_4O_7$) | 96 mM |
| | Boric Acid ($H_3BO_3$) | 304 mM |
| | | |
| SPR buffer | HEPES pH 7.4 | 20 mM |
| | NaCl | 150 mM |
| | Polysorbate 20 | 0.005% |
| | | |
| SDS-PAGE sample buffer (4x) | Tris-HCl pH 6.8 | 200 mM |
| | Glycerol | 40% (v/v) |

Buffers used in FPLC chromatography and surface plasmon resonance (SPR) analyses were sterile filtered and subsequently degassed.

## 2.1.8 Sequences

DNA sequences are displayed from 3'- to 5'-ends with the restriction sites printed in red and the stop codon in blue. The start codon is highlighted by underlining. Translated protein sequences are displayed from N- to C-termini with the protease sites printed in bold green (TEV) or bold purple (Thrombin) and C->S mutations highlighted yellow. Residue numbers refer to the native sequences omitting the initial methionine.

### 2.1.8.1 Lyon IARC polyomavirus VP1

The gene of Lyon IARC polyomavirus (LIPyV) VP1 was codon-optimized for expression in *Escherichia coli* (*E. coli*) and cloned into the pET-15b vector (Novagen) using the restriction sites of NcoI and BamHI as ordered from General Biosystems, Inc. Following a His6-tag and TEV protease site, the synthetic gene encodes the canonical amino acids 43–325 of the LIPyV VP1 with a C121S mutation.

**DNA**   CCATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCGAGAATCTTTATTTTCAG GGCAGCCATATGGGCGGCATTGAAGTGCTGGAAACCGTGAGCGGCCCGGATAGCATTACCCAG ATTGAACTGTTTCTGCAGCCGCGCATGGGCGTGAACACCTATGATATTGAACGCTATAGCAACT GGTATGGCTATAGCTATGAATTTCATCCGAAAGATGTGAACGATGTGCCGAGCCCGGAAACCCT GCCGAGCTATGCGGCGGCGCGCGTGCCGCTGCCGCCGCTGAACGAAGATATGACCAGCAACAT GATTATGATGTGGGAAGCGGTGAGCGTGAAAACCGAAGTGATTGGCATTAACACCCTGATTAA CGTGCATAAAGAAAAACAGCTGGAAATTCCGCAGTATGGCGAACATGCGGCGGGCATTCCGAT TCAGGGCCCGAACTATCATTTCTTTAGCATTGGCGGCGAACCGCTGGATCTGCAGGGCATTGTG GCGGATTATCGCACCATGTATCCGGAAGGCGAAAACAAACCGGTGAACATTAAAACCGTGACC AAAACCCCGATGACCCCGAAAAACCAGGGCATGGACCCGACCGCGAAAACCAAACTGCTGAAA GATGGCTATTATCCGATTGAAATTTGGTGCCCGGACCCGAGCCGCAACGAAAACACCCGCTATT TTGGCAGCTTTACCGGCGGCACCGATACCCCGCCGGTGCTGCAGTTTACCAACACCCTGACCAC CGTGCTGCTGGATGAAAACGGCATTGGCCCGCTGTGCAAAGCGGATGGCCTGTTTGTGGCGGCG GCGGATATTTGCGGCTTCTTTCATCAGGCGAGCGGCATGCGCTATCGCGGCCTGCCGCGCTATT TTAACGTGACCCTGCGCAAACGCTGGGTGAAAAACTAGGGATCC

**Protein**   MGSSHHHHHHSSG**ENLYFQG**SHMGGIEVLETVSGPDSITQIELFLQPRMGVNTYDIERYSN WYGYSYEFHPKDVNDVPSPETLPSYAAARVPLPPLNEDMT<mark>S</mark>NMIMMWEAVSVKTEVIGINTLINVHKEK QLEIPQYGEHAAGIPIQGPNYHFFSIGGEPLDLQGIVADYRTMYPEGENKPVNIKTVTKTPMTPKNQGMD PTAKTKLLKDGYYPIEIWCPDPSRNENTRYFGSFTGGTDTPPVLQFTNTLTTVLLDENGIGPLCKADGLFV AAADICGFFHQASGMRYRGLPRYFNVTLRKRWVKN

### 2.1.8.2 Sea otter polyomavirus VP1

The gene of sea otter polyomavirus (SOPyV) VP1 was codon-optimized for expression in *E. coli* and cloned into the pET-15b vector (Novagen) using the restriction sites of NcoI and BamHI as ordered from General Biosystems, Inc. Following a His6-tag and Tobacco Etch Virus (TEV) protease site, the gene encodes for the canonical amino acids 19–288 of SOPyV VP1 with a C95S mutation.

**DNA**  CCATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCGAGAATCTTTATTTTCAG GGCAGCCATATGGGCGGCATTGAAGTGCTGGAAGTGCGCAGCGGCCCGGAAAGCACCGTGGAA ATTGAACTGTATCTGAACCCGCACATGGGCGCGCCGGAAGATAGCGGCCATCTGTATGGCTTTA GCAACCAGGTGACCCTGAAAGAAGATTTTGGCGATGATACCCCGAACAAAGAAGAACTGCCGA AATATAGCTGCGCGCGCGTGCCGCTGCCGATGCTGAACGAAGATCTGACCAGCGAGAAGCTGC TGATGTGGGAAGCGTATACCTGCAAAACCGAAGTGCTGGGCATTAACATTCTGTGCGATGTGC ATAGCGGCGGCCAGCGCGTGGGCACCAGCGGCATTGGCAACCCGGTGACCGGCATTAACTTTC ACATGTTTGCGGTGGGCGGCGAACCGCTGGATCTTCAGGGCATTGCGTTTAACTATCGCACCCG CTATCCGAGCGGCACCATTGGCCCGAGCGGCATGACCCAGAAAAGCCAGGGCCTGGACCCGAA ACATAAAGCGAAACTGGATAAAGATAACGCGTATCCGGTGGAATGCTGGATTCCGGACCCGAG CCGCAACGAAAACACCCGCTATTTTGGCAGCTATACCGGCGGCGATAGCACCCCGCCGGTGCTT CAGTTTACCAACACCGTGAGCACCGTGCTGCTGGATGAACATGGCGTGGGCCCGCTGTGCAAA GGCGATGGCCTGTATCTGGCGTGCGCGGATATTGTGGGCTTTTATCGCAACAAAAGCAGCACCA TGCAGTTTCGCGGCCTGCCGCGCTATTTTAAAGTGGGCCTGCGCAAACGCTGGGTGAAGAACTA GGGATCC

**Protein**  MGSSHHHHHHSSG**ENLYFQG**SHMGGIEVLEVRSGPESTVEIELYLNPHMGAPEDSGHLYG FSNQVTLKEDFGDDTPNKEELPKYSCARVPLPMLNEDLT**S**EKLLMWEAYTCKTEVLGINILCDVHSGGQ RVGTSGIGNPVTGINFHMFAVGGEPLDLQGIAFNYRTRYPSGTIGPSGMTQKSQGLDPKHKAKLDKDNA YPVECWIPDPSRNENTRYFGSYTGGDSTPPVLQFTNTVSTVLLDEHGVGPLCKGDGLYLACADIVGFYR NKSSTMQFRGLPRYFKVGLRKRWVKN

### 2.1.8.3 Avian polyomavirus VP1

The synthetic gene of avian polyomavirus (APyV) VP1 was codon-optimized for expression in *E. coli* and cloned into the pET-15b vector (Novagen) using the restriction sites of NcoI and BamHI as ordered from General Biosystems, Inc. Following a His6-tag and TEV protease site, the gene encodes for the canonical amino acids 21–287 of APyV VP1 with a C91S mutation.

**DNA** CCATGGGCAGCAGTCATCATCATCATCACCATAGCAGCGGTGAAAATCTGTATTTTCAG
GGCAGCCATATGGGTGGTATTGAAGTTCTGGATGTGAAAAGTGGCCCGGATAGCATTACCACCA
TTGAAGCCTATCTGCAGCCGCGTCCGGGTCAGAAAAATGGCTATAGCACCGTTATTACCGTTCA
GGCAGAAGGCTATCAGGATGCCCCGCATAGTACCGAAGTTCCGTGTTATAGCTGTGCCCGTATT
CCGCTGCCGACCATTAATGATGATATTACCAGCCCGACCCTGCTGATGTGGGAAGCAGTGAGTG
TGAAAACCGAAGTTGTGGGCGTGAGCAGCATTCTGAATATGCATAGCGGTGCATTTCGCGCCTT
TAATGGCTATGGTGGTGGTTTTACCATTTGTGGTCCGCGTATTCATTTCTTTAGTGTGGGTGGCG
AACCGCTGGATCTGCAGGCCTGCATGCAGAATAGTAAAACCGTTTATCCGGCACCGCTGATTGG
TCCGGGTGAAGGTGAACGCCGTGAAACCGCACAGGTGCTGGATACCGGTTATAAAGCCCGTCTG
GATAAAGATGGTCTGTATCCGATTGAATGTTGGTGCCCGGACCCTGCCAAAAATGAAAATACCC
GTTATTATGGTAACCTGACCGGCGGTCCGGAAACCCCGCCGGTTCTGGCTTTTACCAATACCACC
ACCACCATTCTGCTGGATGAAAATGGCGTGGGTCCGCTGTGCAAAGGCGATGGCCTGTTTCTGA
GCGCCGCCGATGTTGCCGGCACCTATGTGGATCAGCGCGGTCGCCAGTATTGGCGTGGTCTGCC
GCGTTATTTTAGCATTCAGCTGCGCAAACGTAATGTTCGCAATTAAGGATCC

**Protein** MGSSHHHHHHSSG**ENLYFQG**SHMGGIEVLDVKSGPDSITTIEAYLQPRPGQKNGYSTVITV
QAEGYQDAPHSTEVPCYSCARIPLPTINDDITSPTLLMWEAVSVKTEVVGVSSILNMHSGAFRAFNGYG
GGFTICGPRIHFFSVGGEPLDLQACMQNSKTVYPAPLIGPGEGERRETAQVLDTGYKARLDKDGLYPIEC
WCPDPAKNENTRYYGNLTGGPETPPVLAFTNTTTTILLDENGVGPLCKGDGLFLSAADVAGTYVDQRG
RQYWRGLPRYFSIQLRKRNVRN

### 2.1.8.4 Chimpanzee polyomavirus VP1

The gene of chimpanzee polyomavirus (ChPyV) VP1 was cloned into the pET-15b vector (Novagen) using the restriction sites of NdeI and BamHI and was already available at the beginning of this work. Following a His6-tag and thrombin protease site, the gene encodes for the canonical amino acids 36–323 of ChPyV VP1 with a C114S mutation.

**DNA**   CCATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCGAGAATCTTTATTTTCAG
GGCAGC CATATG GGCGGCATTGAAGTGCTGGAAGTGCGCAGCGGCCCGGAAAGCACCGTGGAA
ATTGAACTGTATCTGAACCCGCACATGGGCGCGCCGGAAGATAGCGGCCATCTGTATGGCTTTA
GCAACCAGGTGACCCTGAAAGAAGATTTTGGCGATGATACCCCGAACAAAGAAGAACTGCCGA
AATATAGCTGCGCGCGCGTGCCGCTGCCGATGCTGAACGAAGATCTGACCAGCGAGAAGCTGCT
GATGTGGGAAGCGTATACCTGCAAAACCGAAGTGCTGGGCATTAACATTCTGTGCGATGTGCAT
AGCGGCGGCCAGCGCGTGGGCACCAGCGGCATTGGCAACCCGGTGACCGGCATTAACTTTCACA
TGTTTGCGGTGGGCGGCGAACCGCTGGATCTTCAGGGCATTGCGTTTAACTATCGCACCCGCTAT
CCGAGCGGCACCATTGGCCCGAGCGGCATGACCCAGAAAAGCCAGGGCCTGGACCCGAAACAT
AAAGCGAAACTGGATAAAGATAACGCGTATCCGGTGGAATGCTGGATTCCGGACCCGAGCCGC
AACGAAAACACCCGCTATTTTGGCAGCTATACCGGCGGCGATAGCACCCCGCCGGTGCTTCAGT
TTACCAACACCGTGAGCACCGTGCTGCTGGATGAACATGGCGTGGGCCCGCTGTGCAAAGGCGA
TGGCCTGTATCTGGCGTGCGCGGATATTGTGGGCTTTTATCGCAACAAAAGCAGCACCATGCAG
TTTCGCGGCCTGCCGCGCTATTT TAA GATCC[3]

**Protein**   MGSSHHHHHHSSG**LVPRGS**HMGGVEVLNIITGPDSTTEIELYLEPRMGINSPTGDKKEWYG
YSEVIHHADGYDNNLLSIQMPQYSCARVQLPMLNTDMT==S==DTLMMWEAVSCKTEIVGIGSLISVHLLEAKM
AAKEGGDGPSQPIEGMNYHMFAVGGEPLDLQGIESNALTKYASAIPPKTIHPNDIAKLAEEEKPQLQGLV
PKAKARLDKDGFYPIEEWSPDPSRNENSRYFGSFVGGLNTPPNLQFTNAVTTVLLDENGVGPLCKGDG
LFVSAADICGVMVKADNEAIRYRGLPRYFKVTLRKRAVKN

---

[3]The initial G nucleoside of the BamHI site was lost during cloning. Yet, this does not affect protein synthesis as it succeeds the stop codon.

### 2.1.8.5 Sheep polyomavirus VP1

The synthetic gene of sheep polyomavirus (ShPyV) VP1, codon-optimized for expression in *E. coli* and cloned between the restrictions sites of NdeI and BamHI into the pET-15b vector (Novagen), was manufactured by GenScript and already available at the beginning of this work. Following a His6-tag and a thrombin protease site, the gene encodes for the canonical amino acids 20–291 of ShPyV VP1 with a C95S mutation highlighted in yellow.

**DNA**  CCATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGC
CATATGGGCGGCATCGAAGTTCTGGCGGTGCGTACCGGCCCGGACAGCATTACCGAGATTGAAG
CGTACCTGAACCCGCGTATGGGTCAACCGCAGAACGAGGACTTCTACGGCTTTAGCGATAACGT
GACCGTTAGCGACGATTTCGGTAGCGACGCTCCGCCGTGGAAGCAGTTTCCGTGCTATAGCACC
GCGCGTATCAGCCTGCCGATGCTGAACCAAGACATGACCAGCGATACCATCCTGATGTGGGAGG
CGATTAGCTGCCGTACCGAAGTGATGGGCGTTAACATGCTGACCAACGTGCACAGCGCGCAAAA
ACGTGTTTACGAGAACGATCGTGAAGGTACCGGCATCGGTGTTGAGGGCATGGGTTATCACATG
TTCGCGATTGGTGGCGAGCCGCTGGAACTGCAGTTCATGGTGTTTAACCATCGTGCGACCTACCC
GGCGGAAGCGACCGTGATTAAGAACCCGGGTGCGAGCAGCCAAGTTTTCGACCCGAACCTGAA
AGGCACCCTGACCGCGGATGGTGTGTTTCCGGTTGAGGCGTGGGGTCCGGACCCGTTCAAGAAC
GAAAACACCCGTTACTTTGGCCAGTATACCGGTGGCACCCAAACCCCGCCGGTGCTGACCTTCA
CCAACACCCAGACCACCATCCTGCTGGATGAAAACGGCGTTGGTCCGCTGTGCAAAGGCGACGG
TCTGTTTCTGAGCTGCGCGGATATTGTGGGTTTCTTTACCCAACACAACAAGAAAATGAGCTTCC
GTGGTCTGCCGCGCTACTTCCGTGTGACCCTGCGTAAACGTGTTGTGAAAAATTAAGGATCC

**Protein**  MGSSHHHHHHSSG**LVPRGS**HMGGIEVLAVRTGPDSITEIEAYLNPRMGQPQNEDFYGFSD
NVTVSDDFGSDAPPWKQFPCYSTARISLPMLNQDMT<mark>S</mark>DTILMWEAISCRTEVMGVNMLTNVHSAQKRV
YENDREGTGIGVEGMGYHMFAIGGEPLELQFMVFNHRATYPAEATVIKNPGASSQVFDPNLKGTLTADG
VFPVEAWGPDPFKNENTRYFGQYTGGTQTPPVLTFTNTQTTILLDENGVGPLCKGDGLFLSCADIVGFF
TQHNKKMSFRGLPRYFRVTLRKRVVKN

# 2.2 Methods

## 2.2.1 Computational biology

### 2.2.1.1 Multiple sequence alignments

To calculate multiple sequences alignments (MSAs), sequences of nucleic acids or proteins were gathered from online data bank resources, e.g., GenBank or RefSeq [160, 161], and subjected to the MUSCLE web server with default parameters [162]. In case these MSAs were used to calculate phylogenetic trees, hyper-variable sequence regions within the alignments were discarded using Gblocks [163] implemented in SeaView 5 [157].

### 2.2.1.2 Phylogenetic trees

Unrooted maximum-likelihood trees to display phylogenetic relationships between species were inferred by subjecting the prepared MSAs to the IQ-TREE 2 software [150]. Within IQ-TREE 2, the best-fitting model was selected according to the Bayesian Information Criterion by the ModelFinder algorithm [164]. Branch support values were approximated by deploying 1000 iterations of the ultrafast bootstrap method UFBoot2 [165]. For visualization, iTOL version 5 was used [75].

### 2.2.1.3 Structure prediction

To predict the 3D structure of avian polyomavirus (APyV), AlphaFold2 implemented in Google CoLab was used [166, 167]. As the query, the monomeric protein sequence of APyV VP1 was used from residue 21 [GGIEVL...] to residue 287 [...KRNVRN], corresponding to the canonical sequence in the protein expression construct (see Sequence 2.1.8.3). The top-ranked predicted structure was relaxed using the Amber force field [168, 169]. For the prediction, no template information was used. For the MSA and pairing mode, the options "mmseqs2_uniref_env" and "unpaired_paired" were chosen, respectively. In the advanced settings, all options were left default using a single seed to sample predictions. As such, the APyV VP1 structure prediction was conducted three times.

## 2.2.2 Molecular biology

### 2.2.2.1 Mutagenesis

Site-directed mutagenesis was performed using the "quick change" polymerase chain reaction (PCR) protocol (see Table 8). Therein, parental template DNA and a complementary primer pair containing the desired mutation were used to produce linearized DNA. After completion, the (methylated) template DNA was digested by adding 1 µL of DpnI (incubation at 37 °C for an hour). The linear PCR product could directly be transformed into *E. coli* strains such as DH5α or XL-1 Blue.

Table 8: The site-directed mutagenesis protocol.

|  | Temperature (°C) | Duration (s) | No. of cycles |
|---|---|---|---|
| Initial denaturation | 94 | 120 | 1 |
| Denaturation | 94 | 50 | |
| Annealing | 58–67 | 60 | 18 |
| Extention | 72 | 390 | |
| Final extension | 72 | 600 | 1 |
| Cooling | 4 | *indefinite* | N/A |

### 2.2.2.2 Transformation

For the transformation of PCR product (e.g., from site-directed mutagenesis) or isolated plasmids, 5 µL and ~100 ng of DNA, respectively, were added to freshly thawed aliquots of competent *E. coli*, followed by incubation on ice for 30 min. After a temperature shift to 42 °C for not more than a minute ('heat shock'), the bacteria were incubated on ice for another 5 minutes. Subsequently, 950 µL of room temperature Lysogeny broth (LB) medium was added. To establish the vector-encoded antibiotic resistance, the transformation sample was incubated at 37 °C for an hour. Lastly, the transformed bacteria (150–200 µL) were plated on LB-agar plates containing the appropriate antibiotic (here, 100 µg/mL ampicillin or 50 µg/mL kanamycin). In the case of successful transformation, colonies of *E. coli* grew at 37 °C overnight.

### 2.2.2.3 Plasmid isolation

To isolate plasmids, single colonies of transformed *E. coli* were cultivated in 5 mL of liquid LB-antibiotic medium at 37 °C overnight. Then, bacterial cells were separated from

the supernatant by centrifugal sedimentation (5000 rpm for 10 minutes). Subsequently, the plasmid isolation was performed following the manufacturer's protocol (Promega Wizard® Plus SV or Invitrogen PureLink™). In the last step, the plasmid DNA was recovered from the spin column by elution with 50 µL of nuclease-free water.

### 2.2.3 Protein preparation

#### 2.2.3.1 Expression

Overnight cultures of 10–15 mL LB containing the appropriate antibiotic were inoculated with single-picked colonies of transformed BL21 (DE3) *E. coli* and incubated overnight at 37 ℃ and 120 rpm. For expression, overnight cultures were diluted 1:100 in the same medium and incubated at 37 ℃ and 95 rpm until an optical density at 600 nm ($OD_{600}$) of 0.7–1 was reached (SmartSpec plus, BioRad). Subsequently, the temperature was shifted to 20 ℃ before the induction of protein expression by the addition of 0.4 mM IPTG. The expression was incubated overnight until extraction.

#### 2.2.3.2 Extraction

After incubation, the expression culture was centrifuged at 7000 rpm and 4 ℃ for 20 minutes (Sorvall RC 6 Plus with an SLC4000 rotor) to sediment the bacterial cells. Following a thorough resuspension in lysis buffer (see Table 7), the protein was released from the bacteria by sonication (Digital Sonifier, Branson) under cooling in an ice bath. The soluble fraction containing the target protein was recovered in the supernatant by a second centrifugation run at 17,000 rpm and 4 ℃ for 60 minutes (SS34 rotor).

#### 2.2.3.3 IMAC

Immobilized metal ion affinity chromatography (IMAC) purification is based on the chemical interactions between metal ions embedded into the column resin material, here $Nickel^{2+}$-NTA, and the hexahistidine-tag of the protein. In this work, the cleared protein extracts were loaded onto 5 mL HisTrap FF crude columns (Cytiva), applying a flow of 1–2 mL/min via Äkta prime chromatography systems (Cytiva). Undesired protein was removed from the column by a continuous flow (1–2 mL/min) of Buffer $A_{his}$ and a subsequent wash with 10–25 % of Buffer $B_{his}$ until the absorbance reached baseline. Elution of the target protein was carried out using 60 % $B_{his}$ (300 mM imidazole) at 1 mL/min while collecting fractions of 2 mL. Columns were regenerated by purging with 100 % $B_{his}$ and $ddH_2O$. Fractions of interest were analyzed via SDS-PAGE (see 2.2.3.7).

### 2.2.3.4 SEC

The size exclusion chromatography (SEC), at times also referred to as gel filtration, is a method to separate molecules by their hydrodynamic radii. The accessible portion of the total column volume (CV) decreases with molecule size, and large proteins elute earlier than smaller ones. For analytical SEC, Superdex 200 Increase 3.2/300 columns (Cytiva) were used in an Äkta Ettan system (Cytiva) with SEC buffer (see Table 7) applying a flow of 0.55 mL/min. For preparative SEC, a Superdex 200 Hiload 16/60 column (Cytiva) was used in an Äkta Basic system (Cytiva), using the same buffers and a flow of 1 mL/min. Fractions were collected between 40–120 mL, and the eluted protein was analyzed via SDS-PAGE (see section 2.2.3.7).

### 2.2.3.5 Site-directed protein digest

To remove the N-terminal affinity tags from the recombinantly expressed VP1 proteins, thrombin or Tobacco Etch Virus (TEV) protease was used. Thrombin was used with 10 U/mg of protein in concentrations of 1 mg/mL or higher and with 30 U/mg for more dilute protein solutions. TEV protease was used with 1–3 mg per digest (comprising 50–150 mg of VP1). Incubation was carried out at room temperature in the dark on an orbital shaker. Samples of the protein digest were taken at various times for analysis via SDS-PAGE (see section 2.2.3.7).

### 2.2.3.6 Determination of protein concentration

To measure protein concentrations, a Nanodrop ND-1000 spectrophotometer (Thermo Scientific) was used. After blanking the device with the respective buffer solution, the concentration of a protein was determined using its extinction coefficient and molecular mass. The applied sample volume was 2 µL.

### 2.2.3.7 SDS-PAGE

For analysis via sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), protein samples were diluted 1:3 in SDS-PAGE sample buffer (see Table 7) and denatured at 95 °C for 5–10 min. In cases of substantial precipitation, the samples were centrifuged for 1 min at 16,100 g. After loading into the gels, the samples were separated in the running buffer (1x Rotiphorese®, Sigma-Aldrich) by electrophoresis (180V for 60 minutes). Lastly, the gels were stained in InstantBlue® (Abcam) or ReadyBlue™ (Sigma-Aldrich) for at least 15 minutes.

Table 9: Composition of the SDS-PAGE gels.

| Compound | Separating gel (12% V/V) | Stacking gel (4% V/V) |
|---|---|---|
| $H_2O$ | 5.0 mL | 6.1 mL |
| 1.5 M Tris/HCl pH 8.8 | 3.75 mL | - |
| 1.5 M Tris/HCl pH 6.8 | - | 2.5 mL |
| 10% SDS | 150 µL | 100 µL |
| TEMED | 7.5 µL | 10 µL |
| Acrylamide/Bis solution | 6.0 mL | 1.3 mL |
| 10% APS | 150 µL | 100 µL |

APS was used from freshly thawed aliquots.

## 2.2.4 Crystallization

### 2.2.4.1 Robotic screens

To screen for suitable crystallization conditions, the purified VP1s were subjected to commercial crystal screens[4]. Initially, the VP1s were concentrated to 3–15 mg/mL and sterile filtered before pipetting by the Freedom Evo (Tecan) or Gryphon (ARI) robots. Using 96-well sitting drop plates (Intelli-Plate 96, ARI), the protein and mother liquor (ML) were mixed in a 1:1 ratio, applying 200 nL and 300 nL of each solution using the Tecan and ARI robots, respectively. After sealing the plates, the drops were equilibrated against a reservoir of 50 µL of pure ML at 4 °C or 20 °C. The crystallization was monitored over days to weeks by visual inspection of the drops.

### 2.2.4.2 Fine screening and crystal seeding

If positive results from the initial robotic screens occurred, the respective crystallization conditions were manually refined by a two-dimensional fine screening of the salt and precipitant concentrations. Where applicable, the pH was altered as a third dimension. In each case, the conditions were reached using concentrated stocks of the constituent solutions. In 24-well hanging drop plates (Greiner Bio-One), the protein and ML were manually mixed in varying ratios, resulting in drop volumes of 1–3 µL. To promote crystal growth, the drops were inoculated with different dilutions of a crystallization seed (50–200 nL), which was produced from the freeze-thawing of crystals harvested from older screens.

---

[4]A library of crystallization screens was purchased from Hampton Research, Molecular Dimensions, and Jena Biosciences (see Consumables).

## 2.2.5 X-ray crystallography

### 2.2.5.1 Data collection and reduction

The X-ray diffraction data of protein crystals presented in this work were collected at the SLS synchrotron at the Paul Scherrer Institute in Villigen, Switzerland. The frozen crystals were aligned with the photon beam and exposed to the radiation for 0.1 s and a rotation of 0.1° per image. A data set comprised a full rotation, resulting in 3600 diffraction images. In cases where the crystal space group was triclinic (P1), a second complete rotation data set was recorded from another angle and later merged with the first to ensure the completeness of measured reflections. To convert the X-ray diffraction images into a list of distinct reflexes with intensities and standard deviations ("data reduction"), *XDS* software was used [159]. Resolution limits were determined manually based on the data precision indicators signal-to-noise ratio ($I/\sigma I$), $R_{meas}$, and $CC_{1/2}$ [170–172]. Data quality was assessed as a function of the diffraction images, and images with subpar quality were excluded from the data sets. To convert the reflection-intensity data in the format required by structure determination software (see the following two paragraphs), *XDSCONV* was used [159]. Therein, *XDSCONV* also incorporated test reflection labels (~2000–3000 reflections) required for the calculation of the $R_{free}$ value [173].

### 2.2.5.2 Molecular replacement

To solve the X-ray crystal structure of a VP1 pentamer, a search model comprising the single pentameric structure of the closest structurally characterized relative was used for molecular replacement. In this work, the search models for the SOPyV, ChPyV, and LIPyV VP1s were derived from the PDB entries of BKPyV VP1 (4MJ1), NJPyV VP1 (6Y5X), and MCPyV VP1 (4FMG). For the X-ray structures of complexed ShPyV VP1, the search model was taken from the unliganded VP1 pentamer (6Y61). For molecular replacement, *Phaser* crystallographic software implemented in the CCP4 software suite was used [153, 174]. First, the content of the crystallographic unit cell was calculated using Matthews' method [175]. The calculated solvent content and model copy number were entered into the *Phaser* graphical user interface. During the initial placement of the search model, which translates and rotates the structure towards the correct position and orientation within the experimental unit cell, the calculation resolution was lowered, for instance, to 3.5–5 Å. After placement and assessment of the evaluation parameters, the top-scoring solution was extended to the full resolution of the data set, providing coordinates and structure factors for further model and phase refinement.

### 2.2.5.3 Structure factor refinement and model building

Following molecular replacement, the resulting coordinates were used together with the initial diffraction amplitudes from data reduction to refine structure factors (F) and calculate their real-space Fourier transform, the electron density map, in *Refmac5* [176]. During the refinement process in *Refmac5*, the atomic positions in the model are adjusted employing a maximum-likelihood method to align the calculated (calc) structure factors with the observed (obs) data [156]. After that, the model coordinates were adjusted using the *Coot* model building tool [149]. Subsequently, structure factor refinement and model building were performed alternately until the evaluation parameters, i.e., the R-factors ($R = \frac{\sum |F_{obs} - F_{calc}|}{\sum F_{obs}}$), converged. By excluding certain reflections from the refinement process (see section 2.2.5.1), two independent R-factors for the working set ($R_{work}$) and the test set ($R_{free}$) could be calculated.

## 2.2.6 Characterization of protein-ligand interactions

### 2.2.6.1 Glycan array screening

All glycan array screening experiments were performed by our collaborators at the Glycosciences Laboratory of the Imperial College, London. For the protocol, please refer to Rustmeier *et al.* [2]. For further reading, see Liu *et al.* [177] and Palma *et al.* [178].

### 2.2.6.2 Saturation transfer difference NMR

The saturation transfer difference (STD) nuclear magnetic resonance (NMR) method was used in Rustmeier *et al.* [2] attached to this work. For the exact protocols, please refer to the respective Methods section. Briefly summarized, STD NMR can unveil interactions between proteins and ligands in solution with dissociation constants between $10^{-3}$–$10^{-8}$ M [179], which is particularly feasible for protein-carbohydrate complexes. We performed our experiments using micromolar concentrations of VP1 and low millimolar concentrations of glycan ligands at room temperature. In the case of binding, the STD spectrum allows the mapping of the glycan epitope interacting with the protein surface.

### 2.2.6.3 Determination of the crystallographic dissociation constant

To approximate a thermodynamic dissociation constant (Kd) between the ShPyV major capsid protein and Forssman pentaose ($F_P$), VP1 crystals were grown as described in Rustmeier *et al.* [2] and derivatized with a dilution series of $F_P$ similar to the protocols described in Stehle & Harrison [62], Ströh *et al.* [93], and Buch *et al.* [90]. The reference

soaking solution (10 mM $F_P$ in 150 mM KSCN, 20 % (w/v) PEG 3350 and 20 % (v/v) MPD) was diluted in 1:2 steps to a final concentration of 0.0195 mM. For the derivatization at the different $F_P$ concentrations, the VP1 crystals were soaked in parallel for 30 minutes. A separate crystal without $F_P$ for determining the null value was prepared accordingly. Data collection was performed as described in Rustmeier *et al.* [2]. For processing, the unit cell constants of all data sets were treated as isomorphous. All intensities were scaled against the 10 mM $F_P$ reference data set using Scaleit [180]. A single molecular replacement structure solution was sufficient for all data sets. Atomic coordinates of VP1 lacking water and carbohydrate atoms and a control tripeptide (amino acids 143–145) in all ten individual chains were used for simulated annealing refinement using standard parameters in Phenix refine [155]. The resulting bias-reduced positive Fo-Fc electron densities were integrated within 1 Å around the terminal trisaccharide (GalNAcα1-3GalNAcβ1-3Galα) of the refined $F_P$ model and the control tripeptide of VP1, respectively, for all chains in each data set using the software *xdlMAPMAN* [152]. The electron density integrals were plotted against the ligand concentrations using R [181]. The data was interpreted using the R library drc's [182] Michaelis-Menten fit that conforms to the following equation:

$$e = e_{max} * \frac{c}{Kd_c + c} \tag{2.1}$$

Here, the integrated electron densities (e) are a function of the $F_P$ concentration c. The parameters $Kd_c$ and $e_{max}$ represent the crystallographic dissociation constant and the asymptote of the density integrals, respectively.

### 2.2.6.4 Carbohydrate-BSA coupling

For the kinetic characterization of protein-ligand interactions, carbohydrate chains were coupled to BSA to prepare a ligand to be immobilized onto a surface plasmon resonance (SPR) chip surface. To covalently link the reducing ends of oligosaccharides to the lysine sidechains of bovine serum albumine (BSA), reductive amination was performed. The reaction proceeded in batch buffered with alkaline borate (pH 8.5), where the primary amine of the lysine residue reacted with the open aldehyde of the oligosaccharide to form an intermediate hemiaminal, which converted to an imine via the (reversible) dissociation of a water molecule. This process was facilitated by 500 mM of $Na_2SO_4$, shifting the equilibrium towards the product. Subsequently, the double bond of the imine was reduced to an amine the reducing agent sodium cyanoborohydride ($NaBH_3CN$). This procedure was adapted from Gildersleeve *et al.* [183], who suggest the usage of

30 equivalents of oligosaccharide to protein. The incubation (96 hours at 54 °C) was carried out in a PCR thermal cycler with lid heating to prevent condensation. The exact batch composition is minuted in Table 10. After 96 h, the reaction was quenched on ice. Subsequently, the reaction mixture was applied to analytical SEC (see Methods, section 2.2.3.4) for purification of the carbohydrate-coupled BSA and buffer exchange to 20 mM HEPES, pH 7.4 + 20 mM NaCl.

Table 10: Composition of the reductive amination batch reaction.

| Compound | Amount (µL) |
|---|---:|
| $H_2O$ | 4.7 |
| $Na_2SO_4$ (2 M) | 5.6 |
| Sodium borate buffer pH 8.5 (400 mM) | 5.5 |
| BSA (100 mg/mL) | 2 |
| Forssman pentaose (100 mM) | 2 |
| $NaBH_3CN$ (3 M) | 2.2 |

Total volume = 22 µL

### 2.2.6.5 Surface plasmon resonance analysis

For the surface plasmon resonance (SPR) analysis protocol to determine the kinetic binding parameters between ShPyV VP1 and Forssman pentaose, please refer to the Methods in Rustmeier *et al.* [2].

### 2.2.6.6 Hemagglutination assay

Forssman-positive sheep blood and Forssman-negative bovine blood were diluted to reach 2% suspensions in PBS buffer. Human reference erythrocytes in 2% suspensions were purchased (see Materials). For the assay, arrays of 50 µL aliquots of each suspension were pipetted into a 96-well conical microtiter plate. To these, 5 µL volumes of different dilutions of ShPyV VP1 solution in PBS buffer were added. A sample mixed with 5 µL of pure PBS as a negative control was prepared for each blood type. Subsequently, the assay was incubated for 10 minutes at room temperature. The results were evaluated visually by comparing the samples with the negative controls.

# Chapter 3: Results

## 3.1 Sialyl receptor engagement of the *Polyomaviridae*

**Background**

In Ströh *et al.* [1], we structurally characterized the engagement of sialylated ligands by the human New Jersey Polyomavirus (NJPyV) and the animal viruses sheep polyomavirus (ShPyV), goose hemorrhagic polyomavirus (GHPyV), and finch polyomavirus (FiPyV). While the structure of NJPyV VP1 in complex with its sialylated ligand defined a new sialic acid binding mode, the remaining structures showed that polyomaviruses of different genera share a conserved site for sialyl ligands recognition, including the alphapolyomavirus *Trichodysplasia spinulosa*-associated polyomavirus (TSPyV), the betapolyomavirus ShPyV, and the bird gammapolyomaviruses. The presence of this binding site in distantly related PyV species indicates an ancient origin. Yet, the presented sialyl complex structures did not provide insights into the differing host tropisms of these viruses. In consequence, we decided to subject the respective major capsid proteins to glycan array screening, which constituted the basis for most of the findings described in this thesis.

### 3.1.1 VP1 purification for broad-spectrum glycan array analyses

To identify potential receptor candidates for novel polyomaviruses and to unravel the differences in glycan engagement between PyV species, several recombinant VP1s were purified for broad-spectrum glycan array screening as explained in the Methods, aiming for monodisperse protein in the analytical SEC and a single band in the SDS-PAGE gel. The purification results are shown in Figure 10. The glycan array findings will be presented in the viruses' respective Results sections on page 51 for ChPyV and NJPyV VP1, page 66 for LIPyV VP1, Figure 1 in the attachment Rustmeier *et al.* [2] for ShPyV VP1, and page 140 for SOPyV VP1 in the Appendices. For an overview of the array's glycan probes, please refer to Table A2 following page 140 in the Appendices.
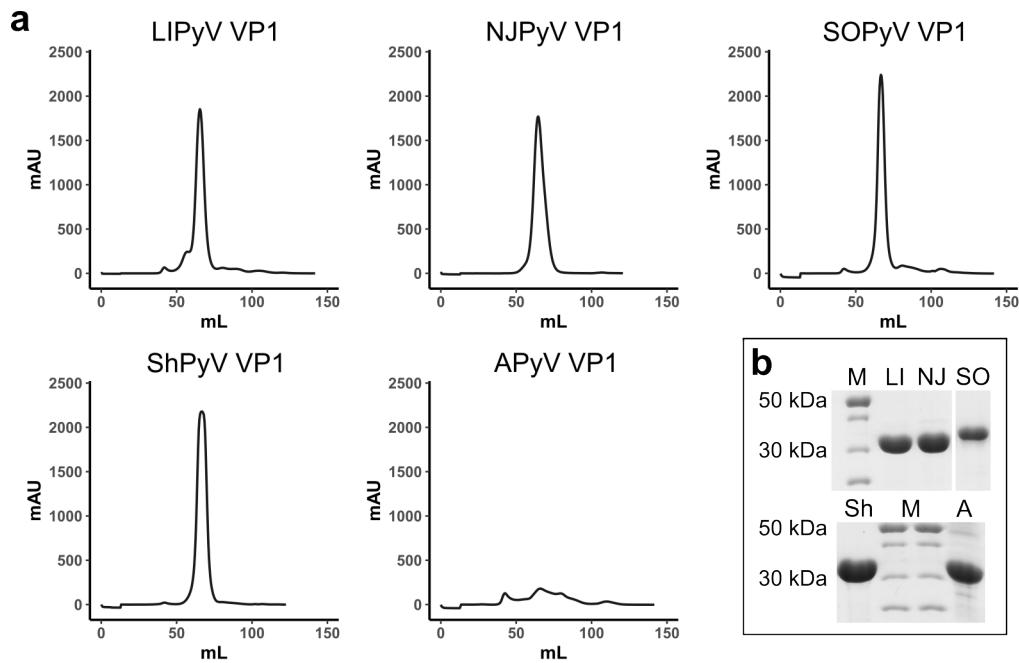
Figure 10: VP1 samples subjected to glycan array screening. **a)** Size exclusion chromatograms of the VP1s subjected to glycan array screening. All VP1s carry an N-terminal His6-tag. **b)** SDS-PAGE analyses of corresponding samples. Marker lanes (M) indicate molecular weights. *Note*: The VP1 of ChPyV analyzed by glycan array screening was purified by Dr. Luisa J. Ströh during her time as a doctoral candidate in our lab and is thus not shown here. The results of the SOPyV VP1 screening are shown under Figure A2 in the Appendices.

## 3.1.2 Sea otter polyomavirus VP1 interacts with the GD3 oligosaccharide

### 3.1.2.1 Background

Sea otter polyomavirus (SOPyV), initially discovered in deceased sea otters (see Introduction, section 1.2.5.4), by sequence analysis, shares a sialyl binding site with its relative betapolyomaviruses SV40 and BKPyV (see Figure 11a). Additionally, the VP1 of SOPyV also displays high conservation in the sialyl binding site commonly used by TSPyV, ShPyV, and GHPyV (see Figure 11b). To investigate if SOPyV sustains two distinct sialyl glycan sites in its VP1, a synthetic gene of SOPyV VP1, optimized for *E. coli* expression, encoding an N-terminal His6-tag followed by a TEV protease site and the capsid protein's native amino acids 19-288 was ordered (see Materials, section 2.1.8.2).

**a**



**b**
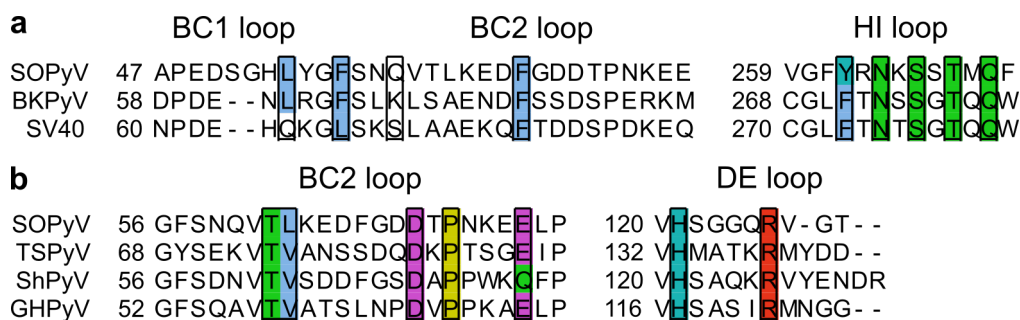


Figure 11: Multiple sequence alignments of the VP1s from SOPyV and other polyoma-viruses. Sequences are truncated to the surface loops responsible for sialyl glycan binding. SOPyV VP1 sequence in comparison with the binding sites found in **a)** BKPyV and SV40 and **b)** TSPyV, ShPyV, and GHPyV. Residue positions contributing to binding are highlighted with Clustal coloring.

### 3.1.2.2 Expression and purification

After transforming the SOPyV VP1 gene-containing pET-15b vector into *E. coli* BL21 (DE3), the protein was expressed at 20 °C overnight following the protocol described in the Methods. After extraction, most of the protein remained in the insoluble fraction (see Figure 12a). However, the following IMAC purification of the His6-tagged SOPyV VP1 yielded 90 mg of pure, soluble protein. The His6-VP1 was dialyzed into a buffer suitable for TEV protease digest (see Materials, Table 7). An initial TEV protease reaction (0.5 mg protease / 20 mL batch), performed at room temperature for 20 hours, did not result in the entire digestion of the protein (see Figure 12b). Following a reversed IMAC, the digest was repeated using 2 mg of fresh protease for 72 hours to complete the process. After another reversed IMAC to remove the His6-tagged protease and cleaved-off His6-tags, the VP1 was concentrated and subjected to SEC to prepare the protein for crystallization (see Figure 12c).

### 3.1.2.3 Crystallization

The purified VP1 of SOPyV was concentrated to 4 mg/mL for initial crystallization condition screening (see Methods, section 2.2.4). During the incubation at 4 °C, crystals with varying morphologies grew within days, including specimens with straight edges in the PEGIon I condition 14 (PEGIon I-14) and the Wizard IV condition 27 (Wiz IV-27, see Table 11). After six days, these crystals surpassed 100 µm in length (see Figure 13). Subsequently, the conditions were manually fine-screened using crystal seeding, which resulted in large and well-defined crystals. Specimens of the fine-screened Wiz IV-27 condition were used for the GD3 oligosaccharide soaking experiment.

Figure 12: SDS-PAGE analyses of the SOPyV VP1 purification. **a)** Control of protein expression (Expr. test) without (-) and with (+) addition of IPTG to the bacterial culture shows a His6-VP1 band of appropriate size (~32 kDa) in the positive sample. Extraction samples of total lysate (T) and pellet (P) show considerable amounts of undissolved protein. However, the supernatant (S) contains remaining soluble VP1. Samples of IMAC column loading show no VP1 band in the flow through (FT) and wash (W) fractions. The major band in these lanes corresponds to an *E. coli* protein that runs slightly above VP1 (compare - and + lanes). IMAC lanes show the elution of pure VP1 between fractions 12 and 20. Throughout the figure, the marker lanes (M) indicate molecular weights. Here, M bands are tagged to enhance visibility. **b)** Analysis of the His6-tag removal using 0.5 mg TEV protease. Digest samples correspond to 20 h incubation (D1), reversed IMAC of D1 (D2), and incubation for 24 h and 72 h (D3 and D4, respectively) after the addition of 2 mg of fresh protease. The size difference between the His6-VP1 and VP1 bands is ~2 kDa. The lanes on the right correspond to the reversed IMAC of the completely digested SOPyV VP1. **c)** Analysis of the SEC as the last step of VP1 purification. Lanes correspond to the load, peak fractions (samples 8-16), and the peak fraction pool. For the raw scans of the SDS-PAGE gels, please refer to Figure 42 in the Appendices.

Table 11: Selection of crystallization conditions for SOPyV VP1.

| Condition | Buffer/ salt solution | Precipitant |
| --- | --- | --- |
| PEGIon I-14 | 0.2 M Potassium thiocyanate | 20% w/v Polyethylene glycol 3350 |
| Wiz IV-27 | 0.1 M TRIS pH 8.5 + 0.2 M trimethylamine *N*-oxide (TMAO) | 20% w/v Polyethylene glycol monomethyl ether 2000 |

### 3.1.2.4 The overall structure of SOPyV VP1 and its interactions with GD3 oligosaccharide

The crystal structure of SOPyV VP1 in complex with 20 mM of GD3 oligosaccharide ligand (Neu5Acα2-8Neu5Acα2-3Galβ1-4Glc) was solved to 1.7 Å via molecular replace-

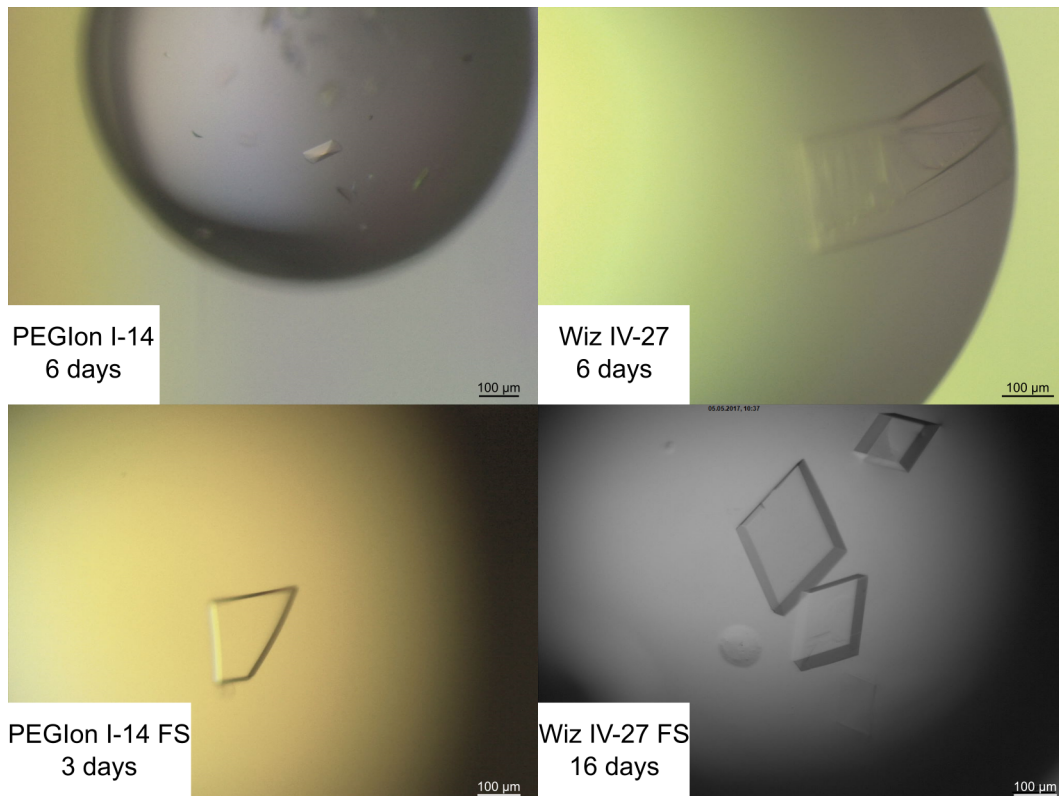Figure 13: The initial crystallization screens of SOPyV VP1 yielded large crystals in the PEGIon I-14 (upper left) and the Wiz IV-27 (upper right panel) conditions within six days. Well-defined crystals were obtained in the crystal-seeded fine screens (FS, lower panels). All crystallization was incubated at 4 °C.

ment (see Methods, section 2.2.5.2) using a single BKPyV VP1 pentamer (PDB 4MJ1) as the search model. The asymmetric unit (ASU) of the SOPyV VP1 crystal comprises four pentamers with protein loops mediating all inter-pentameric contacts.

The individual VP1 pentamers, one of which is displayed in Figure 14a, are structurally almost identical with C$\alpha$ root mean square deviations (RMSDs) between 0.07 Å and 0.14 Å. The resolved residues comprise the non-canonical "HM" dipeptide remaining at the N-terminus after proteolytic digest, followed by the canonical amino acids Gly 19 to Asn 288. The CD loop comprising residues 91-96 is not resolved in ten chains, probably due to structural disorder within the crystal. Structure refinement revealed that all 20 bindings sites are occupied by at least the terminal Neu5Ac. In a few chains, the terminal trisaccharide of GD3 (Neu5Ac$\alpha$2-8Neu5Ac$\alpha$2-3Gal$\beta$) is resolved without artificial crystal contacts, indicating a favored glycan conformation (see Figure 14b). The binding mode resembles the BKPyV VP1-GD3 oligosaccharide complex: the N-acetyl moiety's methyl group of the terminal Neu5Ac (Neu5Ac$\alpha$-8) binds to a hydrophobic depression in the surface of SOPyV VP1, which is mainly constituted by phenylalanine and tyrosine
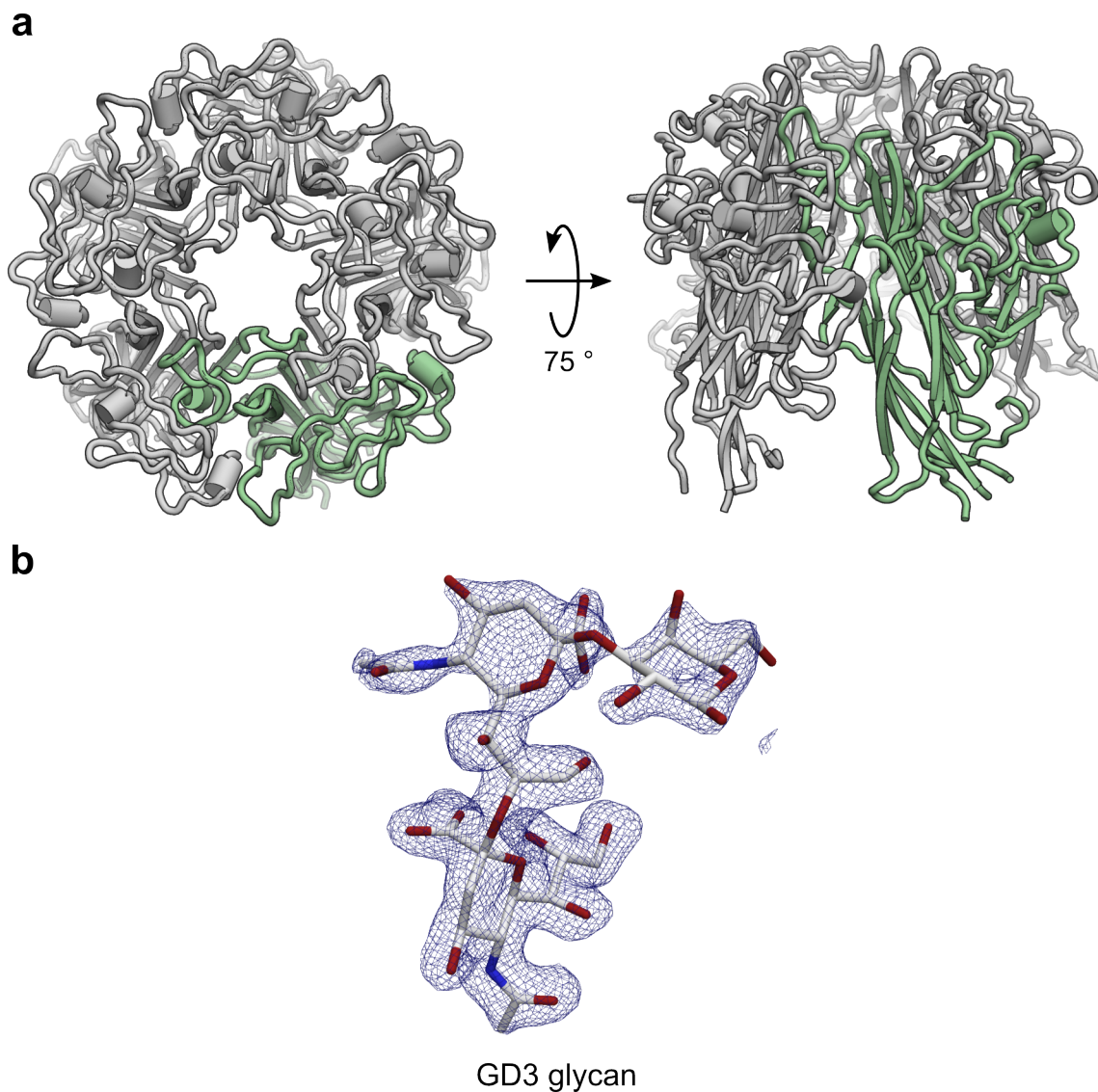
GD3 glycan

Figure 14: Results of the crystallographic SOPyV VP1 complex formation using GD3 oligosaccharide. **a)** The overall structure of the SOPyV VP1 pentamer from above (left) and tilted around the lateral axis (right). The protein is shown in cartoon representation with one monomer in pale green. Alpha helices are displayed as cylinders. **b)** The terminal trisaccharide of the GD3 oligosaccharide (Neu5Acα2-8Neu5Acα2-3Galβ1-4Glc) as resolved at 1.7 Å in the SOPyV VP1 complex structure. Ligand drawn in stick representation with the carbon, oxygen, and nitrogen atoms colored in white, red, and blue, respectively. The Fo-Fc difference map is shown as a blue mesh around the oligosaccharide with a contour level of σ = 2.5 and a radius of 1.6 Å.

