Structural and Functional Studies on the Early Steps of Polyomavirus and Adenovirus Life Cycles

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Dipl. Biochem. Antonio Manuel Liaci

aus Heilbronn

Tübingen

2017

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

Tag der mündlichen Qualifikation:	18.04.2017
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Thilo Stehle
2. Berichterstatter:	Prof. Dr. Niklas Arnberg

ABSTRACT

Polyoma- and adenoviruses are important human pathogens with worldwide significance. Both are non-enveloped DNA viruses, but differ significantly in terms of how they infect humans: Polyomaviruses (PyVs) cause persistent and occasionally fatal diseases involving multiple organs, while adenoviruses (HAdVs) are ubiquitous among societies worldwide and generally cause self-limiting diseases of mucous epithelial tissues that range from common cold symptoms to gastroenteritis and eye infections. The main focus of this dissertation is placed on the fundamental processes underlying the earliest steps of the viral life cycles of both viruses: cell attachment and entry. Most of the work is dedicated to the unraveling of the attachment and entry strategies of specific strains or serotypes, and a comparison of the factors that govern the differences between them.

The cell entry of polyomaviruses is mediated by the interaction of the major capsid protein VP1 with ubiquitous surface glycolipids called gangliosides. I have investigated three well-known strains of the murine polyomavirus (MuPyV) that show remarkable differences in terms of pathogenicity and tissue tropism, but only differ by singular amino acid exchanges located within their ganglioside binding cavities. This work establishes the ganglioside GT1a as a novel functional receptor and discovers minimal changes in the receptor binding affinities among the three strains that presumably lead to the drastically altered *in vivo* behavior.

Unlike PyVs, HAdVs generally employ a two-step mechanism mediated by distinct sets of capsid proteins to enter their target cells. The interaction of the C-terminal knob domain of the so-called fiber with a cellular primary attachment factor selects and tethers the viral particles to the host cell and facilitates cell entry mediated by the interaction of the viral penton base with a secondary entry receptor. These processes are assumed to contribute to the manifestation of viral tropism and host range. Here, the discovery and functional and structural characterization of novel primary attachment factors for two unrelated HAdV types is reported. The first type, HAdV-G52, is a rare and unique serotype that possesses two distinct fibers, which use completely different receptors: the long fiber recognizes the tight junction protein coxsackie- and adenovirus receptor (CAR), while the short fiber shows a specific preference for the glycan polysialic acid using a novel binding site on its fiber knob. The second type, HAdV-D36, is associated with obesity and has the unique ability to infect animals. This work demonstrates that HAdV-D36 uses a yet unidentified protein for attachment, and at the same time possesses a specificity for CAR and a sialic acid variant that is presumably only found in animals. Both projects have implications for the infectious routes of the two viruses.

An additional, third project presents the purification of the penton base protein of HAdV-D09 with the aim of structurally characterizing the interactions with its receptor counterpart, the integrin $\alpha_v\beta_3$, and unraveling the factors that dominate later phases of cell entry. The fourth part of this study addresses ways to interfere with adenoviral infections and to use adenoviruses as vectors for highly specific therapeutic applications. To this end, the development and evaluation of a series of second-generation inhibitors for the ocular pathogen HAdV-D37 is reported, whose design was inspired by its natural receptor, the sialic acid-containing glycan GD1a. Furthermore, this work sets the stage for the development of an HAdV-G52-based viral vector for the oncolytic treatment of somatic cancers expressing the tumor antigen polysialic acid. The last part of this dissertation describes ongoing work for the structural characterization of the adenoviral early gene product E4ORF1, a viral powerhouse protein that deregulates cell metabolism and polarity through various host interactions.

The findings presented in this dissertation have implications for our general understanding of how the differences among virus entry strategies emerge on a structural level, and provide valuable information for the development of efficient antiviral strategies and safer virus-based drugs.

ZUSAMMENFASSUNG

Polyoma- und Adenoviren sind klinisch relevante Krankheitserreger mit weltweiter Bedeutung. Obgleich es sich bei beiden um unbehüllte DNA-Viren handelt, unterscheiden sie sich deutlich im Hinblick darauf, wie sie Menschen infizieren: Polyomaviren (PyVs) verursachen typischerweise persistente und potenziell tödliche Krankheiten mehrerer Organe, während Adenoviren (HAdVs) weltweit verbreitet sind in der Regel selbstlimitierende Infektionen epithelialer Mukosagewebe verursachen, die von gewöhnlichen Erkältungssymptomen über Gastroenteritis bis hin zu Augeninfektionen reichen. Der Schwerpunkt dieser Arbeit liegt auf den Prozessen, die den frühesten Schritte beider viraler Lebenszyklen zugrunde liegen: der Anheftung an und das letzendliche Eindringen in die Wirtszelle. Der größte Teil der Arbeit befasst sich mit der Entdeckung neuer Anheftungsstrategien durch spezielle Virusstämme oder –serotypen sowie dem Vergleich der Faktoren, die für diese Unterschiede verantwortlich sind. Der Zelleintritt von Polyomaviren wird durch die Wechselwirkung ihres Haupt-Kapsidproteins VP1 mit ubiquitären Oberflächen-Glykolipiden, sog. Gangliosiden, vermittelt. Wir haben drei bekannte Stämme des murinen Polyomavirus (MuPyV) untersucht, die bemerkenswerte Unterschiede in ihrer Pathogenität und ihrem Gewebetropismus aufweisen, sich jedoch nur durch einzelne Mutationen in ihrer Rezeptor-Bindetasche unterscheiden. Diese Arbeit etabliert das Gangliosid GT1a als neuartigen funktionellen Rezeptor und zeigt minimale Veränderungen in den Bindungsaffinitäten zwischen den drei Stämmen auf, die vermutlich zu dem drastisch veränderten Verhalten *in vivo* führen. Im Gegensatz zu PyVs nutzen HAdVs im Allgemeinen einen zweistufigen Mechanismus, um in ihre Zielzellen einzudringen, wobei beide Stufen von unterschiedlichen Proteinen vermittelt werden. Die Wechselwirkung der C-terminalen Kopf-Domäne der sogenannten Virus-Faser mit einem primären Zielmolekül dient der Selektion und einem initialen Andocken an die Wirtszelle und ermöglicht so die Wechselwirkung des viralen Penton-Proteins mit einem sekundären Rezeptor, der den Eintritt in die Zielzelle herbeiführt. Diese Prozesse tragen zur Ausprägung des viralen Tropismus und des Wirtsspektrums bei. In dieser Arbeit wurden neue primäre Zielmoleküle für zwei unabhängige HAdV-Serotypen entdeckt und sowohl strukturell als auch funktionell charakterisiert. Das erste Virus, HAdV-G52, ist ein seltener und einzigartiger humaner Serotyp, der zwei unterschiedliche Fasern besitzt. Wir konnten nachweisen, dass beide Fasern völlig unterschiedliche Rezeptoren verwenden: die langen Fasern erkennen das Oberflächenprotein CAR, während die kurze Faser eine spezifische Präferenz für Poly-Sialinsäure besitzt, welche sie durch eine neuartige Bindungsstelle auf der Kopf-Domäne erkennt. Das zweite Virus, HAdV-D36, steht im Verdacht, Fettleibigkeit auszulösen und verfügt über die einzigartige Fähigkeit, Tiere zu infizieren. Diese Arbeit zeigt, dass HAdV D36 ein noch nicht identifiziertes Protein zur Anheftung an Zielzellen verwendet und zugleich eine Spezifität für CAR sowie eine spezielle Sialinsäure-Variante besitzt, die vermutlich nur in Tieren vorkommt. Beide Entdeckungen haben Auswirkungen auf unser Verständnis der Infektionswege beider Viren. In einem weiteren Projekt wird die Reinigung des Penton-Proteins von HAdV-D09 beschrieben - mit dem Ziel, seine Interaktionen mit dem Eintritts-Rezeptor Integrin $\alpha_v\beta_3$ stukturell zu charakterisieren und die Faktoren zu entschlüsseln, die während der späteren Phase des Zelleintritts wichtig sind. Ein weiterer Teil dieser Arbeit befasst sich mit Möglichkeiten, adenovirale Infektionen zu behandeln und Adenoviren als Vektoren für hochspezifische therapeutische Anwendungen zu verwenden. Hier wird die Entwicklung einer neuen Generation von Inhibitoren für HAdV-D37 berichtet, dem Erreger einer Augenkrankheit. Das Design dieser Inhibitoren wurde inspiriert von dem natürlichen Rezeptor des Virus, dem Sialinsäure-haltigen Glykan GD1a. Darüber hinaus legt diese Arbeit den Grundstein für die Entwicklung eines HAdV-G52-basierten onkolytischen Vektors für die gezielte Behandlung von somatischen Krebserkrankungen, die das Tumorantigen Poly-Sialinsäure exprimieren. Der letzte Teil dieser Dissertation beschreibt Vorarbeiten für die strukturelle Charakterisierung des adenoviralen frühen Genprodukts E4ORF1, ein Protein das den Stoffwechsel und die Polarität der Wirtszelle durch verschiedene Wechselwirkungen dereguliert.

Die Ergebnisse die in dieser Dissertation vorgestellt werden haben Auswirkungen auf unser allgemeines Verständnis davon wie spezifische Unterschiede zwischen den Eintrittsmechanismen des Virus entstehen, und liefert wertvolle Informationen für die Entwicklung effizienter antiviraler Strategien sowie sicherer Virus-basierter Medikamente.

INDEX

ABBREVIATIONS

A549 cells	An adenocarcinomic human alveolar basal epithelial cell line
AAV	Adeno-associated virus
AdV	Adenovirus
AIDS	Acquired immune deficiency syndrome
BAK	Bcl-2 homologous antagonist killer
BAX	Bcl-2-associated X protein
Bcl-2	B-cell lymphoma 2, prototype for the Bcl-2 protein family
CAR	Coxsackie- and adenovirus receptor
Cas9	CRISPR-associated 9
CASD1	Cellular apoptosis susceptibility (CAS)-domain containing 1
CAV-2	Canine adenovirus 2
CD	Cluster of differentiation (numbered)
CD spectroscopy	Circular dichroism spectroscopy
CD4+ T cell	T helper cell
CHO cells	Chinese hamster ovary cells
CRISPR	Clustered regularly interspaced short palindromic repeats
cryo-EM	Cryo-electron microscopy
DC	Dendritic cell
Dlg1	Discs large 1
DNA	Desoxyribonucleic acid
DOC	Deoxycholine
DP	Degree of polymerization
dsDNA	double-stranded DNA
DSG-2	Desmoglein-2
DSLNT	Disialyllacto-N-tetraose
ECM	Extracellular matrix
E. coli	Escherichia coli
E4ORF1	Open reding frame 1 of the adenovirus early gener region 4
E4ORF1p	Gene product of E4ORF1
EM	Electron microscopy
EndoF	phage K1F-derived endosialidase NF
ER	Endoplasmic reticulum
ERAD	ER-associated degradation machinery
ERp29	ER protein 29
FACS	Fluorescence activated cell sorting
FimH	Type 1 fimbrial adhesin
FK	Fiber knob, the C-terminal domain of the HAdV fiber protein
Gal	Galactose
GalNAc	N-acetyl galactosamine
GBM	Glioblastoma multiforme
Glc	Glucose

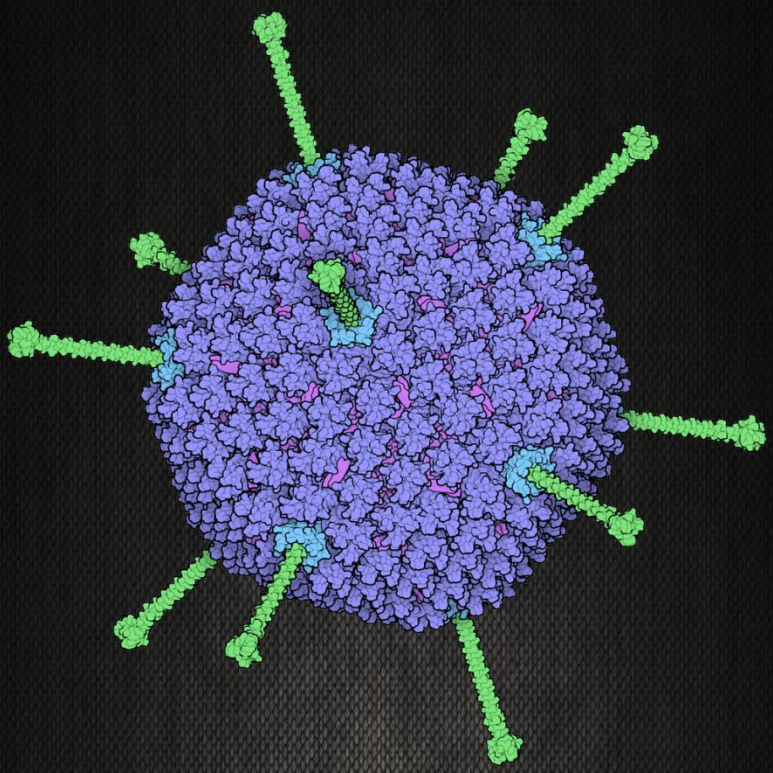
GlcNAc	N-acetyl glucosamine
HAdV	Human adenovirus
HCE cells	Human corneal epithelium cells
HE	Hemagglutinin-esterase
HIV	Human immunodeficiency virus
HPyV	Human polyomavirus
HSPG	Heparan sulfate proteoglycan
HSV	Herpes simplex virus
HSV-TK	HSV thymidylate kinase
ICTV	International Committee on Taxonomy of Viruses
IL	Interleukin (indexed)
ISAV	Infectious salmon anemia virus
kbp	kilobase pairs
kDa	kilodaltons
Kdn	2-Keto-3-deoxynonic acid
KIPyV	Karolinska Institutet Polyomavirus
LFK	Knob domain of the long fiber
LSTc	Sialylneolacto-N-tetraose c
MBP	Maltose-binding protein
MCC	Merkel cell carcinoma
MCPyV	Merkel cell polyomavirus
MD	Molecular dynamics
MEF	Mouse embryo fibroblast
MHC	Major histocompatibility complex
MHV-S	Mouse hepatitis virus strain S
MICA/B	MHC class I polypeptide-related sequence A/B
MIP-1α	Macrophage inflammatory protein 1α
mRNA	Messenger RNA
mTOR	Mechanistic Target of Rapamycin, mammalian target of rapamycin
MuPyV	Murine polyomavirus
MWPyV	Malawi polyomavirus
MYC	Myelocytomatosis oncogene, also named c-Myc
NCAM	Neuronal cell adhesion molecule (numbered)
N. meningitidis	Neisseria meningitidis
Neu	Neuraminic acid (indexed by substitution pattern, e.g. 5Ac, 4,5Ac_2, 5Gc)
NFκB	Nuclear factor 'kappa-light-chain-enhancer' of activated B-cells
NIH	National institutes of health
NJPyV	New Jersey polyomavirus
NK cells	Natural Killer cells
NLS	N-lauroylsarcosine
NPC	Nuclear pore complex
ORF	Open reding frame
p53	Transformation-related protein 53
PBM	PDZ binding motif
PDB	Protein Data Bank

