# Structural and Functional Analysis of Polyomavirus Attachment to Carbohydrate Receptors

**Dissertation**

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Diplom-Biochemikerin Ursula Neu
aus Oberndorf am Neckar

Tübingen
2010

| | |
|---|---|
| Tag der mündlichen Qualifikation: | 21.12.2010 |
| Dekan: | Prof. Dr. Wolfgang Rosenstiel |
| 1. Berichterstatter: | Prof. Dr. Thilo Stehle |
| 2. Berichterstatter: | Prof. Dr. Andrej Lupas |

Why do you make such a fuss about solving structures?
In the end, they all look like potatoes!

# Abstract

Viruses must attach to specific receptors on their host cells in order to initiate entry. Viral spread, tissue and host tropism are determined by precisely regulated contacts and affinities between viral proteins and their cognate ligands at the cell surface.

The present thesis focuses on the attachment proteins and receptor interactions of several polyomaviruses, a group of small, non-enveloped double-stranded DNA viruses that infect a range of mammalian and avian hosts. Simian virus 40 (SV40) has been a paradigm for understanding attachment and entry of non-enveloped viruses, viral DNA replication, and virus assembly, as well as for endocytosis pathways associated with cholesterol. It shares high homology with the widespread human polyomaviruses JC and BK (JCV and BKV, respectively), which establish asymptomatic infections in early childhood, but cause disease in immunocompromised individuals. JCV causes the fatal demyelinating disease Progressive Multifocal Leukoencephalopathy through lytic infection of oligodendrocytes, and BKV is a major cause of kidney allograft loss because it gives rise to Polyomavirus-associated Nephropathy and Hemorrhagic Cystitis in transplant recipients. The emerging human polyomaviruses WU and KI (WUV and KIV, respectively) are not linked to disease yet.

Polyomavirus capsids consist of 72 pentamers of the major capsid protein VP1, which mediates interactions with cell-surface receptors. These are sialylated oligosaccharides for SV40, JCV and BKV. In this thesis, VP1 pentamers were produced that could not assemble into capsids because of truncations in their C-termini. Glycan array screening of such pentamers revealed that SV40 VP1 recognizes the ganglioside GM1, which carries a branched $\alpha$2,3-linked sialic acid with narrow specificity. The $\alpha$2,6-sialylated pentasaccharide LSTc was identified as a specific receptor motif for JCV VP1. BKV VP1 was shown to attach to b-series gangliosides, which share a common $\alpha$2,8-linked disialic acid motif, by glycan array screening, cell-based infection assays and saturation transfer difference (STD) NMR.

The crystal structures of SV40 VP1 in complex with the GM1 oligosaccharide, of JCV VP1 in complex with LSTc, and of BKV VP1 in complex with the oligosaccharide portion of the b-series ganglioside GD3 were determined at 2.25 Å, 2.0 Å and 1.7 Å, respectively. They indicate that carbohydrates are bound in shallow, solvent-exposed grooves at the outer surface of the capsid that are formed by extensive loops of VP1. This is consistent with the relatively low affinity in the millimolar range of one binding site of SV40 VP1 for GM1, which was determined by isothermal titration calorimetry. The three viruses share a common binding site that interacts with the major contact, terminal sialic acid, through a complex

network of interactions. The structures of the binding sites suggest that BKV and JCV prefer the human-type sialic acid N-acetyl neuraminic acid (NeuNAc), while SV40 prefers the monkey-type N-glycolyl neuraminic acid (NeuNGc). Specificity for different carbohydrate sequences underlying the sialic acid is determined by a small number of carefully positioned satellite residues. These form relatively few contacts with additional carbohydrate moieties, but enable engagement of the correct ligand while blocking others.

Comparison of the SV40 and BKV VP1 structures suggested that their specificities for sialic acid in different linkages were determined by the difference in one amino acid. When the corresponding point mutation K68S was introduced into BKV VP1, its specificity was switched from b-series gangliosides to GM1, as demonstrated by STD NMR and glycan array screening. Moreover, the mutant recognized only GM1 carrying NeuNAc and not NeuNGc, hinting at a role for different sialic acid types in host tropism.

Crystal structures of WUV and KIV VP1 were solved at 2.9 and 2.55 Å, respectively. They are distant in evolution from SV40, BKV and JCV VP1 as well as from Murine Polyomavirus (Polyoma) VP1. Comparison of the KIV and WUV VP1 structures with SV40 and Polyoma VP1 revealed a conserved VP1 core and a surprisingly different loop arrangement. However, all four structures feature a stretch of conserved main chain structure in one surface loop, which is at the center of the overlapping SV40 and Polyoma sialic acid binding sites.

# Zusammenfassung

Der erste Schritt bei der Infektion einer Zelle durch ein Virus ist das Andocken an einen spezifischen Rezeptor. Die präzise regulierten Kontakte und Bindungsaffinitäten zwischen Virusproteinen und ihren Liganden auf der Zelloberfläche bestimmen die Ausbreitung des Virus, seine Gewebe- und Wirtsspezifität.

Diese Arbeit untersucht die Anheftungsproteine verschiedener Polyomaviren sowie deren Wechselwirkungen mit Rezeptoren. Polyomaviren sind eine Gruppe kleiner, unbehüllter Viren mit doppelsträngigem DNA-Genom, die eine Reihe von Säugetieren und Vögeln befallen. Simian Virus 40 (SV40) ist ein Modell zur Erforschung der Virusreplikation, der Andock- und Aufnahmewege von unbehüllten Viren sowie von cholesterinabhängigen Endozytosemechanismen. SV40 ähnelt sehr den weit verbreiteten menschlichen Polyomaviren JC und BK (JCV und BKV), die Menschen in früher Kindheit persistent asymptomatisch infizieren, aber bei geschwächtem Immunsystem Erkrankungen auslösen. JCV verursacht durch lytische Infektion von Oligodendrocyten die tödliche demyelinisierende Krankheit Progressive Multifokale Leukenzephalopathie (PML). BKV ist eine der Hauptursachen für das Scheitern von Nierentransplantationen, da es zu Polyomavirus-assoziierter Nephropathie und hämorrhagischer Cystitis führt. Es ist noch nicht bekannt, ob die neu entdeckten menschlichen Polyomaviren WU und KI (WUV und KIV) Krankheiten verursachen.

Polyomavirus-Capside bestehen aus 72 Pentameren des Proteins VP1, welches an Rezeptoren auf der Zelloberfläche bindet. Die Rezeptoren für SV40, BKV und JCV sind Oligosaccharide, die Sialinsäure enthalten. In dieser Arbeit wurden rekombinante VP1 Pentamere mit verkürzten C-Termini produziert, die sich deswegen nicht mehr zu Capsiden assemblieren können. Es wurde mittels Glycanarray-Screening von solchen Pentameren gezeigt, dass SV40 VP1 das Gangliosid GM1 mit hoher Spezifität erkennt, das eine terminale, $\alpha$2,3-verbundene Sialinsäure enthält. Das Pentasaccharid LSTc, das eine terminale, $\alpha$2,6-verbundene Sialinsäure enthält, wurde als spezifisches Rezeptormotiv für JCV VP1 identifiziert. Weiterhin wurde mittels Glycanarray-Screening, zellbasierten Infektionsassays und STD NMR gezeigt, dass BKV an Ganglioside der b-Serie bindet, die alle ein $\alpha$2,8-verbundenes, terminales Disialinsäure-Motiv gemein haben.

Die Kristallstrukturen der Komplexe aus SV40 VP1 und dem Oligosaccharid-Anteil von GM1, JCV VP1 und LSTc sowie BKV VP1 und dem Kohlenhydrat-Anteil des b-Serien-Gangliosids GD3 wurden mit 2,25 Å, 2,0 Å und 1,7 Å Auflösung bestimmt. Sie zeigen, dass

die Kohlenhydrat-Rezeptoren in flachen, exponierten Mulden auf der Außenseite des Capsids gebunden werden. Diese werden in allen drei Komplexen von langen Loops in VP1 gebildet. Dieser Bindungsmodus ist konsistent mit einer relativ geringen, millimolaren Affinität einer Bindungsstelle auf SV40 VP1 für GM1, die mittels Isothermer Titrationskalorimetrie gemessen wurde. Die drei Viren nutzen dieselbe konservierte Bindungsstelle für ihren Hauptkontakt, terminale Sialinsäure, die sie mittels eines komplexen Netzwerks von Interaktionen erkennen. Die Strukturen der Bindungsstelle lassen eine Präferenz von BKV und JCV für den Sialinsäure-Typ vermuten, der in Menschen am häufigsten vorkommt, nämlich N-Acetyl-Neuraminsäure (NeuNAc). SV40 dagegen scheint die in Affen häufigere N-Glycolyl-Neuraminsäure (NeuNGc) zu bevorzugen. Die Unterscheidung von Sialinsäuren in unterschiedlichen glycosidischen Bindungen bzw. verschiedener mit der Sialinsäure verknüpfter Kohlenhydrate wird durch einige wenige geschickt gelegene Aminosäuren außerhalb der Bindungsstelle für terminale Sialinsäure bewerkstelligt. Obwohl diese nur wenige Kontakte mit dem Oligosaccharid ausbilden, ermöglichen sie die Bindung des korrekten Liganden und blockieren Wechselwirkungen mit anderen Kohlenhydraten.

Es wurde gezeigt, dass die unterschiedlichen Spezifitäten von BKV und SV40 für Sialinsäure in verschiedenen Kontexten durch eine einzelne Aminosäure bestimmt wird. Die entsprechende Punktmutation K68S wurde in BKV VP1 erzeugt, was die Spezifität von BKV von Gangliosiden der b-Serie zu GM1 verschob, was mittels STD NMR und Glycanarray-Screening gezeigt wurde. Weiterhin erkennt die Mutante nur GM1 mit NeuNAc, nicht aber mit NeuNGc. Dies deutet darauf hin, dass die Sialinsäuretypen für die Wirtsspezifität von Polyomaviren von Bedeutung sein könnten.

Die Kristallstrukturen von WUV und KIV VP1 wurdem mit Auflösungen von 2,9 Å und 2,55 Å gelöst. WUV und KIV sind evolutionär weit entfernt von SV40 und dem murinen Polyomavirus (Polyoma). Ein Strukturvergleich von WUV und KIV mit SV40 und Polyoma zeigte einen konservierten Kern von VP1 und überraschend diverse Loopstrukturen. Alle vier Strukturen enthielten allerdings auch eine konservierte Konformation der Hauptkette in einem Loop auf der Proteinoberfläche, die sich im Zentrum der überlappenden Sialinsäurebindungsstellen von SV40 und Polyoma befindet.

# Notes

The following parts of this thesis have been published before submission or will be published before publication of this thesis in the articles indicated. They have been modified for inclusion in this thesis.

Neu U., Woellner K., Gauglitz G., and Stehle T. (2008) Structural basis of ganglioside GM1 receptor recognition of Simian Virus 40. Proc. Natl. Acad. Sci. USA; 105(13):5219-5224.
Chapters 3.2, 4.1.1, 4.2.1

Neu U., Stehle T., and Atwood W.J. (2009) The Polyomaviridae: Contributions of virus structure to our understanding of virus receptors and infectious entry. Virology; 384(2):389-399.
Chapters 1.3.1-1.3.3, 4.4

Neu U.*, Maginnis M.S.*, Palma A.S., Ströh, L., Nelson, C.D.S., Feizi T., Atwood W.J., and Stehle T. (2010) Structure-function analysis of the human JC polyomavirus establishes the LSTc pentasaccharide as functional receptor. Cell Host & Microbe; 8(4):309-319.
Chapters 3.3, 4.2.1, 4.3.1, 4.3.2.2

Neu U., Wang J., Macejak D., Garcea R.L., and Stehle T. (2011) Structures of the major capsid proteins of the human Karolinska Institutet and Washington University polyomaviruses. J. Virol.; 85(14):7384-7392
Chapters 3.7, 4.2.4., 4.2.5

Neu U.*, Hengel H.*, Blaum B.S., Schowalter R.M., Macejak D., Gilbert M., Wakarchuk, W.W., Imamura A., Ando H., Kiso M., Arnberg N., Garcea R.L., Peters T., Buck C.B., and Stehle T. (2012) Structures of Merkel cell polyomavirus VP1 complexes define a sialic acid binding site required for infection. PLoS Path.; 8(7):e1002738
Chapter 4.2.2

Neu U.*, Allen S.A.*, Blaum B.S., Liu Y., Frank M., Palma A.S., Feizi T., Peters T., Atwood W.J., and Stehle T. (2013) A structure-guided mutation in the major capsid protein retargets BK Polyomavirus. Manuscript submitted.
Chapters 3.4, 3.5, 3.6.1, 4.1.3, 4.3.1, 4.3.2.2

All of the experiments described in this thesis were performed by me, with exceptions clearly indicated in the text. In all cases of experiments that were performed by someone else, they were suggested and/or designed by me and/or performed using reagents I supplied. In addition, I was involved in troubleshooting experiments and discussing results in each case.

# Contents

# Abbreviations

| | |
|---|---|
| $A_{280}$ | Absorbance at a wavelength of 280 nm |
| Ad37 | Adenovirus serotype 37 fiber knob |
| AIDS | Acquired Immunodeficiency Syndrome |
| ASU | Asymmetric unit |
| BKV | BK Polyomavirus |
| CFG | Consortium for Functional Glycomics |
| CNS | Central nervous system |
| COSY | Correlation spectroscopy |
| DNA | Desoxyribonucleic acid |
| *E. coli* | *Escherichia coli* |
| ER | Endoplasmic reticulum |
| ERAD | ER-associated degradation |
| ESRF | European Synchrotron Radiation Facility, Grenoble, F |
| Gal | Galactose |
| GalNAc | *N*-acetyl-galactosamine |
| Glc | Glucose |
| GlcNAc | *N*-acetyl-glucosamine |
| GlcNAc-LSTc | LSTc carrying GlcNAc instead of Glc |
| HA | Influenza A virus haemagglutinin |
| HIV | Human Immunodeficiency Virus |
| hPyV6 | Human Polyomavirus 6 |
| hPyV7 | Human Polyomavirus 7 |
| HSQC | Heteronuclear single-quantum coherence |
| Influenza | Influenza A Virus |
| ITC | Isothermal titration calorimetry |
| JCV | JC Polyomavirus |
| KIV | Karolinska Institutet Polyomavirus |
| large-T | Large T-antigen |
| small-t | Small T-antigen |
| LPV | B-lymphotropic Papovavirus (African Green Monkey Polyomavirus) |
| MCC | Merkel Cell Carcinoma |
| MCV | Merkel Cell Polyomavirus |
| MWCO | Molecular weight cut-off |

| | |
|---|---|
| NCS | Non-crystallographic symmetry |
| NeuNAc | 5-*N*-acetyl-neuraminic acid |
| NeuNAc-GM1 | GM1 carrying NeuNAc |
| NeuNGc | 5-*N*-glycolyl-neuraminic acid |
| NeuNGc-GM1 | GM1 carrying NeuNGc |
| NMR | Nuclear magnetic resonance |
| $OD_{600}$ | Optical density at a wavelength of 600 nm |
| PCR | Polymerase chain reaction |
| PML | Progressive Multifocal Leukoencephalopathy |
| PVN | Polyomavirus-induced Nephropathy |
| Polyoma | Murine Polyomavirus |
| r.m.s.d. | Root mean square deviation |
| SD | Standard deviation |
| SLS | Swiss Light Source, Villigen, CH |
| STD | Saturation transfer difference |
| SV40 | Simian Virus 40 |
| TOCSY | Total correlation spectroscopy |
| TSV | Trichodysplasia Spinulosa-associated Polyomavirus |
| VLP | Virus-like particle |
| VP1 | Polyomavirus viral protein 1 |
| VP2 | Polyomavirus viral protein 2 |
| VP3 | Polyomavirus viral protein 3 |
| VP4 | Polyomavirus viral protein 4 |
| WUV | Washington University Polyomavirus |

Ganglioside sequences were abbreviated using the Svennerholm nomenclature (Svennerholm, 1963).

# 1. Introduction

Viruses must attach to specific receptors on their host cells in order to initiate entry, but exceedingly strong receptor interactions prevent viral progeny from spreading to new host cells. As a result, attachment and release processes depend on precisely regulated contacts and affinities between viral proteins and their cognate ligands at the cell surface. These interactions determine viral spread and tissue tropism, viral pathogenesis and host tropism. The present thesis focuses on the attachment proteins and receptor interactions of several polyomaviruses, namely the human viruses BK Polyomavirus (BKV), JC Polyomavirus (JCV), WU Polyomavirus (WUV) and KI Polyomavirus (KIV) as well as Simian Virus 40 (SV40), whose natural host is the macaque.

## 1.1. Polyomaviruses

Polyomaviruses make up the *Polyomaviridae* family of viruses, for which currently 23 members are known. They infect a variety of mammalian and avian hosts, such as humans, simians, ungulates and rodents, but also geese, crows and smaller birds. The best-studied polyomaviruses are SV40 and Murine Polyomavirus (Polyoma). They were discovered in 1953 and 1960, respectively (Gross, 1953; Sweet and Hilleman, 1960), and caught interest because they can transform cells in culture and cause tumors in newborn animals, hence the name Polyomaviruses (Polyoma = many tumors) (Eddy et al., 1958; Girardi et al., 1962; Stewart et al., 1958). BKV and JCV were isolated from immunocompromised patients in 1971 and are named after the initials of the patients from whom they were isolated (Gardner et al., 1971; Padgett et al., 1971). Since 2008, several new human polyomaviruses have been discovered, such as KIV (Allander et al., 2007), WUV (Gaynor et al., 2007), Merkel Cell Polyomavirus (MCV) (Feng et al., 2008), Human Polyomavirus 6 (HPyV6) and Human Polyomavirus 7 (HPyV7) (Schowalter et al., 2010), and Trichodysplasia Spinulosa-associated Polyomavirus (TSV) (van der Meijden et al.).

All polyomaviruses are non-enveloped. They feature a small, circular, double-stranded DNA genome of about 5,200 bp that is present as a minichromosome in virions (Fig. 1A). It encodes the major capsid protein VP1, which attaches to receptors and promotes viral entry, the minor capsid proteins VP2 and VP3 as well as the non-structural proteins large T-antigen (large-T) and small t-antigen (small-t), which promote viral replication and are responsible for the transforming properties of polyomaviruses. Transformation of mammalian cells with large-T is often used to immortalize cell lines. Some members of the family feature additional

non-structural proteins, such as a middle T-antigen, an agnoprotein or a VP4 protein (Cole and Conzen, 2001; Daniels et al., 2007).

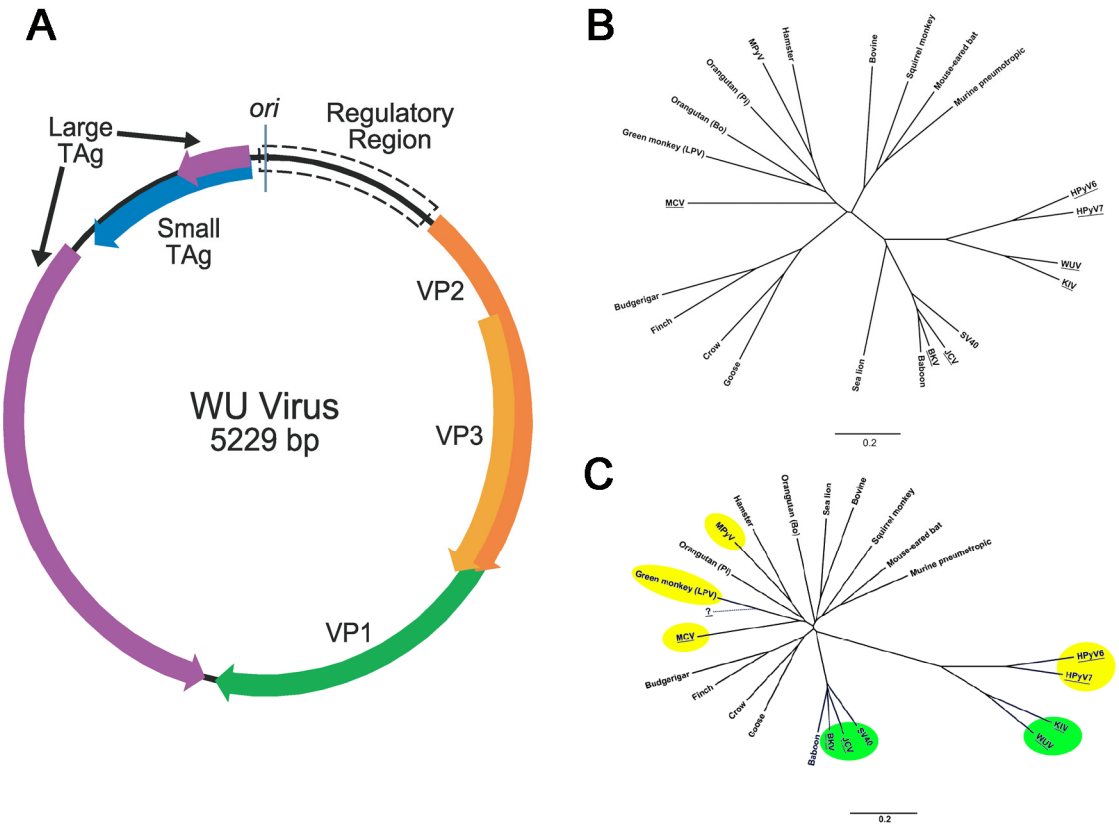



**Fig. 1: Genome organisation and homology of polyomaviruses.**
(A) Genome organisation of WUV. The genomes of all polyomaviruses are very similar. Picture modified from Gaynor et al., PLoS Pathogens, 2010.
(B) and (C) Homology among entire polyomavirus genomes (B) and VP1 open reading frames as determined from nucleotide sequences. Human viruses are underlined. VP1 proteins whose structures were determined in this thesis are shaded green. VP1 proteins whose sequences or structures are used in this thesis are shaded yellow. Picture modified from Schowalter et al., PLoS Pathogens, 2010.

All polyomaviruses are quite homologous to one another (Cole and Conzen, 2001). SV40, BKV and JCV share very high homology, with sequence identity between the SV40 and BKV VP1 proteins as high as 84% (Fig. 1B,C). WUV and KIV are also similar to one another and relatively similar to HPyV6 and HPyV7, but are evolutionarily distant from other polyomaviruses. Polyoma is related to MCV and B-Lymphotropic Papovavirus (LPV), but this group is more diverse than the others (Schowalter et al., 2010).

## 1.2. Epidemiology and pathogenesis

### 1.2.1. SV40

SV40 was isolated from the rhesus monkey cell cultures used to produce the first batches of poliovirus vaccine that had been given to millions of children in the 1950s (Sweet

and Hilleman, 1960). Once its transforming properties and its ability to cause tumors in newborn hamsters were known (Girardi et al., 1962), it was a major concern whether this virus was also tumorigenic in humans. However, no clear link between SV40 and human cancer has been established (Garcea and Imperiale, 2003), even though SV40 DNA has been detected in samples of rare human cancers that correspond to the ones it causes in hamsters (Bergsagel et al., 1992). In addition, evidence for SV40 seroconversion in humans is inconclusive, indicating that it might not be able to persist in humans (Kean et al., 2009). The natural hosts of SV40 are macaques, where the virus establishes a persistent infection and does not cause tumors. In immunocompromised simians, however, it can cause syndromes similar to BKV and JCV pathogenesis in humans (Horvath et al., 1992) (1.2.2, 1.2.3).

SV40 has been used as a paradigm for DNA replication and transcription, viral replication and gene expression (Cole and Conzen, 2001), and for endocytosis pathways associated with lipid rafts and cholesterol (Mercer et al., 2010). It has also been explored as a vector for gene therapy (Buscail and Cordelier, 2007; Strayer et al., 2005).

### 1.2.2. JCV

JCV is a significant human pathogen for which approximately 50-80% of individuals are seropositive (Kean et al., 2009; Knowles et al., 2003). JCV establishes a persistent, mostly asymptomatic infection in the kidney (Dorries, 1998). However, the virus can become reactivated in immunosuppressed hosts, leading to enhanced viral replication and infection of glial cells, including astrocytes and the myelin-producing oligodendrocytes, in the central nervous system (CNS). JCV infection leads to cytolytic destruction of oligodendroglia and causes the fatal disease Progressive Multifocal Leukoencephalopathy (PML) (Khalili and White, 2006; Seth et al., 2003; Silverman and Rubinstein, 1965; Zurhein and Chou, 1965). PML is commonly associated with Human Immunodeficiency Virus (HIV) and Acquired Immunodeficiency Syndrome (AIDS) (Cinque et al., 2009). However, the incidence of PML has increased recently, arising in patients receiving immunosuppressive therapies for autoimmune diseases such as Multiple Sclerosis, Crohn's disease, rheumatoid arthritis and systemic lupus erythematosus (Carson et al., 2009a; Carson et al., 2009b; Major, 2009; Rahmlow et al., 2008; Van Assche et al., 2005). The prognosis of PML is bleak, as it usually proves fatal within 6-12 months of the onset of symptoms. Treatment options for PML are currently inadequate (Kishida, 2007).

### 1.2.3. BKV

More than 70% of the adult human population is seropositive for BKV, which establishes a persistent, non-replicative, asymptomatic infection in the genitourinary tract

(Hirsch and Steiger, 2003; Nickeleit et al., 1999; Shinohara et al., 1993). A key modulator of BKV reactivation is immunosuppression of the host, which leads to an increase in viral replication (Hirsch and Steiger, 2003). Complications of BKV reactivation include the development of Polyomavirus-induced Nephropathy (PVN) in kidney transplant recipients, and hemorrhagic cystitis in bone marrow transplant recipients (Bedi et al., 1995; Hirsch, 2005; Hirsch and Steiger, 2003; Xie et al., 2009). If not diagnosed, PVN has a high mortality rate as it raises the potential for irreversible allograft failure due to increased fibrosis and necrosis of the renal tissue caused by lytic viral replication and the host immune response (Nickeleit et al., 1999). Individuals at risk of developing BKV-related diseases include those undergoing immunosuppressive therapy as a result of organ transplantation or chemotherapy, and individuals who have compromised immune systems due to infection with HIV, or autoimmune diseases (Hirsch and Steiger, 2003; Wiseman, 2009). Currently, treatment of PVN and hemorrhagic cystitis includes reduction of immunosuppression for PVN and hyperhydration, irrigation of the bladder, or surgery for hemorrhagic cystitis (Hirsch and Steiger, 2003; Wiseman, 2009).

### 1.2.4. KIV and WUV

KIV and WUV were sequenced from respiratory tract samples (Allander et al., 2007; Gaynor et al., 2007), implying that they infect the respiratory system and maybe also persist there. The viruses have been confirmed in respiratory specimens worldwide, suggesting that they are widespread in the human populations (Jiang et al., 2009a). 55 % of (Western) adults are seropositive for KIV and 69 % for WUV, with available data suggesting a primary infection during childhood (Kean et al., 2009). However, it is not clear to date whether there is a link between KIV or WUV infection and respiratory disease (Jiang et al., 2009a).

### 1.2.5. Polyoma

Polyoma is not a human pathogen, but an established animal model to study polyomavirus spread, pathogenesis and tumorigenesis. When newborn mice are inoculated with Polyoma, they develop multiple tumors (Eddy et al., 1958). In adult mice, the virus does not have this effect, hinting at a role of a developed immune system in the control of the infection. Different strains of Polyoma exhibit different levels of virulence. Small plaque strains such as the field isolate strain RA form small plaques in cell culture and are only moderately pathogenic in mice. Large plaque strains such as the PTA strain form large plaques in cell culture, spread quickly through the host and cause tumors in multiple sites (Dubensky et al., 1991; Freund et al., 1991a). Interestingly, the different behaviour of the RA and PTA strains has been mapped to a single point mutation in the receptor binding site of VP1 (Freund et al., 1991b)(4.3.2.1).

### 1.2.6. MCV

MCV was isolated from specimens of Merkel Cell Carcinoma (MCC), a rare, but aggressive cancer of the skin (Feng et al., 2008). Merkel cells are sensory cells that function in light-touch detection. MCV is likely to be the etiological agent of this cancer because it was found to be clonally integrated in MCC cells and knock-down of large-T expression led to growth arrest and cell death in MCC-derived cell lines (Houben et al., 2010). In most MCV sequences from tumors, large-T carried mutations that truncate the expressed protein and exhibited altered activity (Shuda et al., 2008). This mutated large-T versions are therefore thought to be tumorigenic.

## 1.3. Polyomavirus structure

The main component of polyomavirus capsids is the major structural protein VP1 that assembles into icosahedral particles with a diameter of roughly 45 nm. Empty capsids, so called virus-like particles (VLPs), of BKV and JCV were analyzed by electron microscopy (Chang et al., 1997; Li et al., 2003). They are very similar to one another and to SV40 and Polyoma capsids, whose structures are known at atomic resolution (Liddington et al., 1991; Stehle et al., 1994). Capsids or infectious particles have not yet been produced for WUV and KIV, but due to sequence conservation, the basic principles extracted from other polyomavirus structures likely apply to them as well.

### 1.3.1. Icosahedral symmetry and quasiequivalence

Icosahedral symmetry is quite common among viruses. It enables the virus to construct a container for its DNA from a large number of identical subunits, eliminating the need to encode many different components (Crick and Watson, 1956). This kind of symmetry is based on the icosahedron, consisting of 20 triangular faces. Icosahedral particles must obey three sets of symmetry requirements: fivefold axes through the vertices of the icosahedron, threefold axes at the centers of each triangle, and twofold axes through the middle of each of the icosahedron's edges. Strict icosahedral symmetry, in which all subunits as well as their interactions with each other are identical, requires each of the 20 faces of the icosahedron to be divided into 3 equivalent subunits. Therefore, it can only apply to capsids consisting 60 identical subunits. However, the need for more complex structures that can package larger genes as well as internal proteins requires for most viruses a capsid that cannot be formed with such a small number of subunits. Therefore, most viruses, including polyomaviruses, build larger capsids constructed from multiples of 60 subunits. In these capsids, subunits must be joined in a quasi-equivalent manner, meaning that they do not form the same contacts with their neighbors throughout the capsid, but exhibit flexibility or alternative

binding modes to accommodate local symmetry mismatches (Caspar and Klug, 1962), such as assembling the same basic subunit into both pentamers and hexamers to form the viral capsid (Fig. 2).

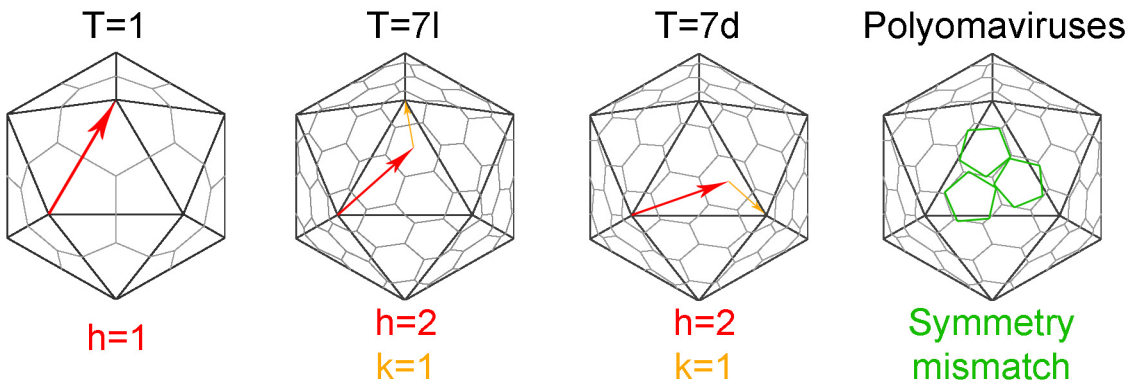



**Fig. 2: Icosahedral lattices and their triangulation numbers.**
Since the polyomavirus capsids do not consist of hexamers and pentamers, but of pentamers only, there is a symmetry mismatch. Icosahedral models are from the icosahedral symmetry server http://viperdb.scripps.edu/icos_server.php.

Icosahedral particles can be described by their triangulation number T. Only specific values are allowed for T, which is defined as $T = h^2 + hk + k^2$, with h and k being integers, and h > k. h and k can be determined by counting the number of subunits when going from one fivefold axis to the next via the sixfold symmetry axes generated by the subunits, with h and k pointing in different directions. The easiest case of icosahedral particle with 60 subunits has a triangulation number of T = 1. For values of T greater than 3, two possibilities exist for the h and k vectors, *dextro* and *laevo* (*d* and *l*, respectively), resulting in icosahedrons that are mirror images of each other (Fig. 2). The number of subunits can be calculated by multiplying the triangulation number by 60.

### 1.3.2. Structure of polyomavirus capsids

Polyomaviruses capsids form a T = 7*d* icosahedral lattice (Finch, 1974). They had therefore been thought to consist of 420 subunits, grouped into 12 pentamers and 60 hexamers of the major capsid protein VP1. However, both the five- and six-coordinated morphological units of the Polyoma capsid were identified as pentamers and the capsid consists of only 360 molecules of VP1 (Fig. 3) (Rayment et al., 1982). This result appeared to challenge the concept of quasiequivalence as the same pentameric module was able to form both pentavalent and hexavalent contacts. The crystal structure of SV40 at 3.8 Å resolution provided the molecular basis for the unexpected capsid construction of polyomaviruses (Liddington et al., 1991). It revealed that the interpentamer contacts are primarily mediated by C-terminal extensions of VP1, termed arms, which protrude in different

directions from each VP1 monomer and contact neighboring pentamers. Both five- and six-coordinated pentamers donate and accept five C-terminal arms to tie the capsid together. The VP1 pentamer cores and the contacts between incoming arm and acceptor monomer are identical throughout the capsid, but the direction of the C-terminal arms is not. Thus, the polyomavirus capsid also uses a form of quasiequivalence as the basis for its construction.

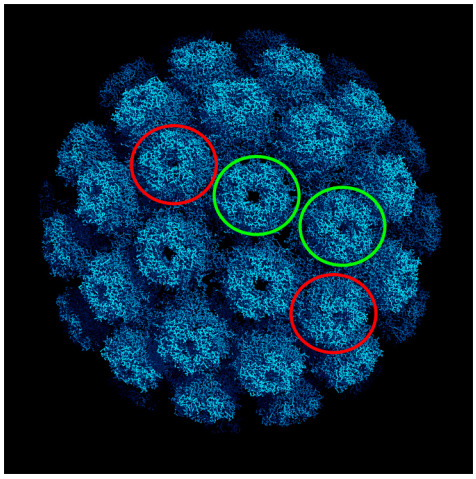


**Fig. 3: Structure of SV40.**
X-ray crystal structure (1SVA) of the complete SV40 virion (Liddington et al., 1991; Stehle et al., 1996). Two five-coordinated VP1 pentamers are circled in red; their five-fold symmetries extend throughout the virion. Two symmetry-mismatched six-coordinated VP1 pentamers are circled in green. Their five-fold symmetry is only valid for their pentamer cores, but not their C-terminal arms nor their neighboring pentamers.

### 1.3.3. Structure of VP1

The structure of VP1 is known from studies with SV40 and Polyoma virions (Liddington et al., 1991; Stehle et al., 1996; Stehle and Harrison, 1996; Stehle et al., 1994) as well as one structure of an unassembled Polyoma pentamer at 1.9 Å resolution (Stehle and Harrison, 1997). Each VP1 monomer folds into two antiparallel β-sheets: one sheet formed by strands B, I, D and G (the BIDG sheet) and another sheet formed by strands C, H, E and F (the CHEF sheet) (Fig. 4A). The b-strands are linked by loops, which are named after the two b-strands they connect. The two β-sheets stack against one another to form a β-sandwich, and are organized into a jelly-roll fold (Fig. 4B), a structural motif that is quite often observed in virus capsid proteins, but is not as common in non-viral proteins. It is, however, found in proteins of the tumor necrosis factor family as well as in carbohydrate-metabolizing enzymes and different types of lectins.

Five VP1 monomers assemble around a fivefold symmetry axis into a ring-shaped pentamer. For clarity, the clockwise and anticlockwise neighbors of a given VP1 monomer in the pentamer will be designated cw and ccw, respectively. Interactions between two monomers within a pentamer are extensive and bury 2,800 Å of surface. The N-terminal part of the G β-strand (G1) of one monomer contributes to the CHEFcw sheet of the clockwise neighboring monomer, and its DE-loop inserts between the EFcw-loop and the CHEFcw sheet of the clockwise neighbor. The cores of VP1 pentamers are therefore very compact and stable against protease digestion as well as against chaotropic salts. Monomeric VP1 has not been isolated in a folded form.

Structural studies with Polyoma revealed that receptors are bound on the outer edge of the capsid, at the top surface of the VP1 pentamer (Stehle and Harrison, 1996, 1997; Stehle et al., 1994). This surface is formed by the BC-, DE- and HI-loops of each VP1 monomer. The extensive BC-loop can be subdivided into two parts facing in different directions (BC1 and BC2). Unlike the well-conserved VP1 core structure, these loops exhibit considerable sequence variability among polyomaviruses, accounting for their different receptor specificities.

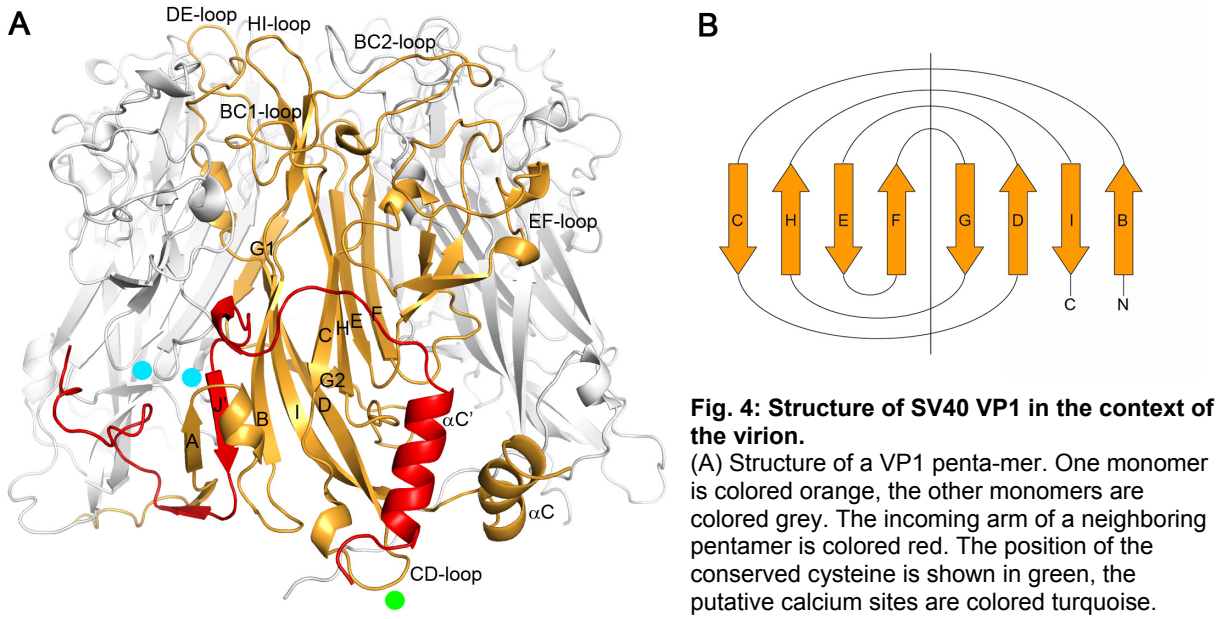



**Fig. 4: Structure of SV40 VP1 in the context of the virion.**
(A) Structure of a VP1 penta-mer. One monomer is colored orange, the other monomers are colored grey. The incoming arm of a neighboring pentamer is colored red. The position of the conserved cysteine is shown in green, the putative calcium sites are colored turquoise.
(B) Schematic of the jelly-roll fold.

### 1.3.4. Structural basis of capsid stability

A C-terminal arm emerges from the base of each VP1 monomer, and interacts with VP1 in a neighboring pentamer, thus tying together the pentamers in the capsid. Each incoming C-terminal arm adds one strand (J') to the edge of the BIDG sheet of a monomer. The J' strand is then fixed by the N-terminus of the accepting monomer, which acts as a clamp. It forms a short helix and augments the β-sheet with another β-strand (A) so that an AJ'BIDG sheet is formed. The assembled structure is stabilized by two calcium ions that are complexed by residues on the incoming C-terminal arm and the N-terminal clamp as well as core residues of the accepting pentamer (Fig. 4).

Polyomavirus capsids are also stabilized by disulfide bonds. A cysteine residue in the CD-loop at the bottom of the VP1 pentamer is strictly conserved among polyomaviruses. In SV40, this C104 forms disulfide bonds with C104 residues on adjacent pentamers, forming a covalently cross-linked capsid. Each of the C104 residues in five-coordinated pentamers is approached by C104 residues from two neighboring six-coordinated pentamers, and two of

the three cysteines form a disulfide bond. C104 residues in six-coordinated pentamers engage in contacts with five-coordinated pentamers, other six-coordinated pentamers or in no contacts. In addition, there are also disulfide bonds between C9 residues in the N-terminal arm in SV40. Polyoma capsids are not cross-linked by disulfide bonds. However, C114 in the Polyoma CD-loop forms a disulfide bond with C19 in the N-terminal arm, which likely serves to hold the N-terminal clamp in place. In line with these findings, SV40 and Polyoma virions can only be disassembled in vitro by treatment with reducing and $Ca^{2+}$-chelating agents (Brady et al., 1977; Salunke et al., 1986). Since both the calcium-binding sites and the cysteines are conserved among all polyomaviruses, the capsids of other polyomaviruses are likely stabilized by similar interactions.

### 1.3.5. VP2 and VP3

The minor capsid proteins VP2 and VP3 share a common C-terminus, with VP2 having additional amino acids at its N-terminus. Moreover, the N-terminus of VP2 is myristylated (Streuli and Griffin, 1987), which is important for infection (Sahli et al., 1993). Both VP2 and VP3 interact with VP1 with their C-termini – they bind into the central pore of VP1 in such a way that one VP1 pentamer binds one molecule of VP2 or VP3 (Chen et al., 1998).

## 1.4. Entry pathways

During cell entry, polyomaviruses must deliver their genome from outside the cell to the nucleus. There, viral genes are transcribed, the viral DNA is replicated and progeny virus is assembled, all assisted by the cellular machineries. While traffic through the nuclear membrane seems to be mediated by nuclear localisation signals on the N-terminal extension of VP1, one major obstacle for polyomaviruses is how to cross the membrane barrier from the outside of the cell or an endocytic compartment into the cytosol. While enveloped viruses can deliver their genome across membranes by membrane fusion, non-enveloped viruses do not have this option. They often undergo a stepwise program of conformational changes in their capsid proteins to enable them to penetrate this membrane barrier. Since viruses often use cellular factors to trigger these changes, their uncoating is highly correlated with movement of the viral particle within the cellular compartments.

### 1.4.1. Internalization and intracellular transport

BKV, SV40 and Polyoma enter cells by two related pathways: cholesterol/lipid raft-mediated endocytosis and caveolar endocytosis (Eash et al., 2004a; Gilbert and Benjamin, 2000; Gilbert et al., 2003; Moriyama et al., 2007; Pelkmans et al., 2002). In both cases, viruses attach to gangliosides on the exoplasmic leaflet of the plasma membrane of host cells (Low et al., 2006; Smith et al., 2003; Tsai et al., 2003). The movement of the freely

diffusing virus-receptor complexes becomes confined by the action of the cortical actin cytoskeleton below the intracellular leaflet of the membrane within 5-10s (Ewers et al., 2005). The viruses initiate their own internalization as multivalent binding of the viral capsid to gangliosides induces curvature of the plasma membrane to yield a tight-fitting invagination that contains one virion (Ewers et al., 2010). This process requires both the long, saturated fatty acid chains of the gangliosides and cholesterol. The ensuing membrane scission depends on cellular fusion factors, actin dynamics and tyrosine kinase signalling (Dangoria et al., 1996; Pelkmans et al., 2005; Pelkmans et al., 2002). This process mostly occurs independently of caveolae, small, cholesterol-rich membrane domains associated with a coat of caveolin-1 (Damm et al., 2005). However, since the viruses tend to bind to caveolae due to their high content in gangliosides and their intrinsic curvature, they can also trigger endocytosis of caveolae, which is dependent on dynamin II and slower than the cholesterol-associated pathway (Mercer et al., 2010; Pelkmans et al., 2002).

The viruses then pass through classical endocytic compartments such as early and late endosomes, which form tubular extensions that finally detach from late endosomes. These then move along microtubules and likely deliver the viruses to the endoplasmic reticulum (ER) (Eash and Atwood, 2005; Gilbert and Benjamin, 2004; Gilbert and Benjamin, 2000; Gilbert et al., 2003; Kartenbeck et al., 1989; Pelkmans et al., 2001; Qian et al., 2009). It is enticing to speculate that clustering of gangliosides in the endolysosome leads to a process similar to internalization at the plasma membrane, which this time packages the virus into vesicles targeted to the ER. The caveosome, described as an intermediate organelle in the entry pathway of SV40 (Pelkmans et al., 2001), is likely a modified late endosome (Mercer et al., 2010). Gangliosides can be purified from ER fractions, suggesting that the virus stays attached to them all the way from the plasma membrane to the ER (Tsai et al., 2003). Interestingly, pentameric bacterial toxins such as shiga toxin or cholera toxin attach to gangliosides and use cholesterol-dependent endocytosis, too (Ewers et al., 2010; Romer et al., 2007). Likewise, they can induce membrane curvature (Ewers et al., 2010; Romer et al., 2007), suggesting that the spacing and orientation of oligosaccharide binding sites on pentamers is conducive to creating membrane invaginations (4.5.2). In addition, even artificial particles that are coated with GD1a antibody were delivered from the plasma membrane to the ER (Qian et al., 2009).

By contrast, JCV enters cells by the classical clathrin-mediated pathway (Pho et al., 2000; Querbes et al., 2004). It then also passes through early and late endosomes, colocalizes with SV40 and is transported to the ER in tubular vesicles on microtubules (Querbes et al., 2006). Since JCV likely does not use a ganglioside as receptor, it is unclear what targets it to the ER.

### 1.4.2. Conformational changes and membrane penetration

Polyomaviruses depend on an intact ER for infection. They accumulate in the smooth endoplasmic reticulum (ER) starting at 3 h post infection (Kartenbeck et al., 1989), and use enzymes involved in protein folding and ER-associated degradation (ERAD) for uncoating and membrane penetration. The ERAD pathway recognizes terminally misfolded proteins and retro-translocates them into the cytosol for degradation in the proteaseome (Hebert and Molinari, 2007). Polyoma, BKV and SV40 require proteins involved in ERAD for infection (Jiang et al., 2009b; Lilley et al., 2006; Schelhaas et al., 2007). It is likely that the action of ER enzymes exposes hydrophobic stretches on polyomavirus capsids, which are then recognized by the ERAD machinery.

The ER chaperone ERp57, a thiol-disulfide oxidoreductase isomerizes the disulfide bonds at the contacts between five- and six-coordinated pentamers in the capsid, where C104 from a VP1 in a five-coordinated pentamer is in close proximity to two C104 residues from VP1 in two adjacent six-coordinated pentamers (Schelhaas et al., 2007). In the virion, a random two of the three cysteins are linked with a disulfide bond. Isomerization by ERp57 uncouples the five-coordinated pentamers from the network of disulfide bonds. The most likely explanation is that after isomerization, the two six-coordinated pentamers are linked by a disulfide bond, while C104 on the five-coordinated pentamer is free. However, the uncoupled pentamers cannot be released because their C-terminal arms are still fixed by two calcium ions each. Their release is therefore thought to occur after retro-translocation into the cytosol, where the calcium level is much lower than in the ER.

In the case of Polyoma, disulfide isomerization by the PDI-like enzyme ERp29 leads to a conformational change that exposes the C-terminal arms, renders the virus more hydrophobic and likely enables it to penetrate the ER membrane (Gilbert et al., 2006; Magnuson et al., 2005). Exposure to the low pH of endolysosomes seems to help that process. It is not clear, however, if Polyoma activated in this manner can translocate through the ER membrane by itself, or whether it is recognized as a substrate for the ERAD machinery.

In addition, the minor capsid proteins might be involved in membrane penetration in two ways: First, VP2 carries a myristic acid at its N-terminus (Streuli and Griffin, 1987) (1.3.5), a modification that was shown to be of importance for membrane interactions of several other viruses. Second, both VP2 and VP3 contain a hydrophobic peptide that can insert into membranes and likely form pores that disrupt the membrane (Daniels et al., 2006). It is not clear whether they function during virus entry or during egress of progeny virus from an infected cell. Interestingly, a short non-structural protein of SV40, VP4, also shares the C-terminus of VP2 and VP3, has properties similar to VP2 and VP3, and is only expressed late in infection (Daniels et al., 2007).

## 1.5. Receptor recognition

Sialylated oligosaccharides have been implicated as receptors for polyomaviruses because treatment of host cells with neuraminidase was found to reduce or abolish infection of Polyoma, BKV and JCV (Cahan and Paulson, 1980; Fried et al., 1981; Liu et al., 1998b; Seganti et al., 1981). Sialic acids are a group of acidic acetylated amino sugars that are present on virtually every cell type in higher vertebrates (Varki, 2007). Sialylation is a terminal modification on glycolipids and glycoproteins, and sialic acids can be attached to a variety of monosaccharides via different glycosidic linkages. Physiological functions include cell attachment, general charge repulsion of blood cells as well as roles in neuronal plasticity, immune regulation and glomerular filtration. The predominant sialic acid in humans is α-5-N-acetyl neuraminic acid (NeuNAc) (Fig. 5A). Simians, the natural hosts of SV40, also feature NeuNAc, but their most common sialic acid is α-5-N-glycolyl-neuraminic acid (NeuNGc) (Varki, 2001). The NeuNGc variant carries a $CH_2$–OH group instead of a $CH_3$ group on the sialic acid amide group.

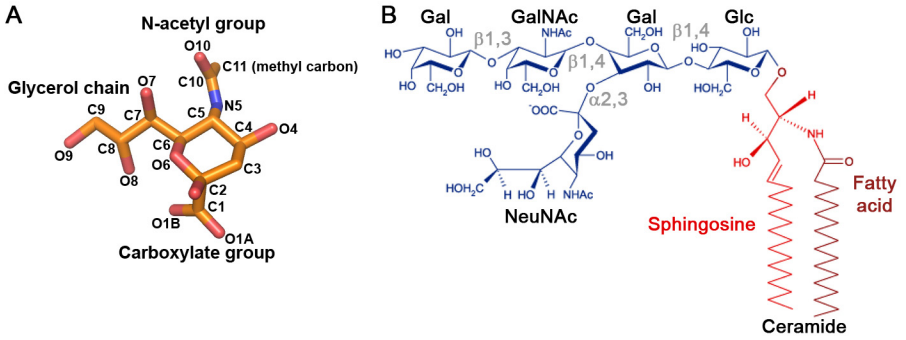



**Fig. 5: Structures of polyomavirus receptor components.**
(A) Structure of NeuNAc.
(B) Chemical structure of the ganglioside GM1.

Gangliosides have been identified as receptors for SV40, Polyoma, BKV and MCV (Tab. 1). Gangliosides are sialylated ceramide-based glycosphingolipids with saturated fatty acyl chains (Fig. 5B). In line with the endocytosis mechanisms of SV40, Polyoma, and SV40, gangliosides in a membrane associate with saturated phospholipids and cholesterol as well as some membrane proteins. They can be found in dynamic local lipid phase separations, sometimes termed “lipid rafts” (Lingwood and Simons, 2010). These transient membrane microdomains can influence membrane fluidity, receptor clustering, and assembly of signaling complexes.

The carbohydrate moieties of gangliosides are synthesized in the Golgi apparatus by stepwise addition of monosaccharides (Fig. 6) (Kolter et al., 2002). Since the synthesizing enzymes are highly specific for both monosaccharides and glycosidic linkages, ganglioside glycans exhibit regular oligosaccharide sequences and linkages. They typically consist of two branches, termed “arms”. One arm, the “right” arm by convention, consists entirely of sialic acids and is used to classify gangliosides into series. Those of the a-series, e.g. GM1, carry

only one NeuNAc on their right arm, while those of the b-series, e.g. GD3 or GT1b, all carry two NeuNAc residues at this location. All gangliosides feature a stem consisting of glucose (Glc) and galactose (Gal) in the sequence Gal-β1,4-Glc. The left arm is synthesized by the stepwise addition of *N*-acetyl-galactosamine (GalNAc), Gal and NeuNAc to the branching Gal residue.

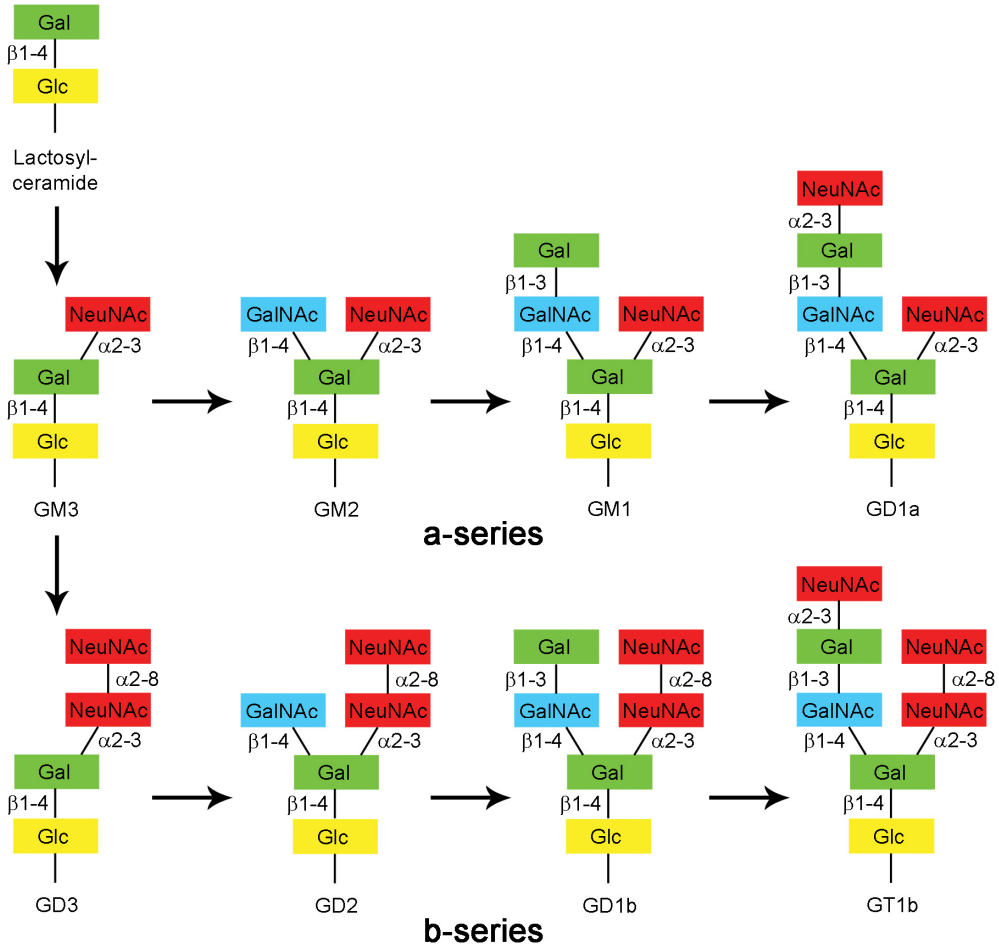



**Fig. 6: Sequences and branching patterns of the most common gangliosides.**
Arrows indicate biosynthesis reactions. The transferases adding monosaccharides to the left arm are the same for all series. The sialyltransferases acting on the right arm are specific.