Table 12: Data and refinement statistics for the SOPyV VP1 crystal structures.

| Data set | GD3 oligosaccharide complex |
|---|---|
| **Data processing** | |
| Wavelength (Å) | 1.0 |
| Resolution range (Å) | 48.3 - 1.70 (1.76 - 1.70) |
| Space group | P 1 |
| Unit cell | |
| a, b, c (Å) | 92.6, 121.3, 139.1 |
| $\alpha$, $\beta$, $\gamma$ (°) | 98.5, 90.1, 91.1 |
| Total reflections | 2,259,561 (227,449) |
| Unique reflections | 636,839 (62,212) |
| Multiplicity | 3.5 (3.7) |
| Completeness (%) | 96.8 (94.6) |
| Mean I/ $\sigma$(I) | 10.7 (1.3) |
| Wilson B-factor (Å$^2$) | 20.5 |
| $R_{meas}$ | 0.09 (0.99) |
| $CC_{1/2}$ | 1.00 (0.66) |
| | |
| **Refinement** | |
| $R_{work}$ / $R_{free}$ (%) | 18.7 / 21.3 |
| No. of atoms | 46,203 |
| Protein | 40,972 |
| Glycans | 932 |
| Solvent | 4,299 |
| RMSDs | |
| Bonds (Å) | 0.009 |
| Angles (°) | 1.61 |
| Ramachandran plot | |
| Favored (%) | 96.59 |
| Allowed (%) | 3.30 |
| Outliers (%) | 0.11 |
| Rotamer outliers (%) | 1.69 |
| Clashscore | 3.13 |
| Average B-factors (Å$^2$) | 28.18 |
| Protein | 27.26 |
| Glycans | 43.48 |
| Solvent | 33.62 |

Values in parentheses refer to the highest resolution bin.

residues of the BC1, HI, and BC2cw loops. The carboxylate of Neu5Ac$\alpha$2-8 interacts with Ser 265 (S265) and Thr 267 (T267) in the HI loop (see Figure 15a). The side chain of Gln 59 (Q59) directly interacts with the carboxylate group of the subterminal Neu5Ac and forms a water-mediated hydrogen bond with its amide nitrogen atom, aligning with K68 in BKPyV VP1 providing an electrostatic interaction (see Figure 15a & b).
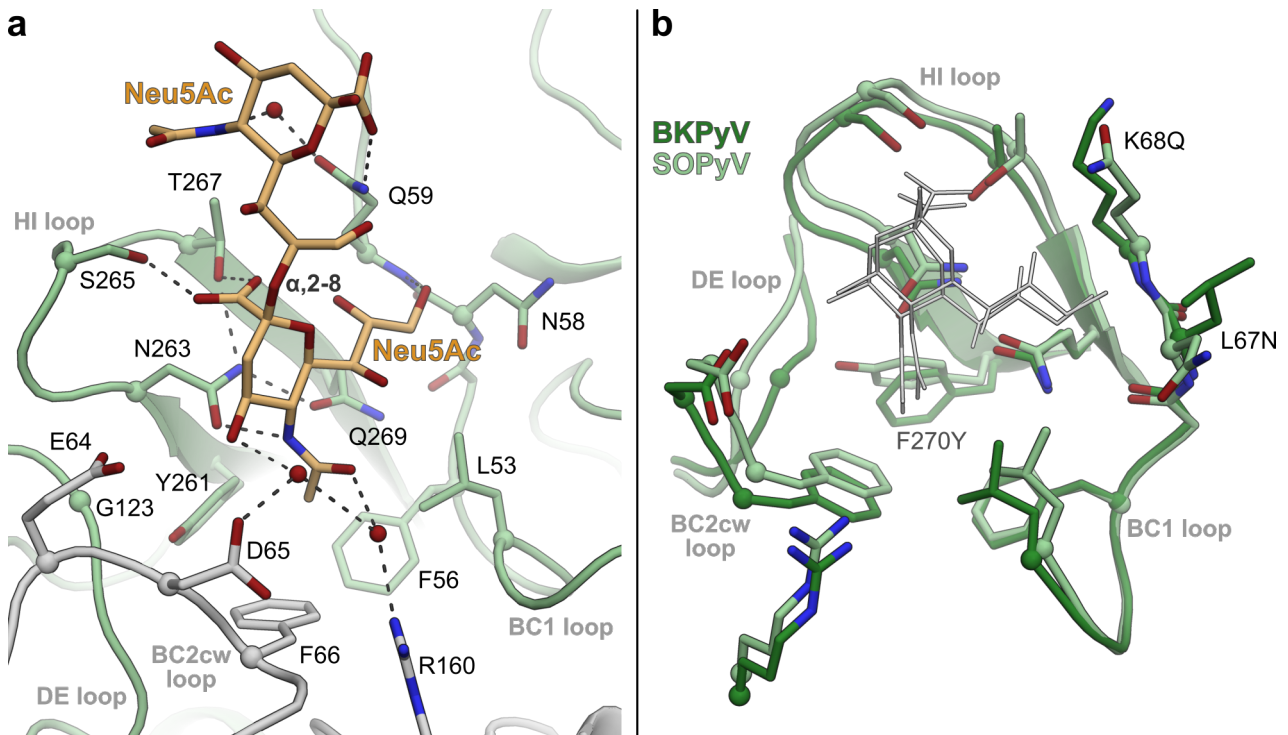
Figure 15: The binding site of SOPyV VP1 in complex with the GD3 oligosaccharide. **a)** The sialyl glycan site is constituted by two VP1 chains distinguished by gray and pale green coloring. The residues contributing to glycan binding and the terminal Neu5Acα2-8Neu5Acα of the GD3 oligosaccharide, with carbons colored in light orange, are shown as sticks. Polar contacts are indicated as dashed lines. **b)** Structural comparison of the sialyl sites in BKPyV and SOPyV. Amino acid substitutions are indicated from the perspective of BKPyV VP1 (PDB 4MJ0). The terminal Neu5Ac moieties in the two complex structures are outlined in white. All oxygen and nitrogen atoms are colored in red and blue, respectively.

### 3.1.2.5 The structure of the SOPyV VP1 BC2 loop in comparison with the Neu5Ac site of TSPyV

The BC2 loop of SOPyV VP1 is well resolved in all chains in the GD3 complex structure. An exception is the side chain of lysine 63 (K63), which projects outwards and lacks electron density (see Figure 16a), probably due to elevated flexibility. In the MSA focusing on the sialyl binding sites of TSPyV and ShPyV VP1s, K63 aligns with the small residues alanine 75 (A75) and serine 63 (S63, see Figure 16b), both of which are facing inwards and thus do not contribute to Neu5Ac binding (see Figure 16c). Contrary to the anticipations, no sialyl ligand binding was detected at this position in the SOPyV VP1-GD3 oligosaccharide complex structure. Instead, the side chain of K63 occupies the space inhabited by the Neu5Ac moiety in the complex structures of TSPyV VP1 and

ShPyV VP1. Hence, this residue may inhibit the engagement of Neu5Ac, although all the remaining binding site is conserved in SOPyV VP1.
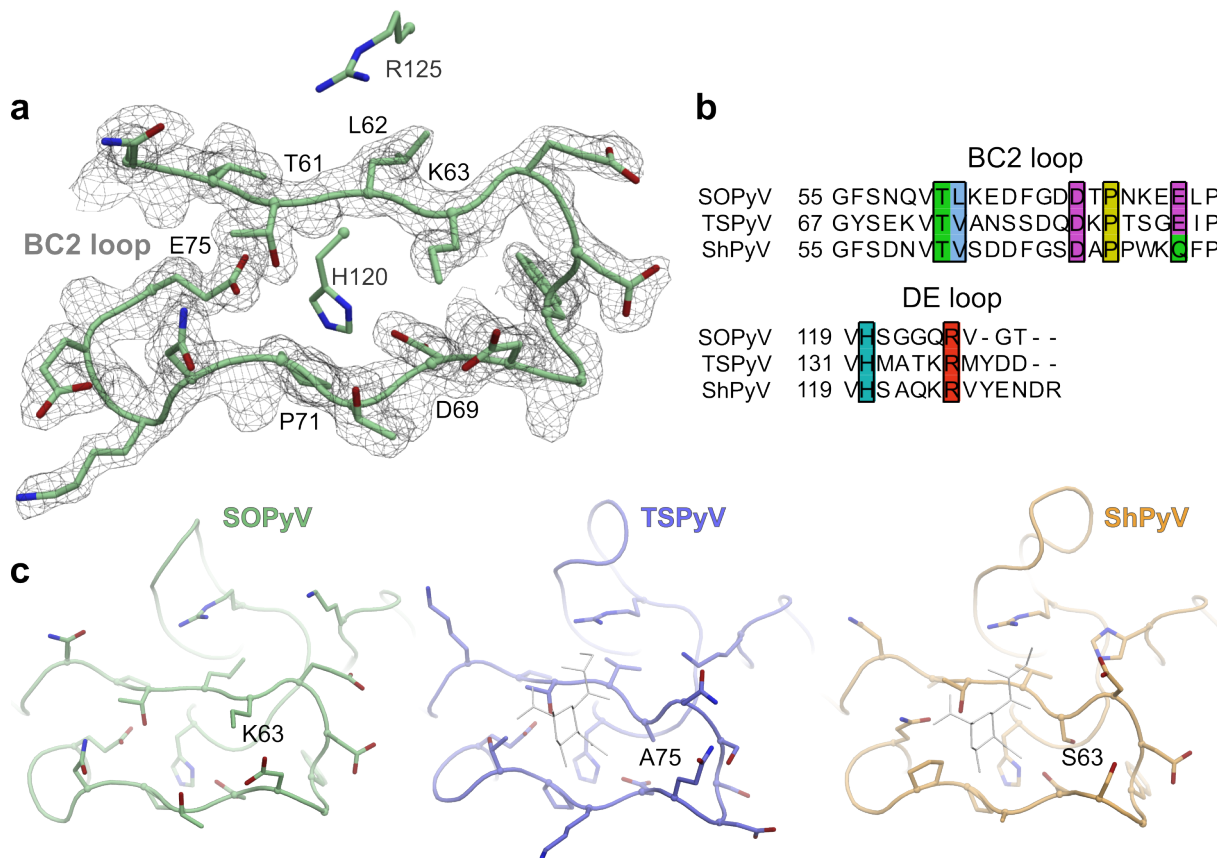


Figure 16: The BC2 loop of SOPyV VP1. **a)** The SOPyV VP1 BC2 loop structure and the 2Fo-FC electron density at a contour level of σ = 1 and a radius of 1 Å around the residues. The side chains of the DE loop residues histidine 120 (H120) and arginine 125 (R125) were added for completeness.
**b)** The MSA of the sialyl binding site in TSPyV. Residue positions contributing to glycan engagement are highlighted in boxes with Clustal coloring. **c)** The BC2 loop structure of SOPyV VP1 in juxtaposition with the sialyl sites in the glycan ligand complex of TSPyV VP1 (PDB 4U60) and ShPyV VP1 (PDB 6Y64) with the bound Neu5Ac moieties outlined. All oxygen and nitrogen atoms are colored red and blue, respectively.

### 3.1.3 Structural elucidation of the avian polyomavirus VP1

*The expression and purification of the avian polyomavirus, also designated budgerigar fledgling disease polyomavirus (BFDPyV), was mainly performed by the students Alexander Herrmann and Joshua C. Müller under my supervision. Since the results of that work were already minuted in the respective Bachelor theses archived in the Stehle lab, I will summarize these findings and further elaborate on their implications in the Discussions.*

#### 3.1.3.1 Background

As mentioned in the Introduction's section about APyV, the primary motivation to investigate the structure APyV VP1 was the comparison to SOPyV VP1. The focus was on the BC2 loop, which constitutes the binding site in the VP1s of TSPyV, ShPyV, and the bird polyomaviruses but didn't engage sialyl ligands in the SOPyV VP1 crystal structure. Furthermore, APyV's ability to infect hosts from several bird species is a clear distinction from most other PyVs and makes it an interesting study subject *per se.*

#### 3.1.3.2 Avian polyomavirus VP1 is not monodisperse in SEC buffer

The expression vector of APyV VP1 was designed similarly to those of the other polyomavirus VP1s (see Materials, section 2.1.8.3). However, the APyV VP1 remained mostly insoluble upon extraction and subsequent purification using the buffer systems for the other VP1s in this work. Although it was possible to obtain small amounts of protein with only minor impurities, APyV VP1 did not run as a single monodisperse species in the analytical SEC (see again Figure 10 on page 38).

#### 3.1.3.3 Crystallization

After fine-tuning the buffer conditions, we obtained monodisperse protein in a TRIS buffer with highly concentrated sodium chloride ($c_{NaCl}$ = 1 M). The APyV VP1 purified under these conditions was concentrated to 3 mg/mL and then subjected to initial robotic crystallization screens, which, however, only yielded salt crystals during the observation period. Hence, the crystal structure of APyV VP1 could not be determined to date.

#### 3.1.3.4 Structure prediction of avian polyomavirus VP1 using AlphaFold

Instead of experimentally solving the structure of the pentameric APyV VP1, its monomer was predicted using AlphaFold2 as described in the Methods. The monomeric VP1

was chosen as the target because the calculation of the pentamer required substantially more computational time and eventually resulted in the same VP1 tertiary structure. Also, using a monomeric model does not hinder predicting the putative sialyl binding site mainly constituted by the BC2 loop of a single VP1 chain. The predicted structure of the monomeric VP1 adopts the inherent beta-barrel or "jelly roll" fold (see Figure 17a), which in the pentameric assembly contributes to inter-chain contacts. The prediction of the jelly roll core domain is highly reliable with confidence measures (plDDT) above 90%. The structure was predicted three times, resulting in only insignificant C$\alpha$ RMSDs of 0.003-0.004 Å. Minor structural deviations only occur at the N-terminus and around the sequence positions 62 (tip of the BC2 loop) and 170 (within the EF loop), where the plDDT drops below 80% as indicated by yellow coloring in Figure 17a. Coincidentally, the number of sequences that align with the EF loop is lowest around residue 170 (position 150 in the query sequence, see Figure 17b).

### 3.1.3.5 Exposed hydrophobic residues may cause aggregation of recombinant APyV VP1

Besides confirming the jelly roll fold, the AlphaFold model shows that APyV VP1 bears four hydrophobic residues at an exposed location of the DE loop, namely phenylalanine 121 (F121), phenylalanine 124 (F124), tyrosine 127 (Y127), and phenylalanine 131 (F131, see Figure 17c). Their hydrophobic nature may cause problems during the folding process and decrease the protein's solubility at higher concentrations, providing a rationale behind the tendency of recombinant APyV VP1 to aggregate in solution.

### 3.1.3.6 The BC2 loop structure of APyV VP1 conforms to the sialyl glycan binding site of TSPyV VP1

Investigating the putative sialyl glycan binding site of APyV VP1 reveals that the BC2 loop residue glutamine 59 (Q59), which aligns with K63 in SOPyV VP1, folds towards the base of the BC2 loop (see Figure 18a). Although Q59 is larger than the aligning alanine 75 (A75) in TSPyV VP1 and serine 63 (S63) in ShPyV VP1 (see Figure 18b), the structural prediction resulted in congruent C$\beta$ coordinates for these residues (see Figure 18c). Consequently, Q59 would not overlap with Neu5Ac in the established TSPyV VP1 sialyl binding mode, contrasting K63 of SOPyV VP1.
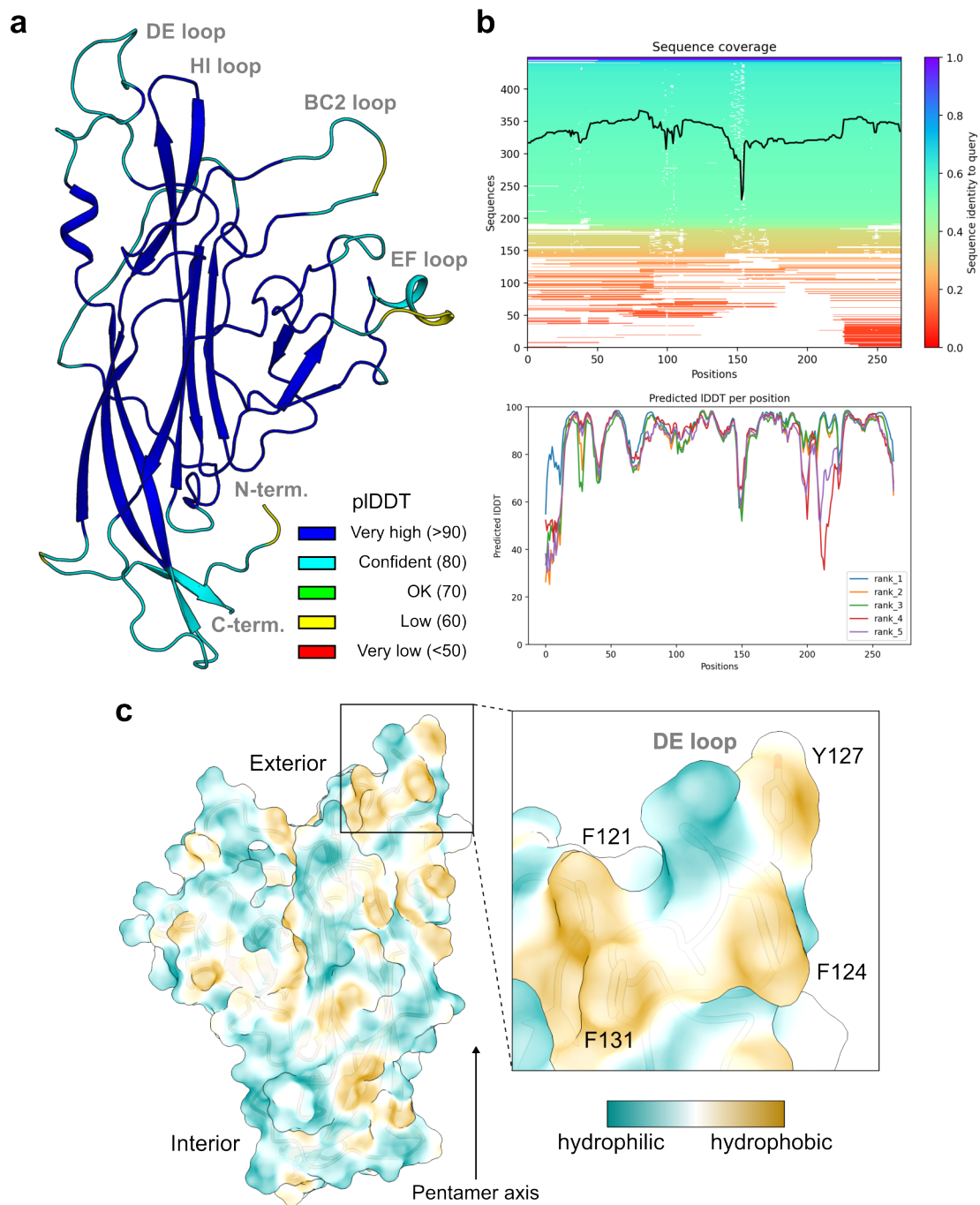
Figure 17: The prediction of the APyV VP1 structure. **a)** The monomer of APyV VP1 as modelled by AlphaFold2. The protein is shown as a cartoon with the coloring indicating the confidence measure (predicted lDDT, plDDT) as illustrated in the legend. **b)** Numbers of aligned sequences and their sequence identities plotted against the APyV VP1 query sequence (above) and plDDT-per-position chart (below). **c)** The APyV VP1 surface colored by hydrophobicity as rendered by ChimeraX [158]. "Interior" and "Exterior" refer to the relative positions in the viral capsid. The arrow indicates the five-fold symmetry axis of the pentameric VP1. The four exposed hydrophobic residues situated in the DE loop are labeled in the magnification panel.
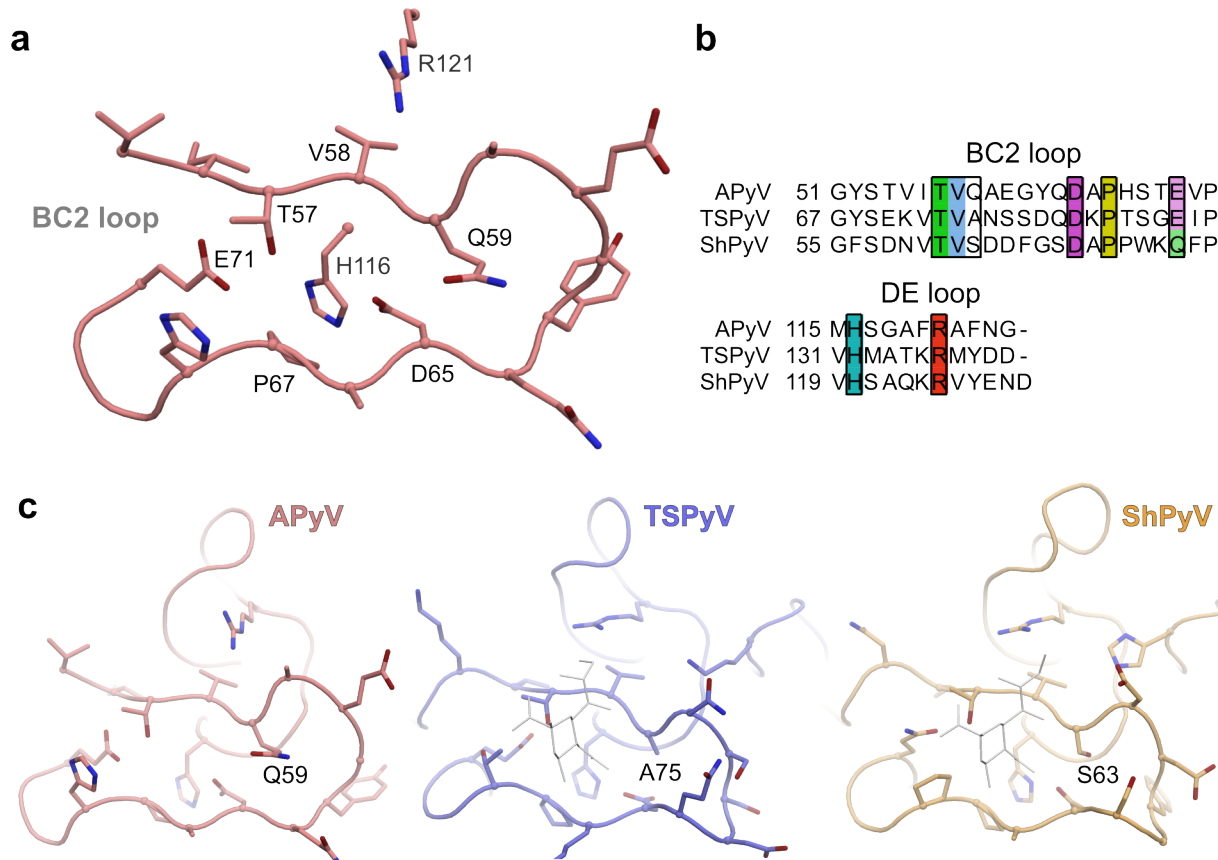
Figure 18: The putative sialyl glycan binding site of APyV VP1. **a)** The BC2 loop structure is shown as predicted by AlphaFold. The side chains of the DE loop residues histidine 116 (H116) and arginine 121 (R121) were added for completeness. **b)** The MSA of the sialyl binding site residues. Positions contributing to glycan engagement in TSPyV VP1 are highlighted in boxes with Clustal coloring. **c)** Model of the APyV VP1 sialyl glycan binding site in juxtaposition with the experimental sialyl complex structures of TSPyV VP1 (PDB 4U60) and ShPyV VP1 (PDB 6Y64) with the bound Neu5Ac moieties outlined. All oxygen and nitrogen atoms are colored red and blue, respectively.

# 3.2 Engagement of non-sialylated glycan ligands by polyomavirus major capsid proteins

## 3.2.1 Chimpanzee polyomavirus

### 3.2.1.1 Background

In Ströh *et al.* [1], we described the unliganded crystal structure of ChPyV VP1. The binding to sialyl glycan ligand 3'SL was not structurally characterized by X-ray crystallography but assessed via STD NMR. We showed that the putative sialyl binding site and the 3'SL STD NMR spectrum of ChPyV VP1 were highly similar to the results obtained for the VP1 of human NJPyV. Thus, NJPyV and ChPyV were considered a "human-simian pair" of viruses, which bears implications for determining the evolutionary routes and the factors guiding viral host tropisms. However, the VP1 structures and STD NMR spectra of NJPyV VP1 and ChPyV VP1 in complex with their ligands were too similar to signify glycan binding preferences. In response, we decided to perform broad-spectrum glycan array analyses on these proteins to gain a more detailed understanding of their glycan binding profiles.

### 3.2.1.2 Glycan array analyses of ChPyV VP1 and NJPyV VP1

*Note: A slightly different glycan array configuration was used for ChPyV VP1 and NJPyV VP1. Numbers in this paragraph do not match the glycan probe list found in Table A2 of the Appendices.*

The His6-tagged VP1s of ChPyV and NJPyV were prepared as described in Ströh *et al.* [1] (see Figure 20). Subjecting the proteins to broad-spectrum glycan array screening resulted in similar glycan binding profiles (see Figure 19). Glycan probes terminating in galactoses yielded higher signals than those bearing terminal sialic acids. Signals for both α- and β-linked galactoses were detected. The strongest signal for NJPyV VP1 derived from the globotriose probe (Gb3, chart pos. 24, Galα-4Galβ-4Glc), whereas ChPyV VP1 preferentially bound to para-lacto-*N*-hexaose (pLNH, pos. 132, Galβ-3GlcNAcβ-3Galβ-3GlcNAcβ-3Galβ-4Glc). Both proteins recognized fucosyl-para-lacto-*N*-hexaose IV (pLNFH-IV, pos. 175, Galβ-3GlcNAcβ-3Galβ-4[Fucα-3]GlcNAcβ-3Galβ-4Glc) followed by lacto-*N*-tetraose (LNT, pos. 72, Galβ-3GlcNAcβ-3Galβ-4Glc) with similar signal strengths. While these glycans are naturally present in sources such as glycolipids and milk, also the artificial oligosaccharide galactotetraose (Gal4) was engaged by both proteins (pos. 697, Galα-3Galβ-4Galα-3Gal).

Among the sialylated glycans, ChPyV VP1 detected 3'-sialyl lacto-*N*-fucopentaose III (SA(3')-LNFP-III, pos. 817, NeuAcα-3Galβ-4[Fucα-3]GlcNAcβ-3Galβ-4Glc) with the highest signal, which NJPyV VP1 did not bind at all. The only sialyl glycan detected by NJPyV VP1 was sialyl-lacto-*N*-tetraose a (LSTa, pos. 805, NeuAcα-3Galβ-3GlcNAcβ-3Galβ-4Glc), which ChPyV VP1 bound much weaker. Except for the Galβ-3/4GlcNAc linkage and the fucosylation, LSTa and SA(3')-LNFP-III share a scaffold, representing a potential basis for tropism differences.

The crystal structure of NJPyV VP1 in complex with 3'SL (NeuAcα-3Galβ-3Glc) is available (PDB 6Y5Y), resembling LSTa without the *N*-acetylation. For the structural investigation of ChPyV VP1 glycan binding, 3'-sialyl Lewis X tetrasaccharide (sLeX, NeuAcα-3Galβ-4[Fucα-3]GlcNAc), an integral part of SA(3')-LNFP-III was chosen. Furthermore, the non-sialylated Gb3 (α-linked Gal) and LNT (β-linked Gal) were selected.
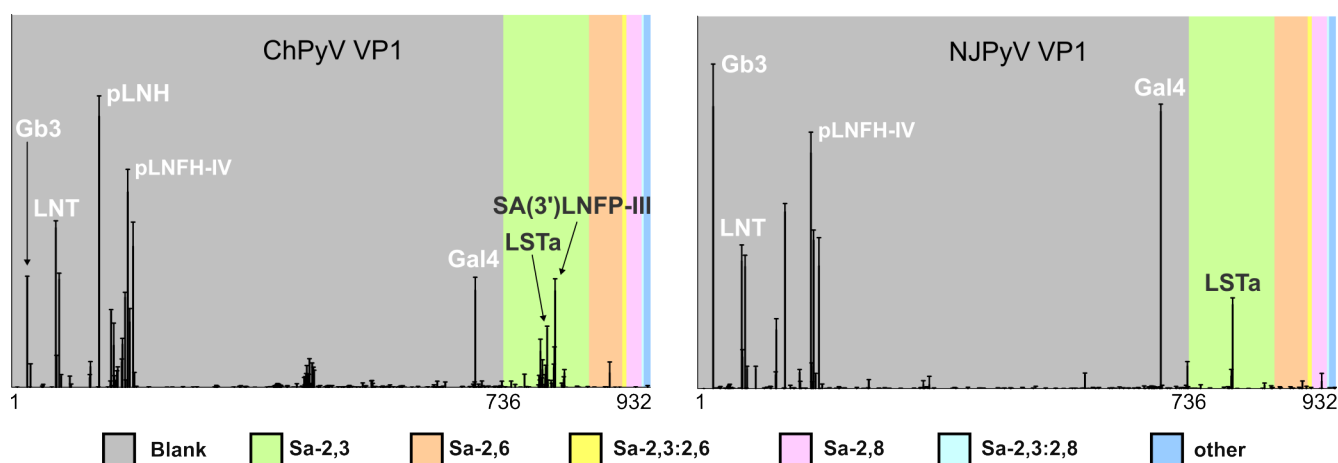


Figure 19: The glycan array binding profiles of ChPyV VP1 and NJPyV VP1. Glycan probes are sorted by sialyl linkages as indicated by the color code. Although the intensities for the individual glycan probes differ, the signal profiles are similar. ChPyV VP1 binds several 2,3-linked sialic acids (green background), whereas NJPyV VP1 only recognizes LSTa (Pos. 805). Charts kindly provided by the Glycosciences Laboratory, Imperial College London.

### 3.2.1.3 Purification of ChPyV VP1

The purification of ChPyV VP1 for crystallization was performed by student Melissa Weldle under my supervision, following the protocol described in the Methods of this thesis. The resulting SDS-PAGE gels are shown in Figure 20 and Figure 43. The purification of the ChPyV VP1 batch used in the glycan array screening was performed by Dr. Luisa J. Ströh without my participation and is not shown here.
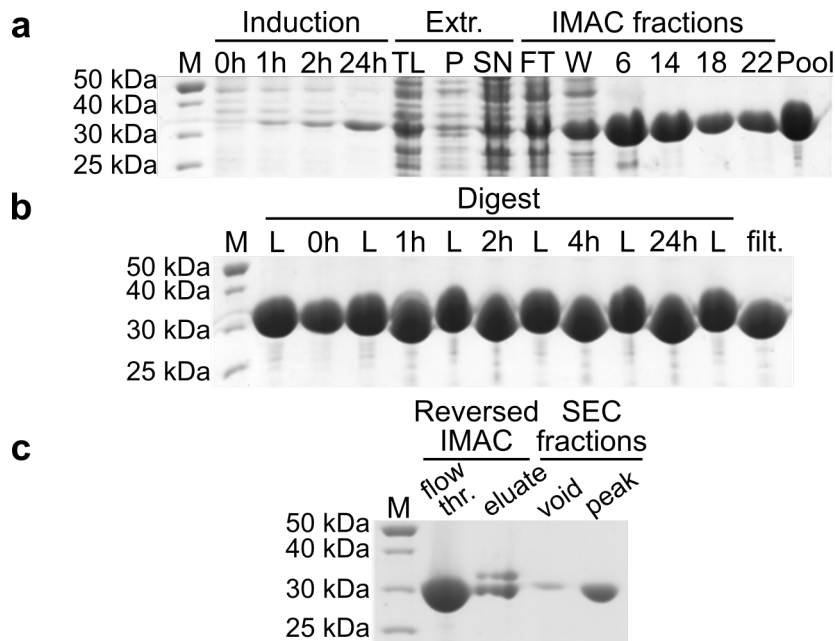
Figure 20: SDS-PAGE analyses of the ChPyV VP1 purification. **a)** Induction, extraction, and IMAC purification. After adding 400 µM IPTG, His6-VP1 bands manifest above 30 kDa with growing amounts during 24h. After extraction, bulk His6-VP1 was present in the total lysate (TL) and supernatant (SN) with minute amounts in the pellet sample (P). The IMAC column overcharged during load, and His-VP1 remained in the flow-through and wash fractions (FT and W, respectively). Large amounts of His6-VP1 eluted using 20% buffer $B_{his}$ (sample 6). The main yield was eluted within fractions 14-22 using 60% buffer $B_{his}$. The main peak fractions were combined (sample pool). **b)** The digest of His6-VP1 (sample L) using TEV protease was monitored over 24 h and yielded completely digested VP1. After 4 h, no traces of His6-VP1 remained visible on the gel. The difference in molecular weight is ~2 kDa. Subsequently, the protein was filtered (sample filt.) for further purification. **c)** The reversed IMAC of the digested VP1 yielded pure ChPyV VP1 (~30 kDa) in the flow through (sample flow thr.). In contrast, partially digested pentamers were eluted at 60% buffer $B_{his}$ (sample eluate). The subsequent SEC, used for buffer exchange and the final purification of ChPyV VP1, resulted in a single species of pure protein (sample peak) with only minor quantities in the void volume (sample void). Throughout the figure, marker lanes (M) indicate the molecular weights. For the raw scans of the SDS-PAGE gels, please refer to Figure 42 in the Appendices.

### 3.2.1.4 ChPyV VP1 complex structures reveal oligosaccharide binding at the outer protein surface

The VP1 of ChPyV was crystallized as describe [1]. The crystal structures of ChPyV VP1 in complex with globotriose (Gb3) and lacto-*N*-tetraose (LNT) could be obtained via soaking. In contrast, crystal derivatization trials using sLeX glycan soaking did not succeed due to blocked binding sites in the crystal structure. Thus, the sLeX tetrasaccharide was preemptively added to the crystallization mixture to perform co-crystallization.

Table 13: Data and refinement statistics for the ChPyV VP1 crystal structures.

| Data set | sLeX | Gb3 | LNT |
|---|---|---|---|
| **Data processing** | | | |
| Wavelength (Å) | 1.0 | 1.0 | 1.0 |
| Resolution range (Å) | 41.5 - 1.40 (1.45 - 1.40) | 44.5 - 1.85 (1.92 - 1.85) | 44.6 - 1.80 (1.86 - 1.80) |
| Space group | P $2_1$ $2_1$ $2_1$ | P 1 | P 1 |
| Unit cell | | | |
| a, b, c (Å) | 94.5, 131.1, 130.0 | 63.5, 82.2, 82.6 | 63.7, 82.3, 82.6 |
| $\alpha$, $\beta$, $\gamma$ (°) | 90, 90, 90 | 69.1, 77.3, 77.6 | 69.2, 77.3, 77.5 |
| Total reflections | 4,662,110 (428,752) | 485,584 (46,342) | 418,468 (31,250) |
| Unique reflections | 314,306 (30,947) | 127,111 (12,515) | 134,523 (12,113) |
| Multiplicity | 14.8 (13.9) | 3.8 (3.7) | 3.1 (2.6) |
| Completeness (%) | 99.8 (99.0) | 99.0 (97.4) | 95.7 (86.2) |
| Mean I/ $\sigma$(I) | 12.7 (1.3) | 8.8 (1.4) | 8.8 (1.2) |
| Wilson B-factor (Å$^2$) | 14.0 | 20.5 | 19.5 |
| $R_{meas}$ | 0.15 (2.00) | 0.14 (1.05) | 0.13 (1.01) |
| $CC_{1/2}$ | 1.00 (0.53) | 1.00 (0.59) | 1.00 (0.45) |
| **Refinement** | | | |
| $R_{work}$ / $R_{free}$ (%) | 16.9 / 18.3 | 17.2 / 21.3 | 16.8 / 20.0 |
| No. of atoms | | | |
| Protein | 11,004 | 10,656 | 10,702 |
| Glycans | 379 | 205 | 207 |
| Solvent | 1,994 | 1,177 | 1,353 |
| RMSDs | | | |
| Bonds (Å) | 0.009 | 0.009 | 0.010 |
| Angles (°) | 1.56 | 1.58 | 1.58 |
| Ramachandran plot | | | |
| Favored (%) | 96.3 | 95.6 | 96.5 |
| Allowed (%) | 3.7 | 4.4 | 3.5 |
| Outliers (%) | 0.0 | 0.0 | 0.0 |
| Rotamer outliers (%) | 0.4 | 1.3 | 0.4 |
| Clashscore | 3.9 | 3.3 | 2.2 |
| Average B-factors (Å$^2$) | | | |
| Protein | 14.9 | 27.7 | 22.6 |
| Glycans | 26.3 | 40.0 | 34.6 |
| Solvent | 29.9 | 35.6 | 32.1 |

Values in parentheses refer to the highest resolution bin.

Diffraction of the resulting VP1-sLeX co-crystals revealed that the complex crystallized in the space group P $2_1$ $2_1$ $2_1$ as opposed to the P 1 space group of the native crystals. The crystal selected for structure elucidation diffracted to a resolution of 1.4 Å (see Table 13). The complex structures of ChPyV VP1 with Gb3 and LNT had similar space group parameters as the apo crystal structure [1]. Again, the crystals chosen for data collection resulted in high-resolution diffraction of 1.85 Å and 1.80 Å, respectively (see Table 13). Molecular replacement resulted in single pentamers of ChPyV VP1 (see Figure 21a) in

the ASUs of all complexes. Structure evaluation confirms that the ChPyV VP1-sLeX oligosaccharide complex maintains the intermolecular contacts of NJPyV VP1-3'SL [1] with an additional fucose-mediated hydrogen bond (see Figure 22a), inducing a slight conformational shift compared to 3'SL (see Figure 22b).
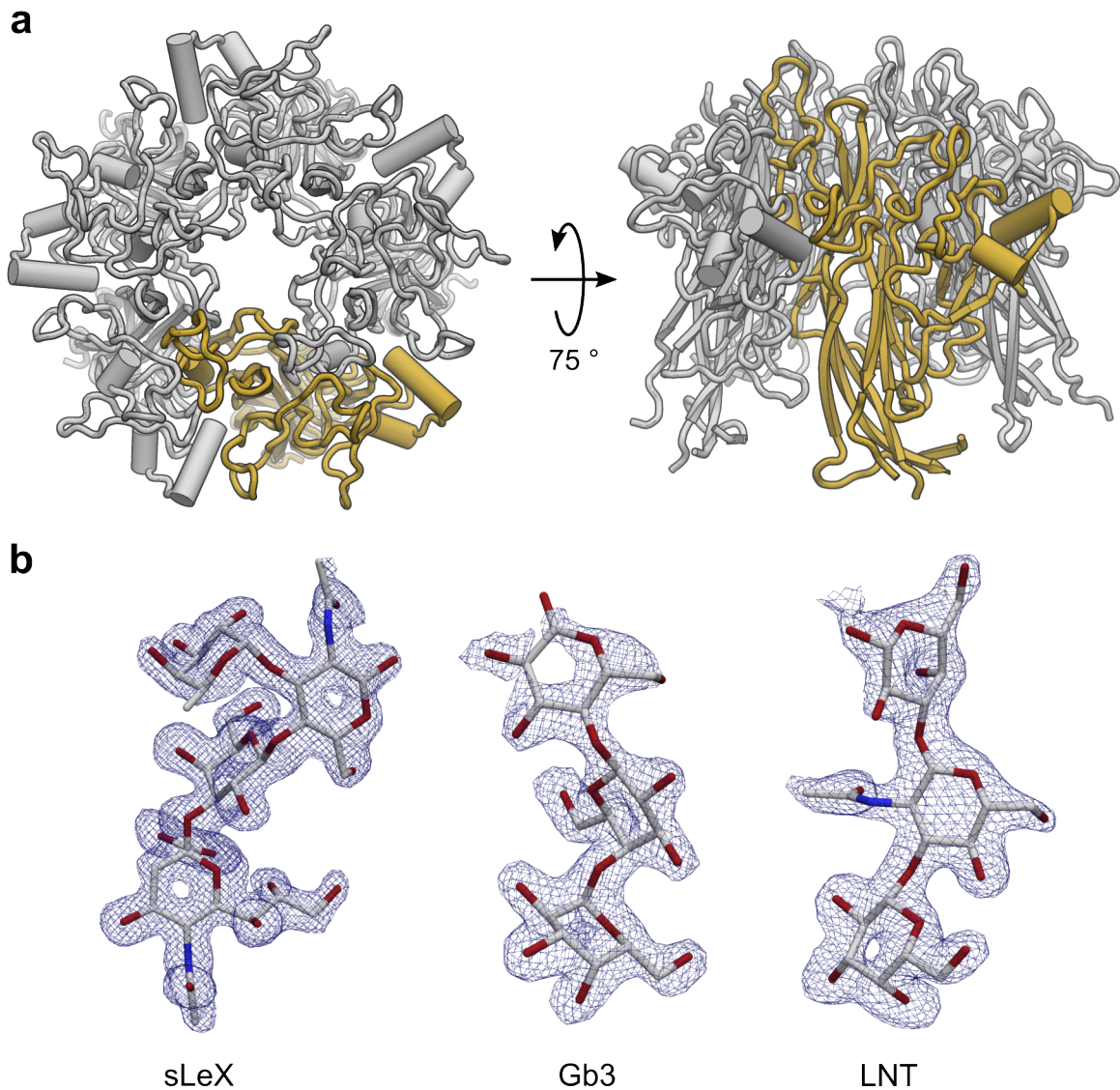
**a**



**b**



| sLeX | Gb3 | LNT |

Figure 21: Results of the crystallographic ChPyV VP1 complex formation using sialylated and neutral oligosaccharides. **a)** The overall structure of the ChPyV VP1 pentamer is shown from above (left) and tilted around the lateral axis (right). The protein is shown in cartoon representation with one monomer highlighted in gold. Alpha helices are displayed as cylinders. **b)** All resolved ligands of ChPyV VP1 are shown side by side, drawn as sticks with the carbon, oxygen, and nitrogen atoms colored in white, red, and blue, respectively. The Fo-Fc difference maps calculated from unliganded VP1 are displayed as blue meshes around the individual oligosaccharides with a contour level of $\sigma = 2.5$ and a radius of 1.6 Å. For the map resolutions, please refer to Table 13.
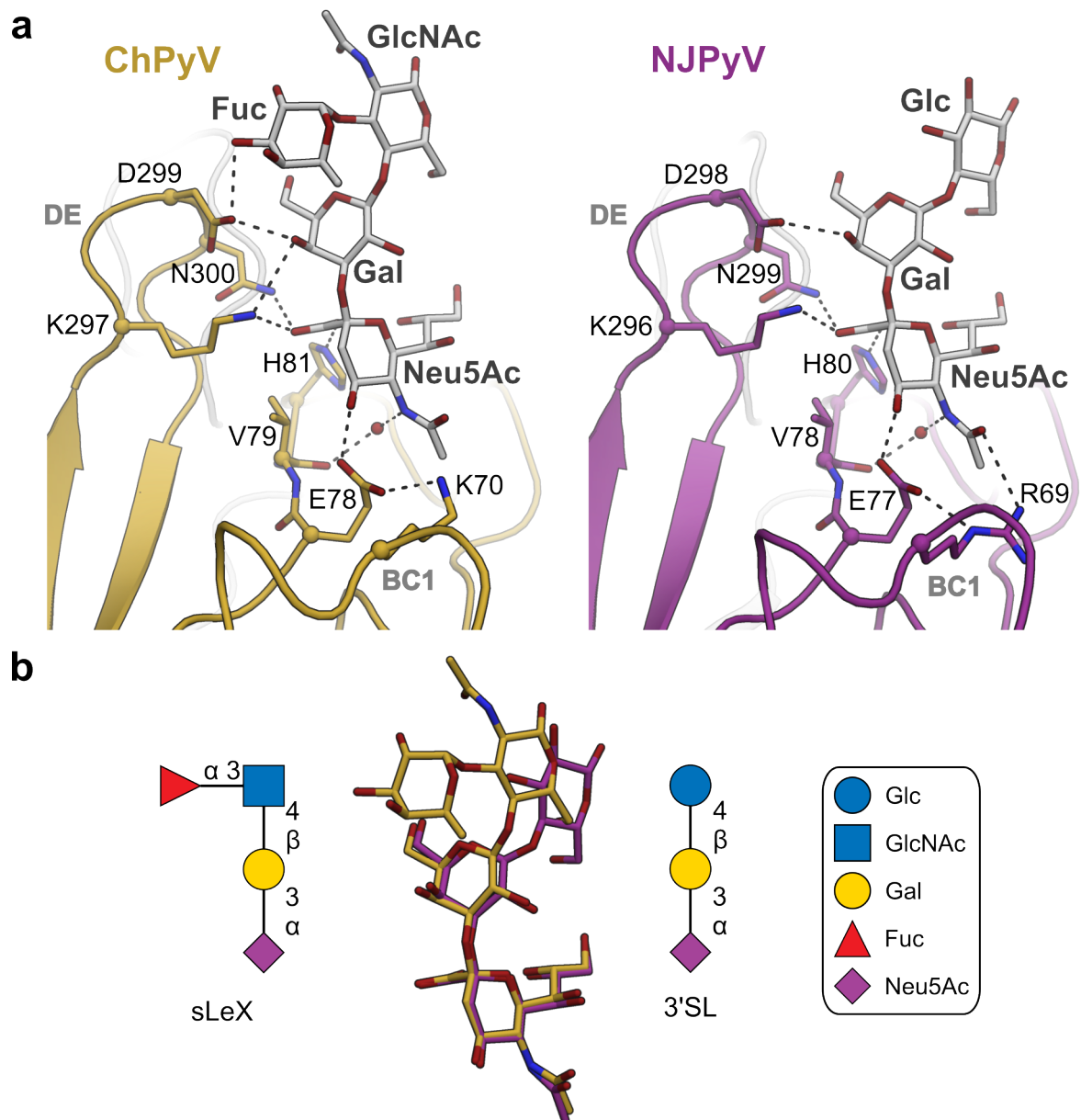
Figure 22: The sialyl oligosaccharide complex structures of ChPyV VP1 and NJPyV VP1.
**a)** Comparison of the sLeX and 3'SL complexes. Oligosaccharides with white carbon atoms and binding site residues are shown as sticks. Loops are labeled in bold gray. Intermolecular contacts are drawn as dashed lines. The fucose in sLeX mediates an additional hydrogen bond to the side chain of Asp 299. **b)** Structural superposition of the complexed sLeX (in gold) and 3'SL (in purple) oligosaccharides. Their symbol formulas are displayed according to the Symbol Nomenclature for Glycans legend. All oxygen and nitrogen atoms are colored red and blue.

**3.2.1.5 ChPyV VP1 engages galactosyl ligands at a novel binding site**

After the Gb3 and LNT complex structures of ChPyV VP1 were solved via molecular replacement, structure refinement revealed a novel binding site for galactosyl ligands. This site is located beside the sialyl binding site, approximately 15 Å apart in a clockwise direction (see Figure 23a). While the sialyl site is mainly formed by a single VP1 chain, the galactosyl site manifests at the interface of two adjacent chains. The terminal galactoses of Gb3 and LNT bind in an aligning position, as shown in the magnified panel of Figure 23a. However, due to their differing α- and β-linkages, the adjoining monosaccharides divert in different directions and establish differing molecular interactions with VP1 (see Figure 23b). For a list of these contacts, please refer to Table 14.

In both complex structures (Gb3 and LNT), the terminal galactose establishes direct hydrogen bonds with the side chain of His 81 and the backbone of Glu 142. Indirect contacts, mediated by water molecules, occur to the side chains of Tyr 86cw and Glu 142 and the backbone atoms of His 82, Leu 140, and Ala 143. Additionally, a less polar CH-π interaction [184] occurs between the methylene group of Gal and the aromatic electron density of Tyr 86cw. The β-Gal of Gb3 does not directly interact with VP1 and only maintains two water-mediated contacts to the side chains of Tyr 86cw and Asp 87cw. The β-GlcNAc of LNT establishes a direct H-bond to the side chain of Asp 299 and water-mediated contacts to backbone atoms of Asp 299 and Ala 143. The remaining monosaccharides of Gb3 and LNT do not interact with VP1. They are visible in the electron densities (see Figure 21b), likely due to adopting preferred conformations within the oligosaccharides.

**3.2.1.6 The galactosyl binding site of ChPyV VP1 is conserved in NJPyV VP1**

Based on the sequences in the NCBI Reference Sequence Database (Refseq) [161], ChPyV VP1 (YP_004046682.1, 497 residues) and NJPyV VP1 (YP_009030020.1, 489 residues) are 83% identical (and 93% similar). In Ströh *et al.* [1], we showed that the sialyl binding sites in NJPyV VP1 and ChPyV VP1 are highly similar, which I confirmed by the structure of sLeX in complex with ChPyV VP1 (see Results, section 3.2.1.4). The novel galactosyl binding site is located at a different part of the ChPyV VP1 surface, again allowing for the assessment of conservation in NJPyV VP1. After deducing the binding site residues in the ChPyV VP1-Gb3 complex structure, structural and sequence alignments show that the galactosyl binding site is conserved in the VP1 of NJPyV (see Figure 24). Implications for host tropism will be discussed in the next chapter.
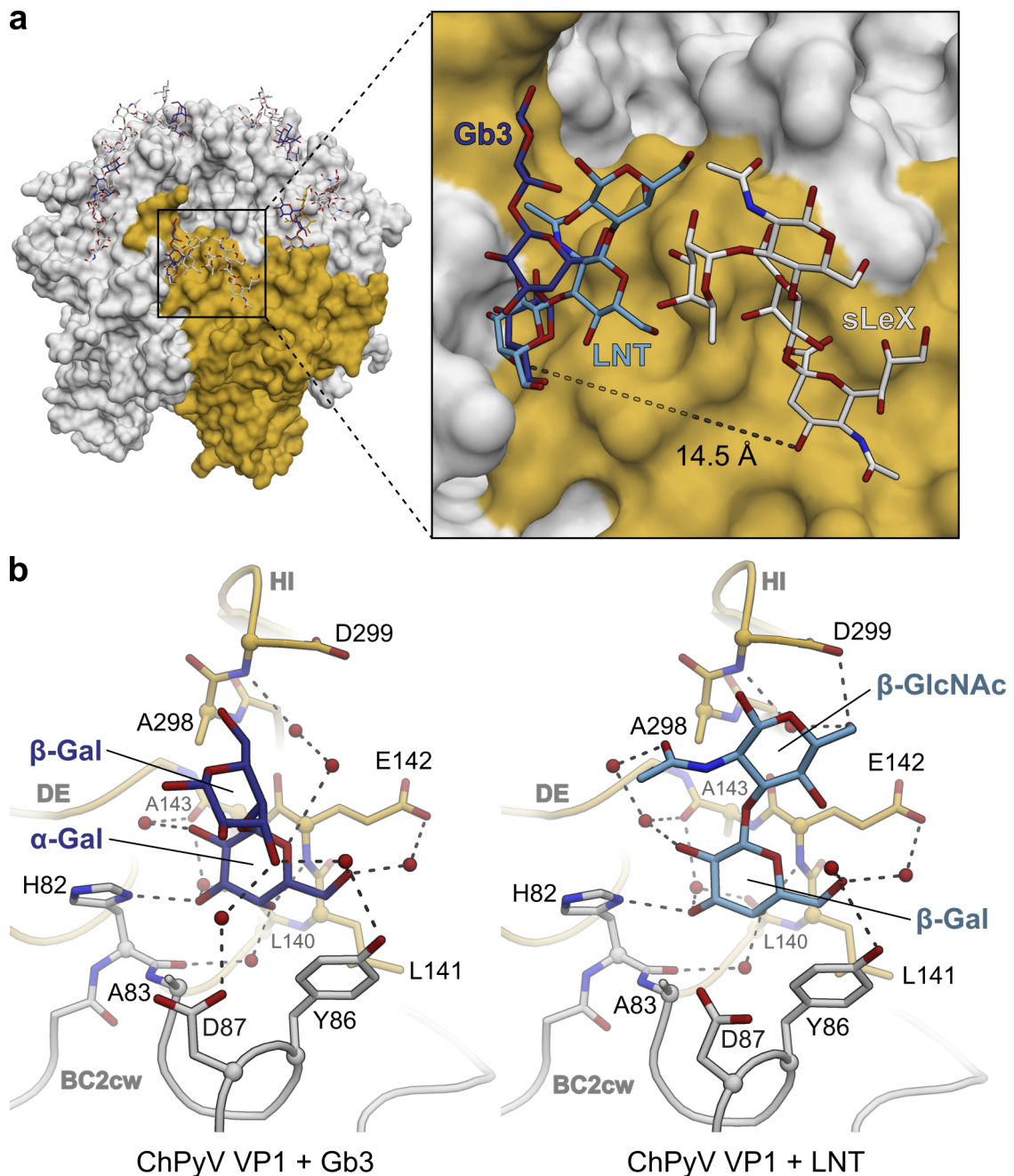
**ChPyV VP1 + Gb3**

**ChPyV VP1 + LNT**

Figure 23: The glycan binding sites of ChPyV VP1. **a)** Structural alignment of the three ligand structures shows that the glycan binding sites are located on the outer surface of the ChPyV VP1 pentamer, displayed with one monomer colored gold. While the binding site of sLeX is mostly constituted by a single monomer, galactosyl ligand binding occurs at the interface of two neighboring VP1 monomers. The distance between the O4 atoms of the sialic acid and the galactoses is indicated as a dashed line. **b)** Molecular interactions between ChPyV VP1 and Gb3 (left) or LNT (right), which are displayed as their terminal disaccharides. Protein loops are labeled in bold gray. Hydrogen bonds are drawn as dashed lines. Water molecules are shown as spheres. All oxygen and nitrogen atoms are colored red and blue, respectively.

Table 14: Intermolecular contacts between ChPyV VP1 and galactosyl ligands.

| | | Gb3 | | | | LNT | |
|---|---|---|---|---|---|---|---|
| | | Contact | Distances (Å) | | | Contact | Distances (Å) |
| α-Gal | O2 | H2O-*A143* | 2.6-2.8 | β-Gal | O2 | H2O-*A143* | 2.6-2.8 |
| | O3 | H82 | 2.7 | | O3 | H82 | 2.8 |
| | | H2O-*L140* | 2.7-2.8 | | | H2O-*L140* | 2.7-2.8 |
| | | H2O-*A143* | 2.7-2.8 | | | H2O-*A143* | 2.7-2.9 |
| | O4 | *E142* | 2.7 | | O4 | *E142* | 2.9 |
| | | H2O-*H82* | 2.6-2.7 | | | H2O-*H82* | 2.6-2.8 |
| | O6 | H2O-Y86 | 2.8-3.1 | | O6 | H2O-Y86 | 2.7-3.2 |
| | | H2O-E142 | 2.8-2.7 | | | H2O-E142 | 2.6-2.8 |
| | | | | | | | |
| β-Gal | O3 | H2O-Y86 | 2.6-3.1 | β-GlcNAc | O6 | D299 | 2.8 |
| | | H2O-D87 | 2.9-3.0 | | | H2O-*D299* | 2.9-3.0 |
| | | | | | O7 | H2O-H2O-*A143* | 2.7-2.7-2.8 |

Distance values are derived from single binding sites and may alter between chains.
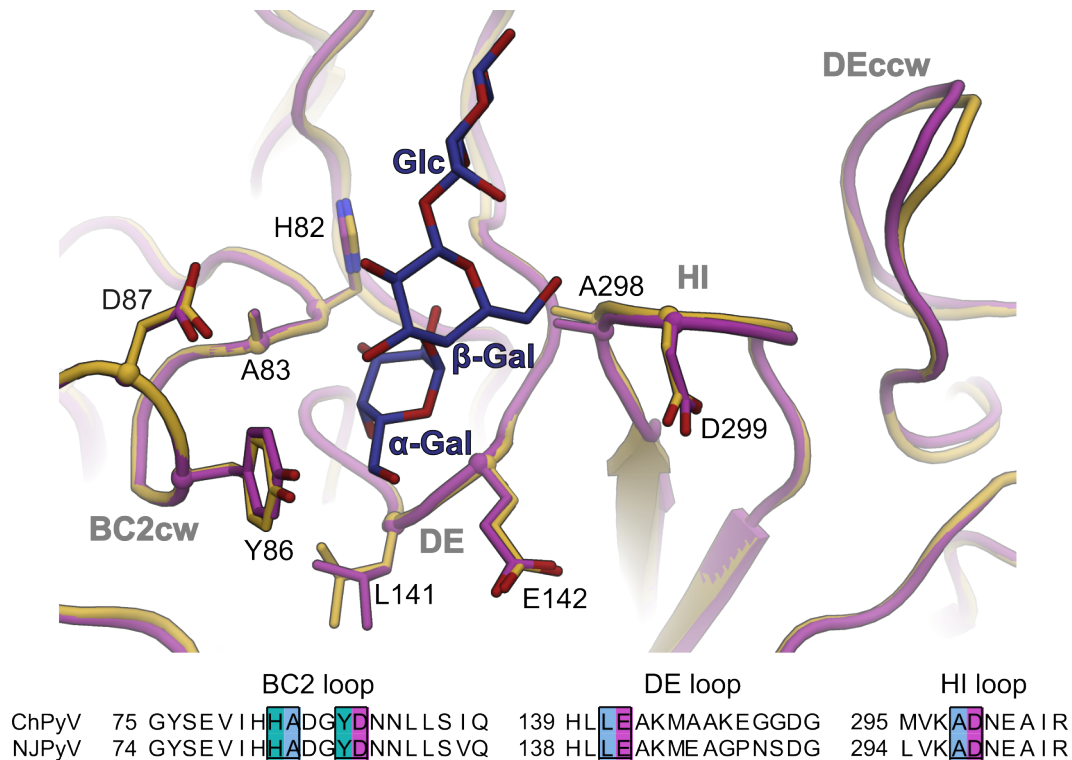Residues contributing backbone atoms are printed in italics.



Figure 24: Structural and sequence alignments of the galactosyl binding sites in ChPyV VP1 (gold) and NJPyV VP1 (pink). VP1 residues 5 Å around the Gb3 trisaccharide (purple) are drawn as sticks. Oxygen and nitrogen atoms are red and blue, respectively. Loops are labeled in bold gray. The ChPyV VP1 galactosyl binding site residues are completely conserved in NJPyV VP1, as highlighted in the pairwise sequence alignment.