PDB ID	PDB identifier
PDZ	a common structural domain; name is an acronym of PSD95, Dlg1, ZO-1
PI3K	Phosphatidylinositol-3-kinase
PML	Progressive multifocal leukoencephalopathy
PPH	Peripentonal hexon
PSia	Polysialic acid
PyV	Polyomavirus
rb	Retinoblastoma
RGD	An arginine-glycine-aspartic acid motif occurring in HAdV penton base
RNA	Ribonucleic acid
SAD	Single wavelength anomalous diffraction
SAdV	Simian adenovirus
SARS	Severe acute respiratory syndrome
SAXS	Small-angle X-Ray scattering
SCID	Severe combined immunodeficiency
SCLS	Small cell lung cancer
SFK	Knob domain of the short fiber
Sf-9	An ovary cell line from *S. frugiperda*, for protein expression
S. frugiperda	*Spodoptera frugiperda*
SPR	Surface plasmon resonance
STD-NMR	Saturation transfer difference nuclear magnetic resonance
STLPyV	St. Louis polyomavirus
SUMO	Small ubiquitin-like modifier
SV40	Simian virus 40
TLR9	Toll-like receptor 9
UV	Ultraviolet light
VCAM	Vascular cell adhesion molecule
VLA-4	Very late antigen-4; an alternative name for integrin $\alpha_4\beta_1$
VLP	Virus-like particle
VP1	Polyomavirus major capsid protein 1
VP2,3	Polyomavirus major capsid proteins 2 and 3
WHO	World Health Organization

Common abbreviations are used for biological, chemical, and physical units and values, amino acids, and bases occurring in DNA or RNA. Gangliosides are abbreviated according to the Svennerholm abbreviation system [1]. If only the carbohydrate portion is referred to, the Svennerholm abbreviation is used with the addition ‘glycan’. Other abbreviations are explained in the text.

CHAPTER

1



1. INTRODUCTION

1.1. History of Virology and Virus Classification

Over millions of years, viruses have established a complex and interwoven relationship with their hosts. Some viruses have co-evolved with early predecessors of mammals, while many other human viruses emerged from contact between humans and animals. Viral infections have accompanied all large waves of human migration and adapted to the existence of a worldwide human society [2]. Today, about 8% of the human genome is of retroviral origin, and some retroviral gene products carry important physiological functions [3,4]. Despite their indisputable importance for mankind, viruses have been recognized as infectious agents for only little more than a century. In the 4th century BC, Hippocrates explained the sudden spread of diseases with a so-called *miasma* (ancient Greek for 'pollution' or 'bad air'). Even in the 19th century, this term was still in use [5]. However, the first sophisticated antiviral treatments date back to the 11th century: at the time, a procedure called *variolation* was used to treat smallpox in China and India [6]. In this process, healthy individuals were treated with infectious material from pox pustules of infected patients. Despite the obvious risks of the treatment, variolation was perceived as a large improvement at the time [2]. Consequently, the practice was adopted in Europe and America in the 18th century and eventually received a significant refinement when physician Edward Jenner realized that the safety of the treatment could be much improved when human samples were replaced with material from cowpox lesions [7,8]. About 85 years later, Louis Pasteur coined the term vaccination (Lat. '*vacca*' = cow) to honor Jenner's achievements [9]. It was Pasteur himself who was the first to design a purposely attenuated vaccine when he passaged cow rabies in rabbits in order to create vaccination material that was then less infectious to cows. Following Jenner and Pasteur, a series of further safety measures was developed, bare all knowledge about the origins of the diseases. Groundbreaking insights into the agents behind the phenomena were only possible after the invention of the light microscope, which opened the eyes of mankind to the world of microorganisms. By the 19th century, their existence was largely accepted, and Pasteur and Robert Koch conducted key experiments to establish the rules by which these new organisms spread and replicate. Koch was the first to ascribe specific microorganisms to specific processes (reviewed in [10]). At the time, these findings were mostly attributed to bacteria and other organisms, as viruses still escaped the human eye. In 1892, however, Dmitri Ivanovsky reported that the agent causing tobacco mosaic disease could not be retained by established methods to filter out bacteria, and six years later Martinus Beijerinck ascribed this phenomenon to an even smaller agent than bacteria, one so small that it could still not be visualized. Friedrich Löffler and Paul Frosch realized that the non-filterable causative agent of foot-and-mouth disease replicated only within the host organism. To sum up these features, Beijerinck eventually named the agent causing tobacco mosaic disease *contagium vivum fluidum*. Eventually, the term '*virus*' (Lat. poison) was used to describe all agents of this sort [2].

The first human virus to be discovered was yellow fever virus in 1901. Following this discovery, many important new human viruses were detected, among them the infectious agents causing smallpox, ebola, AIDS, rabies, hepatitis, dengue fever, and influenza. Today, viruses are defined as obligate intracellular infectious particles whose DNA or RNA genome is replicated within the host cell. Virus particles, called virions, are formed by *de novo* assembly within the host cell and serve to transmit the genomic material between hosts [2]. Usually, viruses lyse their host cells upon completion of their life cycle, although some of them establish a persistent infection. Since they cannot replicate by themselves, viruses are not considered living creatures. In August 2016, the International Committee on Taxonomy of Viruses (ICTV) listed a total of 3704 virus species belonging to 610 genera and 111 families, and it seems probable that a large portion of viruses are still not discovered [11]. The genomic material of viruses can be either single- or double-stranded DNA or RNA. The viral genome can measure 3 to 1200 kbp and codes for a number of functional and structural components that serve to initiate and exert the viral life cycle and to transport the virus progeny from one host cell to another, respectively. The classification of animal viruses is largely based on the identity of the genome and the morphology of the virus particle. The most commonly used

features are the identity of nucleic acids, the capsid symmetry and size, the presence or absence of an envelope, and the genome type and architecture. In an attempt to emphasize the importance of the host's translational machinery, David Baltimore proposed a classification of viruses according to the way that viral mRNA is produced from the original genetic material [12]. Although the evolutionary relationships among viruses are still largely unknown, the classification system serves to highlight similarities among viruses that often result in similar approaches to a specific step of the life cycle.

The viruses representing the main topic of this thesis, polyoma- (PyVs) and adenoviruses (AdVs), both possess a dsDNA genome and non-enveloped icosahedral capsids. Both viruses are grouped into Baltimore class I. Although they differ significantly in some aspects of their life cycles, several similarities exist that will be elaborated on in the following introductory chapters.

1.2. Polyomaviruses

The small, non-enveloped PyVs possess a circular genome of about 5 kbp and a simple genome organization. In addition to the three major capsid proteins VP1-3, the viruses encode several non-structural proteins, including different splice variants of the so-called T-antigen that act together to orchestrate the viral replication cycle [13]. The family *Polyomaviridae* coevolved with animals for at least half a billion years and currently contains about 100 members infecting fish, spiders, scorpions, birds, and mammals [14]. The three genera avi-, ortho-, and wukipolyomaviruses are classified based on host range, genome sequence, and the genetic content [15]. In 1971, the first two human PyVs (HPyVs) were simultaneously discovered and named JCPyV and BKPyV after the initials of the two patients they were isolated from [16,17]. For about four decades, these two viruses were thought to be the only PyV species to infect humans, the only other PyV of clinical importance being the so-called simian virus 40 (SV40) that had been found in a contaminated polio vaccine batch in the 1960's [18,19]. In the course of the last decade, however, a total of eleven new PyV species were discovered and named either after the places they were found at, the order of their discovery, or the diseases they are associated with. Namely, these viruses include the KI and WU polyomaviruses (KIPyV, for Karolinska Institutet [20]; and WUPyV for Washington University [21]), Merkel cell polyomavirus (MCPyV) [22], HPyV 6 and 7 [23], *trichodysplasia spinulosa*-associated polyomavirus (TSPyV) [24], HPyV9 [25], HPyV10 (also called MWPyV after the country of its discovery, Malawi) [26], STLPyV (St. Louis) [27], HPyV12 [28], and finally NJPyV (New Jersey) [29]. The name 'polyoma' (derived from Greek '*many tumors*') was originally selected in order to account for the tumorigenic activity of the first polyomavirus isolates that caused a number of different neoplasms [30,31]. However, these effects are generally not seen in healthy, immunocompetent individuals, and in fact only a minority of PyVs is known to be tumorigenic. So far, causal disease associations have been established for only four HPyV species. BK- and JCPyV Infections frequently occur during early childhood, and the majority of cases are subclinical. Depending on the region, the seroprevalence is about 50-80%, and there is a frequent establishment of asymptomatic latent infections of the kidney [32]. However, both viruses are known to cause life-threatening diseases in immunocompromised hosts. BKPyV is known to cause severe nephropathy after kidney transplantations which potentially leads to allograft loss [33], and JCPyV is the causative agent of a rare, but very dangerous demyelinating disease called progressive multifocal leukoencephalopathy (PML) [34,35]. TSPyV was named after its association with *trichodysplasia spinulosa*, a very rare skin disease associated with immunosuppression after organ transplantations [36]. MCPyV is the causative agent of Merkel cell carcinoma (MCC), an uncommon but highly aggressive form of skin cancer that is usually fatal within a few months and is frequently associated with the integration of viral DNA into the host genome [37]. In contrast, little is known about the pathological profiles and infectious routes of the newly discovered HPyV species (reviewed in [38,39]). To date, MCPyV is the only HPyV species linked to cancer, and its aggressiveness is perhaps only paralleled by specific strains of the murine polyomavirus (MuPyV) in newborn mice and hamsters. MuPyV, one of the viruses of interest to this study, was the first PyV to be discovered. The first MuPyV isolates were reported in 1958 [30,31], five years after virologist Ludwik Gross identified a cancer-transmitting factor in leukemic

extracts which to his surprise caused bilateral neck tumors when inoculated into healthy mice [40]. These severe effects seemed to occur mostly in newborn mice and hamsters, while immunocompetent adult animals experienced a much lighter course of disease. In the following decades, a range of different MuPyV strains have been isolated that differ greatly in the frequency and severity of tumors they cause [41] (**Figure 1**). In particular, three strains have been the topic of extensive research. A laboratory-derived MuPyV strain called RA causes only singular tumors of exclusively mesenchymal origin [42]. Another strain called PTA naturally occurs in feral mice [43]. PTA typically displays a broader tissue tropism than RA and causes different tumors of mesenchymal as well as epithelial origin [44,45]. Lastly, another laboratory-derived strain called LID produces a highly aggressive phenotype that causes a wide range of tumors, and infected mice usually succumb to kidney failure or brain hemorrhages within weeks [46-48].

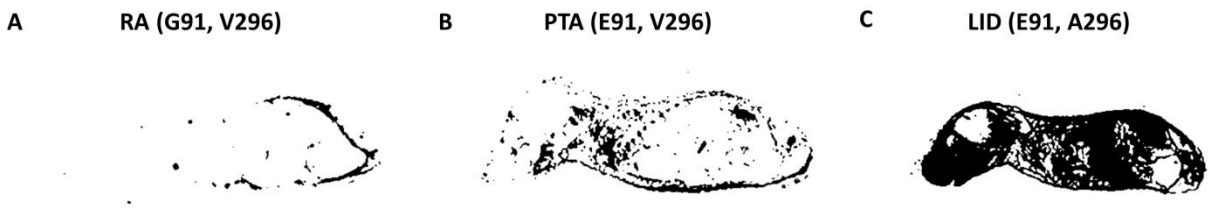


Figure 1 | Phenotypes caused by the three MuPyV strains RA, PTA, and LID. *Whole mouse-sections of newborn mice were obtained 7 days post infection and stained with an anti-VP1 antibody.* ***A*** *RA caused singular tumors of mesenchymal origin.* ***B*** *PTA caused various tumors of mesenchymal and epithelial origin.* ***C*** *LID caused multiple tumors of various origins. Figure modified from [49]. © 1999, American Society for Microbiology.*

Surprisingly, the genetic differences between these three very different MuPyV strains are limited to only nine mutations within the whole genome. While five of these single-base substitutions are silent, two have been mapped to the ORF of the T antigen and have been shown to produce no phenotypic differences. Interestingly, both laboratory-derived strains differ from PTA by only one residue, and both of these remaining two mutations occur in the receptor binding pocket of the major capsid protein VP1 [50]. The PTA strain possesses a glutamate residue at VP1 position 91, as opposed to the RA strain that has substituted the glutamic acid with a glycine residue. The deadly LID strain, in turn, exhibits an even more subtle change, as it replaced valine 296 with an alanine [41,49-53]. These small differences have been unambiguously and causally linked to the differing phenotypes, as the respective phenotype can be transferred by simple site-directed mutagenesis of either position [52]. Similar observations have been independently obtained for two other MuPyV strains that possess analogous residue switches [50].

1.3. Adenoviruses

In 1953, the year of Gross' discovery of polyomaviruses, two similar unknown viruses were isolated from human patients with acute respiratory infections [54]. These new viruses were named adenoviruses, owing to the fact that they were both isolated from adenoid tissue. AdVs are among the largest non-enveloped DNA viruses. The common ancestral lineage of modern adenoviruses likely dates back to before the divergence of bony fish from the vertebrate branch [55]. Today, the family of *Adenoviridae* contains four genera, members of which infect every major branch of vertebrates. Most mammalian AdVs, including all human AdVs (HAdVs), belong to the genus *Mastadenovirus* [11]. To date, more than 70 HAdV types have been identified and grouped into a total of seven species (A-G) [56-58] according to their genome sequence and a number of functional criteria: serology, hemagglutination behavior, as well as their oncogenicity in rodents and capability to transform primary human

cells. According to a new naming convention, newer HAdVs are now named 'types' instead of 'serotypes' and characterized based on geno- rather than serotyping [54,59-62].