Sialylated oligosaccharides have also been shown to be a component of the JCV receptor, but it is unclear whether the sialic acid is attached to a glycoprotein or glycolipid, or which sialic acid linkage JCV recognizes (Tab. 1). An α2,6-linked glycoprotein, the ganglioside GT1b or α2,3-linked sialic acid have all been proposed as receptors. In addition, signalling by the serotonin receptor 5-$HT_{2A}$R is necessary for JCV entry, suggesting it as a receptor candidate. Likewise, a glycoprotein carrying terminal α2,3-linked sialic acid has been suggested as a receptor for BKV, and entry of SV40 depends on the presence of MHC class I molecules in some cell lines (Tab. 1).

**Tab. 1: Overview of polyomavirus receptor candidates**

| Virus | Receptors | References |
|---|---|---|
| SV40 | GM1<br>MHC class I | Tsai et al., 2003<br>Atwood and Norkin, 1989 |
| BKV | GD1b and GT1b<br>$\alpha$2,3-linked glycoprotein | Low et al., 2006<br>Dugan et al., 2005 |
| JCV | $\alpha$2,6-linked glycoprotein<br>$\alpha$2,3-linked sialic acid<br>GT1b<br>Serotonin receptor (5-$HT_{2A}$R) | Liu et al., 1998b<br>Dugan et al., 2008<br>Komagome et al., 2002<br>Elphick et al., 2004 |
| Polyoma | GD1a and GT1b | Smith et al., 2003<br>Tsai et al., 2003 |
| MCV | GT1b | Erickson et al., 2009 |

## 1.6. Aims and significance

The main goal of this thesis was to understand the structural basis of receptor binding and specificity of several related polyomaviruses. To that end, several questions were addressed:

- Which are the oligosaccharide motifs bound by SV40, BKV and JCV?
- How do their VP1 proteins recognize these motifs?
- Which are the structural determinants of the viruses' binding specificities?
- Does receptor binding lead to conformational changes in the viral protein?
- Does the viruses' receptor specificity explain features of their endocytosis, tropism and pathogenesis?
- How do the structures of the only distantly related VP1 proteins of WUV and KIV compare to those of other polyomaviruses?

The approach to answer these questions was to combine glycan microarray screening with structural analysis by X-ray crystallography. In most studies of protein-oligosaccharide interactions, the subject is the interaction of one protein with one defined oligosaccharide. However, in order to understand the oligosaccharide specificity of proteins, it is essential to determine which carbohydrates are not bound. This is often hard to deduce from the atomic structure of the complex alone. In addition, there are relatively few structures of protein-

carbohydrate complexes available to date so that the structures solved in this thesis may help to define unifying principles in carbohydrate recognition.

By comparing the receptor specificities of several related viruses, this study also aimed at understanding how structural changes in a virus enable it to adapt to new host cells, spread to different tissues, or to infect a new host species since all of these are dependent on virus-receptor interactions. Finally, the structures of receptor complexes of pathogenic viruses like BKV and JCV are a first step towards the rational design of drugs that might interfere with the attachment process, thus blocking viral infection before the virus even enters a host cell.

# 2. Materials and Methods

## 2.1. Materials

This chapter lists all materials used for the experiments performed by me for this thesis in the laboratory of Prof. Thilo Stehle. Materials for experiments that were conducted in the laboratories of collaborators are described in the chapters on the experiments (2.4.1-2.4.5).

### 2.1.1. Instruments and tools

#### 2.1.1.1. General

| | |
|---|---|
| Autoclave VX-95 | Systec, Wettenberg, Germany |
| Centrifuge 5414 D | Eppendorf, Wesseling-Bertzdorf, Germany |
| Centrifuge Multifuge 1L-R | Kendro, Langenselbold, Germany |
| Centrifuge Sorvall RC-6 | Kendro, Langenselbold, Germany |
| Gel Doc XR imaging system | Bio-Rad, Munich, Germany |
| Gel dryer Slab 2300 | LKB, Bromma, Sweden |
| Icemaker | Scotsman, Milan, Italy |
| Incubation Shaker Unitron | InforsHT, Bottmingen-Basel, Switzerland |
| Microwave NN-E203 WB | Panasonic, Hamburg, Germany |
| PCR cycler iCycler | Bio-Rad, Munich, Germany |
| pH meter PB-11 | Sartorius, Göttingen, Germany |
| Photometer NanoDrop ND-1000 | Thermo Scientific, Waltham, USA |
| Photometer SmartSpec Plus | Bio-Rad, Munich, Germany |
| Pipette Mettler (2.5 µL) | Mettler Toledo, Columbus, USA |
| Pipettes Pipetman | Gilson, Middleton, USA |
| Quartz cuvette (75 µL) | Hellma, Jena, Germany |
| Scale 323S-OCE | Sartorius, Göttingen, Germany |
| Scale 4202S-OCE | Sartorius, Göttingen, Germany |
| Scale Genius | Sartorius, Göttingen, Germany |
| SDS-PAGE Mini Protean 3 Cell | Bio-Rad, Munich, Germany |
| SDS-PAGE Mini Sub-Cell GT | Bio-Rad, Munich, Germany |
| SDS-PAGE PowerPac Basic | Bio-Rad, Munich, Germany |
| Shaker DOS-10L | NeoLab, Heidelberg, Germany |
| Shaker Intelli-Mixer | NeoLab, Heidelberg, Germany |
| Sorvall Rotor SS34 | Kendro, Langenselbold, Germany |
| Sorvall Rotor SLC4000 | Kendro, Langenselbold, Germany |
| Sorvall Rotor 75002000 | Kendro, Langenselbold, Germany |
| SpeedVac Heto Vacuum Centrifuge | Heto, Allerod, Denmark |
| Transilluminator Universal Hood II | Bio-Rad, Munich, Germany |

#### 2.1.1.2. Protein Purification

| | |
|---|---|
| High pressure homogenizer EmulsiFlex | Avestin, Ottawa, Canada |
| Sonicator Digital Sonifier 250 | Branson, Danbury, USA |
| Analytical FPLC Ettan | GE Healthcare, Uppsala, Sweden |
| FPLC BioLogic Duo Flow | Bio-Rad, Munich, Germany |

| | |
|---|---|
| Peristaltic EconoPump | Bio-Rad, Munich, Germany |
| Pharmacia LKB Controller LCC-501 Plus | GE Healthcare, Uppsala, Sweden |
| Pharmacia LKB Pump P-500 | GE Healthcare, Uppsala, Sweden |
| Pharmacia LKB REC 102 | GE Healthcare, Uppsala, Sweden |
| Affinity column HisTrap 1 mL | GE Healthcare, Uppsala, Sweden |
| Affinity column HisTrap FF 5 mL | GE Healthcare, Uppsala, Sweden |
| Column Sepharose 12 3.2/30 | GE Healthcare, Uppsala, Sweden |
| Column Superdex 200 10/30 | GE Healthcare, Uppsala, Sweden |
| Column Superdex 200 16/60 | GE Healthcare, Uppsala, Sweden |

#### 2.1.1.3. X-ray crystallography

| | |
|---|---|
| Cryosystem X-Stream | Rigaku/MSC, The Woodlands, USA |
| Crystallization Robot Freedom Evo | Tecan, Männedorf, Switzerland |
| Depression wells | Hampton research, Aliso Viejo, USA |
| Detector ADSC Quantum 4 at ESRF | ADSC, Poway, USA |
| Detector Mar345 dtb | Mar research, Hamburg, Germany |
| Detector MarCCD 225 at SLS | Mar research, Hamburg, Germany |
| Detector PILATUS 6M at SLS | Dectris Ltd., Baden, Switzerland |
| Microscope MZ16 | Leica, Wetzlar, Germany |
| Microscope S6E | Leica, Wetzlar, Germany |
| Optics Varimax HF | Rigaku/MSC, The Woodlands, USA |
| Rotating Anode Micromax 007 HF | Rigaku/MSC, The Woodlands, USA |

### 2.1.2. Consumables

All chemicals were purchased from Merck (Darmstadt, Germany), Roth (Mannheim, Germany) or Sigma Aldrich (Munich, Germany) unless stated otherwise.

#### 2.1.2.1. Molecular Biology

| | |
|---|---|
| 6x loading dye for agarose gels | Fermentas, St. Leon-Rot, Germany |
| BamHI | Fermentas, St. Leon-Rot, Germany |
| dNTPs | Promega, Mannheim, Germany |
| DpnI | Fermentas, St. Leon-Rot, Germany |
| LB medium | Sigma Aldrich, Munich, Germany |
| NdeI | Fermentas, St. Leon-Rot, Germany |
| PfuUltra High-Fidelity DNA polymerase | Stratagene, La Jolla, USA |
| T4 DNA ligase | Fermentas, St. Leon-Rot, Germany |
| Tag polymerase LC | Fermentas, St. Leon-Rot, Germany |
| Wizard Plus SV Minipreps | Promega, Mannheim, Germany |
| XhoI | Fermentas, St. Leon-Rot, Germany |

##### 2.1.2.1.1. Plasmids

The plasmid pET15b (Novagen, Darmstadt, Germany) was used for expression of all proteins. A plasmid encoding full-length SV40 VP1 was provided by Prof. Robert L. Garcea (University of Colorado, USA), and plasmids encoding encoding full-length BKV and JCV VP1 were provided by Aisling Dugand and Megan Gasparovich (Brown University, USA).

The expression constructs for KIV and WUV VP1 were made by Dr. Dennis Macejak (University of Colorado, USA) according to my design. Site-directed mutagenesis of WUV and BKV VP1 was performed by Jianbao Wang and Holger Hengel (Eberhard Karls Universität Tübingen), respectively, under my supervision.

#### 2.1.2.1.2. *E. coli* strains

Nova Blue Novagen (Merck), Darmstadt, Germany
endA1 hsdR17 ($r_{K12}^{-}$ $m_{K12}^{+}$) supE44 thi-1 recA1 gyrA96 relA1 lac
F′[proA+B+ lacIqZΔM15::Tn10] ($Tet^{R}$)

BL21 (DE3) Novagen (Merck), Darmstadt, Germany
$F^{-}$ ompT $hsdS_B$ ($r_B^{-}$ $m_B^{-}$) gal dcm (DE3)

#### 2.1.2.1.3. Primers

**Tab. 2: Primers used for clonign, site-directed mutagenesis and sequencing**
All primers were puchased from biomers.net, Ulm, Germany.

| Name | Sequence (5' – 3') |
|---|---|
| Cloning primers | |
| SV40_G31_fwd_Nde | AGTAGTCATATGGGAGGAATAGAAGTTCTAGGA |
| SV40_N298_rev_Bam | ATGGATCCTAGTTTTTCACAGACCGCTTTCTAAG |
| BKV_JCV_fwd_Nde | AGTAGTCATATGGGAGGAGTAGAAGTTCTAGAA |
| BKV_P301_rev_Bam | CTGGATCCTATGGGTAAGGATTCTTTACAGATC |
| JCV_N291_rev_Bam | CTGGATCCTAGTTTTTAACCCTCCTTTTCCT |
| WUV_fwd_Nde | CAAAGAAGGGTTAGAAACCCATATACTTAGGATCCAG |
| WUV_rev_Bam | CTGGATCCTAAGTATATGGGTTTCTAACCCTTCTTTG |
| Primers for site-directed mutagenesis | |
| WUV_R197K_fwd | AGTTCTGAAACACCCAAGGCAAGGGTCACTAAT |
| WUV_R197K_rev | ATTAGTGACCCTTGCCTTGGGTGTTTCAGAACT |
| BKV_K68S_fwd | AGGGGCTTTAGTCTAAGTCTAAGTGCTGAAAATG |
| BKV_K68S_rev | CATTTTCAGCACTTAGACTTAGACTAAAGCCCCT |
| Sequencing primers | |
| T7_fwd | TAATACGACTCACTATAGG |
| T7_rev | GCTAGTTATTGCTCAGCGG |

### 2.1.2.2. Protein Purification

Acrylamide-Bisacrylamide Rotiphorese 30 Roth, Karlsruhe, Germany
Bradford Protein Assay Bio-Rad, Munich, Germany
Centrifugal filters (0.22 µm) Corning, Corning, USA
Concentrators Microcon YM3 Millipore, Schwabach, Germany

| | |
|---|---|
| Concentrators Ultra 4 and Ultra 15 | Millipore, Schwabach, Germany |
| Coomassie Brillant Blue R-250 | Bio-Rad, Munich, Germany |
| Membrane filters (0.45 and 5 µm) | Millipore, Schwabach, Germany |
| Page Ruler protein ladder | Fermentas, St. Leon-Rot, Germany |
| SDS-PAGE 10x buffer Rotiphorese | Roth, Karlsruhe, Germany |
| Thrombin | GE Healthcare, Uppsala, Sweden |

#### 2.1.2.3. Crystallization and X-ray crystallography

##### 2.1.2.3.1. General

| | |
|---|---|
| Capillaries | Hampton Research, Aliso Viejo, USA |
| Cover slides siliconized 22mm | Hampton research, Aliso Viejo, USA |
| Crystallization plate 24 well | Hampton research, Aliso Viejo, USA |
| Crystallization 96 well Intelli-Plates | Art Robbins, Sunnyvale, USA |
| Crystallization sealing foil for 96 well plate | HJ-Bioanalytik, Mönchengladbach, Germany |
| Fiber loops and magnetic bases | Hampton research, Aliso Viejo, USA |
| Syringe top filters (0.2 and 0.45 µm) | VWR International, Vienna, Austria |
| Vacuum grease Baysilone | GE Bayer Silicones, Leverkusen, Germany |

##### 2.1.2.3.2. Crystallization Screens

| | |
|---|---|
| Additive Screen | Hampton research, Aliso Viejo, USA |
| Crystal Screens 1 and 2 | Hampton research, Aliso Viejo, USA |
| Grid Screen | pipetted in the Stehle Lab |
| PEG/Ion Screens 1 and 2 | Hampton research, Aliso Viejo, USA |
| Precipitant Synergy Screen | pipetted in the Stehle Lab |
| Wizard I, II and III Screens | Emerald BioSystems, Bainbridge Island, USA |

#### 2.1.2.4. Oligosaccharides

| | |
|---|---|
| GD3 (= Disialyllactose) sodium salt | Sigma Aldrich, Munich, Germany |
| GM1 oligosaccharide sodium salt | Alexis, CH |
| GT1b | Dr. Akihiro Imamura, Gifu University, Japan |
| LSTc | Dextra Ltd., Reading, UK |

### 2.1.3. Software

#### 2.1.3.1. DNA and protein tools

| | |
|---|---|
| Bio-Rad BioLogic Duo Flow | Bio-Rad, Munich, Germany |
| DNA Calculator | Sigma Genosys, The Woodlands, USA<br>*www.sigma-genosys.com* |
| Fermentas double digest | Fermentas, St. Leon-Rot, Germany<br>*www.fermentas.com* |
| LALIGN | William R. Pearson, University of Virginia, USA<br>*www.expasy.org* |
| NEB Cutter | New England Biolabs Inc., Ipswich, USA<br>*www.neb.com* |
| PeptideCutter | Gasteiger et al., 2005<br>*www.expasy.org* |
| ProtParam | Gasteiger et al., 2005<br>*www.expasy.org* |

#### 2.1.3.2. X-ray crystallography

| | |
|---|---|
| AMoRe | Navaza, 1994 |
| CCP4 | CCP4, 1994 |
| Coot | Emsley and Cowtan, 2004 |
| HKL | HKL research, Charlottesville, USA |
| LSQMAN | Kleywegt, 1996 |
| Mar345dtb | Mar research GmbH, Hamburg, Germany |
| Phaser | McCoy et al., 2007 |
| Phenix | Adams et al., 2010 |
| Pymol | Schrödinger, Portland, USA |
| Refmac | Murshudov et al., 1997 |
| Superpose | Krissinel and Henrick, 2004 |
| TLS Motion Determination | Painter and Merritt, 2006<br>http://*skuld.bmsc.washington.edu/~tlsmd/* |
| XDS | Kabsch, 1993 |

## 2.2. Molecular Biology

### 2.2.1. Bacteria culture and glycerol stocks

*E. coli* strain Nova Blue was used for plasmid production, and *E. coli* strain BL21 (DE3) was used for protein expression. Bacteria transformed with a pET15b-based expression plasmid that contains an ampicillin resistance gene were cultured in LB medium in the presence of 50 µg/mL ampicillin at 37 °C in a shaking incubator. For production of competent cells, bacteria not harbouring a resistance plasmid were grown in SOB medium in the absence of antibiotics. All media were autoclaved, followed by addition of antibiotics, before use.

For overnight cultures, 10 mL of the appropriate medium in 50 mL Falcon tubes were inoculated with bacteria either from a glycerol stock (see below) or from a colony on an agar plate.

To prepare glycerol stocks, 900 µL of confluent bacteria culture were mixed on ice with 300 µL 50 % (v/v) glycerol, and frozen at -80 °C.

### 2.2.2. Isolation of plasmid DNA from *E. coli*

Plasmid DNA was extracted with the Wizard Miniprep kit according to the manufacturer's instructions, with the sole exception that the purified DNA was eluted in 50 µL $H_2O$ instead of 100 µL.

The DNA concentration was determined by the absorbance of the sample at 260 nm. At that wavelength, an absorbance of 1.0 corresponds to a concentration of 50 µg/mL for double-stranded DNA.

### 2.2.3. Polymerase chain reaction

DNA for cloning into an expression vector was amplified using polymerase chain reaction (PCR) with the high-fidelity polymerase Pfu. PCR that verified the presence of the insert after molecular cloning was performed with Taq polymerase.

The compositions of the PCR reactions were calculated according to the manufacturer's instructions for reaction volumes of 50 and 20 µL for preparative and analytical applications, respectively. Reaction mixtures were pipetted while keeping all reagents in tubes on ice. Positive and negative controls were included in each set of reactions. DNA amplified by PCR was recovered by isopropanol precipitation (2.2.5).

The following PCR program was used for all applications except site-directed mutagenesis (2.2.10):

| | | |
|---|---|---|
| 1x | 2 min | 95 °C |
| 30x | 30 s | 95 °C |
| | 30 s | $T_a$ |
| | 1 min | 72 °C |
| 1x | 10 min | 72 °C |
| 1x | ∞ | 4 °C |

### 2.2.4. Agarose gel electrophoresis

Each PCR reaction was verified by electrophoresis in a 1.8 % agarose gel, for which 540 mg agarose were melted in 30 mL TAE buffer in the microwave and cast into the gel chamber. For staining the DNA, 1.5 µl ethidium bromide solution (10 mg/mL) were added before the gel had solidified. A 2 µL aliquot of each reaction was mixed with 2 µL 6x loading dye and 8 µL $H_2O$ before loading into the gel. Gels were run for 1 h at 90 V before visualization under UV light.

50x TAE buffer:
60.5 g Tris
25 mL 0.5 M EDTA pH 8.0
14.3 mL glacial acetic acid
$H_2O$ ad 250 mL

### 2.2.5. DNA precipitation

Isopropanol precipitation was used to purify DNA from PCR samples (2.2.3). Two PCR reactions (100 µL total) were mixed with 110 µL 100 % isopropanol (-20 °C) and 10 µL sodium acetate (pH 5.2; 3 M; 4 °C). The DNA was precipitated for 30 min at -20 °C before it was pelleted by centrifugation for 10 min at 14,000 g (4 °C). After discarding the supernatant, the pellet was washed with 75 µL 70 % (v/v) ethanol (-20 °C), and centrifuged again for 5 min at 14,000 g (4 °C). The supernatant was again discarded and the pellet was dried for 5 min in a vacuum centrifuge at room temperature. The pellet was then resuspended in 20 µL $H_2O$, and the DNA concentration was determined (2.2.2).

Ethanol precipitation was used to purify DNA from restriction samples (2.2.6). One restriction reaction (20 µl) was mixed with 60 µL 100 % ethanol (-20 °C) and 3 µL sodium acetate (pH 5.2; 3 M; 4 °C). DNA was precipitated overnight at -80 °C and recovered by centrifugation (see above). The pellet was resuspended in 10 µL $H_2O$.

## 2.2.6. Restriction

To obtain DNA with „sticky ends“ for positional cloning, both insert DNA amplified by PCR (2.2.3) and vector were treated with two restriction endonucleases. The composition of each 20 µL restriction sample was 1 µg of insert or 2 µg of vector, 1 µL of each restriction enzyme (10 U/µL) and buffer according to the manufacturer's instructions. Restriction reactions were pipetted on ice and incubated for 4 h or overnight at 37 °C. Then, the enzymes were inactivated by incubation at 65 °C for 20 min. Digested vector was dephosphorylated by addition of 3 µL SAP after heat-inactivation of the restriction enzymes. The sample was incubated for 30 min at 37 °C. Then, SAP was heat-inactivated at 65 °C for 20 min. DNA from restriction samples was recovered by precipitation (2.2.5).

## 2.2.7. Ligation

Insert DNA was ligated into the vector using 100 ng of vector for each reaction and molar ratios of insert to vector of 2:1, 3:1 or 5:1. The corresponding amount of insert depended on its length and was calculated as follows:

$$ng[Insert] = ratio\frac{Insert}{Vector} \cdot ng[vector] \cdot \frac{length[insert]}{length[vector]}$$

**Eq. 1: DNA amounts for ligation**

The composition of ligation reactions was calculated according to the manufacturer's instructions for a volume of 10 µL. Reactions were pipetted on ice and incubated overnight at 20 °C. The ligase was heat-inactivated at 65 °C for 20 min. Then, 0.2 µL of XhoI restriction endonuclease were added to the ligation reactions and they were incubated for 30 min at 37 °C. The XhoI site of pET15b lies between the NdeI and BamHI restriction sites used for cloning all contructs in this thesis, and is only present in religated or uncut vector without insert. Therefore, treatment with XhoI after ligation linearized all vector that did not contain insert. After heat-inactivation of XhoI for 20 min at 65 °C, the ligated DNA was transformed into *E. coli* Nova Blue (2.2.9).

## 2.2.8. Preparation of competent cells

SOB medium (100 mL) was inoculated with 1 mL of overnight bacterial culture (2.2.1) without antibiotics. Bacteria were grown to an $OD_{600}$ of 0.45-0.55 at 37 °C, then cooled down

on ice for 10 min and pelleted by centrifugation at 1,700 g at 4 °C in 50 mL Falcon tubes. Bacteria were resuspended in a total of 20 mL buffer TfBI (4 °C) on ice, incubated on ice for 10 min, and then pelleted by centrifugation at 1,700 g at 4 °C. The pellet was resuspended in 4 mL buffer TfBII (4°C) on ice, and resuspended bacteria were frozen in 100 µL aliquots at -80 °C.

Buffers TfBI and TfBII were mixed from autoclaved stock solutions and not autoclaved after mixing. They were adjusted to a pH of 5.8 and 6.8, respectively, with potassium hydroxide at room temperature.

Their compositions were:

| | |
|---|---|
| TfBI: | 100 mM RbCl<br>10 mM $CaCl_2$<br>50 mM $MnCl_2$<br>30 mM potassium acetate<br>15 % Glycerol |
| TfBII: | 10 mM MOPS<br>75 mM $CaCl_2$<br>10 mM RbCl<br>15 % Glycerol |

### 2.2.9. Transformation

For molecular cloning, samples from ligation reactions were transformed into *E. coli* Nova Blue. For protein expression, a sample of purified plasmid was transformed into *E. coli* BL21 (DE3).

For transformation with DNA, 100 µL of chemically competent *E. coli* (2.2.8) were thawed on ice. Ligation reaction (3 – 10 µL) or 1 µL purified plasmid (diluted 1:10) were added to the bacteria, which were incubated on ice for 10 min. Then, bacteria were heat-shocked for 45 s at 42 °C, and cooled on ice for 2 min. LB medium without antibiotics (500 µL) was added, and the mixture was incubated at 37 °C for 30 – 60 min on a shaking incubator. Samples (50 – 200 µL) of the mixture were then plated on LB agar plates containing 50 µg/ml ampicillin and incubated overnight at 37 °C.

To verify successful transformation with ligation reactions, 10 single colonies were used to inoculate overnight cultures (2.2.1), their plasmid DNA was purified (2.2.2) and analyzed by PCR (2.2.3). Once presence of the insert was confirmed, the DNA was sequenced at Eurofins MWG Operon (Ebersberg, Germany) using the T7 forward and reverse primers (2.1.2.1.3).

### 2.2.10. Site-directed mutagenesis

Mutations were introduced into the expression constructs for BKV and WUV VP1 by PCR reactions similar to the QuikChange protocol (Stratagene). Briefly, the entire plasmid was amplified by Pfu high fidelity polymerase with the primers introducing the mutation at

one point. Then, the restriction enzyme DpnI, which is specific for methylated DNA, is used to cut the parental DNA without the mutation so that only mutated plasmid is transformed.

Reaction mixtures were pipetted on ice using 125 ng of each primer and 50 ng of template DNA in a volume of 50 µL.

The following PCR program was used:

| | | |
|---|---|---|
| 1x | 2 min | 95 °C |
| 18x | 30 s | 95 °C |
| | 1 min | 55 °C |
| | 5.5 min | 68 °C |
| 1x | ∞ | 4 °C |

After PCR was complete, 1 µL DpnI (10 U/µL) was added and samples were incubated at 37 °C for 1 h before the enzyme was heat-inactivated. Then, aliquots of 1 – 10 µL of the samples were transformed into competent *E. coli* Nova Blue (2.2.9) and plated on LB-agar plates. Clones were analyzed as described above without the analytical PCR step (2.2.9).

### 2.2.11. Overexpression of recombinant proteins

All VP1 proteins were expressed from pET15b-based expression vectors in *E. coli* BL21 (DE3). 1 L of LB medium containing 50 µg/ml ampicillin in a 2 L Erlenmeyer flask was inoculated with 10 mL of a confluent overnight culture (2.2.1). The bacteria were grown at 37 °C in a shaking incubator (200 rpm). At $OD_{600}$ of 0.6 – 1.0, the temperature of the incubator was set to 20 °C, protein expression was induced with 0.4 mM IPTG, and protein was expressed overnight.

## 2.3. Protein purification and analysis

All protein purification steps were performed at 4 °C or on ice unless stated otherwise. If the purification was interrupted for more than 12 h, protein fractions were frozen at -30 °C.

### 2.3.1. Buffers for protein purification

The following buffers were used for purification of all proteins:

| | |
|---|---|
| $A_{His}$ | 50 mM Tris<br>5 % Glycerol (v/v)<br>250 mM NaCl<br>30 mM imidazole<br>pH adjusted to 7.5 with HCl at 4 °C |
| $B_{His}$ | 20 mM Tris<br>5 % Glycerol (v/v)<br>250 mM NaCl<br>500 mM imidazole<br>pH adjusted to 7.5 with HCl at 4 °C |

| | |
|---|---|
| Dialysis buffer | 20 mM Tris<br>5 % Glycerol (v/v)<br>250 mM NaCl<br>pH adjusted to 7.5 with HCl at 4 °C |
| Gel filtration buffer | 20 mM HEPES<br>150 mM NaCl<br>pH adjusted to 7.5 with NaOH |

All buffers for protein purification were filtered through 0.2 µm membrane filters and kept at 4 °C; those used for FPLC were degassed in addition.

## 2.3.2. Purification of BKV, JCV, KIV and WUV VP1

### 2.3.2.1. Lysis and nickel affinity purification

After overnight protein expression (2.2.11), bacteria were harvested by centrifugation at 4,000 g at 4 °C for 10 min. Bacteral pellets obtained from 2 L of culture were resuspended in 20 mL buffer $A_{His}$. PMSF was added to a final concentration of 1 mM. Cells were lysed by sonication with four runs of a program that repeated bursts of 0.5 s with 30 % duty cycle and 0.5 s pauses for 1 min. During sonication, the cell suspension was cooled in an ice-water bath, and the suspension was allowed to cool for 10 min between runs of the sonication program. After sonication, insoluble cell debris was pelleted by centrifugation at 18,000 g for 30 – 45 min. The supernatant was filtered successively through a 5 µm and a 0.45 µm membrane filter.

The cleared lysate was then loaded onto a HisTrap nickel affinity column that had been equilibrated with buffer $A_{His}$. Then, the column was washed with buffer $A_{His}$ until the absorbance at 280 nm had returned to the level observed before loading the column. The column was eluted with a linear gradient from 0 % buffer $B_{His}$ to 100 % buffer $B_{His}$ (corresponding to an imidazole gradient from 30 mM to 500 mM) over 40 mL or 100 mL for 1 mL and 5 mL columns, respectively. Elution fractions were analyzed with SDS-PAGE (2.3.5), and those containing VP1 in sufficient purity were pooled. The protein concentration was determined using the Bradford method (2.3.4). If the concentration was higher than 1 mg/mL, the concentration was adjusted to 1 mg/mL with dialysis buffer. Then, the pooled fractions were dialysed in a dialysis tube with a MWCO of 6 – 8 kDa against 2 L of dialysis buffer overnight to remove imidazole. To produce His-tagged pentamers for glycan array screening or NMR spectroscopy, the protein was further purified by gel filtration (2.3.2.3).

### 2.3.2.2. Thrombin cleavage and polishing

To remove the His-tag for crystallization, 10 U of thrombin were added to the dialysed protein solution for each mg of protein. If the protein concentration was lower than 1 mg/mL, 30 – 50 U of thrombin were used per mg of protein. The solution was then incubated at 20 °C for 24 – 72 h.

Since cleavage of all five His-tags of a pentamer was essential for homogeneity of the crystallization sample, the protein solution after cleavage was loaded onto a nickel affinity column again after filtration through a 0.45 µm filter. During loading and washing of the column with buffer $A_{His}$, the absorbance of the flowthrough at 280 nm was monitored and protein-containing fractions were collected and pooled. Protein in these fractions was homogeneously cleaved.

#### 2.3.2.3. Gel filtration

To prevent formation of non-natural disulfide-bonded dimers of pentamers, DTT was added to the protein solution to a final concentration of 20 mM. The solution was concentrated with Amicon Ultra spin concentrators (MWCO 50 kDa) to a concentration of 4 – 8 mg/mL, and filtered through 0.2 µm spin filters. 2.5 mL of this solution were loaded onto a Superdex 200 HiLoad column that had been equilibrated with gel filtration buffer. The protein was eluted with 130 mL gel filtration buffer and its molecular weight was estimated by comparison with a standard curve of proteins with known molecular weight. Gel filtration fractions that corresponded to VP1 eluting as a pentamer were pooled and frozen in aliquots at -30 °C. To have the opportunity to assay the protein both in the presence and absence of DTT, a fraction of the protein was supplemented with 20 mM DTT before freezing.

### 2.3.3. Purification of SV40 VP1

SV40 VP1 was purified essentially in the same manner as the other VP1 proteins. However, since SV40 VP1 was stable in the presence of 300 mM imidazole, the eluate of the first nickel affinity column was not dialysed. Instead, thrombin (10 U/mg of SV40 VP1) was added directly to the elution fractions because for this protein, the thrombin digest was more efficient in the presence of imidazole. The solution was then incubated at 20 °C for 24 h. Since cleavage in solution was complete after that time, no second nickel affinity chromatography was performed for this protein. Concentration and gel filtration were performed as for the other proteins.

### 2.3.4. Determination of protein concentration

The concentration of protein solutions were determined using UV absorbance, if the buffer of the solution was known and could be used as a blank, and a Bradford assay (Bradford, 1976), if the buffer composition was not known such as in fractions of nickel affinity chromatography gradient elutions.

For the Bradford assay, 1 – 10 µL protein were mixed with water and Bradford reagent according to the manufacturer's instructions and incubated for exactly 10 min at room temperature. The absorbance at 595 nm was determined against a blank containing water

instead of protein. The protein concentration was calculated from the protein sample volume V with the following equation that was derived from a standard curve determined by Dr. Pierre Schelling (Eberhard Karls Universität Tübingen):

$$c[mg/ml] = \frac{A_{595nm} - 0.004}{0.051 \cdot V[\mu l]}$$

**Eq. 2: Bradford assay**

To determine the protein concentration by UV absorption, the absorbance at 280 nm was measured against a blank of puffer. The molar concentration was calculated from the molar extinction coefficients ε of the proteins calculated with ProtParam (Gasteiger et al., 2005) using the following equation, with d being the path length of the UV beam through the sample:

$$c[\mu M] = 10^6 \cdot \frac{A_{280nm}}{\varepsilon \cdot d}$$

**Eq. 3: UV absorption by proteins**

### 2.3.5. SDS-PAGE

Protein samples taken during purification steps were analyzed by discontinuous SDS-PAGE (Laemmli, 1970), which separates proteins according to their molecular weight.

Four minigels (thickness 0.75 mm) were prepared simultaneously from the following solutions:

| | 12 % separating gel | 4 % stacking gel |
|---|---|---|
| $H_2O$ | 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/Bisacrylamide | 6.0 mL | 1.3 mL |
| 10 % APS (freshly prepared) | 150 µL | 100 µL |

Before electrophoresis, samples were mixed with 4x sample loading buffer, incubated at 95 °C for 2 min and then centrifuged for 2 min at 14,000 g. They were then loaded into the gels. Electrophoresis was carried out at 180 V for 1 h.

| | |
|---|---|
| 4x sample loading buffer: | 20 mL glycerol<br>20 mL 1 M Tris/HCl pH 6.8<br>10 mL 10 % SDS<br>1.63 mL 0.5 M EDTA pH 8.0<br>4 mL mercaptoethanol<br>20 mg bromophenol blue |

Gels were heated for 1 min in the microwave in Coomassie staining solution and then incubated for 15 – 30 min at room temperature. This treatment both fixed and stained the proteins. Gels were then destained overnight in Coomassie destaining solution.

| | |
|---|---|
| Coomassie staining solution: | 250 mL methanol<br>30 mL glacial acetic acid<br>500 mg Coomassie brilliant blue<br>$H_20$ ad 500 mL |
| Coomassie destaining solution: | 25 mL methanol<br>37.5 mL glacial acetic acid<br>$H_20$ ad 500 mL |

## 2.4. Characterization of protein-glycan interactions

### 2.4.1. Glycan microarrays

Glycan array screening of SV40 VP1 were conducted at Core H of the Consortium for Functional Glycomics (CFG) in the laboratory of Prof. David F. Smith (Emory University, Atlanta, USA) (2.4.1.1). Glycan microarray experiments of BKV and JCV as well as K68S VP1 (3.6.1) were conducted in the laboratory of Prof. Ten Feizi (Imperial College London, UK) (2.4.1.2). Glycan array screening experiments were performed by Dr. Angelina de Sa Palma (JCV) and Dr. Yan Liu (BKV and K68S) (both Imperial College London, UK).

#### 2.4.1.1. Covalently linked oligosaccharide array

The glycan array used to screen SV40 VP1 was composed of six glycoproteins and 258 sequence-defined glycans covalently linked to a chip by variable linkers in six replicates each (6.3.1). Many of the sequence-defined glycans did not contain sialic acid.

His-tagged SV40 VP1 pentamers (0.2 or 0.5 mg/ml) were assayed by Core H of the CFG on its printed array (version 2.1) in 20 mM Tris pH 7.5, 150 mM NaCl, 2 mM $CaCl_2$, 2 mM $MgCl_2$, 0.05 % Tween-20, 1 % BSA, 1 mM DTT. After washing, bound protein was detected with an AlexaFluor-conjugated anti-His-tag antibody. For each probe on the array, the highest and lowest fluorescence counts of the six replicates were discarded to reduce bias from extreme outliers. Data were analyzed statistically with Microsoft Excel.

#### 2.4.1.2. Lipid-linked glycan arrays

The microarray used for screening of JCV VP1 was composed of 87 sequence-defined lipid-linked oligosaccharide probes: 81 sialyl-terminating probes and six uncharged probes as negative controls (6.3.2). The probes were robotically printed in duplicate on nitrocellulose-coated glass slides at 2 and 5 fmol/spot using a non-contact instrument (Palma et al., 2006). The microarray used for screening of BKV and K68S VP1 (6.3.3) was composed of 28 lipid-

linked oligosaccharide probes, 8 of which did not carry sialic acid, at four different probe concentrations of 0.3, 0.8, 1.7 and 5 fmol/spot.

JCV and BKV VP1 were pre-complexed with mouse monoclonal anti-poly-histidine (Ab1) and biotinylated anti-mouse IgG antibodies (Ab2) (both from Sigma) in a ratio of 4:2:1 (by weight). In brief, the JCV VP1-His tagged protein-antibody pre-complexes were prepared by pre-incubating Ab1 with Ab2 for 15 min at ambient temperature, followed by addition of VP1 and incubation for a further 15 min on ice. The VP1-antibody complexes were diluted in 5 mM HEPES pH 7.4, 150 mM NaCl, 0.3% (v/v) Blocker Casein (Pierce), 0.3% (w/v) bovine serum albumin (Sigma), 2.5 mM DTT and 5 mM $CaCl_2$, to give a final VP1 concentration of 150 µg/ml, and overlaid onto the arrays at 20 °C for 2 h. K68S VP1 was analyzed without precomplexing, but in the same buffer. In both cases, binding was detected with Alexa Fluor-647-labelled streptavidin (Molecular Probes) and imaging (Palma et al., 2006). Data analysis was done as described (Stoll and Feizi, 2009).

### 2.4.2. Isothermal titration calorimetry

During isothermal titration calorimetry (ITC), a ligand solution is titrated into a solution containing a binding partner. The method detects binding by the heat generated or taken up by the system after each titration if binding takes place. Curve-fitting to the data allows the simultaneous determination of several parameters: stoichiometry of binding, free energy of binding (and consequently affinity), and entropy of binding (Wiseman et al., 1989). For an ideal sigmoidal curve, the concentrations of the protein and ligand solutions are crucial. The protein solution should be 10-times higher than the dissociation constant, and the ligand solution should be 20-times more concentrated than the protein solution. ITC experiments were performed by Dr. Karin Wöllner and me in the laboratory of Prof. Günther Gauglitz (Eberhard Karls Universität Tübingen).

His-tagged SV40 VP1 pentamers were prepared for the experiment by a buffer exchange to 70 mM HEPES pH 7.5, 150 mM NaCl, 5 mM DTT using gel filtration on Superdex-200 (2.3.2.3). The protein was concentrated to about 1 mM (40 mg/mL) with an Amicon Ultra spin concentrator and the exact concentration was determined by UV absorbance (2.3.4). GM1 oligosaccharide was dissolved in the same buffer, i.e. the flowthrough of the last concentration step.

ITC affinity measurements were performed with a Thermal Activity Monitor 2277 (Thermometrics) at 20 °C. 23 additions of 12 µL GM1 oligosaccharide at a concentration of 3.78 mM were made to 270 µL of VP1 solution at 0.91 mM concentration. Signals from a blank experiment of titrating 3.72 mM GM1 into buffer were also collected to establish a base line. Calorimetric signals were processed, integrated and fitted with DigiTAM (Thermometrics). Due to protein solubility and the availability of GM1 oligosaccharide, the chosen

parameters were suboptimal for curve fitting, but allowed me to discern whether binding had taken place or not.

### 2.4.3. Saturation transfer difference NMR

Saturation transfer difference NMR (STD NMR) is a method that allows mapping of binding epitopes on small molecules. The technique exploits the different properties of the large VP1 pentamers and the small carbohydrates in magnetic resonance measurements. Selective excitation of the protein in the on-resonance spectrum is followed by a period in which saturation spreads quickly through the entire protein and, if ligand is bound to the protein, is transferred to the ligand through the intermolecular Nuclear Overhauser Effect. Upon release from the protein, the small ligand maintains the magnetic imprint of the protein's binding site, which can then be detected in the non-selective off-resonance spectrum. Subtraction of on- and off-resonance spectra allows for easy visualization of the effect in the difference spectrum. STD NMR experiments were performed and analyzed by Dr. Bärbel Blaum in the laboratory of Prof. Thomas Peters (Universität zu Lübeck).

All NMR spectra were recorded using 3 mm tubes on a Bruker DRX 500 MHz spectrometer fitted with a 5 mm cryogenic probe at 283 K and processed with TOPSPIN 2.0 (Bruker). For each of the three proteins used for STD NMR (SV40 VP1 and BKV VP1 as well as K68S VP1 (3.6.1)) two NMR samples were prepared, one containing 1 mM GM1 oligosaccharide and one containing 1 mM GD3 oligosaccharide. Protein concentrations were between 19 µM and 22 µM as judged by the absorbance at 280 nm. Additional samples were prepared that contained no protein but 1 mM GM1 or GD3 oligosaccharide. The pure oligosaccharide samples were used to verify that no direct excitation of ligand resonances occurred during STD NMR measurements in the absence of protein, and they served as samples for the spectral assignment. 0.1 mM trimethylsilyl propionate was then added to the GD3 sample to allow $^{1}H$ referencing. The buffer used for all NMR measurements contained 20 mM deutero-Tris pH 7.5, 150 mM NaCl, and 20 mM deutero-DTT. Samples were prepared in $D_2O$ and no additional water suppression was used in order not to affect the anomeric proton signals. The off- and on-resonance frequencies were set to 80 ppm and 7 ppm, respectively. The total relaxation delay was 4 s. A cascade of 40 Gaussian-shaped pulses with 50 ms duration each, corresponding to a strength of 65 Hz, and a saturation time of 2 s was used for selective excitation. A 10 ms continuous-wave spin lock filter with a strength of 3.7 kHz was employed in order to suppress residual protein signals. 32 k points were collected and zero filling to 64 k data points was employed. Spectra were multiplied with an exponential line broadening factor of 1 Hz prior to Fourier transformation.

For assignment of the oligosaccharide proton resonances, series of 1D $^{1}H$-TOCSY and COSY spectra as well as $^{1}H$, $^{13}C$-HSQC spectra were acquired. Literature values on related

oligosaccharides served as additional assignment controls (Brisson et al., 1992; Haselhorst et al., 2007; Houliston et al., 2009; Michon et al., 1987). Assignment of the N-acetyl methyl groups was taken from (Michon et al., 1987) for GD3 and (Houliston et al., 2009) for GM1.

### 2.4.4. Cell-based oligosaccharide competition assay

The cell-based oligosaccharide competition assays were performed by Dr. Melissa Maginnis in the laboratory of Prof. Walter J. Atwood (Brown University, USA) using SVG-A cells, which are a subclone of the human glial cell line SVG transformed with an origin-defective SV40 mutant (Major et al., 1985). Cells were grown in Minimum Essential Medium (MEM) (Gibco) supplemented to contain 10 % fetal calf serum (FCS) (Atlanta Biologicals) and 1 % Penicillin/Streptomycin (P/S) in a humidified incubator at 37 °C. The stock solution of P/S (Cellgro) contained 10,000 U/mL penicillin and 10,000 µg/mL streptomycin.

JCV of the Mad-1/SVE strain (Vacante et al., 1989) was pretreated with 5 mM of LSTb or LSTc (V Labs, Inc.), which had been diluted in sterile $H_2O$, in media containing 2 % fetal calf serum (FCS) on ice for 1 h. SVG-A cells in 96 well plates were pre-chilled at 4 °C for 30 min. JCV-oligosaccharide complexes were added to cells and incubated at 4 °C for 1 h. Cells were washed with phosphate-buffered saline (PBS) twice, and media containing 10 % FCS, 1% P/S, and 1% amphotericin B was added, and cells were incubated at 37 °C for 72 h.

Then, cells were washed in PBS, fixed in cold methanol and incubated at -20 °C. Cells were washed in PBS, permeabilized with 0.5 % TX-100 (USB Corporation) at room temperature for 15 min and blocked with 10 % goat serum (MP Biomedicals) in PBS at RT for 30 min. They were then incubated at 37 °C for 1 h with VP1-specific antibody MAB597 (Atwood et al., 1995), diluted 1:10 in PBS. Samples were washed with PBS, incubated with a goat-anti-mouse Alexa Fluor 488 conjugated antibody (1:1000) in PBS at 37 °C for 1 h and washed again with PBS. Cells were then mounted on slides using VectaShield with Dapi (Vector Laboratories Inc.). Cells were analyzed for nuclear VP1 staining under a 20 X objective using an Eclipse 800 or Eclipse TE2000-U microscope (Nikon) equipped with an ORCA-ER digital camera (Hamamatsu).

### 2.4.5. Cell-based ganglioside supplementation assay

Ganglioside supplementation assays were performed by Stacy-ann A. Allen in the laboratory of Prof. Walter J. Atwood (Brown University, USA). Vero cells (ATCC, Manassas, VA) were maintained at 37 °C in MEM (Gibco) supplemented with 5 % heat inactivated FCS (Atlanta Biologicals) and 1 % P/S (Cellgro).

Vero cells were plated in 24-well dishes at 7.60 x $10^4$ cells per well. Cells were pre-incubated with media, DMSO or gangliosides GM1, GD2, GD3, GD1b, and GT1b at 30, 3 or 0.3 µM for 17 h at 37 °C. Prior to infection, cells were chilled for 20 min at 4 °C and washed

with 2% MEM. Cells were infected with 7.7 x $10^5$ fluorescent forming units of BKV for 1 h at 37 °C. The infectious media was then removed and replaced with fresh growth media. Infection was scored 48 h or 72 h post infection by staining for large-T or VP1, respectively, and analyzed by indirect immunoflourescence.

To detect expression of large-T or VP1, Vero cells were fixed with 2 % para-formaldehyde in PBS for 20 min at 25 °C and permeabilized with 1% Triton-X 100 in PBS for 15 min at 37 °C. Cells were incubated with the primary mouse monoclonal antibody PAb597 (1:10), which recognizes SV40 VP1 and cross-reacts with both JCV and BKV VP1, to stain for VP1 (Dugan et al., 2007). The mouse monoclonal antibody PAb416 (Calbiochem), which recognizes SV40 large-T and cross-reacts with JCV and BKV large-T (Eash and Atwood, 2005), was used at a concentration of 0.008 µg/µL to stain for BKV large-T. After incubation with the primary antibody, cells were washed with PBS and incubated with Alexa Fluor 488-labeled goat anti-mouse antibody (Invitrogen). Primary and secondary antibodies were diluted in PBS. The nucleus was counterstained with Dapi (Sigma).

## 2.5. X-ray crystallography

General information in this chapter is based on the textbooks “Biomolecular Crystallography” by Bernhard Rupp, “Outline of Crystallography for Biologists” by David Blow and “Principles of Protein X-ray Crystallography” by Jan Drenth (Blow, 2002; Drenth, 2007; Rupp, 2010).

### 2.5.1. Protein crystallization

In crystals, proteins are arrayed in a way that a small assembly, termed the unit cell, is translated over and over again in three dimensions to build up the crystal lattice. This arrangement is indispensable for X-ray crystallography because in this way, the unit cells all have the same scattering properties, which amplifies the scattering signal.

The geometry of the unit cell is described by the lengths of its three axes a, b and c as well as the angles $\alpha$, $\beta$ and $\gamma$ between the axes. Depending on whether certain geometric requirements, such as $\alpha = \beta = \gamma = 90°$, are fulfilled, crystal lattices can be classified into seven different crystal systems: triclinic, monoclinic, orthorhombic, tetragonal, trigonal/hexagonal, rhombohedric and cubic. These can be further discriminated according to the centering of the lattice, which can be primitive (P, no centering), face-centered (F), body-centered (I) or C-centered (C). Combination of the crystal systems with the possible centering types for each crystal system gives results in the 14 Bravais lattices.

The unit cell can contain one or several macromolecules. Often, the components of the unit cell are arranged in a symmetrical way and are linked by symmetry elements such as rotations or screw axes. Since proteins are composed of chiral amino acids and have chiral

secondary structure elements, symmetry elements such as inversion centers or mirror planes, which would generate mirror images of the protein, are not found in protein crystals. If the operators describing the internal symmetry of the unit cell apply to the entire crystal and not only certain parts of the unit cell, these operators are called crystallographic symmetry operators. They divide the unit cell into smaller partitions, termed asymmetric units. For structure determination, it is only necessary to solve the structure of the asymmetric unit since the contents of the unit cell can be generated using the crystallographic symmetry operators. The complete crystallographic symmetry of a crystal is described in the so-called space groups, which contain information about crystal system, lattice centering and the crystallographic symmetry operators of the unit cell. However, an asymmetric unit can contain additional symmetry elements that apply only to molecules inside the asymmetric unit and not to the entire crystal, such as the five-fold rotation axis of the VP1 pentamer. These non-crystallographic symmetry elements do not influence the space group. The objects linked by non-crystallographic symmetry (NCS) are not identical by definition, and therefore may be symmetric only in parts so that e.g. loops might adopt different conformations.

Protein crystals differ in many aspects from crystals of smaller molecules. Because of the size of proteins, there are often large, solvent-filled channels in protein crystals, which contribute to their large solvent content of 30-70 %. In the context of a crystal, the protein surface is therefore mostly in contact with water, and most proteins retain their native conformation in the crystal.

Protein crystals are grown from supersaturated protein solutions, in which crystallization nuclei can form spontaneously. Crystals then grow by addition of more molecules to the existing crystal, which is energetically favorable. In the vapour diffusion method, a protein solution is mixed with a precipitant solution, but only to such a concentration that the protein does not yet precipitate. Then, a small drop of protein-precipitant mixture is placed into a sealed compartment together with a big reservoir of precipitant solution. Due to the different vapour pressures above the less concentrated drop and the higher concentrated reservoir, more water will evaporate from the drop than from the reservoir, and the precipitant concentration of the drop will slowly rise until the drop has the same precipitant concentration as the reservoir. If the drop reaches protein supersaturation during equilibration with the reservoir, protein crystals can form.

However, crystallization conditions must be determined empirically for each protein. Therefore, a large number of different precipitants and precipitant concentrations, pH values and salt concentrations must be tested when a protein is first crystallized. Common crystallization screens coarsely sample a large area of crystallization space (i.e. possible crystallization conditions) in a sparse matrix. After initial crystals are obtained, crystal size and quality often can be improved by varying protein, precipitant and salt concentrations, the

pH of the solution, temperature, or by using additives. In addition, the optimal precipitant concentration for nucleation is often higher than that for crystal growth. Seeding techniques bypass the nucleation step by introducing crystallization nuclei generated from earlier obtained crystals into the crystallization drop.

In this thesis, both sitting and hanging drop vapour diffusion experiments were used for protein crystallization. Sitting drop vapour diffusion was mainly used to screen for initial crystallization conditions because of the low protein and buffer consumption. Using a Tecan Freedom Evo robot, 300 nL of pure protein solution (6-12 mg/mL) that had been filtered through a 0.2 µm spin filter were mixed with 300 nL of precipitant solution on the table of a 96-well crystallization plate. The reservoir contained 100 µL of precipitant solution, and the plate as a whole was sealed with foil, generating 96 sealed compartments. Hanging drop vapour diffusion was mainly used to refine initial crystallization conditions as this setup can generate bigger crystals due to the larger drop size and allows for seeding techniques. Here, 1 µL of protein solution was mixed with 1 µL of precipitant solution on a siliconized glass cover slide, which was then placed upside down over a reservoir of 500 µL precipitant solution in one well of a 24-well plate. The rims of the well were sealed with grease. Crystallization plates were inspected under a microscope about 1, 2, 4, and 7 days after setting up the drops, then weekly for the first month, and then once every one or two months.

Once crystals were obtained, they had to be harvested, cryoprotected and frozen for data collection. Since a crystallization drop of 0.3 – 1 µl quickly evaporates, crystals were often harvested with a glass capillary (0.3 – 1 mm, depending on crystal size) into a larger volume of solution for handling and soaking, often 20 µL in a glass depression well. The harvesting solution was identical to the reservoir solution for WUV and KIV VP1 crystals. However, BKV, SV40 and JCV VP1 crystals formed and started to disintegrate during equilibration of the drop. Therefore, the harvesting solution contained less precipitant than the reservoir solution for these crystals (3.2.3, 3.3.3, 3.4.1.2).

For oligosaccharide complex formation by soaking, crystals were transferred with a glass capillary into harvesting solution containing oligosaccharide, and incubated for 1 – 30 min in 10 – 20 µL soaking solution in a covered depression well. Another method for soaking crystals was to transfer them with a cryoloop (0.05 – 1 mm, depending on crystal size) into a 2 µL drop of soaking solution on a cover slide and to soak them above a reservoir of harvesting solution. Oligosaccharide diffused into the crystal through solvent tunnels and bound to the binding sites on VP1 that were not occluded by crystal contacts. Since oligosaccharide binding can distort or even destroy the crystal lattice, leading to loss of diffraction, different soaking times and oligosaccharide concentrations were tested for each type of crystal. In the end, the highest oligosaccharide concentration and longest soaking time that still gave the same resolution as an unliganded crystal were used.

Protein crystals are often cryocooled to -180 °C during data collection to prevent the diffusion of radicals, which are generated when a crystal is illuminated with X-rays. This greatly prolongs the time a crystal retains its diffraction properties in the X-ray beam and enables the collection of complete datasets from a single crystal. However, the formation of ice crystals during crystal freezing must be prevented because (a) growth of ice crystals in the solvent tunnels of the protein crystal can damage the crystal, and (b) diffraction of X-rays by ice crystals overlaps with protein diffraction and renders parts of the diffraction image unusable. Therefore, crystals were transferred into a harvesting or soaking solution supplemented with a cryoprotectant (e.g. glycerol) either with a glass capillary or a fiber loop. Crystals were soaked for 5 – 10 s in the cryoprotectant solution before being scooped up with a fiber loop and flash-frozen in liquid nitrogen or in the cryostream.

### 2.5.2. X-ray diffraction and data collection

When X-ray radiation hits a crystal, it is scattered by the electrons in the crystal. Diffraction is only observed at specific angles, for which the scattered beams from all the unit cells in the crystal are in phase and interfere constructively. Otherwise, the scattered waves cancel each other out. Sir W. Lawrence Bragg noted that diffraction could also be described as “reflection” of the incoming beam on parallel lattice planes in the crystal, which act like semitransparent mirrors (Bragg, 1913). A reflection can only be recorded if the path difference between two beams hitting crystal lattice planes spaced by the distance d under the angle $\theta$ is a multiple of the utilized wavelength.

$$n\lambda = d \cdot \sin\theta$$

**Eq. 4: Bragg’s law**

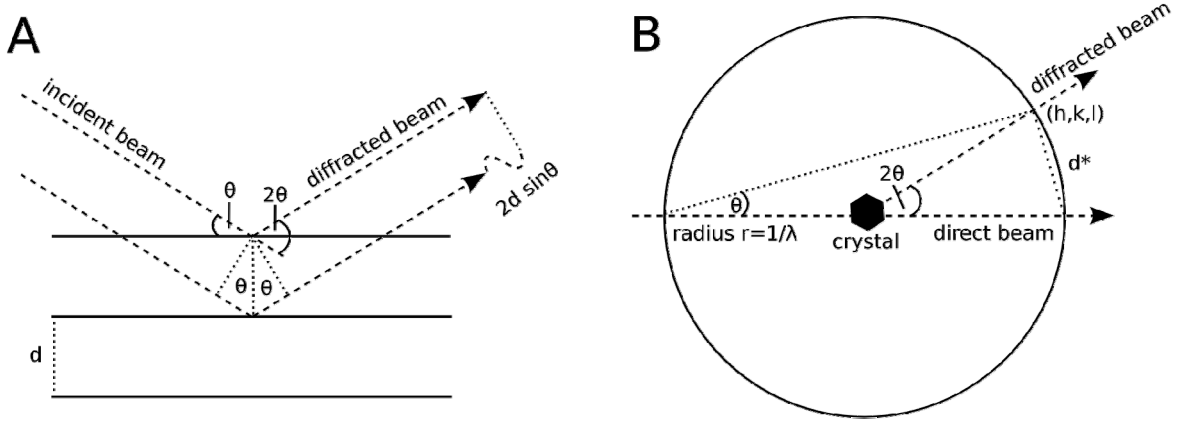



**Fig. 7: Diffraction by a crystal described by Bragg’s law and the Ewald sphere of reflection.**
(A) Graphical representation of Bragg’s law. The path difference of two beams diffracted at parallel lattice planes must be a multiple of the wavelength for constructive interference. d: distance between two lattice planes; $\theta$: angle of incidence.
(B) Ewald sphere of reflection. (h, k, l): coordinates of the reflection in reciprocal space; d*: 1/d; Radius of the sphere: 1/$\lambda$.