## 3.2.2 Lyon IARC polyomavirus

### 3.2.2.1 Background

LIPyV was identified in a human skin sample and thus designated, chronologically, as the 14th human polyomavirus [140]. Only later, it was found that no seroreactivity against LIPyV was traceable in a human cohort [143]. Instead, LIPyV was detected in stool samples of diarrheic cats in Canada and China [141, 142]. Hence, LIPyV is currently considered a feline virus. LIPyV was only distantly related to a structurally characterized polyomavirus, the Merkel cell polyomavirus (MCPyV) with 61% VP1 sequence identity. Besides its initial alleged human origin, the phylogenetic relation of LIPyV to tumor-associated viruses, including MCPyV and a raccoon polyomavirus, prompted interest in characterizing its glycan engagement. Crystallization was successful and yielded well-diffracting crystals. Yet, crystal derivatization trials using a range of sialylated glycan analogs such as 2-*O*-Methyl-α-D-*N*-acetylneuraminic acid (2-*O*-Me-Neu5Ac), 3-sialyllactose (3'SL), and 6-sialyllactose (6'SL) remained fruitless. To identify potential receptor candidates, recombinant LIPyV VP1 was subjected to broad-spectrum glycan array screening, eventually leading to the first neolacto-series glycan complex structures of a polyomavirus VP1.

### 3.2.2.2 Expression and purification

A synthetic LIPyV VP1 gene, optimized for *E. coli* expression and cloned into the pET-15b vector (see Materials, section 2.1.8.1), was transformed into *E. coli* BL21 (DE3). The expression was conducted as described in the Methods. After extraction of the soluble protein fraction, LIPyV VP1 was purified by IMAC (see Figure 25a). For glycan array screening, the His6-tagged protein was directly subjected to SEC (see Figure 10). For crystallization, the His6-tag was removed via site-directed digestion by TEV protease, followed by IMAC and SEC purification (see Figure 25b-d).

### 3.2.2.3 Crystallization

To find a suitable crystallization condition for LIPyV VP1, the protein was concentrated to 9.6 mg/mL and subjected to initial robot-pipetted crystal screening. Using the Crystal Gryphon LCP (Art Robbins Instruments), the sitting drop screening plates were set up with drops comprising protein solution and screening conditions in a 1:1 ratio (200 nL each). After sealing, the equilibration took place at 20 °C. After five days, LIPyV VP1 revealed a propensity to crystallize in various solutions containing PEG 3350. Most of
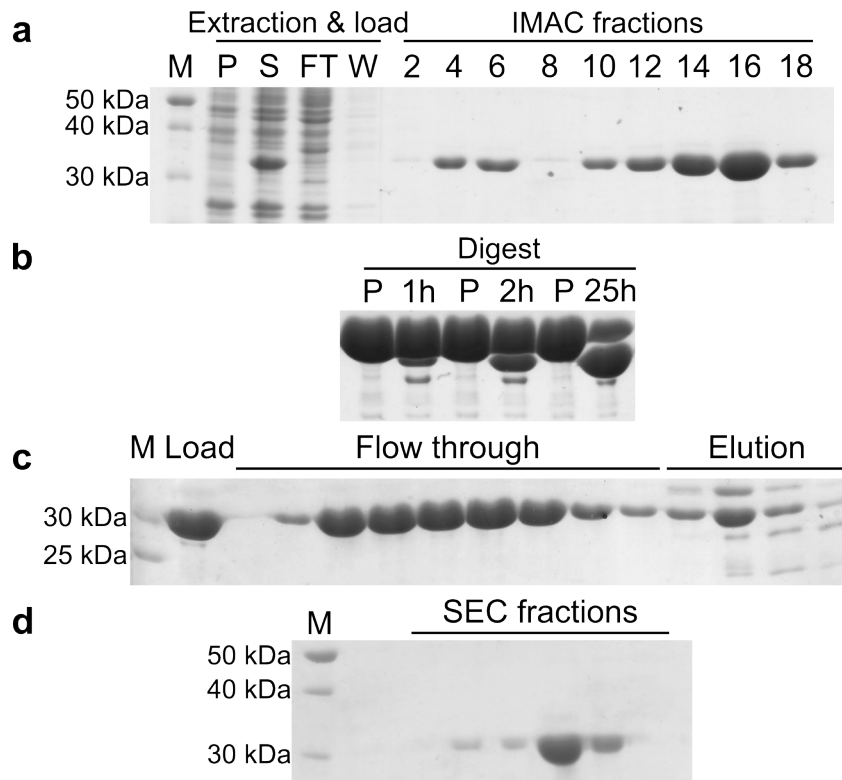
Figure 25: SDS-PAGE analyses of the LIPyV VP1 purification. **a)** Protein extraction and IMAC purification. The His6-VP1 runs above 30 kDa and was present in the soluble fraction (S) of the extraction but not in the pellet sample (P). The protein remained on the column during load and no His6-VP1 was present in the flowthrough (FT) and wash (W) samples. The second wash with 20% buffer $B_{his}$ (fractions 2-8) eluted some His6-VP1, while the major yield occured within fractions 10-18 using 60% buffer $B_{his}$. **b)** The digest of His6-VP1 using TEV protease, monitored over 25 h in comparison to the His6-VP1 IMAC pool (P), yielded ~60% digested VP1. The difference in molecular weight amounts to ~2 kDa. **c)** The reversed IMAC of the digested VP1 (Load fraction) yielded pure LIPyV VP1 (~30 kDa) in the flow through, while partially digested pentamers and impurities were eluted at 60% buffer $B_{his}$ (Elution). **d)** SEC for buffer exchange and final purification resulted in pure LIPyV VP1. Throughout the figure, marker lanes (M) indicate the molecular weights. For the raw scans of the SDS-PAGE gels, please refer to Figure 44 in the Appendices.

the crystal grew from drops with vast initial precipitation. However, the drops cleared up as the crystals continued to grow for the next six months, implying that the precipitation of LIPyV VP1 was reversible (see Figure 26).

### 3.2.2.4 The overall crystal structure of apo LIPyV VP1

After data collection of a native LIPyV VP1 crystal (PEG2-21 condition, see Table 15), the structure was solved to 1.45 Å resolution via molecular replacement using the unli-

Table 15: Selection of crystallization conditions for the VP1 of Lyon-IARC polyomavirus

| Condition | Buffer/ salt solution | Precipitant |
|---|---|---|
| PEGIon1-2 | 0.2 M Potassium fluoride | 20% w/v Polyethylene glycol 3350 |
| PEGIon1-11 | 0.2 M Potassium iodide | 20% w/v Polyethylene glycol 3350 |
| PEGIon1-36 | 0.2 M Sodium tartrate dibasic dihydrate | 20% w/v Polyethylene glycol 3350 |
| PEGIon1-37 | 0.2 M Potassium sodium tartrate tetrahydrate | 20% w/v Polyethylene glycol 3350 |
| PEGIon1-44 | 0.2 M Ammonium phosphate dibasic | 20% w/v Polyethylene glycol 3350 |
| PEGIon2-8 | 0.2 M Sodium malonate pH 7.0 | 20% w/v Polyethylene glycol 3350 |
| PEGIon2-11 | 4% v/v Tacsimate pH 5.0 | 12% w/v Polyethylene glycol 3350 |
| PEGIon2-15 | 4% v/v Tacsimate pH 5.0 | 12% w/v Polyethylene glycol 3350 |
| PEGIon2-16 | 8% v/v Tacsimate pH 7.0 | 20% w/v Polyethylene glycol 3350 |
| PEGIon2-21 | 0.1 M Ammonium citrate tribasic pH 7.0 | 12% w/v Polyethylene glycol 3350 |
| PEGIon2-30 | 0.2 M Ammonium tartrate dibasic pH 7.0 | 20% w/v Polyethylene glycol 3350 |
| PEGIon2-42 | 0.02 M Calcium chloride dihydrate | 20% w/v Polyethylene glycol 3350 |
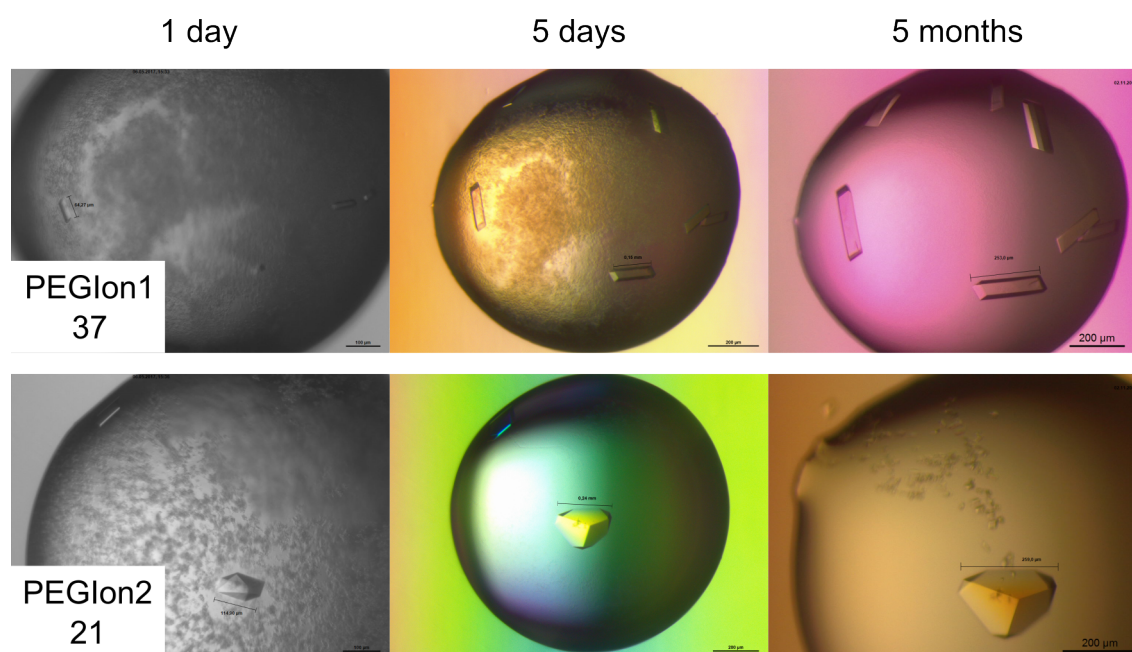


Figure 26: Crystals of LIPyV VP1 are shown for the PEGIon1-37 and PEGIon2-21 conditions in the initial robotic screen. Growth was monitored over five months. Removing the smaller crystal in the PEGIon2-21 drop for seed stock preparation induced additional crystallization.

ganded VP1 pentamer of Merkel cell polyomavirus (MCPyV) as the search model (PDB 4FMG) [95]. For statistics, please refer to Table 16. The refinement yielded a high-quality electron density map, allowing for the manual adjustment of protein sequence mismatches. Subsequently, the LIPyV VP1 structure was refined until the refinement parameters converged (see Methods). The asymmetric unit of the crystal contains a single LIPyV VP1 pentamer. The structurally resolved residues include G43 to D118 and

Table 16: Data and refinement statistics for the LIPyV VP1 crystal structures.

| Data set | apo | 20 mM pLNnH | 50 mM LacNac | 20% Lactose |
|---|---|---|---|---|
| **Data processing** | | | | |
| Wavelength (Å) | 1.0 | 1.0 | 1.0 | 1.0 |
| Resolution range (Å) | 43.2 - 1.45 (1.5 - 1.45) | 43.3 - 1.25 (1.30 - 1.25) | 43.5 - 1.52 (1.57 - 1.52) | 42.8 - 1.80 (1.86 - 1.80) |
| Space group | C 2 | C 2 | C 2 | C 2 |
| Unit cell | | | | |
| a, b, c (Å) | 123.3, 111.6, 134.0 | 123.4, 112.3, 133.8 | 124.0, 112.2, 134.1 | 124.2, 111.7, 134.2 |
| $\alpha$, $\beta$, $\gamma$ (°) | 90, 116.6, 90 | 90, 116.5, 90 | 90, 116.6, 90 | 90, 116.6, 90 |
| Total reflections | 1,918,475 (183,000) | 1,991,900 (193,074) | 1,132,704 (102,804) | 1,028,783 (95,531) |
| Unique reflections | 283,808 (27,459) | 433,623 (41,940) | 248,285 (24,121) | 151,121 (15,050) |
| Multiplicity | 6.8 (6.7) | 4.6 (4.6) | 4.6 (4.3) | 6.8 (6.3) |
| Completeness (%) | 99.2 (96.0) | 96.9 (93.9) | 98.7 (96.1) | 99.9 (99.6) |
| Mean $I/\sigma(I)$ | 16.7 (1.0) | 14.2 (1.3) | 11.8 (1.0) | 9.4 (1.2) |
| Wilson B-factor (Å$^2$) | 21.6 | 14.8 | 23.5 | 24.4 |
| $R_{meas}$ | 0.07 (1.98) | 0.06 (1.24) | 0.07 (1.61) | 0.16 (1.7) |
| $CC_{1/2}$ | 1.00 (0.57) | 1.00 (0.49) | 1.00 (0.53) | 1.00 (0.57) |
| **Refinement** | | | | |
| $R_{work}$ / $R_{free}$ (%) | 17.8 / 20.0 | 16.6 / 17.8 | 17.8 / 21.0 | 17.4 / 20.0 |
| No. of atoms | | | | |
| Macromolecules | 11,154 | 10,933 | 10,860 | 10,901 |
| Glycans | N/A | 213 | 108 | 46 |
| Solvent | 1,526 | 1,826 | 1,439 | 1,451 |
| RMSDs | | | | |
| Bonds (Å) | 0.010 | 0.009 | 0.009 | 0.010 |
| Angles (°) | 1.59 | 1.54 | 1.54 | 1.61 |
| Ramachandran plot | | | | |
| Favored (%) | 95.5 | 96.0 | 96.5 | 96.4 |
| Allowed (%) | 4.3 | 3.9 | 3.5 | 3.6 |
| Outliers (%) | 0.2 | 0.1 | 0.0 | 0.0 |
| Rotamer outliers (%) | 0.8 | 0.3 | 0.3 | 0.7 |
| Clashscore | 4.3 | 3.1 | 2.5 | 2.2 |
| Average B-factors (Å$^2$) | | | | |
| Macromolecules | 25.9 | 16.9 | 27.4 | 27.6 |
| Glycans | N/A | 21.2 | 47.1 | 55.3 |
| Solvent | 36.3 | 29.9 | 37.8 | 37.4 |

Values in parentheses refer to the highest resolution bin.

M123 to N325. The pentameric VP1 is similar to those of other polyomaviruses. The core domain of VP1 adopts a single-JR fold of antiparallel β-sheets that contribute to inter-chain contacts and are connected by extensive protein loops.

### 3.2.2.5 The MCPyV VP1 sialyl binding site is not conserved in LIPyV VP1

The closest structurally characterized relative of LIPyV VP1 is the VP1 of MCPyV with a sequence identity of 61%. Sequence deviations are mainly found in the surface loops: The sequence of LIPyV VP1 is one residue shorter in the BC loop, three residues longer in the DE loop, one residue shorter in the EF loop, and one residue shorter in the HI loop compared to MCPyV VP1 (Figure 27).

```
LIPyV     1 MAPKRKASKPVCSPPPKKDCSQTYVKCPKRCPKVPCVPKLLIKGGIEVLETVSGPDSITQ  60
MCPyV     1 MAPKRKASS--TCKTPKRQC----IPKPGCCPNVASVPKLLVKGGVEVLSVVTGEDSITQ  54

LIPyV    61 IELFLQPRMGVNTYDIERYSNWYGYSYEFHPKDVN-DVPSPETLPSYAAARVPLPPLNED 119
MCPyV    55 IELYLNPRMGVNSPDLPTTSNWYTYTYDLQPKGSSPDQPIKENLPAYSVARVSLPMLNED 114

LIPyV   120 MTCNMIMMWEAVSVKTEVIGINTLINVHKEKQLEIPQYGEHAAGIPIQGPNYHFFSIGGE 179
MCPyV   115 ITCDTLQMWEAISVKTEVVGISSLINVHYWDMKRVHDYG---AGIPVSGVNYHMFAIGGE 171

LIPyV   180 PLDLQGIVADYRTMYPEGEN-KPVNIKTVTKTPMTPKNQGMDPTAKTKLLKDGYYPIEIW 238
MCPyV   172 PLDLQGLVLDYQTEYPKTTNGGPITIETVLGRKMTPKNQGLDPQAKAKLDKDGNYPIEVW 231

LIPyV   239 CPDPSRNENTRYFGSFTGGTDTPPVLQFTNTLTTVLLDENGIGPLCKADGLFVAAADICG 298
MCPyV   232 CPDPSKNENSRYYGSIQTGSQTPTVLQFSNTLTTVLLDENGVGPLCKGDGLFISCADIVG 291

LIPyV   299 FFHQASG-MRYRGLPRYFNVTLRKRWVKNPYGINSLINTLFNNFVPQLKGQPMTGDNSQV 357
MCPyV   292 FLFKTSGKMALHGLPRYFNVTLRKRWVKNPYPVVNLINSLFSNLMPKVSGQPMEGKDNQV 351

LIPyV   358 EEVRVYDGREQLPGGLDLPRIIEQFKPKTFEVNVPVVGDVGDNVGEEHDNDTDNPNNNRD 417
MCPyV   352 EEVRIYEGSEQLPGDPDIVRFLDKFGQEKTVYPKPSVAPAAVTFQSNQQDKGKAPLKGPQ 411

LIPyV   418 NDHDQPLINNVPLITAIL                                          435
MCPyV   412 KASQK------ESQTQEL                                          423
```

Figure 27: The complete pairwise sequence alignment of LIPyV VP1 and MCPyV VP1. The surface loops BC, DE, EF, and HI are highlighted with Clustal coloring.

In structural alignment, the Cα RMSDs between the VP1s of LIPyV and MCPyV amount to 0.50 Å and 1.60 Å, calculated in PyMOL with and without outlier rejection, respectively (Figure 28a). The sialyl binding site of MCPyV VP1 is located at the interface of the BC, DE, and HI loops (see Figure 28b) [95], where the sequence variability is highest. The only two residues contributing to MCPyV VP1 sialyl glycan binding that are conserved in LIPyV VP1 are W81 and Y86 in the BC loop. The other binding site residues are not conserved or do not align due to sequence length mismatches (see Figure 28c & d).
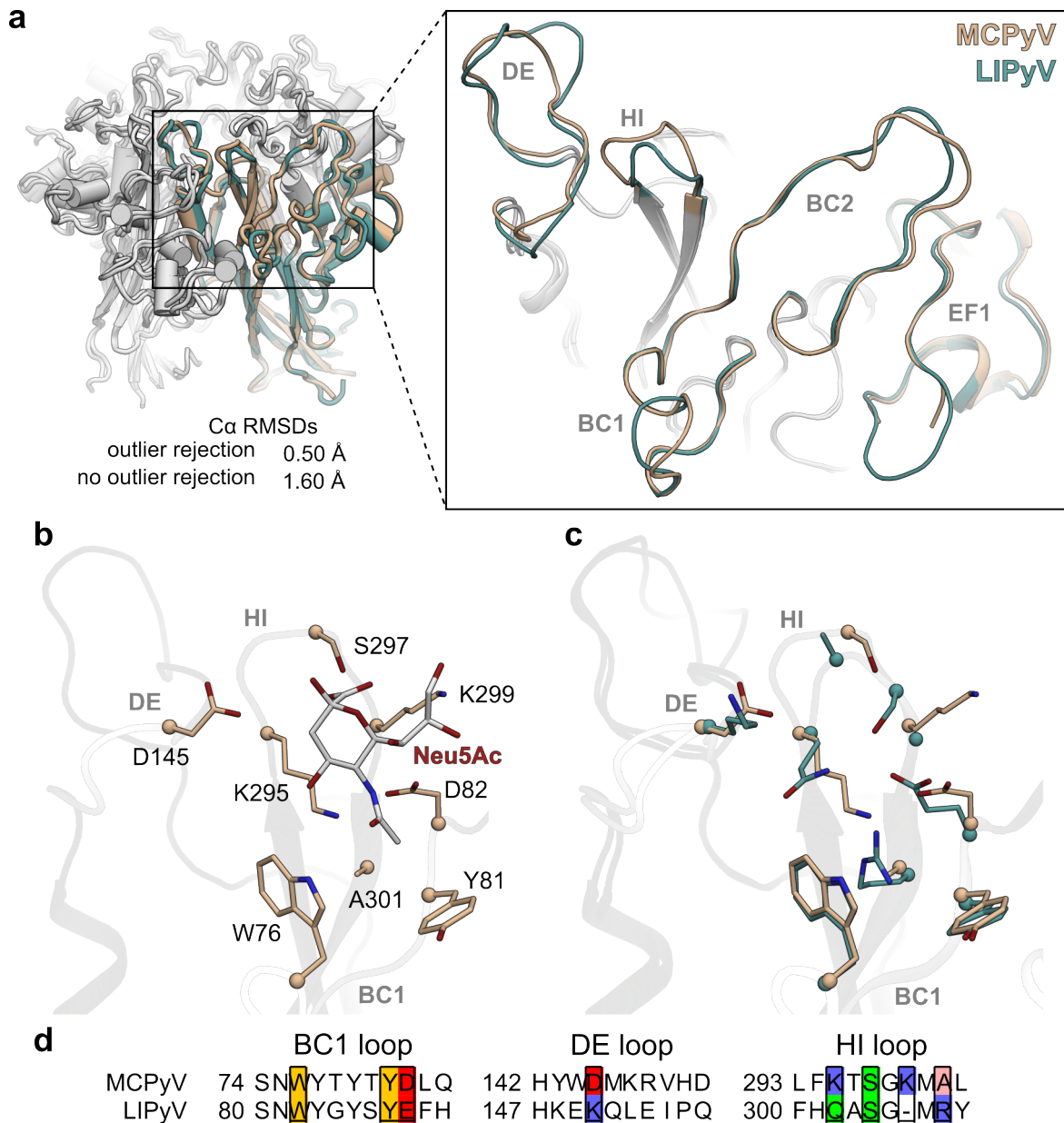
Figure 28: Comparison of the LIPyV and Merkel cell polyomavirus (MCPyV) VP1 structures. **a)** Structural alignment of the pentamers in cartoon representation with one chain colored in apricot (MCPyV VP1) and light teal (LIPyV VP1). Cα RMSD values are indicated as calculated by PyMOL. A magnification focused on the surface loops is shown in the right panel. Protein loops are labeled in bold gray. **b)** The sialyl binding site of MCPyV VP1 (PDB 4FMJ) with the contributing residues and the terminal Neu5Ac moiety (white carbons) drawn as sticks. **c)** The structural alignment of the MCPyV VP1 sialyl binding site and LIPyV VP1 in the same orientation. Protein backbones are faded into the background. Oxygen and nitrogen atoms are colored red and blue, respectively. **d)** The MCPyV VP1 binding site loops in pairwise sequence alignment with LIPyV VP1, with the relevant residues colored by chemical properties, with orange, red, blue, green, and salmon for large and aliphatic, anionic, cationic, neutral, and small and aliphatic, respectively.

### 3.2.2.6 LIPyV VP1 does not bind sialylated glycan analogs *in crystallo*

For the complex formation of LIPyV VP1 with sialylated compounds, a range of different sialyl glycan analogs, including 3'SL, 6'SL, and the GD3 oligosaccharide, were used for co-crystallization and crystal soaking (see Table 17). However, these efforts remained fruitless and repeatedly yielded unliganded LIPyV VP1 indistinguishable from the apo structure. These findings imply that no sialyl binding site is present on the surface of LIPyV VP1. For further investigation of glycan engagement, a sample of LIPyV VP1 was prepared for broad-spectrum glycan array screening (see again Figure 10).

### 3.2.2.7 Glycan array screening signifies neolacto glycan binding by LIPyV VP1

The purified His6-VP1 of LIPyV (see Figure 10) was subjected to broad-spectrum glycan array screening as described in the Methods. For a full list of glycan probes please refer to Table A2 in the Appendices. As mentioned in the previous paragraph, the derivatization trials of LIPyV VP1 crystals signified no binding of sialylated ligands. Yet, the probe of NeuAcα-(6')-lacto-*N*-neooctaose (NeuAcα-(6')LNnO, chart pos. 614, NeuAcα-6Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH) yielded a signal of 3100 ± 400 fluorescence intensity units. More frequently, signals arose from non-sialylated glycan probes, including those presenting Fucα2-Lacto-*N*-neohexaose (Galβ-4GlcNAcβ-3[Fucα-2Galβ-4GlcNAcβ-6]Galβ, pos. 115 & 116, 8,600 ± 1500 and 25,400 ± 130 units, respectively) or Lacto-*N*-neohexaose (Galβ-4GlcNAcβ-3[Galβ-4GlcNAcβ-6]Galβ, pos. 140, 32,000 units). Another signal was detected for para-lacto-*N*-neohexaose (pLNnH) with the sequence Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glc (pos. 109, 3900 ± 1000 units). These glycans correspond to the carbohydrate groups of neolacto-series glycosphingolipid (GSL) with the formula (Galβ-4GlcNAcβ)$_{n,m}$-3/6Galβ-4Glcβ-Cer. Sample 143 with a branched Galα-3Galβ-4GlcNAcβ-3[Galα-3Galβ-4GlcNAcβ-6]Galβ sequence ("B-like decaosylceramide") yielded a signal of 6300 ± 0 units. Lastly, globotriose (Gb3, Galα-4Galβ-4Glc, pos. 23) and galactotetraose (Gal4, Galα-3Galβ-4Galα-3Gal, pos. 451) yielded signals of 3100 ± 400 and 3200 ± 200 units, respectively. Results refer to the array using 5 fmol of immobilized glycan probe per array spot. Fluorescence intensities were rounded to full hundreds.

### 3.2.2.8 Neolacto-series oligosaccharide derivatization yielded LIPyV VP1 complex structures

Crystals of LIPyV VP1 were grown as described above. For a list of oligosaccharides and their concentrations used in crystal derivatization trials, please refer to Table 17.
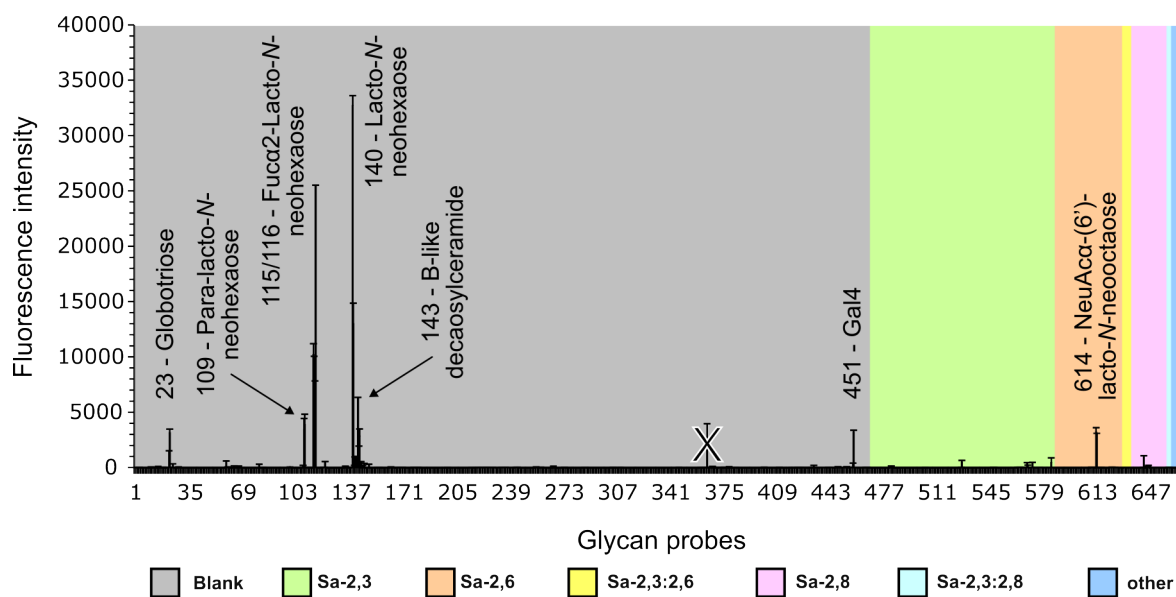
Figure 29: Results of the glycan array screening using LIPyV VP1. Glycan probes are sorted by sialyl linkages as indicated by the color code. Samples with errors >50% are crossed out. Chart kindly provided by the Glycosciences Laboratory, Imperial College London.

Table 17: Oligosaccharides used for the derivatization of LIPyV VP1 crystals.

| | Oligosaccharide | Final concentration | Crystallographic complex formation |
|---|---|---|---|
| Sialylated | 2-*O*-Me-Neu5Ac | 50 mM | × |
| | 3'-sialyl-*N*-acetyllactosamine (3'SLN) | 20 mM | × |
| | 6'-sialyl-*N*-acetyllactosamine (6'SLN) | 20 mM | × |
| | GD3 oligosaccharide | 20 mM | × |
| Neutral | pLNnH | 20 mM | ✓ |
| | LacNAc | 50 mM | ✓ |
| | Lactose | 20% | ✓ |
| | H-antigen triose type 2 | 50 mM | × |
| | Globotriose | 50 mM | × |
| | Galactotetraose | 50 mM | × |

As mentioned in the previous paragraphs, experimental results indicated that LIPyV VP1 does not bind sialyl ligands. Also using the oligosaccharides corresponding to the glycan array hits of globotriose (Gb3), galactotetraose (Gal4), and blood group H antigen triose type 2 for derivatization yielded unliganded crystal structures. Furthermore, branched neolacto-series oligosaccharides, those yielding the highest fluorescence intensities in the glycan array screening, were not commercially available during this work.

In contrast, positive results were obtained from crystals derivatized with 20 mM para-lacto-*N*-neohexaose (pLNnH) or 50 mM *N*-acetyllactosamine (LacNAc) solutions (see Methods). The derivatization with lactose, where the crystal was incubated in a 20% disaccharide solution also serving as the cryo-protectant, yielded complex formation. After data collection, the complex structures were solved using the apo LIPyV VP1 pentamer as the search model. All asymmetric units (ASUs) contained a single pentamer, as is displayed in Figure 30a. For data and refinement statistics, please refer to Table 16. Following simulated-annealing refinement, the Fo-Fc difference electron densities (see Figure 30b) showed binding of pLNnH, LacNAc, and lactose in a novel location designated neolacto-series glycan binding site (see also Figure 31a & b).

### 3.2.2.9 The recessed location of the neolacto-series glycan binding site

The electron density indicating pLNnH binding to LIPyV VP1 occurs in a surface notch towards the "canyon" of VP1, referring to the inter-capsomeric space in the assembled particles. The neolacto-series glycan binding site is exclusively formed by a single VP1 chain (see Figure 31a) at a location more recessed compared to the sialyl binding sites in the VP1s of other polyomaviruses, with a vertical space difference of 5.5 Å (Figure 31b). The accessibility of this recessed location will be addressed in the Discussions chapter.

### 3.2.2.10 Molecular interactions between LIPyV VP1 and neolacto ligands

The intermolecular contacts between LIPyV VP1 and pLNnH were assessed via the atom distances and the orbital geometry. The terminal β-Gal forms hydrogen bonds with the side chain of Asn 71 and backbone atoms of Asn 198 and Gly 230. Its alpha-faced hydrogen atoms form CH-π interactions with the side chain of Tyr 231. The subterminal β-GlcNAc establishes a hydrogen bond between its amide nitrogen atom and the Glu 100 side chain. An electrostatic attraction occurs between its partially negative carbonyl oxygen and the cationic side chain of Lys 199. The third-from-last monosaccharide in pLNnH, another β-Gal, contacts the side chain of Tyr 78 via CH-π interactions. All hydrogen bonds between pLNnH and LIPyV VP1 are listed in Table 18. The remaining monosaccharides do not contact the protein. LacNAc and lactose bind less tightly as signified by the weaker electron densities in Figure 30b and the elevated B-factors in Table 16. Their contacts are those of pLNnH lacking the second β-Gal or, additionally, the *N*-acetyl group.

**a**


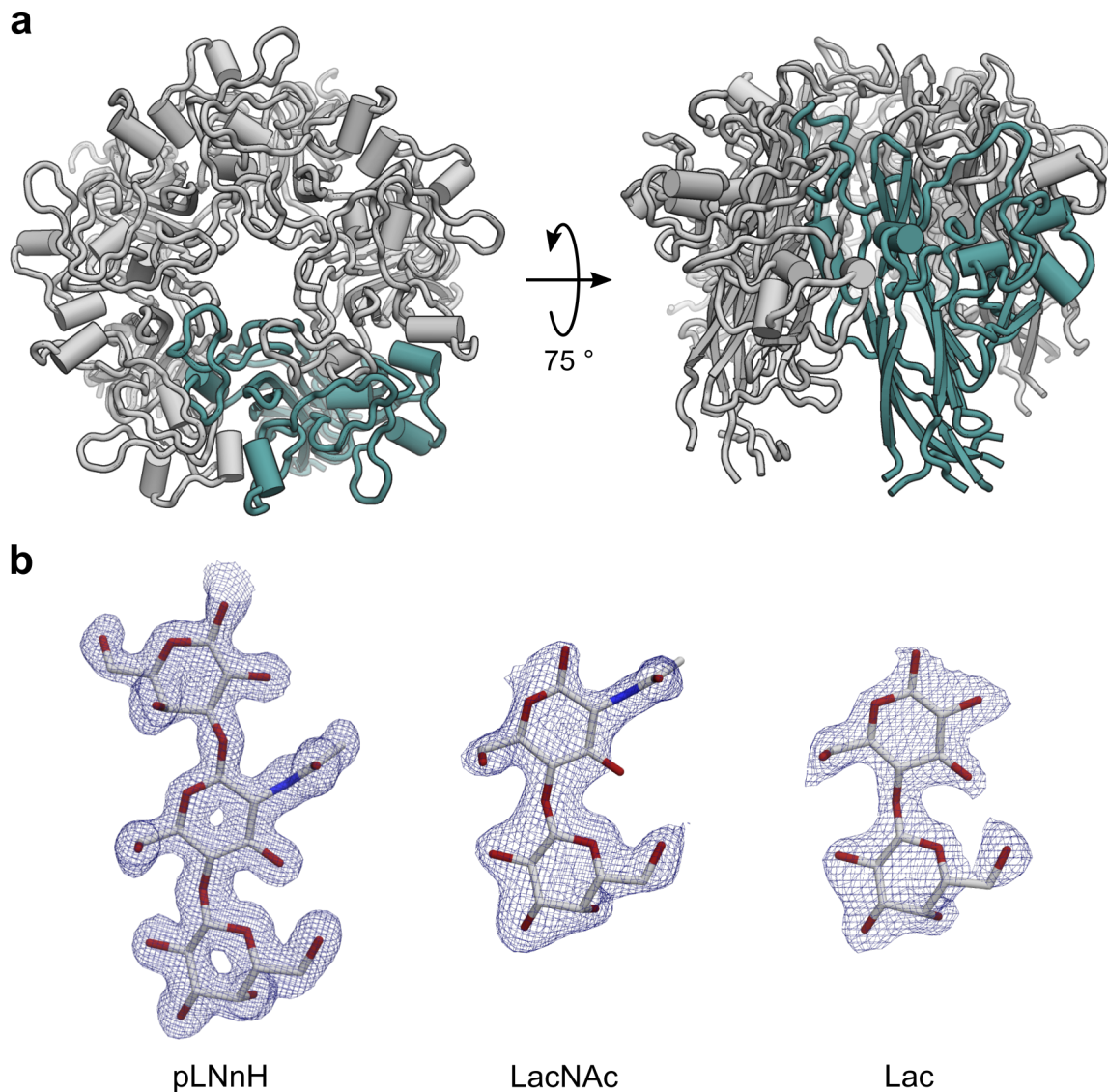
**b**



pLNnH             LacNAc                Lac

Figure 30: Results of the crystallographic LIPyV VP1 complex formation using neutral oligosaccharides. **a)** The overall structure of the LIPyV VP1 pentamer from above (left) and tilted around the lateral axis (right). The protein is shown in cartoon representation with one monomer in light teal. Alpha helices are displayed as cylinders. **b)** All ligands of LIPyV VP1 resolved by X-ray crystallography shown as sticks with the carbon, oxygen, and nitrogen atoms colored in white, red, and blue, respectively. Fo-Fc difference maps are shown as blue meshes around the individual oligosaccharides with a contour level of σ = 2.5 and a radius of 1.6 Å. For the map resolutions, please refer to Table 16.
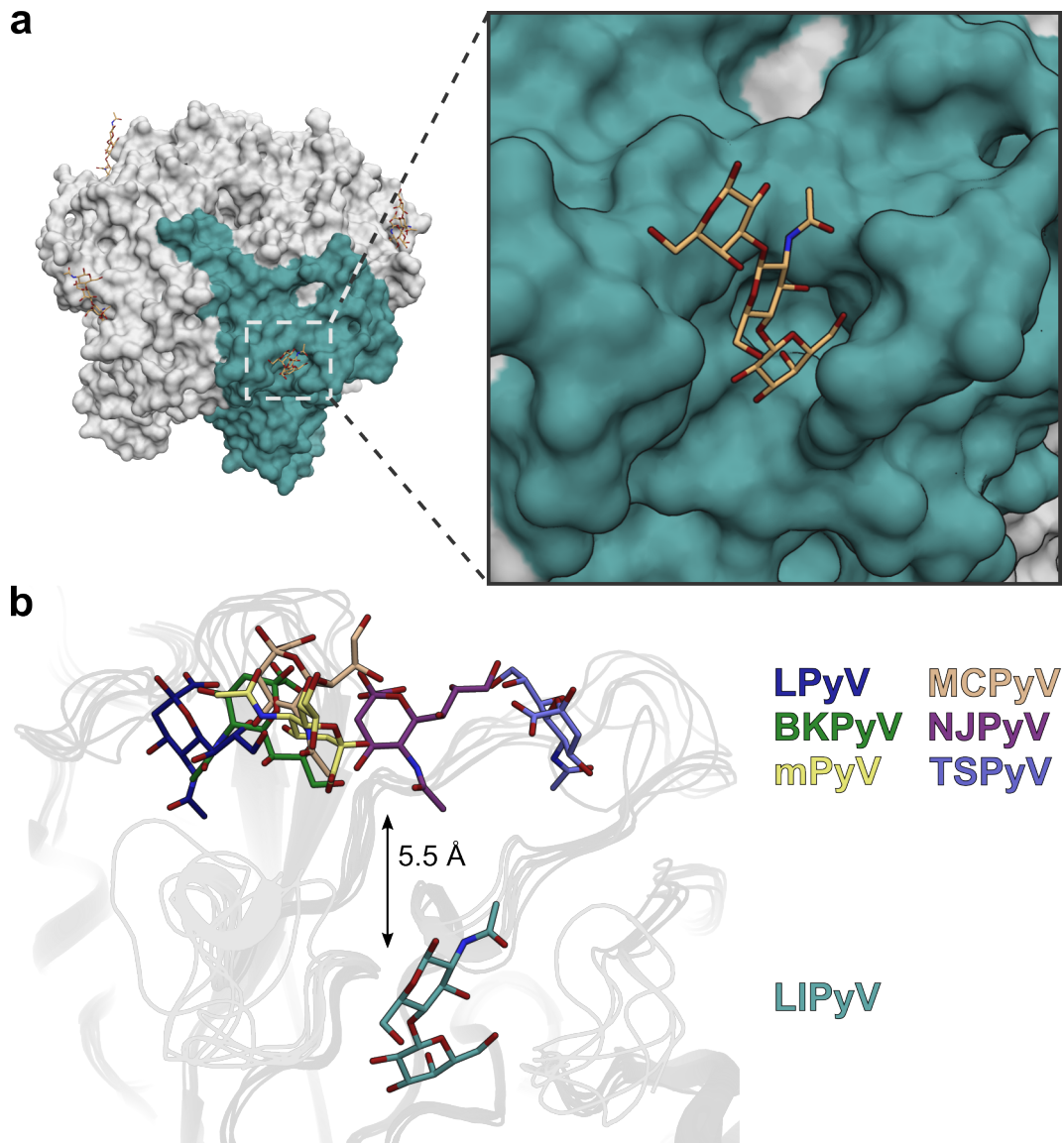
**Figure 31:** The location of the neolacto-series glycan binding site in the VP1 of LIPyV. **a)** The para-lacto-*N*-neohexaose (pLNnH) binds at the outer surface in a recessed location oriented towards the "canyon" of LIPyV VP1. The binding site is solely constituted by a single VP1 monomer, as indicated by the light teal surface coloring. **b)** The recessed location of the neolacto site becomes apparent in comparison with the sialyl sites in the VP1s of B-lymphotropic polyomavirus (LPyV, PDB 4MBZ), BK polyomavirus (BKPyV, PDB 4MJ0), murine polyomavirus 1 (mPyV, PDB 1VPS), Merkel cell polyomavirus (MCPyV, PDB 4FMJ), New Jersey polyomavirus (NJPyV, PDB 6Y5Y), and *Trichodysplasia spinulosa*-associated polyomavirus (TSPyV, PDB 4U60). For the LIPyV VP1-pLNnH complex, the ligands' terminal disaccharide is shown. Cartoon representations of the individual VP1 loops are faded into the background. Oxygen and nitrogen atoms are colored red and blue, respectively.
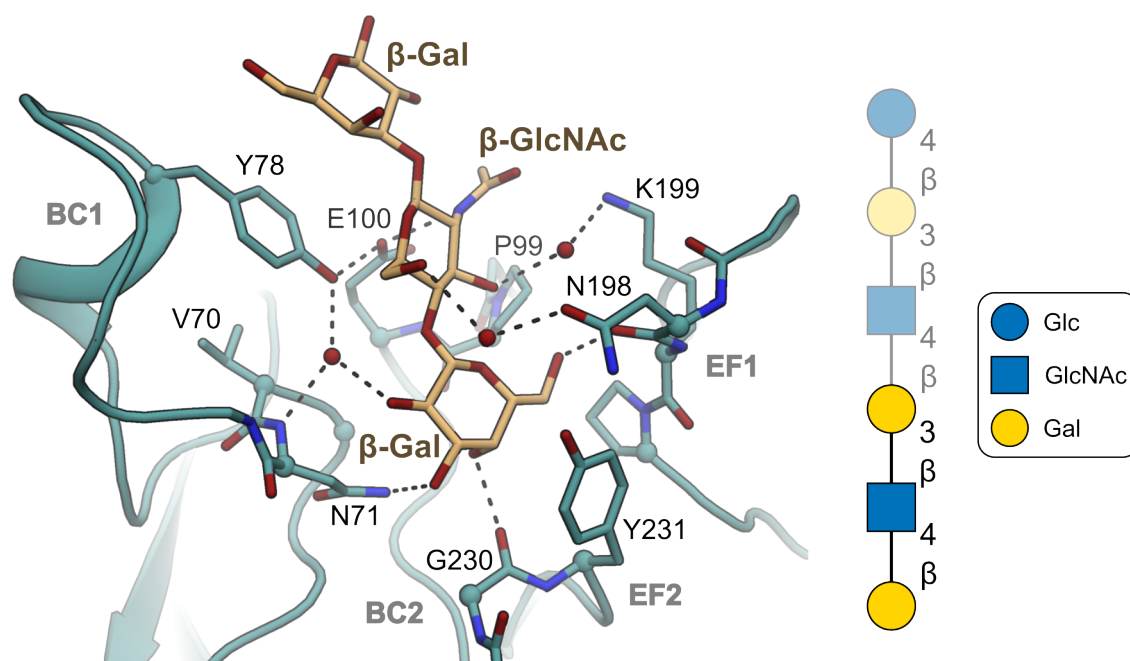
Figure 32: Molecular interactions between LIPyV VP1 and para-lacto-*N*-neohexaose (pLNnH). The protein residues (light teal) involved in binding and the oligosaccharide (light orange) are drawn as sticks. Loops are labeled in bold gray. Waters are displayed as spheres. Hydrogen bonds are indicated as dashed lines. Oxygen and nitrogen atoms are colored red and blue, respectively. On the right, the symbol structure of pLNnH is shown with the Symbol Nomenclature for Glycans legend. Monosaccharides not contacting the protein are shaded gray.

Table 18: Intermolecular contacts between LIPyV VP1 and pLNnH.

| Monosaccharide | | VP1 contacts | |
|---|---|---|---|
| | | Hydrogen-bonds | Distances (Å) |
| β-Gal (term.) | O2 | H2O-*N71* | 3.0-2.8 |
| | | H2O-Y78 | 3.0-3.0 |
| | O3 | N71 | 3.2 |
| | O4 | *G230* | 2.7 |
| | O6 | *N198* | 2.8 |
| | | | |
| β-GlcNAc | N2 | E100 | 2.9 |
| | O3 | H2O-K199 | 2.7-2.9 |
| | O6 | H2O-N198 | 3.0-2.7 |

Residues contributing backbone atoms to contacts are printed in italics.
Distance values are derived from a single binding site and may alter between chains.

### 3.2.2.11 Models of LIPyV VP1 glycan complexes without experimental structures

Glycan binding models of oligosaccharides were prepared to investigate their theoretical interactions with LIPyV VP1. In the first instance, a structure of the branched LNnH comprising six monosaccharides (Galβ-4GlcNAcβ-3[Galβ-4GlcNAcβ-6]Galβ-4GlcNAcβ) was obtained from the PDB entry 6IAL with the title "Porcine *E. coli* enterotoxin B-pentamer in complex with Lacto-*N*-neohexaose". In the enterotoxin complex, the β-1,3 arm of LNnH engages a protein surface depression, while the β-1,6 arm interacts with an exposed arginine side chain [185]. Interestingly, this experimental LNnH oligosaccharide structure aligns well with the pLNnH electron density in its LIPyV VP1 complex. After geometry optimization using Phenix [155], the LNnH model was only slightly shifted compared to the experimental pLNnH conformation (see Figure 33a). The β-1,6-linked arm of LNnH appears to interact with the backbone of Tyr 78 (Y78) and the side chain of arginine 77 (R77, not fully resolved in Figure 33a), which may contribute to the elevated intensities of the branched neolacto-series glycan array signals.

Branched neolacto glycans with terminal 2'-fucosyllactosamine (Fucα-2Galβ-4GlcNAc) moieties gave rise to high signals in the glycan array screening but using the H-antigen type 2 glycan analog for crystal derivatization did not result in complex formation. Thus, a second binding model was built by aligning the 2'-fucosyllactose structure from PDB entry 6HUR with the LIPyV VP1-lactosamine complex, which elucidates the experimental outcome. The model reveals severe overlaps between the terminal fucose and the protein irresolvable through rotation around the glycosidic bond (see Figure 33b).

### 3.2.2.12 The neolacto-series glycan binding site may be conserved in a few polyomaviruses related to LIPyV

Viruses related to LIPyV were identified using its VP1 sequence in BLAST (see Methods) [148], resulting in the puma (*Puma concolor*) and the raccoon (*Procyon lotor*) polyomaviruses as the closest relatives. Other relatives are from the cow (*Bos taurus*), the intermediate roundleaf bat (*Hipposideros larvatus*), the straw-colored fruit bat (*Eidolon helvum*), Daubenton's bat (*Myotis daubentonii*), the Chinese rufous horseshoe bat (*Rhinolophus sinicus*), and the chimpanzee (*Pan troglodytes*). Among LIPyV's relatives, Merkel cell polyomavirus (MCPyV) is the only human polyomavirus, with a VP1 identity of 61%, and the only other structurally characterized polyomavirus (the chimpanzee polyomavirus addressed in previous parts of this thesis belongs to a different virus species). The VP1 sequences of these viruses were subjected to multiple sequence alignment (MSA) to evaluate the conservation of LIPyV VP1's neolacto glycan binding site (see Figure 34a) and to infer their phylogenetic relationship (see Figure 34b).
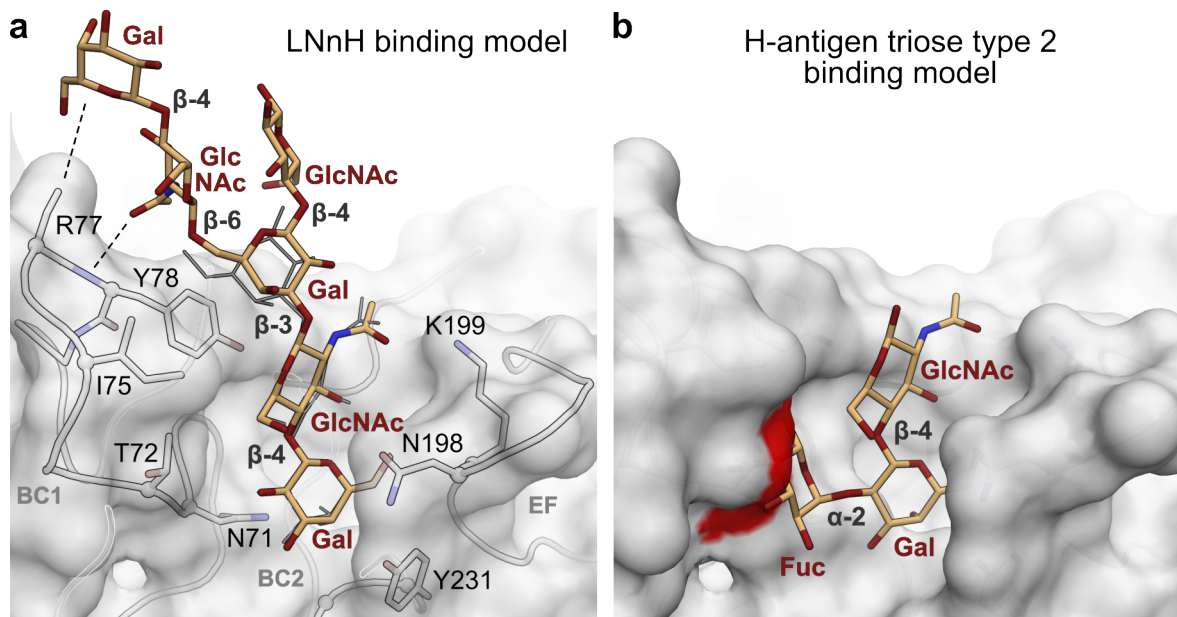
Figure 33: LIPyV VP1 glycan binding models. **a)** Relaxed model of lacto-*N*-neohexaose, derived from PDB entry 6IAL, in association with the neolacto glycan binding site drawn as sticks. Potential interactions not present in the experimental binding conformation of pLNnH (outlined in dark gray) are indicated by dashed lines. Loops are labeled in bold gray. **b)** Model of 2'-fucosyllactosamine, the glycan of H-antigen type 2, associated with the same site. Overlaps between the ligand model and the protein surface are indicated in red. All oxygen and nitrogen atoms are colored dark red and blue, respectively.

The MSA signifies that the neolacto binding site is completely conserved in the puma virus (Pumfec polyomavirus LSF128) VP1 with no apparent steric hindrances (see Figure 34a). The same applies to the raccoon polyomavirus VP1. Despite its N198H variation, the histidine side chain could adopt the water-mediated contact by LIPyV VP1's N198 to pLNnH (see again Figure 32).

Bovine polyomavirus 2 and the bat polyomaviruses have completely conserved binding site residues in the BC1 and BC2 loops but not within the EF1 loop of their VP1s (see Figure 34a). Mutations such as N198G or K199M reduce the number of possible hydrogen bonds, making the engagement of neolacto glycans less likely but not impossible.

In contrast, the sister viruses MCPyV and chimpanzee polyomavirus 1a carry several mutations in their VP1s that make neolacto binding unlikely (see Figure 34a). The P199K substitution may clash with the oligosaccharide, and N198G and K199G remove substantial amounts of hydrogen bonding. Furthermore, Y231N resolves the CH-π interactions with pLNnH's terminal β-Gal and destructs the sandwich-like architecture of the binding site. The latter also applies to the Y78T mutation exclusively found in MCPyV.
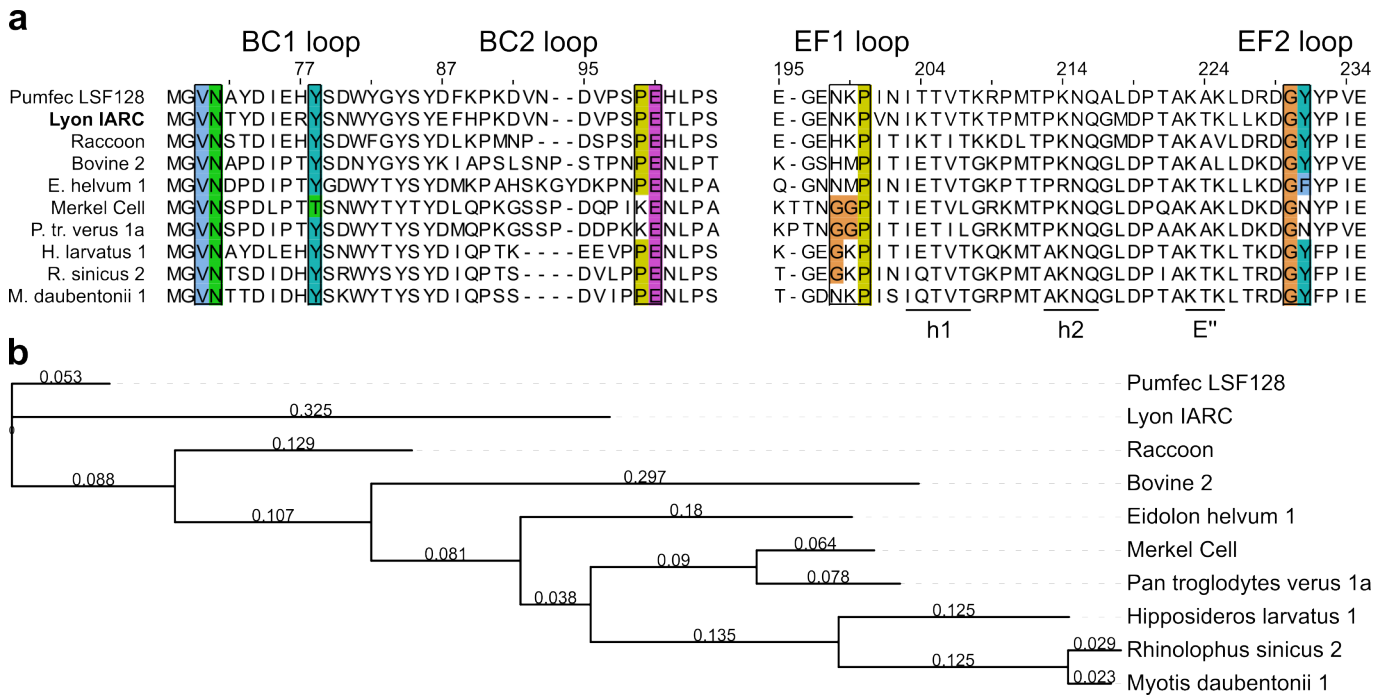
**a**



Figure 34: Conservation of neolacto-series glycan binding in polyomaviruses related to LIPyV. **a)** Multiple sequence alignment of the VP1s of LIPyV (bold) and related species, truncated to the BC and EF loops. Just like the BC loop, the EF loop is divided into an anterior (EF1) and a posterior part (EF2) and encloses a small β-sheet (with strands E', E", and E''') and two short helices (h1 and h2). Strands E' and E''' are located outside the relevant sequences and are not shown in the alignment. Residue numbering corresponds to the LIPyV VP1 sequence. Positions responsible for glycan binding in LIPyV VP1 are highlighted with Clustal coloring. **b)** Unrooted phylogenetic tree based on the same MSA. Evolutionary distances are provided above the individual branches. Virus names[5] are truncated to the hosts for conciseness. The GenBank accession numbers are WAA14057.1 (Pumfec PyV LSF128), YP_009352870.1 (Lyon IARC PyV), YP_009021046.1 (Raccoon PyV), YP_009110709.1 (Bovine PyV 2), YP_007346968.1 (E. helvum PyV 1), YP_009111420.1 (MCPyV), YP_009094189.1 (P. troglodytes[6] verus PyV 1a), BBG61962.1 (H. larvatus PyV 1), BBG62110.2 (R. sinicus PyV 2), and UYR20760.1 (M. daubentonii PyV 1).