HAdVs are important pathogens that cause a wide variety of diseases, generally of mucous epithelial tissues. Depending on the specific type and its tropism, these diseases include diverse infections of the eye as well as the respiratory, gastroenteric, and urinary tracts. Remarkably, even members of the same species can produce largely different pathological outcomes. Infections usually occur *via* respiratory or oral-fecal routes, without the involvement of a vector host [63,64]. While HAdV infections are usually self-limiting in immunocompetent individuals, immunosuppressed patients or children can develop chronic and eventually fatal disease courses that also spread to other tissues and cause severe cases of pneumonia, hepatitis, myocarditis, or encephalitis (reviewed in [65]). HAdVs are typically host-specific, although they are capable of infecting a wide range of mammalian species when administered experimentally [66]. The main reason for the reduced permissiveness in animals is most likely a poor support of viral replication [67,68]. Many HAdV serotypes are highly contagious, and a dose of only five viral particles can already be enough to cause a disease [63]. Since the virus particles are also exceptionally stable and can persist on inanimate surfaces for several weeks [69-71], it is not surprising that HAdV infections are ubiquitous all around the world and occur throughout the year, with peak infection rates during winter and spring [4]. Following the 1962 discovery that HAdV-A12 caused malignant tumors in newborn hamsters, HAdVs were found to transform human cell lines and considered tumor virus candidates [72,73]. In fact, HAdVs infect both dividing and somatic cells and have the ability to drive resting cells into the S phase of the cell cycle. However, no transformative activity has been observed in humans so far [54].

Table 1 | Key features of selected HAdV types. *O = ocular, R = respiratory, GI = gastrointestinal, A = adipose, U = urinary. Attachment factors that were discovered in this thesis in italics. Sources: [74-78].*

HAdV type	Tropism	Associated diseases in otherwise healthy humans	Seroprevalence	Attachment Factor
C05	O, R	• Acute respiratory illnesses • Pharyngocunjunctival fever • Conjunctivitis	60-98%	CAR
D36	GI, A	• Gastroenteritis • Obesity	6-77%	*$Neu4,5Ac_2$, CAR Unknown Protein*
D37	O, U	• Epidemic keratoconjunctivitis	~8%	GD1a glycan
G52	GI	• Gastroenteritis	<1%	*CAR / PSA*

The HAdV types discussed in this study are types C05, D36, D37, and G52. An overview of their most important features can be found in Table 1. HAdV-C05 is perhaps the most well-studied of all AdVs, and most of our knowledge about the HAdV life cycle is derived from experiments with this type [79]. The virus causes several diseases of the respiratory tract and is highly abundant in the human population, with seropositivity rates residing between 60% and 98% depending on the region [74]. HAdV-D37, on the other hand, is found in cervical swabs and is one of the major causative agents of epidemic keratoconjunctivitis (EKC), along with HAdV types D08, D53, D54, D56, and D19 (now retyped as D64 to distinguish it from the non-EKC causing prototype virus HAdV-D19p) [58,80]. Some HAdV types of species B and E also display an ocular tropism, but cause other conjunctivital diseases [81-83]. The biology of HAdV-D36 is less well understood, despite a seroprevalence of up to about 77% depending on region and age of the individuals. The virus is associated with gastroenteritis, since it has been initially isolated from the feces of an enteritis patient [84,85]. As a unique feature, it has been associated with animal and human obesity in a large body of epidemiological and experimental studies, which paradoxically seems to coincide with lowered serum cholesterol and triglyceride levels (reviewed in [75,86]). Although HAdV-C05 and -D37 also increase

adiposity in inoculated animals, HAdV-D36 is the only one of these types to be associated with obesity in humans, especially children [87]. Moreover, the virus is readily spread from infected to healthy chickens through natural routes, and naturally occurring antibody responses have been detected in domesticated rhesus monkeys as well as chickens and rats [85,88-90]. The ability to infect such a wide range of vertebrates is highly unusual for HAdVs, especially given the fact that HAdV-D36 seems to cause obesity in all of them, despite substantial differences in energy metabolism [90]. HAdV-G52 is the only HAdV member of *Mastadenovirus* species G, which otherwise contains only simian AdVs (SAdVs) [91]. HAdV-G52 has so far been isolated from only a handful of individuals during a single outbreak of gastroenteritis in a senior residence in California in 2007, and one stool sample originating from an unrelated case of gastroenteritis in the US. Since then, no signs of the virus have been found in humans, suggesting a very low seroprevalence [77]. Being only distantly related to other HAdVs, HAdV-G52 exhibits several unique features, among them a reduced binding of coagulation factors by its hexons and the presence of two sequentially and functionally distinct fiber molecules (see **Chapter 3.2.1**).

1.4. Structural Features of Non-Enveloped Viruses

Viruses are nature's dedicated minimalists. Each of their components has been streamlined through constant evolutionary pressure, and every one of their gene products fulfills one or more crucial tasks. Perhaps the most impressive example for this sculpting process is the structure of viral capsids. Their main task is to guarantee efficient transmission of the viral genome from the original cell to a different target cell, either within the infected host or to a different host. During the transport, the genetic material needs to be shielded from a generally hostile environment, be it from the host's immune system or - in between hosts - to extreme conditions such as chemical or physical distress, pH changes, UV radiation, or drought stress [92]. To protect the cargo against these risks, viral particles have to be stable and rigid. At the same time, the particles need to uphold a certain degree of flexibility during their assembly, and to finally disassemble readily when triggered by a specific stimulus after cell entry. Hence, the virus particle needs to be maintained in a resilient, but metastable, non-energy-minimal state [93]. In non-enveloped viruses, which lack the additional protection of a lipid envelope, this can only be achieved through remarkable protein stability, dense packing, and a minimalistic design that evades as many host immunity factors as possible. Evolutionary pressure and genetic economy usually dictate that virus capsids are built from a limited number of building blocks, and indeed, most viruses possess only a handful of major capsid proteins [92]. From an architectural point of view, this brings about an inherent symmetry that all protein capsids are subject to, and a remarkable flexibility in the way interfaces between chemically identical subunits can be arranged. As another consequence, common principles apply for all virus capsid structures, as different as they may seem in the first place. All viruses contain at least one protein coat, and with only few exceptions, all of them possess either a helical or an icosahedral symmetry, and even more complicated capsids often contain elements of these basic symmetries. For example, HIV, which possesses one of the most complex viral particles to infect humans, has a complex 'fullerene cone' capsid that shares certain features with icosahedra [94-96]. Helical capsids are mostly found in non-human RNA viruses. These capsids can easily vary in length and only require one type of capsid protein. Yet, they are open at the ends, they do not have an optimal surface-to-volume ratio, which limits their genome packing capability, and they have limited functional flexibility. For these reasons, helical animal viruses usually have an envelope that carries out additional protective functions.

Non-enveloped viruses, on the other hand, usually impose icosahedral symmetry on their capsids. Icosahedra are closed three-dimensional structures with 12 pentameric vertices that are connected by 20 triangular faces. Since these faces do not have to be flat, even round virus capsids follow the icosahedral symmetry, allowing for an optimal packing ratio. In 1962, Caspar and Klug derived a set of rules for icosahedral viruses that - for the main part - still holds up today [97]. As such, three-dimensional closure of an icosahedron requires twelve fivefold vertices, and icosahedral particles have to be made up of at least 60 repetitive structural units. Further, icosahedra possess inherent two-, three-, and fivefold symmetry elements (**Figure 2A**), and thus the otherwise identical

structural units have to be combined by similar, but not identical contacts. This principle of 'quasi-equivalence' is circumvented when there is more than one type of capsid protein. To amend the Caspar-Klug theory, particles of this sort are generally referred to as 'pseudo-icosahedral'. If an icosahedral capsid possesses more than 60 subunits, the additional subunits have to form triangular 'tiles' that fill the space between the vertex positions. In order to achieve dense packing, these building blocks need to form hexagonal structures. The fivefold and sixfold assemblies are called capsomers. An elegant concept to describe the complexity of an icosahedral particle is the so-called 'triangulation number' T, which essentially describes the ratio between the face area and the area of one capsomer. In practice, T is determined by how many capsomers it takes to connect vertex positions following only the two axes h and k (**Figure 2B**).

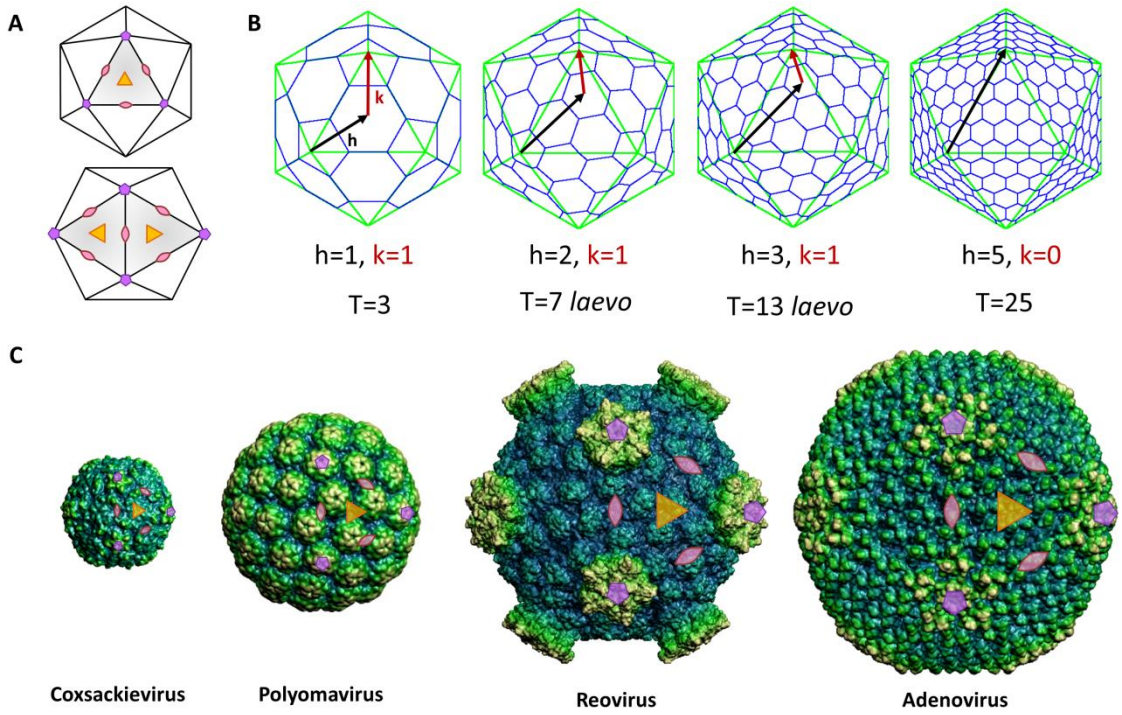


Figure 2 | Icosahedral capsid symmetry. ***A*** *Symmetry elements of icosahedra. Two- three- and fivefold axes are colored accordingly. The two representations are equivalent ways to illustrate icosahedral virus capsids.* ***B*** *Derivatization of triangulation numbers. For T=7 and 13, one of two relative orientations is shown, the other one being dextro.* ***C*** *Capsid structures of icosahedral viruses relevant to this study: coxsackie- (T=pseudo-3), polyoma- (T=7 dextro), reo- (T=13 laevo), and AdVs (T=pseudo-25). Relative sizes are approximated, and icosahedral symmetry operators are depicted schematically. Panel B was generated using the Icosahedral Server of the VIPERdb. Capsid figures of panel C were adapted from VIPERdb [98] .*

Since the triangular faces have to be equilateral, the triangulation number is then calculated by an altered binomial equation:

$$T = h^2 + hk + k^2 \qquad \text{(Equation 1)}$$

The underlying symmetry elements of select non-enveloped viruses relevant to this study are highlighted in **Figure 2C**.

One common structural feature that can be found in many different virus families is the usage of β-barrel folds in their capsid proteins (**Figure 3**). Possible explanations for this preference might be a common evolutionary history or convergent evolution owing to the fact that β-barrels possess excellent stability and a good shape complementarity in both hexameric and pentameric arrangements [92,99-101].

1.4.1. The Polyomavirus Capsid

PyVs possess a T=7 capsid whose outer shell has a diameter of about 45 nm and is made up exclusively from 72 pentameric copies of the major capsid protein VP1 [102]. The VP1 protein is constituted by a central eight-stranded jelly roll β-sandwich, whose antiparallel β-sheets named BIDG and CHEF are connected by loops of differing lengths that make up the capsid exterior and form the receptor recognition sites of the virus (**Figure 3A**). The pentamers possess extensive monomer interfaces, which give the protein excellent stability and force the monomers into to a radial orientation, resulting in distinctive canyons on the capsid surface (**Figure 2C**). Inter-pentamer contacts are mediated through a 'strand-swapping' β-sheet augmentation of the N- and C-terminal parts between pentamers. The VP1 capsomers occupy both penta- and hexameric positions, which results in a remarkable bonding pattern with three different types of threefold or twofold clustering [103]. This pattern differs remarkably from the otherwise very similar papillomavirus capsid structure, in which the 'strand swapping' is replaced by an internal loop reaching over to neighboring capsomers in a 'hanging bridge' arrangement which is then locked by disulfide bridges [104]. Additional stability of the PyV capsid is conferred by disulfide bridges between the CD loops and N-termini of neighboring VP1 proteins as well as Ca^{2+}-mediated bridging interactions. Structural studies of VP1 can be facilitated by using a C- and N-terminally truncated version of VP1 that is unable to form virus-like particles (VLPs) and remains pentameric in solution [105].