Thus, each reflection, or diffraction spot, is indicative of one set of lattice planes. Each set of lattice planes is defined by the three indices h, k and l describing how many lattice planes of the set intersect with the three axes of the unit cell. The resulting reflection is also described by these three indices. The different reflections generated by diffraction at one crystal are also arranged in a lattice, termed the reciprocal lattice, because there is an inverse relationship between the crystal and the reflections (Ewald, 1917). At a given orientation of the crystal, only those reciprocal lattice points (h, k, l) can be observed that lie on the surface of a sphere around the crystal with the radius of $1/\lambda$. This so-called Ewald sphere is another way to visualize Bragg's law (Ewald, 1917). As only a fraction of possible reflections can be recorded at any given orientation of the crystal, the crystal (and with it the reciprocal lattice) is rotated and illuminated many times to record a complete dataset. Internal crystallographic symmetry of the unit cell leads to symmetry within the reciprocal lattice. If crystallographic symmetry is present, only a subset of the entire reciprocal lattice (which corresponds to 360° rotation) needs to be recorded for a complete dataset.

While the direction in which a reflection (h, k, l) can be observed depends on the geometry of the crystal and the diffraction experiment, its intensity and phase depend on the distribution of scattering matter, i.e. electrons, in the unit cell. Each scattered X-ray beam (h, k, l) can be described as a structure factor $F_{hkl}$, which contains both its amplitude $|F_{hkl}|$ and phase $\phi_{hkl}$. $F_{hkl}$ is a complex number and can be represented as a vector in the Gaussian plane.

$$F_{hkl} = |F_{hkl}| \cdot (\cos\varphi_{hkl} + i\sin\varphi_{hkl}) = |F_{hkl}| \exp(i\varphi_{hkl})$$

**Eq. 5: The structure factor as a complex number**

A mathematical operation, termed Fourier transformation, links amplitude and phase of the scattered beams with the scattering contents of the unit cell. If the position of each atom n in the three dimensions x, y and z and its atomic scattering factor $f_n$ (which depends on the electron distribution of the atom) are known, the amplitude and phase of the scattered beam can be calculated for each reflection.

$$F_{hkl} = \sum_{n=1}^{N} f_n \cdot \exp\left[2\pi i\left(hx_n + ky_n + lz_n\right)\right]$$

**Eq. 6: Calculation of structure factors from atomic coordinates by Fourier transformation**

Thus, each reflection contains information about all the atoms in the entire unit cell. In turn, if amplitudes and phases of the scattered X-ray beams are known, the electron density $\rho(x, y, z)$ at each point of the unit cell can also be calculated with a Fourier transformation.

$$\rho(x,y,z) = \frac{1}{V}\sum_{hkl} F_{hkl} \cdot \exp\left[-2\pi i\left(hx+ky+lz\right)\right] = \frac{1}{V}\sum_{hkl} \left|F_{hkl}\right| \cdot \exp\left[-2\pi i\left(hx+ky+lz\right)+i\varphi_{hkl}\right]$$

**Eq. 7: Calculation of electron density from structure factors by Fourier transformation**

Fairly accurate values for the structure factor amplitudes $|F_{hkl}|$ can be determined from the diffraction data as $|F_{hkl}|$ is proportional to the square root of the intensity of a reflection $I_{hkl}$. However, no currently available X-ray sources produce coherent X-rays with a stable phase relationship, and no current detectors can determine the phase of an incoming X-ray beam (see below). Thus, phases cannot be determined directly from the diffraction experiment, which is called the phase problem. Therefore, phases need to be determined indirectly, which is the most important part of structure determination (2.5.4).

Diffraction data for this thesis was collected using either a MicroMax007HF rotating anode X-ray source equipped with a MAR345 image plate detector, or using a synchrotron protein crystallography beamline equipped with either a CCD (charge-coupled device) or PILATUS (Pixel Apparatus for the SLS) detector.

In a rotating anode generator, electrons are accelerated from a cathode using high voltage (40 kV) and hit a copper anode. The anode rotates and is cooled to protect the material from damage because most of the electrons are decelerated by the anode, generating heat and radiation. However, some electrons hitting the anode knock out electrons from the lower electron shells of the copper atoms. Then, electrons of higher electron shells can descend into the holes generated by the electrons, emitting the energy difference between the two shells as radiation. The radiation generated when an L-shell electron descends to the K-shell is called K$\alpha$ radiation, has a wavelength of 1.541789 Å, and is used for data collection. K$\beta$ radiation, generated by an M-shell electron falling down to the K-shell, is filtered out using a nickel filter. Synchrotrons are devices for circulating charged particles such as electrons or positrons at a velocity close to the speed of light. As they are constantly accelerated to keep them on their circular trajectory, the electrons emit a spectrum of intense radiation in the tangential direction according to Maxwell's laws (Maxwell, 1865). Compared to a rotating anode as X-ray source, a synchrotron has two advantages: first, the brighter and more focused beam allows for data collection to higher resolution or enables data collection of weakly diffracting crystals that did not diffract at a rotating anode source, and second, the X-ray beam at a synchrotron is tunable to different wavelengths, e.g. for collection of anomalous dispersion data.

Image plate detectors are phosphorescent screens that store the energy of the incoming X-ray photon as electronic excitation of the detector material. After exposure of the crystal with X-rays, the image plate is scanned with a red laser beam, which causes induced

emission of blue light from all pixels that have stored energy. This signal can be recorded by a photomultiplier and is proportional to the incoming X-ray intensity. CCD detectors are made of the semiconducting material silicon, in which electrons can be either in the valence band (ground state) or the conducting band (excited state). During exposure, X-ray radiation causes electrons to be excited into the valence band, leaving a positively charged hole behind. Due to a voltage applied locally for each pixel, the system cannot return to the ground state. For readout, electron-hole pairs are moved from pixel to pixel towards one corner of the CCD by sequential application of perpendicular voltages. There, the charges are amplified and recorded. CCD detectors feature lower noise as well as faster response to X-rays and faster readout times than image plates. The PILATUS detector is a very large array of single-photon counting pixels that transmit signals for each photon in real time during the measurement. Therefore, the crystal can be continuously exposed without waiting for detector readout. In addition, the PILATUS detector has a significantly higher signal-to-noise ratio and greater special resolution on the detector area than CCD detectors.

### 2.5.3. Data processing

A dataset is collected as a series of diffraction images, in which the distances and angles between the reflections are determined by unit cell parameters and experimental setup and the spot intensities depend on the content of the unit cell. Data processing procedures extract information about crystal geometry and yield a list of numerical values of reflection intensities. Both the HKL2000 and XDS suites were used for data reduction in this thesis, and even though they use different algorithms for different tasks, the overall procedures are very similar (Kabsch, 1993; Otwinowski and Minor, 1997).

First, algorithms search for reflections in a subset of images by looking for coherent groups of pixels with higher intensity than the surroundings. Once enough spot positions are known, this information, together with detector distance and orientation as well as the X-ray wavelength allows calculation of possible unit cell parameters for each lattice type as well as the orientation of the unit cell in the experimental setup. The likely lattice type is chosen by comparing how well the experimental data fulfil the geometrical requirements of the lattice type, e.g. 90° angles for orthorhombic lattices. In turn, this process assigns the reflections their indices h, k and l and is thus called indexing.

While only a subset of images is used for the indexing procedure for time reasons, the integration step determines the intensity for each reflection in the entire dataset, and often concurrently refines unit cell and data collection parameters. However, the same reflection may have been recorded at different values during data collection because of radiation damage to the crystal, intrinsic anisotropy of the crystal, different path lengths of x-rays through the crystal at different positions or intensity fluctuations of the x-ray beam. Scaling

procedures correct for many of these effects by applying scaling factors to data from different images so that the same reflection has a similar value all over the dataset. Moreover, symmetry-related reflections that can be recorded if a crystal features crystallographic symmetry (2.5.1, 2.5.2) are merged. In addition, partial reflections that were not fully recorded on a single diffraction image are added either during integration (XDS) or scaling (HKL2000). A final postrefinement step fits unit cell and data collection parameters using all data in the set, but also takes into account correction factors for radiation damage, X-ray absorption, crystal disorder and anisotropy. The output contains the most accurately determined values for unit cell and data collection parameters as well as statistics to judge the quality of the dataset.

The linear merging R-factor $R_{merge}$ quantifies the overall fraction of experimental error in the intensity readings, and is calculated as follows:

$$R_{merge} = \frac{\sum_{hkl}\sum_{i=1}^{n}\left|I_{hkl,i} - \bar{I}_{hkl}\right|}{\sum_{hkl}\sum_{i=1}^{n} I_{hkl,i}}$$

**Eq. 8: The linear merging R-factor**

The numerator of the equation contains a sum over the deviations of each redundant reflection from the mean intensity value for that reflection, i.e. a sum over all experimental errors. The denominator is a sum over the intensities of all redundant reflections.

The R-factor increases with the redundancy of the dataset. This is counterintuitive because the mean intensity of a reflection should be determined with higher precision, i.e. with a lower R-value, with increasing redundancy. The redundancy-independent R-factor $R_{meas}$ corrects for the inadvertent rise of the R-factor with redundancy (Diederichs and Karplus, 1997):

$$R_{meas} = \frac{\sum_{hkl}\sqrt{\frac{n}{n-1}}\sum_{i=1}^{n}\left|I_{hkl,i} - \bar{I}_{hkl}\right|}{\sum_{hkl}\sum_{i=1}^{n} I_{hkl,i}}$$

**Eq. 9: The redundancy-independent R-factor**

Both $R_{meas}$ and $R_{merge}$ increase with higher resolution because the effect of small irregularities in the crystal increases with higher resolution. Consequently, the signal-to-noise ratio $I/\sigma I$ of the dataset, i.e. the sum over all reflection intensities divided by their standard deviations, decreases with higher resolution. Acceptable data have $R_{meas}$ below 60 % and

I/σI above 1.5 in their highest resolution shell. In addition, data should be as complete as possible; otherwise the electron density maps will not be as detailed as the resolution suggests because their level of detail increases when more structure factors are used to calculate it (Eq.7). However, significant non-crystallographic symmetry, such as present in VP1 pentamers, can compensate for low completeness to some extent.

The space group of the crystal is determined after scaling. As the space group depends on the symmetry operators inside the unit cell, it cannot be derived from the lattice type. If the merging R-factor is low in a given space group, only reflections that are in fact related by symmetry operators have been merged. If it is high (> 10 %) even at low resolution where I/σI is high, reflections have been merged that are not truly symmetry related, and a space group with fewer symmetry operators must be chosen. Screw axes are located, i.e. differentiated from rotation axes at this stage because their presence results in systematically extinct reflections on the (h, 0, 0), (0, k, 0) and (0, 0, l) axes. However, these methods cannot distinguish between enantiomorphic space groups such as $P4_12_12$ and $P4_32_12$, or they can yield false results for pseudosymmetric crystals that diffract only weakly. Therefore, the correct space group is known with certainty only after a successful structure determination.

When the mean intensity of reflections is plotted against the resolution expressed as $(\sin\theta/\lambda)^2$ in a so-called Wilson plot (Wilson, 1949), the intensity generally decreases with higher resolution. A minimum near 5 Å is caused by solvent effects, and a maximum in the range of 4 to 3.5 Å is due to the many interatomic distances in proteins that have this length. The intensity decrease becomes linear at resolutions higher than 3.5 Å and the slope of this linear region is termed the Wilson B-factor. It represents the decrease in diffraction due to thermal vibration and crystal disorder, which play greater roles at high resolution.

For calculation of electron density, the observed reflection intensities $I_{hkl}$ are converted to structure factor amplitudes $|F_{hkl}|$, which are proportional to the square roots of $I_{hkl}$. However, weak reflections might sometimes be observed with negative intensity values because the experimental error, which is in the same order of magnitude as the intensity for weak reflections, can either increase or reduce the intensity of a reflection. Rejecting all reflections with negative intensities would bias the dataset towards positive contributions of experimental error (French and Wilson, 1978). Therefore, routines bases on Bayesian error estimates assign these reflections weak positive intensity values before their square root is calculated. This step was performed with XDSCONV or the uniqueify script in CCP4 (CCP4, 1994; Kabsch, 1993).

### 2.5.4. Structure determination

The determination of a protein structure relies on the determination of the correct phase for each reflection (2.5.2). All structures in this thesis were solved by molecular

replacement. This method relies on two facts: First, structure factors (including phases) can be calculated by Fourier transforms if the positions of the scattering atoms in the asymmetric unit are known (Eq. 6). Second, initial phases derived from a structure (termed the search model) that is similar to the structure to be determined (r.m.s.d. among C$\alpha$ atoms below 1.5 – 2 Å for atomic models) are good enough to obtain an initial electron density map that allows improvement of the model, which, in turn, will lead to better phases. In order for molecular replacement to be successful, the correct orientation and position of the search model in the asymmetric unit of the target structure needs to be found, i.e. the search model must be replaced into the new unit cell (Rossmann and Blow, 1962).

The so-called Matthews coefficient $V_M$ can give an estimate on how many copies n of the molecule need to be placed into the asymmetric unit (Matthews, 1968). It is calculated as follows from the volume of the unit cell $V_{unitcell}$, the molecular weight (MW) of the protein in Da, the number of asymmetric units per unit cell $n_{asu}$ and the number of copies n of the molecule in the asymmetric unit:

$$V_M = \frac{V_{unitcell}}{MW \cdot n_{asu} \cdot n}$$

**Eq. 10: The Matthews coefficient**

Since folded proteins have a relatively uniform density, and since protein crystals typically feature a solvent content of 30 – 70 %, $V_M$ assumes values of 1.75 – 4.2 Å$^3$/Da for most proteins (Kantardjieff and Rupp, 2003), and n can be determined accordingly.

To determine the correct orientation and position of the search model in the asymmetric unit, it needs to be rotated and translated to fit the target structure. In both cases, the so-called Patterson function is used, which is the Fourier transform of $F_{hkl}^2$. $F_{hkl}^2$ stands for the product of the complex number $F_{hkl}$ with its conjugate complex number $F_{hkl}$*, which has the same amplitude as $F_{hkl}$, but a phase $\phi_{hkl}$* that corresponds to the negative value of $\phi_{hkl}$. When the two are multiplied, the product is real and corresponds to the square of the amplitude.

$$F_{hkl}^2 = F_{hkl} \cdot F_{hkl}^* = |F_{hkl}| \exp(i\varphi_{hkl}) \cdot |F_{hkl}| \exp(-i\varphi_{hkl}) = |F_{hkl}|^2 \exp(i\varphi - i\varphi) = |F_{hkl}|^2$$

**Eq. 11: Multiplication of a structure factor with its complex conjugate yields a real number**

A Patterson transformation can thus be calculated without knowing the phases. Patterson space contains all N(N-1) vectors between the N different atoms of a structure, which can also be calculated from an atomic model. The Patterson function of a protein in a crystal contains intramolecular vectors between the atoms of the same protein and cross-

vectors between different copies of the protein in the crystal. The pattern of intramolecular vectors with a length between 15 and 3.5 Å is characteristic of the relative locations of secondary structure elements and is similar for proteins with the same fold. Since it only depends on the orientation of the model, it is used for finding the correct rotation of the search model. The Patterson function of the search model in a large box is calculated to eliminate cross-vectors, and is then rotated in small increments over all angles. The so-called rotation function, which is an overlap function of the model and data Patterson functions, features a peak for the correct rotation angles. A similar overlap function, termed self-rotation function, can be calculated for the Patterson function of the data with itself, and features a peak for rotation angles that superpose molecules linked by non-crystallographic symmetry onto each other. In the case of the pentameric polyomavirus VP1 proteins, five peaks corresponding to its five-fold rotational symmetry were found for each pentamer. When the search model is thus aligned parallel to the target structure, the oriented search model is translated within the asymmetric unit until its cross-vectors, which are indicative of packing contacts, overlap significantly with those in the data. This is quantified by the so-called translation function, which features a peak for the correct translation vector.

In this thesis, molecular replacement was performed with the programs AMoRe and Phaser, which use different search algorithms and different methods for scoring the peaks of their rotation and translation functions (McCoy et al., 2007; Navaza, 1994). AMoRe back-calculates structure factor amplitudes for the search model in different orientations, and then computes a correlation coefficient and an R-factor between the observed and back-calculated structure factor amplitudes. The AMoRe output thus contains a list of the best rotation and translation solutions. Correct solutions are separated from incorrect solutions that arise from noise in the Patterson function because they feature both a higher correlation coefficient and a lower R-factor. Phaser estimates the quality of solutions with maximum-likelihood methods. The log likelihood gain (LLG) quantifies how much better the measured data can be predicted with the model than with a random distribution of the same atoms. Since likelihoods adopt values between 0 and 1, and their differences are very small, the logarithm of the likelihood is used because it is more sensitive to changes. For a correct solution, the LLG of the rotation solution should be higher than “noise” solutions, and it should further increase when the correct translation is determined. The Z-score is the difference between the LLG and the mean LLG for a random sample of orientations (or translations), divided by the r.m.s.d. of a random sample of LLG values from the mean. Like the LLG, the translation Z-score of a solution should be higher than its rotation Z-score. Translation Z-scores between 6 and 8 can indicate correct solutions, and translation Z-scores higher than 8 indicate a certainly correct solution. The maximum-likelihood model implemented in Phaser cannot deal with translational NCS and in that case falsely assigns

incorrect solutions high Z-scores. Therefore, the program should not be used if translational NCS has been detected with a self-rotation function (see above).

### 2.5.5. Structure refinement

Structure refinement is an iterative process, in which the structural model is improved to better fit the experimental data both by computational algorithms and by manual building into electron density. Phases are then calculated from the improved model, which in turn yield an improved electron density map (Eq. 7). The process of refinement can be likened to finding the global minimum (the correct model) on a complex energy landscape with x, y, z coordinates and atomic B-factors for each atom as variables. The search converges when alterations in the model no longer significantly improve its correlation with experimental data.

The progress of refinement is monitored by calculating crystallographic R-factors that describe how well structure factor amplitudes $|F_{calc}|$ that are calcultated from the current model correspond to the experimentally observed structure factor amplitudes $|F_{obs}|$.

$$R = \frac{\sum_{hkl} \left| |F_{obs}| - |F_{calc}| \right|}{\sum_{hkl} |F_{obs}|}$$

**Eq. 12: The crystallographic R-factor**

During refinement, special coordinates x, y, z and atomic B-factors for each atom are fitted to a limited number of experimental data points so that the fit is often only slightly overdetermined. Under these conditions, an incorrect model can be fit to the experimental data with reasonable R-factors by introducing more adjustable variables, which is called overfitting. To prevent this, a so-called “free” set of reflections is chosen to cover the entire dataset in a random manner. It usually encompasses 2 – 10 % of all reflections so that statistically significant values can be calculated even if data are partitioned into 10 resolution bins. Reflections flagged as “free” are set aside after data processing and not used for refining the model. They are then used to cross-validate the model by calculating an R-factor for them, termed $R_{free}$, which describes how well the model predicts these “free” reflections (Brunger, 1992). A difference between $R_{free}$ and $R_{work}$, which is the R-factor for the reflections used in refinement, of greater than 5 % is indicative of overfitting. It should be noted that in the VP1 datasets, which exhibit 5- or 10-fold NCS, the free R-factor is no longer completely independent because reflections in the free and working sets are linked by this symmetry.

Modern refinement programs all optimize target functions that that are based on maximum likelihood, i.e. they try to maximize the probability of observing the data given the (current) model (Afonine et al., 2005; Murshudov et al., 1997). If alterations to the model

make the observations more probable, the likelihood increases. However, different types of refinement are most useful at different stages of refinement. Rigid-body refinement moves big parts of the molecule, e.g. each monomer of a VP1 pentamer, as one entity. This greatly reduces the number of parameters to be computed, which makes it practical as a first robust refinement step after structure determination when the first model is still a relatively poor approximation of the final structure.

Refinement of individual coordinates for each atom is restrained in most cases to reduce the number of individually fitted parameters. Restraints are based on the fact that defined geometrical parameters within proteins, such as bond lengths or bond angles, should have values close to ideal stereochemistry. Thus, target functions for restrained refinement include terms describing the deviation of geometrical parameters from ideal geometry, which are minimized. If NCS is present in a model, NCS restraints limiting how much the same geometric parameter can deviate from its NCS-mate are be applied in a similar way. A special case of restrained coordinate refinement uses the simulated annealing algorithm for optimizing the target function. It corresponds to a molecular dynamics run for the protein structure, which is first given high kinetic energy and then "cooled" in a stepwise way with equilibration of the structure for each cooling step. This allows sidechains or entire loops that are trapped in a local energy minimum to adopt a more realistic conformation.

Refinement of atomic B-factors is restrained in most cases as well, based on the fact that atoms linked by a chemical bond should exhibit similar degrees of thermal motion. In addition, NCS restraints pertain to B-factors as well. In this thesis, all B-factors were modelled as isotropic B-factors, which assume that the thermal motion of each atom is identical in all directions, and require fitting of only one parameter per B-factor. The quality of B-factor refinement can be assessed by how well the average B-factor of the model corresponds to the Wilson B-factor. In addition, B-factors should be relatively low for main chain atoms and increase towards the tip of the side chain due to their greater flexibility.

Electron density maps were calculated during refinement by Fourier transformations using phases $\phi_{calc}$ calculated from the model and $2|F_{obs}|-|F_{calc}|$ as Fourier coefficients. Difference maps were obtained by using $|F_{obs}|-|F_{calc}|$ in the Fourier transformation. These were more sensitive to errors in the model and often displayed where atoms were missing or wrongly oriented. Electron density maps were displayed in Coot (Emsley and Cowtan, 2004) and used to manually improve the model.

Oligosaccharide ligands were located in electron density maps calculated with phases from the unliganded models by using $2|F_{obs,soak}|-|F_{calc,unliganded}|$ and $|F_{obs,soak}|-|F_{calc,unliganded}|$ as Fourier coefficients. They were built residue per residue into the maps using monomers from previously published structures in the pdb. They were refined by restrained coordinate and isotropic B-factor refinement in Refmac5. Restraints for carbohydrate monomers as well as

β1,4- and β1,3-glycosidic bonds were obtained from the CCP4 monomer library. Restraints for the α2,3-, α2,6- and α2,8-glycosidic bonds were user-defined using the definition of the β1,4-glycosidic bond as a starting point.

# 3. Results

## 3.1. Protein expression and purification

### 3.1.1. Construct design

Expression constructs of the different polyomavirus VP1 proteins were based on the free pentamer construct of Polyoma, which had been crystallized earlier (Stehle and Harrison, 1997). The C-terminal arm of VP1 was omitted from the construct to avoid capsid assembly in highly concentrated crystallization setups (Salunke et al., 1986), and the N-terminal extension of VP1 was left out because its presence had inhibited crystallization of the Polyoma free pentamer (Stehle and Harrison, 1997). The new VP1 expression constructs were designed to have similar boundaries as the crystallized Polyoma protein (Fig. 8).

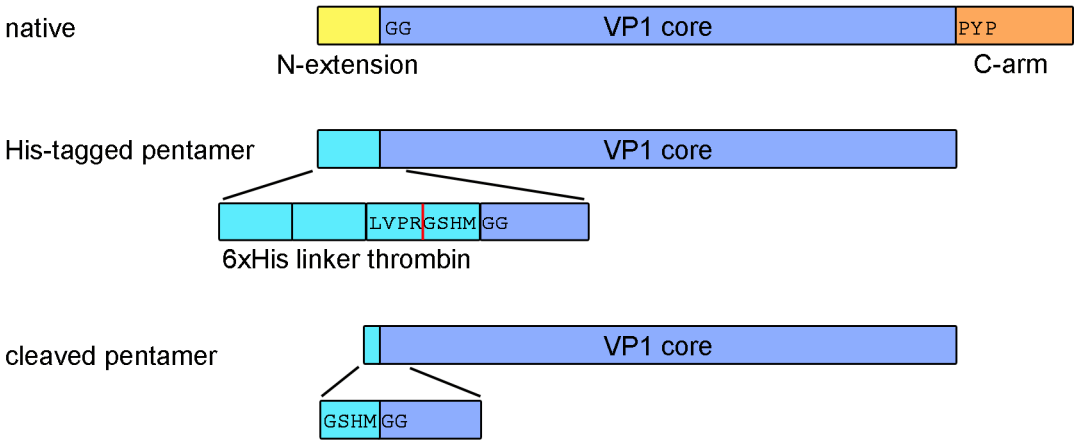



**Fig. 8: Design of expression constructs for the production of unassembled pentamers of polyomavirus VP1 proteins.**
All proteins were expressed as His-tagged pentamers. This protein was purified and used for binding assays. For crystallization, the His-tag was cleaved off with thrombin.

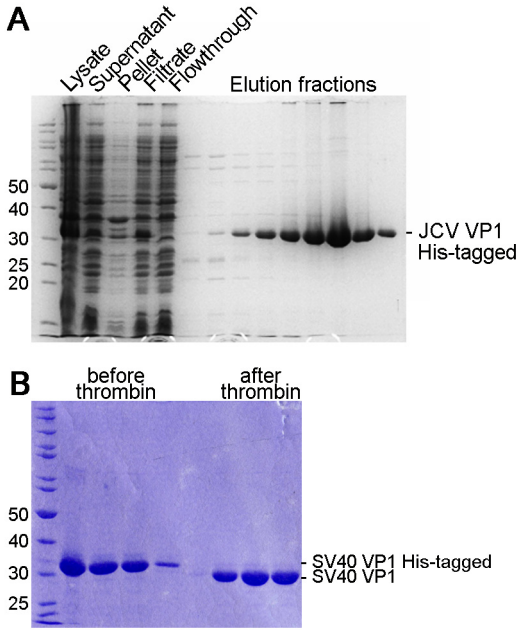
The GG and PYP sequences that were at the start and at the end of the crystallized Polyoma protein are well conserved among all polyomaviruses, and sequence alignment is unambiguous in both regions. However, the DNA coding for PYP in SV40, JCV, WUV and KIV VP1 included sequences leading to primer dimers and primer secondary structure. Since these might have caused problems during PCR cloning, the constructs were truncated right before the PYP motif. In the case of BKV VP1, the PYP sequence was included for similar reasons. For simplicity, the N- and C-terminally truncated, assembly-deficient VP1 core used in this thesis will be referred to as VP1 from here on.

The different VP1 sequences were amplified by PCR, and cloned into the pET15b expression vector in frame with an N-terminal hexahistidine tag (His-tag) and a thrombin cleavage site (2.2). Restriction sites for NdeI and BamHI were added during PCR. Because of the spacing between thrombin cleavage site and restriction sites on pET15b the non-native amino acids GSHM remained at the N-terminus of each VP1 after thrombin cleavage (Fig. 8).

### 3.1.2. Protein expression and purification

The VP1 proteins were overexpressed in *E. coli* and purified by nickel affinity chromatography and gel filtration. The purification protocol was essentially the same for all proteins. Differences in the protocols are given in the Materials and Methods section (2.3).

All VP1 proteins were overexpressed in *E. coli* strain BL21(DE3) after induction with IPTG. The bacteria were lysed by sonification, and a cleared lysate was produced by centrifugation. For all VP1 proteins, a fraction of the ooverexpressed protein partitioned to the pellet and was therefore insoluble (Fig. 9A). This cleared lysate was then filtered and loaded onto a nickel affinity column. Proteins were eluted with an imidazole gradient (40 – 500 mM). This first purification step proved to be quite effective because its five His-tags retained the VP1 pentamer on the column after most impurities were washed away (Fig. 9A). After nickel affinity chromatography, JCV, BKV, KIV and WUV VP1 were dialysed to remove imidazole. SV40 VP1 remained in the elution fractions because it was quite stable in the presence of imidazole.



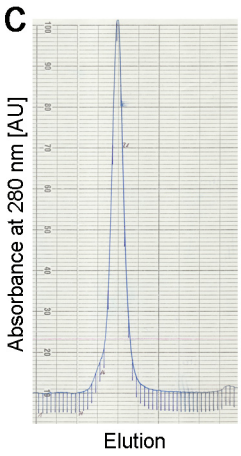




**Fig. 9: Purification of polyomavirus VP1 pentamers**
(A) SDS-PAGE of JCV VP1 solubilization and and nickel affinity purification. The filtered soluble fraction (filtrate) was applied to a nickel affinity column.Elution fractions were taken during the40 – 500 mM imidazole gradient. Molecular weight standards are labelled in kDa on the left.
(B) Cleavage of the His-tag of SV40 in solution by thrombin.
(C) Gel filtration of cleaved BKV VP1 on Superdex-200. The protein eluted at a molecular weight of 130 kDa, which corresponds best to a pentamer. The slight shoulder on the left side of the peak is caused by the formation of disulfide-bonded dimers of pentamers.

To remove the His-tag for crystallization, the proteins were cleaved in solution with thrombin after dialysis. The cleaved proteins were subjected to a second round of nickel affinity chromatography, this time to remove any uncleaved His-tagged VP1 from the flowthrough and to ensure a homogeneous sample for crystallization. The proteins were then concentrated and further purified by gel filtration on Superdex-200 (2.3.2.3). This removed some traces of impurities that were still present and served as a folding control since there is no functional assay for the proteins and since they can only be folded as pentamers.

VP1 eluted from the size-exclusion column at an apparent molecular weight of 130 kDa, corresponding to a pentamer. However, VP1 also eluted at volumes corresponding to dimers of pentamers and aggregates of higher molecular weight. The addition of 20 mM DTT greatly reduced the formation of dimers of pentamers and of higher aggregates, indicating that their formation was mediated by disulfide bonds. Interestingly, interpentamer disulfide bonds between the conserved cysteins of the CD-loop were observed in SV40 and BKV VP1 crystals that had been grown in the absence of DTT (3.2.4, 3.4.1.1). Therefore, proteins were concentrated in the presence of 20 mM DTT. Gel filtration was performed in the absence of DTT to have the option to crystallize the protein under non-reducing conditions. After gel filtration, the protein was stored in aliquots both in the presence and absence of DTT.

During thrombin cleavage of WUV VP1, lower molecular weight species of VP1 were generated because the protein was internally cleaved by thrombin after R197 at a non-canonical cleavage site with the sequence ETPR'ARVT. The protein was then further degraded over time, indicating increased mobility. It could therefore not be used for crystallization. After mutation of R197 to K (2.2.10), which is present in most VP1 proteins at this position, the His-tag could be cleaved without degradation of the protein. This construct was thus used for all following experiments. The mutation did not alter the secondary structure composition of the protein, as shown by CD spectroscopy (data not shown), and thus was unlikely to affect its three-dimensional structure. For simplicity, this WUV VP1 R197K mutant will be referred to as WUV VP1 for the remainder of the text.

## 3.2. SV40 and its receptor GM1

### 3.2.1. Glycan array screening

Although GM1 had been idenfied as a receptor for SV40 (Tsai et al., 2003), it was not clear which parts of GM1 were necessary for the interaction and whether there were other possible receptor candidates. In order to define the spectrum of carbohydrates that can be recognized by SV40 VP1, the His-tagged protein was analyzed by glycan array screening. The experiment was performed by Prof. David F. Smith (CFG, Core H, Emory University,

USA) (2.4.1.1). The array contained 6 glycoproteins and 258 synthetic, physiologically relevant carbohydrates representing glycans from glycoproteins and glycolipids as well as sulphated glycans. The synthesized oligosaccharides had defined sequences and were covalently linked to the carrier surface using different spacers (6.3.1).

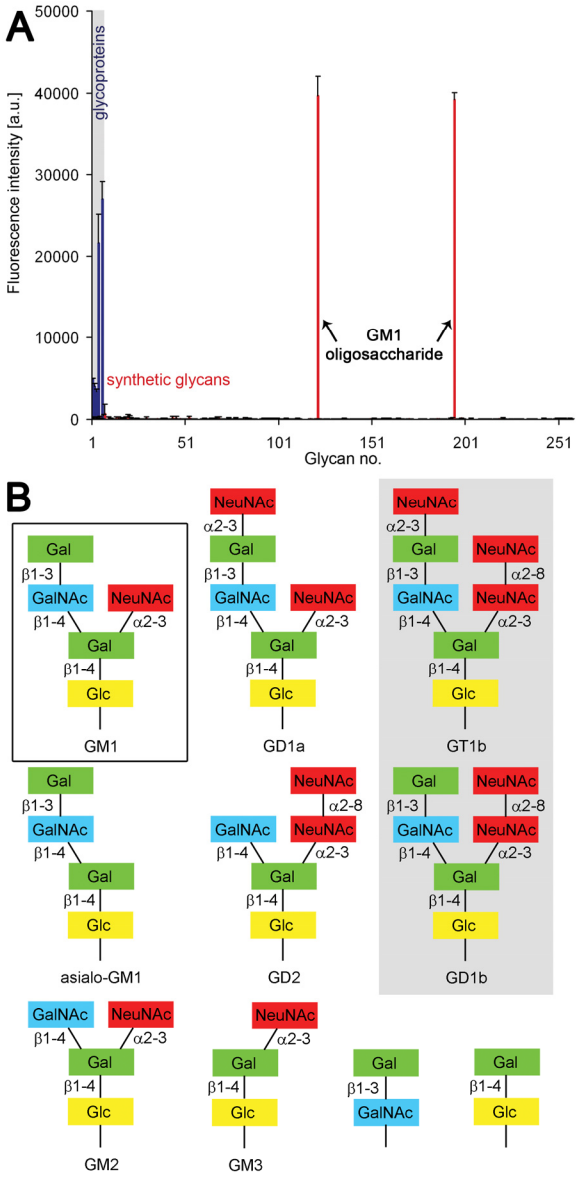
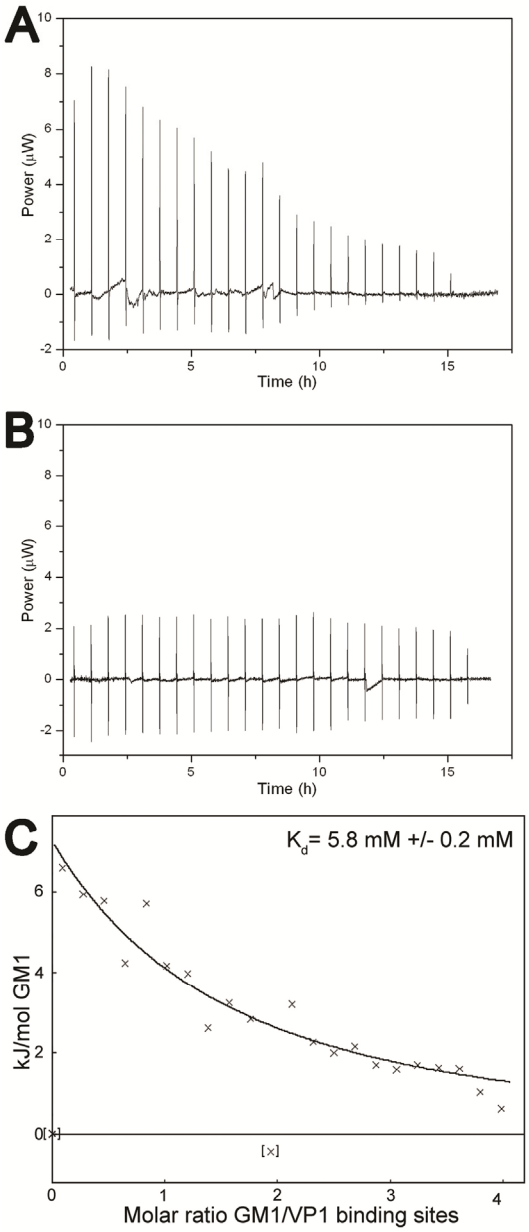



**Fig. 10: Specificity of SV40 VP1 for GM1.**
(A) Analysis of specificity by glycan array screening. Glycan array data are given as mean fluorescence signal for each glycan (see methods). Error bars correspond to the standard error of the mean of six replicates for each glycan, with the highest and lowest signals omitted from analysis to reduce bias from extreme values. Probes on the array were either glycoproteins (blue) or covalently linked synthesized carbohydrates (red).
(B) Schematic view of gangliosides and related structures present on the array. GM1 is highlighted with a black box, and selected gangliosides not included in the array are shaded grey.

Although some binding could be detected for the glycoproteins, the observed signal could not be attributed to a specific carbohydrate structure. The most likely reasons for this are heterogeneous glycosylation of glycoproteins and unspecific protein-protein interactions. The binding signal for synthetic glycans is considered less ambiguous because defined glycans are evaluated. Of the 258 carbohydrates tested, SV40 VP1 exclusively recognized the oligosaccharide portion of ganglioside GM1 (Fig. 10A), which was present twice on the array and which, for reasons of simplicity, will be referred to as GM1 throughout the remaining text. GM1 is a branched compound whose monosaccharide moieties are linked in

the following manner: Gal-β1,3-GalNAc-β1,4-[NeuNAc-α2,3]-Gal-β1,4-Glc. The array contained several GM1-related structures (Fig. 10B), but none of these interacted with SV40 VP1 in a detectable manner. Thus, the complete GM1 epitope was necessary for recognition, and did not tolerate any additions.

### 3.2.2. Isothermal titration calorimetry




Next, the affinity of one binding site of SV40 VP1 to GM1 was determined by isothermal titration calorimetry to be about 5 mM (experiment done together with Karin Wöllner, Eberhard Karls Universität Tübingen) (2.4.2) (Fig. 11). However, data from different measurements and manual integration of data yielded lower dissociation constants between 1 and 5 mM; this scattering is due to the low observed affinity. Consistent with this result, concentrations of about 5 mM were required to obtain complex by soaking VP1 crystals with GM1 (3.2.3). The low affinity of the interaction did not allow us to discern cooperativity of binding.

**Fig. 11: Affinity of SV40 VP1 for GM1.**
Data were obtained through isothermal titration calorimetry. (A) and (B) Titrations of GM1 into 0.9 mM SV40 VP1 (A) or buffer (B) at 20 °C. 23 additions of 12 µl of 3.78 mM and 3.72 mM GM1, respectively, were made.
(C) Plot of integrated calorimetric signals after buffer subtraction against ratio of GM1 to VP1 binding sites.

### 3.2.3. Crystallization and structure determination

Cleaved SV40 VP1 was concentrated to 12 mg/mL without DTT and crystallized using the hanging drop method (2.5.1). Initial crystals of a diameter of 10 µm were obtained with the Wizard I &II suite in a condition containing 100 mM Tris pH 8.5 and 24% (w/v) PEG 3,350 at 20 °C (Fig. 12A). A lower protein concentration of 10.5 – 11.5 mg/mL, crystallization at

12 °C, and using ethylene glycol as additive yielded plate-shaped crystals of 200 – 600 µm (Fig. 12B). Interestingly, in order to obtain these crystals, the drops had to be set up according to the additive screen protocol, i.e. the reservoir solution of 100 mM Tris pH 8.5, 24% (w/v) PEG 3,350 was mixed 4:1 with 30 % (v/v) ethylene glycol, and this was mixed 1:1 with the protein. Crystals were harvested into 100 mM Tris pH 8.5, 20 (w/v) % PEG 3350 and 6 % (v/v) ethylene glycol. For cryoprotection, crystals were incubated for 5-30 s in harvesting solution containing 12.5 % glycerol as a cryoprotectant, then incubated in harvesting solution containing 25 % (v/v) glycerol, and finally flash-frozen in liquid nitrogen. For complex formation, crystals were soaked for 16 min in harvesting solution containing 5 mM GM1. Cryoprotectant solutions were also supplemented with 5 mM GM1.

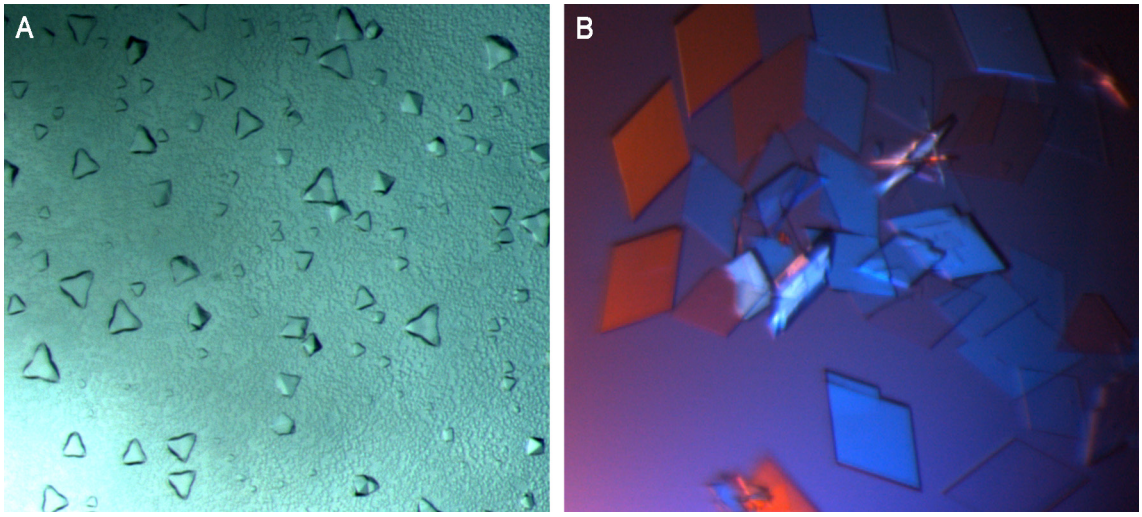



**Fig. 12: Crystallization of SV40 VP1.**
(A) Initial crystals obtained in Wizard Screen.
(B) Refined condition. These crystals were used for soaking with GM1.

Diffraction data were collected at beamline PX I equipped with a MarCCD detector at the SLS (2.5.2). Complete datasets could be obtained from one crystal in each case by translating the crystal in the beam. Data reduction was performed with HKL2000 (Otwinowski and Minor, 1997) (2.5.3).The structures were solved by molecular replacement with AMoRe (Navaza, 1994) using the VP1 core of the SV40 virion structure (1SVA) as a search model (2.5.4). They were refined with Refmac (Murshudov et al., 1997) to 2.3 and 2.25 Å resolution, respectively, for the unliganded and complexed SV40 VP1 pentamers (2.5.5, 6.2.1). The crystals had space group C2 with one complete pentamer in the asymmetric unit, and its fivefold symmetry was used as an NCS restraint throughout refinement. After initial refinement of the native structure, GM1 was located in $F_{obs,soak}$-$F_{calc,nat}$ difference Fourier maps, built using coordinates of monosaccharide fragments from the cholera toxin-GM1 complex structure (3CHB, chain 5) and refined. In the final electron density map, amino acids 43-297 are visible for all five chains.

### 3.2.4. Structure of the SV40 VP1-GM1 complex

The “unassembled” VP1 pentamer structure solved here is very similar to that of the assembled VP1 protein in the structure of the SV40 capsid. Each VP1 monomer assumes the established β-sandwich fold, with two four-stranded antiparallel β-sheets (CHEF and BIDG) forming the core structure (Fig. 13A). Residues and loops of the clockwise and counterclockwise neighbors of a given VP1 monomer are designated cw and ccw, respectively, as before (Fig. 4). The r.m.s.d. between the two proteins is 1.1 Å for the Cα atoms of residues 45 to 297, demonstrating that the recombinantly expressed protein is in agreement with the assembled VP1 protein in the virus. Structural differences map to the CD-loops because VP1 monomers in the context of the unassembled pentamer can be superposed on monomers in the context of the virion with an r.m.s.d. of 0.55 Å between Cα atoms. The CD-loops adopt different conformations in the virion, some of which are engaging in interpentamer disulfide bonds (Liddington et al., 1991; Stehle et al., 1996). In the unassembled pentamer structure, their conformations differ from the ones observed in the virion and they are influenced by crystal contacts. In the crystal, two pentamers stack against one another with their bottom surfaces, and C104 residues in the CD-loops are engaged in interpentamer disulfide bonds in each chain. The geometry of these disulfide bonds differs from those observed in virions because all five C104 residues of one pentamer are engaged to residues from a single pentamer.

The GM1 ligand binds to a shallow groove formed by the BC1-, BC2-, DE- and HI-loops, which all emanate from the β-sandwich at the outer surface of VP1. The BC2cw- and DEccw-loops of neighboring VP1 monomers (viewed from the outside of the virion) complete the binding site at each end (Fig. 13A). Three VP1 monomers bound ligand; access to the GM1 binding sites of the remaining two VP1 monomers is blocked by crystal contacts with neighboring VP1 pentamers. We solved VP1 structures with and without ligand at a similar resolution (6.2.1). Their Cα atoms can be superposed with an r.m.s.d. of 0.3 Å, indicating that no major conformational changes occur upon ligand binding.

The GM1 ligand used in this study contains five carbohydrate residues (Fig. 10B, Fig. 13B). Its overall shape resembles the letter “Y”, with the Gal-β1,4-Glc moiety forming the stem, and the NeuNAc and Gal-β1,3-GalNAc moieties forming the two branches. The oligosaccharide lacks the membrane-anchoring ceramide, which would be attached to the terminal Glc in the GM1 ganglioside. GM1 has essentially the same conformation in all three binding sites, and all five sugar moieties are well defined by electron density (Fig. 13B). The conformation of GM1 in the complex is very similar to its conformation in solution (Acquotti et al., 1990a; Brocca et al., 1998; Richardson et al., 1995), indicating that SV40 VP1 does not induce a major structural change in its ligand. The orientation and position of GM1 on the

SV40 VP1 surface differ from a previously published homology model of their interaction (Campanero-Rhodes et al., 2007).

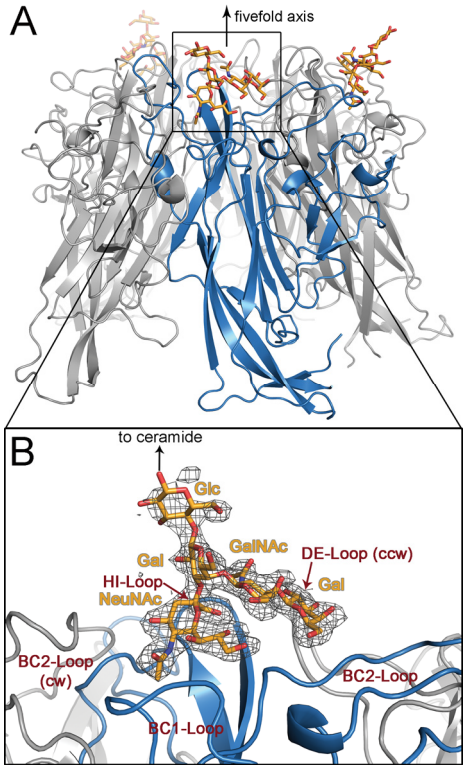



**Fig. 13: Structure of the SV40 VP1-GM1 complex.**
(A) Overvall structure. The five VP1 chains are shown as ribbon tracings, with one VP1 monomer highlighted in blue and the others colored grey. Bound GM1 is shown in stick representation, with carbons drawn in orange, oxygens in red and nitrogens in blue. The GM1 molecules bind on top of the VP1 pentamer, which lies at the outside of the capsid in the virion. Only three monomers carry oligosaccharide; the binding pockets of the two other monomers are occupied by crystal contacts.
(B) Composite annealed omit difference electron density map for GM1, calculated at 2.25 Å resolution, contoured at 2.5 σ and displayed 4 Å around GM1. One VP1 monomer is colored blue, the clockwise (cw) and counterclockwise (ccw) neighbors are colored grey.

### 3.2.4.1. Interaction of VP1 with GM1

Both branches of GM1 make extensive contacts with VP1, and they exhibit temperature factors of around 30 Å$^2$ that are in accordance with those of surrounding amino acids. The Gal-β1,4-Glc stem, which would be attached to the ceramide and anchored into the membrane in a physiological setting, faces away from the protein and does not make any contacts with VP1. The stem exhibits substantially higher temperature factors, indicating higher mobility. The NeuNAc and Gal-β1,3-GalNAc branches of GM1 resemble a bridge that leads over a ridge on the VP1 surface separating the binding sites for the NeuNAc moiety and the terminal Gal (Fig. 14A,C). The ridge is formed primarily by the fully extended side chain of K67, which is stabilized by a salt bridge with D81 at the end of the BC2-loop (Fig. 14C). The interactions bury a total area of 412 Å$^2$ from solvent, 55% of which is contributed by NeuNAc, and 45% by the Gal-β1,3-GalNAc arm. The binding site is formed mainly by one VP1 monomer, but the clockwise and counterclockwise neighbours also contribute.

The NeuNAc arm of GM1 is contacted by residues from the HI- and BC1-loops from one monomer, as well as residues from the BC2cw-loop of the clockwise neighbor. The interaction features a marked shape complementarity between all three protruding functional groups (carboxyl, *N*-acetyl, and glycerol chains) of NeuNAc and sites on VP1 that

accommodate each group (Fig. 14A). The negatively charged carboxyl group points towards the HI-loop and forms hydrogen bonds to S274 and T276 (Fig. 14B). The methyl group of the N-acetyl chain partially inserts into a deep cavity that is lined with the hydrophobic side chains of F270 and L65 as well as F75cw (Fig. 14B). The amide nitrogen of the N-acetyl chain is hydrogen bonded to N272 on the polar rim of the cavity. The glycerol side chain lies in a shallow groove on the surface. It points towards the aliphatic portion of the K67 side chain, and makes extensive polar interactions with its three hydroxyl groups (at O7, O8 and O9) and residues in the BC1- and BC2-loop. On one side, the O7 hydroxyl makes van der Waals interactions with Q62, whereas on the other side the hydroxyl group of S68 can hydrogen bond to the O8 or the O9 hydroxyl of NeuNAc. There is some ambiguity to this interaction as the S68 side chain assumes two alternative conformations. The side chain of Q278 points upward towards NeuNAc and forms the bottom of the binding pocket. It organizes a network of hydrogen bonds that lie beneath the NeuNAc moiety and orients other amino acids for binding (Fig. 14B).

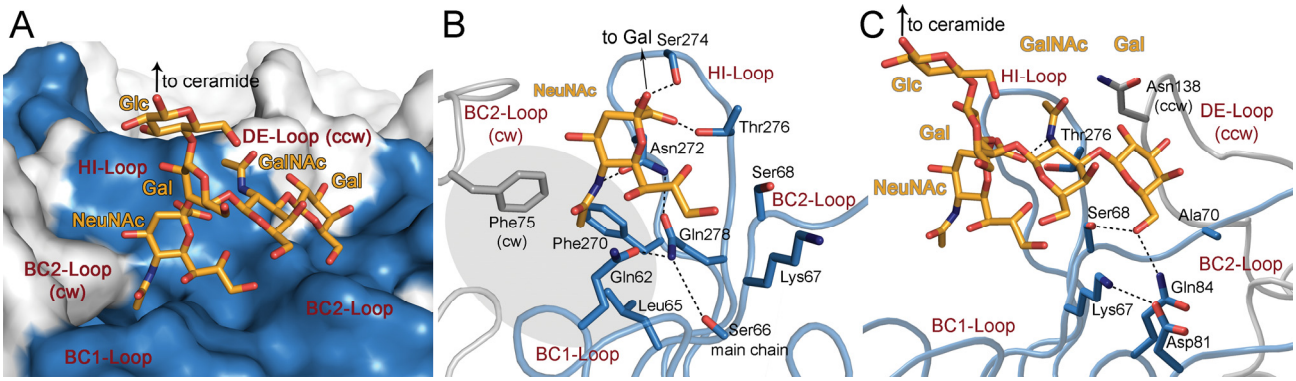



**Fig. 14: Interactions between SV40 VP1 and GM1.**
(A) Surface architecture of VP1 and shape complementarity with GM1. Protein chains are colored using the code in Fig. 2.
(B) Contacts between NeuNAc and SV40 VP1 residues.
(C) Contacts between Gal-(β1,3)-GalNAc and SV40 VP1. In panels B and C, only residues that contact the sugar are shown in stick representation. Hydrogen bonds are shown as broken lines. The cavity between one monomer of VP1 and its clockwise neighbor is shaded grey.

The Gal-β1,3-GalNAc arm of GM1 interacts with residues in the HI- and BC2-loops, and also contacts the DE-loop of the counterclockwise neighboring monomer (Fig. 14C). The terminal Gal is held in place by hydrogen bonds to Ser68 and Gln84 in the BC-loop as well as van der Waals interactions with Ala70 in the BC-Loop and Asn138 in the DE-loop of the counterclockwise neighbor. The GalNAc residue is not involved in direct hydrogen bond formation.

### 3.2.4.2. Structural basis of specificity

Glycan screening demonstrates that binding to GM1 occurs with narrow specificity. Oligosaccharides that contain similar structural motifs in different contexts (linkages,

spacings) do not interact with SV40 VP1. Analysis of the crystal structure confirms these findings: addition of carbohydrate residues to either one of the branches would lead to severe steric clashes with protein residues. Both branches of GM1 are required for interaction with VP1. NMR studies show that GM1 assumes one dominant conformation in solution (Acquotti et al., 1990a; Brocca et al., 1998; Richardson et al., 1995). The GalNAc-$\beta$1,4-[NeuNAc-$\alpha$2,3]-Gal portion of the oligosaccharide forms a rigid entity because of close contacts between the GalNAc and NeuNAc moieties, which are attached to neighboring carbons on the Gal. This arrangement severely limits the rotational freedom of all three sugar residues. By contrast, unbranched oligosaccharides are more likely to adopt different conformations. This is the case, for example, for the linear $\alpha$2,3-sialyllactose, whose conformation has been studied both in solution and complexed to different protein ligands (Breg et al., 1989b; Eisen et al., 1997; Sabesan et al., 1991; Stehle et al., 1994; Wright, 1990). Since the conformation of GM1 in the VP1 complex is very similar to its principal structure in solution, it is likely that interactions of one branch facilitate interactions of the other branch with VP1 residues, and that this bivalent binding mode is a foundation for the high specificity of SV40 for GM1.

## 3.3. JCV and its receptor LSTc

### 3.3.1. Glycan array screening

Purified His-tagged JCV VP1 pentamers were analyzed on glycan microarrays containing 81 lipid-linked sialylated oligosaccharides differing in the sialyl linkage, backbone sequences, chain lengths, and branching patterns. This experiment was performed by Dr. Angelina de Sa Palma (Imperial College London, UK). The microarrays represented carbohydrates found on *N*- and *O*-linked glycoproteins as well as glycolipids (2.4.1.2, 6.3.2).