### 3.2.3  Sheep polyomavirus

*In this section, I present results regarding ShPyV VP1 that were not part of Rustmeier et al. [2], including the determination of a crystallographic dissociation constant (aided by student Alexander Herrmann), a hemagglutination assay, the structural comparison of the glycan binding sites of ShPyV and BKPyV, and the sequence-based assessment of Forssman antigen engagement in other polyomaviruses.*

#### 3.2.3.1  Forssman pentaose binds to ShPyV VP1 with a micromolar crystallographic dissociation constant

In Rustmeier *et al.* [2], I determined a Kd for the ShPyV VP1-$F_P$ interaction, based on the kinetic binding parameters, by surface plasmon resonance (SPR). Prior to that, we performed crystal derivatization trials using serially diluted $F_P$ solutions to estimate a (thermodynamic) crystallographic dissociation constant ($Kd_c$). All crystal complex structures were solved isomorphously. Detailed data and refinement statistics are minuted in Table A1 on page 139 in the Appendices. Evaluation of the control density integrals, which need to be consistent for a valid estimation (see Methods, section 2.2.6.3), resulted in the exclusion of the 0.3125 mM $F_P$ data set (see Appendices, Figure A1). Subsequently, Michaelis-Menten analysis of the remaining data sets' $F_P$ electron density integrals plotted against the oligosaccharide concentrations yielded a $Kd_c$ of about 44 µM (see Figure 35).

#### 3.2.3.2  ShPyV VP1 agglutinates sheep erythrocytes but not those of bovine or human A-type blood

In Rustmeier *et al.* [2], flow cytometry showed that ShPyV VP1 attaches to sheep but not human erythrocytes. Here, as a supplementary experiment, I assessed the interactions between ShPyV VP1 and sheep blood, bovine blood, or human A-type erythrocytes through a hemagglutination assay. At that, the bovine blood served as a Forssman-negative reference from a species related to sheep [186]. The assay resulted in sheep blood hemagglutination between 0.5 µl/µL and 0.001 µg/µL ShPyV VP1, while none of the other erythrocyte samples were agglutinated (see Figure 36a). Since the lowest ShPyV VP1 concentration in the first iteration (0.001 µg/µL $\equiv$ ~6 nM) yielded positive results, a second sheep blood array was prepared. Therein, the concentration of ShPyV VP1 was diluted to femtomolar concentrations, resulting in positive hemagglutination at $3 \times 10^{-11}$ or 30 picomoles (Figure 36b).
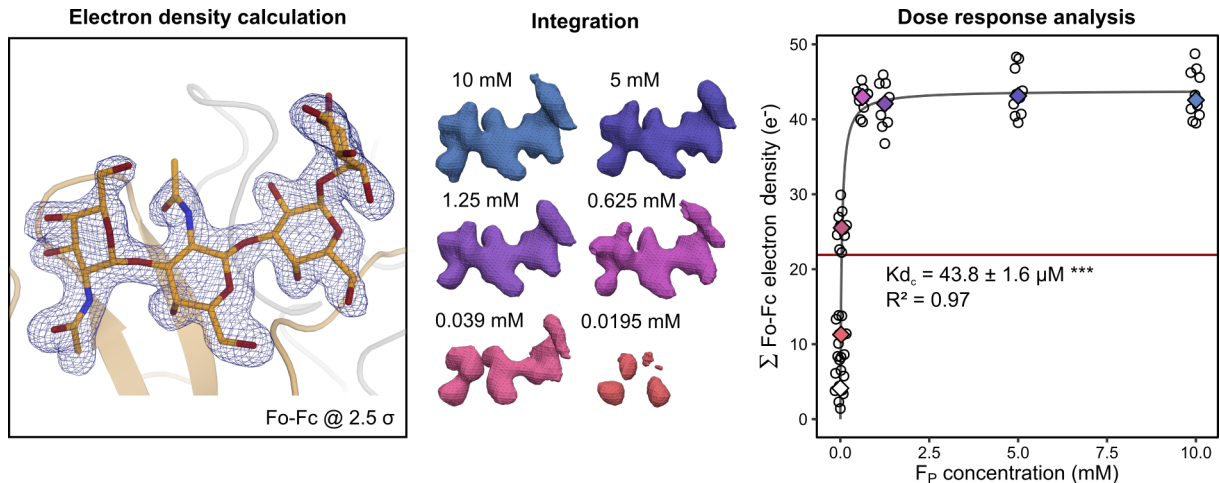
**Figure 35:** Determination of a crystallographic dissociation constant ($Kd_c$) for the ShPyV VP1-$F_P$ complex. Left: The bias-reduced Fo-Fc omit electron density of the Forssman glycan calculated from the 10 mM $F_P$ data set displayed at a contour level of 2.5 σ and with a radius of 1.6 Å around the ligand. The oxygen and nitrogen atoms are colored red and blue, respectively. Middle: Ligand Fo-Fc densities from the dilution series of $F_P$ concentrations in ShPyV VP1 crystals used for integration. Right: Determination of the crystallographic dissociation constant ($Kd_c$) using Michaelis-Menten analysis. Each data set comprises ten integral values. Medians are shown as diamonds colored according to the middle panel with the null value in white. Significance is shown with *** for $p < 0.001$.
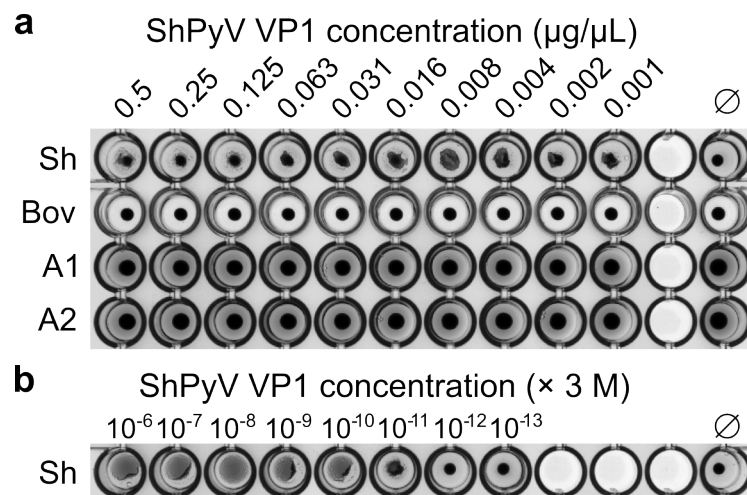


**Figure 36:** Hemagglutination assay using different kinds of mammalian blood and ShPyV VP1 as the agglutinin. **a)** From above: Arrays of sheep blood, bovine blood, and erythrocytes from human type A1 and human type A2 blood. **b)** Array of sheep blood with tenfold dilutions of a 3 M ShPyV VP1 stock solution. Each row features a ∅ sample using buffer instead of protein functions as a negative control.

### 3.2.3.3 Structural analysis reveals a conserved glycan engagement strategy in betapolyomaviruses.

To evaluate the relationship of the ShPyV VP1-$F_P$ binding site to the glycan sites of other polyomaviruses, structural alignments of the respective ligand complexes were prepared. The left panel of Figure 37a shows the six modes for Neu5Ac engagement in polyomavirus VP1s (see also Introduction, Figure 5). A comparison with the ternary complex of ShPyV VP1, Forssman pentaose ($F_P$), and a sialyl moiety (see Figure 37a, right panel) reveals that the $F_P$ molecule covers a surface area that corresponds to five of the sialyl sites. Only the TSPyV site, which is also functional in ShPyV VP1, remains free.

The structural comparison also reveals that the methyl groups in the terminal monosaccharides of the ShPyV VP1-$F_P$ and BKPyV VP1-GD3 complexes spatially align (see Figure 37b), mediated by a shared binding site architecture. The hydrophobic depression for the methyl group engagement is conserved in the two viruses, and sequentially aligning residues confer specificity towards the glycan binding motifs (see Figure 38). Collectively, this topological conservation indicates a common binding site origin, a topic further addressed in the Discussions.

### 3.2.3.4 VP1 sequence analyses indicate Forssman binding in several betapolyomaviruses related to ShPyV

To explore the presence of Forssman binding in the polyomaviruses from other Forssman-positive hosts, such as mice, goats, horses, and dogs, VP1 sequence analyses were performed. These show that within this selection of viruses, the residues required for antigen recognition are only conserved in the murine polyomavirus 2 (mPyV2, see Figure 39a). Beyond, I performed a BLAST run (see Methods) using the ShPyV VP1 BC2 loop, which mainly contributes to Forssman binding, as the query sequence. This search identified several betapolyomaviruses that share a common ancestor with ShPyV and feature a conserved Forssman antigen (FA) binding site, including mPyV2 and viruses from bats, multimammate mouse, lion, and red panda (see Figure 39b). No hits occurred for PyVs of the alpha, gamma, and delta genera, suggesting that Forssman binding exclusively occurs in ShPyV and potentially a few related betapolyomaviruses. The implications of these findings, including an evaluation of these animals' Forssman status, will be discussed in the next chapter.
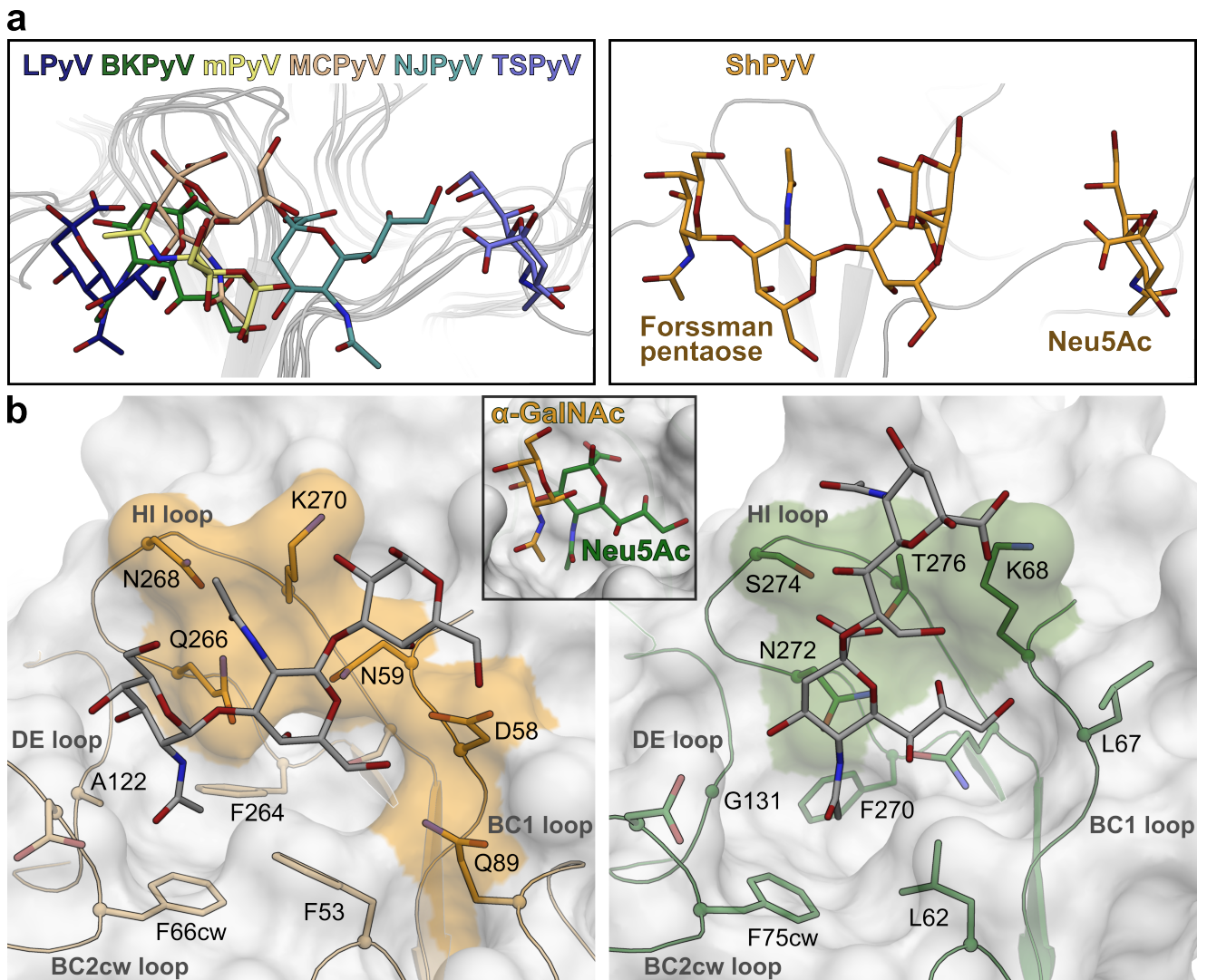
Figure 37: Relationship between the binding sites of Forssman pentaose and sialyl glycans. **a)** Left: The aligned sialic acid moiety binding modes are shown in superposition for the VP1s of B-lymphotropic polyomavirus (LPyV, PDB 4MBZ), BK polyomavirus (BKPyV, PDB 4MJ0), murine polyomavirus 1 (mPyV, PDB 1VPS), Merkel cell polyomavirus (MCPyV, PDB 4FMJ), New Jersey polyomavirus (NJPyV, PDB 6Y5Y), and *Trichodysplasia spinulosa*-associated polyomavirus (TSPyV, PDB 4U60). The individual protein loops are faded into the background. The sialic acids are colored according to the VP1 identifiers. Right: In the same orientation, ShPyV VP1 in complex with Forssman pentaose ($F_P$) and the Neu5Ac of 3'SLN (PDB 7B6V). **b)** The ShPyV VP1-$F_P$ and BKPyV VP1-GD3 glycan complex structures in comparison. The residues establishing polar contacts towards $F_P$ and the Neu5Ac-Neu5Ac motif feature an orange and green surface color, respectively. The small panel shows the superposition of $F_P$'s α-GalNAc and the terminal Neu5Ac moiety aligned to the BKPyV VP1 surface (from PDB 4MJ0). In all images, oxygen and nitrogen atoms are colored red and blue, respectively.
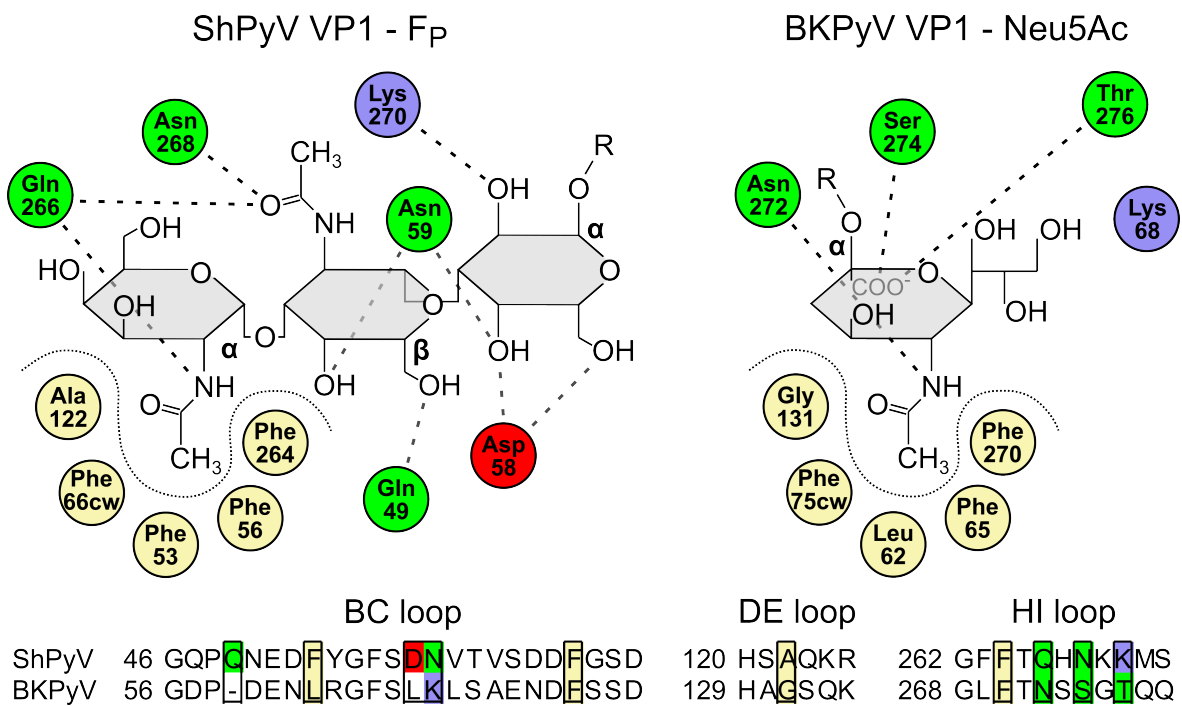
Figure 38: 2D representations of glycan engagement in the VP1s of ShPyV and BKPyV. In the diagrams, amino acids contributing to ligand binding are symbolized as circles. Similar circle positions signify the same residue positioning within the VP1 sequences. Circle colors correspond to residues hydrophobicity (light yellow), polarity (green), positive charge (blue), and negative charge (red). In the pairwise sequence alignment, binding site residues are colored accordingly. Polar and apolar interactions are indicated by thick and finer dashes, respectively. Glutamine 49 aligns with a gap in the BKPyV VP1 sequence and has no equivalent in the right panel. Not shown on the right is leucine 67 of BKPyV VP1, which doesn't interact with the glycan. Also, the subterminal sialic acid moiety interacting with lysine 68 is omitted for clarity.

**a** VP1s of PyVs from selected Forssman-positive hosts

```
ShPyV    21  GGIEVLAVRTGPDSITEIEAYLNPRMGCPQNEDFYGFSDNVTVSDDFGSDAFPWKCFPCYSTARISLPMLNQDMTCDTILMWEAISCRTEVMGVNMLTNV
mPyV2    27  ...Q..D.........Q...F......C.VDS.F.F.DNITVSA.YTQDMPRI.EL....M.K.......E..............K...V..SS...C
CaegPyV  27  .....GL.....T.T..LF.....CSTESEY..F.DNQRGATSRTD.DLISAEL.R..LGVVQ..L..EKL.D.VL.....V..K...V..T..TC
EcabPyV  20  .....DLKS.E...T..CF....V.HATG.A.Y.DR.TVSENFDSTBNPGENEI....C..VP.....D.I..NK.....V.VK...V..SS..YC
CfamPyV  18  ..VS..GL......TIT........BD ATKGSI.SEITVTGTN S.BNPIKSCM....C...Q..LV.E...ANE.V....V.VKS..I..GS...A

ShPyV   121  HSAQKRVYENDREGTGIGVEGMGYHMFAIGGEPLELQFMVFNHRATYPAEATVIKNPGASS      QVFDPNLKGTLTADGVFPVEAWGPDPFKNENTRYFGQY
mPyV2   127  H.AV.RL.D.  .A.FPIQ.LNF.F.SV....A.D..MV.E.Y.CK...GVAALQAAPKAA    ..L..K..AK......K..I...S...A.........T
CaegPyV 127  FGYK.RYSPSAGQ.SAMPI..IN..F.V......I..ICEDFK.P.HPTE...PEKNKITNKS..L..S...I..IDK...Y.I.C.C..S........T
EcabPyV 120  H.LGRRLNDD     ..I.LPI..INF....V......ALLT.G.T..HEKFI.P.PAS.GN    ..LN.HA...R.DK..LY...V.SA..AR...S....S
CfamPyV 116  H.SL.RTNNT   T.A.YVT..PNF.F.SV........AL.I..QMHL.EGV.    RPQ.TINVDA..LNQ....Q.DR..TY...L.HA..S........S

ShPyV   219  YTGGTQTPPVLTFTNTQTTILLDENGVGPLCKGDGLFLSCADIVGFTCHNKKMSFRGLPRYFRVTLRKRVVKNPYPVSHLLNTLFSNMQPTVRGQPMQGE
mPyV2   223  ....L.....QI...T.....N.........R.CQNN..H.......SI....L.......S......E.M.QID..K.E..
CaegPyV 228  ....VS.....Q...V...........A.K.YITA...C.LLCPNDQQQ.......IS....L......IASI.TS..T.SL.P.QS.E.E
EcabPyV 215  L....AS.....Q....L......I......E.....S......L.NRSGT.QW......Q......W.......QS..TS..HST..NLQ....T.D
CfamPyV 213  ....SN.....QW...L..V...................V..C.YC.NQSGAQFY.......N.S...........S.ISS....LN.QIT....V.A

ShPyV   320  DAQVEEVRVYQGVEGLPGDPDMVRFRDQFGQTNTV    VGN
mPyV2   324  ...I.D...R.........I.Y..K...EV..PDIPSNEA
CaegPyV 327  K......I............YVNK...EE.C    IPK
EcabPyV 316  .G........E.M..V.....LA.TLNRY..NI.E    LPQE
CfamPyV 314  NN.I......S.K....A...LH..MGPL.E.R.Q    LPGTTTSVSS
```
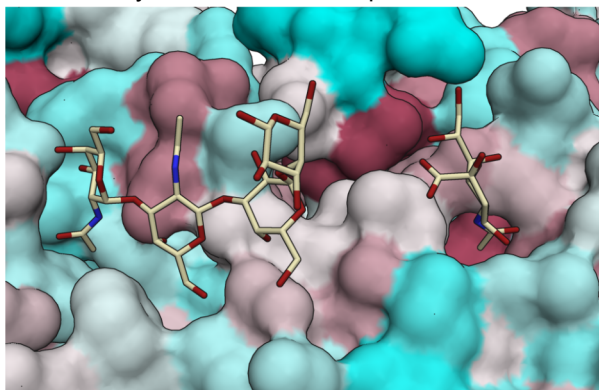
**b** VP1s of PyVs related to ShPyV

```
ShPyV    21  GGIEVLAVRTGPDSITEIEAYLNPRMGCPQNEDFYGFSDNVTVSDDFGSDAFPWKCFPCYSTARISLPMLNQDMTCDTILMWEAISCRTEVMGVNMLTNV
MyoPyV   20  ......S.........T.........C.PES.F..F.DNITVSQSEADQPQI.EI....M.K.N..L..E.I.........VK...V.ISS.C..
MnatPyV  25  ......S.........Q...F......CGPDT.Y..F.DNITVSQ.YTNDRPGI.EL......K.D...E.I.........VK...V.I.S....
mPyV2    27  ...Q..D.........Q...F......C.VDS.F..F.DNITVSA.YTQDMPRI.EL....M.K.......E..............K...V..SS...C
AfulPyV  20  ......S.........Q......C.TDS.F..Y.DNITVSG.VH.DTPRL.EL.Y..M.K.N....E.I..NE.F....V.VK...V.ISS....
PleoPyV  21  ......S.......V.Q......C.D.S.Y..Y.DNITVSQ..SADNPGKAEL....M.K.N......E.I..N.....V.VK...V.ISA....

ShPyV   121  HSAQKRVYEN DREGTGIGVEGMGYHMFAIGGEPLELQFMVFNHRATYPAEATVIKNPGASSQVFDPNLKGTLTADGVFPVEAWGPDPFKNENTRYFGQY
MyoPyV  120  H.AVHREF.    .A.FP.Q.LNF.F..V........MV.D..KCR...GLAALQQAPKTA..L..Q..AN.IK..T...V.C...A...S..Y.S.
MnatPyV 125  H.AV.RE...    .A.FPIQ.LNF.F..V...A....LL.D.Y.C...EGTI.T..LPKTA..L..Q...K.L..K..I.V.S...S........S.
mPyV2   127  H.AV.RL.D.    .A.FPIQ.LNF.F.SV...A.D..MV.E..Y.CK...GVAALQAAPKAA..L..K..AK......K..I...S...A........T.
AfulPyV 120  H.AVTRE.G.TNTD.A.FPIQ.NF.F.SV....DI.ML.G.YL.N..D..N.FRS.PKNA..LN.S..TK.LR..TY......CA..A........S.
PleoPyV 121  H.AVLREFDQ    .A.VPIQ.LNF.........D..MV.N..YKTK..SPLAALQAAPKAA..L..S..AK.LK...IY.I.V.C...S....S....S.

ShPyV   220  TGGTQTPPVLTFTNTQTTILLDENGVGPLCKGDGLFLSCADIVGFTCHNKKMSFRGLPRYFRVTLRKRVVKNPYPVSHLLNTLFSNMQPTVRGQPMQGE
MyoPyV  217  ...LT.....Q...V.....N...........C...R.CDN..Q.......N......L.......S........D.M.K.Q..V.E.D
MnatPyV 222  ...LT....MQ...V.....N...........Y..A...C.RCSNN..Q.......N......L.......S......E.M.KMQ..D.E..
mPyV2   224  ...L......QI...T.....N..........Y.AS.......R.CQNN..H......SI....L.......S......E.M.QID..K.E..
AfulPyV 220  ...LT.....Q...V.....N...........C...AS.......R.CQNHN.H.......N......L.......S......K...RIE..N.E.S
PleoPyV 218  ...LT.....Q...V.....N...........Y.AS....C..R.CST...Q.......S......L.......S......Q...EIA..D.E.K

ShPyV   320  DAQVEEVRVYQGVEGLPGDPDMVRFRDQFGQTNTV    V    GN
MyoPyV  317  K...K.....E...R........I.Y..KY..EK.K  I  PAVA.
MnatPyV 322  G........E...R........YINKY..EV.Q    IP VA.
mPyV2   324  .........I.D...R........I.Y..K...EV..PDI PSNEA
AfulPyV 320  G........D...P......VI.Y.N.L..EEI.K    I PSND
PleoPyV 318  .........E...D........I.Y.N.Y..EEV.K    IP H.
```

**c**

Glycan binding conservation in the VP1s of PyVs from Forssman-postive hosts

Glycan binding conservation in the VP1s of PyVs related to ShPyV
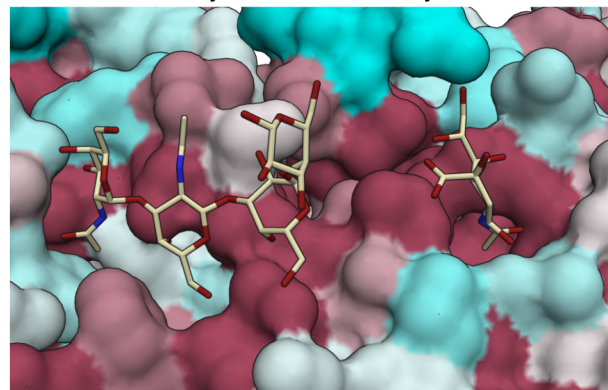


nonconserved    conserved

Figure 39:
*Caption on next page*

Figure 39: *Caption*

Conservation of Forssman glycan engagement in **a)** PyVs from Forssman-positive hosts and **b)** PyVs related to sheep polyomavirus. The MSAs were compiled with ShPyV VP1 (NCBI: YP_009134726.1) as the reference, highlighted in bold. PyV species in the upper MSA derive from house mouse (*Mus musculus*, mPyV2, ALN69893.1), wild goat (*Capra aegagrus*, CaegPyV, YP_010087535.1), horse (*Equus caballus*, EcabPyV, YP_006383691.1), and dog (*Canis familiaris*, CfamPyV, YP_009362021.1). The lower MSA comprises the VP1s of ShPyV, mPyV2, and PyVs from a bat species (*Myotis lucifugus*, MyoPyV, YP_002261488.1), multimammate mouse (*Mastomys natalensis*, MnatPyV, YP_009111357.1), red panda (*Ailurus fulgens*, AfulPyV, AOW44158.1), and lion (*Panthera leo*, PleoPyV, YP_009553611.1). Residue positions relevant for glycan binding are colored according to their conservation score from low (teal) to high (burgundy) as calculated in JalView 2 [187]. **c)** Surface representations of ShPyV VP1 liganded with Forssman pentaose and a sialic acid moiety (PDB 7B6V) with residues coloring according to the MSA conservation scores. Structure images rendered in ChimeraX [158].

# Chapter 4:  Discussions

### 4.0.1  Sea otter and avian polyomaviruses

This work showed that the VP1 of sea otter polyomavirus (SOPyV) and its homolog from BKPyV share a conserved sialyl binding site. SOPyV VP1 also features highly conserved residues of the TSPyV sialyl binding site, initially considered as another potential receptor attachment point. However, my SOPyV VP1-GD3 oligosaccharide crystal structure has revealed that a lysine residue (K63) in this putatively second binding site causes steric hindrance for Neu5Ac engagement. Still, the high conservation of the remaining TSPyV site in SOPyV VP1 is unlikely a product of chance. It is more probable that the ancestor of SOPyV gained the ability to engage sialic acid moieties at the BKPyV sialyl site and, in turn, lost the selection pressure to maintain the TSPyV site, allowing for the introduction of sterically unfavorable residues such as K63. Yet, the significant conservation of the TSPyV site in the SOPyV VP1 implies a recent abandonment, in contrast to related betapolyomaviruses, where the TSPyV sialyl binding site is less conserved (see Figure 40b). These considerations also imply that the sialyl binding site commonly utilized by several members of the alpha-, beta-, and gammapolyomaviruses is phylogenetically older than the sialyl site of BKPyV and its relatives.



**a**

| | | BC2 loop | | | | DE loop | |
|---|---|---|---|---|---|---|---|
| SOPyV | 55 | GFSNQV | T LK | EDFGD | D TP | NKE E LP | 119 V H SGGQ R V - GT - |
| APyV | 51 | GYS TV I | T VQ | AEGYQ | D AP | HS T E VP | 114 M H SGAF R AFNG- |
| TSPyV | 68 | GYSEKV | T VA | NSSDQ | D KP | TSGE I P | 131 V H MATK R MYDD- |
| GHPyV | 52 | GFSQAV | T VA | TSLNP | D VP | PKA E LP | 119 V H SAS I R MNGG- |
| ShPyV | 56 | GFSDNV | T VS | DDFGS | D AP | PWK Q FP | 115 V H SAQK R VYEND |

**b**

| | | BC2 loop | | | | DE loop | |
|---|---|---|---|---|---|---|---|
| SOPyV | 55 | GFSNQV | T L | KEDFGD | D TP | NKE E LP | 119 V H SGGQ R VGTSG |
| JCPyV | 57 | GFSKS I | S I | SDTFES | D SP | NRD M LP | 121 V H SNGQ A THDNG |
| BKPyV | 65 | GFSLKL | S A | ENDFSS | D SP | ERK M LP | 129 L H AGSQ K VHEHG |
| SV40 | 67 | GLSKSL | A A | EKQFTD | D SP | DKE Q LP | 131 L H SGTQ K THENG |

Figure 40: Conservation of the TSPyV VP1 sialyl binding site in the VP1s of **a)** SOPyV, APyV, TSPyV, GHPyV, and ShPyV and **b)** SOPyV, JCPyV, BKPyV, and SV40. The residue position responsible for sialyl glycan engagement in TSPyV VP1 are highlighted in boxes with Clustal coloring.

To better understand the consequences of the SOPyV VP1 K63 residue for sialyl glycan engagement, we sought to structurally characterize APyV VP1, which carries a similarly sized[7] glutamine residue (Q59) at the same residue position (see Figure 40a). As a gammapolyomavirus, such as finch polyomavirus (FiPyV) and goose hemorrhagic polyomavirus (GHPyV), APyV VP1 was expected to engage sialyl glycans at the conserved

---

[7]Both lysine and glutamine feature five non-hydrogen atoms in their side chains.

TSPyV site, which was later (theoretically) confirmed by AlphaFold's structure prediction. The folding model showed that the Q59 side chain, in contrast to K63 in SOPyV VP1, folds inwards and should not inhibit the engagement of sialic acid moieties. However, this structure model may not reflect the physiological fold of APyV VP1 within the assembled pentamer, nor does it provide information on the specific interactions with potential sialyl glycan ligands. Due to its tendency to aggregate, which may be caused by exposed hydrophobic residues (see again Figure 17c), gathering experimental data on the interactions between APyV VP1 and potential sialyl glycan ligands proved challenging. Protein aggregates are amorphous and unlikely to crystallize, which hindered analyses by X-ray crystallography. Other techniques, such as microscale thermophoresis or STD NMR, require stably monodisperse protein, which for APyV VP1 could not be achieved. To overcome this problem, substituting the exposed hydrophobic residues with polar amino acids using site-directed mutagenesis may increase solubility and enable further experimental procedures.

## 4.0.2 Chimpanzee polyomavirus

### 4.0.2.1 The roles of galactosyl and sialyl moieties for receptor engagement

In the Results, I described a novel galactosyl binding site in ChPyV VP1 located approximately 15 Å beside the sialyl glycan binding site (see again Figure 23). I showed that this galactosyl site engages both α- and β-linked galactoses and thus contrasts the rather strict specificities of the glycan binding sites in other PyV VP1s, which usually only accept one configuration of a terminal monosaccharide. One of the characterized ligands, lacto-*N*-tetraose (LNT, Galβ-3GlcNAcβ-3Galβ-4Glc), is similar in sequence to the glycans of gangliosides GM1 (Galβ-3GalNAcβ-4[Neu5Acα-3]Galβ-4Glc) and GD1b (Galβ-3GalNAcβ-4[Neu5Acα-8 Neu5Acα-3]Galβ-4Glc), both of which carry terminal galactose on one of their branches. Thus, it is plausible that the galactosyl binding site contributes to the engagement of ganglioside receptors. Furthermore, LNT is an integral part of the Lc4Cer glycosphingolipid (GSL) and its Lewis A antigen derivate, which may also function as cellular anchor points. Lc4Cer is also the precursor of the ABO and Lewis b GSLs, but these antigens carry fucosylated galactoses that probably do not fit into the binding pocket. The other characterized galactosyl ligand of ChPyV VP1, globotriose (Gb3, Galα-4Galβ-4Glc), corresponds to the glycan of the abundant Gb3Cer glycolipid. However, Gb3Cer is probably too short to function as a direct attachment receptor. Ultimately, to test if the recognition of galactosyl carbohydrates is sufficient for infection, *in vitro* studies using cells depleted of membrane sialylation will be required.

### 4.0.2.2 Differences in the glycan engagement of ChPyV and NJPyV and their implications for host tropism

As discussed in the previous paragraph, the galactosyl binding site of ChPyV VP1 may contribute to cellular attachment and thus influence tropism and viral host range. However, the NJPyV VP1, with a sequence identity of 83% to ChPyV VP1, gave similar glycan array screening results (see again Figure 19). From a structural point of view, it is disputable whether these viruses engage different receptors at all. Therefore, ChPyV and NJPyV contrast other human-simian pairs such as HPyV9 and LPyV or BKPyV and SV40, where subtle variations in the sialyl binding site residues of one member facilitate the preferred engagement of Neu5Gc over Neu5Ac [188]. Because the VP1s of ChPyV and NJPyV seem to function identically, their T antigens must be considered host range factors. Here, the sequence differences are more prominent (78% identity), with the large T antigens (LTAgs) of NJPyV (GenBank YP_009030021.1, 711 residues) being considerably longer than the ChPyV homolog (YP_004046683.1, 615 residues). It is conceivable that both viruses can enter the same cells due to the engagement of similar receptors, where only the virus with the appropriate LTAg, interacting with the hosts' transcription and replication machinery, establishes productive infection. Viral reproduction studies using simian and human cells with determined membrane glycan profiles may clarify the question of host tropism.

## 4.0.3 Lyon IARC polyomavirus

### 4.0.3.1 Implications of LIPyV VP1 glycan engagement

In this work, I described the interactions between the linear neolacto-series oligosaccharide para-lacto-*N*-neohexaose (pLNnH) and LIPyV VP1 in a novel neolacto glycan binding site. I showed that the minimal binding motif appears to be the terminal lactosamine (Galβ-4GlcNAcβ) disaccharide, as free lactosamine yielded decent, albeit weaker, ligand binding, whereas lactose showed only faint association with the protein.

Beyond the linear pLNnH, I, by theoretical models, propose that the elevated signals of branched neolacto-series glycans, as seen in the glycan array screening, are mediated by additional contacts to the surface of LIPyV VP1. Likewise, I show that the neolacto-series binding site of LIPyV VP1 does not accept fucosylated lactosamine moieties on the β-1,3-linked arm. These models significantly complement the experimentally acquired complex structures for understanding the glycan engagement strategy of LIPyV VP1.

Pursuing the crystallographic complex formation of the glycan array hits corresponding

to globotriose and galactotetraose (Gal4), which display terminal α-Gal instead of β-Gal, did not result in ligand binding. Although large portions of the LIPyV VP1 surface are accessible for engagement, a systemic problem cannot be ruled out in this context. To fully confirm that complex formation with these oligosaccharides does not occur, solution-based methods, such as STD NMR, should be employed.

In contrast to ChPyV and ShPyV, which feature active sialyl binding sites beyond those of neutral glycans, LIPyV VP1 seems to not bind to sialylated ligands (the glycan array hit for NeuAcα-(6')-lacto-*N*-neooctaose is based on the neolacto-series core). Consequently, the attachment of LIPyV VP1 to its host cells is potentially solely mediated by the recognition of terminal lactosamine groups present in membrane glycolipids or glycoproteins. The neolacto glycan binding site of LIPyV VP1 is recessed toward the geometrical center of the viral particle. Yet, (neo)lacto-series glycans carry multiple repeats of their disaccharide building blocks, likely enabling viral engagement. However, the physiological role of neolacto-series glycans, which occur in the glycolipids and milk of many mammalian species, including humans, as receptors for LIPyV remains unknown. To obtain a clearer picture, *in vitro* studies examining the role of neolacto glycans during LIPyV infection will be necessary. Therefore, (sialyl glycan-depleted) cell lines with appropriate membrane constitutions will be required.

### 4.0.3.2 Phylogenetics and host tropism

Lyon IARC polyomavirus (LIPyV), initially described as a human polyomavirus, was later found in the feces of domestic cats several times. By evaluating the binding site conservation in viruses related to LIPyV, I showed that a puma and a raccoon polyomavirus may also use neolacto-series glycan receptors. The cat-associated LIPyV and its two closest relatives infect species of the Carnivora order, including the Feliformia suborder (puma and cat), and species native to the New World (puma and raccoon). The phylogenetic relationships of these viruses and their hosts, which also share a common ancestor, may reflect co-evolution [76]. I also showed that nested within the LIPyV-related viruses is Merkel cell polyomavirus (MCPyV), which, by sequence, does not engage neolacto glycans and instead uses glycosaminoglycan (GAG) and sialylated receptors for cell attachment. During evolution, MCPyV and its chimpanzee sister virus may have developed the ability to engage these differing infection receptors, reducing the selection pressure to maintain a neolacto site. The alternative acquisition of the MCPyV VP1 sialyl glycan binding site by recombination with the capsid protein of another (unknown) polyomavirus is less likely, as this site remains unique to MCPyV and, by sequence, its sister virus.

## 4.0.4 Sheep polyomavirus

### 4.0.4.1 The ancient origin of the Forssman antigen binding site

In the Results on ShPyV VP1, I showed that its Forssman glycan binding site shares a glycan engagement strategy with the sialyl site of BKPyV VP1. In addition to the betapolyomaviruses ShPyV and BKPyV, the alphapolyomaviruses B-lymphotropic polyomavirus (LPyV) and human polyomavirus 9 also use a similar but more recessed binding site for sialyl glycans in their VP1 proteins [96, 189]. Their shared binding site architecture may have originated in an ancient VP1 that diverged to recognize different monosaccharides, including the sialic acid in gangliosides or the α-GalNAc in the Forssman antigen. However, the original recognition motif, whether sialic acid, GalNAc, or another monosaccharide, remains unknown.

### 4.0.4.2 The role of Forssman antigen in betapolyomavirus evolution and host tropism

The sheep polyomavirus was initially identified as a genetic contamination in processed lamb meat, and it so far remains unknown if the virus propagates in the animal [76]. It is related to a group of monophyletic betapolyomaviruses with a few members featuring a conserved Forssman binding region. It is conceivable that their last common viral ancestor utilized FA to infect its host before diverging into the recent betapolyomavirus species. During viral evolution, the heterophile distribution of the FA in animals may have facilitated host switching, which is especially plausible for pervasive host species such as bats and rodents. For example, the house mouse is a Forssman-positive animal, and the pneumotropic polyomavirus of mice (*Betapolyomavirus secumuris*, mPyV2), a close relative of ShPyV, is likely able to engage Forssman antigen (see again Figure 37).

Host switching mediated by FA may also explain why these monophyletic betapolyomaviruses comprise members found in phylogenetically distinct hosts such as lion and red panda [77]. However, to my knowledge, it is unknown if bats, lions, and red pandas express FA on their cell membranes. Based on the available knowledge on the genetic determinants for Forssman antigenicity [186], lions should express an active Forssman synthetase (see GenBank XM_042913499.1). In contrast, the genomes of red panda (*Ailurus fulgens*) and the bat *Myotis lucifugus* do not possess a Forssman synthetase (explored using the Forssman synthetase *GBGT1* gene of the house mouse, not shown), suggesting that these species lack Forssman antigen, assuming their genomic information is complete.

### 4.0.4.3 The utility of ShPyV VP1 for Forssman antigenicity typing

In Rustmeier *et al.* [2], flow cytometry experiments demonstrated that fluorescently labeled ShPyV VP1 can detect Forssman antigen on cells, similar to the recognition by an antibody or lectin. A comparable result was achieved through a hemagglutination assay presented in this thesis, showing that ShPyV VP1 only agglutinates Forssman-positive erythrocytes. These results also suggested that the protein's recognition of sialyl glycans is negligible in a cellular context. As shown by the crystal structure, ShPyV VP1 recognizes the GalNAc-GalNAc (Forssman) disaccharide. For this specific recognition motif, no distinct lectin is known, and deploying anti-Forssman IgM antibodies requires a high product consumption. Under these considerations, labeled ShPyV VP1 may prove helpful for typing the Forssman antigenicity in tissue samples. It may help clarify the distribution of Forssman antigen in animals and its occurrence in malignancies of humans.

# Chapter 5: Conclusions

Polyomaviruses (PyVs) infect a wide range of different animals. While the dsDNA viruses were initially discovered in mammals such as mice and primates, including humans, the known host range today also comprises birds and fish. New PyV species are consistently discovered, and although the pathogenic potential of most of these viruses remains unknown, the awareness of emerging viral diseases makes studying PyVs no less significant. The first two isolated human polyomaviruses, BKPyV and JCPyV, are opportunistic pathogens that can cause severe renal and cerebral diseases, respectively. The efforts to determine PyV pathogenicity factors have resulted in what today is an established paradigm: Most polyomaviruses require sialic acid-bearing receptors for infectious entry. As non-enveloped particles, many polyomaviruses use their major and outermost capsid protein to recognize a terminal sialic acid moiety on their target cells' glycolipids and glycoproteins. Subtle differences in the VP1 sequences mediate the recognition of different sialic acid motifs. The most prominent instance is in the sialyl binding sites of BKPyV and the simian PyV SV40. Two single amino acid substitutions enable SV40 to preferentially bind (i) Neu5Gc over Neu5Ac, two sialic acids that differ by hydroxylation at the *N*-acetyl group, and (ii) GM1 over the b-series gangliosides. Such differentiations are intertwined with the co-evolution of hosts and viruses: As humans lost the ability to produce Neu5Gc, engaging only Neu5Ac signified an increased fitness for the human virus. A consequence, which is valid for PyVs and viruses in general, is that the presence or absence of specific cellular receptors may facilitate or prevent viral diseases.

The deployment of glycan array screening has supported many studies of PyV-receptor interactions. In a few examples, a single carbohydrate group (glycan) could be determined as the primary determinant of infection, including the lactoseries tetrasaccharide c (LSTc) for JCPyV. Glycan microarray experiments involve the VP1's screening against a broad library of carbohydrates from various sources, which here, for the first time, indicated interactions of VP1s with neutral glycans, contrasting the negatively charged sialyl glycans or glycosaminoglycans. These findings enabled the characterization of glycan analog complexes by X-ray crystallography, including the structures of chimpanzee polyomavirus (ChPyV) VP1 with galactose-terminating oligosaccharides, Lyon IARC polyomavirus (LIPyV) VP1 with neolacto-series sugars, and, foremost, the interaction of sheep polyomavirus (ShPyV) VP1 with the Forssman antigen pentaose. These interactions attest that the receptor usage within the polyomavirus family is even more versatile than previously assumed.

Investigating the structural determinants of viral receptor binding also helps comprehend their evolution. For instance, I demonstrate that the ShPyV VP1's Forssman glycan binding site is phylogenetically related to the sialyl site in the VP1s of BKPyV and SV40, which provides valuable insights into the glycan engagement of ancestry viruses.

Lastly, my findings suggest that utilizing the outstanding binding affinity of ShPyV VP1 to the Forssman antigen, a longstanding subject in the fields of glycobiology, immunology, and oncology, may hold benefits for future research.

# Bibliography

1. Ströh, L. J. *et al.* Structural basis and evolution of glycan receptor specificities within the polyomavirus family. *mBio* **11,** 1–21 (2020).
2. Rustmeier, N. H. *et al.* A novel and highly specific Forssman antigen-binding protein from sheep polyomavirus. *bioRxiv,* 2023.04.10.536218 (2023).
3. Hendrix, R. W., Lawrence, J. G., Hatfull, G. F. & Casjens, S. The origins and ongoing evolution of viruses. *Trends in Microbiology* **8,** 504–508 (2000).
4. Koonin, E. V., Senkevich, T. G. & Dolja, V. V. The ancient virus world and evolution of cells. *Biology Direct* **1,** 29–29 (2006).
5. Koonin, E. V. & Dolja, V. V. Virus World as an Evolutionary Network of Viruses and Capsid-less Selfish Elements. *Microbiology and Molecular Biology Reviews* **78,** 278–303 (2014).
6. Forterre, P. The origin of viruses and their possible roles in major evolutionary transitions. *Virus Research* **117,** 5–16 (2006).
7. Forterre, P. & Krupovic, M. in *Viruses: Essential Agents of Life* (2012).
8. Krupovic, M. & Koonin, E. V. Multiple origins of viral capsid proteins from cellular ancestors. *Proceedings of the National Academy of Sciences of the United States of America* **114,** E2401–E2410 (2017).
9. Koonin, E. V. & Dolja, V. V. A virocentric perspective on the evolution of life. *Current Opinion in Virology* **3,** 546–557 (2013).
10. Finsterbusch, T. & Mankertz, A. Porcine circoviruses—Small but powerful. *Virus Research* **143,** 177–183 (2009).
11. Raoult, D. *et al.* The 1.2-megabase genome sequence of Mimivirus. *Science* (2004).
12. Arslan, D., Legendre, M., Seltzer, V., Abergel, C. & Claverie, J. M. Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. *Proceedings of the National Academy of Sciences of the United States of America* (2011).
13. Shackelton, L. A. & Holmes, E. C. The evolution of large DNA viruses: combining genomic information of viruses and their hosts. *Trends in Microbiology* **12,** 458–465 (2004).
14. Bergh, Ø., Børsheim, K. Y., Bratbak, G. & Heldal, M. High abundance of viruses found in aquatic environments. *Nature* **340,** 467–468 (1989).
15. Rosario, K. & Breitbart, M. Exploring the viral world through metagenomics. *Current Opinion in Virology* **1,** 289–297 (2011).
16. Eickbush, T. H. Telomerase and Retrotransposons: Which came first? *Science* **277,** 911–912 (1997).
17. Krupovič, M. & Bamford, D. H. Virus evolution: How far does the double $\beta$-barrel viral lineage extend? *Nature Reviews Microbiology* **6,** 941–948 (2008).
18. Benson, S. D., Bamford, J. K., Bamford, D. H. & Burnett, R. M. Does common architecture reveal a viral lineage spanning all three domains of life? *Molecular Cell* **16,** 673–685 (2004).

19. Bamford, D. H., Grimes, J. M. & Stuart, D. I. What does structure tell us about virus evolution? *Current Opinion in Structural Biology* **15,** 655–663 (2005).

20. Krupovic, M. & Bamford, D. H. Double-stranded DNA viruses: 20 families and only five different architectural principles for virion assembly. *Current Opinion in Virology* **1,** 118–124 (2011).

21. Benson, S. D., Bamford, J. K., Bamford, D. H. & Burnett, R. M. Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures. *Cell* **98,** 825–833 (1999).

22. Caspar, D. L. Structure of Bushy Stunt Virus. *Nature 1956 177:4506* **177,** 475–476 (1956).

23. Rayment, I., Baker, T. S., Caspar, D. L. & Murakami, W. T. Polyoma virus capsid structure at 22.5 Å resolution. *Nature* **295,** 110–115 (1982).

24. Olsen, R. H., Siak, J.-S. & Gray, R. H. Characteristics of PRD1, a plasmid-dependent broad host range DNA bacteriophage. *Journal of Virology* **14,** 689–699 (1974).

25. Crick, F. H. & Watson, J. D. Structure of small viruses. *Nature 1956 177:4506* **177,** 473–475 (1956).

26. De Fremery, D. & Knight, C. A. A chemical comparison of three strains of tomato bushy stunt virus. *Journal of Biological Chemistry* **214,** 559–566 (1955).

27. Klug, A. & Finch, J. T. Structure of viruses of the papilloma-polyoma type: I. Human wart virus. *Journal of Molecular Biology* **11,** 403–423 (1965).

28. Caspar, D. L. & Klug, A. Physical principles in the construction of regular viruses. *Cold Spring Harbor symposia on quantitative biology* **27,** 1–24 (1962).

29. Weber, K., Rosenbusch, J. & Harrison, S. C. Structure of tomato bushy stunt virus. *Virology* **41,** 763–765 (1970).

30. Richardson, J. S. The anatomy and taxonomy of protein structure. *Advances in Protein Chemistry* **34,** 167–339 (1981).

31. Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K. & Bricogne, G. Tomato Bushy Stunt Virus at 2.9 Resolution. *Nature* **276,** 368–373 (1978).

32. Rossmann, M. G. *et al.* Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature* (1985).

33. Hogle, J. M., Chow, M. & Filman, D. J. Three-dimensional structure of poliovirus at 2.9 Å resolution. *Science* (1985).

34. Liddington, R. C. *et al.* Structure of simian virus 40 at 3.8-å resolution. *Nature* (1991).

35. McKenna, R. *et al.* Atomic structure of single-stranded DNA bacteriophage $\phi$X174 and its functional implications. *Nature* **355,** 137–143 (1992).

36. Krupovic, M., Makarova, K. S. & Koonin, E. V. Cellular homologs of the double jelly-roll major capsid proteins clarify the origins of an ancient virus kingdom. *Proceedings of the National Academy of Sciences of the United States of America* **119** (2022).

37. Kazlauskas, D., Varsani, A., Koonin, E. V. & Krupovic, M. Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. *Nature Communications* **10,** 1–12 (2019).

38. Johne, R. & Müller, H. Polyomaviruses of birds: etiologic agents of inflammatory diseases in a tumor virus family. *Journal of Virology* **81,** 11554–11559 (2007).

39. Gerits, N. & Moens, U. Agnoprotein of mammalian polyomaviruses. *Virology* **432,** 316–326 (2012).

40. Decaprio, J. A. & Garcea, R. L. A cornucopia of human polyomaviruses. *Nature Reviews Microbiology 2013 11:4* **11,** 264–276 (2013).

41. Moens, U. *et al.* Biology, evolution, and medical importance of polyomaviruses: An update. *Infection, Genetics and Evolution* **54,** 18–38 (2017).

42. Feng, H., Shuda, M., Chang, Y. & Moore, P. S. Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science* **319,** 1096–1100 (2008).

43. Van der Meijden, E. *et al.* Discovery of a new human polyomavirus associated with Trichodysplasia Spinulosa in an immunocompromized patient. *PLoS Pathogens* **6,** 1–10 (2010).

44. Schowalter, R. M., Pastrana, D. V., Pumphrey, K. A., Moyer, A. L. & Buck, C. B. Merkel cell polyomavirus and two previously unknown polyomaviruses are chronically shed from human skin. *Cell Host and Microbe* **7,** 509–515 (2010).

45. Nguyen, K. D. *et al.* Human polyomavirus 6 and 7 are associated with pruritic and dyskeratotic dermatoses. *Journal of the American Academy of Dermatology* **76,** 932–940 (2017).

46. Lehn, H. & Müller, H. Cloning and characterization of budgerigar fledgling disease virus, an avian polyomavirus. *Virology* **151,** 362–370 (1986).

47. Rott, O., Kröger, M., Müller, H. & Hobom, G. The genome of budgerigar fledgling disease virus, an avian polyomavirus. *Virology* **165,** 74–86 (1988).

48. Dela Cruz, F. N. *et al.* Novel polyomavirus associated with brain tumors in free-ranging raccoons, western United States. *Emerging infectious diseases* **19,** 77–84 (2013).

49. Sweet, B. H. & Hilleman, M. R. The vacuolating virus, S.V.40. *Experimental Biology and Medicine* **105,** 420–427 (1960).

50. Eddy, B. E., Borman, G. S., Grubbs, G. E. & Young, R. D. Identification of the oncogenic substance in rhesus monkey kidney cell cultures as simian virus 40. *Virology* **17,** 65–75 (1962).

51. An, P., Robles, M. T. S. & Pipas, J. M. Large T antigens of polyomaviruses: Amazing molecular machines. *https://doi.org/10.1146/annurev-micro-092611-150154* **66,** 213–236 (2012).

52. Forest, C., Czerucka, D., Negrel, R. & Ailhaud, G. Establishment of a human cell line after transformation by a plasmid containing the early region of the SV40 genome. *Cell biology international reports* **7,** 73–81 (1983).

53. Jha, K. K., Banga, S., Palejwala, V. & Ozer, H. L. SV40-mediated immortalization. *Experimental Cell Research* **245,** 1–7 (1998).

54. Stewart, S. E., Eddy, B. E., Gochenour, A. M., Borgese, N. G. & Grubbs, G. E. The induction of neoplasms with a substance released from mouse tumors by tissue culture. *Virology* **3,** 380–400 (1957).

55. Dmochowski, L. & Grey, C. E. Subcellular structures of possible viral origin in some mammalian tumors. *Annals of the New York Academy of Sciences* **68,** 559–615 (1957).

56. Kahler, H., Rowe, W. P., Lloyd, B. J. & Hartley, J. W. Electron microscopy of Mouse Parotid Tumor (Polyoma) Virus. *JNCI: Journal of the National Cancer Institute* **22,** 647–657 (1959).

57. Klug, A. Structure of viruses of the papilloma-polyoma type: II. Comments on other work. *Journal of Molecular Biology* **11,** 424–431 (1965).

58. Finch, J. T. The surface structure of polyoma virus. *The Journal of General Virology* **24,** 359–364 (1974).

59. Salunke, D. M., Caspar, D. L. & Garcea, R. L. Self-assembly of purified polyomavirus capsid protein VP1. *Cell* **46,** 895–904 (1986).

60. Walter, G. & Deppert, W. Intermolecular disulfide bonds: an important structural feature of the polyoma virus capsid. *Cold Spring Harbor symposia on quantitative biology* **39 Pt 1,** 255–257 (1975).

61. Stehle, T., Gamblin, S. J., Yan, Y. & Harrison, S. C. The structure of simian virus 40 refined at 3.1 Å resolution. *Structure* **4,** 165–182 (1996).

62. Stehle, T. & Harrison, S. C. Crystal structures of murine polyomavirus in complex with straight-chain and branched-chain sialyloligosaccharide receptor fragments. *Structure* **4,** 183–194 (1996).

63. Bayer, N. J. *et al.* Structure of Merkel cell polyomavirus capsid and interaction with its glycosaminoglycan attachment receptor. *Journal of Virology* **94** (2020).

64. Chang, D., Haynes, J. I., Brady, J. N. & Consigli, R. A. The use of additive and subtractive approaches to examine the nuclear localization sequence of the polyomavirus major capsid protein VP1. *Virology* **189,** 821–827 (1992).

65. Ishii, N. *et al.* Analysis of a nuclear localization signal of simian virus 40 major capsid protein Vp1. *Journal of Virology* **70,** 1317–1322 (1996).

66. Qu, Q. *et al.* Nuclear entry mechanism of the human polyomavirus JC virus-like particle: role of importins and the nuclear pore complex. *The Journal of biological chemistry* **279,** 27735–27742 (2004).

67. Stehle, T., Yan, Y., Benjamin, T. L. & Harrison, S. C. Structure of murine polyomavirus complexed with an oligosaccharide receptor fragment. *Nature* **369,** 160–163 (1994).

68. Stehle, T. & Harrison, S. C. High-resolution structure of a polyomavirus VP1-oligosaccharide complex: Implications for assembly and receptor binding. *EMBO Journal* **16,** 5139–5148 (1997).

69. Chang, D. *et al.* Production of the antigen and the antibody of the JC virus major capsid protein VP1. *Journal of virological methods* **59,** 177–187 (1996).