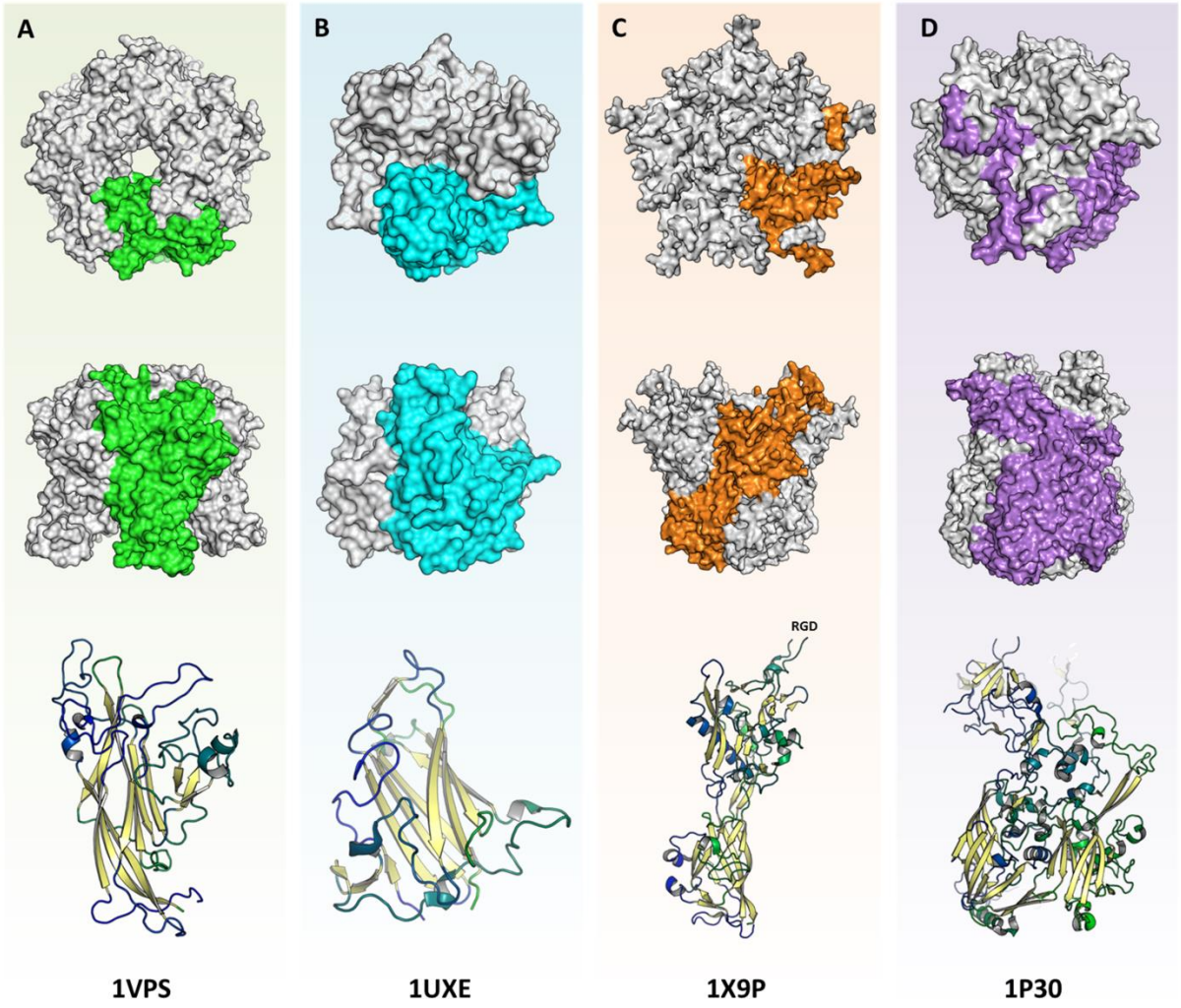


Figure 3 | Major capsid proteins of polyoma- and adenoviruses. *Upper panel: top view of the functional multimers. Middle panel: side view of the functional multimers. Bottom panel: detailed side view of the respective monomers. All proteins form β-barrels.* ***A*** *PyV VP1* ***B*** *AdV fiber knob* ***C*** *AdV penton base. The location of the RGD motif is indicated.* ***D*** *AdV hexon. The PDB IDs of the individual structures are listed.*

The two additional PyV capsid proteins, VP2 and VP3, lie on the inside of the capsid, with their C-termini projecting into the VP1 fivefold axis [106,107]. While VP2 is important for capsid stability and virus entry, the function of VP3, which is essentially a truncated version of VP2, is currently unknown.

1.4.2. The Adenovirus Capsid

AdVs possess a much more complicated capsid structure than PyVs. AdV capsids range between 70 and 90 nm in diameter and exhibit a beautifully visible icosahedral shape (**Figure 2C**). The capsid is made up of a total of 252 capsomers [108-112]. There are two types of capsomers, for both of which structural information is available (**Figure 3B-D**). The vertices of the icosahedron are occupied by the pentameric penton base, harboring the prominent elongated fiber trimer that projects radially from the capsid. Together, penton base and fiber form the penton capsomer. The remaining 240 capsomers are formed by an entirely different protein, the hexon, which makes up the bulk of the capsid. Due to this arrangement, the AdV capsid symmetry is commonly described as pseudo T=25. Although the hexons occupy the hexameric positions, they are actually homotrimers that possess two β-sheets per protein chain (**Figure 3D**) [113-115]. There are 12 hexon trimers per tile (**Figure 2B**), occupying four distinct positions that engage in different inter-hexon contacts. The most distinct position is the one surrounding the penton bases, and hexons of this type are called peripentonal hexons (PPH). The penton base is a homopentameric β-barrel that harbors two hypervariable loops at exposed positions: one loop of approx. 40-100 aa which in almost all types contains an arginine-glycine-aspartic acid (RGD) and is thus called the RGD loop (**Figure 3C**), and a shorter loop simply called 'variable loop' (see **Appendix 6.2.3** for a full sequence alignment) [109,116,117]. The penton base structure has been determined both alone and in a capsid context, and the two structures differ remarkably in their overall morphology. The most prominent difference is the relative twist of the monomers and the resulting cavity along the fivefold axis. The flexibility implied by this discrepancy is coherent with the peculiar functions of the penton base in virus entry and disassembly (see **Chapter 4.1**).

In addition to the three major capsid proteins, AdVs possess seven minor structural proteins. Four of them are called cement proteins and are tightly associated with the capsid either from the inside (proteins VI and VIII) or from the outside (IX and presumably IIIa). Protein IX forms flexible triskelions that glue together the inner parts of the hexon tile. Although there is still some disagreement about function and location of protein IIIa, based on latest structural data it most likely stabilizes inter-tile and PPH interactions from outside the capsid [118]. Proteins V, VI, and VIII form a ternary complex that is closely attached to the PPH ring from the inside of the capsid. Apart from stabilizing functions, protein VI is the key membrane lysis factor during endosomal escape (see **Chapter 1.8.3**). One probable reason for the presence of the cement proteins is that they may help to stabilize the mature capsid while keeping the particle flexible during assembly. In this light, it is not surprising that many of the minor capsid proteins undergo a proteolytic maturation mediated by the AdV protease (reviewed in [119]). The cleavage of DNA-associated core proteins inside the particle detaches the genome from the protein shell, and the increased electrostatic repulsion of the DNA is building up pressure inside the capsid. The result is an increased particle stiffness that is believed to facilitate virus uncoating [120-122]. Remarkably, there are no stabilizing interactions with the penton base, leaving the vertices the weakest part of the capsid that is most sensitive to the increase in capsid pressure. Another intriguing part of the AdV capsid architecture is the match of the trimeric fiber to its anchor point, the pentameric penton base. The fiber consists of a shaft region made up from triple-β-spiral repeats, forming a super-helix which ends in a rigid C-terminal knob domain [123]. There are three different types of fiber proteins among HAdVs (see **Appendix 6.2.2**) [124,125]. Members of species B and D have short, sturdy fibers with a length of about 320-380 amino acids, corresponding to 5-6 triple-β-spiral repeats [126]. In contrast, fibers of species A and C HAdVs have much longer, flexible fibers that contain more than 20 triple-β-spiral repeats and feature a flexible hinge region [127]. The only member of species E, HAdV-E04, possesses a fiber of intermediate length (426 amino acids). The length of the fiber is likely to be one of the determinants of receptor usage (see **Chapter 1.8.1**). The three HAdV types belonging to species F (40/41) and G (52) possess two different fibers simultaneously [91,128], and in the course of this thesis it was shown that HAdV-G52 uses both fibers for the

recognition of two distinct receptors (see **Chapter 3.2.1**). Currently, there are two models for how fiber and penton base interact (**Figure 4**). According to a penton base crystal structure of HAdV-C-02 in complex with an N-terminal peptide of the HAdV-C-02 fiber as well as electron density reconstructions from a full-capsid cryo-electron microscopy map, the fiber does not fit into the cavity formed at the fivefold interface of the penton base (28 Å in diameter) [117,129]. Instead, the conserved and extended N-terminus of the fiber is bound at the outward-facing side of the penton base, at the interface between two monomers. This finding is in agreement with numerous electron microscopy images that show the fibers projecting far from the viral capsid [108]. In contrast to this, the X-Ray structure of the HAdV-C05 capsid pseudotyped with a short HAdV-D35 fiber showed a highly dilated penton cavity (56 Å in diameter), with helical density inside that likely belongs to the fiber shaft [109]. In this arrangement, the fiber could possibly be retracted into the penton base. Depending on the length of the fiber, only the knob domain might be accessible and sitting right on the capsid surface. Both conformations possibly represent physiological states that might occur during different steps of the viral life cycle.

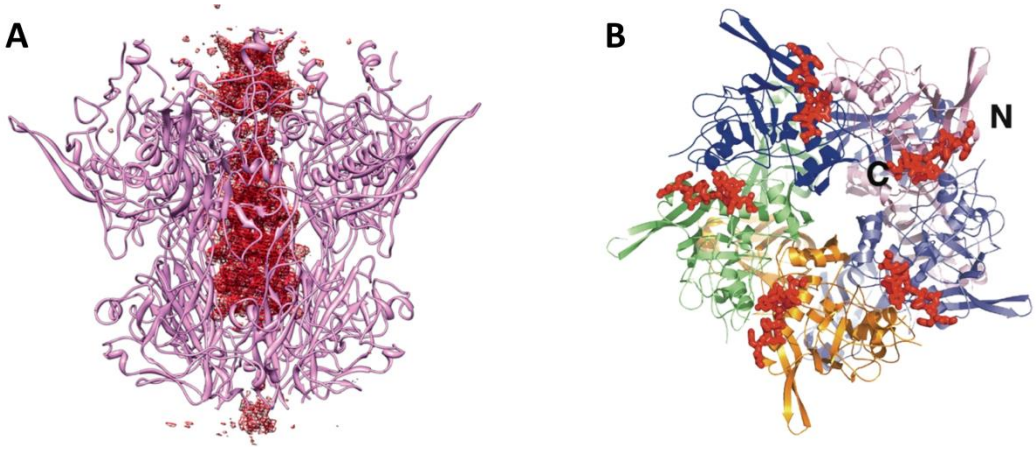


Figure 4 | Models for the penton base – fiber interaction. ***A*** *Fiber electron density as visible in the HAdV-C05 crystal structure pseudotyped with HAdV-B35. Pink cartoon: penton base. Figure adapted from [109]. © 2012, Elsevier B. V.* ***B*** *Interaction model based on the HAdV-C02 cryo-EM structure, analogous to a HADV-C02 penton base crystal structure in complex with an N-terminal fiber peptide. Cartoon: penton base; red sticks: fiber. Figure adapted from [117]. © 2005, Elsevier B. V.*

The fiber knob is a sturdy trimer of globular, 8-stranded β-barrel domains formed by the C-terminus of the fiber protein (**Figure 3B**) [130]. The β-barrels are held together almost exclusively by hydrophobic interactions, and there is an extensive trimer interface featuring hydrophobic and polar contacts alike. The result of this arrangement is high rigidity as well as remarkable overall stability and sturdiness at temperatures beyond 80°C. Similarly to PyV VP1, the AdV fiber knob is the main point of primary attachment receptor recognition. The solvent-accessible surface, including the receptor interfaces, is mainly made up by the loops connecting the β-strands. While the overall knob architecture is conserved among HAdVs, the differences in attachment receptor usage manifest themselves in variable lengths of the main recognition loops (mostly the AB, CD, DG, GH, and IJ loops), and a different relative angle of the monomers within the knob. A full sequence alignment of HAdV types C01-D70 can be found in **Appendix 6.2.1**. Interestingly, both PyVs and AdVs form minimal T=1 structures during virus assembly [116,131-133]. While there is no known function of PyV T=1 particles, in the case of HAdV-B03 so-called 'penton dodecahedra' made up exclusively of fiber-equipped penton units have been shown to bind desmoglein-2 (DSG-2) and destroy tight junctions at the lateral site of epithelial tissues, the main target tissues of these viruses (see also **Chapter 1.8.1**) [134]. Hence, at least for AdVs, these apparently faulty assemblies seem to have an important function during infection and cell egress, although they do not carry any genomic material.

1.5. Viral Entry Strategies

Until the late 1970's, the early stages of the viral life cycle were believed to be subject to purely passive processes. However, extensive work starting in the 1980's made it abundantly clear that all animal viruses specifically attach to receptor molecules on the host surface, which then allow the virus to usurp existing cellular uptake processes. The series of events unfolding when a virus encounters a suitable host cell can be divided into the processes of attachment to the cell surface, passing of the plasma membrane, intracellular trafficking, and ultimately membrane penetration and release of the cargo into the cytosol. The later steps are often, but not always, combined with full or partial uncoating of the virus particle [135]. The attachment process is dominated by the specificity and availability of the receptor molecules that are used as catch-hold hooks to tether the viral particle on the host cell surface. Viruses use a wide range of surface molecules for this purpose, including proteins and glycans. While the presence of a receptor is necessary to render the target cell susceptible to the virus, the permissiveness of the host (i.e. its ability to support virus replication) often depends on additional intracellular factors [135]. As an example, the fact that the highly similar HAdV types D37 and D19p have different pathogenicity profiles despite using the same receptor remains a mystery [136]. Although the receptor usage of a specific virus does not always correlate linearly with tissue tropism and pathogenicity profiles, receptor identity and availability are important prerequisites for the process of infection. In some cases, such as the case of MuPyV, receptor usage can be directly linked to pathogenicity profiles [49].