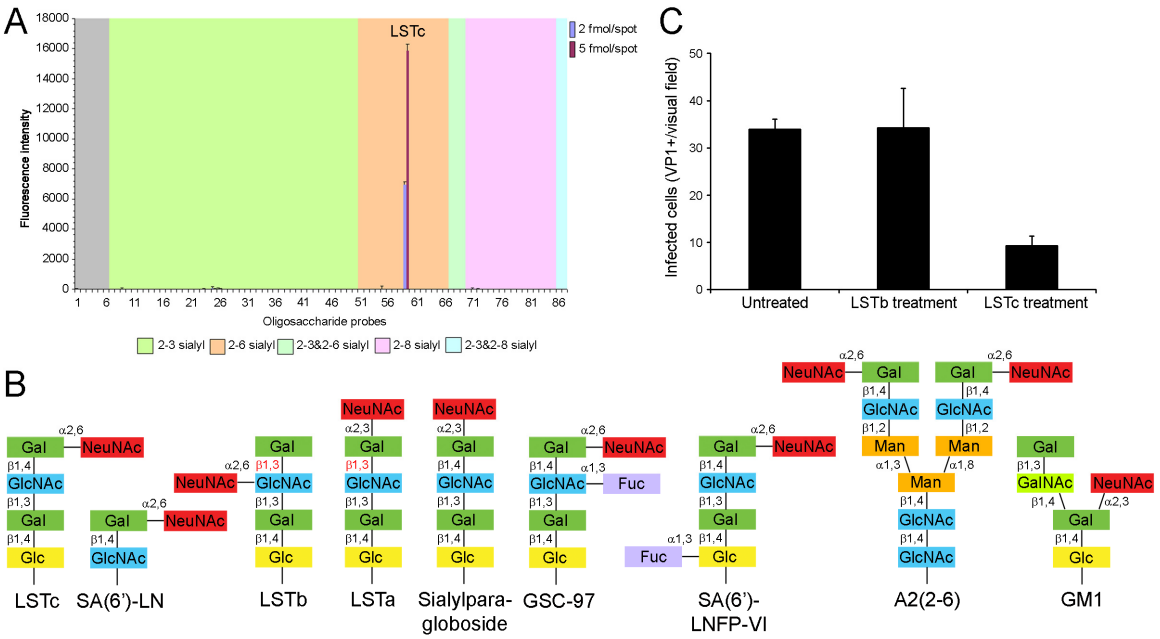



**Fig. 15: LSTc is a functional, specific receptor motif for JCV**
(A) Glycan microarray analysis of JCV VP1 showing highly selective binding to LSTc. Numerical scores for the binding intensity are shown as means of fluorescence intensities of duplicate spots at 2 (in blue) and 5 (in red) fmol/spot. Error bars represent half of the difference between the two values.
(B) Structures of selected glycans present on the glycan microarray. The oligosaccharide sequence of the LSTc probe is shown, as well as those of similar compounds that were not bound.
(C) LSTc inhibits JCV Infection. JC virus was pre-incubated with LSTb or LSTc, and complexes were added to SVG-A cells for infection. Infected cells were quantified based on nuclear VP1 staining. The data represent the average number of infected cells per visual field for 8 fields of view from an experiment performed in triplicate. Error bars indicate SD of triplicate samples.

JCV VP1 bound specifically to α2,6-linked sialyl lactoseries tetrasaccharide c (LSTc) (Fig. 15A). LSTc is a linear pentasaccharide with the sequence NeuNAc-α2,6-Gal-β1,4-GlcNAc-β1,3-Gal-β1,4-Glc. This sequence is of the type that occurs in the peripheral region of glycans of glycoproteins and glycolipids. The glycan array contained several sequences that closely resembled LSTc, but VP1 did not bind to these (Fig. 15A,B).

### 3.3.2. Oligosaccharide competition assays

To determine whether JCV engages LSTc to mediate infection of host cells, JCV was pre-incubated with LSTc, and the mixture was added to SVG-A cells, a glial cell line in which JCV can propagate (2.4.4). The JCV/LSTc mixture blocked JCV infection, while incubation with a JCV/LSTb mixture had no effect on JCV infection (Fig. 15C). LSTb is identical to LSTc in molecular weight and composition but features a branched $\alpha$2,6-linked NeuNAc on the GlcNAc residue of the Gal-$\beta$1,3-GlcNAc backbone (Fig. 15B). Thus, the specific engagement of an LSTc-like sequence on host cells is critical for JCV infection.

### 3.3.3. Crystallization and structure determination

For crystallization, cleaved JCV VP1 was concentrated to 8 mg/mL and screened with the sitting drop vapour diffusion method with the Hampton I & II, Wizard I & II and PEG/Ion I & II crystallization screens (2.5.1). Initial crystals were obtained at 20 °C in a condition of the PEG/Ion screen that contained 20% (w/v) PEG 3,350 and 200 mM KSCN. Bigger crystals that grew more reproducibly were obtained by reducing the protein concentration, microseeding and switching to the hanging drop method that allows for bigger drops. The optimized reservoir solution contained 12% (w/v) PEG 3,350, 100 mM HEPES pH 7.5, and 200 mM KSCN. Drops were set up by mixing 1 µL protein solution at 4.5 mg/mL with 1 µL reservoir solution and 0.2 µL microseeding stock from previously obtained crystals. Crystals grew after 1 – 2 days and were harvested into a solution containing 10% (w/v) PEG 3,350, 100 mM HEPES pH 7.5 and 200 mM KSCN. They were cryoprotected by soaking for 10 s in harvesting solution containing 30% (v/v) glycerol and flash frozen in liquid nitrogen. For complex formation, crystals were soaked for 3 min in harvesting solution supplemented with 5 mM LSTc. The cryoprotecting solution then also contained the same concentration of LSTc.

Diffraction data was collected at beamline PX III of the SLS equipped with a MarCCD detector (2.5.2). Data reduction was performed with XDS (Kabsch, 2010) (2.5.3). The structure was solved by molecular replacement in Phaser (McCoy et al., 2007) using a search model generated from the SV40 VP1 pentamer structure (3BWQ) (2.5.4), and refined at 1.9 and 2.0 Å, respectively, for unliganded and complexed JCV VP1. Refinement was then performed with Phenix and Refmac5 (Adams et al., 2010; Murshudov et al., 1997), and model building was done in Coot (Emsley and Cowtan, 2004) (2.5.5). Five-fold non-crystallographic symmetry restraints were used throughout the refinement for most of the protein. In data from crystals soaked in LSTc, the ligand was located in $2F_{oob,soak}$-$F_{calc,nat}$ and $F_{oobs,soak}$-$F_{calc,nat}$ weighted electron density maps. The ligand was refined using restraints from the CCP4 library, with the exception of the $\alpha$2,6-glycosidic bond, which had to be user-

defined. The final models have good stereochemistry and agree very well with the experimental data (6.2.2).

### 3.3.4. Structure of the JCV VP1-LSTc complex

The overall structure of the JCV VP1 pentamer closely resembles the SV40 VP1 pentamer. LSTc binds on top of the pentamer, on the outer surface of the virion (Fig. 16A,B). Two of the five possible binding sites on VP1 contain LSTc, the other three are occluded by other pentamers in the crystal lattice. As for SV40 VP1, residues and loops of the clockwise and counterclockwise neighbors of a VP1 monomer are designated cw and ccw, respectively. The binding site for LSTc is formed by residues from the BC1-, BC2-, DE- and HI-loops from one monomer as well as the BCcw-loop from its clockwise neighbor ( Fig. 16B). These loops also engage receptors in SV40 and Polyoma VP1, and they are the only parts of VP1 that differ significantly among the three viruses. LSTc binds where the BC1-loop of one monomer approaches the tip of the BC2-loop of its clockwise neighbor, and it interacts with both.

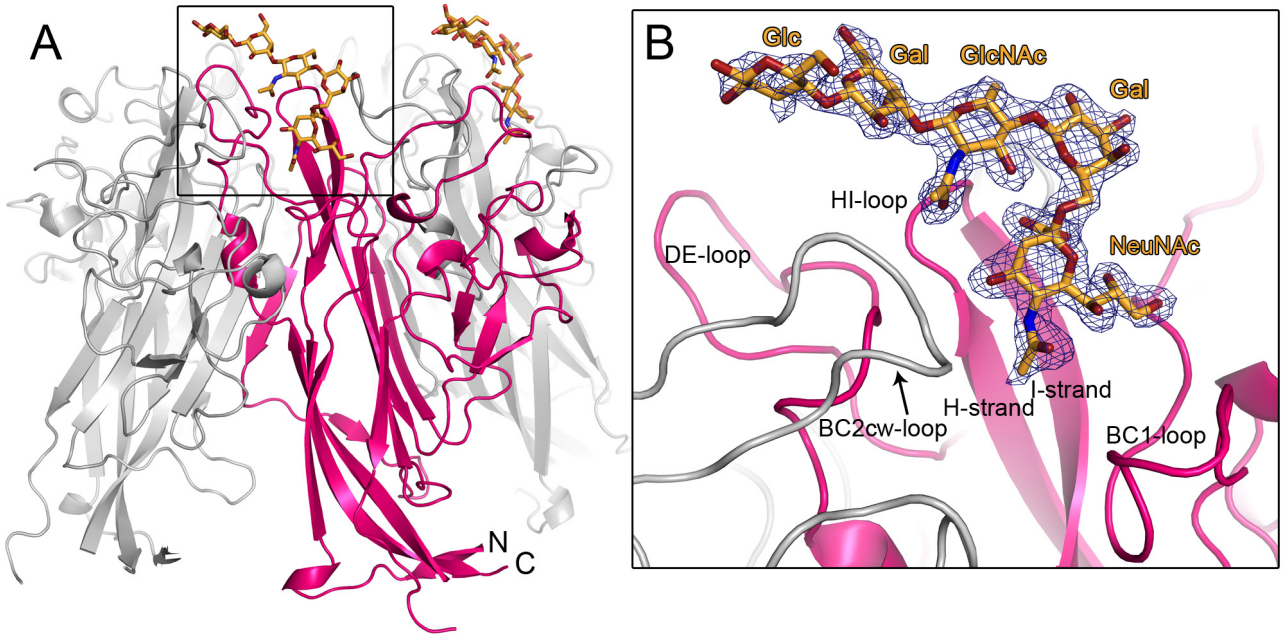



**Fig. 16: Structure of JCV VP1 in complex with LSTc.**
(A) Structure of the JCV VP1 pentamer in complex with LSTc. The protein is shown in cartoon representation, with one VP1 monomer highlighted in pink and the other monomers depicted in gray. The LSTc oligosaccharide is shown in stick representation, with carbons drawn in orange, oxygens in red and nitrogens in blue.
(B) Close-up view of the LSTc binding site. JCV VP1 and LSTc are drawn as in (A). A composite annealed omit difference density map of LSTc is shown contoured at 3.0 $\sigma$ for 2.0 Å around LSTc.

#### 3.3.4.1. Structure of LSTc

The bound LSTc molecule resembles the letter “L”, with the shorter leg formed by NeuNAc and its $\alpha$2,6-linkage to Gal, and the longer leg consisting of the GlcNAc-$\beta$1,3-Gal-

β1,4-Glc trisaccharide (Fig. 16B). In solution, the glycosidic bonds of LSTc can rotate freely due to its unbranched sequence, and the molecule can assume different conformations (Breg et al., 1989a). However, as the NeuNAc and GlcNAc moieties of LSTc are attached to positions 6 and 1 of the same Gal residue, its conformations all feature a kink. JCV VP1 interacts selectively with one of these conformations, contacting sugar moieties in both legs of the L-shaped ligand. The observed conformation of LSTc as well as its orientation with respect to the VP1 surface differs from previously published homology models of JCV interactions with oligosaccharides (Gee et al., 2004; Sunyaev et al., 2009).

#### 3.3.4.2. Interactions between JCV VP1 and LSTc

Most of the interactions between JCV and LSTc involve the three longer functional groups projecting from the terminal NeuNAc residue (Fig. 17A,B). The NeuNAc carboxylate faces towards the HI-loop, and is recognized by hydrogen bonds to the side chains of S266 and S268, as well as via a water-mediated hydrogen bond to the S266 amide group. The extended NeuNAc glycerol chain docks into a shallow depression on the protein surface that is shaped by residues in the BC- and HI-loops. The bottom of the depression is formed by side chain of Q270, which projects upwards from the I-strand. Hydrogen bonds to the amide and hydroxyl groups of S60, located at the rim of the depression, further anchor the glycerol chain. The methyl moiety of the N-acetyl group inserts into a hydrophobic cavity at the intersection between the HI-, BC1- and BC2cw-loops. The N-acetyl group nitrogen is hydrogen-bonded to the side chain carbonyl of N264 of the HI-loop. Finally, the O4 hydroxyl group of NeuNAc forms a hydrogen bond with the side chain of N123 in the DE-loop. The O4 hydroxyl group also makes two water-mediated hydrogen bonds to the backbone of N123 and F67cw.

The majority of contacts between JCV VP1 and the long leg of the L-shaped LSTc compound involve the GlcNAc residue (Fig. 17B). The side chain of DE-loop residue N123 forms hydrogen bonds to both the carbonyl of the GlcNAc N-acetyl group and the side chain of R265 in the HI-loop. In turn, the carbonyl group of the N-acetyl chain forms water-mediated hydrogen bonds with the main chain of S266 and the carboxyl group of the NeuNAc residue. The methyl group of the GlcNAc N-acetyl chain makes van der Waals interactions with the methyl group of T66cw. In addition, the O3 hydroxyl group of GlcNAc forms an intramolecular hydrogen bond with the glycosidic oxygen linking NeuNAc to Gal. There are no hydrogen bonds between JCV VP1 and the remaining three sugar moieties in LSTc.

However, the Glc and two Gal residues contribute to the surface area buried upon complex formation (Fig. 18) and are well-ordered in the structure, indicating low flexibility due to van der Waals interactions with VP1. These interactions are clearly relevant for specific binding of JCV, as demonstrated by the glycan microarray.

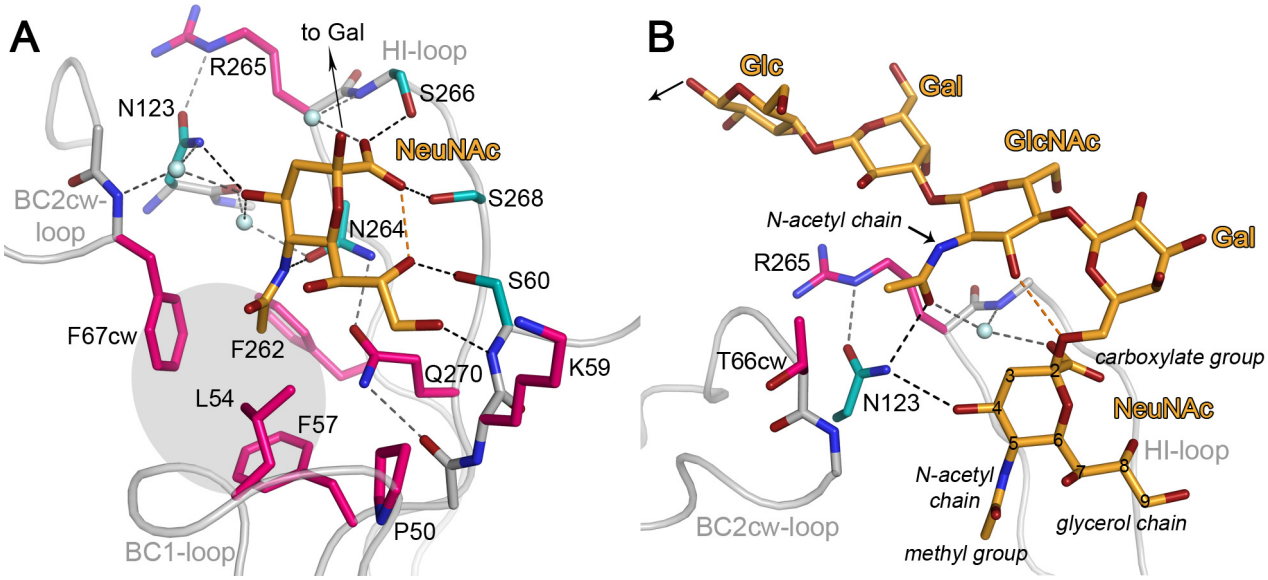



**Fig. 17: Interactions between JCV VP1 and LSTc**
(A) Interactions between JCV VP1 and the terminal NeuNAc of LSTc. JCV VP1 is shown as a cartoon, with side chains interacting with NeuNAc in stick representation. NeuNAc is shown in stick representation, with carbons drawn in orange, oxygens in red and nitrogens in blue. Waters are represented as spheres. Residues forming direct hydrogen bonds to NeuNAc are colored teal and residues forming van der Waals contacts or water-mediated hydrogen bonds are colored pink. Direct hydrogen bonds between JCV VP1 and NeuNAc are shown as black dashed lines, water-mediated hydrogen bonds or bonds between protein atoms are colored gray. Intramolecular hydrogen bonds within the oligosaccharide are orange.
(B) Interactions between JCV VP1 and other parts of LSTc. Atoms are colored as in (A).

### 3.3.4.3. Carbohydrate binding sites are necessary for JCV infection

For the key interactions observed in the crystal structure, mutants were designed to abolish these interactions. These were: S266A/S268A, S266N/S268N and L54Y in the sialic acid binding site, as well as N123A and R265A that probe contacts with the longer leg of LSTc. Viruses with mutations in the sialic acid binding site were unable to propagate due to a severe binding defect (Neu et al., 2010). The N123A mutant featured significantly reduced binding, but propagated to half the level of wild-type, while the R265A mutant propagated like wild-type, but had a milder binding defect. These results corroborate the key roles of sialic acid as major point of attachment, and of N123 as main contact with LSTc.

### 3.3.4.4. Structural Basis for the Narrow Specificity of JCV for LSTc

The crystal structure provides a platform for understanding the narrow specificity of JCV VP1 for the LSTc motif (Fig. 18A,B). As the majority of contacts involve NeuNAc, it is not surprising that compounds lacking this terminal sugar residue do not bind. However, the position of NeuNAc in the larger context of an oligosaccharide is clearly a major determinant of specificity as many close homologs of LSTc fail to interact with JCV VP1 (Fig. 15B). Amino acid N123 is key to the specific engagement of LSTc since it makes hydrogen bonds with

both the GlcNAc and terminal NeuNAc sugar residues (Fig. 17A,B). These interactions are dependent on the observed conformation of the terminal NeuNAc-α2,6-Gal-β1,4-GlcNAc motif. Different linkages of NeuNAc, Gal and GlcNAc would not produce the precise arrangement of NeuNAc and GlcNAc that is required for complex formation. As a result, sialylparagloboside, which carries α2,3-linked NeuNAc but is otherwise identical in oligosaccharide sequence to LSTc, is not bound (Fig. 15B).

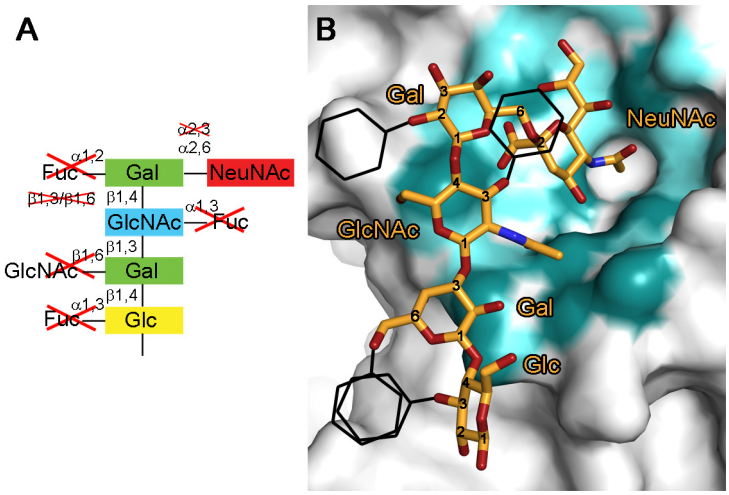



**Fig. 18: Specific recognition of LSTc by JCV VP1**
(A) The cartoon represents structural features of oligosaccharides that are required for JCV binding. These were extracted from our glycan microarray data. Crossed-out sugar residues would produce steric clashes.
(B) Structural basis for JCV VP1 specificity for LSTc. JCV is shown in surface representation, with residues interacting with LSTc colored according to their change in surface accessibility upon LSTc binding (gray < 1 Å$^2$ change, light teal 1-10 Å$^2$ change, dark teal > 10 Å$^2$ change). The branching substitutions at LSTc that abolish binding are indicated as black hexagons, indicating where they would clash with protein or LSTc atoms.

Interestingly, 6-sialyl-lactosamine, whose NeuNAc-α2,6-Gal-β1,4-GlcNAc motif is also present in LSTc, is also not recognized by JCV VP1. As our glycan microarray is (neo)glycolipid-based, the GlcNAc residue of 6-sialyl-lactosamine is directly linked to a lipid tag (6.3.2). Productive engagement of this oligosaccharide might be prevented because of its close proximity to the lipid. Thus, for reasons of accessibility JCV requires at least the LSTc pentasaccharide for binding, defining it as a minimal binding motif (Fig. 18A).

The glycan array tested the effect of additional, branching residues to this minimal binding motif. Compounds fucosylated at O2 of Gal, O3 of GlcNAc, and O3 of Glc were not bound. These fucoses would clash with either JCV VP1 or other LSTc residues (Fig. 18D). Likewise, attachment of a second, branched oligosaccharide chain to O6 of Gal, as in DSLNnH, or direct attachment of the NeuNAc-α2,6-Gal-β1,4-GlcNAc motif to mannose residues, as in short outer chains of N-glycans, would also lead to steric clashes. However, the presence of a GlcNAc residue instead of the terminal Glc of LSTc (GlcNAc-LSTc), as present in most glycans carrying LSTc-like sequences, is unlikely to alter the conformation of the oligosaccharide. Our data therefore provide strong support for a mode of binding that

requires both an α2,6-linked terminal NeuNAc and the extended L-shaped binding conformation of LSTc in the context of a long, linear chain.

### 3.3.4.5. JCV VP1 Undergoes Induced Fit Movements upon Binding LSTc

To determine whether JCV VP1 undergoes conformational changes upon ligand binding, we solved the structure of unliganded JCV VP1 at 1.95 Å resolution (6.2.2). Comparison with liganded VP1 shows that residues in the binding site rearrange upon ligand binding (Fig. 19).

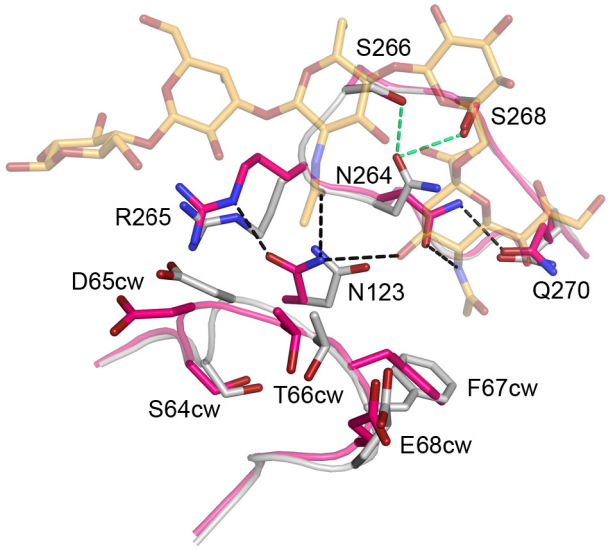



**Fig. 19: Structural changes in JCV VP1 upon LSTc binding.**
The structures of unliganded (gray) and liganded (pink) JCV VP1 were superposed using the β-sandwich core residues. Hydrogen bonds only present in unliganded VP1 are indicated with green dashes, those only present in the complex are colored black.

In the unliganded structure, the N264 side chain is oriented towards the side chains of S266 and S268 and forms hydrogen bonds with both residues. Upon ligand binding, N264 swings around to interact with Q270. This rearrangement enables N264 to form a hydrogen bond with the N-acetyl chain of NeuNAc. In the unliganded structure, the "binding" conformation of N264 is already present as a second conformation, suggesting a dynamic equilibrium between two conformations that is shifted by ligand binding. More drastic changes are seen at residue N123, which is critical for recognizing the L-shaped binding conformation of LSTc. In the absence of LSTc, the N123 side chain points closer towards the NeuNAc binding site, and would clash with the O4 hydroxyl of NeuNAc. Upon LSTc binding, the N123 side chain moves to accommodate NeuNAc. It now forms hydrogen bonds to the O4 hydroxyl of NeuNAc and the GalNAc N-acetyl group. The altered position of N123 leads to clashes with main chain atoms in the BC2cw-loop, and so residues 64-68 of the BC2cw-loop move to accommodate the new conformation of N123. Induced fit movements were not observed in receptor interactions of SV40 and Polyoma (Neu et al., 2008; Stehle and Harrison, 1997) (3.2.4). The induced fit of JCV VP1 may help to bind LSTc in a unique conformation.

#### 3.3.4.6. Additional glycerol in JCV VP1 structure

In both the JCV VP1-LSTc complex and the JCV VP1 structure, ordered glycerol molecules from the cryoprotectant were found bound to the BC2-loop in a way that the BC2-loop wrapped around them. Since glycerol might be indicative of a second sialic acid binding site, crystals were soaked in 20 mM 6'-sialyllactose. To prevent displacement of the oligosaccharide by glycerol, they were frozen with 30 % ethylene glycol as cryoprotectant. However, no electron density for bound sialic acid could be seen outside the previously identified binding site (Luisa J. Ströh, personal communication). Thus, the glycerol either was an artefact or indicates a very weak sialic acid binding site that can only favourably modulate binding of oligosaccharides already bound to the main sialic acid binding site.

## 3.4. BKV and its receptor GD3

### 3.4.1. Crystallization and structure determination

#### 3.4.1.1. Crystal form 1

The first crystal of BKV VP1 (Crystal form 1) was obtained by hanging-drop vapour diffusion (2.5.1) from a condition of the Wizard screens, namely 40% (w/v) PEG 400, 100 mM cacodylate pH 6.5 and 200 mM lithium sulphate. The protein had been screened at a concentration of 10.8 mg/mL in the absence of DTT. A crystal of roughly 200 x 200 x 200 $\mu m^3$ from a similar condition, namely 40 % (w/v) PEG 400, 200 mM lithium sulphate and 100 mM cacodylate pH 6.27 was used for data collection at beamline PX I of the SLS equipped with a MarCCD detector (2.5.2). The crystal had been harvested into mother liquor and quickly frozen in mother liquor containing 30% glycerol as a cryoprotectant.

A complete dataset of 2.9 Å resolution in space group $C222_1$ could be assembled by scaling together data from illuminating several volumes of this crystal by scalepack in HKL2000 (Otwinowski and Minor, 1997) (2.5.3) . The unit cell of this crystal was large with 242.2 x 447.4 x 155.9 $Å^3$, and Matthews coefficient analysis predicted 5 – 8 VP1 pentamers per asymmetric unit (2.5.4). However, only three pentamers could be found by molecular replacement with a search model generated from the SV40 pentamer structure (3BWR) with AMoRe (Navaza, 1994) (2.5.4). The three BKV VP1 pentamers were arranged in the asymmetric unit in a way that gave rise to solvent-filled channels with a diameter of 200 Å that traversed the entire crystal (Fig. 20). This is reflected in the high Matthews coefficient of 4.7 Å/Da and solvent content.

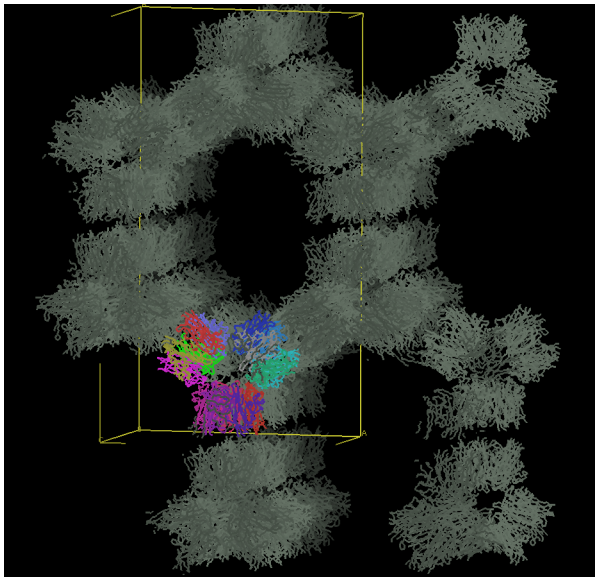


**Fig. 20: Crystal packing of BKV VP1 crystal form 1.** The three VP1 pentamers in the asymmetric unit are shown in color as Cα traces. Crystallographic symmetry mates are colored gray. The tunnels between the pentamers span the entire crystal and have diameters of roughly 200 by 150 Å.

The structure was refined at 2.9 Å resolution by alternating rounds of model building in Coot (Emsley and Cowtan, 2004) and restrained refinement in Refmac (Murshudov et al., 1997) with 15-fold non-crystallographic symmetry restraints (2.5.5). The final electron density map was very clear due to the high level of NCS, and the model contained residues 45 – 98 and 107 – 296 in all 15 chains of BKV VP1. Like in the structure of SV40 VP1 pentamers, the BKV VP1 pentamers packed together with their bottom faces and had formed disulfide bonds mediated by C104 in the CD-loop during crystallization (3.2.4).

For this crystal form, no harvesting conditions could be found in which the crystals retained their diffraction quality for more than 1 min. Since weakly-binding oligosaccharides often need longer times to engage their binding sites in the crystal, no complex with carbohydrate could be obtained from this crystal form. However, ordered sulphate ions from the 200 mM lithium sulphate in the crystallization solution were visible in the electron density. Three sulphate ions per VP1 monomer were observed. All had bound to the top of the VP1 pentamer, and one sulphate was bound at the same position as the sialic acid carboxylate group in the SV40-GM1 complex, suggesting that the sialic acid binding site might be conserved between the two proteins (Fig. 21).

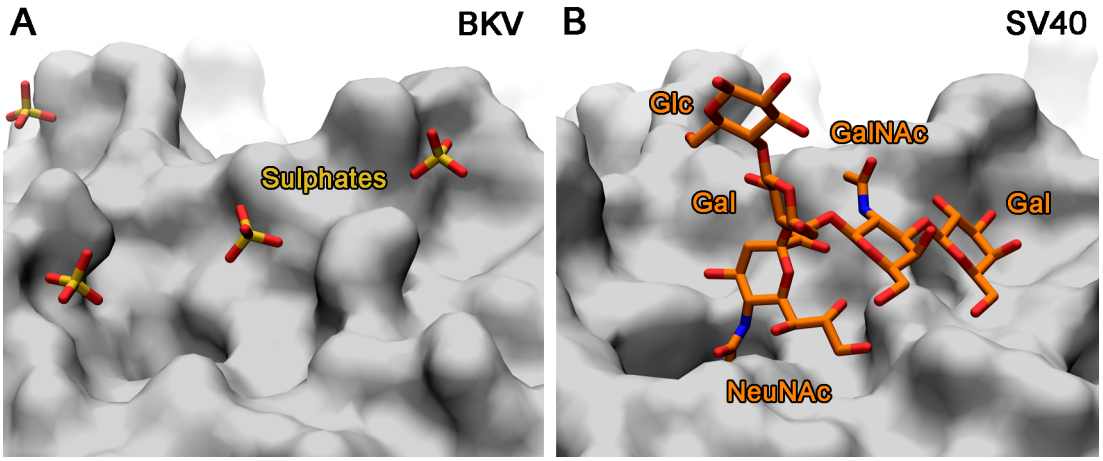



**Fig. 21: Sulphate-binding sites on BKV VP1 might indicate a conserved sialic acid binding site.** (A) + (B) BKV VP1 in complex with sulphate ions (A) and SV40 VP1 in complex with GM1 (B). VP1 is shown in surface representation and colored gray. Ligands are shown in stick representation, with carbons drawn in orange, sulphurs in yellow, oxygens in red and nitrogens in blue. The carboxylate group of the GM1 NeuNAc is bound at the same position as the middle sulphate ion.

### 3.4.1.2. Crystal form 2

Since crystal form 1 could not be used for soaking oligosaccharides, BKV VP1 was also screened with the Hampton, Wizard, Precipitant Synergy and PEG/Ion screens at a lower protein concentration of 6.6 mg/mL in the presence of 20 mM DTT using the sitting-drop method (2.5.1). The best initial crystals grew at 20 °C in a condition of the PEG/Ion screen that contained 20% (w/v) PEG 3,350 and 200 mM lithium chloride. By varying the precipitant and salt concentrations, bigger crystals were obtained with a solution containing

16 – 18 % PEG 3,350, 0.1 M HEPES pH 7.5 and 250 mM lithium chloride. The crystals were harvested into reservoir solution containing only 14-16% (w/v) PEG 3,350, in which they were stable for 15 – 20 min, and cryoprotected by soaking in harvesting solution supplemented with 30% (v/v) glycerol for 10 s. They were then flash-frozen in liquid nitrogen.

Since GT1b was a known receptor for BKV, crystals were first soaked in solutions containing the oligosaccharide portion of GT1b in order to obtain a crystal of the complex. However, no well-diffracting complex crystal could be obtained although different soaking times and oligosaccharide concentrations were used. Crystals were either damaged during soaking and did not diffract beyond 4 Å anymore, or they diffracted well, but had not bound oligosaccharide. Since the GT1b oligosaccharide is a branched heptasaccharide whose size, shape and charge might damage crystals, a smaller compound was identified that might serve as analogon for GT1b binding. Both GT1b and GD1b are receptors for BKV and both carry an α2,8-disialic acid motif on their right arm (Fig. 22). GD3 carries an unbranched tetrasaccharide that consists of only the stem and the conserved right arm of GD1b and GT1b. As it might mimick GT1b in its interaction with BKV, it was used for complex formation. BKV VP1 crystals were soaked in harvesting solution containing 20 mM GD3 oligosaccharide for 15 min before cryoprotection and freezing. The cryoprotectant solutions also contained 20 mM GD3 oligosaccharide.

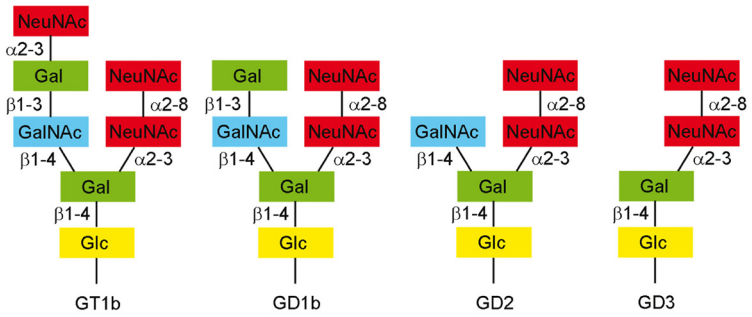



**Fig. 22: Oligosaccharide structures of b-series gangliosides**
GT1b and GD1b are known receptors for BKV. GD3 oligosaccharide was used as an analogon for obtaining a crystal of a BKV VP1-receptor complex.

Diffraction data of unliganded and complexed BKV VP1 crystals were collected at beamlines PX III (equipped with a MarCCD detector) and PX I of the SLS (equipped with a PILATUS detector), respectively (2.5.2). Data were processed with XDS to resolutions of 2.0 and 1.7 Å, respectively (Kabsch, 2010) (2.5.3). The structure was solved by molecular replacement with Phaser (McCoy et al., 2007) using the structure obtained from crystal form 1 as a search model (2.5.4). This second crystal form was in space group $P2_12_12$ and contained one VP1 pentamer in the asymmetric unit. The VP1 pentamers packed against each other top to bottom and had not formed disulfide bonds. With a solvent content of 46.8 %, the packing was much denser than in crystal form 1 (73.8 % solvent content), which might have contributed to the higher resolution. Both the unliganded and complex structures were refined with Phenix and Refmac using the fivefold NCS among the five VP1 monomers

as a restraint (Adams et al., 2010; Murshudov et al., 1997) (2.5.5). Oligosaccharide residues were located in Coot (Emsley and Cowtan, 2004) in weighted $2F_{obs,soak}$-$F_{calc,nat}$ and $F_{obs,soak}$-$F_{calc,nat}$ electron density maps and refined with restraints from the CCP4 monomers library; only the α2,3- and α2,8-glycosidic linkages had to be user-defined. The final models span residues 37 – 98 and 108 – 298 of BKV VP1 for all chains.

### 3.4.2. Structure of the BKV VP1-GD3 complex

The BKV protein adopts the β-sandwich fold established from structures of its homologs SV40 and JCV VP1 (3.2.4, 3.3.4). The surface loops, especially BC2 and DE, feature the same main chain conformation as in SV40 VP1 (r.m.s.d. between Cα atoms of the BC2-loop 0.35 Å).

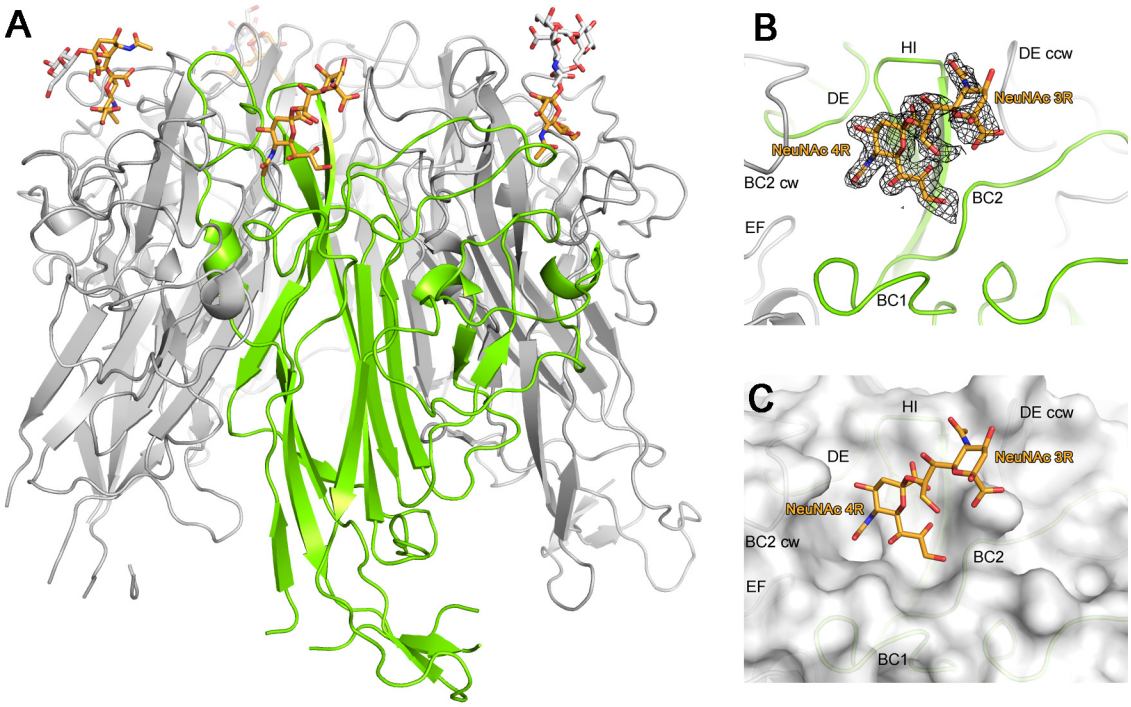



**Fig. 23: Structure of a BKV VP1-GD3 oligosaccharide complex**
(A) Structure of a free BKV VP1 pentamer in complex with GD3. One VP1 monomer is highlighted in green, the others are colored gray. GD3 is shown in stick representation, with oxygens in red and nitrogens in blue. The carbons of the GD3 monosaccharides unbiased by crystal contacts are colored orange, while the ones of monosaccharides in contact with crystal contacts are colored white.
(B) Composite annealed difference electron density for the terminal disialic acid motif of GD3 oligosaccharide bound to BKV VP1 at a σ level of 2.5. The names of VP1 surface loops are indicated. cw/ccw = belonging to the clockwise/counterclockwise neighboring VP1 monomer within the pentamer.
(C) Binding surface for α2,8-disialic acid on BKV VP1

The oligosaccharide engages BKV VP1 at the top of the pentamer, which corresponds to the outer surface of the virion (Fig. 23). Like in other polyomavirus receptor complexes, contacts involve residues in the BC1-, HI- and DE-loops of one monomer, as well as the BC2cw- and DEccw-loops of neighboring monomers. Four of the five binding sites within one

pentamer are occupied by ligand, while one is inaccessible due to crystal packing. Two of the four occupied binding sites, however, are in close proximity to adjacent VP1 pentamers within the crystal. With the exception of the terminal NeuNAc, all of the monosaccharides visible in these two sites participate in crystal contacts with other VP1 pentamers. As their conformations are influenced by these non-physiologic interactions, they will not be considered further. The two remaining GD3 oligosaccharides do not participate in crystal contacts and assume an identical conformation, which therefore should represent a physiologically relevant complex. In both cases, only the terminal NeuNAc-α2,8-NeuNAc motif interacts with BKV VP1, while the Gal-Glc moiety projects into solution. The structures of BKV VP1 alone and in complex with GD3 are virtually identical (r.m.s.d. of 0.12 Å for all atoms), indicating that GD3 binding does not induce a conformational change in the protein.

#### 3.4.2.1. Contacts between BKV VP1 and GD3

The terminal NeuNAc is the main contact of GD3 with BKV VP1. In all four occupied binding sites on the BKV VP1 pentamer, it adopts the same position and conformation, with identical contacts with the protein regardless of the presence or absence of crystal contacts. BKV VP1 makes extensive contact with the three projecting functional groups of NeuNAc. The carboxylate group is recognized by two hydrogen bonds to the side chains of S274 and T276 (Fig. 24A). Additional, water-mediated hydrogen bonds are formed to the side chain of S273 and the backbone of S274. The O4 hydroxyl group of NeuNAc 4R interacts via water-mediated hydrogen bonds with the side chain of N272, the backbone of G131, and the backbone nitrogen of F75cw. The N-acetyl group is bound by a hydrogen bond to N272. Moreover, its methyl group inserts into a tight-fitting, hydrophobic pocket on BKV VP1 that is formed by L62, F65, F270 and F75cw. In addition, the amide oxygen of the N-acetyl group is bound by a water-mediated hydrogen bond to R169cw. The water mediating this contact is part of network of water molecules, with contacts to D59, N61 and D74cw. The glycerol chain of NeuNAc lies in a shallow, rimmed groove and makes van der Waals contacts to both the bottom, formed by Q278, and the rim of the groove, formed by the side chains of P58, L62, L67 and K68. Via its O9 hydroxyl group, it furthermore forms a single hydrogen bond to the K68 backbone nitrogen.

The second sialic acid of GD3, NeuNAc 3R, makes fewer contacts with the protein (Fig. 24B). It is likely to be more mobile, which is reflected in its weaker electron density and elevated temperature factors in comparison to NeuNAc 4R. The carboxylate group of NeuNAc 3R forms a bidentate salt bridge with the positively charged side chain of K68. The methyl group of its N-acetyl chain stacks against a hydrophobic surface created by parts of the side chains of H138ccw, S274 and T276. Without any other contacts to the protein, NeuNAc 3R is suspended between the side chain of K68 and the surface that contacts its N-acetyl group. Again, its conformation is stabilized by an intramolecular hydrogen bond

between its carboxylate group and the O9 hydroxyl group of its glycerol chain. Weak electron density ($\sigma$ <1) is observed for the Gal moieties of the two GD3 molecules that are in the presumed physiological conformation. The electron density did not allow for accurate modelling of this moiety, but indicates that the Gal-$\beta$1,4-Glc moiety projects into solution.

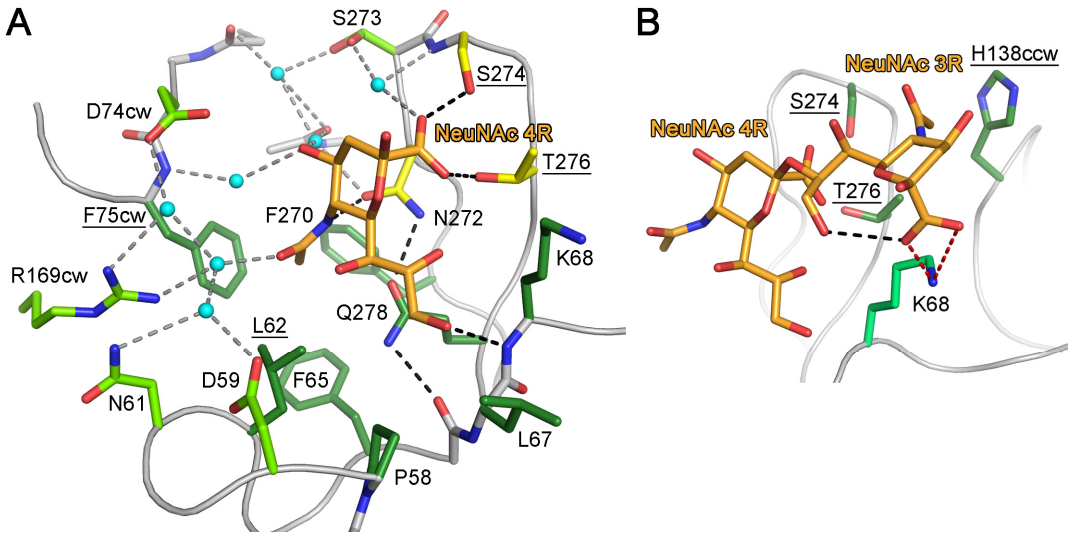



**Fig. 24: Binding site for α2,8-linked disialic acid on BKV**
(A) Interactions of terminal NeuNAc with BKV VP1. GD3 is shown in stick representation, with carbons colored orange, nitrogens blue and oxygens red. The protein residues that make contact with GD3 are shown in stick representation and colored according to atom type. Their carbons are colored by the kind of interaction they make with the sugar: Side chains making hydrogen bonds are colored yellow, those making van der Waals interactions are colored dark green, and those making water-mediated hydrogen bonds are colored light green. The protein backbone is shown as grey cartoon, and water molecules as cyan spheres. Direct hydrogen bonds are indicated as black dashed lines, water-mediated ones are gray. Residues probed by site-directed mutagenesis are underlined.
(B) Contacts of second NeuNAc. The terminal and second sialic acids are denoted 4R and 3R, respectively, due to their positions on gangliosides. Only residues in contact with NeuNAc 3R are shown. Colors are as in A. The salt bridge between NeuNAc2 and K68 is colored red.

#### 3.4.2.2. Carbohydrate binding sites are necessary for BKV infection

Like for JCV, mutants were designed to probe the interactions observed in the complex structure. Virus with mutations in the binding site for the terminal sialic acid (L62W, F75V, F75W, S274A, T276A, S274A/T276A) did not propagate (Stacy-ann A. Allen and Walter J. Atwood, personal communication), demonstrating the physiological importance of this site. Virus with a mutation in the binding site for the internal sialic acid (H138A) propagated, but more slowly than wild-type. K68 was also targeted for mutation and is described below (3.6.1).

### 3.4.3. Saturation transfer difference NMR

Since BKV VP1 had bound the a2,8-disialic acid motif of GD3 in the crystal structure, the interaction was investigated in solution by saturation transfer difference (STD) NMR

spectroscopy (Fig. 25). These experiments were performed by Dr. Bärbel Blaum (Universität zu Lübeck) on samples provided by me.

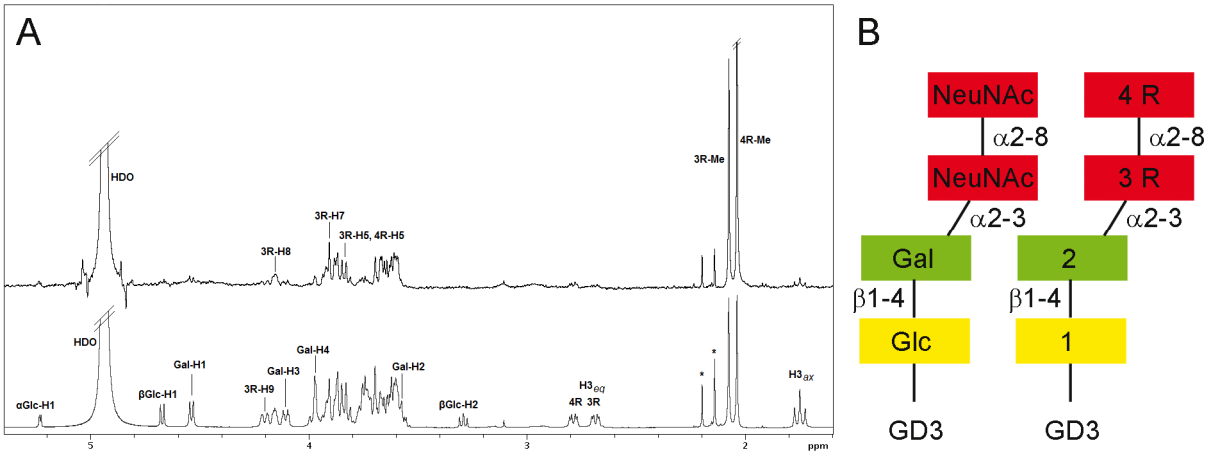



**Fig. 25: BKV VP1 binds to the disialic acid motif found in all b-series gangliosides**
(A) STD difference spectrum (top) and off-resonance spectrum (bottom) of WT BKV VP1 in the presence of 50-fold excess of GD3 oligosaccharide. The off-resonance spectrum was scaled to 3%. GD3 resonances labeled in the difference spectrum receive considerable saturation transfer from the protein while those labeled in the off-resonance spectrum receive no or negligible saturation transfer. Regions with strong signal overlap are not labeled because saturation effects in this region cannot be unambiguously assigned. Signals which were truncated are denoted by diagonal bars, impurities are labeled with asterisks (bottom). The large signal at 4.7 ppm is due to residual water.
(B) Schematic of the GD3 sequence, and of the nomenclature used for assignment of the NMR spectrum.

The strongest saturation transfer from BKV VP1 was observed for the methyl group of the terminal NeuNAc (NeuNAc 4R), followed by the methyl group of the neighboring NeuNAc (NeuNAc 3R) (Fig. 25). No significant transfer was observed to any of the anomeric protons between 4.2 and 5.0 ppm, or to the NeuNAc H3 protons. Interestingly, no transfer was also observed for the entire Glc and Gal residues of GD3 (Fig. 25), indicating that they do not participate in binding BKV VP1. This corresponds well with the X-ray structure of the BKV VP1-GD3 complex, in which the Gal-β1,4-Glc disaccharide projected into solution and did not contact BKV VP1 (3.4.2). Some resonances of the two NeuNAc rings could be unambiguously assigned in the spectral region between 3 and 4 ppm, where heavy signal overlap occurs, among them H5, H7 and H8 from NeuNAc 3R and H5 from NeuNAc 4R. Thus, the NMR experiments confirmed α2,8-disialic acid as the BKV binding epitope on GD3 that had been identified by X-ray crystallography (3.4.2).

### 3.4.4. Glycan array screening

In order to probe binding to GT1b, GD1b and GD3 simultaneously in one assay, His-tagged BKV VP1 pentamers were subjected to glycan array screening by Dr. Yan Liu (Imperial College London, UK) (2.4.1.2). The array contained relevant ganglioside-derived sequences as well as non-sialylated controls at four different concentrations (6.3.3).

BKV VP1 exhibited only very weak binding compared to SV40 and JCV VP1, which both gave 10-fold higher signal. Weak signals could be observed for the GT1b and GD1b oligosaccharides (Fig. 26). No reproducible signal was obtained for GD3 and GD2, which also feature the α2,8-disialic acid motif. It is likely that the interaction of BKV VP1 pentamers with GD3 and GD2 is too weak for detection on the glycan microarray, even after precomplexation. However, binding between GD3 and BKV VP1 had been observed by STD NMR, which is more sensitive for extremely low-affinity systems.

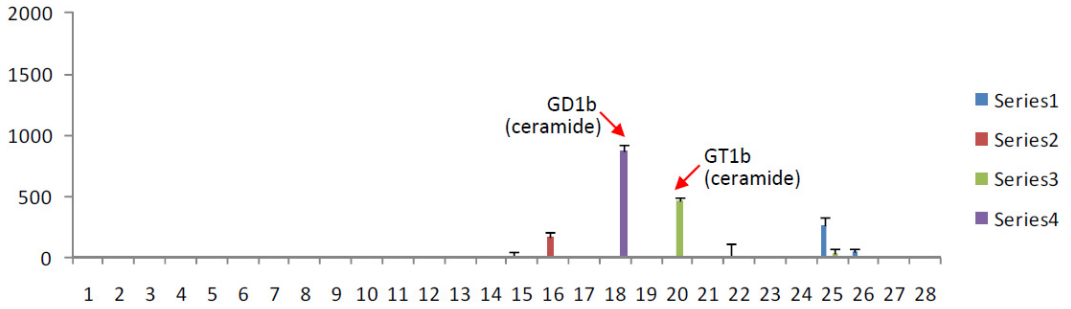



**Fig. 26: Glycan array screening of BKV VP1.**
BKV VP1 showed weak binding to the ceramide-linked GD1b and GT1b oligosaccharides. The protein was analyzed at 150 µg/ml on the dose-dependent probe set and had been precomplexed with primary and secondary antibodies. The fluorescence signal could only be visualized with 100 % laser excitation power. Series 1-4 correspond to different probe concentrations of 0.3, 0.8, 1.7 and 5 fmol/spot.

### 3.4.5. Ganglioside supplementation assays

Next, the different binding partners for BKV VP1 were probed for their capacity to support BKV infection by Stacy-ann A. Allen (Brown University) in a cell-culture based ganglioside supplementation assay (2.4.5). Briefly, Vero cells were incubated with different gangliosides overnight to incorporate them into the plasma membrane of the cells. Then, BKV was added to infect the cells, and infection rates were determined by immunofluorescence microscopy after 3 days. The b-series gangliosides GD3, GD2, GD1b and GT1b that all carry the α2,8-disialic acid epitope were tested with the assay as well as GM1 as a negative control. Consistent with the crystal structure and the NMR experiments (3.4.2, 3.4.3), all b-series gangliosides enhanced infection by BKV while supplementation with GM1 had no effect on infection (Fig. 27). This demonstrates the importance of the α2,8-disialic acid motif as primary epitope for BKV VP1. However, the complex gangliosides GT1b and GD1b supported infection to a higher degree and at lower concentration than GD3 and GD2, indicating that GD1b and GT1b might engage in more contacts with BKV VP1 than GD3 or GD2. This finding is consistent with the results from glycan array screening (3.4.4).