70. Moens, U., Van Ghelue, M., Song, X. & Ehlers, B. Serological cross-reactivity between human polyomaviruses. *Reviews in Medical Virology* **23,** 250–264 (2013).

71. Jelcic, I. *et al.* Broadly neutralizing human monoclonal JC polyomavirus VP1-specific antibodies as candidate therapeutics for progressive multifocal leukoencephalopathy. *Science Translational Medicine* **7** (2015).

72. Krupovic, M. Recombination between RNA viruses and plasmids might have played a central role in the origin and evolution of small DNA viruses. *Bioessays : News and Reviews in Molecular, Cellular and Developmental Biology* **34,** 867–870 (2012).

73. Krupovic, M. Networks of evolutionary interactions underlying the polyphyletic origin of ssDNA viruses. *Current Opinion in Virology* **3,** 578–586 (2013).

74. Tsao, J. *et al.* The three-dimensional structure of canine parvovirus and its functional implications. *Science* **251,** 1456–1464 (1991).

75. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research* **49,** W293–W296 (2021).

76. Buck, C. B. *et al.* The ancient evolutionary history of polyomaviruses. *PLoS Pathogens* **12,** 1005574 (2016).

77. Torres, C. Evolution and molecular epidemiology of polyomaviruses. *Infection, Genetics and Evolution* **79,** 104150 (2020).

78. Calvignac-Spencer, S. *et al.* A taxonomy update for the family Polyomaviridae. *Archives of Virology* **161,** 1739–1750 (2016).

79. Ehlers, B. *et al.* Novel polyomaviruses in mammals from multiple orders and reassessment of polyomavirus evolution and taxonomy. *Viruses* **11,** 930 (2019).

80. Moens, U. *et al.* ICTV virus taxonomy profile: Polyomaviridae. *Journal of General Virology* **98,** 1159–1160 (2017).

81. Stroh, L. J. *et al.* Structure Analysis of the Major Capsid Proteins of Human Polyomaviruses 6 and 7 Reveals an Obstructed Sialic Acid Binding Site. *Journal of Virology* **88,** 10831–10839 (2014).

82. Schuurman, R., Sol, C. & Van Der Noordaa, J. The complete nucleotide sequence of bovine polyomavirus. *Journal of General Virology* **71,** 1723–1735 (1990).

83. Anthony, S. J. *et al.* Identification of a novel cetacean polyomavirus from a common dolphin (Delphinus delphis) with Tracheobronchitis. *PloS one* **8** (2013).

84. Tsai, B. *et al.* Gangliosides are receptors for murine polyoma virus and SV40. *EMBO Journal* **22,** 4346–4355 (2003).

85. Varki, A. *et al.* Essentials of Glycobiology. *Cold Spring Harbor (NY),* 823 (2017).

86. O'Hara, S. D., Stehle, T. & Garcea, R. *Glycan receptors of the Polyomaviridae: Structure, function, and pathogenesis* 2014.

87.  Neu, U., Woellner, K., Gauglitz, G. & Stehle, T. Structural basis of GM1 ganglioside recognition by simian virus 40. *Proceedings of the National Academy of Sciences of the United States of America* **105,** 5219–5224 (2008).

88.  Rustmeier, N. H., Strebl, M. & Stehle, T. *The symmetry of viral sialic acid binding sites-implications for antiviral strategies* 2019.

89.  Neu, U., Wang, J., Macejak, D., Garcea, R. L. & Stehle, T. Structures of the Major Capsid Proteins of the Human Karolinska Institutet and Washington University Polyomaviruses. *Journal of Virology* **85,** 7384–7392 (2011).

90.  Buch, M. H. C. *et al.* Structural and functional analysis of murine polyomavirus capsid proteins establish the determinants of ligand recognition and pathogenicity. *PLoS Pathogens* **11,** e1005104 (2015).

91.  Neu, U. *et al.* Structure-function analysis of the human JC polyomavirus establishes the LSTc pentasaccharide as a functional receptor motif. *Cell Host and Microbe* **8,** 309–319 (2010).

92.  Neu, U. *et al.* A structure-guided mutation in the major capsid protein retargets BK polyomavirus. *PLoS Pathogens* **9,** 1–13 (2013).

93.  Ströh, L. J. *et al.* The greater affinity of JC polyomavirus capsid for $\alpha$2,6-linked lactoseries tetrasaccharide c than for other sialylated glycans is a major determinant of infectivity. *Journal of virology* **89,** 6364–6375 (2015).

94.  Varki, A. N-glycolylneuraminic acid deficiency in humans. *Biochimie* **83,** 615–622 (2001).

95.  Neu, U. *et al.* Structures of merkel cell polyomavirus VP1 complexes define a sialic acid binding site required for infection. *PLoS Pathogens* **8,** 8 (2012).

96.  Neu, U. *et al.* Structures of B-lymphotropic polyomavirus VP1 in complex with oligosaccharide ligands. *PLoS Pathogens* **9** (2013).

97.  Ströh, L. J. *et al.* Trichodysplasia spinulosa-associated polyomavirus uses a displaced binding site on VP1 to engage sialylated glycolipids. *Plos Pathogens* **11,** e1005112–e1005112 (2015).

98.  Schowalter, R. M., Pastrana, D. V. & Buck, C. B. Glycosaminoglycans and sialylated glycans sequentially facilitate merkel cell polyomavirus infectious entry. *PLoS Pathogens* **7,** 1002161 (2011).

99.  Geoghegan, E. M. *et al.* Infectious entry and neutralization of pathogenic JC polyomaviruses. *Cell Reports* **21,** 1169–1179 (2017).

100.  Hurdiss, D. L., Frank, M., Snowden, J. S., Macdonald, A. & Ranson, N. A. The structure of an infectious human polyomavirus and its interactions with cellular receptors. *Structure* **26,** 839 (2018).

101.  Elphick, G. F. *et al.* The human polyomavirus, JCV, uses serotonin receptors to infect cells. *Science* **306,** 1380–1383 (2004).

102.  Maginnis, M. S., Haley, S. A., Gee, G. V. & Atwood, W. J. Role of N-linked glycosylation of the 5-HT 2A receptor in JC virus infection. *Journal of Virology* **84,** 9677–9684 (2010).

103. Assetta, B. *et al.* 5-HT2 receptors facilitate JC polyomavirus entry. *Journal of Virology* **87,** 13490 (2013).

104. Pho, M. T., Ashok, A. & Atwood, W. J. JC virus enters human glial cells by clathrin-dependent receptor-mediated endocytosis. *Journal of Virology* **74,** 2288–2292 (2000).

105. Pelkmans, L., Kartenbeck, J. & Helenius, A. Caveolar endocytosis of simian virus 40 reveals a new two-step vesicular-transport pathway to the ER. *Nature Cell Biology 2001 3:5* **3,** 473–483 (2001).

106. Qian, M., Cai, D., Verhey, K. J. & Tsai, B. A lipid receptor sorts polyomavirus from the endolysosome to the endoplasmic reticulum to cause infection. *Plos Pathogens* **5,** e1000465–e1000465 (2009).

107. Ewers, H. *et al.* GM1 structure determines SV40-induced membrane invagination and infection. *Nature Cell Biology* **12,** 11–18 (2010).

108. Morris-Love, J. *et al.* JC polyomavirus uses extracellular vesicles to infect target cells. *mBio* **10** (2019).

109. Handala, L. *et al.* BK polyomavirus hijacks extracellular vesicles for en bloc transmission. *Journal of Virology* (2020).

110. Pelkmans, L. & Helenius, A. *Endocytosis via caveolae* 2002.

111. Querbes, W., O'Hara, B. A., Williams, G. & Atwood, W. J. Invasion of host cells by JC virus identifies a novel role for caveolae in endosomal sorting of noncaveolar ligands. *Journal of virology* **80,** 9402–9413 (2006).

112. Liebl, D. *et al.* Mouse polyomavirus enters early endosomes, requires their acidic pH for productive infection, and meets transferrin cargo in Rab11-positive endosomes. *Journal of virology* **80,** 4610–4622 (2006).

113. Qian, M. & Tsai, B. Lipids and proteins act in opposing manners to regulate polyomavirus infection. *Journal of virology* **84,** 9840–9852 (2010).

114. Engel, S. *et al.* Role of endosomes in simian virus 40 entry and infection. *Journal of virology* **85,** 4198–4211 (2011).

115. Maginnis, M. S., Nelson, C. D. & Atwood, W. J. JC polyomavirus attachment, entry, and trafficking: unlocking the keys to a fatal infection. *Journal of NeuroVirology 2014 21:6* **21,** 601–613 (2014).

116. Ashok, A. & Atwood, W. J. Contrasting roles of endosomal pH and the cytoskeleton in infection of human glial cells by JC virus and simian virus 40. *Journal of virology* **77,** 1347–1356 (2003).

117. Eash, S. & Atwood, W. J. Involvement of cytoskeletal components in BK virus infectious entry. *Journal of virology* **79,** 11734–11741 (2005).

118. Zila, V., Difato, F., Klimova, L., Huerfano, S. & Forstova, J. Involvement of microtubular network and its motors in productive endocytic trafficking of mouse polyomavirus. **9** (2014).

119. Becker, M. *et al.* Infectious Entry of Merkel Cell Polyomavirus. *Journal of virology* **93** (2019).

120. Magnuson, B. *et al.* ERp29 triggers a conformational change in polyomavirus to stimulate membrane binding. *Molecular cell* **20,** 289–300 (2005).

121. Gilbert, J., Ou, W., Silver, J. & Benjamin, T. Downregulation of protein disulfide isomerase inhibits infection by the mouse polyomavirus. *Journal of virology* **80,** 10868–10870 (2006).

122. Schelhaas, M. *et al.* Simian Virus 40 depends on ER protein folding and quality control factors for entry into host cells. *Cell* **131,** 516–529 (2007).

123. Nelson, C. D., Derdowski, A., Maginnis, M. S., O'Hara, B. A. & Atwood, W. J. The VP1 subunit of JC polyomavirus recapitulates early events in viral trafficking and is a novel tool to study polyomavirus entry. *Virology* **428,** 30–40 (2012).

124. Walczak, C. P. & Tsai, B. A PDI family network acts distinctly and coordinately with ERp29 to facilitate polyomavirus infection. *Journal of virology* **85,** 2386–2396 (2011).

125. Geiger, R. *et al.* BAP31 and BiP are essential for dislocation of SV40 from the endoplasmic reticulum to the cytosol. *Nature cell biology* **13,** 1305–1314 (2011).

126. Nakanishi, A., Shum, D., Morioka, H., Otsuka, E. & Kasamatsu, H. Interaction of the Vp3 nuclear localization signal with the importin alpha 2/beta heterodimer directs nuclear entry of infecting simian virus 40. *Journal of virology* **76,** 9368–9377 (2002).

127. Gasparovic, M. L., Gee, G. V. & Atwood, W. J. JC virus minor capsid proteins Vp2 and Vp3 are essential for virus propagation. *Journal of Virology* **80,** 10858–10861 (2006).

128. Nakanishi, A., Li, P. P., Qu, Q., Jafri, Q. H. & Kasamatsu, H. Molecular dissection of nuclear entry-competent SV40 during infection. *Virus research* **124,** 226–230 (2007).

129. Weis, K. Importins and exportins: how to get in and out of the nucleus. *Trends in biochemical sciences* **23,** 185–189 (1998).

130. Fay, N. & Panté, N. Nuclear entry of DNA viruses. *Frontiers in Microbiology* **6,** 140527 (2015).

131. Fay, N. & Panté, N. *Old foes, new understandings: Nuclear entry of small non-enveloped DNA viruses* 2015.

132. Bennett, S. M., Zhao, L., Bosard, C. & Imperiale, M. J. Role of a nuclear localization signal on the minor capsid Proteins VP2 and VP3 in BKPyV nuclear entry. *Virology* **474,** 110–116 (2015).

133. Soldatova, I., Prilepskaja, T., Abrahamyan, L., Forstová, J. & Huérfano, S. Interaction of the mouse polyomavirus capsid proteins with importins es required for efficient import of viral DNA into the cell nucleus. *Viruses* **10,** 165 (2018).

134. Butin-Israeli, V. *et al.* Simian virus 40 induces lamin A/C fluctuations and nuclear envelope deformation during cell entry. *http://dx.doi.org/10.4161/nucl.2.4.16371* **2,** 320–330 (2011).

135. Mayberry, C. L. & Maginnis, M. S. *Taking the scenic route: Polyomaviruses utilize multiple pathways to reach the same destination* 2020.

136. Cid, E., Yamamoto, M. & Yamamoto, F. Mixed-up sugars: glycosyltransferase cross-reactivity in cancerous tissues and their therapeutic targeting. *ChemBioChem* **23,** e202100460 (2022).

137. Johne, R., Enderlein, D., Nieper, H. & Müller, H. Novel polyomavirus detected in the feces of a chimpanzee by nested broad-spectrum PCR. *Journal of Virology* **79,** 3883–3887 (2005).

138. Deuzing, I. *et al.* Detection and characterization of two chimpanzee polyomavirus genotypes from different subspecies. *Virology Journal* **7,** 1–7 (2010).

139. Mishra, N. *et al.* Identification of a Novel Polyomavirus in a Pancreatic Transplant Recipient With Retinal Blindness and Vasculitic Myopathy. **210,** 1595–1599 (2014).

140. Gheit, T. *et al.* Isolation and characterization of a novel putative human polyomavirus. *Virology* **506,** 45–54 (2017).

141. Fahsbender, E. *et al.* Lyon-IARC polyomavirus DNA in feces of diarrheic cats. *Microbiology Resource Announcements* **8** (2019).

142. Li, Y. *et al.* First detection and complete genome analysis of the Lyon IARC polyomavirus in China from samples of diarrheic cats. *Virus Genes* **1,** 3 (2021).

143. Kamminga, S. *et al.* Development and evaluation of a broad bead-based multiplex immunoassay to measure IgG seroreactivity against human polyomaviruses. *Journal of Clinical Microbiology* **56** (2018).

144. Siqueira, J. D. *et al.* Endemic infection of stranded southern sea otters (Enhydra lutris nereis) with novel parvovirus, polyomavirus, and adenovirus. *Journal of Wildlife Diseases* **53,** 532–542 (2017).

145. Krautwald, M.-E. .-., Müller, H. & Kaleta, E. F. Polyomavirus infection in budgerigars (Melopsittacus undulatus): clinical and aetiological studies. *Zentralblatt fur Veterinarmedizin. Reihe B. Journal of veterinary medicine. Series B* **36,** 459–467 (1989).

146. Varki, A. *et al.* Symbol nomenclature for graphical representations of glycans. *Glycobiology* **25,** 1323 (2015).

147. Neelamegham, S. *et al.* Updates to the Symbol Nomenclature for Glycans guidelines. *Glycobiology* **29,** 620 (2019).

148. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology* **215,** 403–410 (1990).

149. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography* (2004).

150. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **37,** 1530–1534 (2020).

151. Madeira, F. *et al.* Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Research* **50,** W276–W279 (2022).

152. Kleywegt, G. J. & Jones, T. A. xdlMAPMAN and xdlDATAMAN - Programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets. *Acta Crystallographica Section D: Biological Crystallography* **52,** 826–828 (1996).

153. McCoy, A. J. *et al.* Phaser crystallographic software. *Journal of Applied Crystallography* **40,** 658–674 (2007).

154. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *urn:issn:0907-4449* **66,** 213–221 (2010).

155. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallographica Section D: Structural Biology* (2019).

156. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta crystallographica. Section D, Biological crystallography* **53,** 240–255 (1997).

157. Gouy, M., Tannier, E., Comte, N. & Parsons, D. P. Seaview Version 5: A Multiplatform Software for Multiple Sequence Alignment, Molecular Phylogenetic Analyses, and Tree Reconciliation. *Methods in Molecular Biology* **2231,** 241–260 (2021).

158. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science* **30,** 70–82 (2021).

159. Kabsch, W. *et al.* XDS. *Acta Crystallographica Section D Biological Crystallography* (2010).

160. Benson, D. A. *et al.* GenBank. *Nucleic acids research* **41** (2013).

161. O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research* **44,** D733–D745 (2016).

162. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32,** 1792–1797 (2004).

163. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. **56,** 564–577 (2007).

164. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods 2017 14:6* **14,** 587–589 (2017).

165. Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* **35,** 518–522 (2018).

166. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature 2021 596:7873* **596,** 583–589 (2021).

167. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nature Methods 2022 19:6* **19,** 679–682 (2022).

168. Ponder, J. W. & Case, D. A. Force fields for protein simulations. *Advances in Protein Chemistry* **66,** 27–85 (2003).

169. Case, D. A. *et al.* The Amber biomolecular simulation programs. *Journal of Computational Chemistry* **26,** 1668–1688 (2005).

170. Diederichs, K. & Karplus, P. A. Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nature Structural Biology 1997 4:4* **4,** 269–275 (1997).

171. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336,** 1030–1033 (2012).

172. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography* **69,** 1204–1214 (2013).

173. Brünger, A. T. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature 1992 355:6359* **355,** 472–475 (1992).

174. Agirre, J. *et al.* The CCP4 suite: integrative software for macromolecular crystallography. *urn:issn:2059-7983* **79,** 449–461 (2023).

175. Matthews, B. W. Solvent content of protein crystals. *Journal of Molecular Biology* **33,** 491–497 (1968).

176. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography* **67,** 355–367 (2011).

177. Liu, Y. *et al.* Neoglycolipid-based oligosaccharide microarray system: Preparation of ngls and their noncovalent immobilization on nitrocellulose-coated glass slides for microarray analyses. *Methods in Molecular Biology* **808,** 117–136 (2012).

178. Palma, A. S., Feizi, T., Childs, R. A., Chai, W. & Liu, Y. The neoglycolipid (NGL)-based oligosaccharide microarray system poised to decipher the meta-glycome. *Current Opinion in Chemical Biology* **18,** 87–94 (2014).

179. Mayer, M. & Meyer, B. Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *Journal of the American Chemical Society* (2001).

180. Howell, P. L. Identification of heavy-atom derivatives by normal probability methods. *Journal of Applied Crystallography* **25,** 81–86 (1992).

181. Team, R. C. R: A language and environment for statistical computing (2013).

182. Ritz, C., Baty, F., Streibig, J. C. & Gerhard, D. Dose-response analysis using R. *PLOS ONE* **10** (ed Xia, Y.) e0146021 (2015).

183. Gildersleeve, J. C., Oyelaran, O., Simpson, J. T. & Allred, B. Improved procedure for direct coupling of carbohydrates to proteins via reductive amination. *Bioconjugate Chemistry* **19,** 1485–1490 (2008).

184. Hudson, K. L. *et al.* Carbohydrate-Aromatic Interactions in Proteins. *Journal of the American Chemical Society* **137,** 15152–15160 (2015).

185. Heggelund, J. E. *et al.* Specificity of Escherichia coli heat-labile enterotoxin investigated by single-site mutagenesis and crystallography. *International Journal of Molecular Sciences 2019, Vol. 20, Page 703* **20,** 703 (2019).

186. Yamamoto, M., Cid, E. & Yamamoto, F. Molecular genetic basis of the human Forssman glycolipid antigen negativity. *Scientific Reports* **2,** 975 (2012).

187. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25,** 1189–1191 (2009).

188.  Ströh, L. J. & Stehle, T. Glycan engagement by viruses: Receptor switches and specificity. *Annual Review of Virology* **1,** 285–306 (2014).

189.  Khan, Z. M. *et al.* Crystallographic and glycan microarray analysis of human polyomavirus 9 VP1 identifies N-glycolyl neuraminic acid as a receptor candidate. *Journal of Virology* **88,** 6100–6111 (2014).

# Appendices

## 1 Rustmeier *et al.* (2023)

Rustmeier, N. H. *et al.* A novel and highly specific Forssman antigen-binding protein from sheep polyomavirus. *bioRxiv,* 2023.04.10.536218 (2023) was distributed as a preprint (`https://doi.org/10.1101/2023.04.10.536218`) under the CC BY 4.0 deed (`https://creativecommons.org/licenses/by/4.0/`). At the current stage, the material has not undergone a peer-review process. Revised versions of this material may be available online at later times.

For inclusion, the main document of Rustmeier *et al.* (2023) (`https://www.biorxiv.org/content/10.1101/2023.04.10.536218v2.full.pdf`) was scaled to 95%. Pages 1-22, which only contain text, were converted to grayscale.

The Supplementary Tables S1 (`https://www.biorxiv.org/content/biorxiv/early/2023/04/15/2023.04.10.536218/DC1/embed/media-1.xlsx`) and S2 (`https://www.biorxiv.org/content/biorxiv/early/2023/04/15/2023.04.10.536218/DC2/embed/media-2.docx`) were compiled by our Imperial College London collaborators and are available online.

The Supplementary Material S3-S6 (`https://www.biorxiv.org/content/biorxiv/early/2023/04/15/2023.04.10.536218/DC3/embed/media-3.pdf`) was appended printing two pages per page.

Table 19: Author contributions Rustmeier et al. (2023)

| Author | Author position | Scientific ideas % | Data generation % | Analysis & interpretation % | Paper writing % |
|---|---|---|---|---|---|
| Nils H. Rustmeier | 1 | 40 | 70 | 45 | 60 |
| Lisete M. Silva | 2 | 0-5 | 5 | 0-5 | 0 |
| Antonio Di Maio | 3 | 0-5 | 5 | 0-5 | 0 |
| Joshua C. Müller | 4 | 0-5 | 5 | 0-5 | 0 |
| Alexander Herrmann | 5 | 0-5 | 5 | 0-5 | 0 |
| Ten Feizi | 6 | 10 | 0 | 10 | 10 |
| Yan Liu | 7 | 20 | 10 | 15 | 10 |
| Thilo Stehle | 8 | 10 | 0 | 10 | 20 |
| Title of paper: | A novel and highly specific Forssman antigen-binding protein from sheep polyomavirus | | | | |
| Status in publication process: | Preprint available (https://doi.org/10.1101/2023.04.10.536218) | | | | |

1   **A novel and highly specific Forssman antigen-binding protein from**

2   **sheep polyomavirus**

3

4   Nils H. Rustmeier[1], Lisete M. Silva[2], Antonio Di Maio[2], Joshua C. Müller[1], Alexander

5   Herrmann[1], Ten Feizi[2], Yan Liu[2], Thilo Stehle[1]

6

7   [1]Interfaculty Institute of Biochemistry, University of Tübingen, Tübingen, Germany

8   [2]Glycosciences Laboratory, Faculty of Medicine, Imperial College London, London,

9   UK

10

15   **Corresponding author:**

16   Thilo Stehle: thilo.stehle@uni-tuebingen.de

17   Phone: +49-7071-2978090

18

19   **Abstract**

20   Polyomaviruses are small, non-enveloped double-stranded DNA viruses of humans

21   and other mammals, birds, and fish. Infections are usually asymptomatic and result in

22   latency, however, some polyomaviruses can induce severe diseases, including cancer,

23   in immunocompromised individuals. Established cellular receptors for polyomavirus

24   infection are sialylated glycolipids (such as gangliosides), membrane proteins, and

25   glycosaminoglycans. Polyomaviruses are usually highly host specific but the exact

102

26    principles that govern host tropism remain unknown in many cases. Here, glycan array

27    screening shows that the major capsid protein VP1 of sheep polyomavirus (ShPyV)

28    binds to the Forssman glycolipid, an antigen of many vertebrates and a potential tumor

29    marker in humans. Following closer investigation, we can report for the first time that

30    a neutral, non-sialylated glycolipid acts as a polyomavirus receptor. Concurrently, we

31    present the first report of a viral protein that specifically engages the Forssman antigen.

32    We demonstrate that ShPyV VP1 binds to Forssman-positive erythrocytes but not

33    those of human A, B and O blood groups, which is a clear distinction from features

34    thus far described for Forssman lectins. X-ray crystallography and structural analysis

35    of the VP1-Forssman glycan complex define the terminal Forssman disaccharide as

36    the determinant of this protein-receptor interaction. These results strongly suggest that

37    the sheep polyomavirus can use Forssman antigen for infectious cell entry.

38    Furthermore, the ability of ShPyV VP1 to distinguish Forssman-positive from -negative

39    cells may prove useful for monitoring the Forssman-'status' of normal, preneoplastic

40    and neoplastic cells and tissues and establishing the antigen level as a biomarker.

41

42    **Author summary**

43    Elucidation of host cell receptor specificities of viral infection is crucial to understand

44    the pathobiology of associated diseases and develop treatments. However, for many

45    polyomaviruses the receptor engagement as the initial event in infection is poorly

46    understood. In only a few cases polyomavirus tropism has been pinned down to a

47    single type of glycan receptor. While many polyomaviruses utilize sialyl glycans to

48    attach to host cells, the role of non-sialylated glycans as receptors is so far

49    underestimated. Here, we show for the first time that a glycan of neutral charge, in this

50    case the carbohydrate portion of the Forssman antigen, acts as a ligand for a

51    polyomavirus capsid protein and may thus contribute to host tropism and infective cell

52    entry. These results represent a significant addition to knowledge on polyomavirus-

53    glycan interactions and complement general principles of carbohydrate engagement

54    by viruses. Furthermore, as a specific binding protein of Forssman antigen, VP1 may

55    help to determine levels of this antigen in healthy and malignant tissues in humans.

56

57    **Introduction**

58

59    The Forssman antigen (FA), a neutral globo-series glycosphingolipid (GSL, see Table

60    1), is produced by the glycosyltransferase activity of Forssman synthetase (FS)[1,2].

61    Encoded by the *GBGT1* gene, FS catalyzes the addition of αGalNAc to the non-

62    reducing end of globoside (see Table 1). This reaction results in the key immunogenic

63    motif of the Forssman antigen, the terminal GalNAcα1-3GalNAcβ Forssman

64    disaccharide[3]. With few exceptions in humans, *GBGT1* is transcribed but encodes for

65    an inactive FS, resulting in a predominantly Forssman-negative human population[4,5].

66    So far, only a small number of healthy individuals worldwide were found to carry an

67    activating mutation within *GBGT1*. In these individuals, the resulting allele gives rise to

68    a Forssman antigen-positive (FA+) erythrocyte phenotype, which is now recognized as

69    the human FORS blood group[6–8]. Other reports of healthy human FA+ tissues are

70    scarce, and human sera often contain natural anti-FA antibodies[9–11]. More frequently,

71    Forssman antigenicity was described in cancerogenesis and a few other diseases of

72    humans[12–21]. Yet, both the genetics and the precise role of FA expression in disease

73    remain poorly understood.

74

75    The Forssman glycolipid also serves as a receptor for different bacterial infections in

76    animals[22–28], however, it is so far unknown if viruses also utilize FA to target their hosts.

77    In this study, by means of glycan microarray screening, nuclear magnetic resonance

104

78    (NMR), surface plasmon resonance (SPR), and cell binding assays, we show that FA,

79    which is the predominant glycolipid in ovine erythrocytes[29], serves as a cellular

80    attachment receptor for the VP1 of a polyomavirus originally identified in sheep[30]. This

81    designated sheep polyomavirus (ShPyV) is a member of a non-enveloped, double-

82    stranded DNA virus family with icosahedral capsids, scaffolded by the major capsid

83    protein VP1. To initiate an infection many polyomaviruses rely on the interaction

84    between their VP1 and sialic acid-equipped glycolipids in the target cell membranes[31–

85    36]. Although other infection receptors such as glycoproteins and glycosaminoglycans

86    have also been reported for some polyomaviruses[37–42], detailed structural information

87    is so far only available for the interactions between VP1s and sialylated

88    oligosaccharides. Thus, our studies present the first example of a polyomavirus VP1

89    protein interacting with a non-sialyl neutral glycan, complementing the principles of

90    how polyomaviruses recognize their targets.

91

92    **Results**

93

94    <u>Sheep polyomavirus VP1 binds the Forssman epitope with a strong preference over</u>

95    <u>other glycans</u>

96

97    To identify potential attachment factors of ShPyV, we prepared recombinant VP1

98    pentamers as previously described[43]. The protein was analyzed using a neoglycolipid-

99    based glycan screening array encompassing a broad spectrum of lipid-linked glycan

100   probes[44]. A full list of 672 glycan probes is in Table S1, and these include major types

101   of mammalian sequences found on glycoproteins (N- and O-linked), glycolipids, blood

102   group and Lewis antigen related, oligosaccharide fragments of glycosaminoglycan

103   heparin, as well as those derived from polysaccharides of bacteria, fungi, and plants.

104 Among these are 204 sialylated glycan probes variously bound by several

105 polyomavirus VP1s described previously[45–49]. We detected a signal for the Forssman

106 glycolipid that surpasses all other signals by at least one order of magnitude (Fig. 1).

107 A dominance of the terminal GalNAcα1-3GalNAcβ Forssman disaccharide is apparent

108 as the VP1 did not give binding signals with globoside glycolipid (P-antigen, position

109 25, Table S1), the precursor of the Forssman glycolipid which lacks the terminal

110 α-linked GalNAc residue. Furthermore, there were no signals with other GalNAc-

111 terminating glycan probes, including the blood group A related glycan probes, and the

112 two Tn antigen probes GalNAcα-Ser and -Thr (positions 221 and 222, Table S1). The

113 disaccharide probe GalNAcα1-3GalNAc (position 223) with a ring-opened GalNAc

114 core (reduced by reductive amination) was also not bound. These suggest the

115 importance of adjoining monosaccharides beyond the αGalNAc residue for ShPyV VP1

116 binding. Among the sialylated glycans analyzed only one short α2,6-linked sialyl probe

117 with a 9-O-acetylated N-acetylneuraminic acid residue ($Neu5,9Ac_2$) yielded a weak

118 signal (position 607). No binding signals were detected in the negative control

119 experiment with detection antibodies alone in the absence of ShPyV VP1 (Table S1).

120

121 To assess if the glycan binding revealed by the glycan microarray also occurs in

122 solution, we evaluated the interactions between ShPyV VP1 and the Forssman

123 pentaose ($F_P$) in the presence of α2,6-sialyllactose (6'SL) by saturation transfer

124 difference NMR (STD NMR) spectroscopy, which reveals protein-to-ligand

125 magnetization transfer in the case of short-term binding ($10^{-8}$ M < $K_D$ < $10^{-3}$ M)[50,51]. In

126 the first instance, the $^1$H reference resonances of isolated $F_P$ and 6'SL were collected

127 (Fig. 2a and 2b), which allowed us to unambiguously assign the oligosaccharide

128 signals[43,52] (except for the overlaps within 3.5-4.0 ppm) in the $^1$H spectrum of a sample

129 comprising VP1, $F_P$, and 6'SL (Fig. 2c). Subsequently, we used the same sample to

106

130  perform the STD NMR experiment. Evaluation of the resulting spectrum reveals

131  magnetization transfer to both $F_P$ and 6'SL, which infers that VP1 interacts with both

132  glycans in the solution (Fig. 2b). However, while resonances derived from all

133  monosaccharides in $F_P$ (two GalNAc, two Gal, and a Glc) received magnetization

134  transfer, only the sialic acid methyl group of 6'SL re-emerged in the STD spectrum.

135  Yet, these results are in good agreement with the findings from the glycan array

136  analysis of ShPyV VP1, which indicated a much higher affinity for $F_P$ than sialylated

137  glycans (Fig. 1). Strikingly, the methyl group of the αGalNAc in $F_P$ gives a particularly

138  strong peak in the STD spectrum (Fig. 2b), which implies a prominent (and potentially

139  specificity-conferring) role in the VP1-$F_P$ interaction.

140

141  <u>ShPyV VP1 has a nanomolar affinity towards Forssman pentaose-decorated surfaces</u>

142

143  After assessing the complex formation between ShPyV VP1 and $F_P$ in the context of

144  the solid-supported glycan array and solution NMR, we sought to evaluate the

145  underlying binding kinetics by surface plasmon resonance (SPR) and compare the

146  performance of VP1 with the albumen gland agglutinin of *Helix pomatia* (HPA), an

147  αGalNAc-specific lectin with high affinity towards the Forssman antigen[53,54]. Therefore,

148  we prepared an SPR chip decorated with $F_P$ (see Fig. S3) and measured binding

149  responses using different dilutions of VP1 and HPA. The resulting response curves

150  show that within the given association time the pentavalent VP1 and hexavalent HPA

151  bind to the immobilized $F_P$ without reaching a complex formation equilibrium, which

152  suggests that the proteins do not readily dissociate from the surface (Fig. 3a). Indeed,

153  fitting of the kinetic parameters results in extremely low dissociation rates ($k_{off}$) for both

154  proteins, which is a common phenomenon for multivalent binding systems. However,

155  the curves of VP1 result in a higher rate for complex formation ($k_{on}$), determining an

156  overall approximately three-fold lower dissociation constant ($K_D$) compared to HPA

157  (Fig. 3a).

158

159  <u>ShPyV VP1 specifically binds to Forssman-positive erythrocytes</u>

160

161  To confirm the functional relevance of the VP1-F$_P$ complex *in vitro*, we performed cell

162  binding assays using different erythrocytes and compare the cell binding of VP1 with

163  commercial reagents for Forssman antigen detection, i.e., HPA and the monoclonal

164  anti-FA IgM antibody M1/87[55]. We incubated VP1 with Forssman-positive red blood

165  cells (RBCs) from sheep and human RBCs of the different ABO phenotypes and found

166  that ShPyV VP1 bound sheep RBCs with saturation but did not attach to any of the

167  human erythrocytes (Fig. 3b left). In contrast, the αGalNAc-specific lectin HPA bound

168  to the FA on sheep RBCs and to the antigens on human blood group A1 and A2

169  RBCs[56] (Fig. 3b left). To confirm that the carbohydrate portions of the antigens account

170  for binding, we incubated the two proteins with excessive amounts of the terminal

171  trisaccharides (trioses) of FA (GalNAcα1-3GalNAcβ1-3Gal) and A substance

172  (GalNAcα1-3[Fucα1-2]Gal) and assessed the remaining sheep RBCs attachment. As

173  predicted, FA triose blocked VP1 and HPA, while A antigen triose could only inhibit

174  HPA attachment (Fig. 3b right). Lastly, we determined the effective protein

175  concentrations (EC50, see Methods) for sheep RBCs attachment using different

176  dilutions of ShPyV VP1, HPA, and M1/87. Evaluation of the response curves yielded

177  the lowest EC50 for ShPyV VP1 with a value of 1.2 nM, followed by M1/87 and HPA

178  with 3.3 nM and 5.2 nM, respectively (Fig. 3c). Collectively, these results show the

179  remarkable specificity and affinity of ShPyV VP1 towards the Forssman antigen.

180

108

181    <u>Crystal structure of Forssman pentaose bound to VP1 reveals an intricate network of</u>

182    <u>interactions</u>

183

184    To characterize the binding of Forssman antigen by ShPyV VP1 with atomic detail, we

185    solved the crystal structure of the VP1 in complex with 10 mM of Forssman pentaose

186    ($F_P$) at 1.92 Å resolution (Table S4). Initial electron density calculation using

187    unliganded ShPyV VP1 model (PDB 6Y61) shows that five molecules of the

188    oligosaccharide (GalNAcα1-3GalNAcβ1-3Galα1-4Galβ1-4Glc) associate with the

189    outer surface of the pentameric VP1, at a position that does not overlap with the

190    previously described sialic acid binding site[43] (Fig. S5b), in a five-fold symmetric array.

191    All five Forssman binding sites are identical and located between two neighboring VP1

192    monomers (Fig. 4a). In the binding sites without artificial crystal contacts, electron

193    density provides an unambiguous outline of the last four monosaccharides of $F_P$ (Fig.

194    4b). The intermolecular contacts show that the terminal trisaccharide of $F_P$ contributes

195    to binding, whereas the remaining βGal and Glc divert away from VP1 and do not

196    interact with the protein surface. Closer inspection of the binding site reveals that the

197    methyl group of the terminal αGalNAc invades a hydrophobic surface depression

198    constituted by phenylalanine residues of two neighboring VP1 monomers, establishing

199    van der Waals interactions (Fig. 4c). Additionally, the *N*-acetyl moiety of αGalNAc

200    interacts with the Q266 and F66cw residues via hydrogen bonds (Fig. 4c). In βGalNAc,

201    the carbonyl oxygen forms two hydrogen bonds with the side chains of Q266 and N268

202    (Fig. 4c). Apart from these contacts, numerous interactions involve the hydroxyl groups

203    of $F_P$ that are directed towards the protein surface. An overview of all contacts and their

204    distances is provided in Table 2. We also solved the crystal structure of VP1 in complex

205    with the FA glycan precursor, globo-*N*-tetraose, which reveals negligible binding only

206     (Fig. S5). We therefore conclude that the terminal GalNAcα1-3GalNAc disaccharide is

207     essential for the high affinity of the ShPyV VP1-F$_P$ complex.

208

209     **Discussion**

210

211     Sheep polyomavirus (ShPyV) was originally identified as a genomic contaminant in

212     processed lamb meat, and it remains unknown if sheep are the natural hosts of this

213     virus[30]. Our functional data show that ShPyV engages Forssman antigen (FA)-

214     decorated cells with high affinity and specificity, making it plausible that FA initiates

215     sheep polyomavirus infection. However, in sheep the FA was thus far only detected on

216     erythrocytes but not on leukocytes, bone marrow or other tissues[29,57,58]. However, as

217     the studies about the distribution of Forssman antigen in sheep tissues are old, the

218     more sensitive techniques available today could help to elucidate the potential targets

219     cells for viral replication as well as the precise role of Forssman antigen during ShPyV

220     infection.

221

222     The antigenic determinant of Forssman antigen is the GalNAcα1-3GalNAcβ

223     disaccharide. In sheep erythrocytes it is predominantly linked to globotriasolyceramide,

224     giving rise to the linear Forssman pentaosylceramide (see Table 1), but it also appears

225     on the (unbranched) glycans of neolacto-series or Galili glycolipids[29]. In other animals,

226     GalNAcα1-3GalNAcβ occurs on branched structures such as extended H-like antigens

227     (dog gastric mucosa and mucus)[59,60], GM1 ganglioside (English sole fish)[61], isoglobo-

228     ganglio-neolacto glycans (swine)[62], as well as in the form of mucin core 5 (GalNAcα1-

229     3[NeuAcα2-6]GalNAc) in bovine submaxillary mucin[63]. We believe that ShPyV VP1

230     can bind a range of different Forssman glycans, providing that the structural deviations

231     do not result in clashes with the protein. By the assessment of our ShPyV VP1-

110

232　Forssman pentaose ($F_P$) complex structure, the designated minimal glycan binding

233　motif is the GalNAcα1-3GalNAc disaccharide with few additional contacts to the

234　subsequent αGal. From structural perspective, this epitope recognition clearly

235　disallows branching of the glycan at certain positions, e.g., O3 of αGalNAc, N2 of

236　βGalNAc or O4 of αGal. For example, in the blood group A glycan fucose replaces the

237　*N*-acetyl group of the βGalNAc, which would result in overlaps with the protein and

238　accounts for the lack of binding (Fig. S5b). This feature distinguishes VP1 from many

239　other Forssman glycan-binding proteins such as Galectin-9, *Helix pomatia* agglutinin,

240　*Sinularia lochmodes* lectin-2, and *Dolichos biflorus* agglutinin, where recognition is

241　based on the subterminal or terminal GalNAc (see Fig. S6), resulting in a much broader

242　glycan binding capacity[64–67].

243

244　Previously we have shown that ShPyV VP1 interacts with sialylated trisaccharides in

245　crystal structures[43]. Also, our new results from STD NMR and crystallography indicate

246　that sialyl and Forssman glycan binding occur independently from each at separate

247　sites of the protein surface (Fig. S5b). Yet, in our flow cytometry assays we did not

248　detect any binding to human erythrocytes although they are known to be decorated

249　with sialic acids. However, the preferred form of sialic acid for ShPyV VP1, as evinced

250　by our glycan array data, is 9-*O*-acetylated *N*-acetylneuraminic acid (Neu5,9Ac$_2$),

251　which in human erythrocytes is present in too low amounts to result in detectable

252　binding[68]. These findings imply that human cells are incompatible targets for the sheep

253　virus.

254

255　Humans are considered a largely Forssman-negative species and the presence of

256　Forssman antigenicity is mostly associated with human cancer[69]. During the last

257　decade, the elucidation of the genetic basis behind FA reactivation and the

258     establishment of the human FORS blood group underline the continual relevance of

259     Forssman antigen[5,6]. We propose that ShPyV VP1 complements and expands the

260     available resources for the comprehensive detection of Forssman antigen in cellular

261     and histological samples of healthy or malignant provenance. We also anticipate

262     refining mutant VP1 protein to precisely target the Forssman antigen for biomedical

263     use.

264

265     **Summary**

266

267     We show that the Forssman antigen serves as a cellular attachment receptor for the

268     major capsid protein VP1 of ShPyV. After initial detection by glycan array screening,

269     cell binding assays imply functional relevance. Structural analysis of VP1 in complex

270     with the carbohydrate moiety of FA, Forssman pentaose ($F_P$), characterizes the

271     determinants of the underlying interactions with high accuracy. Our findings have

272     significance for the fields of structural virology, glycobiology, and pathobiology as

273     follows:

274     (i) We report the first structural and functional data of a viral entity to specifically

275     interact with the Forssman antigen.

276     (ii) The ShPyV VP1 – $F_P$ complex provides the first detailed information on how

277     a polyomavirus binds to a non-sialylated ligand.

278     (iii) The high specificity and affinity of ShPyV VP1 toward Forssman antigen

279     recommends this protein as a potential biomedical reagent.

280

281     **Materials and methods**

282

283     <u>Protein expression and purification</u>

112

284 The VP1 of sheep polyomavirus was recombinantly produced as previously.[43] Briefly,

285 transformed *E. coli* BL21 (DE3) were grown in LB (Miller) medium supplemented with

286 100 µg/ml ampicillin at 37 °C. Expression was induced at an $OD_{600}$ of 0.7 with 400 µM

287 IPTG and carried out at 20 °C for 16 h. ShPyV VP1 was extracted from bacterial pellets

288 by sonification and subsequent centrifugation (17,000 rpm at 4 °C). The $His_6$-tagged

289 VP1 was purified by immobilized-metal affinity chromatography (IMAC) and gel

290 filtration (20 mM HEPES pH 7.5, 150 mM NaCl). For crystallization of ShPyV VP1 the

291 $His_6$-tag was cleaved off by incubation with 10 U/mg thrombin overnight at room

292 temperature. Tags and thrombin were separated from ShPyV VP1 by IMAC and gel

293 filtration. If not immediately used, the protein was stored at -20 °C.

294

295 <u>Glycan array</u>

296 The $His_6$-tagged ShPyV VP1 was analyzed in the neoglycolipid (NGL)-based

297 microarray system[70]. A broad-spectrum screening glycan microarray set

298 encompassing 672 sequence-defined lipid-linked glycan probes was used (the glycan

299 probes included, and their sequences are given in Supplemental Table S1). Details of

300 the preparation of the glycan probes and the generation of the microarrays are in

301 Supplementary Glycan Microarray Document (Supplemental Table S2) in accordance

302 with the MIRAGE (Minimum Information Required for A Glycomics Experiment)

303 guidelines for reporting of glycan microarray-based data[71]. The microarray analyses

304 were performed essentially as described[47,48]. In brief, after blocking the slides for 1h

305 with HBS buffer (10 mM HEPES, pH 7.4, 150 mM NaCl) containing 0.33% (w/v) blocker

306 Casein (Pierce), 0.3% (w/v) Bovine Serum Albumin (Sigma-Aldrich) and 5 mM $CaCl_2$,

307 the microarray was overlaid with the VP1 protein for 90 minutes as a protein-antibody

308 complex that was prepared by preincubating VP1 with mouse monoclonal anti-

309 polyhistidine and biotinylated anti-mouse IgG antibodies (both from Sigma) at a ratio

310  of 4:2:1 (by weight) and diluted in the blocking solution to provide a final VP1

311  concentration of 150 μg/mL. Binding was detected with Alexa Fluor-647-labelled

312  streptavidin (Molecular Probes) at 1 μg/mL for 30 minutes. Apart from the protein-

313  antibody precomplexation step, which was performed on ice, all other steps were

314  carried out at ambient temperature. Microarray imaging and data analysis are

315  described in the Supplementary MIRAGE document (Supplemental Table S2).

316

317  <u>NMR</u>

318  All nuclear magnetic resonance (NMR) data were collected in a Bruker AVIII-600 MHz

319  spectrometer equipped with a TXI-z probe head at 298 K using 3 mm NMR tubes.

320  Samples were provided in 99 % $D_2O$-PBS containing 20 mM $KH_2PO_4/K_2HPO_4$ (pH 7.4)

321  and 150 mM NaCl. $^1H$ reference spectra of isolate oligosaccharides samples were

322  recorded using pure solutions of 4 mM of Forssman pentaose and 2 mM of 6'SL.

323  Subsequently, the $^1H$ oligosaccharide resonances of a sample containing 1 mM

324  Forssman pentaose and 1 mM 6'SL in the presence of 20 μM ShPyV VP1 were

325  collected under suppression of protein signals. From the same sample, the saturation

326  difference transfer (STD) NMR experiment was performed using previously published

327  pulse protocols[43]. NMR spectra were processed using TOPSPIN 4 (Bruker). The

328  carbohydrates used here were purchased from Biosynth, UK.

329

330  <u>VP1 mutagenesis</u>

331  Site-directed mutagenesis of VP1 was performed in samples of 20 μl containing

332  1x ReproFast amplification buffer (Genaxxon), 20 ng template vector DNA, 50 ng

333  primers, 2 mM dNTP, and 1 unit of ReproFast polymerase (Genaxxon). Initial

334  denaturation was performed at 95 °C for 2 minutes, followed by 18 cycles of 95 °C for

335  1 minute (denaturation), 55-70 °C for 1 minute (annealing), and 72 °C for 6 minutes

114

336    (elongation). Subsequently, DNA synthesis was completed at 72 °C for 10 minutes.

337    Parental template DNA was digested with 2 units of DpnI for two hours at 37 °C. The

338    PCR product was directly transformed into *E. coli* DH5α.

339

340    DNA primers for the site-directed mutagenesis of ShPyV VP1 were ordered from

341    ThermoFisher Scientific and are displayed from 5'- to 3' ends:

342

343    S95C_fwd:          GCTGAACCAAGACATGACCTGCGATACCATCCTGATGTGGGAGG
344    S95C_rev:          CCTCCCACATCAGGATGGTATCGCAGGTCATGTCTTGGTTCAGC

345

346    <u>Fluorescence labeling</u>

347    VP1 was expressed and purified as described above. Surficial cysteine residues were

348    reduced in degassed gel filtration buffer (20 mM HEPES pH 7.4, 150 mM NaCl)

349    supplemented with 80 mM DTT at 4 °C overnight. DTT was removed from the protein

350    solution using a PD-10 desalting column (Cytiva) with DTT-free gel filtration buffer. A

351    20-fold molar excess of cysteine reactive AFDye 488 Maleimide (Fluoroprobes) was

352    added to the protein solutions (0.5 - 0.7 mg/ml) and the reaction was incubated in the

353    dark at 4 °C for 16-20 h. Excess dye was quenched and removed by desalting in gel

354    filtration buffer supplemented with 10 mM DTT. VP1-488 conjugate was eluted in the

355    same buffer. The degree of labeling was photometrically determined according to the

356    manufacturer's protocol (Fluoroprobes).

357

358    <u>Flow cytometry</u>

359    Defibrinated blood from sheep was purchased from TCS Biosciences Ltd and

360    BioTrading Benelux B.V., and human reference erythrocytes from Immucor. Prior to

361    analysis, all blood was diluted to 0.25 % in 1x PBS + 2 % fetal calf serum (FC buffer).

362    For staining, labeled protein was added to samples of $\sim 1 \times 10^6$ cells in a total volume of

363     100 µL. After incubation in the dark for 15 minutes, cells were precipitated by

364     centrifugation (1.5 min at 1900 rpm) and resuspended in fresh FC buffer. Unstained

365     reference samples were prepared accordingly without the addition of labeled protein.

366     For cell attachment inhibition assays, proteins were incubated with oligosaccharides

367     (Forssman antigen trisaccharide: Elicityl, France; Blood group A trisaccharide:

368     Biosynth, UK) for 5 minutes prior to their addition to the cellular samples (~$1 \times 10^6$ cells

369     in 100 µL). To determine the concentration for half maximal cell attachment (effective

370     concentration, EC50), samples comprising serially diluted concentrations of labeled

371     protein, starting from 10 µg/mL for M1/87-FITC (Santa Cruz Biotechnology) and HPA

372     Alexa Fluor$^{TM}$ 488 conjugate (Invitrogen), and 5 µg/mL for VP1-488, were prepared as

373     described above. Here, the wash step was omitted to guarantee consistent protein

374     concentrations and to prevent agglutination. All samples were transferred to the flow

375     cytometer (LSR II, BD Biosciences) and fluorescent fractions were captured at 520 nM.

376     100,000 events were recorded for all samples except those with agglutination, in which

377     case 10,000-20,000 events were recorded. All data were analyzed using FACSdiva

378     software (BD Biosciences). EC50 values were calculated using R[72].

379

380     <u>Surface plasmon resonance</u>

381     Assessment of the kinetic binding parameters was performed via surface plasmon

382     resonance (SPR) in a Biacore X-100 system. Using an adapted protocol[73], BSA

383     decorated with Forssman pentaose (~76 kDa) as the ligand was prepared via reductive

384     amination (see Fig. S3). After rebuffering into a 10 mM acetate buffer at pH 4,

385     approximately 100 response units (RU) of ligand were immobilized onto the CM5

386     sensor chip surface using amine coupling as described in the manufacturer's protocol

387     (Cytiva). For the reference flow cell, native BSA (Sigma-Aldrich) was immobilized

388     similarly. Purified VP1 (162 kDa) and native HPA (79 kDa, Sigma-Aldrich) buffered in

389  SPR assay buffer (20 mM HEPES pH 7.4, 150 mM NaCl, 0.005 % Polysorbate 20)

390  were used as analytes in different concentrations between 0.2 nM and 20 µM. All

391  measurements were performed at 25 °C. During binding (120 s) and dissociation

392  (180 s), a flow rate of 30 µL/min was applied. After each cycle, the surfaces were

393  regenerated applying 10 mM glycine pH 1.5 for 120 s. Experimental curves were

394  interpreted using BIAevaluation software (Cytiva).

395

396  <u>Crystallization and data collection</u>

397  Purified ShPyV VP1 was concentrated to 3.5 mg/ml. Crystallization droplets comprised

398  1 µl protein and 1 µl reservoir solution (150 mM KSCN and 20 % (w/v) PEG 3,350).

399  The droplets were equilibrated against the 500 µl reservoir using the hanging drop

400  vapor diffusion method at 20 °C. For ligand derivatization, VP1 crystals were soaked

401  in crystallization solution supplemented with the respective oligosaccharides (10 mM

402  Forssman pentaose ($F_P$), 10 mM globo-*N*-tetraose, 5 mM $F_P$ + 20 mM 3'SLN, 5 mM $F_P$

403  + 20 mM 6'SLN, 20 mM A antigen trisaccharide (all purchased from Biosynth, UK), and

404  20 mM Tf antigen (TCI Europe), at 20 °C for half an hour. Prior to flash freezing, crystals

405  were transferred to oligosaccharide soaking solutions supplemented with 20 % (v/v)

406  MPD for cryo-protection. X-ray diffraction experiments were performed at the Swiss

407  Light Source (SLS) beamline X06DA (PSI in Villigen, Switzerland). Data were collected

408  at a wavelength of 1 Å with an excitation time and angle of 0.1 s and 0.1° per image.

409

410  <u>Structure determination and refinement</u>

411  X-ray diffraction images were integrated and scaled using *XDS*[74]. For each data set,

412  2,000-3,000 reflections were reserved as a test set for structure validation. Phases

413  were determined by molecular replacement using the native sheep polyomavirus VP1

414  structure as search model (PDB 6Y61) in Phaser (CCP4)[75]. Phase refinement and

*Appendices*

415  model building were performed until validation parameters converged using Phenix

416  and Refmac5, and Coot, respectively[76–78]. Oligosaccharides conformations are in

417  accordance with the restraints from the CCP4 libraries. All protein structure

418  representations were rendered in PyMOL (Schrödinger).

419

420  **Data availability**

421  Atomic coordinates of ShPyV VP1 in complex with carbohydrate ligands are deposited

422  in the RCSB Protein Data Bank (PDB) with the accession codes 7B6S (Forssman

423  pentaose, $F_P$), 7B6T (globo-*N*-tetraose), 7B6U ($F_P$ + 6'SLN), and 7B6V ($F_P$ + 3'SLN).

424  Other data can be provided by the authors upon request.

425

426  **References**

427  1.  B Siddiqui, S. H. A revised structure for the Forssman glycolipid hapten. *J Biol Chem* **246**,
428      5766–5769 (1971).
429  2.  Fraser, B. A. & Mallette, M. F. Structure of forssman hapten glycosphingolipid from sheep
430      erythrocytes. *Immunochemistry* **11**, 581–593 (1974).
431  3.  Haslam, D. B., Baenziger, J. U., Baenzigert, J. U. & Baenziger, J. U. Expression cloning of
432      Forssman glycolipid synthetase: A novel member of the histo-blood group ABO gene family.
433      *Proc. Natl. Acad. Sci. U. S. A.* **93**, 10697–10702 (1996).
434  4.  Xu, H., Storch, T., Yu, M., Elliott, S. P. & Haslam, D. B. Characterization of the human
435      Forssman synthetase gene. An evolving association between glycolipid synthesis and host-
436      microbial interactions. *J. Biol. Chem.* **274**, 29390–29398 (1999).
437  5.  Yamamoto, M., Cid, E. & Yamamoto, F. Molecular genetic basis of the human Forssman
438      glycolipid antigen negativity. *Sci. Rep.* **2**, 975 (2012).
439  6.  Svensson, L. *et al.* Forssman expression on human erythrocytes: Biochemical and genetic
440      evidence of a new histo-blood group system. *Blood* **121**, 1459–1468 (2013).
441  7.  Storry, J. R. *et al.* International Society of Blood Transfusion Working Party on red cell
442      immunogenetics and blood group terminology: Cancun report (2012). *Vox Sang.* **107**, 90–96
443      (2014).
444  8.  Hult, A. K. & Olsson, M. L. The FORS awakens: review of a blood group system reborn.
445      *Immunohematology* **33**, 64–72 (2017).
446  9.  Ono, K. *et al.* Expression of Forssman antigen in human large intestine. *J. Histochem.*
447      *Cytochem.* **42**, 659–665 (1994).
448  10. Jesus, C. *et al.* Prevalence of antibodies to a new histo-blood system: the FORS system. *Blood*
449      *Transfus.* **16**, 178 (2018).
450  11. Hult, A. K., McSherry, E., Möller, M. & Olsson, M. L. GBGT1 is allelically diverse but
451      dispensable in humans and naturally occurring anti-FORS1 shows an ABO-restricted pattern.
452      *Transfusion* **58**, 2036–2045 (2018).
453  12. Kawanami, J. The appearance of Forssman hapten in human tumor. *J. Biochem.* **72**, 783–785
454      (1972).
455  13. Yoda, Y., Ishibashi, T. & Makita, A. Isolation, Characterization, and Biosynthesis of Forssman
456      Antigen in Human Lung and Lung Carcinoma. *J. Biochem.* **88**, 1887–1894 (1980).
457  14. Taniguchi, N., Yokosawa, N., Narita, M., Mitsuyama, T. & Makita, A. Expression of forssman
458      antigen synthesis and degradation in human lung cancer. *J. Natl. Cancer Inst.* **67**, 577–583
459      (1981).