The way that viruses exploit host receptors can be quite different from virus to virus. Some viruses use only one type of receptor that simultaneously fulfills all functions necessary for the different stages of cell entry, while others use distinct receptors for each step or several receptors with equivalent roles. The choice of several receptors facilitates the timing of the uptake and allows the use of receptors that are not suited to directly trigger virus uptake. Additionally, some viruses such as AdVs recognize highly abundant receptors [137,138], while others such as HPyVs specialize on a remarkably small set of receptors with a restricted expression profile in order to evade the host immune system more efficiently [139]. The processes that govern the attachment step are largely probability-driven, and this affects the way viruses choose their receptor molecules. As such, non-enveloped viruses have specialized on exploiting avidity effects using the symmetry of their capsids, and as a result the local crowding and availability of a receptor can be more important than the strength of the interaction itself [140-142]. Some viruses use receptors that are also present on non-permissive cells or positioned in a way that does not allow for cell entry (decoy- or pseudoreceptors), and in this case the number of viral particles released is of major importance for the efficiency of infection [143]. Moreover, there are examples of viruses that target several attachment receptors at the same time, but use them for different purposes at different stages of infections. As such, one of the most prominent attachment receptors of HAdVs, CAR, is most likely not present on the site of initial encounter. Instead, CAR is located as a homodimer in tight junctions and on the basolateral side of epithelial cells, and one of the functions of this receptor targeting is likely the disruption of epithelial integrity to ensure efficient cell egress or the uptake of progeny virus after mounting an infection through epithelium-associated macrophages (see **Chapter 1.8.1**) [144]. Some coxsackieviruses, in turn, initially bind to CD55 on the apical face of the host cell, only to be transported to tight junctions where they eventually enter through interactions with CAR [145]. Regardless of their target tissue, most viruses need to cross epithelial barriers at some point during infection, and so they often interact with epithelial cells without infecting them. These sorts of receptor functions are fundamentally different, but difficult to distinguish experimentally, and the exposure of a receptor in a tissue context can vary from the results obtained in cell-line experiments. The delicate equilibrium established between a virus and its host cell range can be radically altered when the virus is administered through non-natural routes of infection, especially *via* injection into the blood as might be the case in vector-based therapeutic applications such as gene therapy. In this case, the behavior of the virus is essentially unforeseeable. Therefore, these applications usually include a safety switch that either renders the viral particle non- or only conditionally infective. Some viruses such as influenza viruses are known to infect multiple species, and their host range can depend on the usage of species-specific receptors [146]. Similarly, some human pathogens have evolved to use non-human targets as a part of their receptor portfolio [147]. Some viruses, among them noro- and reoviruses, are even known

to specifically attach to commensal gut bacteria in order to hijack a bacteria-induced breach of the epithelial barrier [148,149].

After the initial attachment, the virus has to reach past the plasma membrane and physical host barriers such as the actin cortex in order to reach the cytosol. Viruses evolved to use different cellular pathways to accomplish this task. Enveloped viruses have the possibility to fuse directly with the plasma or an endosomal membrane through the action of a fusion protein and a factor that disassembles the actin cortex. Non-enveloped viruses, on the other hand, have to be taken up through endosomal vesicles and then induce membrane disruption and vesicular escape upon a specific trigger stimulus found in the new environment [135]. This stimulus might be a specific receptor interaction or environmental cues such as the action of endosomal proteases (reoviruses [150] and SARS coronavirus [151]), a drop in pH, a specific membrane lipid content (Dengue virus [152]), or a combination of these factors. These stimuli usually trigger specific conformational changes that weaken the capsid structure. A range of endocytic pathways are possible: phago- or macropinocytosis, clathrin-independent or -dependent endocytosis, ending up in the endosomal pathway, or caveolar, cholesterol-dependent, or dynamin-2-dependent endocytosis [93]. Caveolae are a specific type of lipid rafts associated with caveolin and cavin. It was previously thought that these pathways end up in the so-called caveosome, which later has been demonstrated to be a late endosome with increased caveolin-1 levels [153]. Which of these pathways is chosen is usually determined by the size of the viral particle and the type of surface receptor. Some viruses remain on the cell surface, passively attached to their receptors, while others such as PyVs actively induce endocytosis [154]. In other cases such as Influenza A virus, it is the receptor interaction that triggers endocytosis [155]. Viral entry of some viruses, including AdVs, is initiated through receptor-mediated signaling processes at the plasma membrane, often involving secondary receptors that are recruited and clustered after the initial attachment has taken place [156]. These processes are in some cases accompanied by a lateral movement on the cell surface [157]. Once the endosomal vesicles are disrupted, the genetic material is released into the cytoplasm and transported to its destination by cellular factors, initiating the viral replication cycle.

1.6. Surface Carbohydrate Receptors

Glycans are among the basic building blocks of life and are found on the cell surface of all living organisms. In contrast to proteins and nucleic acids, they are not synthesized from a template, but through the interplay of complex synthetic machineries in the ER and Golgi apparatus. Different types of glycans are generally made by distincts sets of glycosyltransferases [158,159]. Although mammalian glycans are made up from combinations of as little as ten monosaccharides, they are among the most versatile structures of life. Glycans are ubiquitously found on cell surfaces, where they can be attached to proteins, glycolipids, or form long structures in the extracellular matrix (ECM) (**Figure 5A**). Therefore, it is not surprising that carbohydrates execute manifold functions in health and disease. For one, they exert structural and modulatory functions, e.g. as physical barriers or moisteners (glycocalix, mucins) or keepers of growth factor gradients (glycosaminoglycans). In addition, glycans are indispensable for many intrinsic recognition functions, e.g. as self-flags (histo-blood group antigens), immune modulators (e.g. Siglec-7, a negative regulator of NK cells, or selectins), and even as wound healing factors (dermatan/heparan sulfate). Yet, they are exploited by a vast pool of pathogens, either as receptors or by means of molecular mimicry (**Figure 5B**). The interplay of these opposing 'intrinsic' and 'extrinsic' recognition functions likely orchestrates the evolutionary change in an organism's glycan repertoire. As a consequence of the functional diversity and the flexibility of the synthetic machinery, the distribution of many glycans can vary quite significantly and is responsive to tissue type, developmental stages, or even environmental influences or metabolic states [160]. In mammals, most glycans appearing on the cell surface are capped by 9-carbon pyranose moieties called sialic acids. While 'sialic acid' is a common term used for different variants of neuraminic and 2-keto-3-deoxynononic acid, it is also used to describe the most common variant, 5-acetylneuraminic acid (Neu5Ac). With few exceptions, sialic acids are only found in deuterostomes, which makes them important 'self'-markers in these

animals and accounts for many of their vital functions. Some human pathogens, among them *E. coli* K1 and group B meningococci, have gained the ability to synthesize sialic acid through convergent evolution (reviewed in [161,162]) and use them as a disguise from the host's immune system (molecular mimicry). Apart from their exposed positioning on glycans, sialic acids possess several distinctive features compared to other common monosaccharides, namely a carboxyl and a glycerol function, and in most cases also a 5-N-acetyl function. All these functional groups point to the same face of the pyranose ring (**Figure 5C**). The large number of functional groups and the simultaneous display of charged, polar, and hydrophobic parts enables sialic acids to participate in a large number of hydrogen bonds, salt bridges, and non-polar interactions.

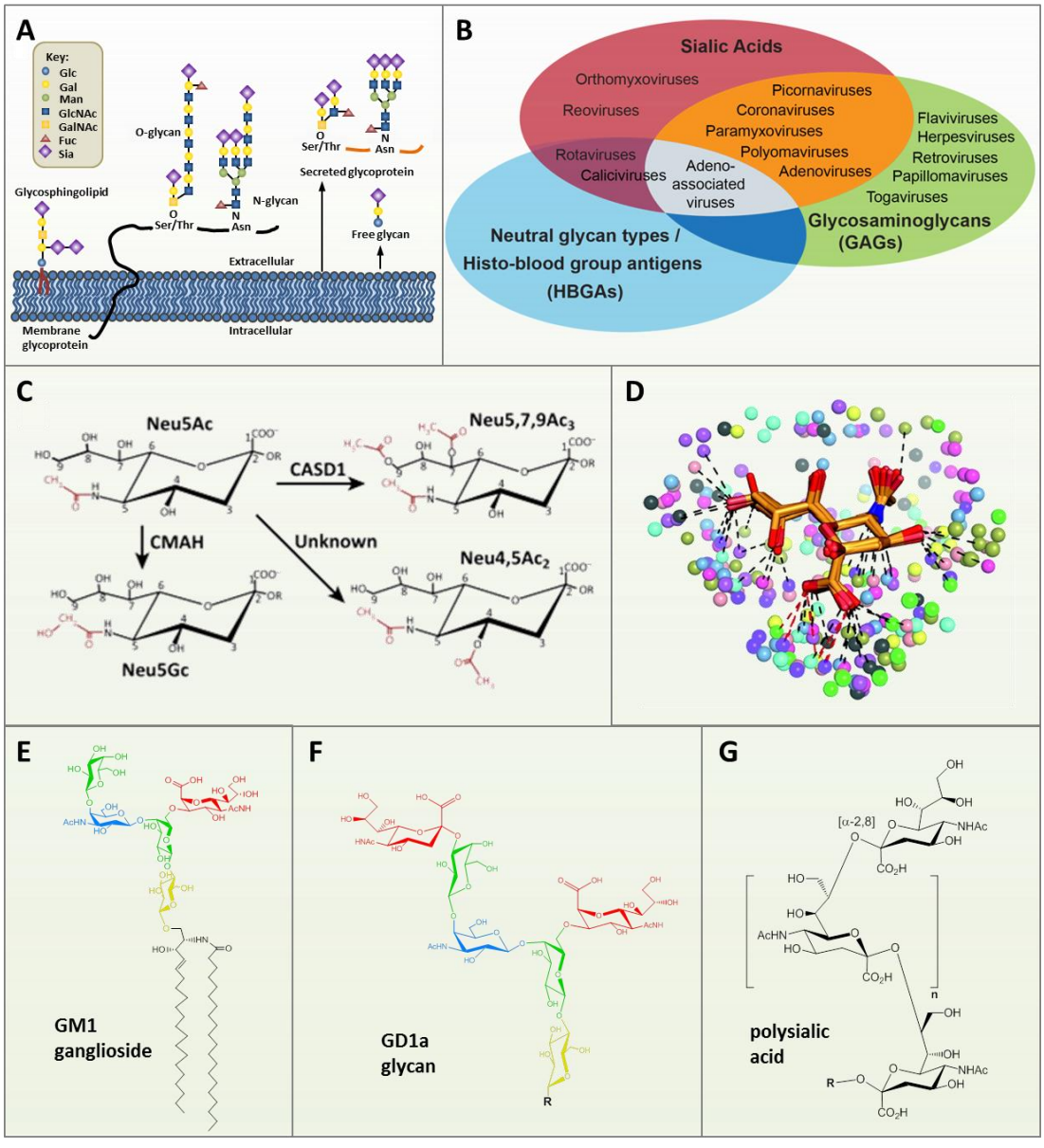


Figure 5 | Glycans as virus entry receptors. ***A*** *The occurrence of sialylated glycans on cell surfaces. Coloring according to the legend.* ***B*** *Overview of the exploitation of glycans by different virus families.* ***C*** *Functional groups of Neu5Ac and enzymes conferring O-acetylation at various positions.* ***D*** *Binding patterns of Neu5Ac by selected viruses including MuPyV, JCPyV, SV40, HAdV-D37, and canine adenovirus 2 (CAV-2). Polar contacts are exclusively mediated by the sugar's functional groups.* ***E*** *Structure of the GM1 ganglioside. Coloring scheme according to [163].* ***F*** *Structure of the GD1a glycan.* ***G*** *Core structure of polysialic acid (PSia). Elongated PSia chains can reach up to ~100 moieties. Panel A based on and panel C taken from [164].© 2016, Elsevier B. V. Panel B modified from [165]. © 2014, Annual Reviews. Panel D modified from [139]. ©2011, Elsevier B. V. All figures were taken with permission.*

As terminal parts of glycans, sialic acids are usually connected to the rest of the glycan by either an α-2,3 or an α-2,6 linkage. If they occupy internal positions, it is mostly because one or more additional sialic acids are attached to their O8 or O9 position *via* an α-2,8 linkage. This distinctive overall profile makes sialic acids prone to be recognized by pathogens. Indeed, they are the major anchor point for an impressive number of viruses, among them influenza-, rota-, polyoma-, reo-, and adenoviruses (reviewed in [139,165]). Although receptor interfaces with sialic acids are inevitably small and their affinities are rather weak compared to protein interfaces, the abundance of sialylated glycans on the host cell surface can compensate for the lack in affinity. Sialic acids are ubiquitous in all tissues and occur in high local concentrations as dense clusters on glycoproteins or in membrane microdomains. Paired with the fact that viruses usually possess a number of identical binding sites on their capsids, this abundance accounts for avidity effects that drastically increase the affinity, often by several orders of magnitude [141,142]. For this reason, small changes in glycan recognition can have severe influences on the behavior of a virus. Some viruses have developed receptor-cleaving enzymes to ensure effective cell egress and spread after the lytic cycle is complete. Other viruses, among them PyVs and AdVs, need to maintain a complex equilibrium between efficient attachment and cell egress. As a direct consequence, these viruses usually have very shallow binding sites with affinities in the millimolar range that form contacts with only one side of the pyranose ring (**Figure 5D**) [139]. These interactions most commonly feature key polar interactions with the sugar's carboxylate and 5-acetyl amide group. Because they are short-ranged and directional in nature, these contacts demand a high shape complementarity between the sugar and its receptor. While directed polar interactions are important for orienting the sugar in the pocket, about 50% of the interaction energy is accounted for by van-der-Waals interactions with the acetyl and glycerol functions of sialic acid. Studies with the bacterial lectin FimH showed that other effects are also important for glycan recognition, among them entropic and electrostatic effects [139,166]. These effects are usually long-ranged and less directed in nature, which makes them harder to assess, and the driving forces behind them are often not well understood.

To battle infections, deuterostomes have evolved over 50 sialic acid variants differing in their substitution patterns, commonly referred to as the sialome. Derivatives of the 4 'core' sialic acid variants Neu5Ac, 5-Glycolylneuraminic acid (Neu5Gc), 2-Keto-3-deoxynonic acid (Kdn), and neuraminic acid (Neu) can be additionally substituted by *O*-acetyl, *O*-methyl, *O*-sulfate, *O*-lactyl, or *O*-phosphate groups. Due to high educt promiscuity in the synthetic machinery, animals can often incorporate sialic acid variants that they cannot synthesize themselves [167]. As an example, humans have lost the ability to synthesize Neu5Gc, but are able to incorporate the glycan through dietary uptake and enzymatic transfer. To date, information about the tissue and species distribution of most sialic acid variants is still sparse, mostly due to the lack of sophisticated and unambiguous *in situ* detection methods (reviewed in [168]). Currently, profiling is done with combinations of more or less variant-specific viral lectins such as the hemagglutinin-esterases (HEs) of infectious salmon anemia virus (ISAV) or the coronavirus mouse hepatitis virus strain S (MHV-S) [169,170]. In addition, the synthetic enzymes for many variants are often not identified or have been identified only very recently (**Figure 5C**) [171,172]. Some animal viruses, among them corona- and influenzaviruses, have developed binding sites that distinctively bind specific sialic acid variants, usually by specific contacts with the additional functional groups [173-175]. For example, the most frequent sialic modification, acetylation at the O9 position, is of critical importance for Influenza virus C infection [176]. In humans, this modification is conferred by the gene product of CASD1 [171]. Acetylation at the O4 position is conferred by a different, still unidentified enzyme that probably acts on already assembled glycans. 4-O-acetylated Neu5Ac (Neu4,5Ac_2) is considered to be absent from human tissues, but plays an inhibitory role for some human Influenza A strains in horse, and is exploited by several animal viruses such as MHV-S or ISAV (reviewed in [164]).