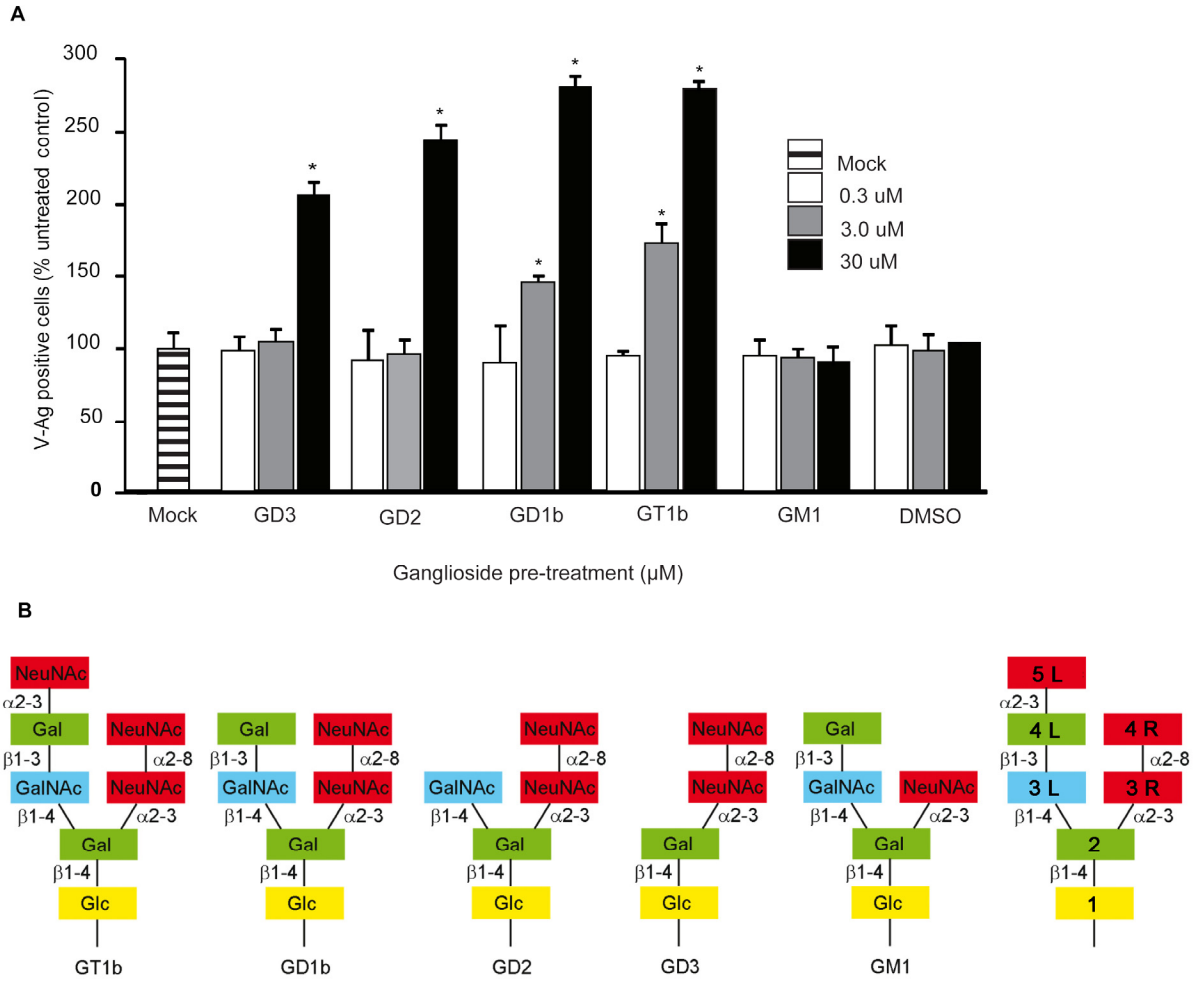



**Fig. 27: BKV uses all b-series gangliosides for infection.**
(A) Infection of BKV after ganglioside supplementation of Vero cells. The average number of VP1 positive cells is plotted compared to controls. Error bars represent the standard deviation for 3 independent experiments and asterisks the p-value (*p<.05).
(B) Schematic view of the ganglioside sequences used in the assay.

## 3.5. Comparison of SV40, BKV and JCV receptor complexes

### 3.5.1. SV40, BKV and JCV feature a conserved sialic acid binding site

The VP1 proteins of BKV, JCV and SV40 share high structure and sequence homology. Their core structures (excluding N- and C-terminal extensions) feature the same overall fold and the same number of amino acids. Yet, different oligosaccharides serve as receptors for these viruses. BKV recognizes $\alpha$2,8-linked disialic acid (3.4), while SV40 is specific for the branched $\alpha$2,3-linked GM1 oligosaccharide (3.2), and JCV attaches to the linear $\alpha$2,6-linked LSTc (3.3).

In all three complexes, the majority of contacts are made to the terminal sialic acid, which is bound in the same orientation and analogous positions on VP1 (Fig. 28). The parallel hydrogen bonds between VP1 and the carboxylate group of sialic acid are conserved, as well as the hydrogen bond to the N-acetyl group of sialic acid. In each complex, the methyl group of sialic acid inserts into a cavity on the protein surface that is formed by residues from two different VP1 monomers. Moreover, all three proteins feature a rimmed depression that accommodates the glycerol chain of sialic acid.

Specificity for different sialic acid types, linkages and underlying carbohydrate sequences arises from a small number of unique contacts that favour interaction with one sialylated receptor and abolish binding to others.

### 3.5.2. Specificity for different sialic acid types

All three proteins exhibit a cavity into which the methyl group of NeuNAc inserts. It is tight-fitting and lined with hydrophobic residues in BKV and JCV, but significantly enlarged and partially hydrophilic in SV40 (Fig. 28). The different cavity sizes arise from amino acid differences at position 166 (BKV and SV40 numbering; the corresponding position in JCV is 158), where BKV and JCV carry a methionine and a phenylalanine, respectively. SV40 carries an alanine at this position. The side chains of M166cw/F158cw close off the cavity of a given monomer from the bottom. In addition, SV40 carries a leucine at position 65, while BKV and JCV both carry a phenylalanine, which also diminishes the size of the cavities.

These differences might reflect the different host species of the viruses. Humans differ from other primates in the predominant sialic acid type they express. A relatively recent deletion prevents humans from metabolizing NeuNAc to NeuNGc. Monkeys therefore carry mostly NeuNGc and small levels of its precursor NeuNAc in their oligosaccharides, while humans carry NeuNAc and only those amounts of NeuNGc that were taken up with nutrition. Consistent with this, glycan array screening of SV40 virus-like particles revealed a

preference of SV40 VP1 to bind NeuNGc-GM1 over NeuNAc-GM1 (Campanero-Rhodes et al., 2007).

The bigger and more hydrophilic glycolyl group of NeuNGc is expected to make interactions with polar residues at the rim of the SV40 cavity. By contrast, modelling shows that the smaller and more hydrophobic cavity of BKV and JCV is unlikely to accommodate the glycolyl group in a similar manner (data not shown).

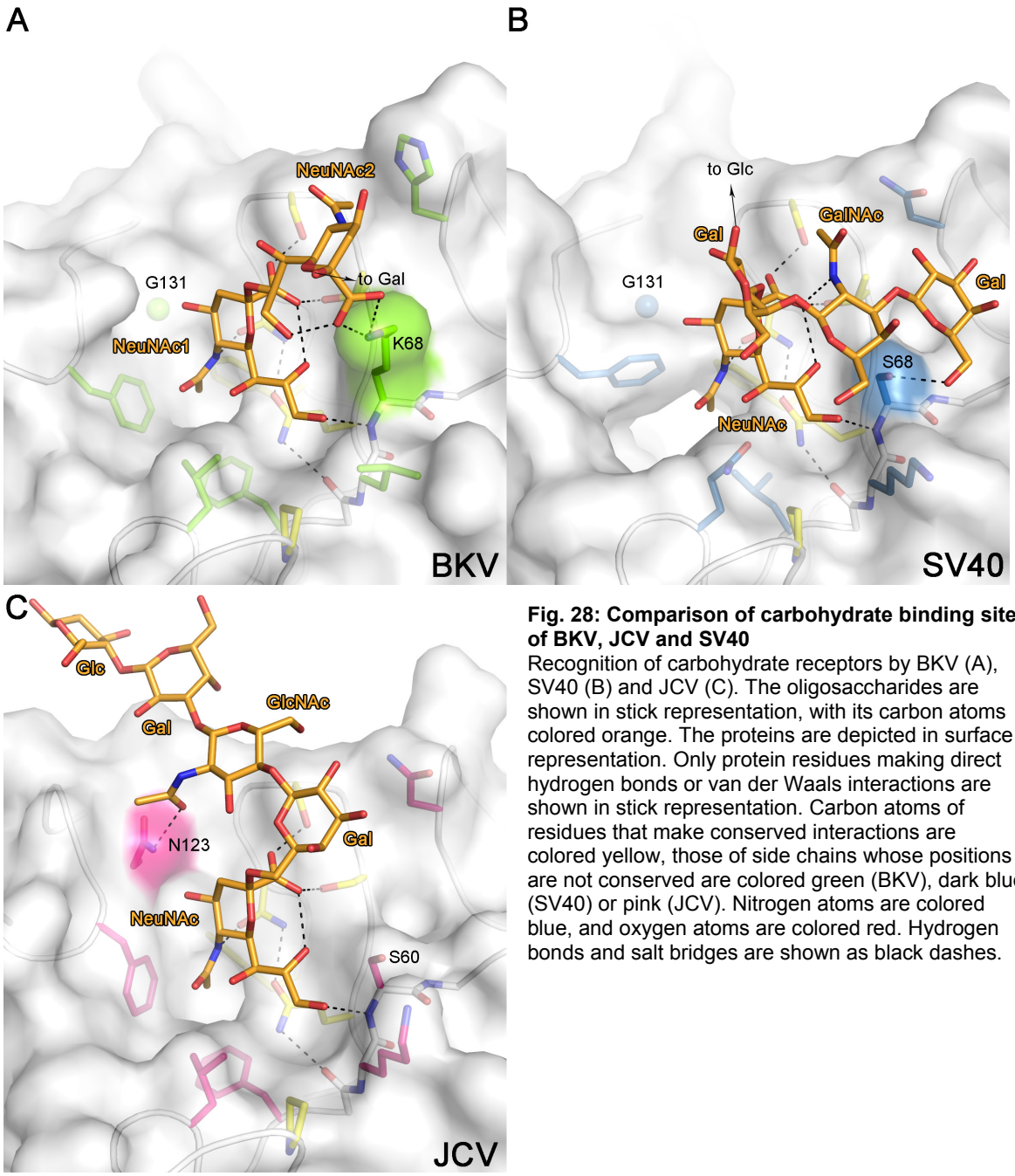



**Fig. 28: Comparison of carbohydrate binding sites of BKV, JCV and SV40**
Recognition of carbohydrate receptors by BKV (A), SV40 (B) and JCV (C). The oligosaccharides are shown in stick representation, with its carbon atoms colored orange. The proteins are depicted in surface representation. Only protein residues making direct hydrogen bonds or van der Waals interactions are shown in stick representation. Carbon atoms of residues that make conserved interactions are colored yellow, those of side chains whose positions are not conserved are colored green (BKV), dark blue (SV40) or pink (JCV). Nitrogen atoms are colored blue, and oxygen atoms are colored red. Hydrogen bonds and salt bridges are shown as black dashes.

### 3.5.3. Specificity for sialic acid in different linkages

BKV VP1 attaches to the α2,8-disialic acid motif of b-series gangliosides. The BKV VP1 residue K68, which is the main contact point for the internal NeuNAc of the disialic acid motif, is not conserved in SV40 and JCV, which both have a serine at the equivalent position (Fig. 28). As the serine cannot engage in a salt bridge with the internal sialic acid, SV40 and JCV are unable to specifically interact with α2,8-disialic acid carrying glycans (3.2.1, 3.3.1). In turn, the long side chain of K68 in BKV would lead to clashes with the second arm of structures that carry a branching α2,3-linked sialic acid such as the SV40 receptor GM1 (Fig. 28). Thus, K68 in BKV both promotes binding of the α2,8-disialic acid motif and blocks binding of branched α2,3-linked compounds.

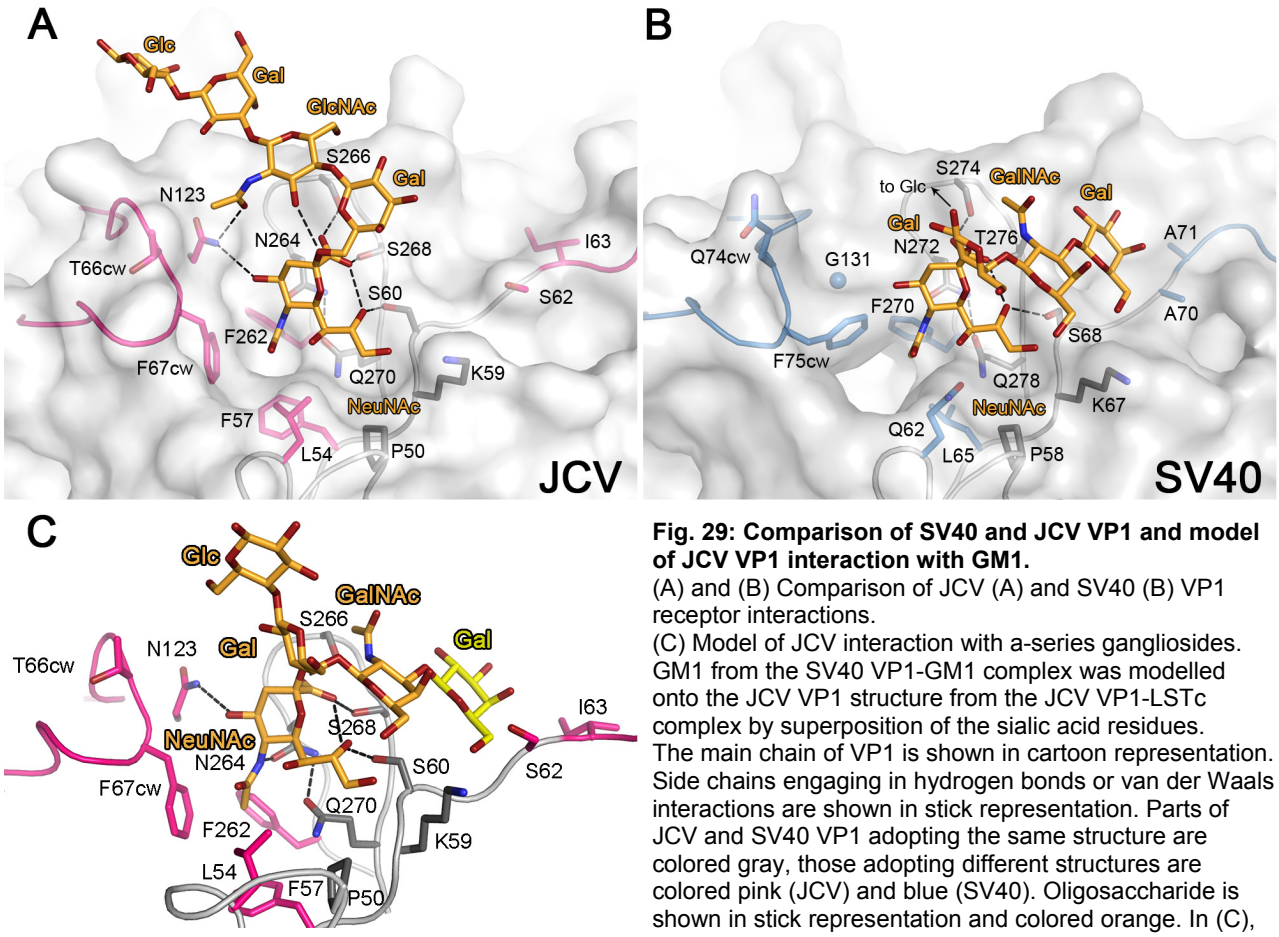



**Fig. 29: Comparison of SV40 and JCV VP1 and model of JCV VP1 interaction with GM1.**
(A) and (B) Comparison of JCV (A) and SV40 (B) VP1 receptor interactions.
(C) Model of JCV interaction with a-series gangliosides. GM1 from the SV40 VP1-GM1 complex was modelled onto the JCV VP1 structure from the JCV VP1-LSTc complex by superposition of the sialic acid residues. The main chain of VP1 is shown in cartoon representation. Side chains engaging in hydrogen bonds or van der Waals interactions are shown in stick representation. Parts of JCV and SV40 VP1 adopting the same structure are colored gray, those adopting different structures are colored pink (JCV) and blue (SV40). Oligosaccharide is shown in stick representation and colored orange. In (C), the Gal specific for GM1 is colored yellow.

JCV recognizes an L-shaped conformation of LSTc that is a consequence of the α2,6-linkage of the terminal NeuNAc to the underlying Gal. The key residue that makes contacts to both legs of the L is N123 that contacts both the NeuNAc and GalNAc moieties of LSTc. BKV and SV40 both have a glycine at the equivalent position that cannot engage in similar

interactions, explaining why these two viruses do not bind to LSTc (Fig. 28). In addition, the BC2-loop of JCV adopts a different conformation to that observed in SV40 or BKV VP1, leading to a pronounced groove on the JCV VP1 surface (Fig. 29A). This is due to the presence of an isoleucine at position 63, while SV40 and BKV carry an alanine at the equivalent position 71. As the bigger isoleucine side chain would clash with the DEcw-loop, the JCV BC2-loop takes up a different conformation that is essential for the induced fit observed in the JCV-LSTc complex, and for van der Waals contacts with LSTc. Conversely, if GM1 were docked on the JCV structure in the same position as in SV40, its terminal Gal would clash with the side chain of S62 of JCV, due to the different conformation of the JCV BC2-loop mentioned above. However, if the terminal Gal residue of GM1 rotated a little bit, a productive, yet energetically somewhat less favourable interaction would be possible. Similarly, GM2, which resembles GM1, but does not carry the terminal Gal, would be able to engage JCV, albeit with lower affinity due to the reduced contact surface. Interestingly, glycan array screening of JCV virus-like particles found strong binding to LSTc and weak binding to GM2 and related, branched compounds (www.functionalglycomics.org). Tab. 3 summarizes the specificity-determining positions in SV40, BKV and JCV VP1.

**Tab. 3: Specificity-determining amino acids in SV40, BKV and JCV VP1**

For BKV and JCV, it is clear that they bind NeuNAc-containing compounds. NeuNGc-containing forms of their receptors were not available and could therefore not be tested. Consequently, the sialic acid preference of these two proteins is in parentheses.

| | SV40 | BKV | JCV | BKV K68S | Verified for BKV K68S? |
|---|---|---|---|---|---|
| Pos. 166 (BK)<br>Pos. 65 (BK) | A<br>L | M<br>F | F<br>F | M<br>F | |
| Cavity | Big | Small | Small | Small | |
| Sialic acid preference | NeuNGc | (NeuNAc) | (NeuNAc) | NeuNAc | Yes |
| Pos. 68 (BK) | S | K | S | S | |
| Bind GM1? | Yes | No | Weakly | Yes | Yes |
| Bind GD3? | No | Yes | No | No | Yes |
| Pos. 131 (BK) | G | G | N | G | |
| Bind LSTc? | No | No | Yes | No | Yes |

## 3.6. Towards retargeting of polyomaviruses to different receptors

### 3.6.1. Retargeting BKV VP1 to bind GM1

The residue K68 in BKV both promotes binding of the α2,8-disialic acid motif and blocks binding of branched α2,3-linked compounds (3.5.3). This is especially interesting given that SV40 and BKV VP1 share 84 % sequence identity. In contrast to JCV, whose main chain conformation differs from the one observed in BKV and SV40 in the BC2-loop, the main chain conformations BKV and SV40 are essentially identical and their Cα atoms can be superposed with an r.m.s.d. of 0.55 Å. Comparison of the BKV and SV40 top surfaces reveals that they carry relatively similar overall surface features (Fig. 30). This suggests that position 68 is the crucial determinant of SV40 and BKV linkage specificity.

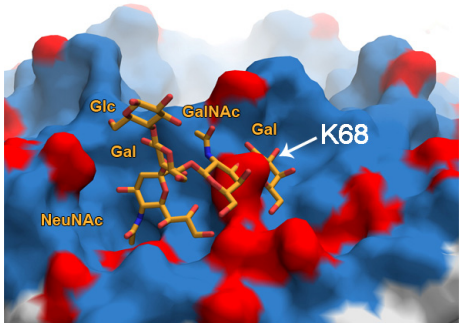



**Fig. 30: Surface conservation between SV40 and BKV VP1 highlights K68 as specificity determinant.**
The BKV VP1 structure is shown in surface representation. GM1 was superposed onto GD3 by superposing the sialic acids and is shown in stick representation. Protein atoms are colored according to conservation with SV40: Atoms identical between the two proteins are colored blue, those unique to BKV are colored red.

To test this hypothesis, K68 in BKV VP1 was targeted for mutation to its counterpart in SV40, a serine. Given the structural data, the BKV K68S mutant was predicted to bind GM1 oligosaccharide, but not oligosaccharides bearing an α2,8-linked disialic acid motif. In addition, since the mutation was done in a BKV background, it should exhibit a preference for NeuNAc-GM1 over NeuNGc-GM1.

#### 3.6.1.1. BKV K68S binds GM1, not GD3

The BKV K68S mutant was first generated in the context of the free pentamer expression construct (2.2.10), and purified in His-tagged form. To determine its specificity, the protein was analyzed using STD NMR (2.4.3) with both GD3 oligosaccharide that binds BKV VP1 (Fig. 25), and GM1 oligosaccharide that binds SV40.

In contrast to BKV VP1 (Fig. 25), saturation transfer from K68S VP1 to GD3 was extremely weak (Fig. 31A), indicating that the mutation abolished binding of BKV to this compound. However, when K68S was assayed with GM1 (Fig. 31B) strong STD signals could be detected for all residues of GM1 except the Glc residue at the stem, indicating interaction of BKV K68S with GM1. Comparison of the K68S + GM1 difference spectrum with

the same spectrum for SV40 + GM1 (Fig. 31C) revealed that the spectra were virtually identical. Saturation transfer is highly sensitive to changes in the interactions of the binding partners, i.e. a different binding site on the protein would lead to a different “imprint” in ligand magnetization and different peak heights in the resulting spectrum. Thus, the fact that the SV40 and K68S difference spectra with GM1 were virtually superposable indicates that both proteins bind GM1 in the same manner, confirming that the K68S mutation converted the BKV oligosaccharide binding site to bind GM1, and did not create a new, distinct binding site.

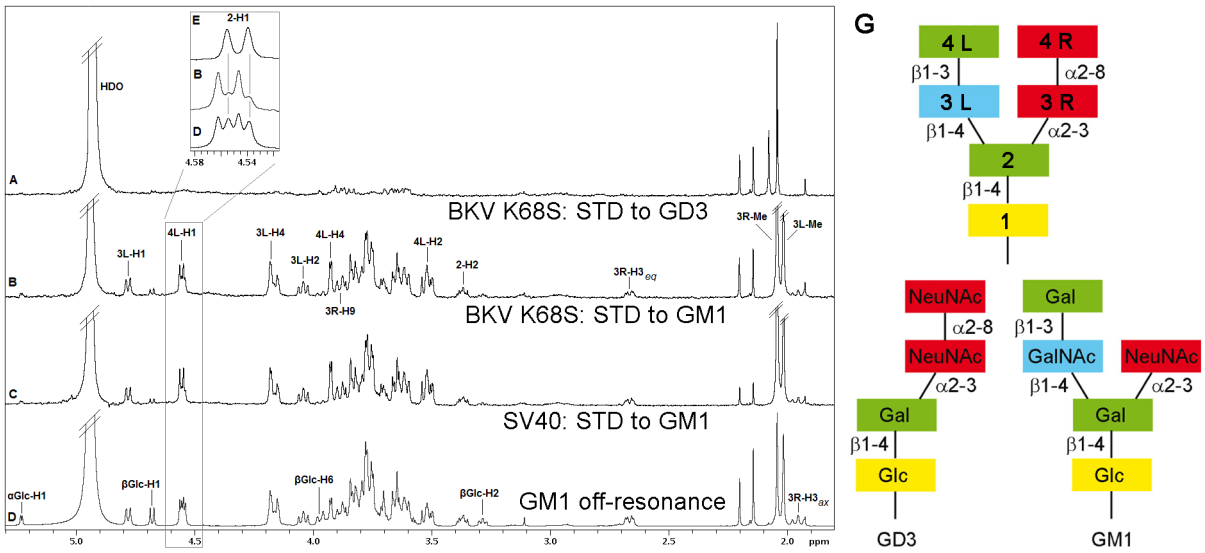



**Fig. 31: BKV K68S binds to GM1, not GD3.**
STD difference spectra of BKV K68S with GD3 (A), BKV K68S and SV40 with GM1 (B and C, respectively) and SV40-GM1 off-resonance spectrum (D).
A 50-fold excess of oligosaccharide was used for each spectrum. The off-resonance spectrum was scaled to 3%. Resonances labeled in the difference spectra with GM1 (B and C) receive considerable saturation transfer from BKV K68S and SV40 while those labeled in the GM1 off-resonance spectrum receive no or negligible saturation transfer. Regions with strong signal overlap are not labeled because they cannot be unambiguously assigned. Binding of BKV K68S to GD3, previously seen for WT BKV (Fig. 2A), is abolished by the mutation (A).
*Enlargement:* A 1D $^{1}$H-TOCSY spectrum of the Gal-2 ring (E) allows distinction of both GM1 galactose H1 resonances despite signal overlap in the off-resonance spectrum (D). The anomeric proton of Gal-4L resonantes at a slightly higher chemical shift and receives more saturation in the STD difference spectra with BKV K68S (B) and SV40 (C). Signals which were truncated are denoted by diagonal bars and impurities are labeled with asterisks. The large signal at 4.7 ppm is due to residual water.
(G) Oligosaccharide sequences used in the experiments, and nomenclature used during assignment.

### 3.6.1.2. BKV K68S binds NeuNAc-GM1, not NeuNGc-GM1

Next, the BKV K68S protein was analyzed by glycan microarray screening (2.4.1.2). The array contained several versions of both NeuNAc-GM1 and NeuNGc-GM1 that differed in their lipid linkers (6.3.3). Consistent with the STD NMR results, K68S did not interact with b-series gangliosides carrying α2,8-disialic acid. Interestingly, K68S interacted only with the human-specific NeuNAc-GM1, but not with the monkey-specific NeuNGc-GM1. This specificity pattern is opposite to the one of SV40 pentamers, which exhibit a preference for NeuGc-GM1, but also tolerate NeuNAc-GM1. Taken together, these results indicate that the

structure of the cavity for the methyl group of NeuNAc directly influences the sialic acid type that can be recognized.

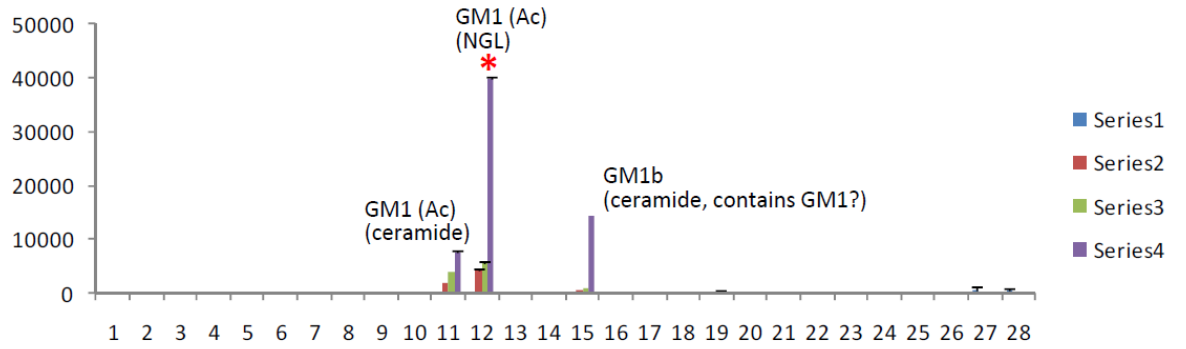



**Fig. 32: BKV K68S binds to NeuNAc-GM1, but not NeuNGc-GM1.**
Binding was observed to the NeuNAc variant of GM1, but not to the NeuNGc forms of GM1, which were present on the array as oligosaccharides #13 and #14.
The protein was analyzed at 150 µg/ml on the dose-dependent probe set and had been precomplexed with primary and secondary antibodies. The fluorescence signal was visualized with 80 % laser excitation power. The asterisk indicates saturadion of the fluorescence detector. Series 1-4 correspond to different probe concentrations of 0.3, 0.8, 1.7 and 5 fmol/spot.

### 3.6.2. Retargeting of JCV to gangliosides

The predominant receptor motif for JCV is the α2,6-linked linear compound LSTc, but the structure suggests that branched, α2,3-linked compounds might bind as well, albeit weakly (3.5.3). This was confirmed by glycan array screening using virus-like particles of the same JCV strain that was used in this thesis (Mad-1) (www.functionalglycomics.org).

However, glycan array screening results by the CFG are also published for a different JCV strain, termed WT1. That strain exhibited relatively strong binding to GM2 and GM2-like compounds that was in the range of the signal intensity observed for LSTc. To define the mechanism of this switch, inspection of the amino acid sequence of WT1 and Mad-1 reveals amino acid differences at five positions, only one of which is in the receptor-binding surface loops. At position 133 in the DE-loop, WT1 carries an alanine while Mad-1 carries a glycine. Even though the change is small, this residue is close to the BC2-loop and might alter its conformation in a small way to be somewhat more conducive to GM2 binding.

Using this information, two mutants were designed in the Mad-1 JCV VP1 background: The first one, G133A, is predicted to bind to LSTc and GM2-type compounds about equally well. The second one, a N123Y G133A double mutant, has the specificity-defining N123 changed to tyrosine, which should abolish interaction with LSTc, but not with sialic acid. It is therefore predicted to bind only GM2-type compounds. These mutants are currently being evaluated by glycan array screening, X-ray crystallography and cell-based infection assays.

The endocytosis pathway will be determined for each mutant by infection assays in the presence of inhibitors of clathrin- or cholesterol-mediated endocytosis.

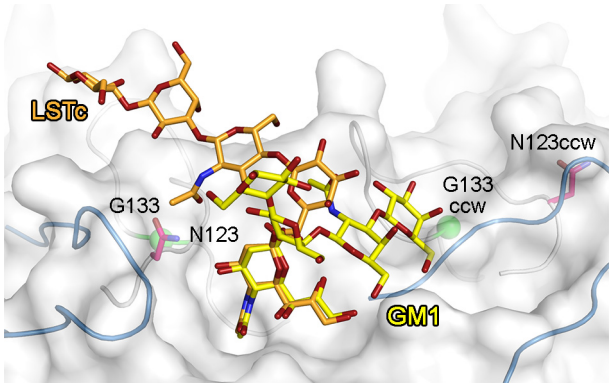



**Fig. 33: Design of JCV VP1 mutants to retarget it to gangliosides.**
JCV VP1 is shown in surface representation. The position of Cα of G133, which is A133 in the WT1 strain, is indicated as a green sphere. Mutation of N123, which is shown in stick represenation, to tyrosine is thought to abolish interaction with the long leg of LSTc, but not with GM1. The DE- and BC2-loops are shown in cartoon representation and colored grey and blue, respectively. LSTc (from the JCV VP1-LSTc) complex and GM1 (modelled) are shown in stick representation. LSTc is colored orange and GM1 is colored yellow.

## 3.7. Structures of KIV and WUV VP1

### 3.7.1. Protein crystallization and structure determination

WUV and KIV VP1 pentamers were screened by the sitting-drop vapor diffusion method (2.5.1) with the Hampton, Wizard and PEG/Ion crystallization screens, both in the presence and absence of 20 mM DTT, and at both 4°C and 20°C. Crystals of WUV VP1 were obtained at 4 °C in the absence of DTT and with a protein concentration of 6.0 mg/mL. The reservoir solution contained 200 mM tri-sodium citrate, 100 mM HEPES pH 7.5 and 20 % (v/v) isopropanol. KIV VP1 was crystallized in the presence of 20 mM DTT at 20 °C and with a protein concentration of 8.6 mg/mL. The reservoir solution contained 200 mM sodium formate pH 7.0 and 20 % (w/v) PEG 3,350. Crystals were harvested into the respective reservoir solutions. For cryoprotection, they were soaked for 10 s in reservoir solution containing 30 % (v/v) glycerol before flash-freezing in liquid nitrogen.

WUV VP1 diffraction data to 2.9 Å were collected at beamline ID14-2 at ESRF (Grenoble, F) with an ADSC Quantum 4 detector. Data for KIV VP1 were first collected at beamline X06DA at SLS with a MarCCD detector, but due to the small crystal size, only data to 3.8 Å resolution were obtained. When the same crystals were exposed at the brighter and more focused beam of beamline X06SA at SLS, which is equipped with a PILATUS detector, they allowed data collection to 2.55 Å resolution (2.5.2). Data were processed with XDS (Kabsch, 2010) (2.5.3), and the structures were solved by molecular replacement with Phaser (McCoy et al., 2007) using the β-sandwich core of the JCV VP1 pentamer structure as search model (2.5.4). The crystals of KIV VP1 belong to space group P1, with two pentamers in their asymmetric unit, while the WUV VP1 crystals contain one pentamer in the asymmetric unit and belong to space group $P4_32_12$. After rigid body and simulated annealing coordinate refinement in Phenix, missing parts of the model such as the surface loops became visible in the electron density, and were built in Coot (Emsley and Cowtan, 2004). Refinement proceeded by alternating rounds of restrained coordinate and isotropic B-factor refinement in Phenix and Refmac5 (Adams et al., 2010; Murshudov et al., 1997) and model building in Coot (2.5.5). The fivefold and tenfold non-crystallographic symmetries among the five VP1 monomers in the WUV VP1 crystal and the ten VP1 monomers in the KIV VP1 crystal, respectively, were used as symmetry constraints throughout refinement.

### 3.7.2. Structures of WUV and KIV VP1

The structures of KIV and WUV VP1 pentamers were solved at 2.55 Å and 2.9 Å resolution, respectively (Fig. 34A, B). Similar to other polyomavirus VP1 proteins, both KIV and WUV VP1 adopt the antiparallel β-sandwich fold typical for viral capsid proteins. Two β-

sheets, formed by β-strands B, I, D and G, and C, H, E and F, respectively, stack against one another in each monomer and form the hydrophobic cores of the proteins. Additional β-strands align with the β-sandwiches from neighboring VP1 molecules to form VP1 pentamers. In both KIV and WUV VP1, the β-sandwiches are decorated with extensive loops that link the β-strands and make up most of the protein surface. The side surfaces of the pentamers are decorated with the very long EF-loops that feature small, three-stranded β-sheets. The CD-loops at the bottom of each pentamer are disordered in most VP1 monomers and only ordered by crystal contacts. They were shown to be flexible in all VP1 pentamer structures determined so far (Stehle and Harrison, 1997) (3.2.4, 3.3.4, 3.4.1.1, 3.4.1.2), and assume different conformations in virions (Liddington et al., 1991; Stehle et al., 1996; Stehle and Harrison, 1996; Stehle et al., 1994).

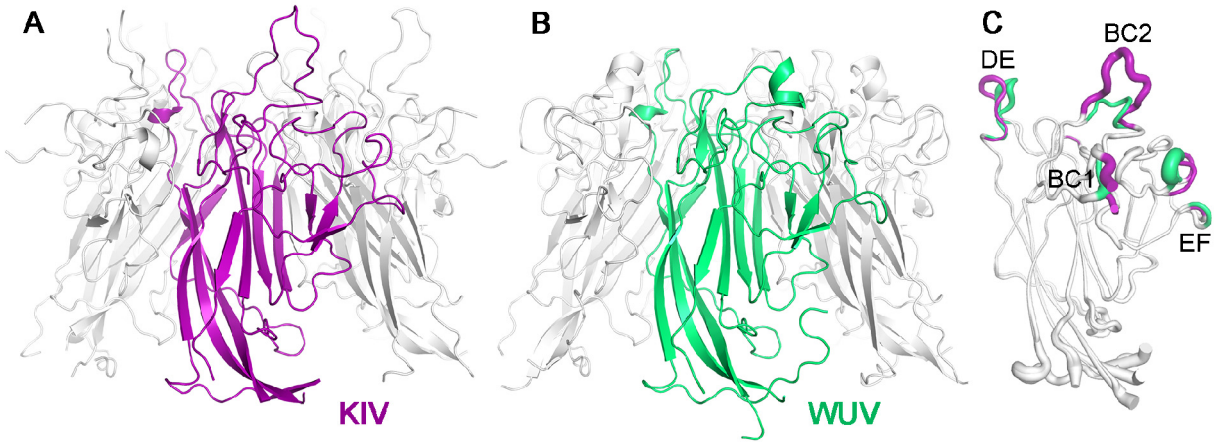



**Fig. 34: Structures of KIV and WUV VP1**
(A)+(B) X-ray structures of unassembled KIV (A) and WUV (B) VP1 pentamers. One monomer is highlighted in color, the others are in grey.
(C) Comparison of the KIV and WUV VP1 monomers. The proteins are shown in tube representation, with the thickness of the tube indicating the B-factor of each amino acid. Protein regions for which the r.m.s.d. for Cα atoms exceeds 0.8 Å are highlighted in color. They align well with the regions that exhibit higher B-factors due to flexibility.

The structures of WUV and KIV VP1 are highly similar to one another. The Cα atoms of one monomer from KIV and WUV VP1 can be superimposed with an r.m.s.d. of 0.79 Å, reflecting their quite high sequence identity of 67.8 %. The main differences between KIV and WUV VP1 locate to relatively short stretches (8 amino acids maximum) in the surface loops, namely the tips of the BC1-, BC2- and DE-loops and parts of the EF-loop (Fig. 34C). If only the conserved cores of VP1 are superposed, the r.m.s.d. between Cα atoms is 0.42 Å, indicating that their differences map to the loop structures. The Cα atoms of entire WUV and KIV VP1 pentamers can be superposed with an r.m.s.d. of 0.86 Å, which is not much higher than for the superposition of VP1 monomers. This suggests that VP1 monomers of KIV and WUV are arranged into pentamers in the same way.

### 3.7.3. Comparison with SV40 and Polyoma VP1

#### 3.7.3.1. Overall structures

Multiple sequence alignment and phylogenetic analysis revealed that VP1 proteins from all polyomaviruses known to date can be classified into three main phylogenetic groups (Fig. 1C) (Schowalter et al., 2010): one encompassing SV40 VP1 and the related BKV and JCV VP1; a big and diverse one containing Polyoma VP1 as well as MCV and LPV VP1; and a third one, which is more distant in evolution to the other two and that consists of WUV VP1, KIV VP1 and of the VP1 proteins from the newly discovered human viruses hPyV6 and hPyV7.

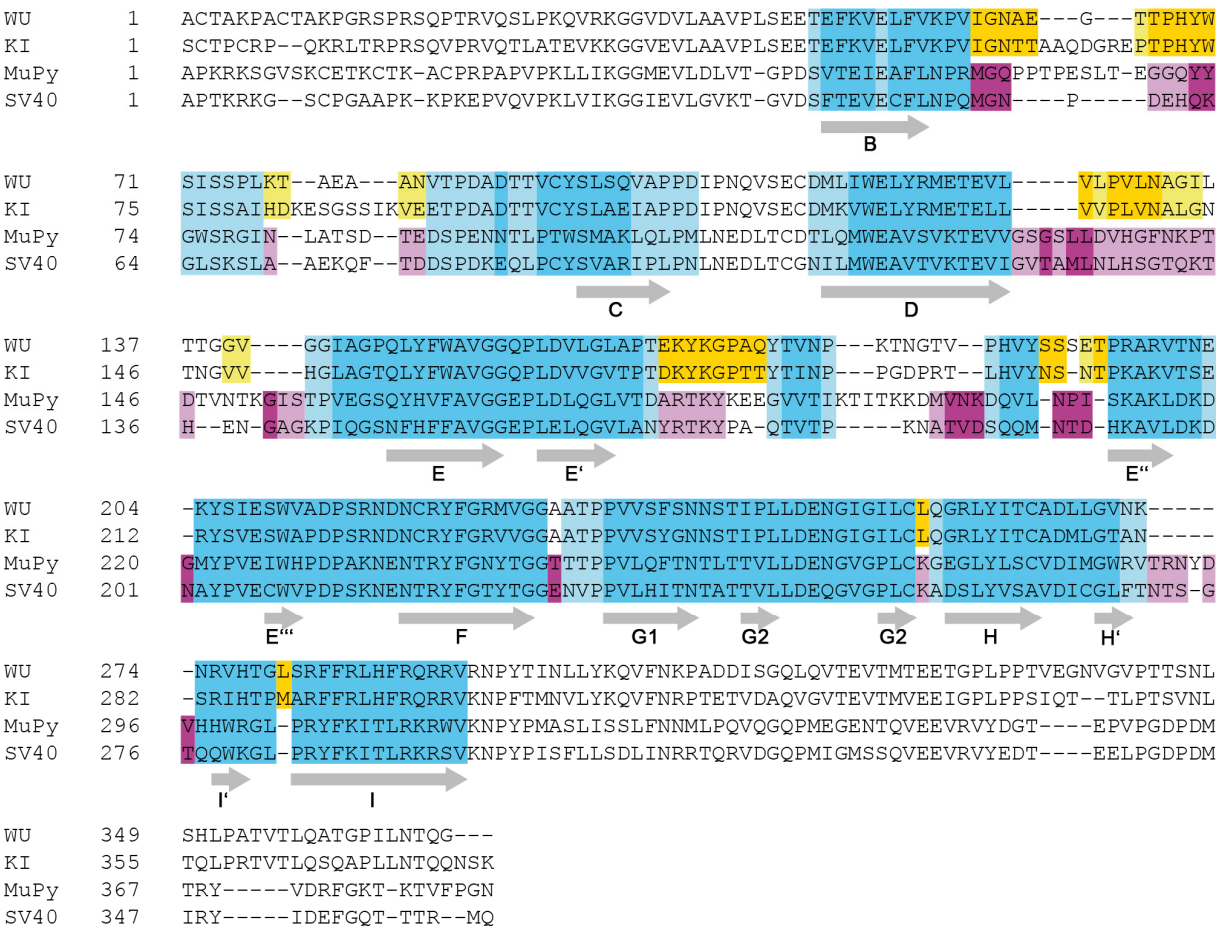



**Fig. 35: Structure-based sequence alignment of WUV, KIV, Polyoma (MuPy) and SV40 VP1**
Protein regions for which all structures align well (r.m.s.d. between Cα atoms < 0.8 Å) or reasonably well (r.m.s.d. between Cα atoms < 1.5 Å) are colored blue and light blue, respectively. Regions where only WUV and KIV (orange and yellow), or SV40 and Polyoma (dark and light purple) align with each other, are colored with the same scheme. β-strands are indicated with grey arrows.

We therefore compared the KIV and WUV VP1 structures to iconic members of the other two groups, namely SV40 and Polyoma VP1, by superposing them using different residue ranges and calculating r.m.s.d. values for their Cα atoms (Tab. 4). The monomer superpositions for the different proteins were then manually combined into a structure-based

sequence alignment (Fig. 35). For reference, the SV40-BKV VP1 pair was also compared because these two VP1 proteins feature the highest level of sequence identity in the polyomavirus family.

The structure of the β-sandwich core is conserved among all proteins compared. However, the β-sandwich cores of KIV and WUV differ somewhat from the cores of SV40 and Polyoma because the r.m.s.d. values of the KIV-SV40 and KIV-Polyoma core comparisons are markedly higher than the ones of the SV40-Polyoma and WUV-KIV pairs (Tab. 4).

**Tab. 4: Comparison between WUV, KIV, SV40, Polyoma and BKV VP1 structures.**
Entire VP1 monomers or pentamers of the indicated proteins were superposed on one another with Superpose in CCP4 (CCP4, 1994) using a secondary-structure matching algorithm. Conserved core structures were identified from the structure-based sequence alignment of WUV, KIV, SV40 and Polyoma as stretches of sequence with more than 5 amino acids for which the average r.m.s.d. of the KIV-SV40, KIV-Polyoma and Polyoma-SV40 pairs was smaller than 0.8 Å. For SV40, these were amino acids 44-54, 86-92, 109-120, 144-166, 193-199, 202-227, 232-254, 257-269, 277-282 and 283-295. These were then superposed with LSQMAN (Kleywegt, 1996).

| Cα r.m.s.d. [Å] | Core (monomer) | Monomer | Pentamer |
|---|---|---|---|
| WUV – KIV | 0.42 | 0.79 | 0.86 |
| Polyoma – KIV | 0.72 | 1.57 | 1.98 |
| SV40 – KIV | 0.76 | 1.48 | 1.90 |
| SV40 – Polyoma | 0.46 | 1.11 | 1.07 |
| SV40 – BKV | 0.32 | 0.69 | 0.72 |

The differences in r.m.s.d. become more pronounced when entire monomers are compared, indicating that the loop structures are more diverse than the cores in each case, even for the closely related SV40 and BKV VP1. For example, the Cα atoms of the Polyoma and SV40 VP1 monomers can be superposed onto KIV VP1 with r.m.s.d. values of 1.57 Å and 1.48 Å, respectively. However, the Polyoma and SV40 VP1 monomers are not very similar to one another either, with an r.m.s.d. of 1.11 Å among Cα atoms. This is in agreement with the phylogenetic observation that they are more distant to each other than WUV VP1 is to KIV VP1. The difference between SV40 and Polyoma VP1 monomers maps mostly to the surface loops, since their β-sandwich cores feature a relatively low r.m.s.d. of 0.46 Å.

This observation is in line with the structure-based sequence alignment, which reveals that more and longer sequences are different between SV40, Polyoma, WUV and KIV VP1 than between WUV and KIV VP1 alone. However, the sites that differ between WUV and KIV

VP1 (tips of BC1-, BC2-, DE-loops and one stretch of EF-loop) also differ between Polyoma and SV40 VP1, defining these as the most variable regions within VP1.

When entire VP1 pentamers are compared, the r.m.s.d. values are quite similar to those calculated for monomers in the case of the WUV-KIV, SV40-BKV and SV40-Polyoma pairs. Thus, each of these pairs likely features a similar arrangement of monomers within the pentamer. In contrast, the pentamer structures of Polyoma and SV40 VP1 can be superposed onto the KIV VP1 pentamer only with significantly higher r.m.s.d. values than the monomers. This indicates that the WUV and KIV VP1 pentamers might be arranged in a slightly different way than those of polyomaviruses from other families.

#### 3.7.3.2. Loop structures

The top surface of the pentamer, corresponding to the outer surface of the virion, is formed almost entirely of loop structures, namely the BC1-, BC2-, DE- and HI-loops, which are highly diverse among the different viruses. The most striking differences between VP1 proteins from different viruses are found in the BC1- and BC2-loops. Unlike the BC1-loops of SV40 and Polyoma, those of KIV and WUV VP1 do not point towards the clockwise neighboring VP1 monomer in the same pentamer, but point away from the fivefold central axis of the pentamer, towards neighboring pentamers within the virion (Fig. 36). The BC1-loops thus point into orthogonal directions in the two groups. While the BC2-loops of Polyoma and SV40 VP1 lie relatively flat on the surface of the protein, the tip of the WUV VP1 BC2-loop forms a short a-helix that is tilted upwards. The BC2-loop of KIV VP1 is longer than those of the other proteins and forms long protrusions from the pentamer surface that render the KIV VP1 surface very rugged (Fig. 36).

The DE-loops of WUV and KIV VP1 are shorter than those of SV40 and Polyoma and, although they point into the same general direction, they engage in different interactions with the core of the proteins (Fig. 36). In addition, they contain a short helix that is not found in SV40 or Polyoma VP1. Like the HI-loops of SV40 and Polyoma, those of WUV and KIV contain two short β-strands, H' and I' that are linked with a β-turn. However, the HI-loops of KIV and WUV VP1 are six amino acids shorter than those of SV40 or Polyoma VP1. Thus, the KIV and WUV HI-loops do not protrude as far to the surface and virtually form a groove between other, longer loops (Fig. 36). The small β-sheet consisting of strands E', E'' and E''' and inserted into the EF-loop is structurally quite conserved between all four viruses.

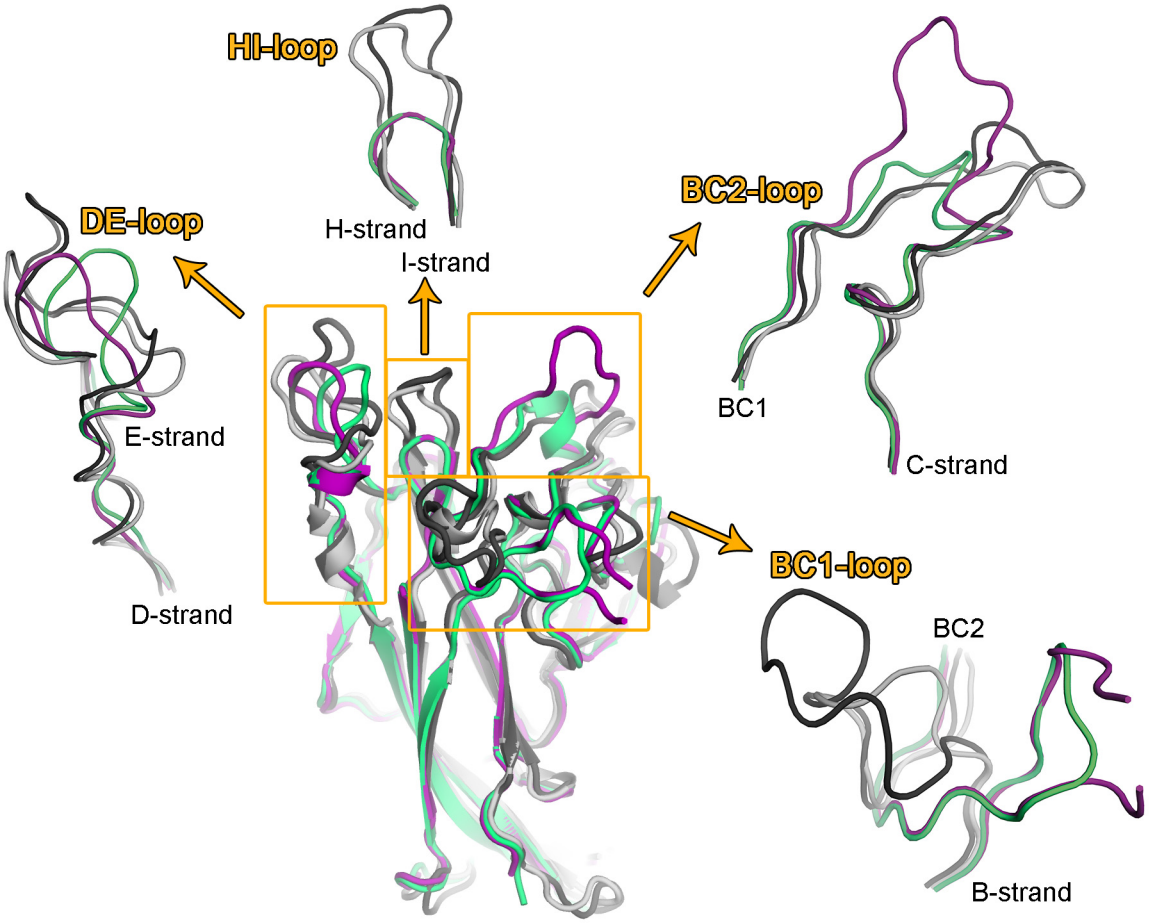



**Fig. 36: Structural differences in the surface loops of VP1**
(A) to (D) Ca-traces of the surface loop structures of KIV (purple), WUV(green), SV40 (light grey) and Polyoma (dark grey) VP1. The overlays are based on the conserved core structure of VP1.

#### 3.7.3.2.1. The BC1-loop

The BC1-loop of KIV is six amino acids longer than the BC1-loop of WUV VP1. Most of these additional residues are disordered in all 10 copies of KIV VP1 in the crystal structure, suggesting that they are also flexible in solution. However, the N-and C-terminal parts of the WUV and KIV BC1-loops feature the same conformation and point in a different direction than the BC1-loops of SV40 and Polyoma VP1. The loops are stabilized by interactions between residues at the beginning and the end of the BC1-loops, which are conserved between WUV and KIV VP1 (Fig. 37). Most prominent is the conserved $\pi$-stacking interaction between N59, H72 and W74 in KIV VP1 (N61, H68 and W70 in WUV VP1). The asparagine and histidine residues of this motif are further stabilized by conserved hydrogen bonds to main chain atoms.

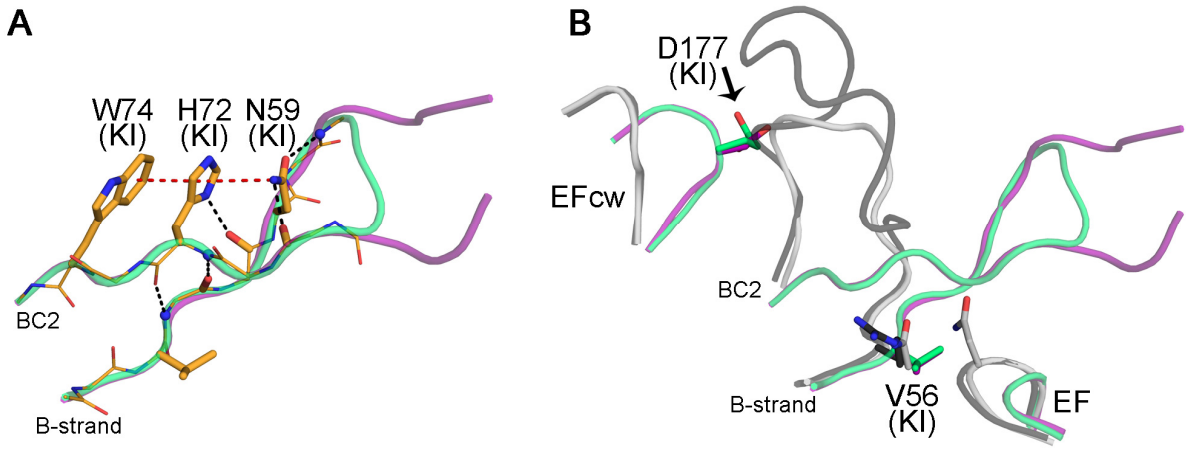



**Fig. 37: Structural basis of the different loop BC1-loop structures of WUV and KIV compared to SV40 and Polyoma**
(A) Interactions that stabilize the BC1-loop conformation of WUV and KIV. The Cα-traces of WUV (green) and KIV (purple) are shown in cartoon representation. Conserved carbon atoms are colored orange. Residues engaging in interactions are shown in stick representation.
(B) The conformation of the BC1-loop depends on specific interactions with residues in the EF-loops of the same and clockwise neighboring monomer. WUV and KIV are colored green and purple, SV40 and Polyoma are colored light and dark grey, respectively.

The different conformation of the WUV and KIV BC1-loops compared to SV40 and Polyoma are also reflected in subtle differences in other loops that make it possible to accommodate the different BC1-loop conformations. For instance, the conserved V58 in WUV VP1 (V56 in KIV VP1) in the BC1-loop would clash with the main chain of N201 in SV40 (G220 in Polyoma) in the EF-loop of the same VP1 monomer, thus guiding their BC1-loop away from the conformation seen in KIV and WUV VP1. This region is one amino acid shorter in WUV and KIV VP1 (Fig. 35, Fig. 37B), enabling the BC1-loop conformation observed in these proteins. The conformations of the BC1-loop in Polyoma and SV40 would likewise not be possible in the context of WUV and KIV VP1 because of clashes with the EFcw-loop (Fig. 37B).

#### 3.7.3.2.2. The BC2-loop

The BC2-loops of both WUV and KIV VP1 are stabilized by salt bridges that link the beginning and the end of the loop (Fig. 38). In addition, both feature hydrogen bonds from the main chain of the BC-loop to the main chain of the EF-loop. The tip of the BC2-loop of WUV VP1 forms a short helix. The long protrusions in the KIV BC2-loop are rather flexible, but they are held together by hydrogen bonds of the main chain and side chain of S87.