118

460  15.  Yokota, M., Warner, G. A. & Hakomori, S. Blood Group A-like Glycolipid and a Novel Forssman
461      Antigen in the Hepatocarcinoma of a Blood Group O Individual. *Cancer Res.* **41**, (1981).
462  16.  Mori, T., Sudo, T. & Kano, K. Expression of Heterophil Forssman Antigen on Cultured
463      Malignant Cell Lines. *JNCI J. Natl. Cancer Inst.* **70**, 811–818 (1983).
464  17.  Hakomori, S. Tumor-associated carbohydrate antigens. *Annu. Rev. Immunol.* **2**, 103–126
465      (1984).
466  18.  Uemura, K., Hattori, H., Ono, K., Ogata, H. & Taketomi, T. Expression of Forssman glycolipid
467      and blood group-related antigens A, Le(x) and Le(y) in human gastric cancer and in fetal
468      tissues. *Jpn. J. Exp. Med.* **59**, 239–249 (1989).
469  19.  Säljö, K., Thornell, A., Jin, C., Norlén, O. & Teneberg, S. Characterization of human medullary
470      thyroid carcinoma glycosphingolipids identifies potential cancer markers. *Int. J. Mol. Sci.* **22**,
471      10463 (2021).
472  20.  Ariga, T., Yoshida, T., Mimori, T. & Yu, R. K. Autoantibodies against Forssman glycolipids in
473      Graves' disease and Hashimoto's thyroiditis. *Clin. Exp. Immunol.* **86**, 483–488 (1991).
474  21.  Hakomori, S., Wang, S. M. & Young, W. W. Isoantigenic expression of Forssman glycolipid in
475      human gastric and colonic mucosa: Its possible identity with 'A like antigen' in human cancer.
476      *Proc. Natl. Acad. Sci. U. S. A.* **74**, 3023–3027 (1977).
477  22.  Stromberg, N. *et al.* Host-specificity of uropathogenic Escherichia coli depends on differences
478      in binding specificity to Galα1-4Gal-containing isoreceptors. *EMBO J.* **9**, 2001–2010 (1990).
479  23.  Stromberg, N., Nyholm, P. G., Pascher, I. & Normark, S. Saccharide orientation at the cell
480      surface affects glycolipid receptor function. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 9340–9344
481      (1991).
482  24.  Johanson, I., Lindstedt, R. & Svanborg, C. Roles of the pap- and prs-encoded adhesins in
483      Escherichia coli adherence to human uroepithelial cells. *Infect. Immun.* **60**, 3416–3422 (1992).
484  25.  Lanne, B. *et al.* Glycoconjugate receptors for P-fimbriated Escherichia coli in the mouse: An
485      animal model of urinary tract infection. *J. Biol. Chem.* **270**, 9017–9025 (1995).
486  26.  Elliott, S. P., Yu, M., Xu, H. & Haslam, D. B. Forssman synthetase expression results in
487      diminished Shiga toxin susceptibility: A role for glycolipids in determining host-microbe
488      interactions. *Infect. Immun.* **71**, 6543–6552 (2003).
489  27.  Müthing, J. *et al.* Promiscuous Shiga toxin 2e and its intimate relationship to Forssman.
490      *Glycobiology* **22**, 849–862 (2012).
491  28.  Nasir, W., Nilsson, J., Olofsson, S., Bally, M. & Rydell, G. E. Parvovirus B19 VLP recognizes
492      globoside in supported lipid bilayers. *Virology* **456–457**, 364–369 (2014).
493  29.  Santos, L. *et al.* Characterization of sheep erythrocyte glycosphingolipids recognized by human
494      anti-Forssman antibodies. *Glycobiology* (2020) doi:10.1093/glycob/cwaa032.
495  30.  Buck, C. B. *et al.* The ancient evolutionary history of polyomaviruses. *PLoS Pathog.* **12**,
496      1005574 (2016).
497  31.  Fried, H., Cahan, L. D. & Paulson, J. C. Polyoma virus recognizes specific
498      sialyloligosaccharide receptors on host cells. *Virology* **109**, 188–192 (1981).
499  32.  Tsai, B. *et al.* Gangliosides are receptors for murine polyoma virus and SV40. *EMBO J.* **22**,
500      4346–4355 (2003).
501  33.  Smith, A. E., Lilie, H. & Helenius, A. Ganglioside-dependent cell attachment and endocytosis of
502      murine polyomavirus-like particles. *FEBS Lett.* (2003) doi:10.1016/S0014-5793(03)01220-1.
503  34.  Erickson, K. D., Garcea, R. L. & Tsai, B. Ganglioside GT1b is a putative host cell receptor for
504      the Merkel cell polyomavirus. *J. Virol.* **83**, 10275–10279 (2009).
505  35.  Gilbert, J. *et al.* Ganglioside GD1a Restores Infectibility to Mouse Cells Lacking Functional
506      Receptors for Polyomavirus. *J. Virol.* **79**, 615–618 (2005).
507  36.  Blaum, B. S. & Stehle, T. Sialic acids in nonenveloped virus infections. in *Advances in
508      Carbohydrate Chemistry and Biochemistry* vol. 76 65–111 (Academic Press Inc., 2018).
509  37.  Elphick, G. F. *et al.* The human polyomavirus, JCV, uses serotonin receptors to infect cells.
510      *Science (80-. ).* **306**, 1380–1383 (2004).
511  38.  Assetta, B. *et al.* 5-HT2 receptors facilitate JC polyomavirus entry. *J. Virol.* **87**, 13490 (2013).
512  39.  Schowalter, R. M., Pastrana, D. V. & Buck, C. B. Glycosaminoglycans and sialylated glycans
513      sequentially facilitate merkel cell polyomavirus infectious entry. *PLoS Pathog.* **7**, 1002161
514      (2011).
515  40.  Hurdiss, D. L., Frank, M., Snowden, J. S., Macdonald, A. & Ranson, N. A. The structure of an
516      infectious human polyomavirus and its interactions with cellular receptors. *Structure* **26**, 839
517      (2018).
518  41.  Geoghegan, E. M. *et al.* Infectious entry and neutralization of pathogenic JC polyomaviruses.
519      *Cell Rep.* **21**, 1169–1179 (2017).
520  42.  Assetta, B. *et al.* Genetic and functional dissection of the role of individual 5-HT2 receptors as
521      entry receptors for JC polyomavirus. *Cell Rep.* **27**, 1960-1966.e6 (2019).

*Appendices*

| | | |
|---|---|---|
| 522<br>523 | 43. | Ströh, L. J. *et al.* Structural basis and evolution of glycan receptor specificities within the polyomavirus family. *MBio* **11**, 1–21 (2020). |
| 524<br>525<br>526 | 44. | Palma, A. S., Feizi, T., Childs, R. A., Chai, W. & Liu, Y. The neoglycolipid (NGL)-based oligosaccharide microarray system poised to decipher the meta-glycome. *Curr. Opin. Chem. Biol.* **18**, 87–94 (2014). |
| 527<br>528 | 45. | Campanero-Rhodes, M. A. *et al.* N-glycolyl GM1 ganglioside as a receptor for simian virus 40. *J. Virol.* **81**, 12846–12858 (2007). |
| 529<br>530 | 46. | Neu, U. *et al.* Structure-function analysis of the human JC polyomavirus establishes the LSTc pentasaccharide as a functional receptor motif. *Cell Host Microbe* **8**, 309–319 (2010). |
| 531<br>532 | 47. | Neu, U. *et al.* A structure-guided mutation in the major capsid protein retargets BK polyomavirus. *PLoS Pathog.* **9**, 1–13 (2013). |
| 533<br>534<br>535 | 48. | Khan, Z. M. *et al.* Crystallographic and glycan microarray analysis of human polyomavirus 9 VP1 identifies N-glycolyl neuraminic acid as a receptor candidate. *J. Virol.* **88**, 6100–6111 (2014). |
| 536<br>537 | 49. | Sorin, M. N. *et al.* Structural and functional analysis of natural capsid variants suggests sialic acid-independent entry of BK polyomavirus. *CellReports* **42**, 112114 (2023). |
| 538<br>539<br>540 | 50. | Mayer, M. & Meyer, B. Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J. Am. Chem. Soc.* (2001) doi:10.1021/ja0100120. |
| 541<br>542<br>543<br>544 | 51. | Blaum, B. S., Neu, U., Peters, T. & Stehle, T. Spin ballet for sweet encounters: Saturation-transfer difference NMR and X-ray crystallography complement each other in the elucidation of protein-glycan interactions. *Acta Crystallographica Section F: Structural Biology Communications* vol. 74 451–462 (2018). |
| 545<br>546<br>547 | 52. | Grönberg, G., Nilsson, U., Bock, K. & Magnusson, G. Nuclear magnetic resonance and conformational investigations of the pentasaccharide of the Forssman antigen and overlapping di-, tri-, and tetra-saccharide sequences. *Carbohydr. Res.* **257**, 35–54 (1994). |
| 548<br>549<br>550 | 53. | Wu, A. M. & Sugii, S. Coding and classification of D-galactose, N-acetyl-D-galactosamine, and β-D-Galp-[1→3(4)]-β-D-GlcpNAc, specificities of applied lectins. *Carbohydr. Res.* **213**, 127–143 (1991). |
| 551<br>552 | 54. | Sanchez, J. F. *et al.* Biochemical and structural analysis of Helix pomatia agglutinin. A hexameric lectin with a novel fold. **281**, 20171–20180 (2006). |
| 553<br>554<br>555 | 55. | Springer, T., Galfrè, G., Secher, D. S. & Milstein, C. Monoclonal xenogeneic antibodies to murine cell surface antigens: identification of novel leukocyte differentiation antigens. *Eur. J. Immunol.* **8**, 539–551 (1978). |
| 556<br>557 | 56. | Hauser, R., Fechner, G. & Brinkmann, B. A1 and A2 blood group substances: are there structural differences? *Z. Rechtsmed.* **103**, 587–591 (1990). |
| 558<br>559 | 57. | Tanaka, N. & Leduc, E. H. A study of the cellular distribution of forssman antigen in various species. *J. Immunol.* **77**, 198–212 (1956). |
| 560<br>561 | 58. | Koizumi, N., Hara, A., Uemura, K. & Taketomi, T. Glycosphingolipids in sheep liver, kidney, and various blood cells. *Jpn. J. Exp. Med.* **58**, 21–31 (1988). |
| 562<br>563 | 59. | Slomiany, B. L. & Slomiany, A. Forssman glycolipid variants of dog gastric mucosa. Structure of a branched ceramide octasaccharide. *Eur. J. Biochem.* **83**, 105–111 (1978). |
| 564<br>565<br>566 | 60. | Slomiany, B. L., Banas-Gruszka, Z., Zdebska, E. & Slomiany, A. Characterization of the Forssman-active oligosaccharides from dog gastric mucus glycoprotein isolated with the use of a monoclonal antibody. *J. Biol. Chem.* **257**, 9561–9565 (1982). |
| 567<br>568<br>569<br>570 | 61. | Ostrander, G. K., Levery, S. B., Hakomori, S. & Holmes, E. H. Isolation and characterization of the major acidic glycosphingolipids from the liver of the English sole (Parophrys vetulus). Presence of a novel ganglioside with a Forssman antigen determinant. *J. Biol. Chem.* **263**, 3103–3110 (1988). |
| 571<br>572<br>573 | 62. | Yamamoto, H. *et al.* Isolation and characterization of a novel Forssman active acidic glycosphingolipid with branched isoglobo-, ganglio-, and neolacto-series hybrid sugar chains. *J. Biochem.* **125**, 923–930 (1999). |
| 574<br>575<br>576 | 63. | Savage, A. V., Donoghue, C. M., D'Arcy, S. M., Koeleman, C. A. M. & van den Eijnden, D. H. Structure determination of five sialylated trisaccharides with core types 1, 3 or 5 isolated from bovine submaxillary mucin. *Eur. J. Biochem.* **192**, 427–432 (1990). |
| 577<br>578<br>579 | 64. | Hamelryck, T. W. *et al.* Carbohydrate binding, quaternary structure and a novel hydrophobic binding site in two legume lectin oligomers from Dolichos biflorus. *J. Mol. Biol.* **286**, 1161–1177 (1999). |
| 580<br>581 | 65. | Lescar, J. *et al.* Structural basis for recognition of breast and colon cancer epitopes Tn antigen and Forssman disaccharide by Helix pomatia lectin. *Glycobiology* **17**, 1077–1083 (2007). |
| 582<br>583 | 66. | Nagae, M. *et al.* Structural analysis of the human galectin-9 N-terminal carbohydrate recognition domain reveals unexpected properties that differ from the mouse orthologue. *J.* |

120

584      *Mol. Biol.* **375**, 119–135 (2008).

585   67.  Kita, A. *et al.* Crystal structure of octocoral lectin SLL-2 complexed with Forssman antigen
586       tetrasaccharide. *Glycobiology* **27**, 696–700 (2017).

587   68.  Bulai, T., Bratosin, D., Pons, A., Montreuil, J. & Zanetta, J. P. Diversity of the human
588       erythrocyte membrane sialic acids in relation with blood groups. *FEBS Lett.* **534**, 185–189
589       (2003).

590   69.  Cid, E., Yamamoto, M. & Yamamoto, F. Mixed-Up Sugars: Glycosyltransferase Cross-
591       Reactivity in Cancerous Tissues and Their Therapeutic Targeting. *ChemBioChem* **23**,
592       e202100460 (2022).

593   70.  Liu, Y. *et al.* Neoglycolipid-based oligosaccharide microarray system: Preparation of ngls and
594       their noncovalent immobilization on nitrocellulose-coated glass slides for microarray analyses.
595       *Methods Mol. Biol.* **808**, 117–136 (2012).

596   71.  Liu, Y. *et al.* The minimum information required for a glycomics experiment (MIRAGE) project:
597       improving the standards for reporting glycan microarray-based data. *Glycobiology* **27**, 280–284
598       (2017).

599   72.  Ritz, C., Baty, F., Streibig, J. C. & Gerhard, D. Dose-response analysis using R. *PLoS One* **10**,
600       e0146021 (2015).

601   73.  Gildersleeve, J. C., Oyelaran, O., Simpson, J. T. & Allred, B. Improved procedure for direct
602       coupling of carbohydrates to proteins via reductive amination. *Bioconjug. Chem.* **19**, 1485–
603       1490 (2008).

604   74.  Kabsch, W. *et al.* XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2010).

605   75.  McCoy, A. J. *et al. Phaser* crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

606   76.  Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and
607       electrons: Recent developments in Phenix. *Acta Crystallogr. Sect. D Struct. Biol.* (2019)
608       doi:10.1107/S2059798319011471.

609   77.  Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures.
610       *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2011) doi:10.1107/S0907444911001314.

611   78.  Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr.*
612       *Sect. D Biol. Crystallogr.* (2004) doi:10.1107/S0907444904019158.

613

614

615

616 **Acknowledgements**

628     We thank the Max-Planck Institute for Biology in Tübingen for access to the NMR

629     facility and Dr. Vincent Truffault for his assistance in NMR data acquisition. We also

630     thank Prof. Barbara Mulloy (Imperial College London) for critical review of the

631     manuscript. We thank the lab of Prof. Schulze-Osthoff at the University of Tübingen for

632     the access to the flow cytometry appliances and especially Dr. Anja Schmitt for her

633     help with the experiments.

634

635     **Author information**

636     Authors and affiliations

637     **Interfaculty Institute for Biochemistry, University of Tübingen, Tübingen, Germany**

638     Nils H. Rustmeier, Alexander Herrmann, Joshua C. Müller, Thilo Stehle

639     **Glycosciences Laboratory, Faculty of Medicine, Imperial College London, London, UK**

640     Lisete M. Silva, Antonio Di Maio, Ten Feizi, Yan Liu

641

642     Contributions

643     N.H.R. and T.S. conceptualized the study. A.H., J.C.M., and N.H.R. prepared and

644     crystallized the protein. L.M.S and A.D.M. performed the glycan array screening

645     experiments. A.D.M., Y.L. and T.F. performed the glycan array analyses and

646     interpreted the glycan interaction data. N.H.R. performed and analyzed the NMR, SPR

647     and flow cytometry experiments. A.H., J.M., and N.H.R. collected the crystallographic

648     data and solved the structures. N.H.R. prepared the first draft of the manuscript and all

649     contributed to the final revision. T.S. provided funding. All authors approved publication

650     of the manuscript.

651
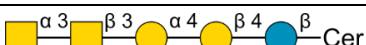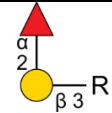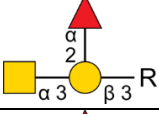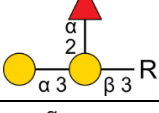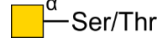
652     **Ethics declarations**

653     The authors declare no competing interests. Animal blood was acquired from vendors

654     guaranteeing animal wellbeing. Human erythrocytes were derived from anonymous

655     donors consenting research purposes.

656

*Appendices*

**Table 1** Carbohydrate antigens cited in this report.

| Class | Name | Symbolic structure | Sequence |
|---|---|---|---|
| Globoseries glycosphingolipids | LacCer | | Galβ1-4Glcβ-Cer |
| | Gb3Cer | | Galα1-4Galβ1-4Glcβ-Cer |
| | Gb4Cer (P antigen) | | GalNAcβ1-3Galα1-4Galβ1-4Glcβ-Cer |
| | Forssman pentaosylceramide | | GalNAcα1-3GalNAcβ1-3Galα1-4Galβ1-4Glcβ-Cer |
| Histo blood group antigens | H antigen | | Fucα1-2Galβ1-R |
| | A antigen | | GalNAcα1-3(Fucα1-2)Galβ1-R |
| | B antigen | | Galα1-3(Fucα1-2)Galβ1-R |
| O-Glycans | Tn antigen | | GalNAcα-Ser/Thr |
| | TF antigen | | Galβ1-3GalNAcα-Ser/Thr |

Lac - lactose, Cer - ceramide, Gb3 - globotriose, Gb4 - globotetraose, R - glycan backbone linked to protein or lipid carriers

○ - Gal  □ - GalNAc  ● - Glc  ■ - GlcNAc  ▲ - Fuc

657

124

658
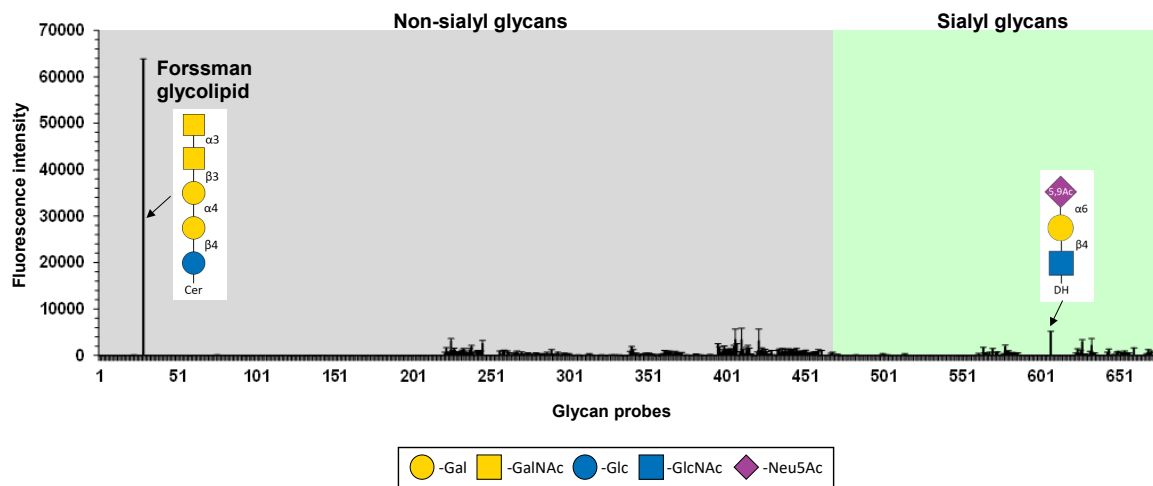
659 **Figure 1 Glycan microarray screening analysis of ShPyV VP1 reveals selective**

660 **engagement of Forssman glycolipid.** The results are the means of fluorescence

661 intensities of duplicate spots, printed at 5 fmol per spot. The error bars represent half

662 of the difference between the two values. In the glycan array the 672 lipid-linked probes

663 are grouped into non-sialylated and sialylated glycans as annotated by the colored

664 panels. The list of glycan probes, their sequences, and binding scores are in Table S1.

665
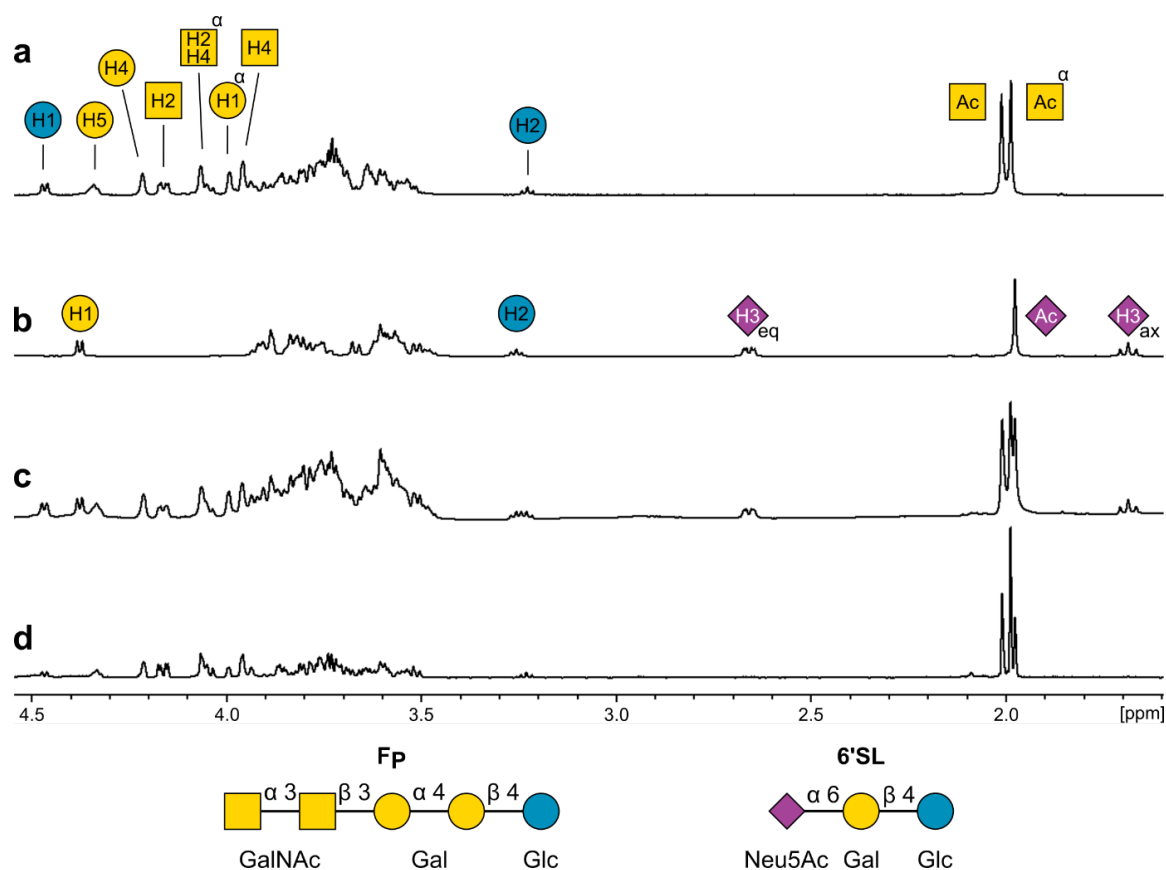
**Figure 2 Assessment of the sheep polyomavirus VP1 – Forssman pentaose interaction in solution by STD NMR**. $^1$H spectra of **a** Forssman pentaose (F$_P$) and **b** 6'sialyllactose (6'SL) in their isolated forms. Proton resonances are labeled according to the carbon positions in the monosaccharides. **c** $^1$H spectrum of a sample comprising 20 µM ShPyV VP1, 1 mM F$_P$, and 1 mM 6'SL collected under protein suppression. **d** The STD NMR spectrum that was recorded from the same sample.
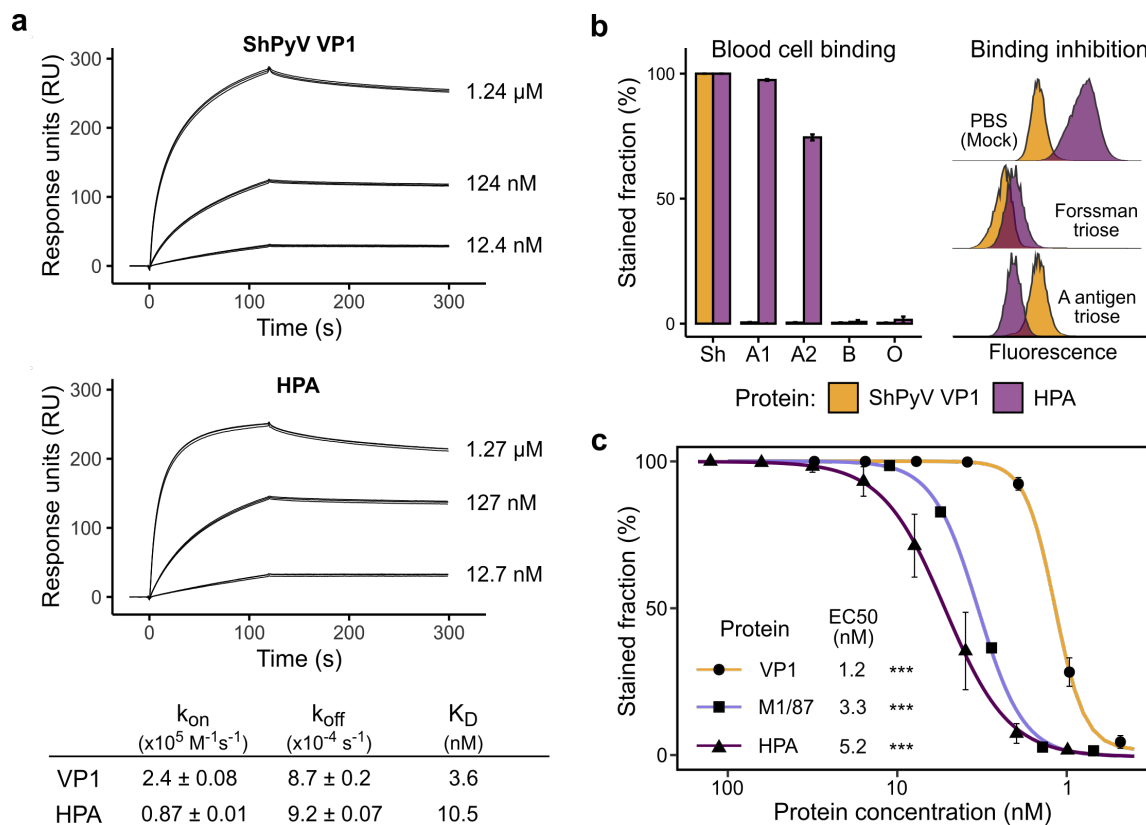
**Figure 3 Flow cytometry and surface plasmon resonance analyses of Forssman antigen engagement. a** Determination of the binding kinetics towards a Forssman-decorated surface using ShPyV VP1 and HPA. Response curves are shown as triplicates for each protein concentration. Kinetic parameters and the dissociation constant are provided in the table below. **b** Absolute cell attachment of ShPyV VP1 and *Helix pomatia* agglutinin (HPA) to Forssman antigen-positive red blood cells (RBCs) from sheep (Sh) and human erythrocytes of the blood groups A1, A2, B, and O (left). Effects of Forssman and blood group A antigen trioses on ShPyV VP1 and HPA attachment to sheep RBCs. Results are shown as concatenated and normalized histograms (right). **c** EC50 determination using ShPyV VP1, HPA, and anti-Forssman IgM M1/87 to sheep blood cells. Significance is indicated with *** for $p < 0.001$.
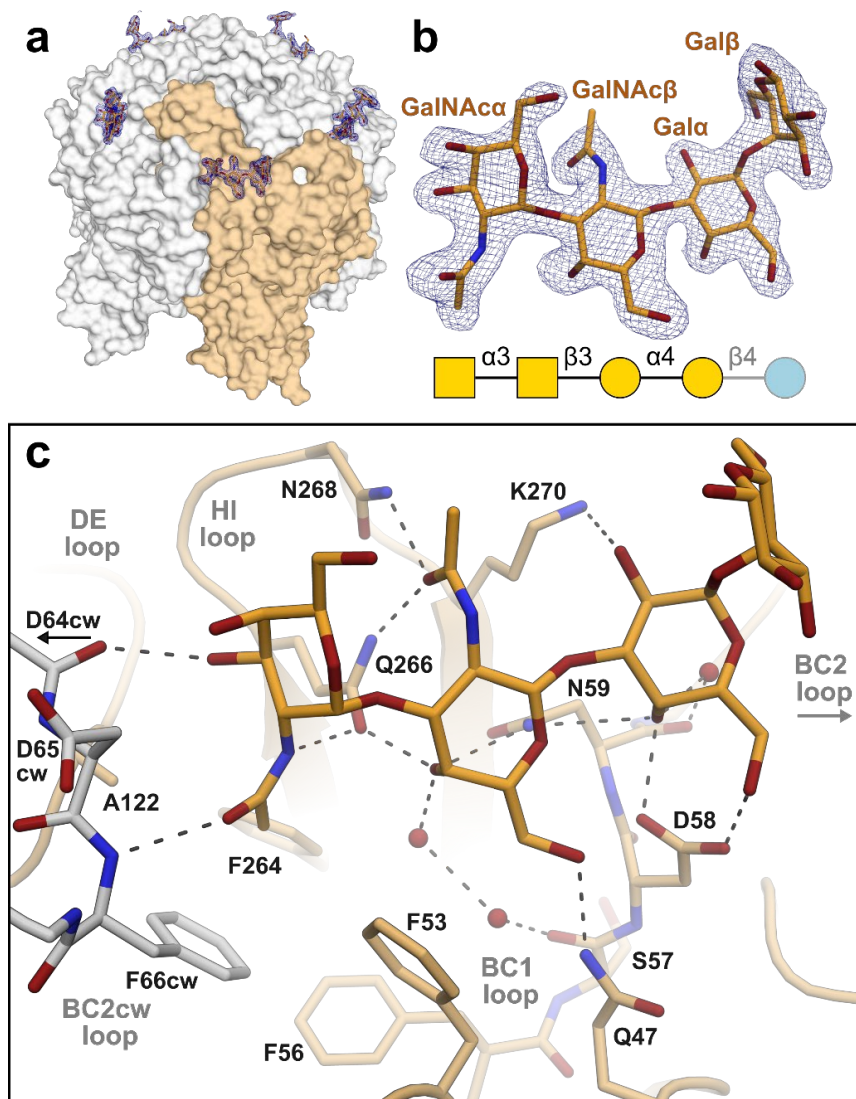
**Figure 4 Forssman pentaose binds to the external surface of ShPyV VP1 with numerous contacts**. **a** The crystal structure of a single VP1 pentamer is shown with five molecules of Forssman pentaose ($F_P$) associated to its outer surface. The bias-reduced Fo-Fc omit electron density map of $F_P$ is displayed as blue meshes around the oligosaccharides, with a contour level of 2.5 σ and a radius of 1.6 Å. The individual $F_P$ binding sites locate between adjoining protein chains, indicated by the orange coloring of one VP1 monomer. **b** Magnified view of a bound $F_P$ molecule with its symbol structure. The glucose of $F_P$, which is not visible in the crystal structure, is grayed out in the latter. **c** Molecular interactions between VP1 and $F_P$ are displayed with the hydrogen bonds drawn as dashed lines. VP1 coloring corresponds to panel **a**. Water

128

698    molecules are shown as spheres. Throughout the figure, oxygen and nitrogen atoms

699    are colored red and blue, respectively.

700

*Appendices*

701     **Table 2** Intermolecular contacts between ShPyV VP1 and Forssman pentaose.

| Monosaccharide | | VP1 contacts | |
|---|---|---|---|
| | | Van der Waals | |
| αGalNAc | Methyl | F53, F56, F66cw, A122, F264 | |
| | | Hydrogen-bonds | Distance (Å) |
| αGalNAc | O3 | *D64cw* | 2.74 |
| | Amide N | Q266 | 2.70 |
| | Acetyl O | *F66cw* | 2.90 |
| βGalNAc | O4 | $H_2O$-$H_2O$-*S57* | 2.95-2.90-2.63 |
| | O4 | N59 | 3.06 |
| | O4 | Q266 | 2.55 |
| | O6 | Q47 | 2.95 |
| | Acetyl O | Q266 | 2.96 |
| | Acetyl O | N268 | 2.78 |
| αGal | O2 | K270 | 3.31 |
| | O4 | D58 | 2.67 |
| | O4 | N59 | 3.45 |
| | O4 | $H_2O$-*N59* | 2.61-2.83 |
| | O6 | D58 | 2.54 |

Residues using their backbone atoms to establish contacts are printed in italic. Distance values are derived from a single binding site and may alter between chains.

702

130
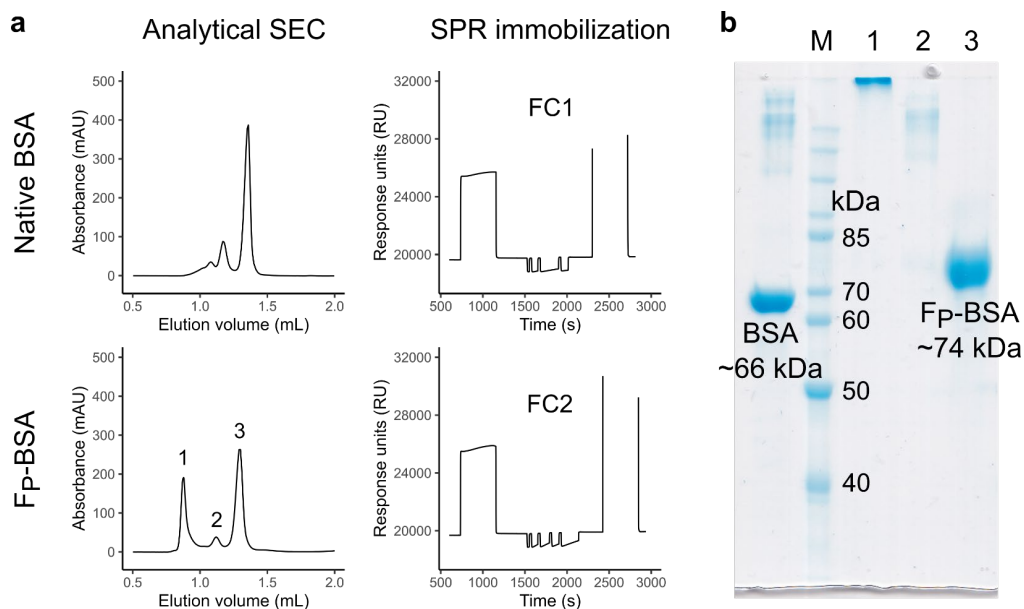
**Figure S3 a** The left panels show the analytical size exclusion chromatography (SEC) results of native BSA and FP-BSA using a SD200 Increase 3.2/300 column (Cytiva). The right panels display the respective surface plasmon resonance (SPR) immobilization procedures. For the second flow cell (FC2), fraction 3 of the FP-BSA SEC was used. **b** SDS-PAGE of native BSA (left lane) and the three fractions of the analytical SEC of FP-BSA. The mass difference between BSA and FP-BSA (lane 3) corresponds to a coupling of ~8 molecules of Forssman pentaose to one molecule of BSA.

**Table S4** Data collection and refinement statistics for the glycan complex crystal structures of ShPyV VP1.

| Data set of derivate(s) | Forssman pentaose | Globo-*N*-tetraose | Forssman pentaose + 6'SLN | Forssman pentaose + 3'SLN |
|---|---|---|---|---|
| PDB accession code | 7B6S | 7B6T | 7B6U | 7B6V |
| **Data collection** | | | | |
| Space group | $P3_1$ | $P3_1$ | $P3_1$ | $P3_1$ |
| a, b, c (Å) | 130.5, 130.5, 222.4 | 130.3, 130.3, 222.2 | 129.6, 129.6, 220.4 | 130.4, 130.4, 221.2 |
| α, β, γ (°) | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 |
| Resolution (Å) | 42.71 - 1.92 (2.04 - 1.92) | 48.91 - 1.70 (1.80 - 1.70) | 49.47 - 1.90 (2.02 - 1.90) | 48.85 - 1.80 (1.91 - 1.80) |
| Total reflections | 2'214'763 | 2'676'281 | 1'728'348 | 3'864'575 |
| Unique reflections | 321'069 | 462'792 | 325'151 | 390'469 |
| $R_{meas}$ (%) | 9.3 (68.4) | 10.7 (108.5) | 14.6 (122.3) | 16.2 (127.3) |
| I/σI | 16.67 (2.91) | 13.51 (1.61) | 9.46 (1.33) | 11.02 (1.51) |
| $CC_{1/2}$ (%) | 99.9 (78.6) | 99.8 (48.5) | 99.6 (44.1) | 99.7 (47.1) |
| Completeness (%) | 99.8 (98.7) | 99.9 (99.7) | 99.8 (98.7) | 99.8 (98.9) |
| Wilson B-factor (Å²) | 32.9 | 27.9 | 31.3 | 28.5 |
| | | | | |
| **Refinement** | | | | |
| $R_{work}$ / $R_{free}$ (%) | 14.9 / 17.6 | 14.3 / 16.5 | 14.6 / 17.3 | 15.2 / 17.7 |
| No. of atoms | | | | |
| Protein | 20'468 | 20'343 | 20'393 | 20'322 |
| Water | 2'254 | 3'256 | 2'031 | 2'373 |
| Glycan | 564 | 411 | 761 | 636 |
| B-factor (Å²) | | | | |
| Protein | 29.5 | 22.8 | 28.0 | 24.5 |
| Water | 37.4 | 34.8 | 35.8 | 30.8 |
| Glycan | 32.8 | 37.2 | 34.0 | 26.4 |
| R.m.s.d | | | | |
| Bond length (Å) | 0.0093 | 0.0095 | 0.0090 | 0.0092 |
| Bond angle (°) | 1.65 | 1.64 | 1.65 | 1.72 |
| Ramachandran outliers (%) | 0.08 | 0.08 | 0.04 | 0.15 |
| Clashscore | 1.94 | 1.80 | 1.89 | 2.96 |
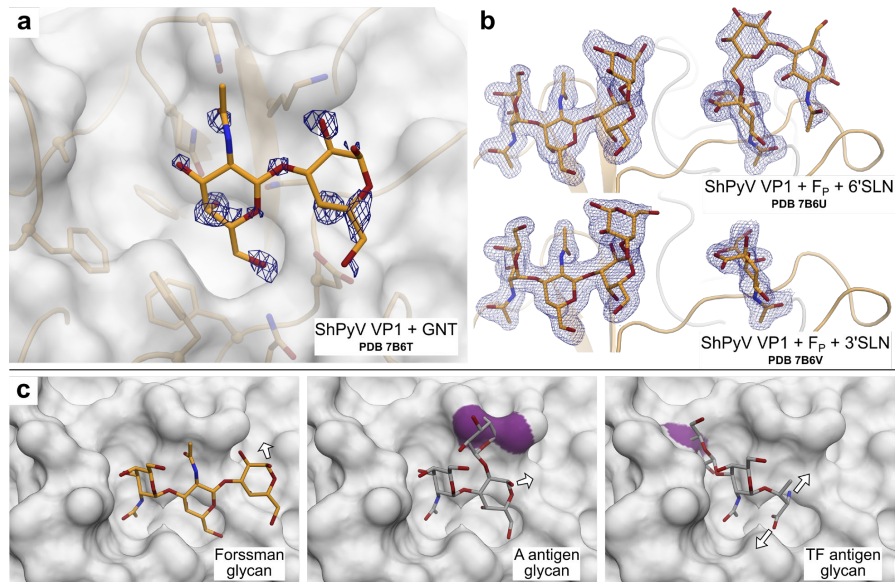
**Figure S5 a** The crystal structure of ShPyV VP1 in complex with 10 mM globo-*N*-tetraose (GNT, PDB 7B6T), the glycan of the Forssman precursor molecule globoside, reveals only negligible binding. **b** The co-complexes of ShPyV VP1 with Forssman pentaose ($F_P$), and 6'sialyllactosamine (6'SLN, PDB ID 7B6U) or 3'sialyllactosamine (3'SLN, PDB ID 7B6V) show that the different glycan species bind to different sites on the protein surface with no close contact between them. In **a** + **b**, the Fo-Fc electron densities are shown as blue meshes around the individual oligosaccharides with a contour level of 2.5 σ and a radius of 1.6 Å. **c** Based on αGalNAc superposition, the side-by-side depiction of the Forssman antigen glycan (GalNAcα1-3GalNacβ1-3Galα), A antigen glycan (GalNAcα1-3[Fucα1-2]Galβ), and the TF antigen (Galβ1-3GalNAcα1-Ser/Thr) in association with the VP1 surface. In the A antigen and TF antigen glycan models the fucose and galactose, respectively, result in overlaps with the protein (highlighted in purple). The Tn antigen, which corresponds to TF truncated by the terminal galactose, also cannot bind to VP1 as the glycoprotein proceeds horizontally to VP1, which again would result in overlaps. In all panels of **c**, arrows indicate the linkage to the remaining antigen portion. Throughout the figure, oxygen and nitrogen atoms are colored red and blue, respectively.
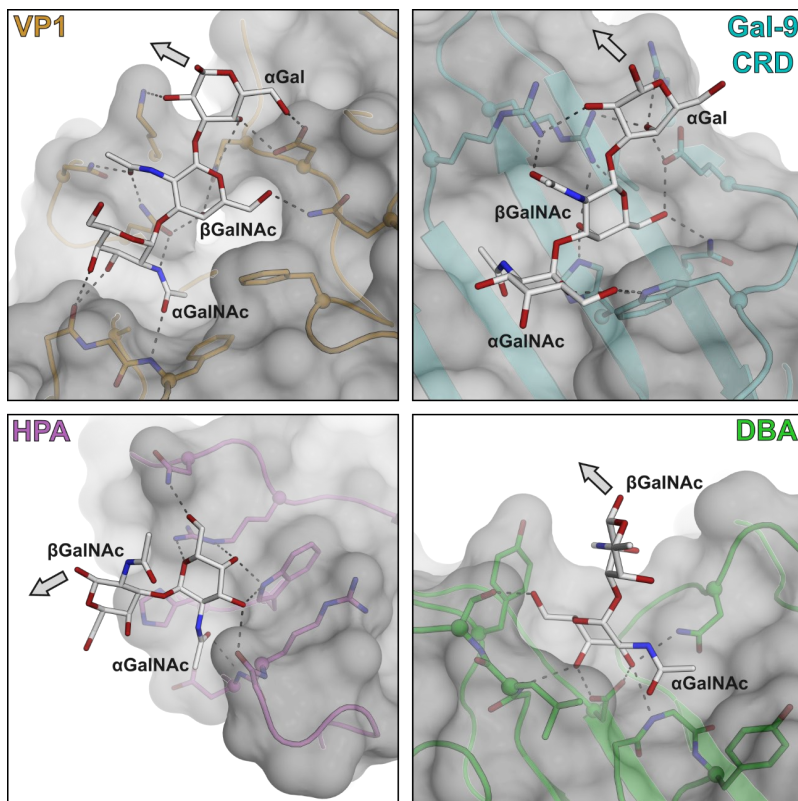


**Figure S6** Different modes of Forssman glycan binding in ShPyV VP1, Galectin-9 carbohydrate recognition domain (Gal-9 CRD), *Helix pomatia* agglutinin (HPA), and *Dolichos biflorus* agglutinin (DBA). VP1 recognizes the GalNAcα1-3GalNAcβ1-3Galα trisaccharide sequence of Forssman pentaose ($F_P$) via multiple hydrogen bonds to each sugar unit (upper left panel). Human Gal-9 CRD (PDB 5X4A) mainly interacts with the βGalNAc of $F_P$ (upper right panel). HPA (PDB 2CGY) and DBA (PDB 1LU1) interact exclusively with the terminal αGalNAc but not βGalNAc (bottom panels). All proteins are shown as transparent surfaces with an underlying cartoon representation. Binding site residues (colored according to the protein) and Forssman glycans (light grey) are shown as sticks. Arrows indicate the direction of the linkage to the next monosaccharide. Oxygen and nitrogen atoms are colored red and blue, respectively. Hydrogen bonds are indicated as dashed black lines.

Forssman glycan binding by *Sinularia lochmodes* lectin-2 (PDB 5X4A) is homologous to HPA and not shown.

# 2 Raw scans
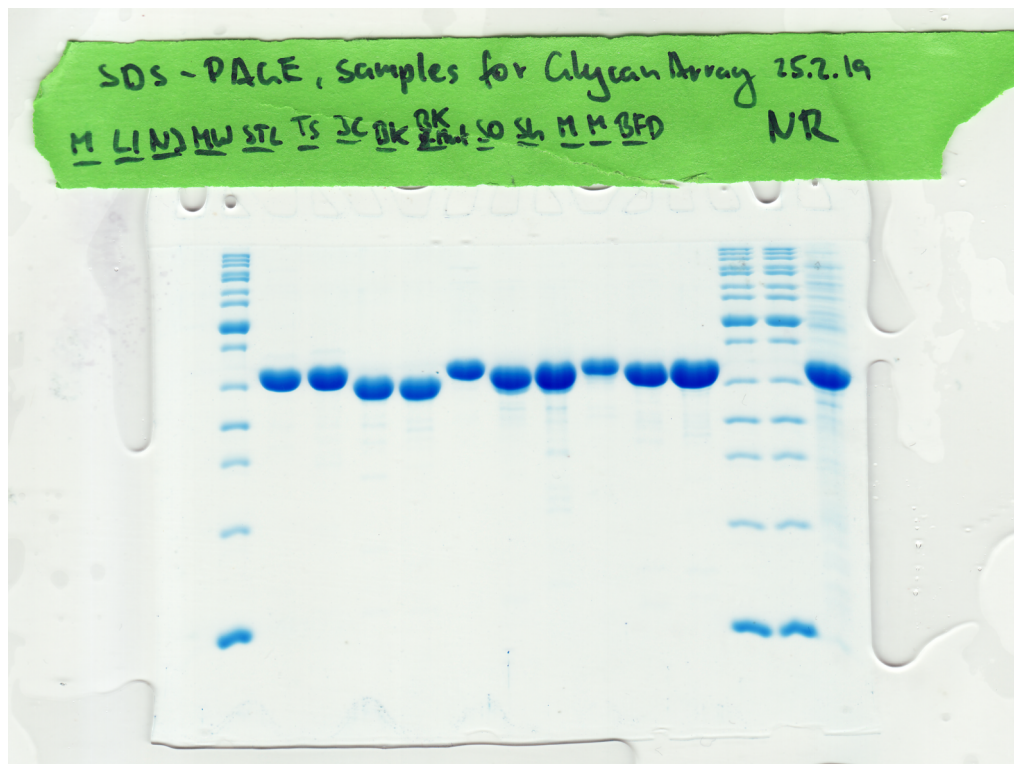


Figure 41: Raw scan of the SDS-PAGE gels of purified VP1 samples for the glycan array. This scan is the basis for Figure 10 on page 38.
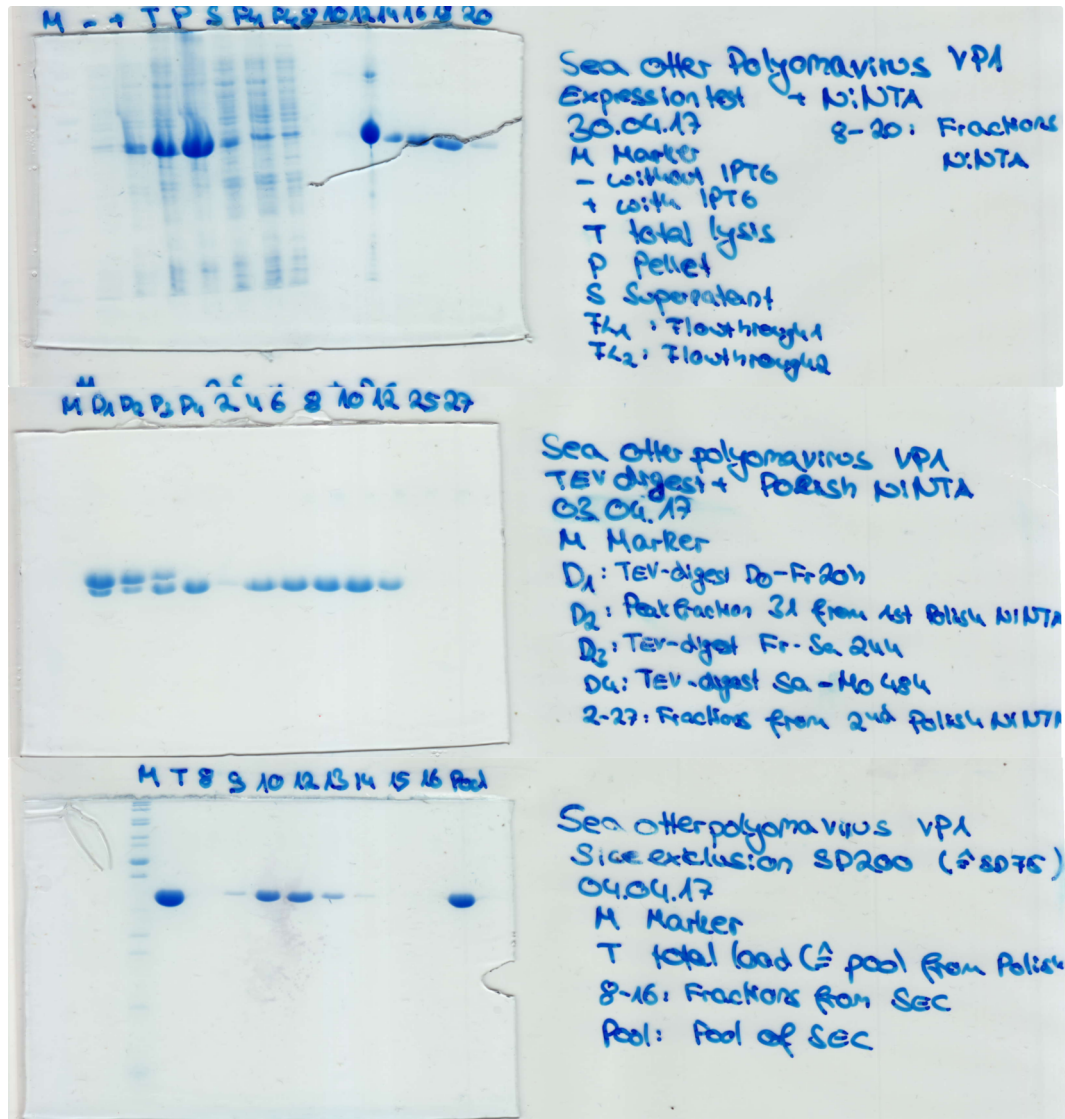
Figure 42: Raw scans of the SDS-PAGE gels of the SOPyV VP1 purification. This scan is the basis for Figure 12 on page 40.
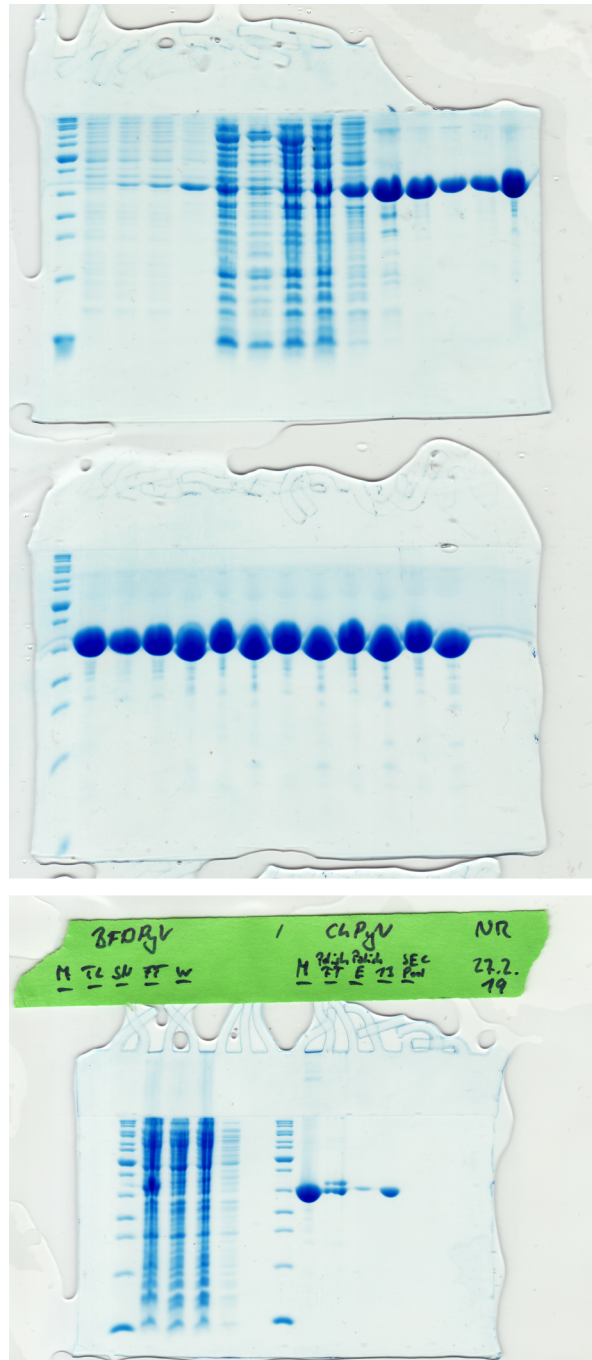
Figure 43: Raw scans of the SDS-PAGE gels of the ChPyV VP1 purification. These scans are the basis for Figure 20 on page 52.
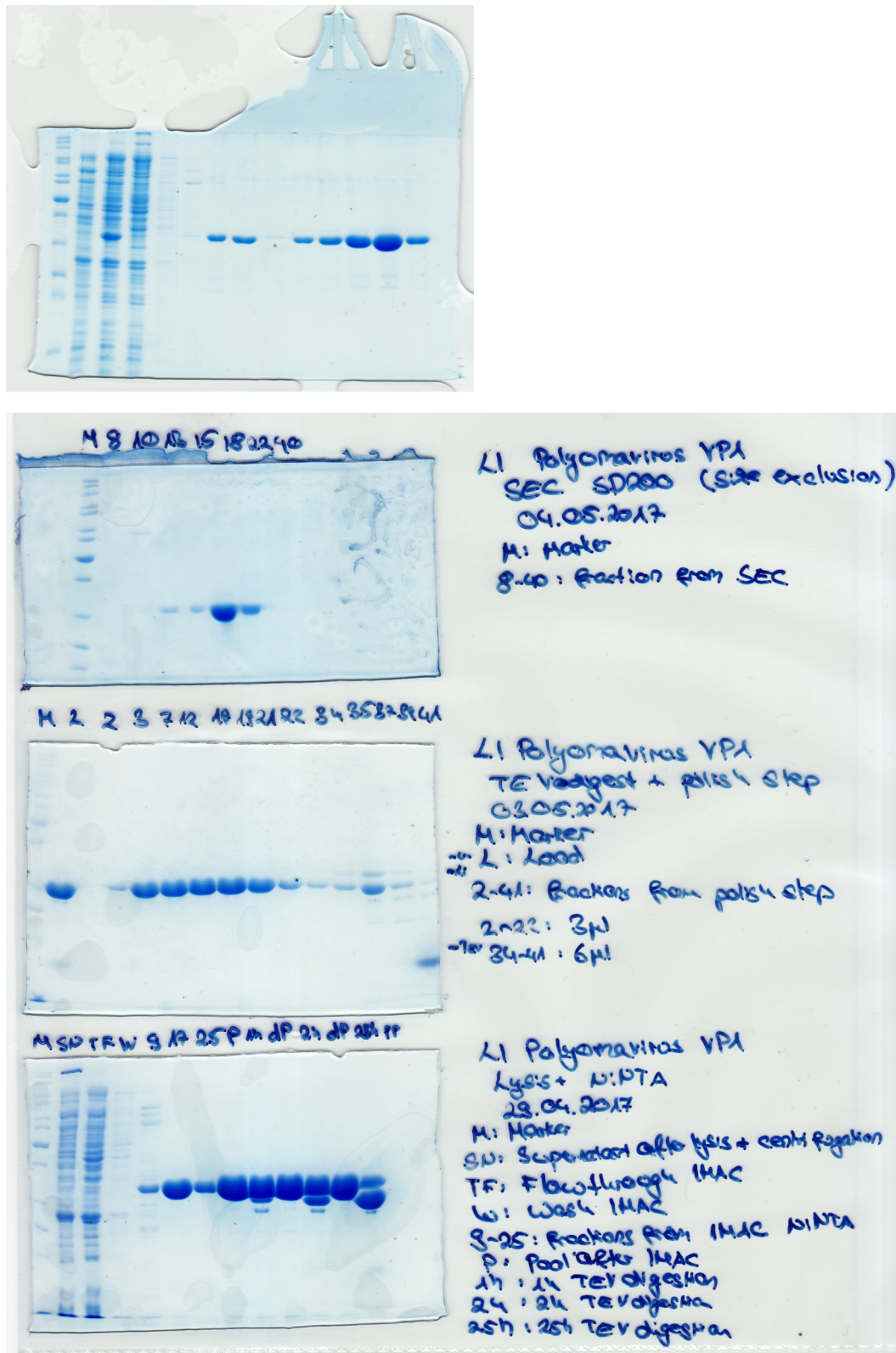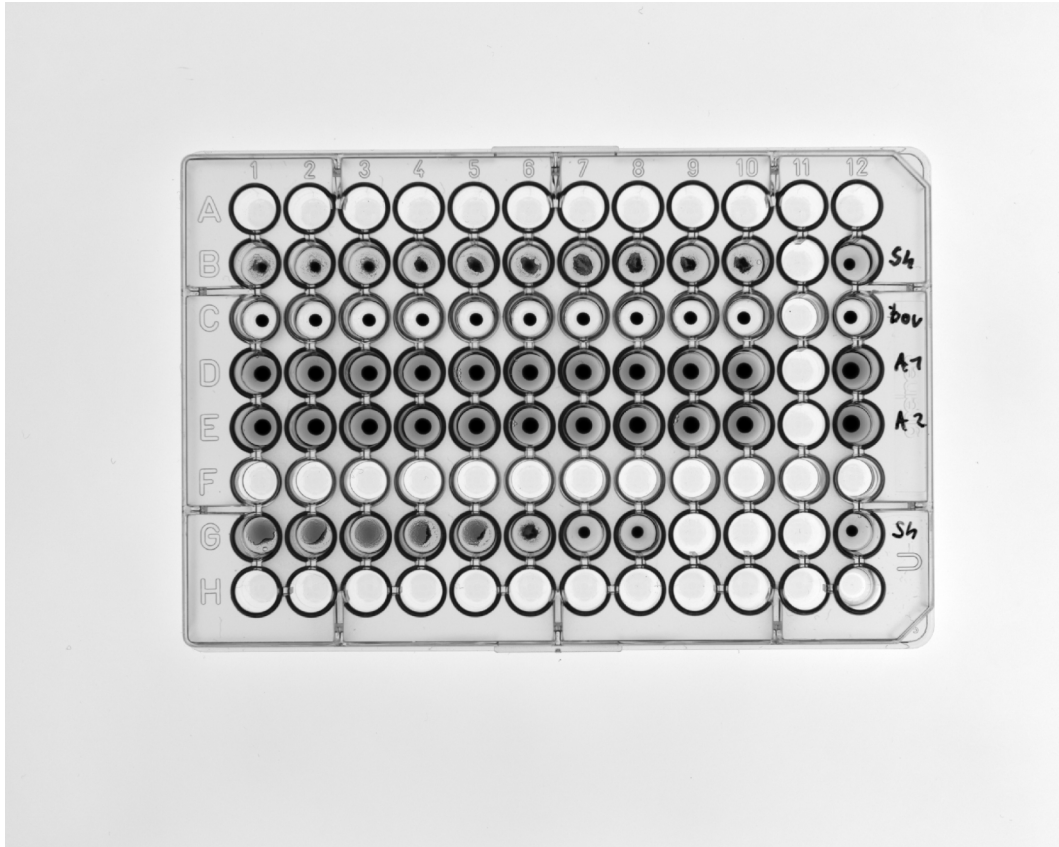
Figure 44: Raw scans of the SDS-PAGE gels of the LIPyV VP1 purification. These scans are the basis for Figure 25 on page 60.