Some carbohydrate-containing molecules occurring on the cell surface have affinities for membrane microdomains, caveolae, or lipid rafts, causing the local crowding at cell entry points discussed above [177-181]. One example of molecules with raft affinities are so-called gangliosides (**Figure 5E**), ubiquitous sphingolipids that are prominently expressed in the nervous system [182]. Gangliosides consist of glycan structures that are attached to ceramide anchors [183]. Depending on the organization of the glycan core, a number of different series exist, and most gangliosides are capped by one to three sialic acid moieties that are transferred by one of 5 human sialyl-

transferase isoforms ([184], a comprehensive overview figure can be found in [163]). Gangliosides are exploited as the main attachment and entry receptors for PyVs, as has been shown in knock-out mice lacking one or more sialyl-transferases [185]. Quite commonly, glycans are permanently attached to proteins *via N*- or *O*-glycosylation (**Figure 5A,F,G**) [186]. These glycosylations have broad implications on protein stability or interactions. Protein glycosylation is performed by a different set of glycosyltransferases than that used for ganglioside synthesis [158,159]. Nonetheless, some glycan epitopes can appear as patterns on both proteins and gangliosides. One example is the glycan portion of the ganglioside GD1a, which is also commonly found on *O*-linked glycoproteins [76]. This coexistence of recognition patterns on several molecule types may in some cases explain the seemingly contradictory behavior of viruses, since one or more of them may act as 'decoy' receptors that only allow for unproductive attachment of virus particles. Some glycans attached to proteins have more specialized functions. As an example, polysialic acid (PSia), a highly charged linear homosialic acid polymer of up to 90 moieties **(Figure 5G),** is almost exclusively attached to the neural cell adhesion molecule (NCAM) in the developing brain, and in the adult it is mainly found in specific brain locations that need constant reorganization [187-190]. Its primary role is in the maintenance of neuronal plasticity, probably by weakening cellular adhesion through the increase of the hydrodynamic radius of the underlying protein through water absorption, and through repulsive effects due to its highly negative charge. However, several tumors exploit this function and upregulate PSia. PSia-expressing tumors can be found even outside the brain, e.g. in small cell lung cancer (SCLC), and PSia expression is associated with metastatic tumor activity [191-199]. Several pathogenic bacteria also display PSia on their surface to protect themselves from the host immune system [190].

1.7. MuPyV Cell Attachment and Entry

The cell entry of MuPyV has been the subject of decades of investigation, and the fact that MuPyV engages sialic-acid containing glycans as receptors has been established in the 1980s. All MuPyV strains are able to recognize sialic acid attached to a galactose moiety *via* an [α-2,3]-linkage. In contrast, only some strains possess the ability to detect glycans with an additional branching [α-2,6]-linked sialic acid [200,201]. In 1997, Stehle and Harrison were the first to structurally characterize the receptor interactions between RA VP1 and the branched glycan DSLNT [105]. In 2003, Tsai *et al.* showed that the naturally occurring receptors of MuPyV are in fact the two gangliosides GD1a and GT1b [185]. In the course of this thesis, the receptor repertoire was expanded by a similar ganglioside called GT1a (see **Chapter 3.1**). In contrast, glycoproteins that harbor the same glycan structures as the receptor gangliosides do not act as infectious receptors. Instead, they seem to act as decoy receptors that promote a non-productive internalization followed by an immune response of the host [143]. The usage of gangliosides as attachment and entry receptors is quite common among PyVs. Examples include BKPyV and SV40, the latter of which displays a clear specificity for the GM1 ganglioside [185]. In contrast, JCPyV was shown to use the glycoprotein-borne glycan motif LSTc along with the serotonin receptor for cell entry [202-204]. The glycan binding site located on the VP1 protein is formed by the BC1, BC2, DE, and HI loops reviewed in [139]). Despite the limited number of contacts, the receptor site is remarkably conserved across species (**Figure 6**). To date, the only known exception is TSPyV, whose glycan binding site is located at a different location and is likely the result of convergent evolution [205]. The high conservation of the binding details appears to reflect the importance of receptor binding for the viral phenotype, as shown by the drastic changes caused by the subtle alterations of MuPyV VP1 (**Chapter 1.2**). A second illustrative example is the behavior of several JCPyV variants that show differences in their LSTc binding pocket [202]. These mutations appear to abrogate the binding of both the natural receptor and neutralizing antibodies, and instead seem to promote an infection of the brain through the action of another, yet unknown receptor and may be a key factor for the development of PML [206-209].

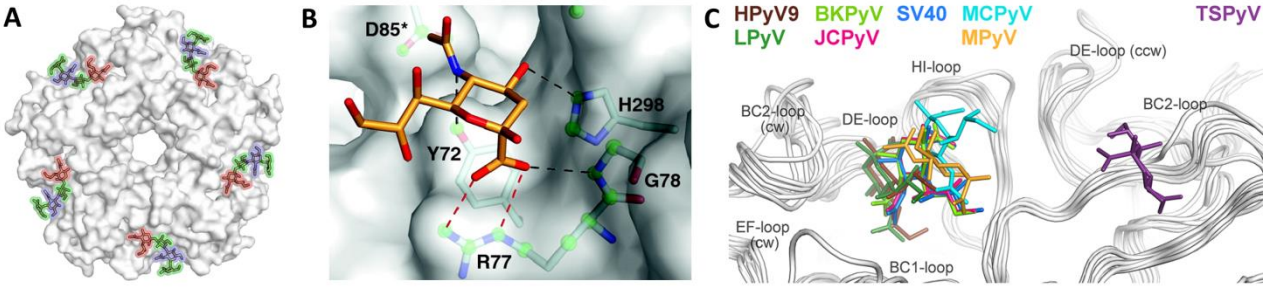


Figure 6 |Receptor interactions between polyomaviruses and sialic acid. ***A*** *Location of the glycan binding site on MuPyV VP1. Carbohydrates are colored according to Figure 5. PDB-ID: 5CQ0.* ***B*** *Polar interactions of MuPyV VP1 with sialic acid. Modified from [139]. © 2011, Elsevier B.V.* ***C*** *Conservation of the sialic acid binding pocket among different PyVs. MuPyV is denoted as 'MPyV' in the figure. Taken from [205]. © 2015, Ströh et al.*

The later entry steps of MuPyV are surprisingly unique and versatile in comparison to other PyV species, and can occur either through caveolae or clathrin-independent smooth monopinocytotic vesicles [210-212] (**Figure 7**). Although the exact mechanism of internalization is unclear, both mechanisms appear to depend on the local crowding of the receptor gangliosides in lipid raft microdomains, which presumably serve to directly induce membrane curvature in a manner similar to SV40 [154]. The latter accumulates its receptor GM1 using its large number of binding sites, which directly results in lipid phase separation and actively induces membrane tubulation. Similarly, MuPyV VLPs induce tubular invaginations in GD1a-containing vesicles. Additionally, MuPyV has been reported to engage the $\alpha_4\beta_1$ integrin (VLA-4) in this process [213]. The key factors of vesicular behavior, such as the importance of pH and VP2 myristoylation, are still a topic of debate [214-216]. After cell entry, PyVs use an atypical transport route. The viral particles are directly transported to the endoplasmic reticulum (ER), presumably together with their receptor gangliosides [210,217]. Within the ER, the action of the disulfide isomerase ERp29 alters the VP1 structure and exposes hydrophobic loops that facilitate attachment to the ER membrane [218]. From there, viral particles escape to the cytosol through an unknown mechanism involving parts of the ER-associated degradation machinery (ERAD) [219]. Finally, the virus is targeted to the nucleus by a nuclear localization signal located at the N-terminus of VP1, and virus replication is initiated [220,221]. Later stages of the viral life cycle will not be discussed in this thesis.

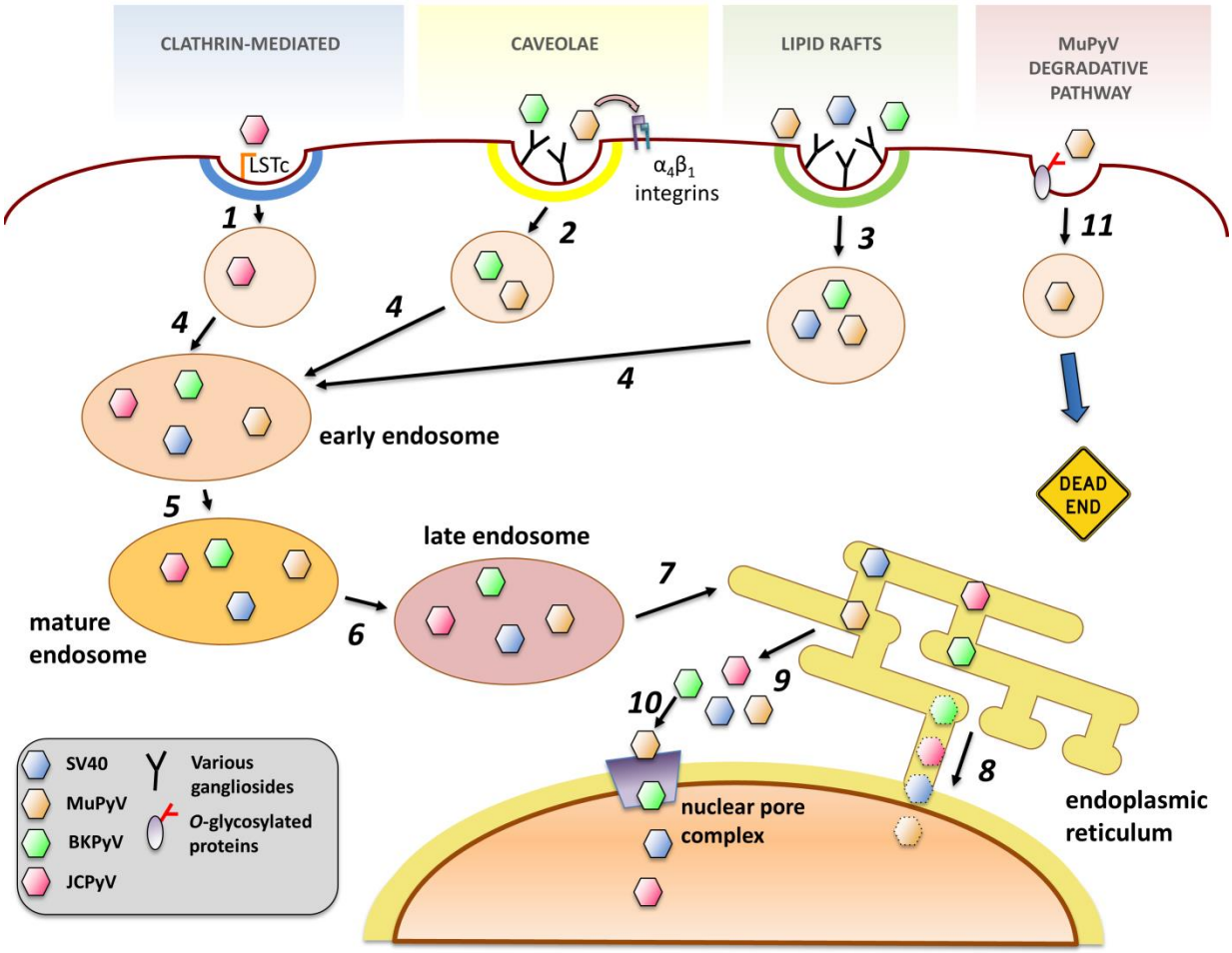


Figure 7 | Polyomavirus cell entry. *PyV particles are colored according to the legend. Regardless of the endocytosis mechanisms (1-3), PyVs end up in endosomes (4) and are eventually transported to the endoplasmic reticulum (ER) (5-7). Uncoating may occur inside the ER, followed by direct transport into the nucleus (8), or at the nuclear pore complex after ERAD-assisted escape into the cytosol (9-10). Figure adapted and modified from [222], with permission. © 2016, University of Tübingen. Original figure based on [223].*

1.8. Adenovirus Life Cycle

1.8.1. Entry Receptors

Generally, HAdVs employ a sequential two-step entry mechanism to enter cells. First, a primary attachment step is mediated by interactions between the fiber knob domain and a cellular primary attachment factor. This primary attachment is at least partially responsible for the viral tropism by selecting possible target cells [57,137]. In addition, the initial attachment serves to tether the virus particle to the target cell, and enables the local crowding of secondary receptors that interact with the viral penton base in order to initiate the entry step. While the secondary receptor is an integrin, the primary attachment proteins differ even among HAdV types of the same species (reviewed in [224,225]; **Figure 8**).