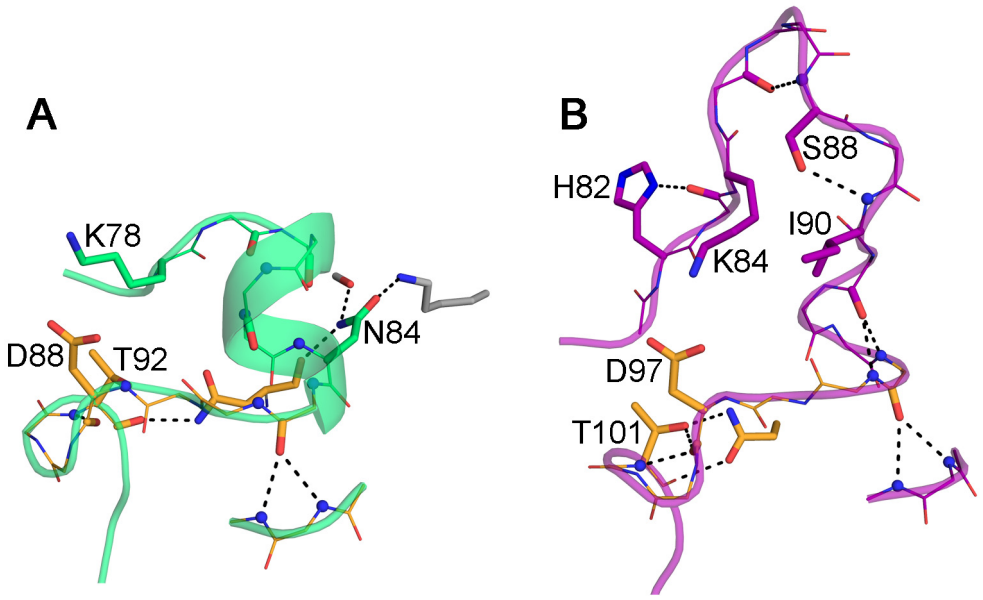



**Fig. 38: Interactions stabilizing the BC2-loop conformations of WUV and KIV.**
The Cα-traces of WUV (green) and KIV (purple) are shown in cartoon representation. The main chain is shown as lines, and residues engaging in interactions are shown in stick representation. Atoms are colored according to atom type: oxygens in red, nitrogens in blue and carbons in green (unique to WUV), purple (unique to KIV), orange (common to both WUV and KIV) or grey (from a different VP1 monomer).

#### 3.7.3.2.3. The BC-linker

Even though both the BC1- and BC2-loops take up completely different conformations in the WUV, KIV, SV40 and Polyoma proteins, a stretch of six amino acids between BC1- and BC2-loops is structurally conserved between all four proteins (Fig. 35). The name "BC-linker" is proposed here for that region. All of its residues feature torsion angles characteristic for β-strands and form main chain hydrogen bonds to the I'-strand of the HI-loop. However, the BC-linker does not form an antiparallel β-sheet with the I'-strand because the side chain of the conserved serine at the third position of the BC-linker makes a hydrogen bond to the main chain of the I'-strand, which disrupts the pattern of main chain hydrogen bonds (Fig. 39). As it also makes a hydrogen bond to the main chain of the following residue in the BC-linker, it causes a pronounced kink in the structure of the BC-linker. In addition, two conserved hydrophobic residues, one before and one after the kink, keep the linker in place: the consensus sequence of this motif is ϕ – S/T – x – x – ϕ. The two hydrophobic residues make van der Waals contacts with the side chains of two consecutive hydrophobic residues on the I'-strand. The first of these residues is strictly hydrophobic, the second one is only hydrophobic close to the main chain in some proteins. In addition, the residue on the I'-strand preceding the two hydrophobic ones often donates a hydrogen bond to the main chain of the conserved serine. All of these features of the linker are conserved in polyomavirus VP1 sequences from all different species, highlighting their importance. For hPyV6 and

hPyV7, automated sequence alignments in this region are ambivalent because of their low level of sequence similarity even to their closest relatives WUV and KIV VP1. However, they do possess a sequence consistent with features of the linker, including the serine, between the sequences for the B- and C-strands. Interestingly, the two residues denoted x in the linker sequence, which protrude into solution, are located in the conserved sialic acid binding site of BKV, JCV and SV40 (4.2.3).

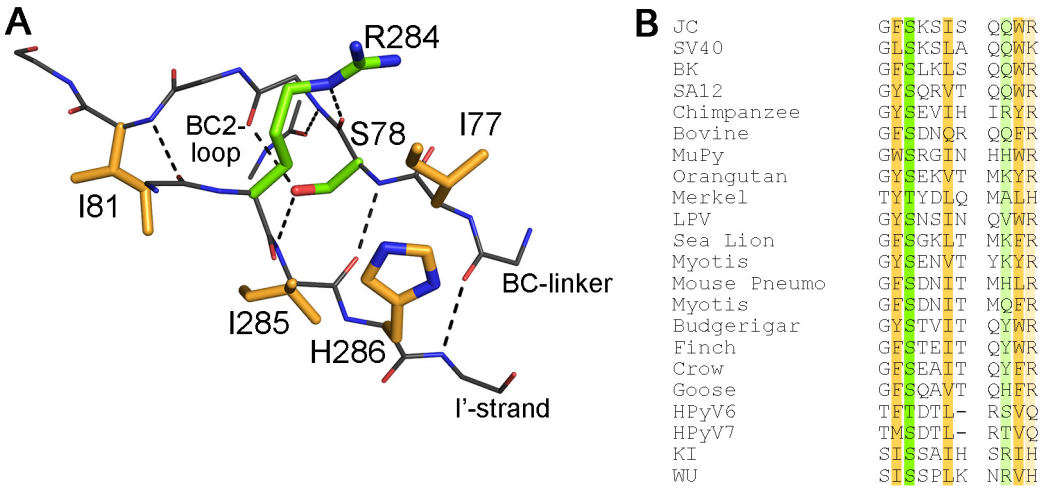



**Fig. 39: Structure of the BC-linker**
(A) BC-linker of KIV VP1. The main chains of the BC-linker, the end of the BC2-loop and the I'-strand are shown as thin lines. Side chains engaging in hydrogen bonds are shown in stick representation and colored green. Side chains engaging in van der Waals interactions are colored orange. Nitrogen atoms are colored blue and oxygen atoms are colored red.
(B) Multiple sequence alignment of the BC-linker (left) and I'-strand (right) of all polyomaviruses. Residues are colored as in A, which residues that are not well conserved shaded pale. The alignment was performed with Muscle (ref), and sequences were then arranged according to the structure-based sequence alignment described here (Fig. 35). The sequences of hPyV6 and hPyV7 had aligned to the BC-linker sequences of KI and WU, but they feature a sequence motif consistent with the BC-linker.

# 4. Discussion

This thesis defines the receptor specificities of the three polyomaviruses SV40, JCV and BKV. Crystal structures of each attachment protein in complex with the respective oligosaccharide receptor reveal a conserved sialic acid binding site and inform on the structural basis of the different receptor specificities. Based on this knowledge, a single amino acid switch was identified that modulates the receptor specificities of SV40 and BKV. This information was then used to retarget BKV to use the SV40 receptor. Moreover, structures of the VP1 proteins of the newly discovered polyomaviruses WUV and KIV were determined. Although the receptors for these viruses are not yet known, the structures already reveal interesting features such as a conserved core structure, a surprisingly different loop arrangement, and a stretch of conserved main chain structure at the center of the receptor binding site.

## 4.1. Relevance of the receptor specificities of SV40, BKV and JCV

### 4.1.1. SV40

Our results demonstrating the narrow specificity of SV40 for its receptor GM1 are in agreement with previous data showing that SV40 can use GM1, but not GD1a or GT1b, as a receptor (Tsai et al., 2003), and they also confirm data on SV40 receptor binding that were generated using a different glycan microarray (Campanero-Rhodes et al., 2007). However, while our crystal structure clearly points to the branching NeuNAc as major point of attachment, earlier reports had claimed that sialic acid was not involved in SV40 attachment because SV40 did not hemagglutinate red blood cells, and because treatment of cells with *V. cholerae* neuraminidase did not reduce SV40 infection (Clayson and Compans, 1989; Shah et al., 1978).

The highly specific recognition of the GM1 oligosaccharide explains why SV40 does not hemagglutinate erythrocytes. Since the GM1 motif is only found in gangliosides and does not occur on glycoproteins, it is likely not expressed on erythrocytes even though they carry sialic acid. Epitopes for other polyomaviruses, e.g. 3'-sialyllactose for Polyoma, however, are present on erythrocytes, and consequently, these viruses do hemagglutinate. The specificity of *V. cholerae* neuraminidase offers clues as to why treatment of cells with this enzyme did not reduce SV40 binding. First, the enzyme is not able to cleave the branching NeuNAc off the GM1 oligosaccharide, most likely due to steric constraints in the active site (Miller-Podraza et al., 1982). Second, when the enzyme cleaves the linear NeuNAc motifs from

more complex gangliosides such as GT1b, the resulting product is GM1 (Svennerholm, 1963). Thus, even though treatment of cells with *V. cholerae* neuraminidase reduced the overall sialylation of cells, it effectively increased the amount of GM1 on the cell surface. This was probably not noticed because the lectins commonly used for NeuNAc staining on cells, *Maackia amurensis* and *Sambucus nigra* agglutinins recognize unbranched linear α2,3- and α2,6-linked NeuNAc, but not GM1 (www.functionalglycomics.org).

The results presented here also give a structural basis for the preference of SV40 for NeuNGc-GM1 over NeuNAc-GM1 (3.5.2). However, the cavity that accommodates the methyl or glycolyl groups of the sialic acid could easily accommodate more atoms, not only an additional hydroxyl group. There are several naturally occurring sialic acids that carry longer substituents at N5 of the sialic acid. It would be interesting to test these in the context of GM1 for SV40 binding.

### 4.1.2. JCV

The results presented here demonstrate that glycans terminating in the LSTc motif serve as main receptors for JCV and suggest that branched α2,3-linked oligosaccharides are bound weakly (3.3.1, 3.5.3). This is consistent with unpublished glycan array screening results generated with JCV virus-like particles (VLPs) by Leonid Gorelik and the CFG (available online at www.functionalglycomics.org). In their experiments, strong binding was observed to LSTc-containing sequences and weak binding was observed for GM2, GM1 and similar branched, α2,3-linked sequences. It is likely that the glycan array screening of JCV presented here, which used unassembled JCV VP1 pentamers, did not pick up these low-affinity interactions due to the lower avidity of pentamers compared to VLPs. The weaker interactions with GM1 and GM2 appear not to be functional because SV40, which does use GM1 as a receptor, does not compete with JCV for binding sites (Liu et al., 1998a), and because GM1 oligosaccharide does not block JCV infection at a concentration of 5 mM (Melissa S. Maginnis, personal communication).

The identification of LSTc as a receptor motif for JCV is also consistent with biochemical assays involving natively glycosylated cells (Liu et al., 1998b). Moreover, the results explain why JCV had been found to bind to α2,3-linked compounds under certain experimental conditions (Dugan et al., 2008). In one experiment, cells were stripped of sialic acid with neuraminidase and glycans were reconstituted using α2,3- or α2,6-specific sialyltransferases, which in each case increased infection. These experiments did not reconstitute the native glycome of the cell, but over- and underrepresented certain glycans according to enzyme specificity. It is likely that the branched α2,3-linked structures that are weakly bound by JCV were overexpressed in the experiment. The overexpression may have compensated for the weaker affinity, allowing JCV to infect the cells.

The LSTc motif is expressed on cell surfaces as a component of glycoproteins and glycolipids. Homologs of LSTc, in which a GlcNAc replaces the terminal Glc (GlcNAc-LSTc) are present in the human body as the termini of long oligosaccharide chains of the 'i' antigen type (Feizi, 1985). All of the glycans terminating in GlcNAc-LSTc contain repeating units of [3-Gal-β1,4-GlcNAc-β1-]$_n$ (polyLacNAc repeats), and fulfill the steric condition of a long carbohydrate that is unbranched at its tip. These polyLacNAc repeats can be found on *N*-linked glycans, mucin-type *O*-linked glycans as well as glycolipids of the neo-lacto series, which are defined by a Gal-β1,4-GlcNAc motif attached to the membrane anchored lactosyl ceramide. They are attached to the mannose-based or GalNAc-based core structures or the ceramide membrane anchor, respectively. It has been suggested that JCV attaches to an *N*-linked glycoprotein (Liu et al., 1998a) since treatment of SVG-A cells with the *N*-glycosylation inhibitor tunicamycin, but not with the *O*-glycosylation inhibitor benzylGalNAc inhibited infection. The serotonin receptor 5-HT$_{2A}$R is necessary for JCV infection and is *N*-glycosylated, but LSTc is likely not attached directly to one of its *N*-glycan chains (Maginnis et al., 2010). *N*-glycan structures are branched, but if the LSTc motif were attached at the tip of one arm of a long, branched *N*-glycan chain so that JCV would not come near the branching point at all, such an oligosaccharide could very likely be used as a receptor by JCV.

### 4.1.3. BKV

The results presented here show that all b-series gangliosides can be used as receptors by BKV (3.4.2-3.4.5). This expands previous reports that showed GT1b and GD1b as BKV receptors that did not test other b-series gangliosides. However, GD2 and GD3 do not bind to BKV as well as GD1b and GT1b, which correlates well with the ability of the respective gangliosides to enhance BKV infection (3.4.4, 3.4.5). The observed differences suggest that regions of GD1b and GT1b that are not present in GD2 and GD3 might make additional contacts with BKV VP1. Since GT1b and GD1b bind equally well, it is likely that BKV can engage in contacts with the Gal residue on the left arm of GD1b and GT1b. However, given that GD3 also supports infection, such contacts would be minor, and only modulate binding in comparison to the main contact point sialic acid. The validity of this interpretation is also demonstrated by mutagenesis studies, which illuminated the crucial role the sialic acid binding site identified here plays in infection (3.4.2.2). A scenario in which GT1b and GD1b attach to BKV VP1 in a completely different manner compared to the binding mode of GD3 is also a formal, although probably unlikely, possibility as it would also explain the observed differences in binding. Such a scenario can only be ruled out through the structure determination of BKV VP1 in complex with the oligosaccharide portions of GT1b or GD1b. Since attachment of one BKV virion probably requires many simultaneous

interactions with carbohydrates and since plasma membranes contain all the different b-series gangliosides, the virus likely engages not only one type of ganglioside with its multiple binding sites, but a mixture of receptors depending on the composition of the plasma membrane.

In addition to b-series gangliosides, a linear motif containing a terminal $\alpha$2,3-linked sialic acid has also been implicated as a receptor for BKV (Dugan et al., 2005). While it is sterically possible that BKV has some affinity for $\alpha$2,3-linked NeuNAc, additional interactions with the second sialic acid would be lacking, and the affinity of the interaction would be expected to be very weak. The $\alpha$2,8-disialic acid motif bound by BKV is not only present on gangliosides, but also on glycoproteins. However, the entry route of BKV via caveolae points to gangliosides as entry receptors, which can transmit the signal for endocytosis via their lipid tails (Ewers et al., 2010) (1.4.1, 4.3.1).

### 4.1.4. Relevance for structure-aided design of BKV and JCV entry inhibitors

The structures of the human BKV and JCV attachment proteins in complex with their receptors enable the design of entry inhibitors. These would ideally be small molecules that bind tightly to the oligosaccharide binding sites of the two viruses, and thus prevent viruses from engaging their receptors. Such a small molecule should engage the sialic acid binding site since it is the main point of attachment for the virus, and confers most contacts. Mutagenesis studies of infectious JCV and BKV showed that single point mutations in the sialic acid binding site are enough to stop the virus from propagating in cell culture (3.3.4.3, 3.4.2.2). While the virus might escape an inhibitor targeted against the periphery of the oligosaccharide binding site by a single mutation, it would be much more difficult to modify the sialic acid binding site without adversely affecting binding. In addition, since the sialic acid binding site is rather conserved in BKV and JCV, it might be possible to develop one inhibitor that is effective against both viruses.

As the virus uses multivalent binding to enter cells, a strategy based on multivalent inhibitors might be applied to achieve high-affinity binding. A pentavalent inhibitor, termed STARFISH, has been synthesized for *E. coli* shiga-like toxin, a pentameric bacterial toxin that binds glycolipids. The multivalent ligands were $10^6$-$10^7$-fold more potent than their monovalent counterparts in inhibiting oligosaccharide binding, and protected cells from the cytotoxic effects of shiga-like toxin (Kitov et al., 2000).

## 4.2. Implications for other polyomaviruses

### 4.2.1. Comparison of SV40, BKV and JCV VP1 with Polyoma VP1

Like SV40, JCV and BKV VP1, Polyoma VP1 binds to sialylated oligosaccharides, namely, the gangliosides GD1a and GT1b (Tsai et al., 2003). In each case, sialic acid serves as a major contact, which is recognized at similar locations on the outer surface of VP1 (Stehle et al., 1996; Stehle and Harrison, 1996; Stehle et al., 1994; Stehle and Harrison, 1997) (3.2.4, 3.3.4, 3.4.1.1, 3.4.1.2). In addition, both SV40 and Polyoma VP1 bind to sialic acid in an $\alpha$2,3-linkage to the underlying Gal residue. Since the sialic acid binding sites of BKV, JCV and SV40 are very similar (3.5.1), comparison of the SV40 oligosaccharide complex with Polyoma will also be representative of the BKV and JCV interactions.

In both SV40 and Polyoma, VP1 residues at equivalent positions are used for interactions with NeuNAc (Fig. 40A,B). However, despite these similarities, the NeuNAc moiety is bound in different orientations by the two proteins. In Polyoma, the NeuNAc carboxylate faces away from the five-fold axis of the pentamer, forming a key salt bridge with R77, and the glycerol side chain points away from the virion into solution. By contrast, the glycerol chain of NeuNAc in GM1 faces towards VP1, whereas its carboxylate group faces towards the five-fold axis and does not engage in a salt bridge that would neutralize its charge (Fig. 40C,D). The residue equivalent to R77 in SV40, K67, does not directly contact NeuNAc but instead forms the ridge that separates the binding pockets for the two branches of GM1. The conformation of R77 in Polyoma is stabilized by Q59 and Y72, whereas K67 in SV40 is held in place by a salt bridge to D81.

The Polyoma protein is specific for NeuNAc in an $\alpha$2,3-linkage to the underlying Gal residue because an equatorial hydroxyl group at C4 of the second sugar ring such as in Glc would clash with the main chain of the BC-linker. In addition, any residue but G at position 78 in Polyoma would clash with with the observed conformation of the oligosaccharide. Moreover, the conformation of the NeuNAc-$\alpha$2,3-Gal linkage differs between the Polyoma and SV40 complexes. The conformation of NeuNAc and Gal in the Polyoma binding site does not support a branch at the axial O4 of the Gal residue, whereas this branch and the rigid conformation it imposes are recognized by SV40. Thus, residues at equivalent positions in the sequence and in space have different contacts in SV40 and Polyoma, leading to different surfaces and ligand specificities.

Taken together, these differences in sialic acid binding between Polyoma and SV40 (and consequently also BKV and JCV) explain why attempts to model the interaction of SV40, BKV and JCV VP1 with sialylated compounds yielded results that are incompatible with our crystallographic analyses (Gee et al, 2004; Dugan et al., 2007; Campanero-Rhodes et al., 2007; Sunyaev et al., 2009). All of these were based on the complex structures of

Polyoma with sialylated compounds, which differ significantly from the SV40, BKV and JCV binding sites and therefore were unsuitable starting points.

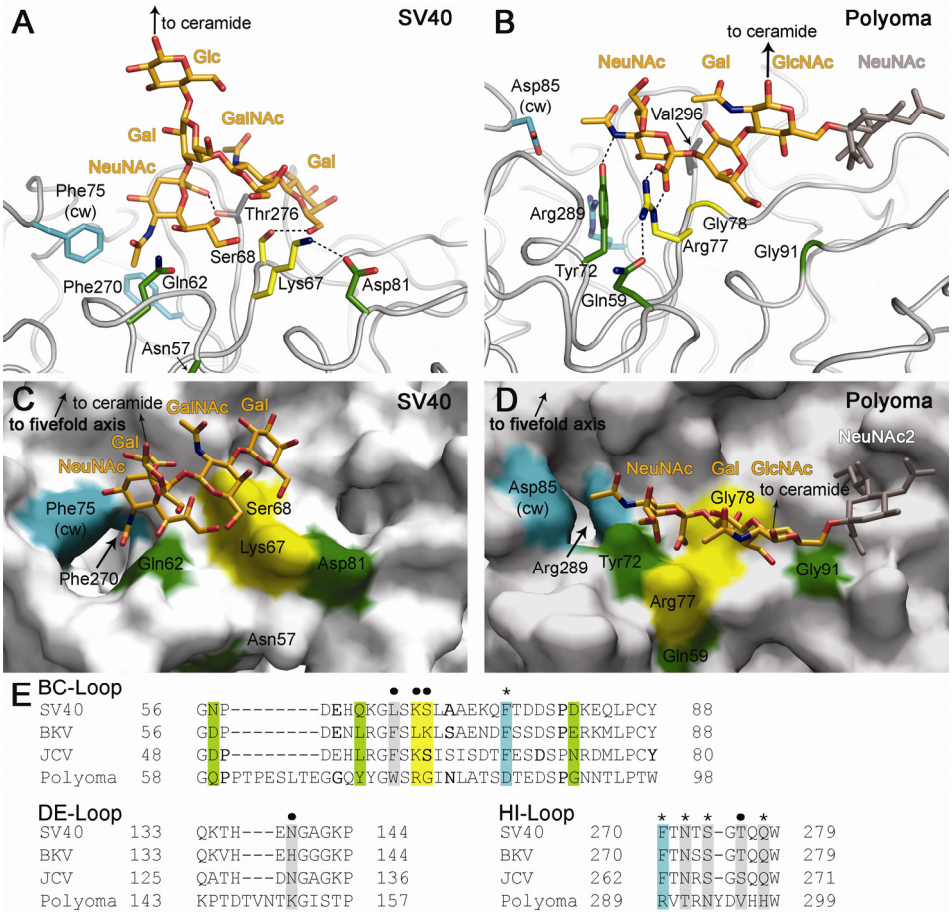



**Fig. 40: SV40 and Polyoma bind their ganglioside receptors in different orientations and conformations.** (A) and (B) Comparison of specific interactions between VP1 from SV40 (A) and Polyoma (B) with their ligands. Polyoma was crystallized with an oligosaccharide that contained a fragment also present in GD1a and GT1b (shown in color). The additional α2,6-linked NeuNAc present in the ligand, but not on a ganglioside receptor, is shown in grey. The same view from the side of VP1 was used for both proteins, and residues at equivalent positions in both proteins have the same color. The RG motif of Polyoma and its SV40 counterpart are colored yellow, the residues holding them in place are green. Hydrophobic residues lining the deep cavity in SV40 and their charged Polyoma equivalents are turquoise. Additional residues contacting NeuNAc in both complexes are grey. (C) and (D) Different binding surfaces of SV40 (C) and Polyoma VP1 (D) for their ligands. The same view from top of the pentamer was used for both complexes, and the color scheme is as in (A) and (B). The terminal Glc was omitted from the SV40 VP1-GM1 complex for clarity. (E). Structure-based alignment of the VP1 sequences of SV40, Polyoma, BKV and JCV. SV40 residues identical to both BKV and JCV are marked with an asterisk, those identical to either BKV or JCV are marked with a dot. The color scheme is as in (A) and (B), with additional SV40 residues in contact with GM1 in grey.

## 4.2.2. Comparison with MCV VP1

The structure of MCV VP1 was recently solved at 1.8 Å in complex with 3'-sialyllactosamine (Holger Hengel, Ursula Neu, Thilo Stehle, unpublished data). In this complex, the terminal sialic acid serves as major contact point as in the Polyoma and SV40 complexes. Interestingly, the sialic acid is again bound at the same position on VP1, but in a new orientation that differs from the two previously known ones (Fig. 41).

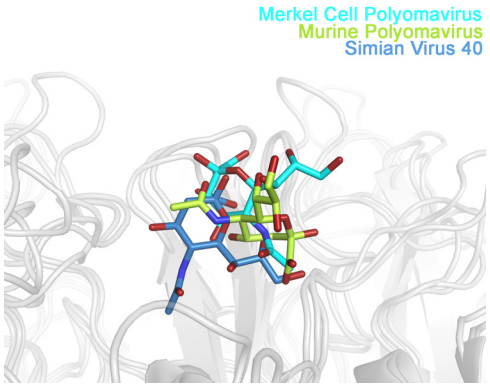



**Fig. 41: Comparison of sialic acid orientations in complexes of MCV, Polyoma and SV40 VP1**

VP1 structures were superposed onto each other with Superpose (CCP4, 1994). VP1 is shown in cartoon representation and colored grey. Sialic acid is shown in stick representation and colored turquoise, green and blue for the MCV, Polyoma and SV40 complexes, respectively. In addition, oxygen atoms are colored red and nitrogen atoms are colored blue.

In the MCV complex, the sialic acid is rotated by 90 ° in the counterclockwise direction with respect to its orientation in the SV40 VP1 complex. The glycerol chain therefore points towards the DEccw-loop of the counterclockwise neighboring VP1 monomer. The carboxylate group is oriented towards the central five-fold symmetry axis as in the SV40 complex, but points with a somewhat different orientation. The *N*-acetyl group points away from the five-fold symmetry axis in the MCV complex.

This site on polyomavirus VP1 proteins therefore seems to be a flexible platform for sialic acid binding that can be adapted not only to different sialic acid types and underlying linkages, but also to binding sialic acid in different orientations. It is therefore possible that other sialic-acid binding polyomaviruses such as LPV bind sialic acid in yet another orientation.

### 4.2.3. Are there fold requirements for oligosaccharide binding?

The structures of KIV and WUV VP1 and the comparison to the distantly related SV40 and Polyoma VP1 highlight the role of the VP1 core as a stable carrier for flexible loop arrangements (3.7.3). The conserved BC-linker identified here likely stabilizes the structures of the BC1- and BC2-loops. However, the two residues of the linker (x in the $\phi$ – S/T – x – x – $\phi$ motif) that make no interactions with other amino acids, but protrude into solution play important roles in oligosaccharide binding. In Polyoma, these residues form the RG motif that recognizes the NeuNAc-$\alpha$2,3-Gal moiety. In BKV and SV40 VP1, the specificity-defining residue K/S68 (S60 in JCV VP1) is the second amino acid denoted x. In the recently solved complex of MCV VP1 with 3'-sialyllactose, the two residues of the BC-linker are both involved in contacts with sialic acid as well.

There are several reasons why this might be the case. For instance, these residues might be relatively easy to mutate without disrupting the correct fold of the protein. On the other hand, the kinked all-$\beta$ conformation of the linker and of the adjacent I'-strand might be conducive to oligosaccharide binding. Interestingly, a superficial comparison of lectin families

frequently revealed oligosaccharide binding sites located on or at the edge of antiparallel b-sheets, often complemented by loop regions. No purely α-helical oligosaccharide binding site was found, and binding sites with contributions by α-helices are relatively rare. Furthermore, many carbohydrate-metabolizing enzymes and lectins such as galectins and L-type lectins feature jelly-roll folds, which are relatively rare except for viral capsid proteins. The dominance of the β-conformation held true for all oligosaccharide binding proteins, not only for those specific to sialic acids. It is tempting to speculate that the particular spacing of main chain atoms on antiparallel β-sheets is in line with the spatial requirements of carbohydrate recognition. However, a much more in depth bioinformatics analysis is necessary to obtain safe conclusions on this matter.

### 4.2.4. Possible receptors for WUV and KIV

As there is currently no cell-culture data on WUV and KIV, nothing is known about their receptors. None of the sialic-acid binding residues of SV40, Polyoma or MCV are conserved in the WUV and KIV VP1 structures. In addition, their BC1-loops point into a different direction than those of SV40 and Polyoma, which are involved in sialic acid binding. Moreover, the residues at the tip of the HI-loop that contact the oligosaccharide in the SV40, Polyoma and MCV complexes are not present in WUV and KIV VP1, whose HI-loops are five and six amino acids shorter, respectively. Thus, it can be ruled out that either of the two proteins binds sialic acid in the same way as SV40, Polyoma or MCV. If they bind sialic acid, it is in an orientation that differs from the three observed ones. However, it is entirely possible that they do not bind sialic acid at all. WUV VP1 features a large, positively charged patch on the surface formed by residues of the HI- and BC2cw-loops, which might suggest binding to highly charged molecules such as sulphated glycosaminoglycans. On the other hand, the rugged surface of KIV VP1 might indicate a peptide receptor or a very long carbohydrate chain. Unfortunately, the structural data presented here only allows very few conclusions about receptor binding until receptors are identified by other means, such as glycan array screening, which is currently ongoing, soaking trials with random oligosaccharides, glycolipid pull-down assays, or infection assays in cell culture.

### 4.2.5. Implications for hPyV6 and hPyV7 VP1

The newly identified human polyomaviruses hPyV6 and hPyV7 share homology with the WUV and KIV VP1 proteins, but are more distant to other polyomavirus VP1 proteins (Fig. 1). Alignment of their sequences to other VP1 sequences reveals the presence of a sequence motif consistent with the structure of the BC-linker between the B- and C-strands of hPyV6 and hPyV7 (Fig. 39). If this sequence were to form a BC-linker in these two proteins, they would feature a very long BC1-loop and an extremely short “BC2-loop” of four

residues that would link the BC-linker directly to the C-strand. In addition, the sequences of hPyV6 and hPyV7 VP1 both contain HI-loop sequences that are much longer than for all other polyomaviruses. It is therefore possible that these long HI-loops fold back on the VP1 core in these two proteins and make up the surface of VP1 that is formed by the BC2-loops in other VP1 proteins. Structure determination of hPyV6 and hPyV7 VP1 is currently ongoing, and will test this prediction.

## 4.3. Polyomavirus properties determined by receptor specificity

### 4.3.1. Entry pathways

In contrast to SV40, Polyoma, and BKV, JCV does not enter cells by cholesterol-dependent endocytosis, but instead uses clathrin-mediated endocytosis (1.4.1). It is enticing to speculate that this difference is reflected in the different types of carbohydrate receptors used by these viruses. SV40, Polyoma, and BKV all use gangliosides (1.5), which, at least in the case of SV40, initiate entry by inducing curvature of the plasma membrane (Ewers et al., 2010). To induce such curvature upon binding, the oligosaccharide portion of the receptor must be somewhat rigid, which is the case for short, branched gangliosides (Acquotti et al., 1994; Brocca et al., 1998; Richardson et al., 1995), but not for long, flexible polyLacNAc extensions, to which LSTc is attached (Sabesan et al., 1991). Thus, the few amino acid changes between JCV and related polyomaviruses might target it not only to a different receptor, but also to a different entry pathway. Given the residual affinity of JCV for gangliosides such as GM2 (3.5.3, 4.1.2), it is likely that JCV arose from a ganglioside-binding ancestor that used cholesterol-dependent endocytosis, and acquired the different entry pathway upon mutations in the receptor binding site that enabled it to bind LSTc.

How viral uptake by cholesterol-mediated endocytosis is triggered upon attachment to gangliosides has been elucidated recently (Ewers et al., 2010). However, how attachment to LSTc triggers endocytosis of JCV by the clathrin pathway in unclear, especially since the serotonin receptor likely does not bind directly to JCV (Maginnis and Atwood, unpublished data). In addition, LSTc is likely not attached to the serotonin receptor (Maginnis et al., 2010). As is the case for SV40, attachment of JCV to cells likely involves several copies of glycoproteins or gangliosides, and would make use of many of the 360 available oligosaccharide binding sites on the virus particle. This scenario is reminiscent of galectin-induced glycoprotein lattices on cells. Galectins are mostly β–Gal specific lectins that often attach to polyLacNAc chains on *N*-glycans. They either exist as bivalent lectins or multimerize, which enables them to organize their *N*-glycan ligands into lattices on the cell surface. Such clustering of cellular receptors often modulates signaling processes as observed for instance in cells of the immune system or insulin producing cells (Rabinovich et

al., 2007). The multivalent JCV particle is very likely able to cluster its receptors into lattices as well, with consequences on modulating the signal transduction of host cells. Thus, an oligomeric cluster of JCV receptors might modulate serotonin receptor signaling in a way conducive to infection even though LSTc is likely not directly linked to the serotonin receptor. This would be consistent with the findings that serotonin receptors colocalize with JCV particles on SVG cells, and that signalling by the serotonin receptor is crucial for JCV uptake (Elphick et al., 2004).

### 4.3.2. Tissue tropism

#### 4.3.2.1. Polyoma

Oligosaccharide receptors can be crucial for viral spread and tropism as demonstrated by a point mutation in Polyoma VP1 that alters its receptor specificity and dramatically changes the virus' tropism and pathogenesis (1.2.5). The field isolate RA strain carries a glycine at position 91 in VP1 and is only moderately pathogenic. The PTA strain, which carries a glutamate at position 91, causes a variety of tumors in different organs of mice. The point mutation in VP1 is located in the oligosaccharide binding site of Polyoma VP1 and gives rise to different carbohydrate specificities. The RA strain binds to both linear $\alpha$2,3-linked sialylated compounds such as NeuNAc-$\alpha$2,3-Gal-$\beta$1,3-GlcNAc and to branched oligosaccharides featuring a second, $\alpha$2,6-linked sialic acid such as NeuNAc-$\alpha$2,3-Gal-$\beta$1,3-[NeuNAc-$\alpha$2,6-]-GlcNAc. By contrast, the more pathogenic PTA binds only to linear $\alpha$2,3-linked sialylated oligosaccharides (Bauer et al., 1999; Cahan et al., 1983; Freund et al., 1991b; Fried et al., 1981; Stehle and Harrison, 1996, 1997; Stehle et al., 1994). The functional receptors for Polyoma are the gangliosides GD1a and GT1b, which both carry $\alpha$2,3-linked sialic acid on their left arm (Smith et al., 2003; Tsai et al., 2003). Since the PTA strain exhibits faster spread than the RA strain both in tissue culture and in mice, it is likely that the branched $\alpha$2,6-linked oligosaccharides act as decoy ligands that delay viral spread and are not conducive to infection.

#### 4.3.2.2. BKV

The finding that b-series gangliosides are uniquely important for BKV infection may have implications for BKV tropism and pathogenesis. Biochemical analysis of kidney glycosphingolipids indicates that the kidney is rich in diverse sphingolipids and is particularly abundant with gangliosides (Holthofer et al., 1994; Shayman and Radin, 1991). Furthermore, GD3, GD2, GD1b and GT1b are resident gangliosides of the human adult kidney and are found differentially expressed in cortical tubular, medullary and glomerular tissues (Holthofer et al., 1994). The developmental expression of these gangliosides has not been evaluated in human kidney tissue, but studies with bovine kidney indicate that there are distinct patterns

of ganglioside distribution associated with fetal, newborn and adult kidneys, with simpler gangliosides such as GD3 being readily abundant in fetal kidney tissue, and more complex gangliosides such as GD2, GD1b and GT1b being increasingly expressed in the newborn and adult kidney. A similar expression pattern is seen for ganglioside expression in the developing brain of mice and humans (Yu et al., 2009). The observation that complex gangliosides such as GD1b and GT1b are more efficient in supporting infection leads us to question whether developmental changes associated with ganglioside distribution may play a role in latency and reactivation. This is highlighted by the fact that SV40 and Polyoma are more pathogenic in newborn animals – they would consequently attach to gangliosides that are more prominent in tissues of newborn animals.

Furthermore, a search of the more than 1000 DNA sequences for BKV VP1 in GenBank revealed that the K68S mutation had not yet been identified in nature. Some sequences feature an arginine or histidine in place of the lysine, but these are unlikely to switch the receptor binding properties of BKV. While this might have a number of reasons, many of which are not linked to receptor binding, it is tempting to speculate that GM1 is not a useful receptor for BKV that might target it to tissues where BKV cannot propagate as efficiently.

#### 4.3.2.3. JCV

The tissue tropism of JCV may also be linked to its oligosaccharide specificity. Glycan profiling by mass spectrometry has identified long glycans consistent with the LSTc motif in the *N*-glycan fractions of several human tissues including kidney, lung, and spleen (http://www.functionalglycomics.org), but tissues of the entire CNS were not analyzed. In addition, the 'i' antigen is expressed on a high proportion of human peripheral lymphocytes (Feizi et al., 1980). Lectin staining of human tissues revealed that $\alpha$2,6-linked NeuNAc is present on the surface of B lymphocytes of the tonsils and spleen, on kidney and lung cells, as well as on astrocytes and oligodendrocytes (Eash et al., 2004b). Only very low-level lectin staining was seen for neurons and T lymphocytes. Notably, neurons, which express the serotonin receptor, are not susceptible to infection with JCV, most likely because they lack the oligosaccharide receptor. As the expression pattern of LSTc-carrying glycans corresponds to sites of JCV persistence, the kidney, and pathogenesis, the glial cells of the CNS, it is likely that the observed specificity of JCV for LSTc contributes to targeting the virus to these sites.

Statistical analyses of JCV sequences demonstrate that 52% of PML patients carry JCV with one mutation in VP1 (L54F, K59M/E/N, N264D/T, S266F/L, or S268F/Y/C) that likely arose from positive selection (Sunyaev et al., 2009). Residues L54, N264, S266 and S268 mediate direct contacts with NeuNAc (Fig. 42). Mutation of any of these four residues therefore adversely affects sialic acid binding (Fig. 42, Fig. 17). Consistently, the L54F and

S268F mutants have reduced hemagglutination activity (Sunyaev et al., 2009). Although residue K59 also contacts NeuNAc, its replacement with M, E or N would still allow for similar contacts. Not all of the mutations can therefore be attributed to a selective pressure on the virus to alter its receptor binding site.

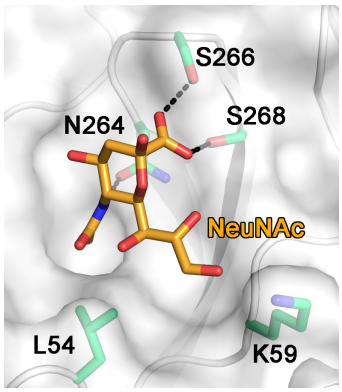



**Fig. 42: JCV residues whose mutation is attributed to positive selection.** JCV VP1 is shown in cartoon and surface representation and is colored white. The sialic acid of LSTc is shown in stick representation. Side chains of the residues whose mutation seems to be positively selected in PML patients are shown in stick representation. Nitrogens atoms are colored blue, oxygens red and carbons green and orange for JCV VP1 and LSTc, respectively.

It is interesting to speculate about how mutations that should block LSTc binding would be selected for in individuals with PML. It is conceivable that JCV featuring a mutated oligosaccharide binding site can spread more rapidly to the CNS from its site of persistence in the kidney because of diminished attachment to carbohydrates. This scenario would be similar to a point mutation in Polyoma VP1 that abolishes binding to non-functional pseudo-receptors, thereby increasing viral spread and rendering the virus more pathogenic (Freund et al., 1991b). However, a defect in sialic acid binding would abolish engagement of both LSTc-carrying glycans and branched $\alpha$2,3-linked compounds. Unpublished glycan array screening data by Leonid Gorelik and the CFG confirms this (www.functionalglycomics.org). One also cannot ascertain whether JCV develops mutations in VP1 while replicating in remote sites of infection such as the kidney or if they develop in the brain. Moreover, it is not clear if patients carrying mutant JCV also carry wild-type JCV. Finally, viruses with the mutations L54Y, S266A/S268A, and S266N/S268N in the sialic acid binding site were unable to infect cells and propagate in a long-term growth assay because of a cell binding defect (3.3.4.3). Since the same residues that were identified in PML patients were mutated, albeit to different residues, virus carrying the mutations present in PML patients would be predicted to not propagate efficiently in glial cells.

In addition, the activity of glycosyl transferases is tightly regulated and is involved in many immunomodulatory processes. It is tempting to speculate that immunosuppression, which is a prerequisite of PML development, might result in an upregulation of LSTc (e.g. on lymphocytes), which enables JCV to spread to sites of pathogenesis.

### 4.3.3. Host tropism

The data on JCV, BKV, SV40 and K68S VP1 presented here suggests that preference of these viruses for different sialic acid types contributes strongly to their host range. Preliminary results from the further characterization of the BKV K68S mutant in cell culture strengthen this point (Stacy-ann A. Allen & Walter J. Atwood, personal communication).

His-tagged K68S pentamers were subjected to a flow cytometry-based cell-binding assay with simian Vero and human HEK cells supplemented with NeuNAc-GM1 ganglioside. NeuNAc-GM1 supplementation increased binding of K68S pentamers to human HEK cells. It did not increase binding of BKV pentamers that were used as a control. The K68S binding was specific for GM1 because competition with cholera toxin B, which is known to bind to GM1 with sub-micromolar affinity, blocked binding of K68S. NeuNAc-GM1 supplementation of simian Vero cells did not increase binding of K68S, indicating that NeuNAc-GM1might be metabolized to NeuNGc-GM1 in these cells.

The K68S mutation was also introduced into an infectious clone of BKV, to assess its growth and infection properties. The K68S mutant did not propagate in the simian Vero cell line, not even when cells were supplemented with NeuNAc-GM1. However, K68S propagated like BKV in human HEK cells, regardless of whether cells were supplemented with NeuNAc-GM1 or not. Taken together, these results demonstrate that K68S attaches specifically to NeuNAc-GM1 on host cells, and suggest that its specificity for one sialic acid type is linked to its host range. However, quantitative data on the relative concentrations of the different gangliosides in the plasma membrane would be very helpful to rule out that the observed effects stem from different concentrations of GM1 and b-series gangliosides. Moreover, the presence of the different sialic acid types on these cells should be verified either by mass spectrometry or by staining with antibodies specific for different sialic acid types. Since the current data stem from only two cell lines, more different cell lines, or primary cultures if available, should be investigated.

To corroborate the structural basis of the different sialic acid type specificities of BKV, JCV and SV40, it might be helpful to introduce mutations such as A166F and L65F into SV40 VP1 pentamers. These mutations would diminish the size of the sialic acid type-specific cavity of SV40 VP1 to the size observed in BKV and JCV and therefore be expected to render SV40 VP1 specific for NeuNAc-GM1.

The structural basis for the species tropism of BKV, JCV and SV40 proposed here differs from the determinants of species specificity in influenza viruses, which also attach to sialylated oligosaccharides and use sialic acid as major contact point. The adaptation of avian HAs to human hosts occurs by a switch in binding specificity from avian-like $\alpha$2,3-sialylated glycans to the longer $\alpha$2,6-sialylated glycans that are present in the upper respiratory tract epithelia of humans (Russell et al., 2006; Skehel and Wiley, 2000; Stevens

et al., 2006; van Riel et al., 2007). Species tropism is thus determined by linkage specificity, not by specificity for a different sialic acid type.

On a related note, most polyomaviruses known so far infect different groups of mammals, in which they cause disease mostly in immunocompromised individuals. By contrast, one group of polyomaviruses infects birds, in which the viruses cause much more severe diseases such as the hemorrhagic fevers caused by Goose hemorrhagic fever polyomavirus (Bozeman et al., 1981; Guerin et al., 2000). It is conceivable that polyomaviruses have switched to birds as hosts relatively recently in evolution. This might explain why disease is so severe, similarly to avian viruses, such as avian influenza viruses, that cause severe disease in humans. Perhaps the properties of avian polyomaviruses can be explained by a similar switch in receptor binding specificity.

## 4.4. Recognition of sialylated oligosaccharides by viral attachment proteins

High-resolution structural information on the interaction of viral attachment proteins with sialylated carbohydrates is available for the following systems: Influenza virus A hemagglutinin (Influenza HA) in complex with oligosaccharides containing $\alpha$2,3-linked and $\alpha$2,6-linked NeuNAc (Eisen et al., 1997; Gamblin et al., 2004; Ha et al., 2001, 2003; Russell et al., 2006; Stevens et al., 2006; Stevens et al., 2004; Weis et al., 1988), Polyoma VP1 with a fragment of ganglioside GD1a (Stehle and Harrison, 1997), rhesus, swine and human rotavirus VP8* (Rotavirus VP8*) with methyl-$\alpha$2,3-sialoside (Blanchard et al., 2007; Dormitzer et al., 2002b), adenovirus Ad37 fiber knob (Ad37) with $\alpha$2,3- and $\alpha$2,6-sialyllactose (Burmeister et al., 2004) and SV40, BKV and JCV VP1 with their respective receptors (Neu et al., 2008; Neu et al., 2010) (3.4.2). Paramyxoviruses bind to their host cells with their haemagglutinin-neuraminidase protein that uses the same binding site for attachment to NeuNAc and cleavage of its glycosidic bond (Crennell et al., 2000). Since the sialic acid binding site of these proteins reflects on both of its functions, they were not considered further in this comparison.

Although most of these attachment proteins are not structurally homologous and found on unrelated enveloped or non-enveloped viruses, there are some general similarities in the modes of attachment to sialylated oligosaccharides. First, NeuNAc is the major point of contact in all complexes compared. Second, all viral attachment complexes bind NeuNAc in shallow depressions rather than deep grooves on the protein surface. Third, the attachment protein complexes all feature NeuNAc in its chair conformation, which is the preferred conformation in solution. Fourth, there are similarities in the specific contacts with sialic acid.

### 4.4.1. Contacts with sialic acid

Sialic acid possesses three protruding, recognizable functional groups – carboxylate chain, N-acetyl or N-glycolyl group, and glycerol chain. Most of these are not or rarely present in simple saccharides such as galactose, and they are not as easy to replace with water molecules as the hydroxyl groups differentiating most other sugars.

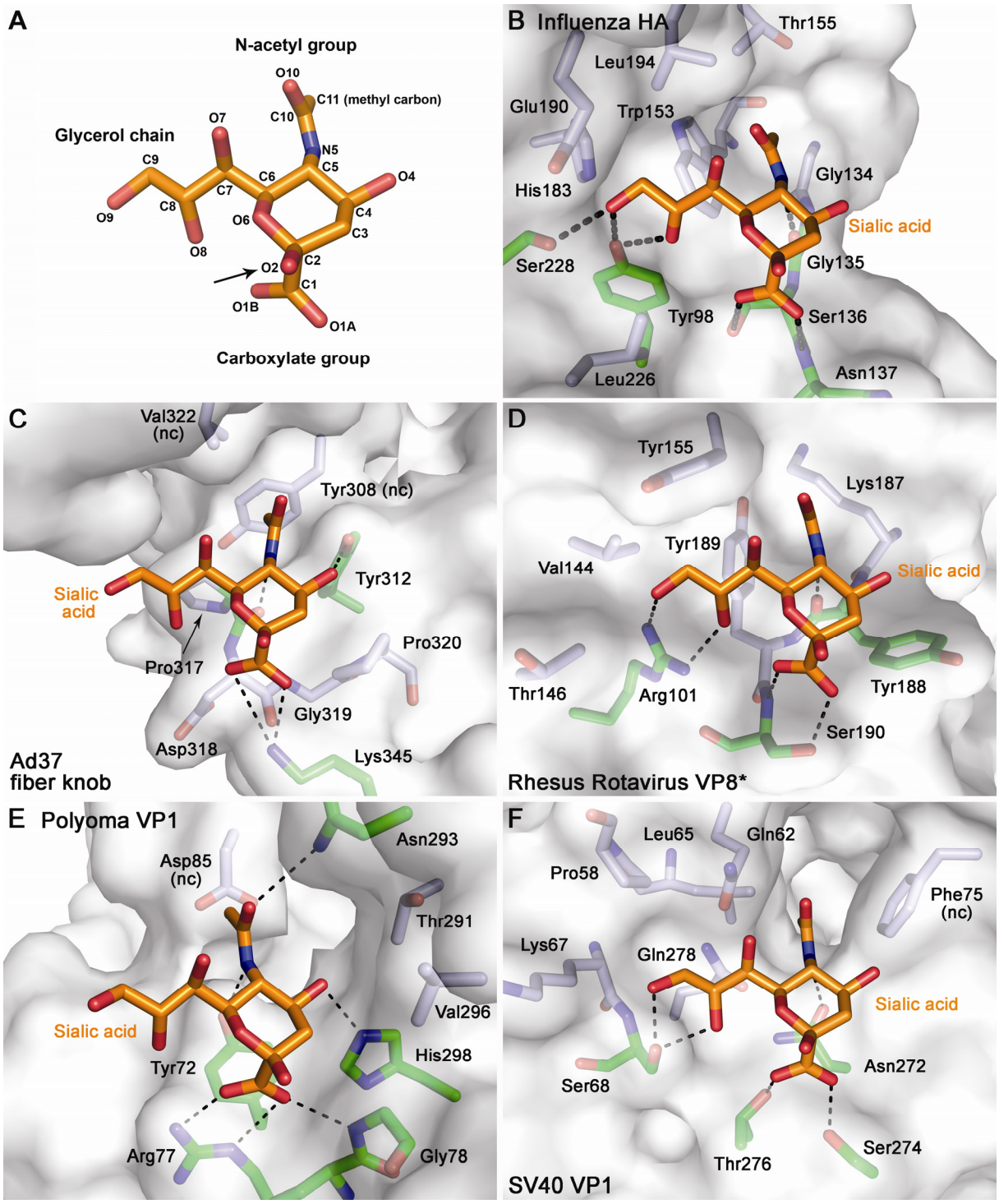



**Fig. 43: Sialic acid recognition by viral attachment proteins.**
(A) Schematic of N-acetyl neuraminic acid.
(B)-(F). Contacts of NeuNAc with Influenza HA (B), Ad37 (C), Rotavirus VP8* (D), Polyoma VP1 (E) and SV40 VP1 (F). NeuNAc is depicted in orange, the protein residues forming hydrogen bonds to NeuNAc are colored green and the residues making van der Waals contacts are colored grey. The viral surface is shown in grey.

All viral proteins investigated here engage the negatively charged carboxylate group of sialic acid, but they employ two different strategies. On the one hand, Polyoma VP1 and Ad37 form salt bridges from the positively charged side chains of R and K residues to the carboxylate group (Fig. 43C and E, respectively). On the other hand, Influenza HA, SV40 VP1 and Rotavirus VP8* donate two parallel hydrogen bonds to the carboxylate group. Hydrogen bond donors are either the hydroxyl groups of S and T side chains or backbone amido groups (Fig. 43B, D, F). For each of the parallel hydrogen bonds, the angle between the carboxylate C1, the carboxylate O that is engaged and the hydrogen bond donor approaches 120°.

The N-acetyl group, situated on the opposite side of the sugar ring as the carboxylate group, serves as additional “handle” on NeuNAc and it provides specificity by additional favourable contacts with the correct ligand. The N5 atom of the N-acetyl group is contacted by a hydrogen bond in each of the complexes. Polyoma VP1 forms an additional hydrogen bond to the carbonyl group of the acetyl function. With the sole exception of Polyoma VP1, the binding sites of all proteins also feature a depression into which the methyl group of the N-acetyl moiety inserts. This depression is shallow and rather hydrophobic in the case of Ad37, Influenza HA, Rotavirus VP8*, as it is lined with hydrophobic side chains (Fig. 43B, C, D). SV40 deviates from this pattern because it preferably recognizes oligosaccharides terminating in NeuNGc (Campanero-Rhodes et al., 2007).

Most attachment complexes feature contacts with the polar and flexible glycerol moiety. This group protrudes away from the protein in the Polyoma VP1 and Ad37 complexes (Fig. 43C and E), but is recognized by a number of contacts in the other complexes. There are hydrogen bonds to O8 and O9 by each of SV40 VP1, Influenza HA and Rotavirus VP8* (Fig. 43B, D, E). In addition, the glycerol chain lies in a shallow pocket in all of these complexes and makes van der Waals contacts to residues that confer shape specificity.

### 4.4.2. Low affinity with high avidity

The affinities of attachment proteins for their oligosaccharide receptors are only in the millimolar range, which is consistent with the relatively small and exposed binding surfaces (Burmeister et al., 2004; Dormitzer et al., 2002a; Eisen et al., 1997; Neu et al., 2008; Stehle and Harrison, 1996). Binding interactions occur in the rather polar environment of the protein surface. Polar and charged interactions are therefore probably quite weak. Hydrogen bonds to an oligosaccharide can be replaced by hydrogen bonds to water molecules without much difference in energy.

The low affinities of viral proteins for their glycan receptors may have a role in providing specificity. The shallow binding sites of attachment proteins do not exclude binding of all incorrect ligands by steric clashes. Specificity is also provided by favorable interactions that

are spaced in a way that only a correct ligand can engage in all contacts. In the cases of SV40, BKV and JCV VP1, sialic acid by itself is not sufficient for binding (3.2.1, 3.3.1, 3.4.4) although sialic acid sterically fits into the binding site. Thus, additional sugars are needed for the interaction.

Since all of the investigated viruses feature multiple copies of their attachment proteins, they bind to cell surfaces in a highly cooperative way, which greatly increases the avidity of the interaction. For example, an SV40 virion attaches to cells with a dissociation contant of 3.78 pM despite the millimolar affinity of each binding site (Clayson and Compans, 1988). Due to membrane fluidity, a virus particle may even organize its own high-affinity binding site by recruiting and engaging more and more receptors. These effects render the binding curve of a complete virion very steep, which amplifies the effect of small modulations in each binding site. In light of the highly cooperative cell attachment process, the relatively low affinity of each binding site may in fact be crucial for viral spread as extremely tight attachment to the host cell will prevent progeny virus from being released.

### 4.4.3. Switching of specificity between different sialylated oligosaccharides

SV40, JCV and BKV all feature a conserved platform of core residues that allows them to engage terminal sialic acid in a similar manner. Nevertheless, each of these viruses achieves its distinct receptor specificity (Fig. 44) with a small number of carefully positioned satellite residues that form very few contacts with additional carbohydrate moieties (3.5.1, 3.5.3). Likewise, a mutation outside the main sialic acid binding site acts as a switch for the receptor specificity of Polyoma VP1 that determines spread and virulence of the virus (1.2.5, 4.3.2.1).

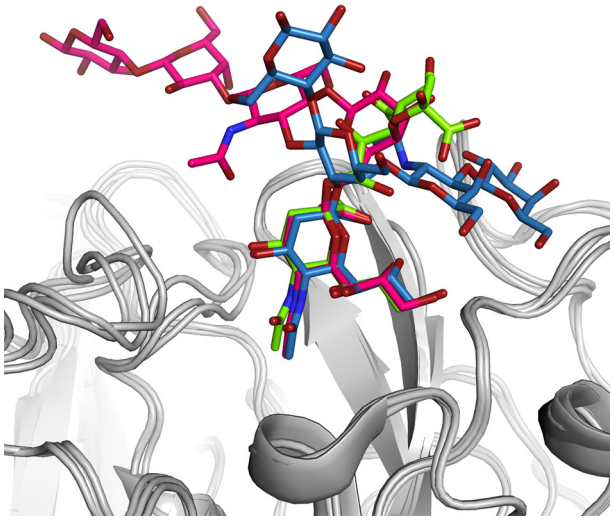


**Fig. 44: BKV, JCV and SV40 VP1 share a common sialic acid binding site, but attach to sialic acid in different linkages.**
VP1 proteins are depicted in cartoon representation and colored grey. The oligosaccharides are shown in stick representation and colored pink (LSTc), green (a2,8-disialic acid) and blue (GM1). Oxygen atoms are colored red and nitrogen atoms are colored blue.