Figure 45: Raw scan of the hemagglutination assay using ShPyV VP1 as coagulant. This scan is the basis for Figure 36 on page 75.

# 3 Supplementary crystallographic statistics
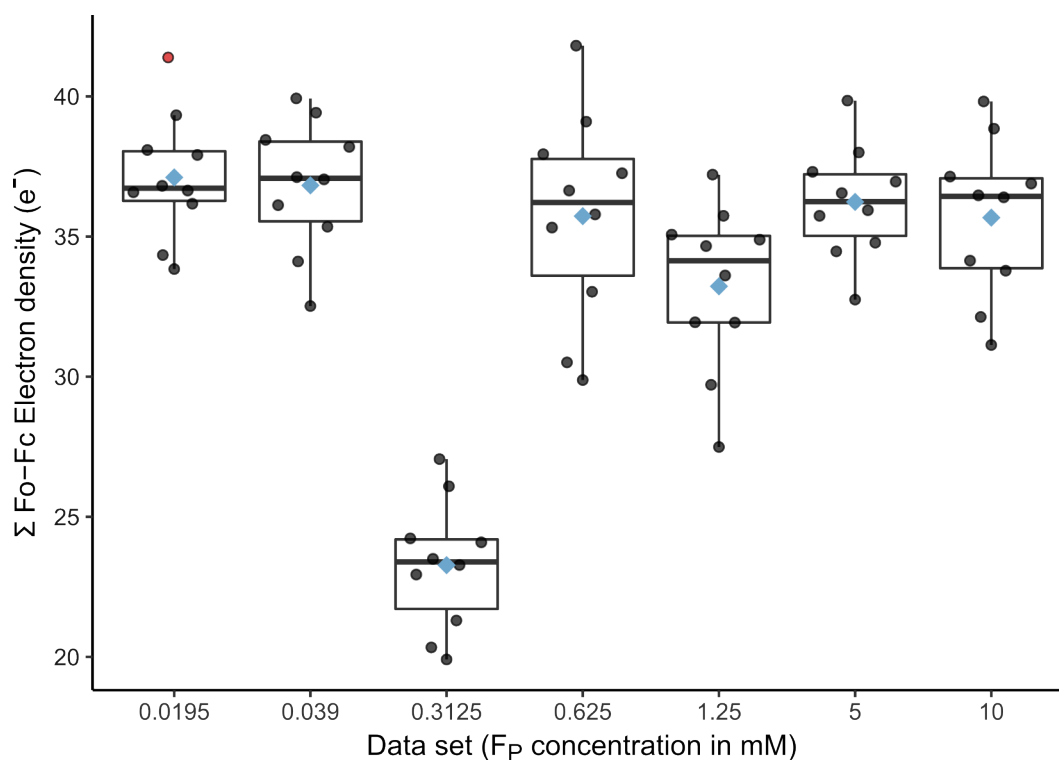


Figure A1: Boxplots of omitted tripeptide (residues 143–145) reference Fo-Fc density integrals for the crystallographic ShPyV VP1-$F_P$ dissociation constant calculation. Outliers are shown in red, and the median and mean values are indicated as bold horizontal lines and blue diamonds, respectively. The boxed areas comprise the interquartile range (IQR) of the data, with the vertical lines extending towards non-outlying extreme values (within 1.5 * IQR).

Table A1: Data and refinement statistics for the crystal structures of ShPyV VP1 with the Forssman pentaose dilution series.

| Data set | 10 mM | 5 mM | 1.25 mM | 0.625 mM | 0.3125 mM | 0.039 mM | 0.0195 mM |
|---|---|---|---|---|---|---|---|
| **Data collection** | | | | | | | |
| Space group | | | | P31 | | | |
| a, b, c (Å) | | | | 131, 131, 222 | | | |
| α, β, γ (°) | | | | 90, 90, 120 | | | |
| Resolution (Å) | 42.9 – 2.2 | 49.9 – 2.2 | 49.1 – 2.2 | 49.9 – 2.2 | 49.1 – 2.2 | 49.9 – 2.2 | 49.9 – 2.2 |
| Total reflections | 1,429,271 | 2,233,502 | 2,261,265 | 2,240,680 | 2,153,610 | 2,220,054 | 2,221,606 |
| Unique reflections | 214,373 | 212,193 | 214,900 | 216,291 | 215,471 | 212,369 | 213,407 |
| $R_{meas}$ (%) | 10.5 (47.5) | 25.1 (169.2) | 21.8 (139.4) | 20.4 (105.6) | 18.9 (101.6) | 13.3 (65.8) | 26.0 (138.7) |
| I/σI | 15.17 (4.14) | 10.11 (1.60) | 10.43 (1.72) | 12.27 (2.58) | 12.21 (2.34) | 15.23 (3.87) | 9.68 (1.59) |
| $CC_{1/2}$ (%) | 99.7 (89.4) | 99.4 (42.3) | 99.5 (55.7) | 99.5 (73.7) | 99.5 (62.8) | 99.7 (83.6) | 99.2 (53.9) |
| Completeness (%) | 99.1 (94.7) | 99.7 (98.4) | 99.8 (98.8) | 100 (100) | 99.7 (98.0) | 99.6 (97.9) | 99.8 (98.7) |
| Wilson B-factor (Å²) | 36.0 | 37.5 | 38.3 | 34.7 | 34.3 | 36.8 | 33.5 |
| | | | | | | | |
| **Refinement** | | | | | | | |
| $R_{work}$ / $R_{free}$ (%) | 21.5 / 24.1 | 22.0 / 24.5 | 20.5 / 23.0 | 19.7 / 21.8 | 19.2 / 21.4 | 19.3 / 22.0 | 21.4 / 23.6 |
| Main twin fractions (%) | | | | | | | |
| H, K, L / H, -H-K, -L | 79.4 / 20.6 | 72.4 / 27.6 | 60.5 / 39.5 | 47.6 / 34.7 | 52.4 / 47.6 | 71.3 / 28.7 | 70.1 / 29.9 |
| Protein B-factor (Å²) | 29.9 | 29.0 | 32.6 | 28.6 | 28.0 | 27.5 | 29.7 |

Values in parentheses refer to the highest resolution bin.

# 4 Supplementary glycan array material
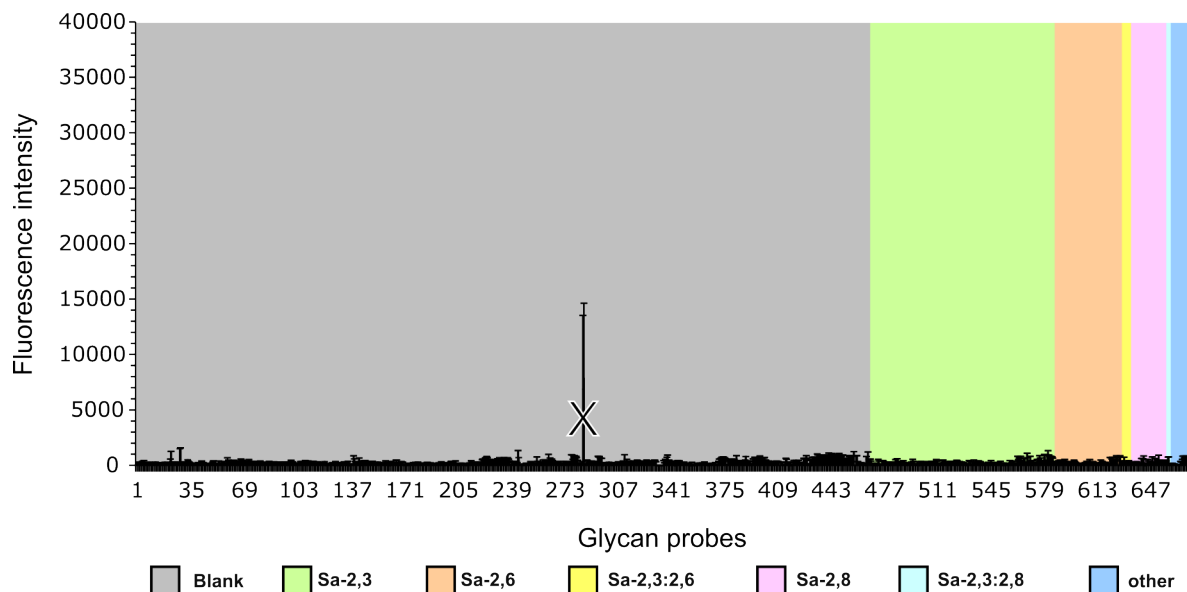


Figure A2: Graph of the SOPyV VP1 glycan array screening results. Glycan probes are sorted by sialyl linkages as indicated by the color code. The crossed-out signal arose from an artifact on the array slide and is erroneous. No preferential glycan binding was detectable for SOPyV VP1.

## 4.1 Glycan probe list

A list of the probes in this work's glycan arrays was deduced from the Supplementary Table S1 (`https://www.biorxiv.org/content/biorxiv/early/2023/04/15/2023.04.10.536218/DC1/embed/media-1.xlsx`) of Rustmeier *et al.* [2], distributed under the CC BY 4.0 deed (`https://creativecommons.org/licenses/by/4.0/`). It was converted to grayscale and attached as Table A2 on the following pages.

**Table A2. List of glycan probes in the broad-spectrum glycan screening array.**

| Pos[a] | Probe[b] | Sequence[c] | GlyTouCan ID[d] |
|---|---|---|---|
| | | *Non-sialyl* | |
| 1 | **Hanfland 1** | Glcβ-Cer | G71142DF |
| 2 | **Galactosylceramide** | Galβ-Cer | G65889KE |
| 3 | **Galactocerebrosides** | Galβ-Cer | G65889KE |
| 4 | **H-Di** | Fucα-2Gal-DH | G66086EX |
| 5 | **A-Tri** | GalNAcα-3Gal-DH<br>    \|<br>    Fucα-2 | G55958KV |
| 6 | **A-tri-AO** | GalNAcα-3Gal-AO<br>    \|<br>    Fucα-2 | G55958KV |
| 7 | **B-Tri** | Galα-3Gal-DH<br>    \|<br>    Fucα-2 | G58612ZQ |
| 8 | **B-Tri-AO** | Galα-3Gal-AO<br>    \|<br>    Fucα-2 | G58612ZQ |
| 9 | **GSC-426** | (3-deoxy,3-carboxymethyl)Galβ-C30 | N/A |
| 10 | **Sulfatide** | SU-3Galβ-Cer | G72548RZ |
| 11 | **GSF-1** | SU-3Galβ-C30 | G72548RZ |
| 12 | **GSC-209** | GlcAβ-3Galβ-Cer42 | G99041PQ |
| 13 | **GSC-210** | SU-3GlcAβ-3Galβ-Cer42 | G53954AN |
| 14 | **Glucocerebrosides** | Glcβ-Cer | G71142DF |
| 15 | **GSF-19** | SU-6Glcβ-C30 | G21194GL |
| 16 | **Lactocerebrosides** | Galβ-4Glcβ-Cer | G84224TW |
| 17 | **Lac** | Galβ-4Glc-DH | G74621DY |
| 18 | **Lac-AO** | Galβ-4Glc-AO | G74621DY |
| 19 | **GSC-432** | (3-deoxy,3-carboxymethyl)Galβ-4Glcβ-C30 | N/A |
| 20 | **GSC-296** | GlcAβ-3Galβ-4Glcβ-C30 | G49110FU |
| 21 | **GSC-353** | SU-3GlcAβ-3Galβ-4Glcβ-C30 | G08069IM |
| 22 | **GalNAcα-3Galβ-4Glc** | GalNAcα-3Galβ-4Glc-DH | G95341UA |
| 23 | **Globotri-AO** | Galα-4Galβ-4Glc-AO | G06000DZ |
| 24 | **Ceramide trihexoside** | Galα-4Galβ-4Glcβ-Cer | G00059MO |
| 25 | **Globoside (P-antigen)** | GalNAcβ-3Galα-4Galβ-4Glcβ-Cer | G00061MO |
| 26 | **Globo-H-Hexa** | Fucα-2Galβ-3GalNAcβ-3Galα-4Galβ-4GlcNAc-DH | G27318OT |
| 27 | **Globo-A-Hepta** | GalNAcα-3Galβ-3GalNAcβ-3Galα-4Galβ-4GlcNAc-DH<br>    \|<br>    Fucα-2 | G28008YF |
| 28 | **Globo-B-Hepta** | Galα-3Galβ-3GalNAcβ-3Galα-4Galβ-4GlcNAc-DH<br>    \|<br>    Fucα-2 | G26270KR |
| 29 | **Forssmann glycolipid** | GalNAcα-3GalNAcβ-3Galα-4Galβ-4Glcβ-Cer | G80062GG |
| 30 | **Fuc(3)-Lac-AO** | Galβ-4Glc-AO<br>    \|<br>    Fucα-3 | G06210XB |
| 31 | **GSC-430** | (3-deoxy,3-carboxymethyl)Galβ-3Glcβ-C30<br>          \|<br>          Fucα-4 | N/A |

| # | Name | Structure | ID |
|---|---|---|---|
| 32 | GSC-260 | 3-deoxy,3-carboxymethyl-Galβ-4Glcβ-C30<br>&#124;<br>Fucα-3 | N/A |
| 33 | GSC-150 | SU-3Galβ-4Glcβ-C30<br>&#124;<br>Fucα-3 | G14344KX |
| 34 | GSC-160 | SU-3Galβ-4Glcβ-Cer36<br>&#124;<br>Fucα-3 | G14344KX |
| 35 | LacNAc(1-3) | Galβ-3GlcNAc-DH | G04616ST |
| 36 | LacNAc(1-3)-AO | Galβ-3GlcNAc-AO | G04616ST |
| 37 | LacNAc | Galβ-4GlcNAc-DH | G51331BY |
| 38 | LacNAc-AO | Galβ-4GlcNAc-AO | G51331BY |
| 39 | H-Tri-T2-AO | Fucα-2Galβ-4GlcNAc-AO | G39064VB |
| 40 | A-Tetra-T1-AO | GalNAcα-3Galβ-3GlcNAc-AO<br>&#124;<br>Fucα-2 | G58507AZ |
| 41 | A-Tetra-T2-AO | GalNAcα-3Galβ-4GlcNAc-AO<br>&#124;<br>Fucα-2 | G17108EX |
| 42 | B-Tetra-T1-AO | Galα-3Galβ-3GlcNAc-AO<br>&#124;<br>Fucα-2 | G26934CO |
| 43 | B-Tetra-T2-AO | Galα-3Galβ-4GlcNAc-AO<br>&#124;<br>Fucα-2 | G25651TK |
| 44 | B-penta-AO | Galα-3Galβ-4Glc-AO<br>&#124;    &#124;<br>Fucα-2  Fucα-3 | G06557IC |
| 45 | Galα-4Galβ-4GlcNAc | Galα-4Galβ-4GlcNAc-DH | G55205ZV |
| 46 | SU(3')-LN | SU-3Galβ-4GlcNAc-DH | G11658OG |
| 47 | Lea-Tri | Galβ-3GlcNAc-DH<br>&#124;<br>Fucα-4 | G00047MO |
| 48 | Lea-Tri-AO | Galβ-3GlcNAc-AO<br>&#124;<br>Fucα-4 | G00047MO |
| 49 | Leb tetra-AO | Fucα-2Galβ-3GlcNAc-AO<br>&#124;<br>Fucα-4 | G39211DH |
| 50 | Lex-Tri | Galβ-4GlcNAc-DH<br>&#124;<br>Fucα-3 | G01187XC |
| 51 | Lex-Tri-AO | Galβ-4GlcNAc-AO<br>&#124;<br>Fucα-3 | G01187XC |
| 52 | Lex-Tri-(Me)AO | Galβ-4GlcNAc-(Me)AO<br>&#124;<br>Fucα-3 | G01187XC |
| 53 | SU(3')-Lea-Tri | SU-3Galβ-3GlcNAc-DH<br>&#124;<br>Fucα-4 | G36845IZ |
| 54 | Su(3')-Lea tri-AO | SU-3Galβ-3GlcNAc-AO<br>&#124;<br>Fucα-4 | G36845IZ |
| 55 | SU(3')-Lex-Tri | SU-3Galβ-4GlcNAc-DH<br>&#124;<br>Fucα-3 | G26555CS |
| 56 | Su(3')-Lex tri-AO | SU-3Galβ-4GlcNAc-AO<br>&#124;<br>Fucα-3 | G26555CS |
| 57 | Orsay-1-AO | Galβ-4GlcNAcβ-6Gal-AO | G59901EA |

| 58 | Orsay-2-AO | Galβ-4GlcNAcβ-3Gal-AO | G92536LD |
|---|---|---|---|
| 59 | Orsay-3-AO | Galβ-3GlcNAcβ-6Gal-AO | G24010VW |
| 60 | Orsay-4-AO | Galβ-3GlcNAcβ-3Gal-AO | G23041XK |
| 61 | GSC-225 | (3-deoxy,3-carboxymethyl)Galβ-4GlcNAcβ-3Galβ-Cer36<br>&#124;<br>Fucα-3 | N/A |
| 62 | GSC-236 | SU-3Galβ-4GlcNAcβ-3Galβ-C30<br>&#124;<br>Fucα-3 | G68845UH |
| 63 | DLNN | GlcNAcβ-3Galβ-4Glc-DH | G87727XV |
| 64 | LNT | Galβ-3GlcNAcβ-3Galβ-4Glc-DH | G45827GY |
| 65 | Paragloboside | Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer | G58896AZ |
| 66 | LNnT | Galβ-4GlcNAcβ-3Galβ-4Glc-DH | G48059CD |
| 67 | B-like pentaosylceramide | Galα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer | G60645OK |
| 68 | Klaus glycolipid | Galβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer | G70213NR |
| 69 | GSC-207 | GlcAβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-C30 | G87867IZ |
| 70 | GSC-191 | GlcAβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36 | G87867IZ |
| 71 | GSC-189 | GlcAβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer42 | G87867IZ |
| 72 | SU(3')-Tri | SU-3Galβ-4GlcNAcβ-3Gal-DH | G69745TK |
| 73 | GSC-208 | SU-3GlcAβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-C30 | G39133HC |
| 74 | GSC-192 | SU-3GlcAβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36 | G39133HC |
| 75 | GSC-190 | SU-3GlcAβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer42 | G39133HC |
| 76 | Led-II pentaosylceramide | Fucα-2Galβ-3GlcNAcβ-3Galβ-4Glcβ-CerA | G00050MO |
| 77 | Led-I pentaosylceramide | Fucα-2Galβ-3GlcNAcβ-3Galβ-4Glcβ-CerB | G00050MO |
| 78 | LNFP-I | Fucα-2Galβ-3GlcNAcβ-3Galβ-4Glc-DH | G01650PH |
| 79 | LNFP-I-AO | Fucα-2Galβ-3GlcNAcβ-3Galβ-4Glc-AO | G01650PH |
| 80 | B-hexaosylceramide | Galα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer<br>&#124;<br>Fucα-2 | G77593JU |
| 81 | A-Hexa | GalNAcα-3Galβ-3GlcNAcβ-3Galβ-4Glc-DH<br>&#124;<br>Fucα-2 | G11928TH |
| 82 | A-Hepta | Fucα-4<br>&#124;<br>GalNAcα-3Galβ-3GlcNAcβ-3Galβ-4Glc-DH<br>&#124;<br>Fucα-2 | G13040AD |
| 83 | LNFP-II | Galβ-3GlcNAcβ-3Galβ-4Glc-DH<br>&#124;<br>Fucα-4 | G98173LG |
| 84 | LNFP-II-AO | Galβ-3GlcNAcβ-3Galβ-4Glc-AO<br>&#124;<br>Fucα-4 | G98173LG |
| 85 | LNDFH-II | Galβ-3GlcNAcβ-3Galβ-4Glc-DH<br>&#124;     &#124;<br>Fucα-4   Fucα-3 | G70115XG |
| 86 | Leb-hexaosylceramide | Fucα-2Galβ-3GlcNAcβ-3Galβ-4Glcβ-Cer<br>&#124;<br>Fucα-4 | G09552LA |
| 87 | LNDFH-I | Fucα-2Galβ-3GlcNAcβ-3Galβ-4Glc-DH<br>&#124;<br>Fucα-4 | G50540JU |
| 88 | LNDFH-I-AO | Fucα-2Galβ-3GlcNAcβ-3Galβ-4Glc-AO<br>&#124;<br>Fucα-4 | G50540JU |

| # | Name | Structure | Code |
|---|------|-----------|------|
| 89 | **LNTFH-I** | Fucα-2Galβ-3GlcNAcβ-3Galβ-4Glc-DH<br>　　　　　\|　　　　　\|<br>　　　Fucα-4　　Fucα-2 | G47970KN |
| 90 | **LNFP-III** | Galβ-4GlcNAcβ-3Galβ-4Glc-DH<br>　　　\|<br>　Fucα-3 | G83916HL |
| 91 | **LNFP-III-AO** | Galβ-4GlcNAcβ-3Galβ-4Glc-AO<br>　　　\|<br>　Fucα-3 | G83916HL |
| 92 | **LNnFP-I** | Fucα-2Galβ-4GlcNAcβ-3Galβ-4Glc-DH | G42553JB |
| 93 | **LNnFP-I-AO** | Fucα-2Galβ-4GlcNAcβ-3Galβ-4Glc-AO | G42553JB |
| 94 | **A-Hexa-T1** | GalNAcα-3Galβ-3GlcNAcβ-3Galβ-4Glc-DH<br>　　　　\|<br>　　Fucα-2 | G11928TH |
| 95 | **A-Hexa-T2** | GalNAcα-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH<br>　　　　\|<br>　　Fucα-2 | G06031DD |
| 96 | **B-Hexa-T1** | Galα-3Galβ-3GlcNAcβ-3Galβ-4Glc-DH<br>　　　\|<br>　Fucα-2 | G22270QF |
| 97 | **B-Hexa-T2** | Galα-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH<br>　　　\|<br>　Fucα-2 | G54585VU |
| 98 | **LNnDFH-I** | Fucα-2Galβ-4GlcNAcβ-3Galβ-4Glc-DH<br>　　　　\|<br>　　Fucα-3 | G94480CC |
| 99 | **LNnDFH-I-AO** | Fucα-2Galβ-4GlcNAcβ-3Galβ-4Glc-AO<br>　　　　\|<br>　　Fucα-3 | G94480CC |
| 100 | **LNnDFH-II** | Galβ-4GlcNAcβ-3Galβ-4Glc-DH<br>　　\|　　　　　\|<br>Fucα-3　　　　Fucα-3 | G11636EE |
| 101 | **LNnDFH-V** | Galβ-4GlcNAcβ-3Galβ-4Glc-DH<br>　　\|　　　\|<br>Fucα-3　Fucα-2 | G24183XZ |
| 102 | **LNnTFH-I** | Fucα-2Galβ-4GlcNAcβ-3Galβ-4Glc-DH<br>　　　　\|　　　\|<br>　　Fucα-3　Fucα-2 | G77137FD |
| 103 | **SU(3')-LNFP-II** | SU-3Galβ-3GlcNAcβ-4Galβ-4Glc-DH<br>　　　\|<br>　Fucα-4 | G54723TY |
| 104 | **SU(6')-LNFP-II** | SU-6Galβ-3GlcNAcβ-3Galβ-4Glc-DH<br>　　　\|<br>　Fucα-4 | G45702SF |
| 105 | **SU(3')-LNFP-III** | SU-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH<br>　　　\|<br>　Fucα-3 | G00777SR |
| 106 | **SU(6')-LNFP-III** | SU-6Galβ-4GlcNAcβ-3Galβ-4Glc-DH<br>　　　\|<br>　Fucα-3 | G82272GJ |
| 107 | **SU(3',6)-LNFP-III** | 　SU-6<br>　　\|<br>SU-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH<br>　　　\|<br>　Fucα-3 | G97370PT |
| 108 | **pLNH-b** | Galβ-3GlcNAcβ-3Galβ-3GlcNAcβ-3Galβ-4Glc-DH | G91182AF |
| 109 | **pLNnH** | Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH | G85278ER |
| 110 | **MFLNH-I** | 　Galβ-4GlcNAcβ-6<br>　　　　　\|<br>　　　Galβ-4Glc-DH<br>　　　　　\|<br>Fucα-2Galβ-3GlcNAcβ-3 | G24504JY |
| 111 | **GSC-216** | GlcAβ-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer42 | G81408RV |
| 112 | **GSC-217** | SU-3GlcAβ-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer42 | G10503MP |
| 113 | **GSC-218** | GlcAβ-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36 | G81408RV |

| 114 | **GSC-219** | `SU-3GlcAβ-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36` | G10503MP |
|---|---|---|---|
| 115 | **GSC-915** | <pre>Fucα-2Galβ-4GlcNAcβ-6<br>              &#124;<br>              Galβ-4GlcNAcβ-3Galβ-4Glc-DH<br>              &#124;<br>  Galβ-4GlcNAcβ-3</pre> | G54388ZL |
| 116 | **GSC-915-AO** | <pre>Fucα-2Galβ-4GlcNAcβ-6<br>              &#124;<br>              Galβ-4GlcNAcβ-3Galβ-4Glc-AO<br>              &#124;<br>  Galβ-4GlcNAcβ-3</pre> | G54388ZL |
| 117 | **GSC-915-2** | <pre>Fucα-2Galβ-4GlcNAcβ-6<br>              &#124;<br>              Galβ-4GlcNAcβ-3Galβ-4Glc-DH<br>              &#124;<br>        GlcNAcβ-3</pre> | G95493AR |
| 118 | **GSC-915-3** | `Fucα-2Galβ-4GlcNAcβ-6Galβ-4GlcNAcβ-3Galβ-4Glc-DH` | G61756VD |
| 119 | **GSC-915-4 (new)** | `Galβ-4GlcNAcβ-6Galβ-4GlcNAcβ-3Galβ-4Glc-DH` | G95676DI |
| 120 | **GSC-915-5** | `GlcNAcβ-6Galβ-4GlcNAcβ-3Galβ-4Glc-DH` | G85628BP |
| 121 | **PSM-F1H2HN3-OY** | <pre>Fucα-2Galβ-4GlcNAcβ-6<br>              &#124;<br>              Galβ-4GlcNAcβ-OY<br>              &#124;<br>        GlcNAcβ-3</pre> | G66089AB |
| 122 | **LNnO** | `Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH` | G85735NR |
| 123 | **Orsay-5-AO** | <pre>  GlcNAcβ-6<br>          &#124;<br>          Gal-AO<br>          &#124;<br>Galβ-3GlcNAcβ-3</pre> | G99775NW |
| 124 | **Orsay-6-AO** | <pre>Galβ-4GlcNAcβ-6<br>              &#124;<br>              Gal-AO<br>              &#124;<br>Galβ-3GlcNAcβ-3</pre> | G00594CD |
| 125 | **Orsay-7-AO** | <pre>Galβ-4GlcNAcβ-6<br>              &#124;<br>              Gal-AO<br>              &#124;<br>Galβ-4GlcNAcβ-3</pre> | G46885EU |
| 126 | **LNH** | <pre>Galβ-4GlcNAcβ-6<br>              &#124;<br>              Galβ-4Glc-DH<br>              &#124;<br>Galβ-3GlcNAcβ-3</pre> | G62056VT |
| 127 | **iLNO** | <pre>Galβ-3GlcNAcβ-3Galβ-4GlcNAcβ-6<br>                              &#124;<br>                              Galβ-4Glc-DH<br>                              &#124;<br>              Galβ-3GlcNAcβ-3</pre> | G16900UB |
| 128 | **iLNO-AD** | <pre>Galβ-3GlcNAcβ-3Galβ-4GlcNAcβ-6<br>                              &#124;<br>                              Galβ-4Glc-AD<br>                              &#124;<br>              Galβ-3GlcNAcβ-3</pre> | G16900UB |
| 129 | **LND** | <pre>Galβ-4GlcNAcβ-6<br>              &#124;<br>              Galβ-4GlcNAcβ-6<br>              &#124;<br>Galβ-3GlcNAcβ-3        Galβ-4Glc-DH<br>                              &#124;<br>              Galβ-3GlcNAcβ-3</pre> | G06511XS |
| 130 | **LNnH** | <pre>Galβ-4GlcNAcβ-6<br>              &#124;<br>              Galβ-4Glc-DH<br>              &#124;<br>Galβ-4GlcNAcβ-3</pre> | G54021WX |

| | | | |
|---|---|---|---|
| 131 | **MFiLNO-IV-AD** | Galβ-3GlcNAcβ-3Galβ-4GlcNAcβ-6<br>      \|<br>     Fucα-3     Galβ-4Glc-AD<br>    Galβ-3GlcNAcβ-3 | G48120WX |
| 132 | **DFiLNO** | Galβ-3GlcNAcβ-3Galβ-4GlcNAcβ-6<br>      \|     \|<br>     Fucα-3     Galβ-4Glc-DH<br>  Fucα-2Galβ-3GlcNAcβ-3 | G41868HL |
| 133 | **DFiLNO-AD** | Galβ-3GlcNAcβ-3Galβ-4GlcNAcβ-6<br>      \|     \|<br>     Fucα-3     Galβ-4Glc-AD<br>  Fucα-2Galβ-3GlcNAcβ-3 | G41868HL |
| 134 | **TFiLNO(1-2,2,3)** | Fucα-2Galβ-3GlcNAcβ-3Galβ-4GlcNAcβ-6<br>       \|     \|<br>      Fucα-3     Galβ-4Glc-DH<br>   Fucα-2Galβ-3GlcNAcβ-3 | G22402PS |
| 135 | **TFiLNO(1-2,2,3)-AD** | Fucα-2Galβ-3GlcNAcβ-3Galβ-4GlcNAcβ-6<br>       \|     \|<br>      Fucα-3     Galβ-4Glc-AD<br>   Fucα-2Galβ-3GlcNAcβ-3 | G22402PS |
| 136 | **Nonaosylceramide** | GlcNAcβ-6<br>   \|<br>GlcNAcβ-6     Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer<br>  \|<br> Galβ-4GlcNAcβ-3<br>  \|<br>GlcNAcβ-3 | G36571WY |
| 137 | **I-octaosylceramide** | Galβ-4GlcNAcβ-6<br>    \|<br>  Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer<br>    \|<br>Galβ-4GlcNAcβ-3 | G98726UE |
| 138 | **H3 (with H3-Fuc)*** | Fucα-2Galβ-4GlcNAcβ-6<br>      \|<br>    Galβ-4GlcNAcβ-3Galβ-4Glc-Cer<br>      \|<br>Fucα-2Galβ-4GlcNAcβ-3 | G41082WI |
| 139 | **Ad (with Ad+Fuc)*** | GalNAcα-3Galβ-4GlcNAcβ-6<br>      \|<br>   Fucα-2     Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer<br>       \|<br>GalNAcα-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3<br> \|<br> Fucα-2 | G79776GW |
| 140 | **I-dodecaosylceramide** | Galβ-4GlcNAcβ-6<br>    \|<br>Galβ-4GlcNAcβ-6     Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer<br>  \|<br> Galβ-4GlcNAcβ-3<br>  \|<br>Galβ-4GlcNAcβ-3 | G75919PR |
| 141 | **I-hexadecaosylceramide** | Galβ-4GlcNAcβ-6<br>    \|<br>Galβ-4GlcNAcβ-6     Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer<br>  \|<br>Galβ-4GlcNAcβ-6     Galβ-4GlcNAcβ-3<br>  \|<br> Galβ-4GlcNAcβ-3<br>  \|<br>Galβ-4GlcNAcβ-3 | G09960SS |

| 142 | **I-eicosaosylceramide** | Galβ-4GlcNAcβ-6<br>        \|<br>    Galβ-4GlcNAcβ-6             Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer<br>        \|                 \|<br>Galβ-4GlcNAcβ-6        Galβ-4GlcNAcβ-3<br>    \|            \|<br>Galβ-4GlcNAcβ-6    Galβ-4GlcNAcβ-3<br>    \|       \|<br>    Galβ-4GlcNAcβ-3<br>    \|<br>Galβ-4GlcNAcβ-3 | G15713JH |
| 143 | **B-like decaosylceramide** | Galα-3Galβ-4GlcNAcβ-6<br>        \|<br>    Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer<br>    \|<br>Galα-3Galβ-4GlcNAcβ-3 | G14974EY |
| 144 | **B-like pentadecaosylceramide** | Galα-3Galβ-4GlcNAcβ-6<br>        \|<br>Galα-3Galβ-4GlcNAcβ-6       Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer<br>    \|<br>    Galβ-4GlcNAcβ-3<br>    \|<br>Galα-3Galβ-4GlcNAcβ-3 | G43365VM |
| 145 | **B-like eicosaosylceramide** | Galα-3Galβ-4GlcNAcβ-6<br>        \|<br>Galα-3Galβ-4GlcNAcβ-6      Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer<br>    \|            \|<br>Galα-3Galβ-4GlcNAcβ-6    Galβ-4GlcNAcβ-3<br>    \|<br>    Galβ-4GlcNAcβ-3<br>    \|<br>Galα-3Galβ-4GlcNAcβ-3 | G12460AY |
| 146 | **B-like pentaeicosaosylceramide** | Galα-3Galβ-4GlcNAcβ-6<br>        \|<br>Galα-3Galβ-4GlcNAcβ-6      Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer<br>    \|            \|<br>Galα-3Galβ-4GlcNAcβ-6    Galβ-4GlcNAcβ-3<br>    \|       \|<br>Galα-3Galβ-4GlcNAcβ-6    Galβ-4GlcNAcβ-3<br>    \|<br>    Galβ-4GlcNAcβ-3<br>    \|<br>Galα-3Galβ-4GlcNAcβ-3 | G28658YC |
| 147 | **pLNFH-IV** | Galβ-3GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH<br>          \|<br>        Fucα-3 | G25471IM |
| 148 | **DFpLNH-II** | Galβ-3GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH<br>  \|           \|<br>Fucα-4      Fucα-3 | G66892DM |
| 149 | **TFpLNH-I** | Fucα-2Galβ-3GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH<br>      \|           \|<br>    Fucα-4      Fucα-3 | G50182HI |
| 150 | **MFLNH-III** | Galβ-4GlcNAcβ-6<br>  \|        \|<br>Fucα-3     Galβ-4Glc-DH<br>          \|<br>Galβ-3GlcNAcβ-3 | G58399HX |
| 151 | **DFLNH(b)** | Galβ-4GlcNAcβ-6<br>  \|        \|<br>Fucα-3     Galβ-4Glc-DH<br>          \|<br>Galβ-3GlcNAcβ-3<br>    \|<br>Fucα-4 | G32679VL |
| 152 | **DFLNH(c)** | Galβ-4GlcNAcβ-6<br>        \|<br>    Galβ-4Glc-DH<br>      \|<br>Fucα-2Galβ-3GlcNAcβ-3<br>     \|<br>    Fucα-4 | G17140YY |