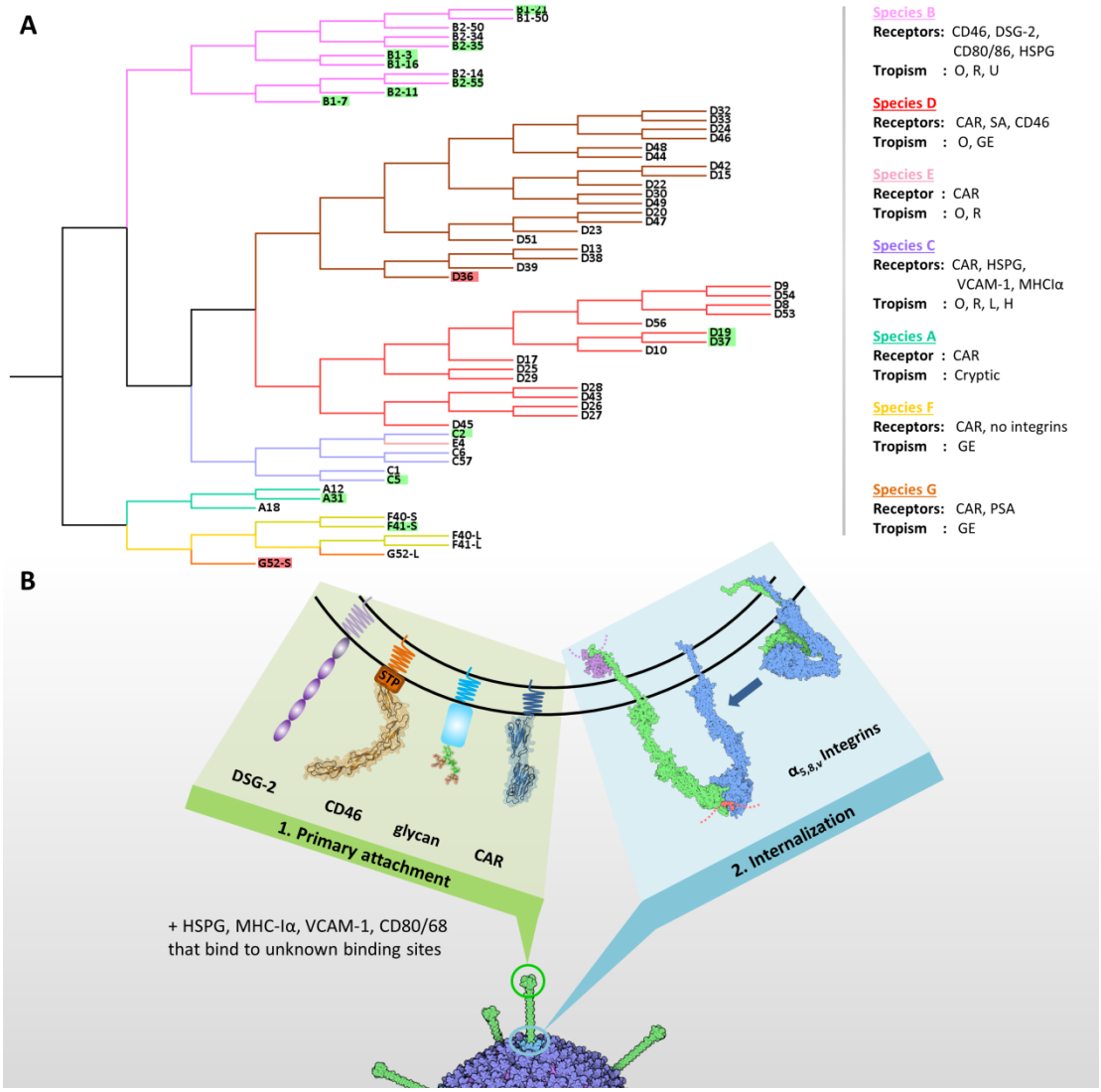


Figure 8 | Analysis of HAdV fiber knob sequences and entry receptors. ***A*** *phylogenetic alignment based on the fiber knob sequences of HAdV types C01-G52. The branches are colored according to the respective species. Two major clusters of species D are colored in red and brown, respectively. Types with published knob structures are highlighted in green, types whose structures were newly solved in this thesis are highlighted in red. O = ocular; R = respiratory; U = urinary; GE = gastrointestinal; L = lymphoid; H = hepatic.* ***B*** *Summary of known HAdV entry receptors. Extracellular portions of receptors are depicted as crystal structures if available, while intracellular domains were left out for clarity. PDB IDs: 3O8E (CD46); 3N0I (GD1α glycan); 3JZ7 (CAR). Parts of the figure were based on [224] and the RCSB PDB-101 'Molecule of the Month' issues of 12/2010 and 02/2011 [226]. STP=Serine, threonine, and proline rich region.*

Initially, CAR was believed to be the only attachment receptor of HAdVs [227,228]. During the last decades, however, an increasing amount of other HAdV primary receptors has been identified, demonstrating that the earliest step of the life cycle is already much more sophisticated than anticipated and contributes to the surprisingly variable tropism and pathogenicity of different HAdV types [137]. Members of each HAdV species except species B have been demonstrated to use CAR as a receptor, thereby disrupting the intercellular homotypic CAR-CAR interface. The intracellular signaling domain of CAR is not needed for efficient infection. However, this domain appears to elicit an immune response upon virus entry [229]. Upon complex formation, the sides of the AdV fiber knob form a large interface with the C-terminal domains of up to three CAR molecules per knob (**Figure 9A**).

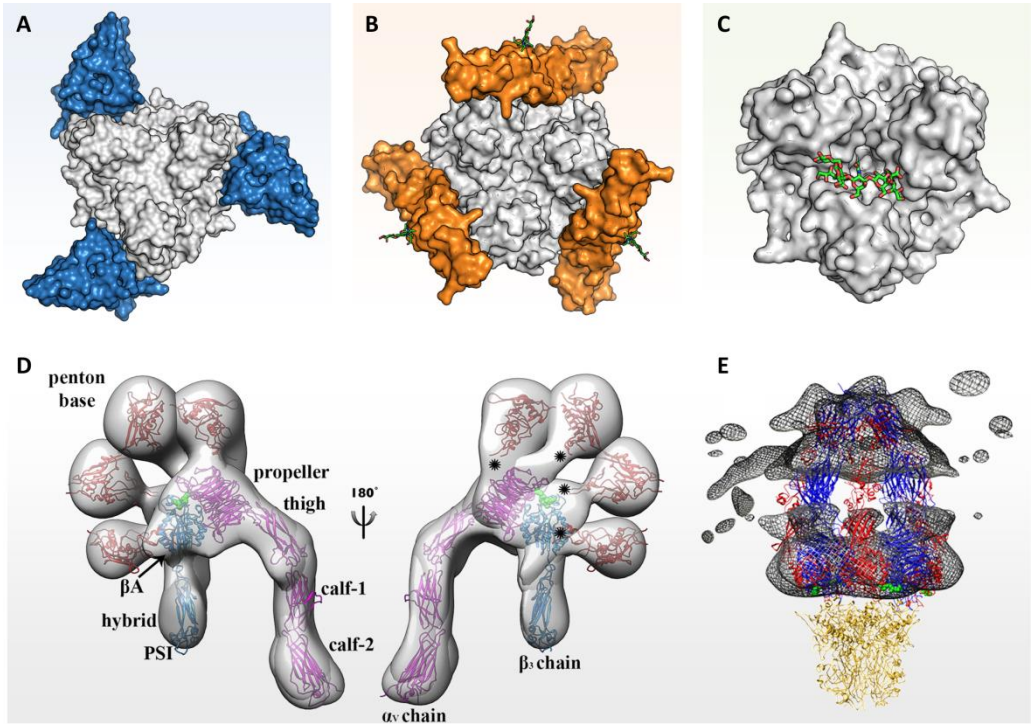


Figure 9 | Interactions of HAdV capsid proteins and their entry receptors. ***A*** *Interactions between the HAdV-A12 fiber knob (gray) and CAR (blue) occur through a large interface at the sides of the knob (PDB-ID 1KAC).* ***B*** *The fiber knob of HAdV-B11 engages the two terminal domains of CD46 at a large, water-filled interface that is formed by two fiber monomers (PDB-ID 2O39).* ***C*** *HAdV-D37 engages its receptor, the GD1a glycan, using a binding pocket near the trimer interface and the two terminal sialic acids of the glycan. The trimeric symmetry of the knob makes it possible to accommodate up to three sialic acid moieties at a time (PDB-ID 3N0I).* ***D*** *Possible interaction modes between the $\alpha_v\beta_3$ integrin (pink and blue) and the distal domain of the HAdV-D09 penton base (red) identified by cryo-electron microscopy (cryo-EM). Panel modified from [230]. © 2014, National Academy of Sciences.* ***E*** *Cryo-EM model of the interaction between HAdV-A12 penton base (yellow) and $\alpha_v\beta_5$ integrin (red/blue and electron density) based on a complex with the entire viral capsid. Icosahedral symmetry was applied for structure solution. Panel modified from [231]. © 2009, American Society for Microbiology.*

The resulting avidity effect produces a high binding affinity ranging from 1 nM (HAdV-C02) to 20 nM (HAdV-D37) [232,233]. In the course of a natural infection, however, the access to this very potent receptor is usually limited, since CAR is mainly expressed at the inaccessible basolateral side of the epithelial target tissue [234]. Several mechanisms have been proposed to explain how HAdVs may circumvent this receptor shortage. For one, the virus might first infiltrate non-polarized epithelial immune cells and then start a lateral attack on its target cells, or require pre-existing epithelial lesions to do so [235]. Upon infection, the virus possibly induces the expression of a CAR splice variant with an apical expression profile [236]. Lastly, members of HAdV species B have replaced CAR with entirely different receptors. While approximately half of the species B HAdVs evolved to use the ubiquitously expressed inhibitory complement receptor CD46 [138,237,238] (**Figure 9B**), the other half recognizes the membrane glycoprotein DSG-2 [239]. The length of the fiber shaft seems to be another prerequisite for receptor usage [124,125]. While CD46-binding viruses generally have short and sturdy fibers containing only five to six β-spiral repeats, CAR-binding fibers are much longer and more flexible. The fiber length presumably determines the effective distance of the attached viral particle from the membrane and facilitates subsequent interactions with the entry receptor [225]. HAdV type D37 was found to engage CAR, but possesses a relatively short fiber and is thus likely not capable of efficiently mounting CAR-dependent infections [124,240]. Instead, the virus was shown to rely on sialylated glycoproteins for productive infection. The trimeric fiber knob engages the sialic acid portions of the bidentate GD1a-glycan in a positively charged pocket at its central trimer interface (see also **Chapter 3.2**) [76,240-242] (**Figure 9C**). In the course of this thesis, similar findings were obtained for HAdV-D36 and G52, which also engage sialic acid-containing glycans for primary cell attachment. According to several studies, HAdV-C05 is

believed to engage heparan sulfate proteoglycans (HSPGs) by a lysine-lysine-threonine-lysine (KKTK) motif located on the fiber shaft, leading to increased liver clearance of the virus [243-246]. In addition, binding of the hexon by coagulation factors can result in altered tropism and retarget the virus to liver cells [247]. A range of additional receptors have been proposed to serve in HAdV infections, among them the immune receptors MHC-1α and VCAM-1 as well as CD80/86 (reviewed in [225]). Furthermore, species C HAdVs engage lactoferrin, a soluble protein which acts as an adapter for Nucleolin on the cell surface. Newer results indicate that this interaction may be common among most HAdVs and is most likely mediated by the hexon protein and therefore fiber-independent ([248] and Persson D, personal communication). The importance of initial attachment factor recognition for viral tropism is still a topic of debate. This correlation might be species-specific, and is still not understood for most HAdV types. Interestingly, when grouped according to their fiber knob sequence, the EKC-causing HAdV types D08, D37, D53, D54, D56, and D64 appear to form a distinct clade together with HAdV-D09, D10, D19p, D59, and D65, none of which causes EKC (**Figure 8A**). In contrast, the amino acid sequences of the hexon and penton base do not seem to correlate with pathogenicity. The knobs of this clade undergo selective pressure at several sites [58]. It is currently unclear why some HAdV species have evolved to use several primary attachment receptors simultaneously with interactions mediated by the same knob, while others recognize only one attachment factor or possess two fibers with distinct binding specificities.

After successful attachment, the viral particle induces local crowding of its secondary receptor, either $\alpha_v\beta_3$ or $\alpha_v\beta_5$ integrin [249]. This process involves lateral movement of the attached viral particle on the cell surface in order to find the best spot for endocytosis and to provide a mechanical force for the initiation of fiber loss and uncoating of mature virus particles [250,251]. Integrins are heterodimeric type I membrane proteins that confer signals both from the outside to the inside of the cell and *vice versa*, effectively integrating the cell with its surrounding [252]. In humans, there are 24 known combinations of the different α- and β-chains. Integrins containing the α-chains IIb, 5, 8, and v recognize extracellular ligands bearing a prominent arginine-glycine-aspartate (RGD) tripeptide motif, typically found on proteins of the extracellular matrix such as vitronectin or fibrinogen [253]. Except for species F, all HAdV types possess such a motif on a surface-exposed loop of the pentameric penton base protein, and all of these types use α_v integrins as the main entry receptors (**Figure 9D+E**). Several models have been developed for this interaction, however exact information about the stoichiometry and the flexibility of the binding is still lacking [230,231]. Upon receptor recognition, integrins transit from an inactive bent state to their active extended conformation [254]. The recognition of the penton base has several effects on the virus capsid. For one, it rapidly induces the loss of the fiber on the cell surface [255]. Secondly, the interaction seems to alter the conformation of the penton base itself, making the whole viral particle more flexible and unstable and preparing the virus for uncoating [109,117,122,231]. Thirdly, the clustering induces integrin transmembrane signaling that upregulates Rho family GTPases and results in localized actin polymerization. This in turn leads to the formation of clathrin-coated pits that ultimately endocytose the viral particle [235,256]. In species B HAdVs, endocytosis is performed through macropinocytosis, although this process is also mediated by α_v integrins [257]. Within the endosome, several stimuli act together to promote viral escape. The physical stress induced by the receptors eventually leads to a loss of the penton base and peripentonal hexon proteins, along with the proteins that are located underneath the icosahedral vertices such as protein VI (reviewed in [56]). Protein VI then lyses the endosome via its amphipathic lytic domain and promotes viral escape, about 15 minutes after initial attachment [258]. In species C HAdVs, partial disassembly and release of protein VI at the plasma membrane leads to an alteration of the host cell's lipid landscape, increasing the level of the cone-shaped ceramide that further promotes endocytosis [259]. Additionally, these lesions seem to appear in early endosomes and prevent endosome acidification, but not maturation [260,261]. The reducing environment of the maturing endosome reactivates the AdV protease, which promotes endosomal escape and final virus uncoating [262]. The importance of this sequence of events is shown by the efficacy of host defensins that stabilize the virion by binding to the icosahedral vertices. The stabilized virions are unable to release their lytic factors and eventually end up in lysosomes where they are degraded [263]. Following endosomal escape, the subviral particles are transported to the nuclear pore complex (NPC) by dyneins that directly bind the viral hexon shell [264]. At the NPC, the genome is completely dismantled and transported into the nucleus, where it initiates the viral replication cycle [54].