Although they bind to sialylated oligosaccharides in a much more promiscuous way, a similar strategy is employed by influenza HA from different strains that recognize sialic acid in different contexts. Avian strains prefer short α2,3-linked glycans while human strains recognize longer α2,6-linked oligosaccharides. Like in SV40, JCV and BKV, NeuNAc is bound in the same orientation and by the same residues and specificity is conferred by satellite residues in contact with additional monosaccharides. Recent data indicate that that the different overall topology of the glycans is recognized (Chandrasekaran et al., 2008). Given the preferred torsion angles of the NeuNAc-α2,3-Gal-β1,3/4-GlcNAc motif typical of short α2,3-linked glycans, the Gal-β1,3/4-GalNAc part is fairly linear and can occupy the space delineated by a cone with NeuNAc at its tip. Sugar moieties farther away from the NeuNAc mostly do not interact with the protein in this conformation. Glycans terminating in α2,6-linked NeuNAc can also adopt topology similar to the cone-shaped one. However, these also sample a topology with a kink at the Gal position, including the one observed in the JCV VP1-LSTc complex (3.3.4.1). The space that can be sampled by the glycan resembles an umbrella, with NeuNAc-α2,6-Gal as the holder and the rest of the glycan as the umbrella. α2,6-linked trisaccharides that can sample both cone- and umbrella-shaped topologies prefer the cone-shaped one. However, longer α2,6-linked glycans prefer the umbrella-shaped topologies, leading to additional contacts with the protein.

In the attachment proteins of both polyomaviruses and influenzaviruses, the NeuNAc-binding core residues mediate the vast majority of interactions, and would therefore mostly account for the binding affinity. While the satellite residues in some cases only add one or two contacts and thus would contribute little to affinity, they define the context in which a terminal sialic acid can be bound.

## 4.5. Comparison with non-viral lectins

Many lectins of different types can bind to sialylated oligosaccharides. Two examples are discussed here because they feature interesting parallels to polyomavirus VP1 proteins.

### 4.5.1. Interaction of Siglecs with sialylated oligosaccharides

Siglecs are a group of mammalian immune signalling molecules with immunoglobulin (Ig) superfamily folds. They are single-span transmembrane proteins, whose extracellular portions contain a variable number of C2-set Ig domains and terminate in a modified V-set Ig domain that binds to sialylated oligosaccharides. The presence of sialic acid on oligosaccharides seems to be a crucial requirement for engaging all Siglec family members. This is unusual for an entire family of non-viral lectins. Most types of lectins use a monosaccharide other than sialic acid as primary contact. Crystal structures of two Siglecs

with sialylated receptors reveal a common sialic acid binding site whose residues engaging sialic acid are conserved in all Siglec sequences (Attrill et al., 2006; May et al., 1998). An arginine residue makes a salt bridge to the sialic acid carboxylate group, and a tryptophane side chain stacks against the unpolar face of the sialic acid glycerol chain. In addition, there are conserved main chain hydrogen bonds to the N-acetyl and glycerol functions. The Siglec sialic acid binding site is rather shallow, and presumably mediates low-affinity interactions, which is similar to viral sialic acid binding sites. Given the relatively large contact surfaces between entire cells, it is likely that multivalency plays an important role in Siglec-mediated cell-cell contacts. Siglecs certainly differ in their specificities for sialylated compounds, some of which are gangliosides, but existing data do not allow clear conclusions about the structural determinants of their different binding specificities.

### 4.5.2. Bacterial $AB_5$ toxins

$AB_5$ toxins such as cholera toxin, shiga toxin and heat-labile enterotoxin consist of a pentamer of B-subunits, which mediate receptor binding, and a single A-subunit that possesses enzymatic activity. The $B_5$ subunit of these toxins forms a ring-shaped pentamer that binds the monomeric A-subunit at its center, similar to a polyomavirus VP1 pentamer one molecule of VP2/3 (Barouch and Harrison, 1994; Merritt and Hol, 1995). Both proteins therefore exhbibit an unusual $AB_5$ symmetry mismatch (Fig. 45A, C).

Like SV40, BKV and Polyoma, AB5 toxins attach to glycolipids and enter cells via cholesterol/lipid rafts-dependent endocytosis, which is rather uncommon among infectious agents. Both cholera toxin, which uses GM1 as a receptor, and shiga toxin, which binds to the non-sialylated globosylceramide, induce membrane curvature in a way similar to SV40 virions and unassembled pentamers of SV40 (Ewers et al., 2010; Romer et al., 2007). In both VP1 and toxin receptor complexes, the glycolipid receptors are bound on a pentamer surface perpendicular to the five-fold axis (Merritt et al., 1994). The spacing between neighboring binding sites on the pentamer is quite similar in both cholera toxin-GM1 and SV40 VP1-GM1 complexes with 43 Å and 41 Å distance from the glucose moiety of GM1 to its neighbor, respectively. Neighboring binding sites in shiga toxin are separated by 38 Å. The specific spacing of oligosaccharide binding sites on the pentameric glycolipid-binding proteins seems to be crucial for induction of membrane curvature since anti-GM1 antibodies do not induce curvature.

During cell entry, both bacterial toxins and many polyomaviruses use their glycolipid receptors to travel to the endoplasmic reticulum (Fujinaga et al., 2003). There, both systems are thought to undergo a massive conformational change that exposes the central, monomeric protein and leads to membrane penetration (Tsai et al., 2003). In both cases, the ERAD machinery seems to be involved in this step.

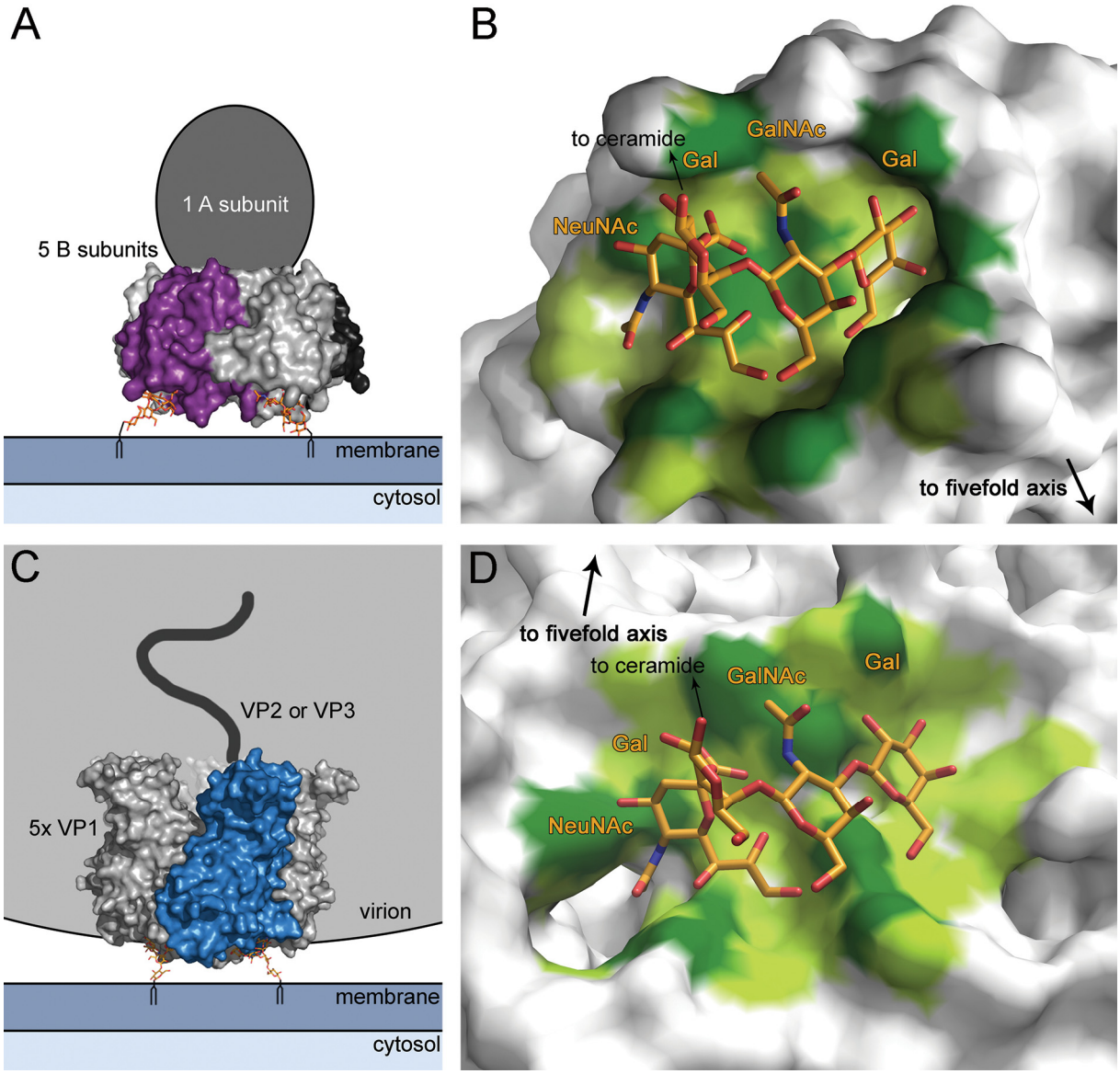



**Fig. 45: Comparison between GM1 binding by SV40 and cholera toxin.**
(A)+(C) Schematic views of the physiological context of cholera toxin B (A) and SV40 VP1 (C) binding to GM1. One pentamer of cholera toxin B (pdb: 3CHB, (Merritt et al., 1994) and SV40 VP1, respectively, is shown in surface representation. The oligosaccharide part of GM1 is shown in stick representation.
(B)+(D) Different GM1 binding modes by cholera toxin B (B) and SV40 VP1 (D). The proteins are shown in surface representation, GM1 in stick representation. In both cases, the terminal Glc of GM1 is left out for clarity. Protein atoms are colored according to changes in solvent-accessible surface area upon binding. Area differences above 10 $Å^2$ (dark green) indicate substantial contributions to binding, differences below 10 $Å^2$ (light green) indicate weaker contributions to binding.

However, even though both SV40 and cholera toxin recognize the solution conformation of GM1 (Acquotti et al., 1990b; Brocca et al., 1998; Merritt et al., 1994), their GM1 binding sites are not conserved. The main interaction of cholera toxin is with the terminal Gal of GM1, which inserts deeply into a pocket on the protein and stacks against a tryptophan side chain with the hydrophobic face of the pyranose ring (Fig. 45B). There are additional hydrogen bonds between cholera toxin and the Gal residue that take place in the relatively deep binding site, somewhat shielded from solvent. Consequently, the interaction

of one binding site of cholera toxin with GM1 has an affinity in the high nanomolar range (Turnbull et al., 2004), in contrast to the millimolar affinity of SV40 VP1 (3.2.2). This is not surprising, given that the multimeric unit of cholera toxin is a pentamer, while SV40 assembles into a viral particle with 360 subunits that binds with considerably higher affinity (Clayson and Compans, 1988). In contrast to the numerous binding sites on viruses, bacterial $AB_5$ toxins feature five carbohydrate binding sites, and the interactions between these and cellular receptors must be sufficient to attach to cells, induce membrane curvature and provide cell entry.

## 4.6. Outlook

This study has provided insights into the carbohydrate-binding properties of polyomaviruses and their importance for viral infection, spread, tropism and pathogenesis. It has also prompted new questions that ask for new experiments to be conducted.

- The interactions of **SV40** VP1 with its simian receptor NeuNGc-GM1 are not experimentally validated at the moment, and might differ from the ones observed with NeuNAc-GM1. If sufficient quantities of NeuNGc-GM1 are available, they can be probed by STD NMR and a co-crystal structure using the established soaking conditions for SV40 VP1. Furthermore, the very big cavity on the SV40 VP1 surface suggests that GM1 carrying sialic acids modified at the acetyl function might bind even better than NeuNGc-GM1, and thus function as better receptors.
- LSTc was identified as the functional attachment motif for **JCV**, and it is likely attached to a longer glycan on a glycoprotein. The identity of the glycoprotein(s) carrying LSTc on JCV host cells might shed light into how endocytosis is triggered by JCV. One approach would be to use LSTc-specific lectins and/or JCV VP1 to pull down LSTc-carrying proteins from membrane fractions of glial cell lines, and to identify these by mass spectrometry.
- Glycan array screening of **BKV** VP1 has only yielded weak signals to date even though the virus likely does not bind with reduced affinity compared to its close relatives. Improvement of probe presentation on the glycan array by different lipid linkers or carrier lipids will be important to differentiate between weakly interacting ligands such as GD3 and non-binders such as GM1. In addition, several lines of evidence point to additional contacts between BKV and its receptors GT1b and GD1b that are not visible in the BKV VP1-GD3 complex presented here. Since soaking of BKV VP1 crystals in GT1b solution did not yield a complex, co-crystallization of the two binding partners might work. In addition, the somewhat smaller GD1b oligosaccharide could be used for soaking, depending on its availability.
- The high-resolution structures of the human pathogens BKV and JCV in complex with their oligosaccharide receptors provide a basis for the rational **design** of VP1-targeting **drugs** that might block attachment.
- Cell attachment data on **K68S** highlighted the importance of species-specific recognition of different sialic acid types, and raised the possibility that NeuNAc-GM1 might be metabolized to NeuNGc-GM1 in simian cells. It would be interesting to further investigate these findings with NeuNAc- and NeuNGc-specific antibodies. It would also be important to analyze the correlation between receptor usage of BKV and K68S and

the ability to propagate in cells from different **tissues**. This will inform attempts to use oligosaccharides to retarget viruses to different host cells e.g. for gene delivery.

- Analysis of the JCV mutants that might bind glycolipids instead of LSTc might inform on how viruses are routed to different uptake mechanisms, and might lay the basis for the **retargeting** of viruses to different **entry pathways**.
- The common features of viral sialic acid binding sites, i.e. the strict spatial requirements and the relatively few possible interactions might be used to create a search algorithm for sialic acid binding sites on viral proteins that uses **structural templates** of known binding sites. This might be facilitated by the fact that most oligosaccharide binding sites occur on residues in β-conformation.

Overall, the current study has greatly benefited from the availability of glycan screening platforms and substantial quantities of complex oligosaccharides, and these will undoubtedly advance the understanding of glycan binding by other viruses as well. However, in order to correlate detailed data about receptor specificity and affinity with biological properties of viruses, it will be crucial to obtain qualitative and quantitative information on the glycan profiles of different tissues and different species, preferably in different stages of development or disease.

# 5. References


Acquotti, D., Cantu, L., Ragg, E., and Sonnino, S. (1994). Geometrical and conformational properties of ganglioside GalNAc-GD1a, IV4GalNAcIV3Neu5AcII3Neu5AcGgOse4Cer. Eur J Biochem *225*, 271-288.

Acquotti, D., Poppe, L., Dabrowski, J., Von der Lieth, C.W., Sonnino, S., and Tettamanti, G. (1990a). Three-dimensional structure of the oligosaccharide chain of GM1 ganglioside revealed by a distance-mapping procedure: a rotating and laboratory frame nuclear overhauser enhancement investigation of native glycolipid in dimethyl sulfoxide and in water-dodecylphosphocholine solutions. J Am Chem Soc *112*, 7772-7778.

Acquotti, D., Poppe, L., Dabrowski, J., von der Lieth, C.W., Sonnino, S., and Tettamanti, G. (1990b). Three-dimensional structure of the oligosaccharide chain of GM1 ganglioside revealed by a distance-mapping procedure: a rotating and laboratory frame nuclear overhauser enhancement investigation of native glycolipid in dimethyl sulfoxide and in water-dodecylphosphocholine solutions. J Am Chem Soc *112*, 7772-7778.

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W.*, et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213-221.

Afonine, P.V., Grosse-Kunstleve, R.W., and Adams, P.D. (2005). The Phenix refinement framework. CCP4 Newsletter *42*, contribution 8.

Allander, T., Andreasson, K., Gupta, S., Bjerkner, A., Bogdanovic, G., Persson, M.A., Dalianis, T., Ramqvist, T., and Andersson, B. (2007). Identification of a third human polyomavirus. J Virol *81*, 4130-4136.

Attrill, H., Imamura, A., Sharma, R.S., Kiso, M., Crocker, P.R., and van Aalten, D.M. (2006). Siglec-7 undergoes a major conformational change when complexed with the alpha(2,8)-disialylganglioside GT1b. J Biol Chem *281*, 32774-32783.

Atwood, W.J., and Norkin, L.C. (1989). Class I major histocompatibility proteins as cell surface receptors for simian virus 40. J Virol *63*, 4474-4477.

Atwood, W.J., Wang, L., Durham, L.C., Amemiya, K., Traub, R.G., and Major, E.O. (1995). Evaluation of the role of cytokine activation in the multiplication of JC virus (JCV) in human fetal glial cells. J Neurovirol *1*, 40-49.

Barouch, D.H., and Harrison, S.C. (1994). Interactions among the major and minor coat proteins of polyomavirus. J Virol *68*, 3982-3989.

Bauer, P.H., Cui, C., Liu, W.R., Stehle, T., Harrison, S.C., DeCaprio, J.A., and Benjamin, T.L. (1999). Discrimination between sialic acid-containing receptors and pseudoreceptors regulates polyomavirus spread in the mouse. J Virol *73*, 5826-5832.

Bedi, A., Miller, C.B., Hanson, J.L., Goodman, S., Ambinder, R.F., Charache, P., Arthur, R.R., and Jones, R.J. (1995). Association of BK virus with failure of prophylaxis against hemorrhagic cystitis following bone marrow transplantation. J Clin Oncol *13*, 1103-1109.

Bergsagel, D.J., Finegold, M.J., Butel, J.S., Kupsky, W.J., and Garcea, R.L. (1992). DNA sequences similar to those of simian virus 40 in ependymomas and choroid plexus tumors of childhood. N Engl J Med *326*, 988-993.

Blanchard, H., Yu, X., Coulson, B.S., and von Itzstein, M. (2007). Insight into host cell carbohydrate-recognition by human and porcine rotavirus from crystal structures of the virion spike associated carbohydrate-binding domain (VP8*). J Mol Biol *367*, 1215-1226.

Blow, D. (2002). Outline of Crystallography for Biologists (Oxford, Oxford University Press).

Bozeman, L.H., Davis, R.B., Gaudry, D., Lukert, P.D., Fletcher, O.J., and Dykstra, M.J. (1981). Characterization of a papovavirus isolated from fledgling budgerigars. Avian Dis *25*, 972-980.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem *72*, 248-254.

Brady, J.N., Winston, V.D., and Consigli, R.A. (1977). Dissociation of polyoma virus by the chelation of calcium ions found associated with purified virions. J Virol *23*, 717-724.

Bragg, W.L. (1913). The diffraction of short electromagnetic waves by a crystal. Proc Camb Phil Soc *17*, 43-57.

Breg, J., Kroon-Batenburg, L.M., Strecker, G., Montreuil, J., and Vliegenthart, J.F. (1989a). Conformational analysis of the sialyl alpha(2-3/6)N-acetyllactosamine structural element occurring in glycoproteins, by two-dimensional NOE 1H-NMR spectroscopy in combination with energy calculations by hard-sphere exo-anomeric and molecular mechanics force-field with hydrogen-bonding potential. Eur J Biochem *178*, 727-739.
Breg, J., Kroon-Batenburg, L.M., Strecker, G., Montreuil, J., and Vliegenthart, J.F. (1989b). Conformational analysis of the sialyl alpha(2----3/6)N-acetyllactosamine structural element occurring in glycoproteins, by two-dimensional NOE 1H-NMR spectroscopy in combination with energy calculations by hard-sphere exo-anomeric and molecular mechanics force-field with hydrogen-bonding potential. Eur J Biochem *178*, 727-739.
Brisson, J.R., Baumann, H., Imberty, A., Perez, S., and Jennings, H.J. (1992). Helical epitope of the group B meningococcal alpha(2-8)-linked sialic acid polysaccharide. Biochemistry *31*, 4996-5004.
Brocca, P., Berthault, P., and Sonnino, S. (1998). Conformation of the oligosaccharide chain of G(M1) ganglioside in a carbohydrate-enriched surface. Biophys J *74*, 309-318.
Brunger, A.T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature *355*, 472-475.
Burmeister, W.P., Guilligay, D., Cusack, S., Wadell, G., and Arnberg, N. (2004). Crystal structure of species D adenovirus fiber knobs and their sialic acid binding sites. J Virol *78*, 7727-7736.
Buscail, L., and Cordelier, P. (2007). Potential of recombinant SV40-based vectors for gene therapy. Recent Pat DNA Gene Seq *1*, 93-99.
Cahan, L.D., and Paulson, J.C. (1980). Polyoma virus adsorbs to specific sialyloligosaccharide receptors on erythrocytes. Virology *103*, 505-509.
Cahan, L.D., Singh, R., and Paulson, J.C. (1983). Sialyloligosaccharide receptors of binding variants of polyoma virus. Virology *130*, 281-289.
Campanero-Rhodes, M.A., Smith, A., Chai, W., Sonnino, S., Mauri, L., Childs, R.A., Zhang, Y., Ewers, H., Helenius, A., Imberty, A.*, et al.* (2007). N-glycolyl GM1 ganglioside as a receptor for simian virus 40. J Virol *81*, 12846-12858.
Carson, K.R., Evens, A.M., Richey, E.A., Habermann, T.M., Focosi, D., Seymour, J.F., Laubach, J., Bawn, S.D., Gordon, L.I., Winter, J.N.*, et al.* (2009a). Progressive multifocal leukoencephalopathy after rituximab therapy in HIV-negative patients: a report of 57 cases from the Research on Adverse Drug Events and Reports project. Blood *113*, 4834-4840.
Carson, K.R., Focosi, D., Major, E.O., Petrini, M., Richey, E.A., West, D.P., and Bennett, C.L. (2009b). Monoclonal antibody-associated progressive multifocal leucoencephalopathy in patients treated with rituximab, natalizumab, and efalizumab: a Review from the Research on Adverse Drug Events and Reports (RADAR) Project. Lancet Oncol *10*, 816-824.
Caspar, D.L., and Klug, A. (1962). Physical principles in the construction of regular viruses. Cold Spring Harb Symp Quant Biol *27*, 1-24.
CCP4 (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr *50*, 760-763.
Chandrasekaran, A., Srinivasan, A., Raman, R., Viswanathan, K., Raguram, S., Tumpey, T.M., Sasisekharan, V., and Sasisekharan, R. (2008). Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin. Nat Biotechnol *26*, 107-113.
Chang, D., Fung, C.Y., Ou, W.C., Chao, P.C., Li, S.Y., Wang, M., Huang, Y.L., Tzeng, T.Y., and Tsai, R.T. (1997). Self-assembly of the JC virus major capsid protein, VP1, expressed in insect cells. J Gen Virol *78 ( Pt 6)*, 1435-1439.
Chen, X.S., Stehle, T., and Harrison, S.C. (1998). Interaction of polyomavirus internal protein VP2 with the major capsid protein VP1 and implications for participation of VP2 in viral entry. Embo J *17*, 3233-3240.
Cinque, P., Koralnik, I.J., Gerevini, S., Miro, J.M., and Price, R.W. (2009). Progressive multifocal leukoencephalopathy in HIV-1 infection. Lancet Infect Dis *9*, 625-636.
Clayson, E.T., and Compans, R.W. (1988). Entry of simian virus 40 is restricted to apical surfaces of polarized epithelial cells. Mol Cell Biol *8*, 3391-3396.
Clayson, E.T., and Compans, R.W. (1989). Characterization of simian virus 40 receptor moieties on the surfaces of Vero C1008 cells. J Virol *63*, 1095-1100.
Cole, C.N., and Conzen, S.D. (2001). Polyomaviridae: the viruses and their replication. In Fields virology, D.M. Knipe, and P.M. Howley, eds. (Philadelphia, PA, Lippincott-Raven Publishers).
Crennell, S., Takimoto, T., Portner, A., and Taylor, G. (2000). Crystal structure of the multifunctional paramyxovirus hemagglutinin-neuraminidase. Nat Struct Biol *7*, 1068-1074.
Crick, F.H., and Watson, J.D. (1956). Structure of small viruses. Nature *177*, 473-475.
Damm, E.M., Pelkmans, L., Kartenbeck, J., Mezzacasa, A., Kurzchalia, T., and Helenius, A. (2005). Clathrin- and caveolin-1-independent endocytosis: entry of simian virus 40 into cells devoid of caveolae. J Cell Biol *168*, 477-488.

Dangoria, N.S., Breau, W.C., Anderson, H.A., Cishek, D.M., and Norkin, L.C. (1996). Extracellular simian virus 40 induces an ERK/MAP kinase-independent signalling pathway that activates primary response genes and promotes virus entry. J Gen Virol *77 ( Pt 9)*, 2173-2182.
Daniels, R., Rusan, N.M., Wadsworth, P., and Hebert, D.N. (2006). SV40 VP2 and VP3 insertion into ER membranes is controlled by the capsid protein VP1: implications for DNA translocation out of the ER. Mol Cell *24*, 955-966.
Daniels, R., Sadowicz, D., and Hebert, D.N. (2007). A very late viral protein triggers the lytic release of SV40. PLoS Pathog *3*, e98.
Diederichs, K., and Karplus, P.A. (1997). Improved R-factors for diffraction data analysis in macromolecular crystallography. Nat Struct Biol *4*, 269-275.
Dormitzer, P.R., Sun, Z.Y., Blixt, O., Paulson, J.C., Wagner, G., and Harrison, S.C. (2002a). Specificity and affinity of sialic acid binding by the rhesus rotavirus VP8* core. J Virol *76*, 10512-10517.
Dormitzer, P.R., Sun, Z.Y., Wagner, G., and Harrison, S.C. (2002b). The rhesus rotavirus VP4 sialic acid binding domain has a galectin fold with a novel carbohydrate binding site. EMBO J *21*, 885-897.
Dorries, K. (1998). Molecular biology and pathogenesis of human polyomavirus infections. Dev Biol Stand *94*, 71-79.
Drenth, J. (2007). Principles of Protein X-ray Crystallography, 3rd edn (New York, Springer).
Dubensky, T.W., Freund, R., Dawe, C.J., and Benjamin, T.L. (1991). Polyomavirus replication in mice: influences of VP1 type and route of inoculation. J Virol *65*, 342-349.
Dugan, A.S., Eash, S., and Atwood, W.J. (2005). An N-linked glycoprotein with alpha(2,3)-linked sialic acid is a receptor for BK virus. J Virol *79*, 14442-14445.
Dugan, A.S., Gasparovic, M.L., and Atwood, W.J. (2008). Direct correlation between sialic acid binding and infection of cells by two human polyomaviruses (JC virus and BK virus). J Virol *82*, 2560-2564.
Dugan, A.S., Gasparovic, M.L., Tsomaia, N., Mierke, D.F., O'Hara, B.A., Manley, K., and Atwood, W.J. (2007). Identification of amino acid residues in BK virus VP1 that are critical for viability and growth. J Virol *81*, 11798-11808.
Eash, S., and Atwood, W.J. (2005). Involvement of cytoskeletal components in BK virus infectious entry. J Virol *79*, 11734-11741.
Eash, S., Querbes, W., and Atwood, W.J. (2004a). Infection of vero cells by BK virus is dependent on caveolae. J Virol *78*, 11583-11590.
Eash, S., Tavares, R., Stopa, E.G., Robbins, S.H., Brossay, L., and Atwood, W.J. (2004b). Differential distribution of the JC virus receptor-type sialic acid in normal human tissues. Am J Pathol *164*, 419-428.
Eddy, B.E., Stewart, S.E., Young, R., and Mider, G.B. (1958). Neoplasms in hamsters induced by mouse tumor agent passed in tissue culture. J Natl Cancer Inst *20*, 747-761.
Eisen, M.B., Sabesan, S., Skehel, J.J., and Wiley, D.C. (1997). Binding of the influenza A virus to cell-surface receptors: structures of five hemagglutinin-sialyloligosaccharide complexes determined by X-ray crystallography. Virology *232*, 19-31.
Elphick, G.F., Querbes, W., Jordan, J.A., Gee, G.V., Eash, S., Manley, K., Dugan, A., Stanifer, M., Bhatnagar, A., Kroeze, W.K.*, et al.* (2004). The human polyomavirus, JCV, uses serotonin receptors to infect cells. Science *306*, 1380-1383.
Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.
Erickson, K.D., Garcea, R.L., and Tsai, B. (2009). Ganglioside GT1b is a putative host cell receptor for the Merkel cell polyomavirus. J Virol *83*, 10275-10279.
Ewald, P.P. (1917). Zur Begründung der Kristalloptik. Ann Phys Lpz *349*, 519.
Ewers, H., Romer, W., Smith, A.E., Bacia, K., Dmitrieff, S., Chai, W., Mancini, R., Kartenbeck, J., Chambon, V., Berland, L.*, et al.* (2010). GM1 structure determines SV40-induced membrane invagination and infection. Nat Cell Biol *12*, 11-18; sup pp 11-12.
Ewers, H., Smith, A.E., Sbalzarini, I.F., Lilie, H., Koumoutsakos, P., and Helenius, A. (2005). Single-particle tracking of murine polyoma virus-like particles on live cells and artificial membranes. Proc Natl Acad Sci U S A *102*, 15110-15115.
Feizi, T. (1985). Demonstration by monoclonal antibodies that carbohydrate structures of glycoproteins and glycolipids are onco-developmental antigens. Nature *314*, 53-57.
Feizi, T., Kapadia, A., and Yount, W.J. (1980). I and i antigens of human peripheral blood lymphocytes cocap with receptors for concanavalin A. Proc Natl Acad Sci U S A *77*, 376-380.
Feng, H., Shuda, M., Chang, Y., and Moore, P.S. (2008). Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science *319*, 1096-1100.
Finch, J.T. (1974). The surface structure of polyoma virus. J Gen Virol *24*, 359-364.
French, S., and Wilson, K. (1978). On the Treatment of Negative Intensity Observations. Acta Crystallogr A Found Crystallogr *34*, 517-525.

Freund, R., Calderone, A., Dawe, C.J., and Benjamin, T.L. (1991a). Polyomavirus tumor induction in mice: effects of polymorphisms of VP1 and large T antigen. J Virol *65*, 335-341.
Freund, R., Garcea, R.L., Sahli, R., and Benjamin, T.L. (1991b). A single-amino-acid substitution in polyomavirus VP1 correlates with plaque size and hemagglutination behavior. J Virol *65*, 350-355.
Fried, H., Cahan, L.D., and Paulson, J.C. (1981). Polyoma virus recognizes specific sialyligosaccharide receptors on host cells. Virology *109*, 188-192.
Fujinaga, Y., Wolf, A.A., Rodighiero, C., Wheeler, H., Tsai, B., Allen, L., Jobling, M.G., Rapoport, T., Holmes, R.K., and Lencer, W.I. (2003). Gangliosides that associate with lipid rafts mediate transport of cholera and related toxins from the plasma membrane to endoplasmic reticulm. Mol Biol Cell *14*, 4783-4793.
Gamblin, S.J., Haire, L.F., Russell, R.J., Stevens, D.J., Xiao, B., Ha, Y., Vasisht, N., Steinhauer, D.A., Daniels, R.S., Elliot, A.*, et al.* (2004). The structure and receptor binding properties of the 1918 influenza hemagglutinin. Science *303*, 1838-1842.
Garcea, R.L., and Imperiale, M.J. (2003). Simian virus 40 infection of humans. J Virol *77*, 5039-5045.
Gardner, S.D., Field, A.M., Coleman, D.V., and Hulme, B. (1971). New human papovavirus (B.K.) isolated from urine after renal transplantation. Lancet *1*, 1253-1257.
Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In The Proteomics Protocols Handbook, J.M. Walker, ed. (Humana Press), pp. 571-607.
Gaynor, A.M., Nissen, M.D., Whiley, D.M., Mackay, I.M., Lambert, S.B., Wu, G., Brennan, D.C., Storch, G.A., Sloots, T.P., and Wang, D. (2007). Identification of a novel polyomavirus from patients with acute respiratory tract infections. PLoS Pathog *3*, e64.
Gee, G.V., Tsomaia, N., Mierke, D.F., and Atwood, W.J. (2004). Modeling a sialic acid binding pocket in the external loops of JC virus VP1. J Biol Chem *279*, 49172-49176.
Gilbert, J., and Benjamin, T. (2004). Uptake pathway of polyomavirus via ganglioside GD1a. J Virol *78*, 12259-12267.
Gilbert, J., Ou, W., Silver, J., and Benjamin, T. (2006). Downregulation of protein disulfide isomerase inhibits infection by the mouse polyomavirus. J Virol *80*, 10868-10870.
Gilbert, J.M., and Benjamin, T.L. (2000). Early steps of polyomavirus entry into cells. J Virol *74*, 8582-8588.
Gilbert, J.M., Goldberg, I.G., and Benjamin, T.L. (2003). Cell penetration and trafficking of polyomavirus. J Virol *77*, 2615-2622.
Girardi, A.J., Sweet, B.H., Slotnick, V.B., and Hilleman, M.R. (1962). Development of tumors in hamsters inoculated in the neonatal period with vacuolating virus, SV-40. Proc Soc Exp Biol Med *109*, 649-660.
Gross, L. (1953). Neck tumors, or leukemia, developing in adult C3H mice following inoculation, in early infancy, with filtered (Berkefeld N), or centrifugated (144,000 X g), Ak-leukemic extracts. Cancer *6*, 948-958.
Guerin, J.L., Gelfi, J., Dubois, L., Vuillaume, A., Boucraut-Baralon, C., and Pingret, J.L. (2000). A novel polyomavirus (goose hemorrhagic polyomavirus) is the agent of hemorrhagic nephritis enteritis of geese. J Virol *74*, 4523-4529.
Ha, Y., Stevens, D.J., Skehel, J.J., and Wiley, D.C. (2001). X-ray structures of H5 avian and H9 swine influenza virus hemagglutinins bound to avian and human receptor analogs. Proc Natl Acad Sci U S A *98*, 11181-11186.
Ha, Y., Stevens, D.J., Skehel, J.J., and Wiley, D.C. (2003). X-ray structure of the hemagglutinin of a potential H3 avian progenitor of the 1968 Hong Kong pandemic influenza virus. Virology *309*, 209-218.
Haselhorst, T., Blanchard, H., Frank, M., Kraschnefski, M.J., Kiefel, M.J., Szyczew, A.J., Dyason, J.C., Fleming, F., Holloway, G., Coulson, B.S.*, et al.* (2007). STD NMR spectroscopy and molecular modeling investigation of the binding of N-acetylneuraminic acid derivatives to rhesus rotavirus VP8* core. Glycobiology *17*, 68-81.
Hebert, D.N., and Molinari, M. (2007). In and out of the ER: protein folding, quality control, degradation, and related human diseases. Physiol Rev *87*, 1377-1408.
Hirsch, H.H. (2005). BK virus: opportunity makes a pathogen. Clin Infect Dis *41*, 354-360.
Hirsch, H.H., and Steiger, J. (2003). Polyomavirus BK. Lancet Infect Dis *3*, 611-623.
Holthofer, H., Reivinen, J., and Miettinen, A. (1994). Nephron segment and cell-type specific expression of gangliosides in the developing and adult kidney. Kidney Int *45*, 123-130.
Horvath, C.J., Simon, M.A., Bergsagel, D.J., Pauley, D.R., King, N.W., Garcea, R.L., and Ringler, D.J. (1992). Simian virus 40-induced disease in rhesus monkeys with simian acquired immunodeficiency syndrome. Am J Pathol *140*, 1431-1440.
Houben, R., Shuda, M., Weinkam, R., Schrama, D., Feng, H., Chang, Y., Moore, P.S., and Becker, J.C. (2010). Merkel cell polyomavirus-infected Merkel cell carcinoma cells require expression of viral T antigens. J Virol *84*, 7064-7072.

Houliston, R.S., Jacobs, B.C., Tio-Gillen, A.P., Verschuuren, J.J., Khieu, N.H., Gilbert, M., and Jarrell, H.C. (2009). STD-NMR used to elucidate the fine binding specificity of pathogenic anti-ganglioside antibodies directly in patient serum. Biochemistry *48*, 220-222.
Jiang, M., Abend, J.R., Johnson, S.F., and Imperiale, M.J. (2009a). The role of polyomaviruses in human disease. Virology *384*, 266-273.
Jiang, M., Abend, J.R., Tsai, B., and Imperiale, M.J. (2009b). Early events during BK virus entry and disassembly. J Virol *83*, 1350-1358.
Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J Appl Cryst *26*, 795-800.
Kabsch, W. (2010). Integration, scaling, space-group assignment and post-refinement. Acta Crystallogr D Biol Crystallogr *66*, 133-144.
Kantardjieff, K.A., and Rupp, B. (2003). Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. Protein Sci *12*, 1865-1871.
Kartenbeck, J., Stukenbrok, H., and Helenius, A. (1989). Endocytosis of simian virus 40 into the endoplasmic reticulum. J Cell Biol *109*, 2721-2729.
Kean, J.M., Rao, S., Wang, M., and Garcea, R.L. (2009). Seroepidemiology of human polyomaviruses. PLoS Pathog *5*, e1000363.
Khalili, K., and White, M.K. (2006). Human demyelinating disease and the polyomavirus JCV. Mult Scler *12*, 133-142.
Kishida, S. (2007). [Progressive multifocal leukoencephalopathy--epidemiology, clinical pictures, diagnosis and therapy]. Brain Nerve *59*, 125-137.
Kitov, P.I., Sadowska, J.M., Mulvey, G., Armstrong, G.D., Ling, H., Pannu, N.S., Read, R.J., and Bundle, D.R. (2000). Shiga-like toxins are neutralized by tailored multivalent carbohydrate ligands. Nature *403*, 669-672.
Kleywegt, G.J. (1996). Use of non-crystallographic symmetry in protein structure refinement. Acta Crystallogr D Biol Crystallogr *52*, 842-857.
Knowles, W.A., Pipkin, P., Andrews, N., Vyse, A., Minor, P., Brown, D.W., and Miller, E. (2003). Population-based study of antibody to the human polyomaviruses BKV and JCV and the simian polyomavirus SV40. J Med Virol *71*, 115-123.
Kolter, T., Proia, R.L., and Sandhoff, K. (2002). Combinatorial ganglioside biosynthesis. J Biol Chem *277*, 25859-25862.
Komagome, R., Sawa, H., Suzuki, T., Suzuki, Y., Tanaka, S., Atwood, W.J., and Nagashima, K. (2002). Oligosaccharides as receptors for JC virus. J Virol *76*, 12992-13000.
Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr *60*, 2256-2268.
Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-685.
Li, T.C., Takeda, N., Kato, K., Nilsson, J., Xing, L., Haag, L., Cheng, R.H., and Miyamura, T. (2003). Characterization of self-assembled virus-like particles of human polyomavirus BK generated by recombinant baculoviruses. Virology *311*, 115-124.
Liddington, R.C., Yan, Y., Moulai, J., Sahli, R., Benjamin, T.L., and Harrison, S.C. (1991). Structure of simian virus 40 at 3.8-A resolution. Nature *354*, 278-284.
Lilley, B.N., Gilbert, J.M., Ploegh, H.L., and Benjamin, T.L. (2006). Murine polyomavirus requires the endoplasmic reticulum protein Derlin-2 to initiate infection. J Virol *80*, 8739-8744.
Lingwood, D., and Simons, K. (2010). Lipid rafts as a membrane-organizing principle. Science *327*, 46-50.
Liu, C.K., Hope, A.P., and Atwood, W.J. (1998a). The human polyomavirus, JCV, does not share receptor specificity with SV40 on human glial cells. J Neurovirol *4*, 49-58.
Liu, C.K., Wei, G., and Atwood, W.J. (1998b). Infection of glial cells by the human polyomavirus JC is mediated by an N-linked glycoprotein containing terminal alpha(2-6)-linked sialic acids. J Virol *72*, 4643-4649.
Low, J.A., Magnuson, B., Tsai, B., and Imperiale, M.J. (2006). Identification of gangliosides GD1b and GT1b as receptors for BK virus. J Virol *80*, 1361-1366.
Maginnis, M.S., Haley, S.A., Gee, G.V., and Atwood, W.J. (2010). Role of N-linked glycosylation of the 5-HT2A receptor in JC virus infection. J Virol *84*, 9677-9684.
Magnuson, B., Rainey, E.K., Benjamin, T., Baryshev, M., Mkrtchian, S., and Tsai, B. (2005). ERp29 triggers a conformational change in polyomavirus to stimulate membrane binding. Mol Cell *20*, 289-300.
Major, E.O. (2009). Reemergence of PML in natalizumab-treated patients--new cases, same concerns. N Engl J Med *361*, 1041-1043.

Major, E.O., Miller, A.E., Mourrain, P., Traub, R.G., de Widt, E., and Sever, J. (1985). Establishment of a line of human fetal glial cells that supports JC virus multiplication. Proc Natl Acad Sci U S A *82*, 1257-1261.
Matthews, B.W. (1968). Solvent content of protein crystals. J Mol Biol *33*, 491-497.
Maxwell, J.C. (1865). A Dynamical Theory of the Electromagnetic Field. Philos Trans R Soc Lond *155*, 459-512.
May, A.P., Robinson, R.C., Vinson, M., Crocker, P.R., and Jones, E.Y. (1998). Crystal structure of the N-terminal domain of sialoadhesin in complex with 3' sialyllactose at 1.85 A resolution. Mol Cell *1*, 719-728.
McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658-674.
Mercer, J., Schelhaas, M., and Helenius, A. (2010). Virus entry by endocytosis. Annu Rev Biochem *79*, 803-833.
Merritt, E.A., and Hol, W.G. (1995). AB5 toxins. Curr Opin Struct Biol *5*, 165-171.
Merritt, E.A., Sarfaty, S., van den Akker, F., L'Hoir, C., Martial, J.A., and Hol, W.G. (1994). Crystal structure of cholera toxin B-pentamer bound to receptor GM1 pentasaccharide. Protein Sci *3*, 166-175.
Michon, F., Brisson, J.R., and Jennings, H.J. (1987). Conformational differences between linear alpha (2----8)-linked homosialooligosaccharides and the epitope of the group B meningococcal polysaccharide. Biochemistry *26*, 8399-8405.
Miller-Podraza, H., Bradley, R.M., and Fishman, P.H. (1982). Biosynthesis and localization of gangliosides in cultured cells. Biochemistry *21*, 3260-3265.
Moriyama, T., Marquez, J.P., Wakatsuki, T., and Sorokin, A. (2007). Caveolar endocytosis is critical for BK virus infection of human renal proximal tubular epithelial cells. J Virol *81*, 8552-8562.
Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr *53*, 240-255.
Navaza, J. (1994). AMoRe: an Automated Package for Molecular Replacement. Acta Crystallogr A Found Crystallogr *50*, 157-163.
Neu, U., Woellner, K., Gauglitz, G., and Stehle, T. (2008). Structural basis of GM1 ganglioside recognition by simian virus 40. Proc Natl Acad Sci U S A *105*, 5219-5224.
Neu, U., Stehle, T., and Atwood, W.J. (2009) The polyomaviridae: Contributions of virus structure to our understanding of virus receptors and infectious entry. Virology *384*(2), 389-399
Neu, U., Maginnis, M.S., Palma, A.S., Ströh, L..J., Nelson, C.D.S., Feizi, T., Atwood, W.J., and Stehle, T. Cell Host Microbe *8*(4), 309-319
Nickeleit, V., Hirsch, H.H., Binet, I.F., Gudat, F., Prince, O., Dalquen, P., Thiel, G., and Mihatsch, M.J. (1999). Polyomavirus infection of renal allograft recipients: from latent infection to manifest disease. J Am Soc Nephrol *10*, 1080-1089.
Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol *302*, 205-217.
Otwinowski, Z., and Minor, W. (1997). Processing of X-Ray Diffraction Data Collected in Oscillation Mode. In Methods in Enzymology, C.W.J. Carter, and R.M. Sweet, eds. (New York, Academic Press), pp. 307-326.
Padgett, B.L., Walker, D.L., ZuRhein, G.M., Eckroade, R.J., and Dessel, B.H. (1971). Cultivation of papova-like virus from human brain with progressive multifocal leucoencephalopathy. Lancet *1*, 1257-1260.
Painter, J., and Merritt, E.A. (2006). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr D Biol Crystallogr *62*, 439-450.
Palma, A.S., Feizi, T., Zhang, Y., Stoll, M.S., Lawson, A.M., Diaz-Rodriguez, E., Campanero-Rhodes, M.A., Costa, J., Gordon, S., Brown, G.D*., et al.* (2006). Ligands for the beta-glucan receptor, Dectin-1, assigned using "designer" microarrays of oligosaccharide probes (neoglycolipids) generated from glucan polysaccharides. J Biol Chem *281*, 5771-5779.
Pelkmans, L., Fava, E., Grabner, H., Hannus, M., Habermann, B., Krausz, E., and Zerial, M. (2005). Genome-wide analysis of human kinases in clathrin- and caveolae/raft-mediated endocytosis. Nature *436*, 78-86.
Pelkmans, L., Kartenbeck, J., and Helenius, A. (2001). Caveolar endocytosis of simian virus 40 reveals a new two-step vesicular-transport pathway to the ER. Nat Cell Biol *3*, 473-483.
Pelkmans, L., Puntener, D., and Helenius, A. (2002). Local actin polymerization and dynamin recruitment in SV40-induced internalization of caveolae. Science *296*, 535-539.
Pho, M.T., Ashok, A., and Atwood, W.J. (2000). JC virus enters human glial cells by clathrin-dependent receptor-mediated endocytosis. J Virol *74*, 2288-2292.
Qian, M., Cai, D., Verhey, K.J., and Tsai, B. (2009). A lipid receptor sorts polyomavirus from the endolysosome to the endoplasmic reticulum to cause infection. PLoS Pathog *5*, e1000465.

Querbes, W., Benmerah, A., Tosoni, D., Di Fiore, P.P., and Atwood, W.J. (2004). A JC virus-induced signal is required for infection of glial cells by a clathrin- and eps15-dependent pathway. J Virol *78*, 250-256.
Querbes, W., O'Hara, B.A., Williams, G., and Atwood, W.J. (2006). Invasion of host cells by JC virus identifies a novel role for caveolae in endosomal sorting of noncaveolar ligands. J Virol *80*, 9402-9413.
Rabinovich, G.A., Toscano, M.A., Jackson, S.S., and Vasta, G.R. (2007). Functions of cell surface galectin-glycoprotein lattices. Curr Opin Struct Biol *17*, 513-520.
Rahmlow, M., Shuster, E.A., Dominik, J., Deen, H.G., Jr., Dickson, D.W., Aksamit, A.J., Jr., Robles, H.A., and Freeman, W.D. (2008). Leflunomide-associated progressive multifocal leukoencephalopathy. Arch Neurol *65*, 1538-1539.
Rayment, I., Baker, T.S., Caspar, D.L., and Murakami, W.T. (1982). Polyoma virus capsid structure at 22.5 A resolution. Nature *295*, 110-115.
Richardson, J.M., Milton, M.J., and Homans, S.W. (1995). Solution dynamics of the oligosaccharide moiety of ganglioside GM1: comparison of solution conformations with the bound state conformation in association with cholera toxin B-pentamer. J Mol Recognit *8*, 358-362.
Romer, W., Berland, L., Chambon, V., Gaus, K., Windschiegl, B., Tenza, D., Aly, M.R., Fraisier, V., Florent, J.C., Perrais, D.*, et al.* (2007). Shiga toxin induces tubular membrane invaginations for its uptake into cells. Nature *450*, 670-675.
Rossmann, M.G., and Blow, D. (1962). The detection of sub-units within the crystallographic asymmetric unit. Acta Crystallogr *15*, 24-31.
Rupp, B. (2010). Macromolecular Crystallography (New York, Garland Science).
Russell, R.J., Stevens, D.J., Haire, L.F., Gamblin, S.J., and Skehel, J.J. (2006). Avian and human receptor binding by hemagglutinins of influenza A viruses. Glycoconj J *23*, 85-92.
Sabesan, S., Bock, K., and Paulson, J.C. (1991). Conformational analysis of sialyloligosaccharides. Carbohydr Res *218*, 27-54.
Sahli, R., Freund, R., Dubensky, T., Garcea, R., Bronson, R., and Benjamin, T. (1993). Defect in entry and altered pathogenicity of a polyoma virus mutant blocked in VP2 myristylation. Virology *192*, 142-153.
Salunke, D.M., Caspar, D.L., and Garcea, R.L. (1986). Self-assembly of purified polyomavirus capsid protein VP1. Cell *46*, 895-904.
Schelhaas, M., Malmstrom, J., Pelkmans, L., Haugstetter, J., Ellgaard, L., Grunewald, K., and Helenius, A. (2007). Simian Virus 40 depends on ER protein folding and quality control factors for entry into host cells. Cell *131*, 516-529.
Schowalter, R.M., Pastrana, D.V., Pumphrey, K.A., Moyer, A.L., and Buck, C.B. (2010). Merkel cell polyomavirus and two previously unknown polyomaviruses are chronically shed from human skin. Cell Host Microbe *7*, 509-515.
Seganti, L., Mastromarino, P., Superti, F., Sinibaldi, L., and Orsi, N. (1981). Receptors for BK virus on human erythrocytes. Acta Virol *25*, 177-181.
Seth, P., Diaz, F., and Major, E.O. (2003). Advances in the biology of JC virus and induction of progressive multifocal leukoencephalopathy. J Neurovirol *9*, 236-246.
Shah, K.V., Valis, J.D., and Daniel, R.W. (1978). Hemagglutination with simian papovavirus SA12. J Clin Microbiol *7*, 396-398.
Shayman, J.A., and Radin, N.S. (1991). Structure and function of renal glycosphingolipids. Am J Physiol *260*, F291-302.
Shinohara, T., Matsuda, M., Cheng, S.H., Marshall, J., Fujita, M., and Nagashima, K. (1993). BK virus infection of the human urinary tract. J Med Virol *41*, 301-305.
Shuda, M., Feng, H., Kwun, H.J., Rosen, S.T., Gjoerup, O., Moore, P.S., and Chang, Y. (2008). T antigen mutations are a human tumor-specific signature for Merkel cell polyomavirus. Proc Natl Acad Sci U S A *105*, 16272-16277.
Silverman, L., and Rubinstein, L.J. (1965). Electron microscopic observations on a case of progressive multifocal leukoencephalopathy. Acta Neuropathol *5*, 215-224.
Skehel, J.J., and Wiley, D.C. (2000). Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. Annu Rev Biochem *69*, 531-569.
Smith, A.E., Lilie, H., and Helenius, A. (2003). Ganglioside-dependent cell attachment and endocytosis of murine polyomavirus-like particles. FEBS Lett *555*, 199-203.
Stehle, T., Gamblin, S.J., Yan, Y., and Harrison, S.C. (1996). The structure of simian virus 40 refined at 3.1 A resolution. Structure *4*, 165-182.
Stehle, T., and Harrison, S.C. (1996). Crystal structures of murine polyomavirus in complex with straight-chain and branched-chain sialyloligosaccharide receptor fragments. Structure *4*, 183-194.
Stehle, T., and Harrison, S.C. (1997). High-resolution structure of a polyomavirus VP1-oligosaccharide complex: implications for assembly and receptor binding. EMBO J *16*, 5139-5148.

Stehle, T., Yan, Y., Benjamin, T.L., and Harrison, S.C. (1994). Structure of murine polyomavirus complexed with an oligosaccharide receptor fragment. Nature *369*, 160-163.
Stevens, J., Blixt, O., Tumpey, T.M., Taubenberger, J.K., Paulson, J.C., and Wilson, I.A. (2006). Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. Science *312*, 404-410.
Stevens, J., Corper, A.L., Basler, C.F., Taubenberger, J.K., Palese, P., and Wilson, I.A. (2004). Structure of the uncleaved human H1 hemagglutinin from the extinct 1918 influenza virus. Science *303*, 1866-1870.
Stewart, S.E., Eddy, B.E., and Borgese, N. (1958). Neoplasms in mice inoculated with a tumor agent carried in tissue culture. J Natl Cancer Inst *20*, 1223-1243.
Stoll, M.S., and Feizi, T. (2009). Software tools for storing, processing and displaying carbohydrate microarray data. Paper presented at: Beilstein Symposium on Glyco-Bioinformatics (Potsdam, Germany, Beilstein Institute for the Advancement of Chemical Sciences).
Strayer, D.S., Cordelier, P., Kondo, R., Liu, B., Matskevich, A.A., McKee, H.J., Nichols, C.N., Mitchell, C.B., Geverd, D.A., White, M.K.*, et al.* (2005). What they are, how they work and why they do what they do? The story of SV40-derived gene therapy vectors and what they have to offer. Curr Gene Ther *5*, 151-165.
Streuli, C.H., and Griffin, B.E. (1987). Myristic acid is coupled to a structural protein of polyoma virus and SV40. Nature *326*, 619-622.
Sunyaev, S.R., Lugovskoy, A., Simon, K., and Gorelik, L. (2009). Adaptive mutations in the JC virus protein capsid are associated with progressive multifocal leukoencephalopathy (PML). PLoS Genet *5*, e1000368.
Svennerholm, L. (1963). Chromatographic Separation of Human Brain Gangliosides. J Neurochem *10*, 613-623.
Sweet, B.H., and Hilleman, M.R. (1960). The vacuolating virus, S.V. 40. Proc Soc Exp Biol Med *105*, 420-427.
Tsai, B., Gilbert, J.M., Stehle, T., Lencer, W., Benjamin, T.L., and Rapoport, T.A. (2003). Gangliosides are receptors for murine polyoma virus and SV40. EMBO J *22*, 4346-4355.
Turnbull, W.B., Precious, B.L., and Homans, S.W. (2004). Dissecting the cholera toxin-ganglioside GM1 interaction by isothermal titration calorimetry. J Am Chem Soc *126*, 1047-1054.
Vacante, D.A., Traub, R., and Major, E.O. (1989). Extension of JC virus host range to monkey cells by insertion of a simian virus 40 enhancer into the JC virus regulatory region. Virology *170*, 353-361.
Van Assche, G., Van Ranst, M., Sciot, R., Dubois, B., Vermeire, S., Noman, M., Verbeeck, J., Geboes, K., Robberecht, W., and Rutgeerts, P. (2005). Progressive multifocal leukoencephalopathy after natalizumab therapy for Crohn's disease. N Engl J Med *353*, 362-368.
van der Meijden, E., Janssens, R.W., Lauber, C., Bouwes Bavinck, J.N., Gorbalenya, A.E., and Feltkamp, M.C. Discovery of a new human polyomavirus associated with trichodysplasia spinulosa in an immunocompromized patient. PLoS Pathog *6*, e1001024.
van Riel, D., Munster, V.J., de Wit, E., Rimmelzwaan, G.F., Fouchier, R.A., Osterhaus, A.D., and Kuiken, T. (2007). Human and avian influenza viruses target different cells in the lower respiratory tract of humans and other mammals. Am J Pathol *171*, 1215-1223.
Varki, A. (2001). Loss of N-glycolylneuraminic acid in humans: Mechanisms, consequences, and implications for hominid evolution. Am J Phys Anthropol *Suppl 33*, 54-69.
Varki, A. (2007). Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. Nature *446*, 1023-1029.
Weis, W., Brown, J.H., Cusack, S., Paulson, J.C., Skehel, J.J., and Wiley, D.C. (1988). Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid. Nature *333*, 426-431.
Wilson, A.J.C. (1949). X-ray diffraction by random layers: ideal profiles and determination of structure amplitudes from observed line profiles. Acta Crystallogr *2*, 245-251.
Wiseman, A.C. (2009). Polyomavirus nephropathy: a current perspective and clinical considerations. Am J Kidney Dis *54*, 131-142.
Wiseman, T., Williston, S., Brandts, J.F., and Lin, L.N. (1989). Rapid measurement of binding constants and heats of binding using a new titration calorimeter. Anal Biochem *179*, 131-137.
Wright, C.S. (1990). 2.2 A resolution structure analysis of two refined N-acetylneuraminyl-lactose--wheat germ agglutinin isolectin complexes. J Mol Biol *215*, 635-651.
Xie, Y., Han, Y., Wu, D.P., Sun, A.N., Cen, J.N., Yao, L., and Ruan, C.G. (2009). [The BK virus load in urine in association with the development of hemorrhagic cystitis after hematopoietic stem cell transplantation]. Zhonghua Xue Ye Xue Za Zhi *30*, 524-527.
Yu, R.K., Nakatani, Y., and Yanagisawa, M. (2009). The role of glycosphingolipid metabolism in the developing brain. J Lipid Res *50 Suppl*, S440-445.
Zurhein, G., and Chou, S.M. (1965). Particles Resembling Papova Viruses in Human Cerebral Demyelinating Disease. Science *148*, 1477-1479.