| | | | |
|---|---|---|---|
| 153 | **DFLNH(a)** | ```
                    Galβ-4GlcNAcβ-6
                          |
         Fucα-3              Galβ-4Glc-DH
                               |
Fucα-2Galβ-3GlcNAcβ-3
``` | G42436GS |
| 154 | **TFLNH** | ```
                    Galβ-4GlcNAcβ-6
                          |
         Fucα-3              Galβ-4Glc-DH
                               |
Fucα-2Galβ-3GlcNAcβ-3
            |
          Fucα-4
``` | G51855ZY |
| 155 | **MFiLNO-IV** | ```
Galβ-3GlcNAcβ-3Galβ-4GlcNAcβ-6
                     |
         Fucα-3          Galβ-4Glc-DH
                          |
       Galβ-3GlcNAcβ-3
``` | G48120WX |
| 156 | **TFiLNO** | ```
Galβ-3GlcNAcβ-3Galβ-4GlcNAcβ-6
         |           |        |
     Fucα-4      Fucα-3    Galβ-4Glc-DH
                              |
          Galβ-3GlcNAcβ-3
                 |
               Fucα-4
``` | G32773JJ |
| 157 | **MFLND** | ```
Galβ-4GlcNAcβ-6
       |         |
    Fucα-3    Galβ-4GlcNAcβ-6
               |              |
Galβ-3GlcNAcβ-3            Galβ-4Glc-DH
                              |
          Galβ-3GlcNAcβ-3
``` | G94431LX |
| 158 | **MFLNnH(a)** | ```
Galβ-4GlcNAcβ-6
       |       |
    Fucα-3   Galβ-4Glc-DH
              |
Galβ-4GlcNAcβ-3
``` | G09962WA |
| 159 | **DFLNnH** | ```
Galβ-4GlcNAcβ-6
       |       |
    Fucα-3   Galβ-4Glc-DH
              |
Galβ-4GlcNAcβ-3
       |
    Fucα-3
``` | G63053GR |
| 160 | **B-III dodecaosylceramide** | ```
Galα-3Galβ-4GlcNAcβ-6
       |             |
    Fucα-2         Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer
Galα-3Galβ-4GlcNAcβ-3
       |
    Fucα-2
``` | G15059WI |
| 161 | **B-IV tetradecaosylceramide** | ```
              Galα-3Galβ-4GlcNAcβ-6
                     |              |
                  Fucα-2          Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer
Galα-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3
                     |
                  Fucα-2
``` | G20265RJ |
| 162 | **Man2(α2)** | `Manα-2Man-DH` | G29313LW |
| 163 | **Man2(α3)** | `Manα-3Man-DH` | G07490EK |
| 164 | **Man2(α6)** | `Manα-6Man-DH` | G51379UR |
| 165 | **Man3(α3,α6)** | ```
Manα-6Man-DH
   |
 Manα-3
``` | G80926OA |
| 166 | **Man5(α3,α6)** | ```
       Manα-3
          |
Manα-6Manα-6Man-DH
          |
        Manα-3
``` | G51964GO |
| 167 | **Man1GN1** | `Manβ-4GlcNAc-DH` | G24464VD |

| 168 | **Man2GN1** | Manα-3Manβ-4GlcNAc-DH | G47023PF |
|---|---|---|---|
| 169 | **Man2aGN2** | Manα-6Manβ-4GlcNAcβ-4GlcNAc-DH | G05108UT |
| 170 | **Man3GN2** | Manα-6<br>  \|<br>  Manβ-4GlcNAcβ-4GlcNAc-DH<br>  \|<br>Manα-3 | G20624LQ |
| 171 | **Man4aGN2** | Manα-3Manα-6<br>  \|<br>  Manβ-4GlcNAcβ-4GlcNAc-DH<br>  \|<br>  Manα-3 | G40915GX |
| 172 | **Man4bGN2** | Manα-6<br> \|<br>Manα-3Manα-6<br>  \|<br>  Manβ-4GlcNAcβ-4GlcNAc-DH | G56756FT |
| 173 | **Man5GN2** | Manα-6<br> \|<br>Manα-3Manα-6<br>  \|<br>  Manβ-4GlcNAcβ-4GlcNAc-DH<br>  \|<br>  Manα-3 | G03652TR |
| 174 | **Man6GN2** | Manα-6<br> \|<br>Manα-3Manα-6<br>  \|<br>  Manβ-4GlcNAcβ-4GlcNAc-DH<br>  \|<br>Manα-2Manα-3 | G19747OA |
| 175 | **Man7(D1)GN2** | Manα-6<br> \|<br>Manα-3Manα-6<br>  \|<br>  Manβ-4GlcNAcβ-4GlcNAc-DH<br>  \|<br>Manα-2Manα-2Manα-3 | G40198OD |
| 176 | **Man7(D1)GN2-AO** | Manα-6<br> \|<br>Manα-3Manα-6<br>  \|<br>  Manβ-4GlcNAcβ-4GlcNAc-AO<br>  \|<br>Manα-2Manα-2Manα-3 | G40198OD |
| 177 | **Man7(D3)GN2** | Manα-2Manα-6<br> \|<br>Manα-3Manα-6<br>  \|<br>  Manβ-4GlcNAcβ-4GlcNAc-DH<br>  \|<br>Manα-2Manα-3 | G86053IC |
| 178 | **Man8(D1D3)GN2** | Manα-2Manα-6<br> \|<br>Manα-3Manα-6<br>  \|<br>  Manβ-4GlcNAcβ-4GlcNAc-DH<br>  \|<br>Manα-2Manα-2Manα-3 | G86385RC |
| 179 | **Man9GN2** | Manα-2Manα-6<br> \|<br>Manα-2Manα-3Manα-6<br>  \|<br>  Manβ-4GlcNAcβ-4GlcNAc-DH<br>  \|<br>Manα-2Manα-2Manα-3 | G56202TA |

| | | | |
|---|---|---|---|
| 180 | **Man9GN2-AO** | Manα-2Manα-6<br>       &#124;<br>Manα-2Manα-3Manα-6<br>             &#124;<br>           Manβ-4GlcNAcβ-4GlcNAc-AO<br>             &#124;<br>Manα-2Manα-2Manα-3 | G56202TA |
| 181 | **Glc1Man9GN2** |     Manα-2Manα-6<br>          &#124;<br>        Manα-6<br>         &#124;   &#124;<br>    Manα-2Manα-3   Manβ-4GlcNAcβ-4GlcNAc-DH<br>                 &#124;<br>Glcα-3Manα-2Manα-2Manα-3 | G82207VX |
| 182 | **Glc1Man9GN2-AO** |     Manα-2Manα-6<br>          &#124;<br>        Manα-6<br>         &#124;   &#124;<br>    Manα-2Manα-3   Manβ-4GlcNAcβ-4GlcNAc-AO<br>                 &#124;<br>Glcα-3Manα-2Manα-2Manα-3 | G82207VX |
| 183 | **Glc2Man7(D1)GN1-AO** |         Manα-6<br>          &#124;<br>      Manα-3Manα-6<br>             &#124;<br>          Manβ-4GlcNAc-AO<br>          &#124;<br>Glcα-3Glcα-3Manα-2Manα-2Manα-3 | G59454TE |
| 184 | **Glc3Man7(D1)GN1-AO** |         Manα-6<br>          &#124;<br>      Manα-3Manα-6<br>             &#124;<br>          Manβ-4GlcNAc-AO<br>          &#124;<br>Glcα-2Glcα-3Glcα-3Manα-2Manα-2Manα-3 | G40350YP |
| 185 | **Man3XylGN2** | Manα-6<br>  &#124;<br>Xylβ-2Manβ-4GlcNAcβ-4GlcNAc-DH<br>   &#124;<br>  Manα-3 | G80513FH |
| 186 | **N1** | Galβ-4GlcNAcβ-2Manα-6      Fucα-6<br>            &#124;         &#124;<br>         Manβ-4GlcNAcβ-4GlcNAc-DH<br>          &#124;<br>      Manα-3 | G39893KL |
| 187 | **N2** |         Manα-6<br>         &#124;<br>     Manβ-4GlcNAcβ-4GlcNAc-DH<br>       &#124;<br>Galβ-4GlcNAcβ-2Manα-3 | G86386QT |
| 188 | **N2** |         Manα-6<br>         &#124;<br>     Manβ-4GlcNAcβ-4GlcNAc-DH<br>       &#124;<br>Galβ-4GlcNAcβ-2Manα-3 | G86386QT |
| 189 | **N4** | Galβ-4GlcNAcβ-2Manα-6<br>            &#124;<br>         Manβ-4GlcNAcβ-4GlcNAc-DH<br>          &#124;<br>      Manα-3 | G10106TL |
| 190 | **GlcNac2Man3-AO** | GlcNAcβ-2Manα-6<br>         &#124;<br>      Man-AO<br>      &#124;<br>GlcNAcβ-2Manα-3 | G63059VZ |
| 191 | **N3** |      GlcNAcβ-2Manα-6      Fucα-6<br>              &#124;        &#124;<br>Galβ-4         Manβ-4GlcNAcβ-4GlcNAc-DH<br>           &#124;<br>     GlcNAcβ-2Manα-3 | G17389UM |

| 192 | NGA2 | ```
GlcNAcβ-2Manα-6
              |
               Manβ-4GlcNAcβ-4GlcNAc-DH
              |
GlcNAcβ-2Manα-3
``` | G88876JQ |
|-----|------|------|----------|
| 193 | NGA2B | ```
GlcNAcβ-2Manα-6
            |
             GlcNAcβ-4Manβ-4GlcNAcβ-4GlcNAc-DH
            |
GlcNAcβ-2Manα-3
``` | G76241TN |
| 194 | NGA3B | ```
GlcNAcβ-2Manα-6
            |
             GlcNAcβ-4Manβ-4GlcNAcβ-4GlcNAc-DH
            |
GlcNAcβ-4Manα-3
       |
  GlcNAcβ-2
``` | G45256SO |
| 195 | NGA4 | ```
  GlcNAcβ-6
       |
GlcNAcβ-2Manα-6
              |
               Manβ-4GlcNAcβ-4GlcNAc-DH
              |
GlcNAcβ-2Manα-3
       |
  GlcNAcβ-4
``` | G91247WI |
| 196 | NGA5B | ```
  GlcNAcβ-2
       |
GlcNAcβ-4Manα-6
       |      |
GlcNAcβ-6     |
              |
             GlcNAcβ-4Manβ-4GlcNAcβ-4GlcNAc-DH
              |
GlcNAcβ-4Manα-3
       |
  GlcNAcβ-2
``` | G01685IB |
| 197 | GNMan5BGN2 | ```
  Manα-6
     |
  Manα-3Manα-6
            |
             GlcNAcβ-4Manβ-4GlcNAcβ-4GlcNAc-DH
            |
GlcNAcβ-2Manα-3
``` | G98602UY |
| 198 | NA2 | ```
Galβ-4GlcNAcβ-2Manα-6
                    |
                     Manβ-4GlcNAcβ-4GlcNAc-DH
                    |
Galβ-4GlcNAcβ-2Manα-3
``` | G66741YQ |
| 199 | NA3 | ```
Galβ-4GlcNAcβ-2Manα-6
                    |
                     Manβ-4GlcNAcβ-4GlcNAc-DH
                    |
Galβ-4GlcNAcβ-4Manα-3
             |
  Galβ-4GlcNAcβ-2
``` | G24604HV |
| 200 | NA4 | ```
Galβ-4GlcNAcβ-6
             |
Galβ-4GlcNAcβ-2Manα-6
                    |
                     Manβ-4GlcNAcβ-4GlcNAc-DH
                    |
Galβ-4GlcNAcβ-4Manα-3
             |
  Galβ-4GlcNAcβ-2
``` | G09568JT |
| 201 | Fuc-GlcNAc | Fucα-6GlcNAc-DH | G46208LD |
| 202 | Man3FGN2 | ```
Manα-6        Fucα-6
     |           |
      Manβ-4GlcNAcβ-4GlcNAc-DH
     |
Manα-3
``` | G45995IV |

| | | | |
|---|---|---|---|
| 203 | **Man3FXylGN2** | ```
                    Manα-6
                     |
             Xylβ-2Manα-4GlcNAcβ-4GlcNAc-DH
                     |               |
                   Manα-3         Fucα-3
``` | G89349EJ |
| 204 | **NGA2F** | ```
GlcNAcβ-2Manα-6            Fucα-6
           |               |
           Manβ-4GlcNAcβ-4GlcNAc-DH
           |
GlcNAcβ-2Manα-3
``` | G65984FE |
| 205 | **NA2F** | ```
Galβ-4GlcNAcβ-2Manα-6          Fucα-6
                 |             |
                 Manβ-4GlcNAcβ-4GlcNAc-DH
                 |
Galβ-4GlcNAcβ-2Manα-3
``` | G00998NI |
| 206 | **NA2F-AO** | ```
Galβ-4GlcNAcβ-2Manα-6          Fucα-6
                 |             |
                 Manβ-4GlcNAcβ-4GlcNAc-AO
                 |
Galβ-4GlcNAcβ-2Manα-3
``` | G00998NI |
| 207 | **NA2FB** | ```
Galβ-4GlcNAcβ-2Manα-6          Fucα-6
                 |             |
                 GlcNAcβ-4Manβ-4GlcNAcβ-4GlcNAc-DH
                 |
Galβ-4GlcNAcβ-2Manα-3
``` | G74325TG |
| 208 | **NA3-Lex** | ```
    Galβ-4GlcNAcβ-2Manα-6
                     |
Fucα-3               Manβ-4GlcNAcβ-4GlcNAc-DH
    Galβ-4GlcNAcβ-4Manα-3
                     |
      Galβ-4GlcNAcβ-2
``` | G57601XR |
| 209 | **Glc2Man-AO** | Glcα-3Glcα-3Man-AO | G42254JC |
| 210 | **Glc2Man2-AO** | Glcα-3Glcα-3Manα-2Man-AO | G88860KA |
| 211 | **Glc2Man3-AO** | Glcα-3Glcα-3Manα-2Manα-2Man-AO | G66733GQ |
| 212 | **(6P)Man5GN2.AO** | ```
P-6Manα-6
       |
      Manα-6
       |     |
     Manα-3  Manβ-4GlcNAcβ-4GlcNAc-AO
             |
           Manα-3



P-6Manα-6
       |
      Manα-6
       |     |
     Manα-3  Manβ-4GlcNAcβ-4GlcNAc-AO
             |
      Manα-2Manα-3
``` | G90966QP<br>G88621MY |
| 213 | **(6P)Man6GN2.AO** | ```
P-6Manα-6
       |
      Manα-6
       |     |
     Manα-3  Manβ-4GlcNAcβ-4GlcNAc-AO
             |
      Manα-2Manα-3
``` | G88621MY |
| 214 | **SM3** | SU-3Galβ-4Glcβ-Cer | G20044TD |
| 215 | **Asialo-GM2** | GalNAcβ-4Galβ-4Glcβ-Cer | G90562PB |
| 216 | **SB2** | ```
SU-3GalNAcβ-4Galβ-4Glcβ-Cer
         |
       SU-3
``` | G28868PL |
| 217 | **Asialo-GM1** | Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer | G21856LC |
| 218 | **Asialo-GM1-Tetra** | Galβ-3GalNAcβ-4Galβ-4Glc-DH | G00407UJ |
| 219 | **SM1a** | ```
Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer
         |
       SU-3
``` | G59508TM |

| | | | |
|---|---|---|---|
| 220 | SB1a | SU-3Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer<br>&#124;<br>SU-3 | G63813EJ |
| 221 | GalNAc-Ser | GalNAcα-Ser-DH | G57321FI |
| 222 | GalNAc-Thr | GalNAcα-Thr-DH | G57321FI |
| 223 | GalNAcα-3GalNAc | GalNAcα-3GalNAc-DH | G45388GV |
| 224 | Galβ-3GalNAc | Galβ-3GalNAc-DH | G78436BB |
| 225 | Galβ-3GalNAc-AO | Galβ-3GalNAc-AO | G78436BB |
| 226 | Galβ-6GalNAc | Galβ-6GalNAc-DH | G27589TC |
| 227 | Galβ-6GalNAc-AO | Galβ-6GalNAc-AO | G27589TC |
| 228 | Man-Ser | Manα-Ser-DH | G61491DK |
| 229 | Man-Ser-Succ | Manα-Ser-Succ-DH | G61491DK |
| 230 | Man-Thr | Manα-Thr-DH | G61491DK |
| 231 | Man-Thr-Succ | Manα-Thr-Succ-DH | G61491DK |
| 232 | A8/1 | GlcNAcα-4Galβ-OX | G94473FP |
| 233 | A8/2 | SU-6<br>&#124;<br>Fucα-3GlcNAcβ-OY | G82872XE |
| 234 | A15/1 | SU-6GlcNAcβ-OY | G17441VE |
| 235 | A15/3 | GlcNAcα-4Galβ-3Galβ-OX<br>&#124;<br>Fucα-2 | G56757GN |
| 236 | B13/a | GlcAβ-3Galβ-3GlcNAcβ-OX | G83866MJ |
| 237 | Notch-1 | Fucα-Thr-DH | G96881BQ |
| 238 | Notch-2 | GlcNAcβ-3Fucα-Thr | G26238GL |
| 239 | Notch-3 | Galβ-4GlcNAcβ-3Fucα-Thr-DH | G60605ZN |
| 240 | XylGlc-AO | Xylα-3Glc-AO | G61224OH |
| 241 | Xyl2Glc-AO | Xylα-3Xylα-3Glc-AO | G86535WT |
| 242 | GlcNAcβ-3Fuc-AO | GlcNAcβ-3Fuc-AO | G85661BQ |
| 243 | GlcNAcβ1-2Fuc-AO | GlcNAcβ-2Fuc-AO | G32795LG |
| 244 | GlcNAcβ1-4Fuc-AO | GlcNAcβ-4Fuc-AO | G10008BD |
| 245 | GlcNAcβ-2Man-AO | GlcNAcβ-2Man-AO | G65609XC |
| 246 | Hep-2-AO* | SU-2<br>&#124;<br>ΔUA-4GlcNS-AO<br>&#124;<br>SU-6 | G01871HO |
| 247 | Hep-4-AO* | SU-2<br>&#124;<br>ΔUA-4GlcNSα-4IdoAα-4GlcNS-AO<br>&#124;    &#124;    &#124;<br>SU-6  SU-2  SU-6 | G58161DR |
| 248 | Hep-6-AO* | SU-2<br>&#124;<br>ΔUA-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNS-AO<br>&#124;   &#124;   &#124;   &#124;   &#124;<br>SU-6  SU-2  SU-6  SU-2  SU-6 | G23622XF |
| 249 | Hep-8-AO* | SU-2<br>&#124;<br>ΔUA-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNS-DH<br>&#124;  &#124;  &#124;  &#124;  &#124;  &#124;  &#124;<br>SU-6 SU-2 SU-6 SU-2 SU-6 SU-2 SU-6 | G73721GC |
| 250 | Hep-10-AO* | SU-2<br>&#124;<br>ΔUA-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNS-AO<br>&#124;  &#124;  &#124;  &#124;  &#124;<br>SU-6 SU-2 SU-6 SU-2 SU-6 | G10593KB |

| | | | |
|---|---|---|---|
| 251 | **Hep-12-AO*** | SU-2<br>　\|<br>ΔUA-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNS-AO<br>　　\|　　　\|　　　\|　　　\|　　　\|　　　\|　　　\|　　　\|　　　\|　　　\|　　　\|<br>　SU-6　SU-2　SU-6　SU-2　SU-6　SU-2　SU-6　SU-2　SU-6　SU-2　SU-6 | G28885LI |
| 252 | **Hep-14-AO*** | SU-2<br>　\|<br>ΔUA-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNS-AO<br>　　\|　　\|　　\|　　\|　　\|　　\|　　\|　　\|　　\|　　\|　　\|　　\|　　\|<br>　SU-6　SU-2　SU-6　SU-2　SU-6　SU-2　SU-6　SU-2　SU-6　SU-2　SU-6　SU-2　SU-6 | G93186IG |
| 253 | **Hep-16-AO*** | SU-2<br>　\|<br>ΔUA-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNS-AO<br>　　\|　　\|　　\|　　\|　　\|　　\|　　\|　　\|　　\|　　\|　　\|　　\|　　\|<br>　SU-6　SU-2　SU-6　SU-2　SU-6　SU-2　SU-6　SU-2　SU-6　SU-2　SU-6　SU-2　SU-6 | G63800GK |
| 254 | **Hep-18-AO*** | SU-2<br>　\|<br>ΔUA-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNS-AO<br>　　\|　　\|　　\|　　\|　　\|　　\|　　\|　　\|　　\|　　\|　　\|　　\|　　\|　　\|<br>　SU-6　SU-2　SU-6　SU-2　SU-6　SU-2　SU-6　SU-2　SU-6　SU-2　SU-6　SU-2　SU-6　SU-2　SU-6 | G39322DH |
| 255 | **Hep-20-AO*** | SU-2<br>　\|<br>ΔUA-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNSα-4IdoAα-4GlcNS-AO<br>　　\|　　\|　　\|　　\|　　\|　　\|　　\|　　\|　　\|　　\|　　\|　　\|　　\|　　\|　　\|<br>　SU-6　SU-2　SU-6　SU-2　SU-6　SU-2　SU-6　SU-2　SU-6　SU-2　SU-6　SU-2　SU-6　SU-2　SU-6　SU-2　SU-6 | G66524OT |
| 256 | **GN2-AO** | GlcNAcβ-4GlcNAc-AO | G07982RO |
| 257 | **GN3** | GlcNAcβ-4GlcNAcβ-4GlcNAc-DH | G81262CB |
| 258 | **GN3-AO** | GlcNAcβ-4GlcNAcβ-4GlcNAc-AO | G81262CB |
| 259 | **GN4-AO*** | GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAc-AO | G98511QW |
| 260 | **GN5-AO*** | GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAc-AO | G90142YY |
| 261 | **GN7-AO*** | GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAc-AO | G45618KR |
| 262 | **GN8-AO*** | GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAcβ-4GlcNAc-AO | G92110HD |
| 263 | **GlcN4-AO (chitosan)** | GlcNβ-4GlcNβ-4GlcNβ-4GlcN-AO | G16905EJ |
| 264 | **GlcN5-AO (chitosan)** | GlcNβ-4GlcNβ-4GlcNβ-4GlcNβ-4GlcN-AO | G19249DC |
| 265 | **GlcN6-AO (chitosan)** | GlcNβ-4GlcNβ-4GlcNβ-4GlcNβ-4GlcNβ-4GlcN-AO | G40222HN |
| 266 | **GlcNAc2(1-6)-AO** | GlcNACβ-6GlcNAc-AO | G33817CT |
| 267 | **Malto-2-AO** | Glcα-4Glc-AO | G41366RG |
| 268 | **Malto-3-AO** | Glcα-4Glcα-4Glc-AO | G33575EC |
| 269 | **Malto-4-AO** | Glcα-4Glcα-4Glcα-4Glc-AO | G78800HY |
| 270 | **Malto-6-AO** | Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glc-AO | G03116ET |
| 271 | **Malto-7-AO** | Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glc-AO | G49764YU |
| 272 | **Malto-8-AO*** | Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glc-AO | G59303OX |
| 273 | **Malto-9-AO*** | Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glc-AO | G25434YA |
| 274 | **Malto-10-AO*** | Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glc-AO | G04793KB |
| 275 | **Malto-11-AO*** | Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glc-AO | G88822KL |
| 276 | **Malto-12-AO*** | Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glc-AO | G85465ZH |
| 277 | **Malto-13-AO*** | Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glc-AO | G86630CJ |
| 278 | **Malto-2** | Glcα-4Glc-DH | G41366RG |
| 279 | **Malto-3** | Glcα-4Glcα-4Glc-DH | G33575EC |
| 280 | **Malto-4** | Glcα-4Glcα-4Glcα-4Glc-DH | G78800HY |
| 281 | **Malto-6** | Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glc-DH | G03116ET |
| 282 | **Malto-7** | Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glcα-4Glc-DH | G49764YU |

| 283 | **Dext-2-AO** | Glcα-6Glc-AO | G39326SA |
| 284 | **Dext-3-AO** | Glcα-6Glcα-6Glc-AO | G33569DD |
| 285 | **Dext-4-AO** | Glcα-6Glcα-6Glcα-6Glc-AO | G86640EE |
| 286 | **Dext-5-AO*** | Glcα-6Glcα-6Glcα-6Glcα-6Glc-AO | G42008QK |
| 287 | **Dext-6-AO*** | Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glc-AO | G73348SI |
| 288 | **Dext-8-AO*** | Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glc-AO | G39376AB |
| 289 | **Dext-9-AO*** | Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glc-AO | G16611XN |
| 290 | **Dext-10-AO*** | Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glc-AO | G13237DH |
| 291 | **Dext-11-AO*** | Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glc-AO | G21485BC |
| 292 | **Dext-12-AO*** | Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glc-AO | G48520ZB |
| 293 | **Dext-13-AO*** | Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glc-AO | G84102VI |
| 294 | **Dext-2** | Glcα-6Glc-DH | G39326SA |
| 295 | **Dext-3** | Glcα-6Glcα-6Glc-DH | G33569DD |
| 296 | **Dext-4** | Glcα-6Glcα-6Glcα-6Glc-DH | G86640EE |
| 297 | **Dext-6*** | Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glc-DH | G73348SI |
| 298 | **Dext-7*** | Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glcα-6Glc-DH | G00936LV |
| 299 | **Lam-2-AO** | Glcβ-3Glc-AO | G52148JS |
| 300 | **Lam-3-AO** | Glcβ-3Glcβ-3Glc-AO | G44414OJ |
| 301 | **Lam-4-AO** | Glcβ-3Glcβ-3Glcβ-3Glc-AO | G45140IE |
| 302 | **Lam-5-AO** | Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO | G59360UC |
| 303 | **Lam-6-AO*** | Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO | G28165JK |
| 304 | **Lam-7-AO** | Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO | G81788HU |
| 305 | **Curd-8-AO*** | Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO | G37467FA |
| 306 | **Curd-9-AO*** | Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO | G02401HG |
| 307 | **Curd-10-AO*** | Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO | G47594QI |
| 308 | **Curd-11-AO*** | Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO | G89899RZ |
| 309 | **Curd-12-AO*** | Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO | G17807TO |
| 310 | **Curd-13-AO*** | Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO | G76359NE |
| 311 | **Lam-2** | Glcβ-3Glc-DH | G52148JS |
| 312 | **Lam-3** | Glcβ-3Glcβ-3Glc-DH | G44414OJ |
| 313 | **Lam-6*** | Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-DH | G28165JK |
| 314 | **Lam-7** | Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-DH | G81788HU |
| 315 | **Cello-3-AO** | Glcβ-4Glcβ-4Glc-AO | G41078PD |
| 316 | **Cello-4-AO** | Glcβ-4Glcβ-4Glcβ-4Glc-AO | G65718MY |
| 317 | **Cello-5-AO*** | Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glc-AO | G49725SZ |
| 318 | **Cello-6-AO*** | Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glc-AO | G22490RY |
| 319 | **Cello-7-AO*** | Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glc-AO | G97077OZ |
| 320 | **Cello-8-AO*** | Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glc-AO | G49540SV |
| 321 | **Cello-9-AO*** | Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glc-AO | G89083OA |
| 322 | **Cello-10-AO*** | Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glc-AO | G74022VE |
| 323 | **Cello-11-AO*** | Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glc-AO | G10534ET |
| 324 | **Cello-12-AO*** | Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glc-AO | G66324UC |
| 325 | **Cello-13-AO*** | Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glcβ-4Glc-AO | G34839AN |
| 326 | **Cellobiose** | Glcβ-4Glc-DH | G12460QS |

| 327 | Pust-3-AO | Glcβ-6Glcβ-6Glc-AO | G87091WI |
|-----|-----------|---------------------|----------|
| 328 | Pust-4-AO | Glcβ-6Glcβ-6Glcβ-6Glc-AO | G52763AS |
| 329 | Pust-5-AO | Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glc-AO | G54780QZ |
| 330 | Pust-6-AO | Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glc-AO | G52505EM |
| 331 | Pust-7-AO* | Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glc-AO | G37879VE |
| 332 | Pust-8-AO* | Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glc-AO | G85252SY |
| 333 | Pust-9-AO* | Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glc-AO | G47833CT |
| 334 | Pust-10-AO* | Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glc-AO | G20094BR |
| 335 | Pust-11-AO* | Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glc-AO | G84174KV |
| 336 | Pust-15-AO* | Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glc-AO | G44062ZH |
| 337 | Pust-15a-AO | Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glcβ-6Glc-AO | G44062ZH |
| 338 | Kojibiose | Glcα-2Glc-DH | G86562DX |
| 339 | Cyano-2-AO | Glcα-2Glc-AO | G86562DX |
| 340 | Cyano-3-AO | Glcα-2Glcα-2Glc-AO | G62300UR |
| 341 | Cyano-4-AO | Glcα-2Glcα-2Glcα-2Glc-AO | G61681XZ |
| 342 | Cyano-5-AO | Glcα-2Glcα-2Glcα-2Glcα-2Glc-AO | G60080EX |
| 343 | Cyano-6-AO* | Glcα-2Glcα-2Glcα-2Glcα-2Glcα-2Glc-AO | G97485LN |
| 344 | Cyano-7-AO* | Glcα-2Glcα-2Glcα-2Glcα-2Glcα-2Glcα-2Glc-AO | G92944JP |
| 345 | Cyano-9-AO* | Glcα-2Glcα-2Glcα-2Glcα-2Glcα-2Glcα-2Glcα-2Glc-AO | G02032CC |
| 346 | Nigerose | Glcα-3Glc-DH | G56660NC |
| 347 | Nigerose-AO | Glcα-3Glc-AO | G56660NC |
| 348 | Poria-3-AO | Glcα-3Glcα-3Glc-AO | G54604OG |
| 349 | Poria-4-AO | Glcα1-3Glcα1-3Glcα1-3Glc-AO | G99708XO |
| 350 | Poria-5-AO | Glcα-3Glcα-3Glcα-3Glc-AO | G33834HE |
| 351 | Poria-6-AO | Glcα-3Glcα-3Glcα-3Glcα-3Glc-AO | G00736XN |
| 352 | Poria-7-AO | Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glc-AO | G11787AF |
| 353 | Poria-8-AO* | Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glc-AO | G09921EA |
| 354 | Poria-9-AO* | Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glc-AO | G05698ZI |
| 355 | Poria-10-AO* | Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glα-3Glc-AO | G07079VC |
| 356 | Poria-11-AO* | Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glα-3Glcα-3Glc-AO | G10484MA |
| 357 | Poria-12-AO* | Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glα-3Glcα-3Glcα-3Glc-AO | G85852IT |
| 358 | Poria-13-AO* | Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glcα-3Glα-3Glcα-3Glcα-3Glcα-3Glc-AO | G64473RB |
| 359 | i-Pano-3-AO | Glcα-4Glcα-6Glc-AO | G07348OG |
| 360 | Pullu-4-AO | Glcα-6Glcα-4Glcα-4Glc-AO | G55967QW |
| 361 | Pullu-6-AO | Glcα-4Glcα-4Glcα-6Glcα-4Glcα-4Glc-AO | G78151IM |
| 362 | CβG-2-AO | Glcβ-2Glc-AO | G63873ID |
| 363 | CβG-3-AO | Glcβ-2Glcβ-2Glc-AO | G10430NG |
| 364 | CβG-4-AO | Glcβ-2Glcβ-2Glcβ-2Glc-AO | G21401UQ |
| 365 | CβG-5-AO* | Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glc-AO | G52982UK |
| 366 | CβG-6-AO* | Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glc-AO | G13923FW |
| 367 | CβG-7-AO* | Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glc-AO | G13975CX |
| 368 | CβG-8-AO* | Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glc-AO | G50790RU |
| 369 | CβG-10-AO* | Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glc-AO | G30846SA |
| 370 | CβG-11-AO* | Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glc-AO | G22031CK |

| 371 | **CβG-12-AO*** | Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glc-AO | G54563TN |
|-----|----------------|-----------------------------------------------------------------------------------|----------|
| 372 | **CβG-13-AO*** | Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glcβ-2Glc-AO | G73670AS |
| 373 | **Barley-3-AO** | Glcβ-4Glcβ-3Glc-AO | G25144YW |
| 374 | **Barley-3a-AO** | Glcβ-3Glcβ-4Glc-AO | G70151NU |
| 375 | **Barley-3b-AO** | Glcβ-4Glcβ-3Glc-AO | G25144YW |
| 376 | **Barley-4a-AO** | Glcβ-3Glcβ-4Glcβ-4Glc-AO | G40028TR |
| 377 | **Barley-4b-AO** | Glcβ-4Glcβ-4Glcβ-3Glc-AO | G03074ZG |
| 378 | **Barley-5-AO*** | Glcβ-4Glcβ-4Glcβ-4Glcβ-3Glc-AO | G38036YP |
| 379 | **Barley-5a-AO** | Glcβ-3Glcβ-4Glcβ-6Glcβ-4Glc-AO | G03985SG |
| 380 | **Barley-6-AO*** | (Glcβ-3/4Glcβ)2-4Glcβ-3Glc-AO | G70453EL |
| 381 | **Barley-6a-AO** | Glcβ-3Glcβ-4Glcβ-6Glcβ-4Glcβ-3Glc-AO; and Glcβ-3Glcβ-4Glcβ-4Glcβ-4Glcβ-3Glc-AO (ratio ~1:1) | G19625CFG17119OZ |
| 382 | **Barley-8-AO*** | (Glcβ-4/3Glcβ)3-4Glcβ-3Glc-AO | G39884NV |
| 383 | **Barley-9-AO*** | (Glcβ-4/3Glcβ)4-3Glc-AO | G81330VE |
| 384 | **Barley-10-AO*** | (Glcβ-4/3Glcβ)4-4Glcβ-3Glc-AO | G25123LE |
| 385 | **Barley-11-AO*** | (Glcβ-4/3Glcβ)5-3Glc-AO | G12314ZO |
| 386 | **Barley-12-AO*** | (Glcβ-4/3Glcβ)5-4Glcβ-3Glc-AO | G84562LP |
| 387 | **Barley-13-AO*** | (Glcβ-4/3Glcβ)6-3Glc-AO | G80842OU |
| 388 | **Barley-14-AO*** | (Glcβ-4/3Glcβ)6-4Glcβ-3Glc-AO | G51450PK |
| 389 | **Barley-15-AO*** | (Glcβ-4/3Glcβ)7-3Glc-AO | G02330UW |
| 390 | **Barley-16-AO*** | (Glcβ-4/3Glcβ)7-4Glcβ-3Glc-AO | G74548HF |
| 391 | **Grifo-3-AO*** | Glc3(β-3/β-6)-AO | N/A |
| 392 | **Grifo-4-AO*** | Glc4(β-3/β-6)-AO | N/A |
| 393 | **Grifo-6-AO*** | Glc6(β-3/β-6)-AO | N/A |
| 394 | **Grifo-7-AO*** | Glc7(β-3/β-6)-AO | N/A |
| 395 | **Grifo-8-AO*** | Glc8(β-3/β-6)-AO | N/A |
| 396 | **Grifo-9-AO*** | Glc9(β-3/β-6)-AO | N/A |
| 397 | **Grifo-10-AO*** | Glc10(β-3/β-6)-AO | N/A |
| 398 | **Grifo-11-AO*** | Glc11(β-3/β-6)-AO | N/A |
| 399 | **Grifo-12-AO*** | Glc12(β-3/β-6)-AO | N/A |
| 400 | **Grifo-13-AO*** | Glc13(β-3/β-6)-AO | N/A |
| 401 | **Grifo-14-AO*** | Glc14(β-3/β-6)-AO | N/A |
| 402 | **Grifo-15-AO*** | Glc15(β-3/β-6)-AO | N/A |
| 403 | **Grifo-16-AO*** | Glc16(β-3/β-6)-AO | N/A |
| 404 | **Lentin-2-AO** | Glc2(β-3/β-6)-AO | G67098NW |
| 405 | **Lentin-3-AO** | Glc3(β-3/β-6)-AO | N/A |
| 406 | **Lentin-4-AO*** | Glc4(β-3/β-6)-AO | N/A |
| 407 | **Lentin-5-AO*** | Glc5(β-3/β-6)-AO | N/A |
| 408 | **Lentin-6-AO*** | Glc6(β-3/β-6)-AO | N/A |
| 409 | **Lentin-7-AO*** | Glc7(β-3/β-6)-AO | N/A |
| 410 | **Lentin-8-AO*** | Glc8(β-3/β-6)-AO | N/A |
| 411 | **Lentin-9-AO*** | Glc9(β-3/β-6)-AO | N/A |
| 412 | **Lentin-10-AO*** | Glc10(β-3/β-6)-AO | N/A |
| 413 | **Lentin-11-AO*** | Glc11(β-3/β-6)-AO | N/A |
| 414 | **Lentin-12-AO*** | Glc12(β-3/β-6)-AO | N/A |

| | | | |
|---|---|---|---|
| 415 | **Lentin-13-AO*** | Glc13(β-3/β-6)-AO | N/A |
| 416 | **NSG-11-AO*** | Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO | G89899RZ |
| 417 | **HE-8-AO** | Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO | G37467FA |
| 418 | **HE-9-AO** | Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO | G02401HG |
| 419 | **HE-10-AO** | Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO | G47594QI |
| 420 | **HE-9B7-AO** | Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO<br>    \|<br>  Glcβ-6 | G37927YI |
| 421 | **HE-10B2-AO** | Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO<br>            \|<br>          Glcβ-6 | G20249DD |
| 422 | **HE-10B3-AO** | Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO<br>         \|<br>       Glcβ-6 | G55512MQ |
| 423 | **HE-10B5-AO** | Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO<br>     \|<br>   Glcβ-6 | G50784KX |
| 424 | **HE-10B7-AO** | Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO<br>   \|<br> Glcβ-6 | G44438BG |
| 425 | **Gu-6B1/3-AO** | Glcβ-3Glcα-3Glcβ-3Glc-AO<br>  \|      \|<br>Glcβ-6   Glcβ-6 | G22277CS |
| 426 | **HE-11B3/6-AO** | Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glcβ-3Glc-AO<br>         \|                \|<br>       Glcβ-6            Glcβ-6 | G38565XH |
| 427 | **Man4(β4)** | Manβ-4Manβ-4Manβ-4Man-DH | G01132FQ |
| 428 | **Man6(β4)** | Manβ-4Manβ-4Manβ-4Manβ-4Manβ-4Man-DH | G33969TN |
| 429 | **Xyl5(β4)** | Xylβ-4Xylβ-4Xylβ-4Xylβ-4Xyl-DH | G17192MX |
| 430 | **Xyl6(β4)** | Xylβ-4Xylβ-4Xylβ-4Xylβ-4Xylβ-4Xyl-DH | G95210WK |
| 431 | **Ara6(α5)** | Araα-5Araα-5Araα-5Araα-5Araα-5Ara-DH | G06678GR |
| 432 | **Ara7(α5)** | Araα-5Araα-5Araα-5Araα-5Araα-5Araα-5Ara-DH | G08367GY |
| 433 | **Gal** | Gal-DH | G68158BT |
| 434 | **Gal-AO** | Gal-AO | G68158BT |
| 435 | **GalNAc** | GalNAc-DH | G27025MB |
| 436 | **GalNAc-AO** | GalNAc-AO | G27025MB |
| 437 | **Glc** | Glc-DH | G15021LG |
| 438 | **Glc-AO** | Glc-AO | G15021LG |
| 439 | **GN** | GlcNAc-DH | G64581RP |
| 440 | **GN-AO** | GlcNAc-AO | G64581RP |
| 441 | **Man-AO** | Man-AO | G70323CJ |
| 442 | **Fuc** | Fuc-DH | G82576YO |
| 443 | **Fuc-AO** | Fuc-AO | G82576YO |
| 444 | **Rha** | Rha-DH | G07006BT |
| 445 | **Rha-AO** | Rha-AO | G07006BT |
| 446 | **Galα-6Glc-AO** | Galα-6Glc-AO | G09165XB |
| 447 | **(6P)-Glc-AO** | P-6Glc-AO | G97131OU |
| 448 | **(6P)-Man** | P-6Man-DH | G56756HZ |
| 449 | **(6P)-Man-AO** | P-6Man-AO | G56756HZ |
| 450 | **(6P)-Fructose-AO** | P-6Fru-AO | G00100EG |
| 451 | **SU-Tyr** | SU-Tyr-DH | Non glycan |
| 452 | **SU-Cholesterol** | SU-Cholesterol | Non glycan |

| | | | |
|---|---|---|---|
| 453 | **GN-Asn** | GlcNAcβ-Asn-DH | G49108TO |
| 454 | **Xyl3Glc4** | Xylα-6<br>\|<br>Glcβ-4Glcβ-4Glcβ-4Glc-DH<br>\|      \|<br>Xylα-6     Xylα-6 | G23659XQ |
| 455 | **Glc4(α6,α4,α4)** | Glcα-6Glcα-4Glcα-4Glc-DH | G55967QW |
| 456 | **Glc(α6,α4,α4)-AO** | Glcα-6Glcα-4Glcα-4Glc-AO | G55967QW |
| 457 | **O1-AO** | GlcNAcβ-6<br>\|<br>Galβ-AO<br>\|<br>GlcNAcβ-3 | G57681KL |
| 458 | **Rutinose-AO** | Rhaα-6Glc-AO | G85868UC |
| 459 | **Gal4** | Galα-3Galβ-4Galα-3Gal-DH | G26927BM |
| 460 | **Ceramides** | Cer | Non glycan |
| 461 | **Galα-4Glc-AO** | Galα-4Glc-AO | G00709ZM |
| 462 | **Carra-Tetra-1S** | aGalα-3Galβ-4aGalα-3Gal-DH<br>\|<br>SU-4 | N/A |
| 463 | **Carra-Tetra-2S** | aGalα-3Galβ-4aGalα-3Gal-DH<br>\|      \|<br>SU-4    SU-4 | N/A |
| 464 | **Carra-Hexa-3S** | aGalα-3Galβ-4aGalα-3Galβ-4aGalα-3Gal-DH<br>\|    \|    \|<br>SU-4   SU-4   SU-4 | N/A |
| 465 | **Carra-Hexa-4S** | aGalα-3Galβ-4aGalα-3Galβ-4aGalα-3Gal-DH<br>\|   \|   \|   \|<br>SU-4  SU-2 SU-4   SU-4 | N/A |
| 466 | **Carra-Octa-4S** | aGalα-3Galβ-4aGalα-3Galβ-4aGalα-3Galβ-4aGalα-3Gal-DH<br>\|    \|    \|    \|<br>SU-4   SU-4   SU-4   SU-4 | N/A |
| 467 | **Gal2GlcNAc(1-3)-AO** | Galα-3Galβ-3GlcNAc-AO | G74668VC |
| 468 | **GalManNAc(1-4)-AO** | Galβ-4ManNAc-AO | G61994TY |
| *α 2-3-Sialy* | | | |
| 469 | **GSC-187** | NeuAcα-3Galβ-C29 | G30207PZ |
| 470 | **GSC-40** | NeuAcα-(S)-3Galβ-Cer42 | N/A |
| 471 | **NeuAcα-(3')Lac** | NeuAcα-3Galβ-4Glc-DH | G56516VH |
| 472 | **NeuAcα-(3')Lac-AO** | NeuAcα-3Galβ-4Glc-AO | G56516VH |
| 473 | **Neu4,5Ac-(3')Lac** | (4-OAc)NeuAcα-3Galβ-4Glc-DH | G69801MT |
| 474 | **Neu4,5Ac-(3')Lac-AO** | (4-OAc)NeuAcα-3Galβ-4Glc-AO | G69801MT |
| 475 | **GSC-16** | NeuAcα-3Galβ-4Glcβ-Cer32 | G91237TK |
| 476 | **GSC-178** | NeuAcα-3Galβ-4Glcβ-Cer34 | G91237TK |
| 477 | **GSC-17** | NeuAcα-3Galβ-4Glcβ-Cer36 | G91237TK |
| 478 | **GSC-18** | NeuAcα-3Galβ-4Glcβ-Cer42 | G91237TK |
| 479 | **GSC-197** | KDNα-3Galβ-4Glcβ-Cer28 | G31170TM |
| 480 | **GSC-199** | KDNα-3Galβ-4Glcβ-C30 | G31170TM |
| 481 | **GSC-198** | KDNα-3Galβ-4Glcβ-Cer34 | G31170TM |
| 482 | **GSC-75** | (4-deoxy)NeuAcα-3Galβ-4Glcβ-Cer36 | G84113WS |
| 483 | **GSC-76** | (7-deoxy)NeuAcα-3Galβ-4Glcβ-Cer36 | G95173VV |
| 484 | **GSC-77** | (8-deoxy)NeuAcα-3Galβ-4Glcβ-Cer36 | G76422JK |
| 485 | **GSC-153** | (4,8-deoxy)NeuAcα-3Galβ-4Glcβ-Cer36 | G31733IY |
| 486 | **GSC-51** | (9-deoxy)NeuAcα-3Galβ-4Glcβ-Cer36 | G83530GE |

| 487 | GSC-78 | (4-OMe)NeuAcα-3Galβ-4Glcβ-Cer36 | G01376ME |
|---|---|---|---|
| 488 | GSC-79 | (9-OMe)NeuAcα-3Galβ-4Glcβ-Cer36 | G91884QW |
| 489 | GSC-23 | (C7)NeuAcα-3Galβ-4Glcβ-Cer36 | G35990AX |
| 490 | GSC-24 | (C8)NeuAcα-3Galβ-4Glcβ-Cer36 | G96486QL |
| 491 | GSC-50 | (C8 diastereoisomer)NeuAcα-3Galβ-4Glcβ-Cer36 | G66544BX |
| 492 | Neuα-(3')Lac | Neuα-3Galβ-4Glc-DH | G28472JY |
| 493 | Neuα-(3')Lac-AO | Neuα-3Galβ-4Glc-AO | G28472JY |
| 494 | NeuAcβ-(3')Lac | NeuAcβ-3Galβ-4Glc-DH | G79930NT |
| 495 | NeuAcβ-(3')Lac-AO | NeuAcβ-3Galβ-4Glc-AO | G79930NT |
| 496 | GSC-161 | NeuAcα-3Galβ-4Glcβ-C30<br>&#124;<br>Fucα-3 | G84173YM |
| 497 | GSC-162 | NeuAcα-3Galβ-4Glcβ-Cer36<br>&#124;<br>Fucα-3 | G84173YM |
| 498 | SAA#1 | NeuAcα-3Galβ-3GlcNAc-DH | G59168YE |
| 499 | SAA#1-AO | NeuAcα-3Galβ-3GlcNAc-AO | G59168YE |
| 500 | NeuAcα-(3')LN | NeuAcα-3Galβ-4GlcNAc-DH | G10203DW |
| 501 | NeuAcα-(3')LN-AO | NeuAcα-3Galβ-4GlcNAc-AO | G10203DW |
| 502 | NeuGcα-(3')LN | NeuGcα-3Galβ-4GlcNAc-DH | G48236KP |
| 503 | PI-1 | NeuAcα-3(6-NAc)Galβ-4GlcNAc-DH | G36141OA |
| 504 | PI-1-AO | NeuAcα-3(6-NAc)Galβ-4GlcNAc-AO | G36141OA |
| 505 | PI-2 | NeuAcα-3(6-NBz)Galβ-4GlcNAc-DH | N/A |
| 506 | PI-2-AO | NeuAcα-3(6-NBz)Galβ-4GlcNAc-AO | N/A |
| 507 | SA(3')-Lea-Tri | NeuAcα-3Galβ-3GlcNAc-DH<br>&#124;<br>Fucα-4 | G81971FM |
| 508 | SA(3')-Lea-Tri-AO | NeuAcα-3Galβ-3GlcNAc-AO<br>&#124;<br>Fucα-4 | G81971FM |
| 509 | SA(3')-Lex-Tri | NeuAcα-3Galβ-4GlcNAc-DH<br>&#124;<br>Fucα-3 | G90387AM |
| 510 | SA(3')-Lex-Tri-AO | NeuAcα-3Galβ-4GlcNAc-AO<br>&#124;<br>Fucα-3 | G90387AM |
| 511 | GSC-440 | NeuAcα-3Galβ-4GlcNAcβ-C30<br>&#124;<br>Fucα-3 | G00054MO |
| 512 | GSC-512 | (4-OAc)NeuAcα-3Galβ-4GlcNAcβ-C30<br>&#124;<br>Fucα-3 | G28073VX |
| 513 | GSC-513 | (9-OAc)NeuAcα-3Galβ-3GlcNAcβ-C30<br>&#124;<br>Fucα-4 | G20189ZO |
| 514 | GSC-511 | (9-OAc)NeuAcα-3Galβ-4GlcNAcβ-C30<br>&#124;<br>Fucα-3 | G83904OG |
| 515 | GSC-479 | NeuAcα-3Galβ-4GlcNAcβ-3Galβ-C30<br>&#124;<br>Fucα-3 | G32081WS |
| 516 | GSC-105 | NeuAcα-3Galβ-4GlcNAcβ-3Galβ-Cer36<br>&#124;<br>Fucα-3 | G32081WS |
| 517 | GSC-121 | NeuAcα-3Galβ-4GlcNAcβ-3Galβ-Cer36<br>&#124;<br>(3-deoxy)Fucα-3 | G99300JR |
| 518 | GSC-123 | NeuAcα-3Galβ-4GlcNAcβ-3Galβ-Cer36<br>&#124;<br>(4-deoxy)Fucα-3 | G87091PM |

| | | | |
|---|---|---|---|
| 519 | **GSC-133** | NeuAcα-3Galβ-4GlcNAcβ-3Galβ-Cer36<br>\|<br>(2-OMe)Fucα-3 | G92310MO |
| 520 | **GSC-131** | NeuAcα-3Galβ-4GlcNAcβ-3Galβ-Cer36<br>\|<br>Quvα-3 | G04643OZ |
| 521 | **GSC-163** | NeuAcα-3Galβ-4GlcNAcβ-3Galβ-Cer36<br>\|<br>Rhaα-3 | G25419XX |
| 522 | **GSC-127** | NeuAcα-3Galβ-4GlcNAcβ-3Galβ-Cer36<br>\|<br>(6-deoxy)Talα-3 | G50973AA |
| 523 | **GSC-341** | KDNα-3Galβ-4GlcNAcβ-3Galβ-C30<br>\|<br>Fucα-3 | G41295AY |
| 524 | **GSC-177** | NeuGcα-3Galβ-4GlcNAcβ-3Galβ-Cer36<br>\|<br>Fucα-3 | G16924VT |
| 525 | **GSC-175** | NeuAcα-3(4-deoxy)Galβ-4GlcNAcβ-3Galβ-Cer36<br>\|<br>Fucα-3 | G75345JW |
| 526 | **GSC-176** | NeuAcα-3(6-deoxy)Galβ-4GlcNAcβ-3Galβ-Cer36<br>\|<br>Fucα-3 | G62486QL |
| 527 | **GSC-257** | NeuAcα-3(4,6-deoxy)Galβ-4GlcNAcβ-3Galβ-Cer36<br>\|<br>Fucα-3 | G95228YK |
| 528 | **LSTa** | NeuAcα-3Galβ-3GlcNAcβ-3Galβ-4Glc-DH | G21066IE |
| 529 | **LSTd** | NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH | G28800UK |
| 530 | **GSC-272** | NeuAcα-3Galβ-3GlcNAcβ-3Galβ-4Glcβ-C30 | G79536OW |
| 531 | **GSC-147** | KDNα-3Galβ-3GlcNAcβ-3Galβ-4Glc-Cer36 | G03118WZ |
| 532 | **GSC-396** | NeuGcα-3Galβ-3GlcNAcβ-3Galβ-4Glcβ-C30 | G93871LV |
| 533 | **Sialylparagloboside** | NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer | G01767JJ |
| 534 | **GSC-273** | NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-C30 | G01767JJ |
| 535 | **GSC-31** | NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36 | G01767JJ |
| 536 | **GSC-516B** | Neuα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36<br>\|<br>SU-6 | G38781WC |
| 537 | **SA(3')-LNFP-II** | NeuAcα-3Galβ-3GlcNAcβ-3Galβ-4Glc-DH<br>\|<br>Fucα-4 | G91326YS |
| 538 | **GSC-533** | NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36<br>\|<br>Fucα-3 | G22545YE |
| 539 | **GSC-64** | NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36<br>\|<br>Fucα-3 | G96401FK |
| 540 | **GSC-472** | Neuα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36<br>\|<br>Fucα-3 | G53444FS |
| 541 | **GSC-314** | KDNα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-C30<br>\|<br>Fucα-3 | G99429GJ |
| 542 | **GSC-149** | KDNα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-C30<br>\|<br>Fucα-3 | G99429GJ |
| 543 | **GSC-311** | KDNα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-C30<br>\|<br>Rhaα-3 | G38543TA |
| 544 | **GSC-268** | SU-6<br>\|<br>NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36<br>\|<br>Fucα-3 | G23225ON |

| 545 | **GSC-268 deNAc** | SU-6<br>\|<br>Neuα-3Galβ-4GlcNβ-3Galβ-4Glcβ-Cer36<br>\|<br>Fucα-3 | G15122EO |
|---|---|---|---|
| 546 | **GSC-269** | SU-6<br>\|<br>NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36<br>\|<br>Fucα-3 | G85894LP |
| 547 | **GSC-406** | SU-6<br>\|<br>Neuα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36<br>\|<br>Fucα-3 | G88332ZA |
| 548 | **GSC-270** | SU-6  SU-6<br>\|    \|<br>NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36<br>\|<br>Fucα-3 | G78236BI |
| 549 | **NeuAcα-(3')LNnO** | NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH | G30764XF |
| 550 | **MSMFLNH** | Galβ-4GlcNAcβ-6<br>\|         \|<br>Fucα-3    Galβ-4Glc-DH<br>\|<br>NeuAcα-3Galβ-3GlcNAcβ-3 | G87485AO |
| 551 | **GSC-221** | NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36<br>\|<br>Fucα-3 | G01296LJ |
| 552 | **GSC-220** | NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36<br>\|         \|<br>Fucα-3      Fucα-3 | G45215IJ |
| 553 | **C4U** | NeuAcα-3Galβ-4GlcNAcβ-3Galβ-3GlcNAc-DH<br>\|      \|     \|<br>SU-6    SU-6  SU-6 | G46196NW |
| 554 | **A2F(2-3)** | NeuAcα-3Galβ-4GlcNAcβ-2Manα-6         Fucα-6<br>\|            \|<br>Manβ-4GlcNAcβ-4GlcNAc-DH<br>\|<br>NeuAcα-3Galβ-4GlcNAcβ-2Manα-3 | G90889KG |
| 555 | **P6-1 (GTP 4N(2,3)-4A+F)** | NeuAcα-3Galβ-4GlcNAcβ-6<br>\|<br>NeuAcα-3Galβ-4GlcNAcβ-2Manα-6         Fucα-6<br>\|            \|<br>Manβ-4GlcNAcβ-4GlcNAc-DH<br>\|<br>NeuAcα-3Galβ-4GlcNAcβ-2Manα-3<br>\|<br>NeuAcα-3Galβ-4GlcNAcβ-4 | G72131JM |
| 556 | **P7-2 (GTP 4N(2,3)-4A+1R+F)** | NeuAcα-3Galβ-4GlcNAcβ-6<br>\|<br>NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-2Manα-6         Fucα-6<br>\|            \|<br>Manβ-4GlcNAcβ-4GlcNAc-DH<br>\|<br>NeuAcα-3Galβ-4GlcNAcβ-2Manα-3<br>\|<br>NeuAcα-3Galβ-4GlcNAcβ-4 | G92594VG |
| 557 | **P8-1 (GTP 4N(2,3)-4A+2R+F)** | NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-6<br>\|<br>NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-2Manα-6         Fucα-6<br>\|            \|<br>Manβ-4GlcNAcβ-4GlcNAc-DH<br>\|<br>NeuAcα-3Galβ-4GlcNAcβ-2Manα-3<br>\|<br>NeuAcα-3Galβ-4GlcNAcβ-4 | G64894MA |

| | | | |
|---|---|---|---|
| 558 | P22-1 (GTP 3N(2,3)-3A(2,6)+F) | NeuAcα-3Galβ-4GlcNAcβ-6<br>     \|<br>NeuAcα-3Galβ-4GlcNAcβ-2Manα-6      Fucα-6<br>             \|           \|<br>           Manβ-4GlcNAcβ-4GlcNAc-DH<br>            \|<br>NeuAcα-3Galβ-4GlcNAcβ-2Manα-3 | G25902AL |
| 559 | GM4 | NeuAcα-3Galβ-Cer | G30207PZ |
| 560 | Haematoside | NeuAcα-3Galβ-4Glcβ-Cer | G91237TK |
| 561 | GM3 | NeuAcα-3Galβ-4Glcβ-Cer | G91237TK |
| 562 | GM3(Gc) | NeuGcα-3Galβ-4Glcβ-Cer | G75789AR |
| 563 | GM2 | GalNAcβ-4Galβ-4Glcβ-Cer<br>    \|<br>  NeuAcα-3 | G79389NT |
| 564 | GSC-576 | GalNAcβ-4Galβ-3Glcβ-C30<br>    \|<br>  NeuAcα-3 | G46127ND |
| 565 | GSC-108 | GalNAcβ-4Galβ-4Glcβ-Cer36<br>    \|<br>  NeuAcα-3 | G79389NT |
| 566 | GSC-193 | GalNAcβ-4Galβ-4Glcβ-Cer36<br>    \|<br>  KDNα-3 | G05517EK |
| 567 | GM1b | NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer | G03277YI |
| 568 | GSC-335 |        SU-6<br>       \|<br>NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer36 | G54000YH |
| 569 | GM1 | Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer<br>    \|<br>  NeuAcα-3 | G48558GR |
| 570 | GM1-penta | Galβ-3GalNAcβ-4Galβ-4Glc-DH<br>    \|<br>  NeuAcα-3 | G46613JI |
| 571 | GM1(Gc) | Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer<br>    \|<br>  NeuGcα-3 | G24034CH |
| 572 | GM1(Gc)-penta | Galβ-3GalNAcβ-4Galβ-4Glc-DH<br>    \|<br>  NeuGcα-3 | G28604UG |
| 573 | GM1b-DH | NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glc-DH | G14633CL |
| 574 | GD1a | NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer<br>          \|<br>       NeuAcα-3 | G46677TE |
| 575 | GD1a-hexa | NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glc-DH<br>          \|<br>       NeuAcα-3 | G69277LC |
| 576 | GalNAc-GD1a(Ac,Gc) | GalNAcβ-4Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer<br>    \|           \|<br>  NeuGcα-3      NeuAcα-3<br><br>GalNAcβ-4Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer<br>    \|           \|<br>  NeuAcα-3      NeuGcα-3 | G43817ZU<br>G70381UX |
| 577 | GSC-195 | KDNα-3Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer36<br>         \|<br>      KDNα-3 | G87111CL |
| 578 | GSC-488 | NeuAcα-3Galβ-3GalNAcβ-C30 | G02684WR |
| 579 | GSC-491 | NeuAcα-3Galβ-3(6-deoxy-6-carboxymethyl)GalNAcβ-C30 | N/A |
| 580 | GSC-489 |        SU-6<br>       \|<br>NeuAcα-3Galβ-3GalNAcβ-C30 | G32167RW |
| 581 | GSC-284 | GalNAcβ-6Galβ-4Glcβ-Cer36<br>    \|<br>  NeuAcα-3 | G83276OP |

| # | Name | Structure | ID |
|---|------|-----------|-----|
| 582 | **GSC-575** | GalNAcβ-4Galβ-3Galβ-C30<br>&#124;<br>NeuAcα-3 | G72475US |
| 583 | **GSC-154** | NeuAcα-3Galβ-4GlcNAcβ-6Galβ-4Glcβ-Cer36<br>&#124;<br>Fucα-3 | G42256GD |
| 584 | **GSC-446** | NeuAcα-3Galβ-4GlcNAcβ-6GalNAcα-3Galβ-4Glc-C30 | G33350YB |
| 585 | **GSC-441** | NeuAcα-3Galβ-4GlcNAcβ-6GalNAcα-3Galβ-4Glcβ-C30 | G75262ZI |
| 586 | **GSC-384** | NeuAcα-3Galβ-4GlcNAcβ-4GalNAcβ-3Galβ-4Glcβ-C30<br>&#124;<br>Fucα-3 | G76759MR |

### *α2-6-Sialy*

| # | Name | Structure | ID |
|---|------|-----------|-----|
| 587 | **GSC-27** | NeuAcα-6Galβ-Cer36 | G63069TR |
| 588 | **GSC-144** | KDNα-6Galβ-Cer36 | G51324IF |
| 589 | **GSC-13** | NeuAcα-(S)-6Galβ-Cer36 | N/A |
| 590 | **GSC-72** | NeuAcα-(S)-6Galβ-(S)-Cer36 | N/A |
| 591 | **GSC-60** | NeuAcα-6Glcβ-Cer36 | G59170KM |
| 592 | **GSC-9** | NeuAcα-(S)-6Glcβ-Cer36 | N/A |
| 593 | **GSC-59** | NeuAcα-6GlcNAcβ-Cer36 | G07539OP |
| 594 | **GSC-95** | NeuAcα-(S)-6GlcNAcβ-Cer36 | N/A |
| 595 | **NeuAcα-(6')Lac** | NeuAcα-6Galβ-4Glc-DH | G69893KP |
| 596 | **NeuAcα-(6')Lac-AO** | NeuAcα-6Galβ-4Glc-AO | G69893KP |
| 597 | **GSC-61** | NeuAcα-6Galβ-4Glcβ-Cer36 | G28868ZC |
| 598 | **GSC-12** | NeuAcα-(S)-6Galβ-4Glcβ-Cer36 | N/A |
| 599 | **GSC-234** | NeuAcα-(S)-6Galβ-(S)-4Glcβ-Cer36 | N/A |
| 600 | **GSC-73** | NeuAcα-(S)-6Galβ-4Glcβ-(S)-Cer36 | N/A |
| 601 | **Neuα-(6')Lac** | Neuα-6Galβ-4Glc-DH | G50804BU |
| 602 | **Neuα-(6')Lac-AO** | Neuα-6Galβ-4Glc-AO | G50804BU |
| 603 | **NeuAcβ-(6')Lac** | NeuAcβ-6Galβ-4Glc-DH | G18288MA |
| 604 | **NeuAcβ-(6')Lac-AO** | NeuAcβ-6Galβ-4Glc-AO | G18288MA |
| 605 | **NeuAcα-(6')LN** | NeuAcα-6Galβ-4GlcNAc-DH | G09376MR |
| 606 | **NeuAcα-(6')LN-AO** | NeuAcα-6Galβ-4GlcNAc-AO | G09376MR |
| 607 | **Neu5,9Ac-(6')LN** | (9-OAc)NeuAcα-6Galβ-4GlcNAc-DH | G30945GQ |
| 608 | **LSTb** | Galβ-3GlcNAcβ-3Galβ-4Glc-DH<br>&#124;<br>NeuAcα-6 | G19017MP |
| 609 | **GSC-397** | NeuGcα-6Galβ-3GlcNAcβ-3Galβ-4Glcβ-C30 | G96427WI |
| 610 | **LSTc** | NeuAcα-6Galβ-4GlcNAcβ-3Galβ-4Glc-DH | G72506RN |
| 611 | **SA(6')-LNFP-VI** | NeuAcα-6Galβ-4GlcNAcβ-3Galβ-4Glc-DH<br>&#124;<br>Fucα-3 | G95310NT |
| 612 | **GSC-97** | NeuAcα-6Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36<br>&#124;<br>Fucα-3 | G09046VD |
| 613 | **MSDFLNnH-AO** | Fucα-2Galβ-4GlcNAcβ-6<br>&#124;   &#124;<br>Fucα-3   Galβ-4Glc-AO<br>&#124;<br>NeuAcα-6Galβ-4GlcNAcβ-3 | G92965OP |
| 614 | **NeuAcα-(6')LNnO (F1)** | NeuAcα-6Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH  (position of NeuAc not fully defined) | N/A |
| 615 | **MSLNH** | NeuAcα-6Galβ-4GlcNAcβ-6<br>&#124;<br>Galβ-4Glc-DH<br>&#124;<br>Galβ-3GlcNAcβ-3 | G08723AU |

| 616 | MSLNnH-I | Galβ-4GlcNAcβ-6<br>&#124;<br>Galβ-4Glc-DH<br>&#124;<br>NeuAcα-6Galβ-3GlcNAcβ-3 | G58274PS |
|---|---|---|---|
| 617 | DSLNnH | NeuAcα-6Galβ-4GlcNAcβ-6<br>&#124;<br>Galβ-4Glc-DH<br>&#124;<br>NeuAcα-6Galβ-4GlcNAcβ-3 | G59924XR |
| 618 | MFMSLNnH | Galβ-4GlcNAcβ-6<br>&#124;  &#124;<br>Fucα-3 Galβ-4Glc-DH<br>&#124;<br>NeuAcα-6Galβ-3GlcNAcβ-3 | G84742FE |
| 619 | A2(2-6) | NeuAcα-6Galβ-4GlcNAcβ-2Manα-6<br>&#124;<br>Manβ-4GlcNAcβ-4GlcNAc-DH<br>&#124;<br>NeuAcα-6Galβ-4GlcNAcβ-2Manα-3 | G73866ZM |
| 620 | AGP-Bi-Ac2 | NeuAcα-6Galβ-4GlcNAcβ-2Manα-6<br>&#124;<br>Manβ-4GlcNAcβ-4GlcNAc-DH<br>&#124;<br>NeuAcα-6Galβ-4GlcNAcβ-2Manα-3 | G73866ZM |
| 621 | AGP-Bi-Gc2 | NeuGcα-6Galβ-4GlcNAcβ-2Manα-6<br>&#124;<br>Manβ-4GlcNAcβ-4GlcNAc-DH<br>&#124;<br>NeuGcα-6Galβ-4GlcNAcβ-2Manα-3 | G98223HN |
| 622 | AGP-Bi-AcGc | NeuGc(Ac)α-6Galβ-4GlcNAcβ-2Manα-6<br>&#124;<br>Manβ-4GlcNAcβ-4GlcNAc-DH<br>&#124;<br>NeuAc(Gc)α-6Galβ-4GlcNAcβ-2Manα-3 | G95112IM |
| 623 | GSC-442 | GalNAcβ-4Galβ-4Glcβ-Cer36<br>&#124;<br>NeuAcα-6 | G29907JM |
| 624 | GSC-68 | NeuAcα-6Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer36 | G48236RC |
| 625 | GSC-155 | Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer36<br>&#124;<br>NeuAcα-6 | G54721WC |
| 626 | GSC-107 | NeuAcα-6Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer36<br>&#124;<br>NeuAcα-6 | G30725CV |
| 627 | BSM-Di-A1-AO | NeuGcα-6GalNAc-AO | G75206XT |
| 628 | BSM-Di-A2-AO | NeuAcα-6GalNAc-AO | G08371CD |
| 629 | GSC-70 | NeuAcα-6Galβ-6GalNAcβ-4Galβ-4Glcβ-Cer36 | G71659FD |
| α2-3- and α2-6-Sialy | | | |
| 630 | DSLNT | NeuAcα-3Galβ-3GlcNAcβ-3Galβ-4Glc-DH<br>&#124;<br>NeuAcα-6 | G38710SX |
| 631 | A3 | NeuAcα-3Galβ-4GlcNAcβ-2Manα-6<br>&#124;<br>Manβ-4GlcNAcβ-4GlcNAc-DH<br>&#124;<br>NeuAcα-3Galβ-4GlcNAcβ-4Manα-3<br>&#124;<br>NeuAcα-6Galβ-4GlcNAcβ-2 | G28330DZ |
| 632 | GSC-118 | NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer36<br>&#124;<br>NeuAcα-6 | G45714BQ |
| 633 | DST | NeuAcα-3Galβ-3GalNAc-DH<br>&#124;<br>NeuAcα-6 | G65191ST |
| 634 | DST-AO | NeuAcα-3Galβ-3GalNAc-AO<br>&#124;<br>NeuAcα-6 | G65191ST |

| | | | |
|---|---|---|---|
| 635 | **GSC-490** | NeuAcα-3Galβ-3GalNAcβ-C30<br>&#124;<br>NeuAcα-6 | G04267XF |

### α2-8-Sialy

| | | | |
|---|---|---|---|
| 636 | **GSC-230** | NeuAcα-8NeuAcα-3Galβ-Cer36 | G43195TR |
| 637 | **GSC-231** | NeuAcα-8NeuAcα-6Galβ-Cer36 | G27002ZI |
| 638 | **GSC-439** | NeuAcα-8NeuAcα-8NeuAcα-6Galβ-Cer36 | G84520VA |
| 639 | **GSC-232** | NeuAcα-8NeuAcα-6Glcβ-Cer36 | G95904OG |
| 640 | **GSC-229** | NeuAcα-8NeuAcα-3Galβ-4Glcβ-Cer36 | G98544DH |
| 641 | **GSC-437** | NeuAcα-8NeuAcα-8NeuAcα-3Galβ-4Glcβ-Cer36 | G93899SO |
| 642 | **GD3** | NeuAcα-8NeuAcα-3Galβ-4Glcβ-Cer | G98544DH |
| 643 | **GD3-tetra** | NeuAcα-8NeuAcα-3Galβ-4Glc-DH | G81928IQ |
| 644 | **GD3-tetra-AO** | NeuAcα-8NeuAcα-3Galβ-4Glc-AO | G81928IQ |
| 645 | **GD2** | GalNAcβ-4Galβ-4Glcβ-Cer<br>&#124;<br>NeuAcα-8NeuAcα-3 | G02657AK |
| 646 | **GD1b** | Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer<br>&#124;<br>NeuAcα-8NeuAcα-3 | G37184KW |
| 647 | **GD1b-DH** | Galβ-3GalNAcβ-4Galβ-4Glc-DH<br>&#124;<br>NeuAcα-8NeuAcα-3 | G97612UN |
| 648 | **GQ1b** | NeuAcα-8NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer<br>&#124;<br>NeuAcα-8NeuAcα-3 | G18625KA |
| 649 | **SA2(α8)** | NeuAcα-8NeuAc-DH | G96368TK |
| 650 | **SA3(α8)** | NeuAcα-8NeuAcα-8NeuAc-DH | G09290IZ |
| 651 | **SA4(α8)** | NeuAcα-8NeuAcα-8NeuAcα-8NeuAc-DH | G55973YK |
| 652 | **SA5(α8)*** | NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAc-DH | G83157EH |
| 653 | **SA6(α8)*** | NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAc-DH | G12986PF |
| 654 | **SA7(α8)*** | NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAc-DH | G46657CU |
| 655 | **SA8(α8)*** | NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAc-8NeuAcα-8NeuAc-DH | G14163TR |
| 656 | **SA9(α8)*** | NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAc-8NeuAcα-8NeuAc-8NeuAcα-DH | G32997PD |
| 657 | **SA10(α8)*** | NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAc-DH | G38397KQ |

### α2-3- and α2-8-Sialy

| | | | |
|---|---|---|---|
| 658 | **GT1a** | NeuAcα-8NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer<br>&#124;<br>NeuAcα-3 | G68110IF |
| 659 | **GT1b** | NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer<br>&#124;<br>NeuAcα-8NeuAcα-3 | G40183QN |
| 660 | **GT1c-DH** | Galβ-3GalNAcβ-4Galβ-4Glc-DH<br>&#124;<br>NeuAcα-8NeuAcα-8NeuAcα-3 | G50762GU |

### Other sialy linkages

| | | | |
|---|---|---|---|
| 661 | **SA(3/6)LNFP-I** | NeuAcα-3/6Galβ-3GlcNAcβ-3Galβ-4Glc-DH<br>&#124;<br>Fucα-2 | G54312KX |
| 662 | **GSC-96** | NeuAcα-9NeuAcα-3Galβ-4Glcβ-Cer36 | G80544IB |
| 663 | **SA2(α2-9)** | NeuAcα-9NeuAc-DH | G56026PA |
| 664 | **SA3(α2-9)** | NeuAcα-9NeuAcα-9NeuAc-DH | G82952PX |
| 665 | **SA4(α2-9)** | NeuAcα-9NeuAcα-9NeuAcα-9NeuAc-DH | G73663JA |
| 666 | **SA5(α2-9)** | NeuAcα-9NeuAcα-9NeuAcα-9NeuAcα-9NeuAc-DH | G90126NY |
| 667 | **SA6(α2-9)** | NeuAcα-9NeuAcα-9NeuAcα-9NeuAcα-9NeuAcα-9NeuAc-DH | G56108YR |

| | | | |
|---|---|---|---|
| 668 | **NeuAc** | NeuAc-DH | G95058CG |
| 669 | **NeuAc-AO** | NeuAc-AO | G95058CG |
| 670 | **NeuGc** | NeuGc-DH | G43724HA |
| 671 | **NeuGc-AO** | NeuGc-AO | G43724HA |
| 672 | **GSC-62** | NeuAcα–2Glcβ–Cer36 | G74222PO |

[a], Probe position in the histogram charts (Supplementary Figure 2).

[b], The oligosaccharide probes are all lipid-linked. Asterisks that follow the names of certain probes indicate that predominant components are shown.

[c], Unless otherwise indicated they are neoglycolipids (NGLs) prepared from reducing oligosaccharides by reductive amination with the amino lipid, 1,2-dihexadecyl-*sn*-glycero-3-phosphoethanolamine (DHPE). AO, NGLs prepared from reducing oligosaccharides by oxime ligation with an aminooxy (AO) functionalized DHPE (Liu et al., Chem. Biol. 14, 847–859, 2007); AD, fluorescent NGLs prepared from reducing oligosaccharides by reductive amination with a fluorescence labelled lipid, the anthracene-modified DHPE (ADHP) (Stoll et al., Eur. J. Biochem. 267(6), 1795-804, 2000); Cer, natural glycolipids with various ceramide moieties; Cer36, synthetic glycolipids with ceramide having a total of 36 carbon atoms; C30, a synthetic lipid [2-(tetradecyl)hexadecanol] with 30 carbon atoms. OX and OY designate, respectively, the C1-4 fragment and the C5-6 fragments of GalNAcol of reduced oligosaccharides after mild periodate oxidation followed by reductive amination with DHPE (Chai et al., Methods Enzymol. 362, 160-195, 2003). ΔUA, 4,5-unsaturated hexuronic acid; aMan, 2,5-anhydro-mannose; aGal, 3,6-anhydro-galactose.

[d], ID in the international glycan structure repository GlyTouCan https://glytoucan.org/.
N/A, structure that will be registered in the GlyTouCan repository in due course.