# 6. Appendix

## 6.1. Sequences and structures

The following sequences and structures were used in this thesis:

| | Sequence | pdb |
|---|---|---|
| SV40 virion | P03087 (UniProt) | 1VPS |
| SV40 pentamer (+GM1) | P03087 (UniProt) | 3BWQ (3BWR) |
| JCV Mad-1 pentamer (+LSTc) | P03089 (UniProt) | 3NXG (3NXD) |
| JCV WT1 | AAQ88264 (GenBank) | |
| BKV pentamer (+GD3) | ABC18005 (GenBank) | |
| Polyoma virion RA<br>+ 3'-sialyllactose<br>(+ branched oligosaccharide) | P49302 (UniProt) | 1SID (1SIE) |
| Polyoma pentamer<br>(+ branched oligosaccharide) | P49302 (UniProt) | 1VPN (1VPS) |
| Polyoma pentamer + VP2 | P49302 (UniProt) | 1CN3 |
| Polyoma VP1 PTA | AAA93236 (GenBank) | |
| Influenza HA<br>+ sialylated oligosaccharide | P03437 (UniProt) | 1HGG |
| Ad37 + 3'-sialyllactose | Q64823 (UniProt) | 1UXA |
| Rotavirus VP8*<br>+ Methyl-sialoside | P12473 (UniProt) | 1KQR |
| Cholera toxin + GM1 | P01556 (UniProt) | 3CHB |
| Shiga toxin mutant + Gb3 | P69178 (UniProt) | 1CQF |
| Shiga-like toxin + STARFISH | P69178 (UniProt) | 1QNU |
| Siglec-7 + GT1b analogon | Q9Y286 (UniProt) | 2HRL |
| Sialoadhesin +3-sialyllactose | Q62230 (UniProt) | 1QFO |

## 6.2. Crystallographic statistics

### 6.2.1. SV40 VP1 and SV40 VP1-GM1

| | SV40 VP1 | | SV40 VP1-GM1 | |
|---|---|---|---|---|
| Unit cell axes [Å] | 153.2, 93.0, 119.2 | | 153.3, 93.9, 119.4 | |
| Unit cell angles [°] | 90, 105.5, 90 | | 90, 105.6, 90 | |
| Space group | C2 | | C2 | |
| Matthews coefficient [$Å^3$/Da] | 2.8 | | 2.8 | |
| Pentamers per ASU | 1 | | 1 | |
| Dataset quality | | | | |
| Resolution | 50-2.3 | 2.38-2.30 | 30-2.25 | 2.33-2.25 |
| Reflections total | 308,478 | 4,392 | 597,014 | 12,384 |
| Reflections unique | 63,895 | 2,440 | 68,387 | 2,580 |
| Rmerge [%] | 9.9 | 26.0 | 9.7 | 27.6 |
| Completeness [%] | 86.8 | 52.2 | 89.8 | 53.8 |
| I/σI | 6.7 | 1.4 | 11.2 | 1.9 |
| Refinement quality | | | | |
| $R_{work}$ [%] | 22.0 | 29.9 | 18.7 | 24.4 |
| $R_{free}$ [%] | 26.4 | 34.8 | 23.4 | 33.0 |
| r.m.s.d. bond lengths [Å] | 0.011 | | 0.005 | |
| r.m.s.d. bond angles [°] | 1.39 | | 0.8 | |
| Protein atoms | 10,157 | | 10,214 | |
| Average B-factor [$Å^2$] | 39.2 | | 33.8 | |
| Carbohydrate atoms | - | | 204 | |
| Average B-factor [$Å^2$] | - | | 41.7 | |
| Solvent atoms | 584 | | 925 | |
| Average B-factor [$Å^2$] | 33.2 | | 33.9 | |
| B-factor Wilson [$Å^2$] | 33.0 | | 33.9 | |

### 6.2.2. JCV VP1 and JCV VP1-LSTc

| | JCV VP1 | | JCV VP1-LSTc | |
|---|---|---|---|---|
| Unit cell axes [Å] | 149.6, 96.0, 128.6 | | 149.7, 96.3, 128.2 | |
| Unit cell angles [°] | 90, 110.4, 90 | | 90, 110.3, 90 | |
| Space group | C2 | | C2 | |
| Matthews coefficient [$Å^3$/Da] | 2.9 | | 2.9 | |
| Pentamers per ASU | 1 | | 1 | |
| Dataset quality | | | | |
| Resolution [Å] | 20 – 1.95 | 2.00 – 1.95 | 35 – 2.0 | 2.05 – 2.00 |
| Reflections total | 419,669 | 15,867 | 437,707 | 30,603 |
| Reflections unique | 119,853 | 8,429 | 114,938 | 8,302 |
| $R_{merge}$ [%] | 9.3 | 32.4 | 11.7 | 48.3 |
| Completeness [%] | 96.5 | 91.7 | 99.5 | 97.5 |
| I/σI | 10.9 | 2.5 | 9.6 | 2.7 |
| Refinement quality | | | | |
| $R_{work}$ [%] | 18.0 | 21.3 | 18.9 | 22.6 |
| $R_{free}$ [%] | 20.8 | 24.7 | 21.2 | 26.1 |
| r.m.s.d. bond lengths [Å] | 0.007 | | 0.006 | |
| r.m.s.d. bond angles [°] | 1.0 | | 0.9 | |
| Protein atoms | 10,135 | | 10,200 | |
| Average B-factor [$Å^2$] | 15.7 | | 16.6 | |
| Carbohydrate atoms | - | | 136 | |
| Average B-factor [$Å^2$] | - | | 24.3 | |
| Solvent atoms | 1,054 | | 980 | |
| Average B-factor [$Å^2$] | 26.4 | | 25.4 | |
| B-factor Wilson [$Å^2$] | 22.0 | | 23.5 | |

### 6.2.3. BKV VP1 crystal form 1

| BKV VP1 crystal form 1 | | |
|---|---|---|
| Unit cell axes [Å] | a=242.2, b=447.4, c=155.9 | |
| Unit cell angles [°] | 90, 90, 90 | |
| Space group | $C222_1$ | |
| Matthews coefficient [$Å^3$/Da] | 4.7 | |
| Pentamers per ASU | 3 | |
| Dataset quality | | |
| Resolution | 30-2.9 | 3.0-2.9 |
| Reflections total | 573,925 | 15,721 |
| Reflections unique | 171,461 | 8,274 |
| $R_{merge}$ [%] | 14.9 | 36.8 |
| Completeness [%] | 94.5 | 63.8 |
| I/σI | 5.7 | 1.2 |
| Refinement quality | | |
| $R_{work}$ [%] | 22.1 | 32.6 |
| $R_{free}$ [%] | 24.2 | 35.0 |
| r.m.s.d. bond lengths [Å] | 0.007 | |
| r.m.s.d. bond angles [°] | 1.04 | |
| Protein atoms | 32,550 | |
| Average B-factor [$Å^2$] | 50.2 | |
| Sulfate atoms | 185 | |
| Average B-factor [$Å^2$] | 90.1 | |
| Solvent atoms | 130 | |
| Average B-factor [$Å^2$] | 35.6 | |
| B-factor Wilson [$Å^2$] | 63.2 | |

### 6.2.4. BKV VP1 crystal form 2 and BKV VP1-GD3

| | BKV VP1 crystal form2 | | BKV VP1-GD3 | |
|---|---|---|---|---|
| Unit cell axes [Å] | 144.9, 152.3, 62.8 | | 144.7, 152.6, 63.2 | |
| Unit cell angles [°] | 90, 90, 90 | | 90, 90, 90 | |
| Space group | $P2_12_12$ | | $P2_12_12$ | |
| Matthews coefficient [$Å^3$/Da] | 2.3 | | 2.3 | |
| Pentamers per ASU | 1 | | 1 | |
| Dataset quality | | | | |
| Resolution [Å] | 50 – 2.0 | 2.05 – 2.00 | 48 – 1.7 | 1.74 – 1.70 |
| Reflections total | 413,405 | 30,097 | 1,103,641 | 64,273 |
| Reflections unique | 92,929 | 6,779 | 154,169 | 11,266 |
| $R_{merge}$ [%] | 15.4 | 58.0 | 6.5 | 63.0 |
| Completeness [%] | 97.2 | 97.0 | 100.0 | 99.8 |
| $I/\sigma I$ | 8.88 | 2.92 | 19.44 | 2.66 |
| Refinement quality | | | | |
| $R_{work}$ [%] | 18.2 | 22.2 | 17.4 | 23.6 |
| $R_{free}$ [%] | 21.8 | 26.9 | 20.2 | 26.4 |
| r.m.s.d. bond lengths [Å] | 0.011 | | 0.006 | |
| r.m.s.d. bond angles [°] | 1.3 | | 1.1 | |
| Protein atoms | 10,011 | | 10,338 | |
| Average B-factor [$Å^2$] | 21.6 | | 20.2 | |
| Ligand atoms | - | | 186 | |
| Average B-factor [$Å^2$] | - | | 34.7 | |
| Solvent atoms | 934 | | 1,297 | |
| Average B-factor [$Å^2$] | 23.9 | | 22.2 | |
| B-factor Wilson [$Å^2$] | 27.9 | | 26.4 | |

### 6.2.5. WUV and KIV VP1

| | KIV VP1 | | WUV VP1 | |
|---|---|---|---|---|
| Unit cell axes [Å] | 70.3, 82.8, 142.1 | | 166.2, 166.2, 127.6 | |
| Unit cell angles [°] | 87.0, 98.2, 108.8 | | 90, 90, 90 | |
| Space group | P1 | | $P4_32_12$ | |
| Matthews coefficient [$Å^3$/Da] | 2.6 | | 2.9 | |
| Pentamers per ASU | 2 | | 1 | |
| Dataset quality | | | | |
| Resolution [Å] | 50 – 2.55 | 2.62 – 2.55 | 40 – 2.9 | 2.98 – 2.90 |
| Reflections total | 237,505 | 16,282 | 159,898 | 10,139 |
| Reflections unique | 93,982 | 6,639 | 39,647 | 2,908 |
| $R_{merge}$ [%] | 5.4 | 60.2 | 9.6 | 57.8 |
| Completeness [%] | 95.9 | 91.1 | 98.8 | 99.4 |
| I/σI | 11.9 | 1.4 | 11.1 | 1.8 |
| Refinement quality | | | | |
| $R_{work}$ [%] | 22.7 | 34.5 | 22.2 | 33.0 |
| $R_{free}$ [%] | 24.4 | 36.7 | 25.0 | 35.7 |
| r.m.s.d. bond lengths [Å] | 0.009 | | 0.009 | |
| r.m.s.d. bond angles [°] | 1.2 | | 1.2 | |
| Protein atoms | 19,845 | | 9,966 | |
| Average B-factor [$Å^2$] | 56.2 | | 58.0 | |
| Ligand atoms | - | | - | |
| Average B-factor [$Å^2$] | - | | - | |
| Solvent atoms | - | | - | |
| Average B-factor [$Å^2$] | - | | - | |
| B-factor Wilson [$Å^2$] | 60.6 | | 58.3 | |

## 6.3. Oligosaccharide sequences used in glycan array screening

### 6.3.1. CFG printed array version 2.1

| Position | Sequence |
|---|---|
| 1 | Alpha1-acid glycoprotein (AGP) |
| 2 | AGP-A (AGP ConA flowthrough) |
| 3 | AGP-B (AGP ConA bound) |
| 4 | Ceruloplasmin |
| 5 | Fibrinogen |
| 6 | Transferrin |
| 7 | α-D-Gal–Sp8 |
| 8 | α-D-Glc–Sp8 |
| 9 | α-D-Man–Sp8 |
| 10 | α-GalNAc–Sp8 |
| 11 | α-L-Fuc–Sp8 |
| 12 | α-L-Fuc–Sp9 |
| 13 | α-L-Rhα–Sp8 |
| 14 | α-Neu5Ac–Sp8 |
| 15 | Neu5Acα1-2–Sp82 |
| 16 | b-Neu5Ac-Sp8 |
| 17 | β-D-Gal–Sp8 |
| 18 | β-D-Glc–Sp8 |
| 19 | β-D-Man–Sp8 |
| 20 | β-GalNAc–Sp8 |
| 21 | β-GlcNAc–Sp0 |
| 22 | β-GlcNAc–Sp8 |
| 23 | b-GlcN(Gc)-Sp8 |
| 24 | (Galβ1-4GlcNAcβ)$_2$-3,6-GalNAcα–Sp8 |

| Position | Sequence |
|---|---|
| 25 | (GlcNAcb1-3(GlcNAcb1-6)GlcNAcb1-4)GlcNAc-Sp8 |
| 26 | [3OSO3][6OSO3]Galb1-4[6OSO3]GlcNAcb-Sp0 |
| 27 | [3OSO3][6OSO3]Galb1-4GlcNAcb-Sp0 |
| 28 | [3OSO3]Galb1-4Glcb-Sp8 |
| 29 | [3OSO3]Galβ1-4(6OSO3)Glcβ–Sp0 |
| 30 | [3OSO3]Galβ1-4(6OSO3)Glcβ–Sp8 |
| 31 | [3OSO3]Galβ1-3(Fucα1-4)GlcNAcβ–Sp8 |
| 32 | [3OSO3]Galβ1-3GalNAcα–Sp8 |
| 33 | [3OSO3]Galβ1-3GlcNAcβ–Sp8 |
| 34 | [3OSO3]Galβ1-4(Fucα1-3)GlcNAcβ–Sp8 |
| 35 | [3OSO3]Galb1-4[6OSO3]GlcNAcb-Sp8 |
| 36 | [3OSO3]Galβ1-4GlcNAcβ–Sp0 |
| 37 | [3OSO3]Galb1-4GlcNAcb-Sp8 |
| 38 | [3OSO3]Galβ–Sp8 |
| 39 | [4OSO3][6OSO3]Galb1-4GlcNAcb-Sp0 |
| 40 | [4OSO3]Galb1-4GlcNAcb-Sp8 |
| 41 | 6-$H_2PO_3$Manα–Sp8 |
| 42 | [6OSO3]Galβ1-4Glcβ–Sp0 |
| 43 | [6OSO3]Galβ1-4Glcβ–Sp8 |
| 44 | [6OSO3]Galβ1-4GlcNAcβ–Sp8 |
| 45 | [6OSO3]Galb1-4[6OSO3]Glcb-Sp8 |
| 46 | NeuAcα2-3[6OSO3]Galβ1-4GlcNAcβ–Sp8 |
| 47 | [6OSO3]GlcNAcβ–Sp8 |
| 48 | 9NAcNeu5Aca-Sp8 |
| 49 | 9NAcNeu5Aca2-6Galb1-4GlcNAcb-Sp8 |
| 50 | Mana1-3(Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Gly |
| 51 | GlcNAcb1-2Mana1-3(GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Gly |

| Position | Sequence |
|---|---|
| 52 | Galb1-4GlcNAcb1-2Mana1-3(Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Gly |
| 53 | Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Gly |
| 54 | Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp8 |
| 55 | Fuca1-2Galb1-3GalNAcb1-3Gala-Sp9 |
| 56 | Fuca1-2Galb1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp9 |
| 57 | Fucα1-2Galβ1-3(Fucα1-4)GlcNAcβ–Sp8 |
| 58 | Fucα1-2Galβ1-3GalNAcα–Sp8 |
| 59 | Fuca1-2Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp0 |
| 60 | Fucα1-2Galβ1-3GalNAcβ1-4(Neu5Acα2-3)Galβ1-4Glcβ-Sp9 |
| 61 | Fucα1-2Galβ1-3GlcNAcβ1-3Galβ1-4Glcβ–Sp10 |
| 62 | Fucα1-2Galβ1-3GlcNAcβ1-3Galβ1-4Glcβ–Sp8 |
| 63 | Fucα1-2Galβ1-3GlcNAcβ–Sp0 |
| 64 | Fucα1-2Galβ1-3GlcNAcβ–Sp8 |
| 65 | Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0 |
| 66 | Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0 |
| 67 | Fucα1-2Galβ1-4(Fucα1-3)GlcNAcβ–Sp0 |
| 68 | Fucα1-2Galβ1-4(Fucα1-3)GlcNAcβ–Sp8 |
| 69 | Fucα1-2Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAc–Sp0 |
| 70 | Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0 |
| 71 | Fucα1-2Galβ1-4GlcNAcβ–Sp0 |
| 72 | Fucα1-2Galβ1-4GlcNAcβ–Sp8 |
| 73 | Fucα1-2Galβ1-4Glcβ–Sp0 |
| 74 | Fucα1-2Galβ–Sp8 |
| 75 | Fuca1-2GlcNAcb-Sp8 |
| 76 | Fucα1-3GlcNAcβ–Sp8 |
| 77 | Fucα1-4GlcNAcβ–Sp8 |

| Position | Sequence |
|---|---|
| 78 | Fucb1-3GlcNAcb-Sp8 |
| 79 | GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb-Sp0 |
| 80 | GalNAca1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp0 |
| 81 | GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0 |
| 82 | GalNAcα1-3(Fucα1-2)Galβ1-4GlcNAcβ–Sp8 |
| 83 | GalNAca1-3(Fuca1-2)Galb1-4Glcb-Sp0 |
| 84 | GalNAcα1-3(Fucα1-2)Galβ–Sp8 |
| 85 | GalNAcα1-3GalNAcb–Sp8 |
| 86 | GalNAcα1-3Galb–Sp8 |
| 87 | GalNAca1-4(Fuca1-2)Galb1-4GlcNAcb-Sp8 |
| 88 | GalNAcb1-3GalNAca–Sp8 |
| 89 | GalNAcb1-3(Fuca1-2)Galb-Sp8 |
| 90 | GalNAcb1-3Gala1-4Galb1-4GlcNAcb-Sp0 |
| 91 | GalNAcb1-4(Fuca1-3)GlcNAcb-Sp0 |
| 92 | GalNAcβ1-4GlcNAcβ–Sp0 |
| 93 | GalNAcβ1-4GlcNAcβ–Sp8 |
| 94 | Galα1-2Galβ–Sp8 |
| 95 | Gala1-3(Fuca1-2)Galb1-3GlcNAcb-Sp0 |
| 96 | Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp0 |
| 97 | Gala1-3(Fuca1-2)Galb1-4GlcNAc-Sp0 |
| 98 | Gala1-3(Fuca1-2)Galb1-4Glcb-Sp0 |
| 99 | Galα1-3(Fucα1-2)Galβ–Sp8 |
| 100 | Gala1-3(Gala1-4)Galb1-4GlcNAcb-Sp8 |
| 101 | Gala1-3GalNAca-Sp8 |
| 102 | Galα1-3GalNAcβ–Sp8 |
| 103 | Galα1-3Galβ1-4(Fucα1-3)GlcNAcβ–Sp8 |
| 104 | Gala1-3Galb1-3GlcNAcb-Sp0 |

| Position | Sequence |
|---|---|
| 105 | Galα1-3Galβ1-4GlcNAcβ–Sp8 |
| 106 | Galα1-3Galβ1-4Glcβ–Sp0 |
| 107 | Galα1-3Galβ–Sp8 |
| 108 | Gala1-4(Fuca1-2)Galb1-4GlcNAcb-Sp8 |
| 109 | Galα1-4Galβ1-4GlcNAcβ–Sp0 |
| 110 | Galα1-4Galβ1-4GlcNAcβ–Sp8 |
| 111 | Galα1-4Galβ1-4Glcβ–Sp0 |
| 112 | Galα1-4GlcNAcb–Sp8 |
| 113 | Gala1-6Glcb-Sp8 |
| 114 | Galβ1-2Galβ–Sp8 |
| 115 | Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0 |
| 116 | Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4GlcNAcb-Sp0 |
| 117 | Galβ1-3(Fucα1-4)GlcNAc–Sp0 |
| 118 | Galβ1-3(Fucα1-4)GlcNAc–Sp8 |
| 119 | Galβ1-3(Fucα1-4)GlcNAcβ–Sp8 |
| 120 | Galβ1-3(Galβ1-4GlcNAcβ1-6)GalNAcα-Sp8 |
| 121 | Galβ1-3(GlcNAcβ1-6)GalNAcα-Sp8 |
| 122 | Galb1-3(Neu5Aca2-6)GalNAca-Sp8 |
| 123 | Galb1-3(Neu5Acb2-6)GalNAca-Sp8 |
| 124 | Galb1-3(Neu5Aca2-6)GlcNAcb1-4Galb1-4Glcb-Sp10 |
| 125 | Galβ1-3GalNAcα-Sp8 |
| 126 | Galβ1-3GalNAcβ–Sp8 |
| 127 | Galb1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp0 |
| 128 | Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glcb-Sp0 |
| 129 | Galβ1-3GalNAcβ1-4Galβ1-4Glcβ–Sp8 |
| 130 | Galβ1-3Galβ–Sp8 |
| 131 | Galb1-3GlcNAcb1-3Galb1-4GlcNAcb-Sp0 |

| Position | Sequence |
|---|---|
| 132 | Galβ1-3GlcNAcβ1-3Galβ1-4Glcβ–Sp10 |
| 133 | Galβ1-3GlcNAcβ–Sp0 |
| 134 | Galβ1-3GlcNAcβ–Sp8 |
| 135 | Galβ1-4(Fucα1-3)GlcNAcb–Sp0 |
| 136 | Galβ1-4(Fucα1-3)GlcNAcb–Sp8 |
| 137 | Galb1-4(Fuca1-3)GlcNAcb1-4Galb1-4(Fuca1-3)GlcNAcb-Sp0 |
| 138 | Galβ1-4(Fucα1-3)GlcNAcβ1-4Galβ1-4(Fucα1-3)GlcNAcβ1-4Galβ1-4(Fucα1-3)GlcNAcβ–Sp0 |
| 139 | Galβ1-4[6OSO3]Glcβ–Sp0 |
| 140 | Galβ1-4[6OSO3]Glcβ–Sp8 |
| 141 | Galb1-4GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8 |
| 142 | Galb1-4GalNAcb1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8 |
| 143 | Galb1-4GlcNAcb1-3(Galb1-4GlcNAcb1-6)GalNAca-Sp8 |
| 144 | Galβ1-4GlcNAcβ1-3GalNAcα–Sp8 |
| 145 | Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0 |
| 146 | Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ–Sp0 |
| 147 | Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ–Sp0 |
| 148 | Galβ1-4GlcNAcβ1-3Galβ1-4Glcβ–Sp0 |
| 149 | Galβ1-4GlcNAcβ1-3Galβ1-4Glcβ–Sp8 |
| 150 | Galβ1-4GlcNAcβ1-6(Galβ1-3)GalNAcα–Sp8 |
| 151 | Galβ1-4GlcNAcβ1-6GalNAcα–Sp8 |
| 152 | Galβ1-4GlcNAcβ–Sp0 |
| 153 | Galβ1-4GlcNAcβ–Sp8 |
| 154 | Galβ1-4Glcβ–Sp0 |
| 155 | Galβ1-4Glcβ–Sp8 |
| 156 | GlcNAca1-3Galb1-4GlcNAcb-Sp8 |
| 157 | GlcNAca1-6Galb1-4GlcNAcb-Sp8 |
| 158 | GlcNAcβ1-2Galβ1-3GalNAcα–Sp8 |

| Position | Sequence |
|---|---|
| 159 | GlcNAcβ1-3(GlcNAcβ1-6)GalNAcα–Sp8 |
| 160 | GlcNAcβ1-3(GlcNAcβ1-6)Galb1-4GlcNAcb–Sp8 |
| 161 | GlcNAcβ1-3GalNAcα–Sp8 |
| 162 | GlcNAcb1-3Galb-Sp8 |
| 163 | GlcNAcb1-3Galb1-3GalNAca-Sp8 |
| 164 | GlcNAcβ1-3Galβ1-4GlcNAcβ–Sp0 |
| 165 | GlcNAcb1-3Galb1-4GlcNAcb-Sp8 |
| 166 | GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0 |
| 167 | GlcNAcβ1-3Galβ1-4Glcβ–Sp0 |
| 168 | GlcNAcb1-4MDPLys (bacterial cell wall) |
| 169 | GlcNAcb1-4(GlcNAcb1-6)GalNAca-Sp8 |
| 170 | GlcNAcb1-4Galb1-4GlcNAcb-Sp8 |
| 171 | $(GlcNAcb1\text{-}4)_6$β-Sp8 |
| 172 | $(GlcNAcb1\text{-}4)_5$β-Sp8 |
| 173 | GlcNAcβ1-4GlcNAcβ1-4GlcNAcβ–Sp8 |
| 174 | GlcNAcβ1-6(Galβ1-3)GalNAcα–Sp8 |
| 175 | GlcNAcβ1-6GalNAcα–Sp8 |
| 176 | GlcNAcb1-6Galb1-4GlcNAcb-Sp8 |
| 177 | Glcα1-4Glcβ–Sp8 |
| 178 | Glcα1-4Glca–Sp8 |
| 179 | Glcα1-6Glcα1-6Glcβ-Sp8 |
| 180 | Glcb1-4Glcb-Sp8 |
| 181 | Glcb1-6Glcb-Sp8 |
| 182 | G-ol-amine |
| 183 | GlcAa-Sp8 |
| 184 | GlcAb-Sp8 |
| 185 | GlcAb1-3Galb-Sp8 |

| Position | Sequence |
|---|---|
| 186 | GlcAb1-6Galb-Sp8 |
| 187 | KDNα2-3Galβ1-3GlcNAcβ–Sp0 |
| 188 | KDNα2-3Galβ1-4GlcNAcβ–Sp0 |
| 189 | Mana1-2Mana1-2Mana1-3Mana-Sp9 |
| 190 | Mana1-2Mana1-3(Mana1-2Mana1-6)Mana-Sp9 |
| 191 | Mana1-2Mana1-3Mana-Sp9 |
| 192 | Mana1-6(Mana1-2Mana1-3)Mana1-6(Manα2Manα1-3)Manb1-4GlcNAcb1-4GlcNAcb-Asn |
| 193 | Manα1-2Manα1-6(Manα1-3)Manα1-6(Manα2Manα2Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAcβ-Asn |
| 194 | Manα1-2Manα1-2Manα1-3(Manα1-2Manα1-3(Manα1-2Manα1-6)Manα1-6)Manβ1-4GlcNAcβ1-4GlcNAcβ-Asn |
| 195 | Manα1-3(Manα1-6)Manα–Sp9 |
| 196 | Mana1-3(Mana1-2Mana1-2Mana1-6)Mana-Sp9 |
| 197 | Manα1-6(Manα1-3)Manα1-6(Manα2Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAcβ-Asn |
| 198 | Mana1-6(Mana1-3)Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4 GlcNAcb-Asn |
| 199 | Man5-9mix-Asn |
| 200 | Manb1-4GlcNAcb-Sp0 |
| 201 | Neu5Aca2-3(Galb1-3GalNAcb1-4)Galb1-4Glcb-Sp0 |
| 202 | Neu5Aca2-3Galb1-3GalNAca-Sp8 |
| 203 | NeuAca2-8NeuAca2-8NeuAca2-8NeuAca2-3(GalNAcb1-4)Galb1-4Glcb-Sp0 |
| 204 | Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3(GalNAcb1-4)Galb1-4Glcb-Sp0 |
| 205 | Neu5Aca2-8Neu5Aca2-8Neu5Acα2-3Galβ1-4Glcβ–Sp0 |
| 206 | Neu5Aca2-8Neu5Acα2-3(GalNAcβ1-4)Galβ1-4Glcβ–Sp0 |
| 207 | Neu5Acα2-8Neu5Acα2-8Neu5Acα-Sp8 |
| 208 | Neu5Acα2-3(6-O-Su)Galβ1-4(Fucα1-3)GlcNAcβ–Sp8 |
| 209 | Neu5Aca2-3(GalNAcb1-4)Galb1-4GlcNAcb-Sp0 |
| 210 | Neu5Aca2-3(GalNAcb1-4)Galb1-4GlcNAcb-Sp8 |
| 211 | Neu5Acα2-3(GalNAcβ1-4)Galβ1-4Glcβ–Sp0 |

| Position | Sequence |
|---|---|
| 212 | NeuAca2-3(NeuAca2-3Galb1-3GalNAcb1-4)Galb1-4Glcb-Sp0 |
| 213 | Neu5Acα2-3(Neu5Acα2-6)GalNAcα–Sp8 |
| 214 | Neu5Acα2-3GalNAcα–Sp8 |
| 215 | Neu5Aca2-3GalNAcb1-4GlcNAcb-Sp0 |
| 216 | Neu5Acα2-3Galβ1-3(6OSO3)GlcNAc-Sp8 |
| 217 | Neu5Acα2-3Galb1-3(Fucα1-4)GlcNAcβ–Sp8 |
| 218 | NeuAca2-3Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0 |
| 219 | Neu5Acα2-3Galβ1-3(Neu5Acα2-3Galβ1-4)GlcNAcβ-Sp8 |
| 220 | Neu5Aca2-3Galb1-3[6OSO3]GalNAca-Sp8 |
| 221 | Neu5Acα2-3Galβ1-3(Neu5Acα2-6)GalNAcα–Sp8 |
| 222 | Neu5Aca2-3Galb-Sp8 |
| 223 | NeuAca2-3Galb1-3GalNAcb1-3Gala1-4Galb1-4Glcb-Sp0 |
| 224 | NeuAca2-3Galb1-3GlcNAcb1-3Galb1-4GlcNAcb-Sp0 |
| 225 | Neu5Acα2-3Galβ1-3GlcNAcβ–Sp0 |
| 226 | Neu5Acα2-3Galβ1-3GlcNAcβ–Sp8 |
| 227 | Neu5Aca2-3Galb1-4[6OSO3]GlcNAcb-Sp8 |
| 228 | Neu5Acα2-3Galβ1-4(Fucα1-3)(6OSO3)GlcNAcβ–Sp8 |
| 229 | Neu5Acα2-3Galβ1-4(Fucα1-3)GlcNAcβ1-3Galβ1-4(Fucα1-3)GlcNAcβ1-3Galβ1-4(Fucα1-3)GlcNAcβ–Sp0 |
| 230 | Neu5Acα2-3Galβ1-4(Fucα1-3)GlcNAcβ–Sp0 |
| 231 | Neu5Acα2-3Galβ1-4(Fucα1-3)GlcNAcβ–Sp8 |
| 232 | Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb-Sp8 |
| 233 | Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4GlcNAcb-Sp8 |
| 234 | Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAc-Sp0 |
| 235 | Neu5Acα2-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ–Sp0 |
| 236 | Neu5Acα2-3Galβ1-4GlcNAcβ–Sp0 |
| 237 | Neu5Acα2-3Galβ1-4GlcNAcβ–Sp8 |

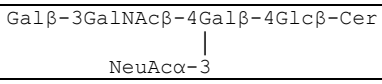
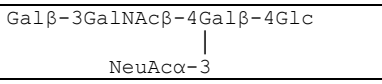
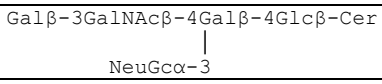
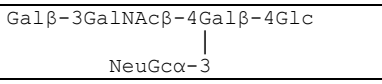
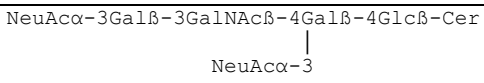
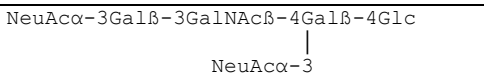
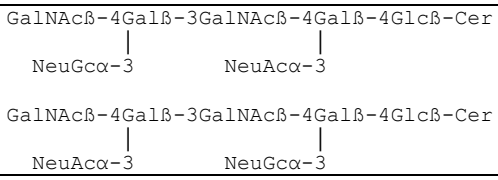
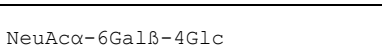
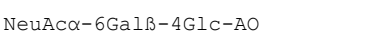
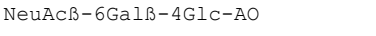
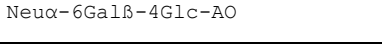
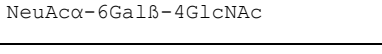
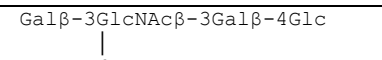
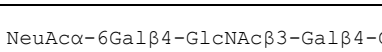
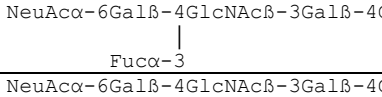
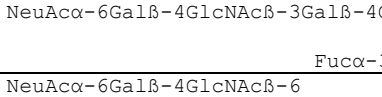
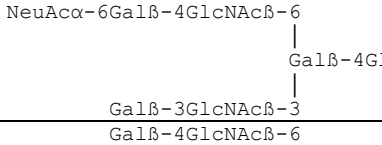
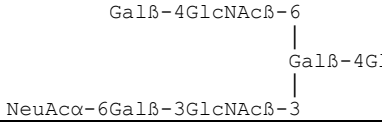
| Position | Sequence |
|---|---|
| 238 | Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0 |
| 239 | Neu5Acα2-3Galβ1-4Glcβ–Sp0 |
| 240 | Neu5Acα2-3Galβ1-4Glcβ–Sp8 |
| 241 | Neu5Acα2-6(Galβ1-3)GalNAcα–Sp8 |
| 242 | Neu5Acα2-6GalNAcα–Sp8 |
| 243 | Neu5Aca2-6GalNAcb1-4GlcNAcb-Sp0 |
| 244 | Neu5Aca2-6Galb1-4[6OSO3]GlcNAcb-Sp8 |
| 245 | Neu5Acα2-6Galβ1-4GlcNAcβ–Sp0 |
| 246 | Neu5Acα2-6Galβ1-4GlcNAcβ–Sp8 |
| 247 | Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0 |
| 248 | Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0 |
| 249 | Neu5Acα2-6Galβ1-4Glcβ–Sp0 |
| 250 | Neu5Acα2-6Galβ1-4Glcβ–Sp8 |
| 251 | Neu5Acα2-6Galβ–Sp8 |
| 252 | Neu5Acα2-8Neu5Acα-Sp8 |
| 253 | Neu5Acα2-8Neu5Acα2-3Galβ1-4Glcβ–Sp0 |
| 254 | Neu5Acβ1-6GalNAcα–Sp8 |
| 255 | Neu5Acb2-6Galb1-4GlcNAcb-Sp8 |
| 256 | Neu5Acβ2-6(Galβ1-3)GalNAcα–Sp8 |
| 257 | Neu5Gca2-3Galb1-3(Fuca1-4)GlcNAcb-Sp0 |
| 258 | Neu5Gca2-3Galb1-3GlcNAcb-Sp0 |
| 259 | Neu5Gca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp0 |
| 260 | Neu5Gcα2-3Galβ1-4GlcNAcβ–Sp0 |
| 261 | Neu5Gcα2-3Galβ1-4Glcβ–Sp0 |
| 262 | Neu5Gcα2-6GalNAcα–Sp0 |
| 263 | Neu5Gcα2-6Galβ1-4GlcNAcβ–Sp0 |
| 264 | Neu5Gcα–Sp8 |

| Spacer | Sequence |
|---|---|
| Sp0 | -$CH_2CH_2NH_2$ |
| Sp8 | -$CH_2CH_2CH_2NH_2$ |
| Sp9 | -$CH_2CH_2CH_2CH_2CH_2NH_2$ |
| Sp10 | -$NHCOCH_2NH$ |
| Sp11 | -$OCH_2C_6H_4$-p-$NHCOCH_2NH$ |
| mMDPLys | -Mur-L-Ala-D-iGlnb-$(CH_2)_4NH_2$ |
| T | Threonine |
| N | Asparagine |

## 6.3.2. Lipid-linked array of sialylated oligosaccharides

| Position | Probe | Sequence |
|---|---|---|
| 1 | LacNAc | Galß-4GlcNAc |
| 2 | LNT | Galß-3GlcNAcß-3Galß-4Glc |
| 3 | LNnT | Galß-4GlcNAcß-3Galß-4Glc |
| 4 | LNFP-III | Galß-4GlcNAcß-3Galß-4Glc<br>\|<br>Fucα-3 |
| 5 | LNH | Galß-4GlcNAcß-6<br>\|<br>Galß-4Glc<br>\|<br>Galß-3GlcNAcß-3 |
| 6 | NA2 | Galß-4GlcNAcß-2Manα-6<br>\|<br>Manß-4GlcNAcß-4GlcNAc<br>\|<br>Galß-4GlcNAcß-2Manα-3 |
| 7 | GM4 | NeuAcα-3Galβ-Cer |
| 8 | GM3 | NeuAcα-3Galß-4Glcß-Cer |
| 9 | GM3(Gc) | NeuGcα-3Galß-4Glc-Cer |
| 10 | Haematoside | NeuAcα-3Galß-4Glcß-Cer |
| 11 | NeuAcα-(3')Lac | NeuAcα-3Galß-4Glc |
| 12 | NeuAcα-(3')Lac-AO | NeuAcα-3Galß-4Glc-AO |
| 13 | NeuAcß-(3')Lac-AO | NeuAcß-3Galß-4Glc-AO |
| 14 | Neuα-(3')Lac-AO | Neuα-3Galß-4Glc-AO |
| 15 | Neu4,5Ac-(3')Lac | Neu4,5Acα-3Galß-4Glc |
| 16 | Neu4,5Ac-(3')Lac-AO | Neu4,5Acα-3Galß-4Glc-AO |
| 17 | NeuAcα-(3')LN1-3 | NeuAcα-3Galß-3GlcNAc |
| 18 | NeuAcα-(3')LN1-3-AO | NeuAcα-3Galß-3GlcNAc-AO |
| 19 | NeuAcα-(3')LN | NeuAcα-3Galß-4GlcNAc |
| 20 | NeuAcα-(3')LN-AO | NeuAcα-3Galß-4GlcNAc-AO |
| 21 | SA(3')-Lea-Tri-AO | NeuAcα-3Galß-3GlcNAc-AO<br>\|<br>Fucα-4 |
| 22 | SA(3')-Lex-Tri-AO | NeuAcα-3Galß-4GlcNAc-AO<br>\|<br>Fucα-3 |
| 23 | LSTa | NeuAcα-3Galß-3GlcNAcß-3Galß-4Glc |
| 24 | Sialylparagloboside | NeuAcα-3Galß-4GlcNAcß-3Galß-4Glcß-Cer |

| Position | Probe | Sequence |
|---|---|---|
| 25 | GSC-31 | NeuAcα-3Galß-4GlcNAcß-3Galß-4Glcß-Cer36 |
| 26 | GSC-516B | Neuα-3Galß-4GlcNAcß-3Galß-4Glcß-Cer36<br>SU-6 |
| 27 | C4U | NeuAcα-3Galß-4GlcNAcβ-3Galβ-3GlcNAc<br>SU-6 SU-6 SU-6 |
| 28 | SA(3')-LNFP-II | NeuAcα-3Galß-3GlcNAcß-3Galß-4Glc<br>Fucα-4 |
| 29 | SA(3')-LNFP-III | NeuAcα-3Galß-4GlcNAcß-3Galß-4Glc<br>Fucα-3 |
| 30 | GSC-105 | NeuAcα-3Galß-4GlcNAcß-3Galß-Cer36<br>Fucα-3 |
| 31 | GSC-64 | NeuAcα-3Galß-4GlcNAcß-3Galß-4Glcß-Cer36<br>Fucα-3 |
| 32 | GSC-472 | Neuα-3Galß-4GlcNAcß-3Galß-4Glcß-Cer36<br>Fucα-3 |
| 33 | GSC-268 | SU-6<br>NeuAcα-3Galß-4GlcNAcß-3Galß-4Glcß-Cer36<br>Fucα-3 |
| 34 | GSC-268 deNAc | SU-6<br>Neuα-3Galß-4GlcNß-3Galß-4Glcß-Cer36<br>Fucα-3 |
| 35 | GSC-269 | SU-6<br>NeuAcα-3Galß-4GlcNAcß-3Galß-4Glcß-Cer36<br>Fucα-3 |
| 36 | GSC-406 | SU-6<br>Neuα-3Galß-4GlcNAcß-3Galß-4Glcß-Cer36<br>Fucα-3 |
| 37 | GSC-270 | SU-6 SU-6<br>NeuAcα-3Galß-4GlcNAcß-3Galß-4Glcß-Cer36<br>Fucα-3 |
| 38 | GSC-220 | NeuAcα-3Galß-4GlcNAcß-3Galß-4GlcNAcß-3Galß-4Glcß-Cer36<br>Fucα-3 Fucα-3 |
| 39 | GSC-221 | NeuAcα-3Galß-4GlcNAcß-3Galß-4GlcNAcß-3Galß-4Glcß-Cer36<br>Fucα-3 |
| 40 | MSMFLNH | Galß-4GlcNAcß-6<br>Fucα-3 Galß-4Glc<br>NeuAcα-3Galß-3GlcNAcß-3 |
| 41 | A2F(2-3) | NeuAcα-3Galß-4GlcNAcß-2Manα-6 Fucα-6<br>Manß-4GlcNAcß-4GlcNAc<br>NeuAcα-3Galß-4GlcNAcß-2Manα-3 |
| 42 | GM2 | GalNAcβ-4Galβ-4Glcβ-Cer<br>NeuAcα-3 |
| 43 | GM1b | NeuAcα-3Galβ-3GalNAcß-4Galß-4Glcß-Cer |

| Position | Probe | Sequence |
|---|---|---|
| 44 | GM1 |  |
| 45 | GM1-penta |  |
| 46 | GM1(Gc) |  |
| 47 | GM1(Gc)-penta |  |
| 48 | GD1a |  |
| 49 | GD1a-hexa |  |
| 50 | GalNAc-GD1a(Ac,Gc) |  |
| 51 | NeuAcα-(6')Lac |  |
| 52 | NeuAcα-(6')Lac-AO |  |
| 53 | NeuAcß-(6')Lac-AO |  |
| 54 | Neuα-(6')Lac-AO |  |
| 55 | NeuAcα-(6')LN |  |
| 56 | NeuAcα-(6')LN-AO |  |
| 57 | Neu5,9Ac-(6')LN |  |
| 58 | LSTb |  |
| 59 | LSTc |  |
| 60 | GSC-97 |  |
| 61 | SA(6')-LNFP-VI |  |
| 62 | MSLNH |  |
| 63 | MSLNnH-I |  |

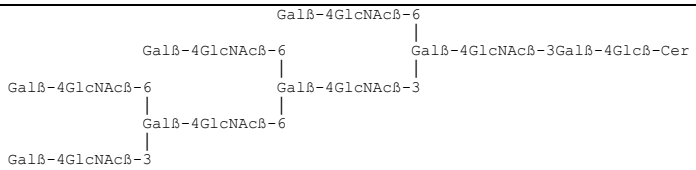
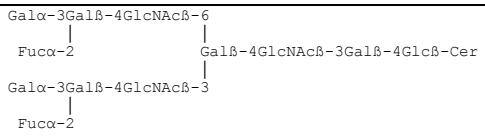
| Position | Probe | Sequence |
|---|---|---|
| 64 | DSLNnH | NeuAcα-6Galß-4GlcNAcß-6<br>\|<br>Galß-4Glc<br>\|<br>NeuAcα-6Galß-4GlcNAcß-3 |
| 65 | MFMSLNnH | Galß-4GlcNAcß-6<br>\|<br>Fucα-3 Galß-4Glc<br>\|<br>NeuAcα-6Galß-3GlcNAcß-3 |
| 66 | A2(2-6) | NeuAcα-6Galß-4GlcNAcß-2Manα-6<br>\|<br>Manß-4GlcNAcß-4GlcNAc<br>\|<br>NeuAcα-6Galß-4GlcNAcß-2Manα-3 |
| 67 | DST | NeuAcα-3Galß-3GalNAc<br>\|<br>NeuAcα-6 |
| 68 | DSLNT | NeuAcα-3Galß-3GlcNAcß-3Galß-4Glc<br>\|<br>NeuAcα-6 |
| 69 | A3 | NeuAcα-3Galß-4GlcNAcß-2Manα-6<br>\|<br>Manß-4GlcNAcß-4GlcNAc<br>\|<br>NeuAcα-3Galß-4GlcNAcß-4Manα-3<br>\|<br>NeuAcα-6Galß-4GlcNAcß-2 |
| 70 | GD3 | NeuAcα-8NeuAcα-3Galß-4Glcß-Cer |
| 71 | GD3-tetra | NeuAcα-8NeuAcα-3Galß-4Glc |
| 72 | GD3-tetra-AO | NeuAcα-8NeuAcα-3Galß-4Glc-AO |
| 73 | GD2 | GalNAcß-4Galß-4Glcß-Cer<br>\|<br>NeuAcα-8NeuAcα-3 |
| 74 | GD1b | Galß-3GalNAcß-4Galß-4Glcß-Cer<br>\|<br>NeuAcα-8NeuAcα-3 |
| 75 | GQ1b | NeuAcα-8NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer<br>\|<br>NeuAcα-8NeuAcα-3 |
| 76 | SA2(α8) | NeuAcα-8NeuAc |
| 77 | SA3(α8) | NeuAcα-8NeuAcα-8NeuAc |
| 78 | SA4(α8) | NeuAcα-8NeuAcα-8NeuAcα-8NeuAc |
| 79 | SA5(α8) | NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAc*** |
| 80 | SA6(α8) | NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAc*** |
| 81 | SA7(α8) | NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAc*** |
| 82 | SA8(α8) | NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAc-8NeuAcα-8NeuAc*** |
| 83 | SA9(α8) | NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAc-8NeuAcα-8NeuAc-8NeuAcα*** |
| 84 | SA10(α8) | NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAc-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAc*** |

| Position | Probe | Sequence |
|---|---|---|
| 85 | SA11(α8) | NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAcα-8NeuAc-8NeuAcα-8NeuAαc-8NeuAcα-8NeuAcα-8NeuAc*** |
| 86 | GT1a | NeuAcα-8NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer<br>NeuAcα-3 |
| 87 | GT1b | NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer<br>NeuAcα-8NeuAcα-3 |

The oligosaccharide probes are all lipid-linked, and are from the collection assembled in the course of research in the lab of Prof. Ten Feizi. DH, designates NGLs prepared from reducing oligosaccharides by reductive amination with the amino lipid, 1,2-dihexadecyl-*sn*-glycero-3-phosphoethanolamine (DHPE; AO, neoglycolipids prepared from reducing oligosaccharides by oxime ligation with an aminooxy (AO) functionalized DHPE; Cer, natural glycolipids with various ceramide moieties; Cer36, synthetic glycolipids with ceramide having a total of 36 carbon atoms.

### 6.3.3. Lipid-linked dose-dependent array of selected sialylated oligosaccharides

| Position | Probe | Sequence |
|---|---|---|
| 1 | GM4 | NeuAcα-3Galß-Cer |
| 2 | GSC-230 | NeuAcα-8NeuAcα-3Galß-Cer36 |
| 3 | GM3 | NeuAcα-3Galß-4Glcß-Cer |
| 4 | GM3(Gc) | NeuGcα-3Galß-4Glc-Cer |
| 5 | GD3 | NeuAcα-8NeuAcα-3Galß-4Glcß-Cer |
| 6 | Asialo-GM2 | GalNAcβ-4Galβ-4Glcβ-Cer |
| 7 | GM2 | GalNAcβ-4Galβ-4Glcβ-Cer<br>\|<br>NeuAcα-3 |
| 8 | GD2 | GalNAcß-4Galß-4Glcß-Cer<br>\|<br>NeuAcα-8NeuAcα-3 |
| 9 | Asialo-GM1 | Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer |
| 10 | Asialo-GM1-Tetra | Galß-3GalNAcß-4Galß-4Glc |
| 11 | GM1 | Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer<br>\|<br>NeuAcα-3 |
| 12 | GM1-penta | Galβ-3GalNAcβ-4Galβ-4Glc<br>\|<br>NeuAcα-3 |
| 13 | GM1(Gc) | Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer<br>\|<br>NeuGcα-3 |
| 14 | GM1(Gc)-penta | Galβ-3GalNAcβ-4Galβ-4Glc<br>\|<br>NeuGcα-3 |
| 15 | GM1b | NeuAcα-3Galβ-3GalNAcß-4Galß-4Glcß-Cer<br>**(also contains a proportion of GM1)** |
| 16 | GD1a | NeuAcα-3Galß-3GalNAcß-4Galß-4Glcß-Cer<br>\|<br>NeuAcα-3 |
| 17 | GD1a-hexa | NeuAcα-3Galß-3GalNAcß-4Galß-4Glc<br>\|<br>NeuAcα-3 |
| 18 | GD1b | Galß-3GalNAcß-4Galß-4Glcß-Cer<br>\|<br>NeuAcα-8NeuAcα-3 |
| 19 | GT1a | NeuAcα-8NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer<br>\|<br>NeuAcα-3 |
| 20 | GT1b | NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer<br>\|<br>NeuAcα-8NeuAcα-3 |
| 21 | GQ1b | NeuAcα-8NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer<br>\|<br>NeuAcα-8NeuAcα-3 |
| 22 | LSTc | NeuAcα-6Galβ4-GlcNAcβ3-Galβ4-Glc |
| 23 | LSTa | NeuAcα-3Galß-3GlcNAcß-3Galß-4Glc |

| Position | Probe | Sequence |
|---|---|---|
| 24 | GSC-189 | GlcAß-3Galß-4GlcNAcß-3Galß-4Glcß-Cer42 |
| 25 | GSC-190 | SU-3GlcAß-3Galß-4GlcNAcß-3Galß-4Glcß-Cer42 |
| 26 | GSC-217 | SU-3GlcAß-3Galß-4GlcNAcß-3Galß-4GlcNAcß-3Galß-4Glcß-Cer42 |
| 27 | I-hexadecaosylceramide |  |
| 28 | B-III dodecaosylceramide |  |

# Acknowledgments

The work presented here is part of a big collaborative effort and I want to thank everybody involved:

Thilo – Thank you for giving me the opportunity to work on this project, for your optimism, patience and support, for prodding me not to be hesitant about crystal fishing (or anything else, for that matter), and for piling on ever more VP1 proteins to this project...

Luisa, Ben, Holger, Michael and Marco – I'm very lucky you work(ed) on the polyomavirus project! Thanks for being such good colleagues!

Walter – Thank you for your enthusiasm about these crystal structures, for setting only almost impossible deadlines, and for the reminder to not "massage things to death".

Melissa, Stacy and Chris – You are such rockstars! It's been a pleasure writing papers to almost impossible deadlines with you. Hope we stay in touch!

Bob – Thanks to you for ever more VP1 proteins, too, and for your tough and helpful questions.

Dennis – Thank you for tirelessly cloning perfect VP1 expression constructs.

Angelina, Yan and Ten – Thank you for being invested to getting glycan array results for each and every VP1 pentamer, and for so many interesting discussions.

David Smith – Thank you for the first glycan array screening experiments, and for nominating SV40 VP1 as glycan binding protein of the year!

Thomas Peters and Bärbel – Thank you for doing such a great job with the STD NMR experiments and for helping me understand the results!

Günther Gauglitz and Karin – Thanks for not giving up during the countless tedious ITC measurements we had to do for SV40 and Polyoma.

In addition, I would like to thank everybody in the Stehle lab for coffee, tea, chocolate, cake and "Kekse" breaks.

Thanks to everybody for doing their share of common duties and for ensuring machines work as they're supposed to so that everybody can get on with their projects.

Thanks to everybody for lab advice, X-ray discussions, helping me out with reagents, listening to my frustrations and being great colleagues.

And a special thanks to Georg, Dirk and Niko for weaving a certainly magic Linux network, and to everybody involved with the robot for keeping this bitchy machine running!

I also want to thank Lila Gierasch and Joanna Swain for setting such great examples on how science is done well and for surprising me by taking me serious even as an undergrad.

Karin, Stephan und Julian – Vielen Dank, dass ihr es mir ermöglicht habt, die Ruhe wiederzufinden, die ich zum Schreiben dieser Arbeit brauchte.

Ancilla und Achim – Ich kann mir nicht vorstellen, dass es irgendwo eine bessere Schwester oder einen besseren Schwager gibt.

Matthias – Ich bin so froh, dass es dich gibt!

# List of Academic Teachers

## Eberhardt Karls Universität Tübingen

| | |
|---|---|
| Prof. Dr. K. Albert | NMR spectroscopy of organic compounds |
| Prof. Dr. C.-F. Bardele | Zoology |
| Dr. E. O. Bayer | General and organic biochemistry |
| Prof. Dr. H. Bisswanger | Enzyme kinetics |
| Prof. Dr. P. Bohley | Biochemistry of cell organelles |
| PD Dr. D. Christen | Physical chemistry |
| Prof. Dr. G. Dodt | Cell culture |
| Prof. Dr. M. Duszenko | General biochemistry |
| Prof. Dr. K. Eisele | Biochemistry |
| Prof. Dr. G. Gauglitz | Physical chemistry |
| Prof. Dr. G. Häfelinger | Organic chemistry |
| Prof. Dr. B. Hamprecht | Biochemistry |
| Prof. Dr. W. Jäger | Law for chemists |
| Prof. Dr. E. Lindner | Inorganic chemistry |
| Prof. Dr. F. Madeo | Yeast genetics |
| Dr. Mayer | Botany |
| Prof. Dr. H. Ninnemann | Plant biochemistry |
| Prof. Dr. A. Nordheim | Cell biology |
| Prof. Dr. H. Oberhammer | Physical chemistry |
| Prof. Dr. C. Ochsenfeld | Theoretical chemistry |
| PD Dr. H. Pommer | Mathematics for chemists |
| Prof. Dr. H. Probst | Biochemistry and molecular biology |
| Prof. Dr. H.-G. Rammensee | Immunology |
| Prof. Dr. K. Reutter | Anatomy |
| Dr. W. Sarrazin | Radioisotopes |
| Prof. Dr. H. Schott | Organic chemistry |
| Prof. Dr. T. Stehle | Structural biology |
| Dr. G. Steinbrück | Zoology |
| Prof. Dr. S. Stevanovic | Immunology |
| Prof Dr. Dr. h.c. J. Strähle | Inorganic chemistry |
| PD Dr. S. Verleysdonk | Biochemistry |

| | |
|---|---|
| Prof. Dr. Dr. h. c. mult. W. Voelter | Physical biochemistry |
| Prof. Dr. G. J. Wagner | Physics |
| Prof. Dr. U. Weser | Inorganic chemistry |
| Prof. Dr. K.-P. Zeller | Spectroscopy of organic compounds |
| Prof. Dr. T. Ziegler | Organic chemistry |

## University of Massachusetts, Amherst

| | |
|---|---|
| Prof. Dr. J. P. Burand | Microbiology and virology |
| Prof. Dr. L. M. Gierasch | Biophysics |
| Prof. Dr. D. N. Hebert | Biochemistry and molecular biology |
| Prof. Dr. R. O. Karlstrom | Developmental neurobiology |
| Prof. Dr. L. C. Norkin | Virology |
| Prof. Dr. R. W. Phillis | Developmental neurobiology |
| Prof. Dr. S. J. Sandler | Microbiology |
| Prof. Dr. D. J. Schnell | Biochemistry and plant biochemistry |
| Dr. J. F. Swain | Protein purification and protein NMR |
| Prof. Dr. P. Wadsworth | Cell biology |
| Prof. Dr. R. M. Weis | Biophysical chemistry |
| Prof. Dr. R. A. Zimmermann | Molecular biology |