1.8.2. Genome

HAdVs possess a linear dsDNA genome of 35-36 kbp in length. The virus hijacks the human RNA polymerase II and splicing machinery for transcription of viral mRNAs, but codes for its own DNA polymerase. The AdV genome is organized in 'transcription units', gene cassettes that are each transcribed from a common promoter element (Figure 10). With a few exceptions, this organization is very similar among HAdVs [55]. Our understanding of the adenoviral life cycle is largely extrapolated from experiments with only two HAdV types, C02 and C05 [79]. During the life cycle, viral gene expression is subject to a minute timing. The transcription units are grouped according to their peak expression time point, relative to the onset of viral DNA replication. In total, the viral genome codes for five early (E1A, E1B, E2-4), four intermediate, and one late transcription unit. Except for the latter, which harbors five different mRNA families, all transcription units code for several gene products that are transcribed from the same promoter. With the exception of the E4 unit, the gene products of a respective cassette carry out related and synergistic functions. While some gene products of a transcription unit share sequence stretches, others coded within the same unit may have a completely different sequence. The bundling into transcription units and the extensive use of alternative splicing likely helps the virus with control and timing of the expression of simultaneously needed gene products at each given time point in the life cycle. There is currently no naming convention for the respective proteins, but naming is usually kept consistent within a transcription unit [54].

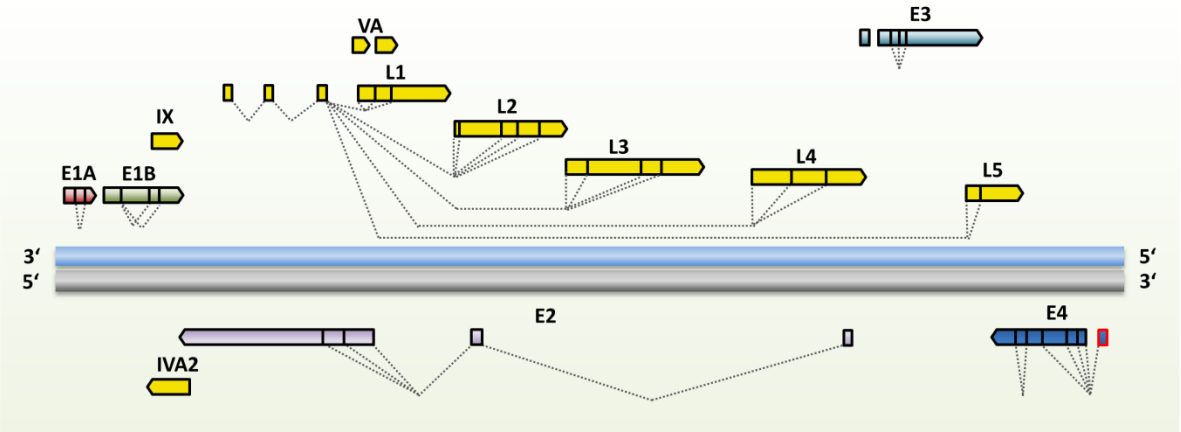


Figure 10 | Adenovirus genome organization. The genome organization of HAdV-C05 is shown representatively. *Early gene transcripts are annotated as in Figure 11, late and structural genes in yellow. Different splice variants are indicated by dashed lines. E4ORF1 is depicted with a red outline. HAdVs of species F and G possess two fibers that are both coded in the L5 gene cassette. Figure based on [265].*

1.8.3. Life Cycle

Once the viral genome has been delivered to the nucleus, the early phase of viral gene expression is started, lasting until the onset of viral DNA replication (Figure 11). The three main goals during early phase are A) to drive the host cell into the S phase and to mobilize metabolites for an effective replication; B) to circumvent host defense mechanisms, mostly immune reactions and apoptosis; and C) to effectively synthesize the viral gene products needed to advance the life cycle [54]. To accomplish these tasks, viral early phase proteins influence key players of cell homeostasis such as p53, retinoblastoma (rb), growth factor receptors, NFκB, and Bcl-2 family proteins.

The first proteins to be expressed are the two main gene products of the E1A cassette, large and small E1A, through the action of an exceptionally strong enhancer upstream of the E1A promoter. Both proteins act together to initiate the S phase of the host cell (i. a. by directly inactivating the rb tumor repressor protein), and they act as

transcription factors for other early viral genes such as E1B and E3 [54]. Some of these gene products further help to activate the host cell metabolism. As an example, the E4ORF1 gene product (E4ORF1p) influences the host metabolism by activation of the protein kinase mTOR through the PI3-kinase pathway, resulting in increased nucleotide and protein biosynthesis [266,267]. At the same time, E4ORF1p supplies the building blocks needed for replication. It increases anabolic glucose uptake by activating the MYC transcription factor and dysregulating several growth factor receptors [268]. Most of these functions are exerted by an association of E4ORF1p with the cellular PDZ protein Discs Large 1 (Dlg1), which stimulates both PI3K and growth factor receptors, leading to an increased MYC expression [269]. However, E4ORF1p is also believed to directly interact with MYC in the nucleus [270,271]. Additionally, E4ORF1p deregulates the host cell polarity by association with several adherens and tight junction proteins through its PDZ-domain binding motif (PBM), essentially weakening the epithelial barrier and potentially leading to altered signal transduction that deregulates cell growth [272].

The entry of the host cell into the S phase and increased metabolic activity inevitably generates several proapoptotic stimuli within the host cell. For example, the abnormal stimulation of the cell cycle leads to increased levels of the proapoptotic tumor suppressor protein p53, and an increased occurrence of dsDNA ends during the viral replication might potentially be perceived as heavy DNA damage and trigger DNA repair and apoptotic reactions [273,274]. In response to this, the E1B transcription unit expresses two main gene products that act together to block apoptosis of the host cell. The smaller E1B protein, E1B19kDa, is a mimic of an anti-apoptotic Bcl-2 family member and directly inhibits the proapoptotic agents BAK and BAX [275,276]. The second gene product, E1B55kDa, has various remarkable functions, among them the inhibition of the host checkpoint response to apparent DNA damage, and various means of inhibiting the tumor suppressor p53, partially aided by the E4-ORF6 gene product [277]. These mechanisms are crucial to the survival of the host cell. Viruses that lack either the rb-binding domain of E1A large or have an E1B55kDa knockout can only replicate in cells with deregulated apoptotic pathways such as tumor cells. This realization has led to the development of so-called oncolytic AdVs that conditionally replicate in rb- or p53-deficient tumor cells, respectively (reviewed in [278-281]). At the same time, the host cell activates the transcription factor NFκB, which stimulates innate and acquired host immune responses. As a response, the viral E3 transcription unit is activated through an NFκB-driven promoter. Its gene products counteract the effect of NFκB-induced immunity, thus serving as a negative feedback loop [54]. The E3 gene products are mostly species-specific and can be membrane-bound or even secreted, but all of them have immunomodulatory functions. As an example, the genus-conserved protein E3/19K directly interacts with newly synthesized MHC class I molecules, blocking their transport to the cell surface and thereby preventing antigen presentation to CD8+ T cells [282-284]. In addition, E3/19K down-regulates the MHC-like NK cell activators MICA and MICB [285]. Species D HAdVs possess a unique E3 protein called E3/49K, which is shed through proteolysis from the plasma membrane and acts as a soluble, extracellular NK and CD4+ T cell suppressor by engaging the pan-leukocyte marker CD45 [286,287]. E2 encodes three proteins that are necessary for DNA replication, among these the viral DNA polymerase. As a result, the peak transcription of the three E2 gene products is slightly delayed in comparison to the other early genes, setting in when a replication-friendly environment is reached. The onset of viral replication triggers the expression of the late gene products which mainly code for structural proteins and the AdV protease, which is essential for endosomal escape. Within the host cell, the AdV protease weakens the intermediate filament system by cleaving the cellular cytokeratin K18, preparing the host cell for its lysis. Additionally, an E3 gene product called adenovirus death protein kills and lyses the host cells as it accumulates during the late phase [54]. Before and upon cell lysis, a high amount of fiber-loaded penton dodecahedra and/or excess fiber trimers is released to the basolateral side of the epithelium, which presumably aids in the disruption of tight junctions in the epithelial environment and CAR- or DSG-2-mediated attachment to neighboring cells [134,288].

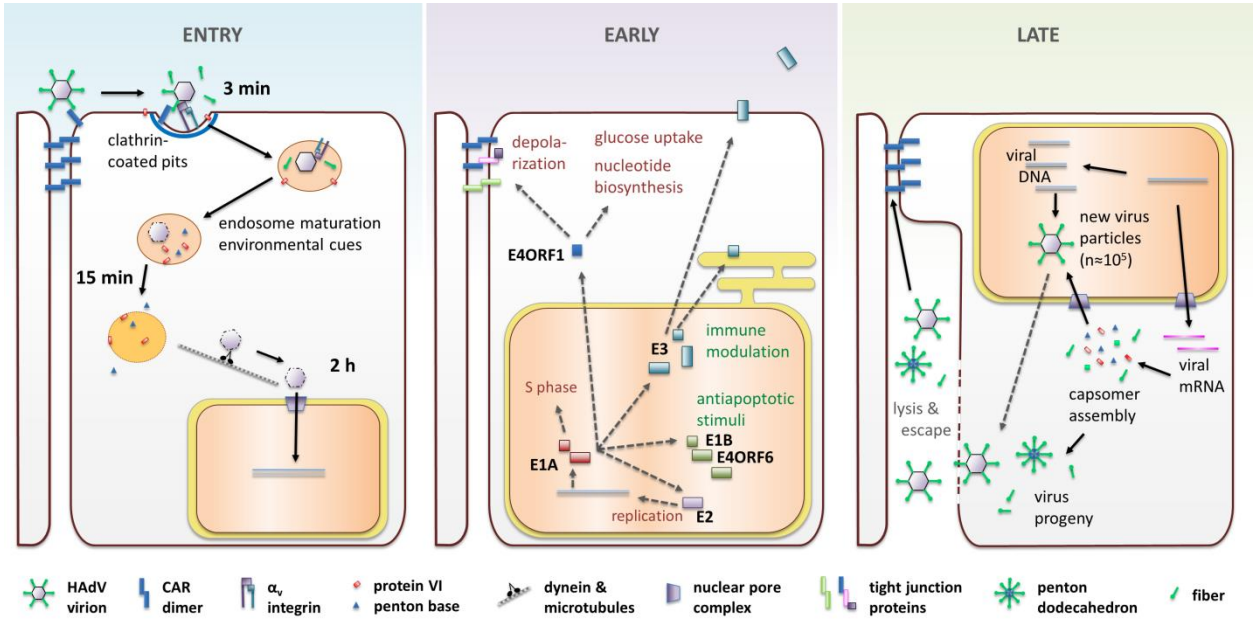


Figure 11 | Major steps of the adenoviral life cycle. *The process of replication and steps of minor interest to this study were omitted for clarity. Arrows that symbolically represent causal correlations are depicted as dashed lines.*

1.9. Viruses in a Clinical Context

Despite the apparent simplicity of viruses, our understanding of and interference with viral infections is often limited to preventive environmental measures or the treatment of symptoms. Viruses harbor the potential to enter the highly networked human population in one fell swoop, and one of the most recent examples, the emergence of HIV-1, dates back to less than a century ago [289]. Owing to a lacking adaptation phase, newly emerging virus mutants tend to be quite aggressive towards their hosts [2]. As a consequence, sudden epidemic outbreaks of virus infections can have devastating consequences, as was seen recently for ebola- and zikavirus epidemics (current WHO reports: [290,291]). For this reason, it is important to understand the physiology of viruses even if they appear to have only a minor impact on public health. In addition, viruses can also be used to combat diseases, as their capsids are essentially highly specialized delivery vehicles that can be engineered with relative ease.

1.9.1. Strategies of Antiviral Therapy

Four main approaches are being pursued to combat viral infections: hygiene methods, vaccination, immunotherapy, and drug development. The strategies employed for vaccine development include live-attenuated viruses, chimeric or recombinant virus particles or proteins, DNA-based vaccines, and the use of viral vectors. Vaccination approaches have been developed to treat both HPyV and HAdV infections. JCPyV vaccines for the treatment of progressive multifocal leukoencephalopathy are currently being developed based on VLPs and pentameric VP1, and both approaches show promise in first human trials [292,293]. Live vaccines of HAdV-E4 and B7 were successfully applied in the 1970's in order to tackle frequent epidemic outbreaks among military recruits. Nonetheless, the production of the vaccine was eventually put on hiatus in the 1990's [294], only to be picked up in 2011 after HAdV-related respiratory infections had resurged. The renewed treatment eventually resulted in the rapid elimination of the disease [295], but was never adapted to the civilian population. While vaccination has shown its enormous potential in the past, e.g. with the demise of the smallpox virus in the 20th century, its success

largely depends on the hosts immune response. The processes that orchestrate these responses, in turn, are dictated by the behavior of the virus itself. Yet, these processes are generally not well understood, and the development of efficient vaccines can be very complicated [296]. As an example, the development of an efficient dengue virus vaccine was hampered by a partial cross-reactivity of the antibody response - which enabled other serotypes to effectively infect host immune cells. It was only very recently that an efficient chimeric viral vaccine was approved in Brazil [297]. In order to obtain a safe and efficient vaccine, immunogenic parts of the target viruses had to be incorporated into a non-infectious yellow fever virus used as a vector. In a similar approach, HAdV-based vaccines carrying immunogenic HIV antigens have been investigated in a high-profile campaign by Merck, but failed to protect patients against HIV [298-301].

In cases of acute infection, treatment options are limited to either boosting the immune response through immunotherapeutic approaches, or to direct interference using drugs. Immunotherapy strategies against adenoviruses such as adoptive T-cell transfer are usually applied for immunocompromised patients that develop more severe infections, particularly following hematopoietic stem cell transplantation. Adoptive T cell transfer approaches rely on the immunogenicity of HLA-I restricted viral peptide epitopes. In the case of adenoviruses, several conserved peptides that are naturally presented on common HLA types have been described [302]. Immunocompetent patients, in contrast, should ideally be treated with appropriate drugs, which are generally easy to apply and do not require frequent interference by medical personnel. The first requirement for developing a drug is the identification of appropriate 'druggable' targets that fulfill several criteria. As such, they need to be crucial factors for the viral life cycle, function in a way that can be inhibited by a drug, and at the same time be sufficiently different from host factors to allow for a certain degree of specificity. While most well-known antivirals such as the pro-drug ganciclovir block viral replication, antiviral agents can theoretically interfere with any step of the viral life cycle [296]. The selection of a promising approach is a complicated task to begin with, since most stages of the viral life cycle are intricately interwoven with the host physiology. Hence, most of the eligible viral targets are either expressed within the host cell or are in fact deregulated host proteins, making drug toxicity one of the major obstacles in the quest for 'druggable' targets. The only available anti-adenovirus drugs to date are forms of the nucleoside phosphonate cidofovir, which unfortunately only have a limiting effect on infections and can have severe nephropathic side effects [303]. This inherent risk is partially overcome when focusing on the very first and last steps of the viral life cycle, cell attachment/entry and post-lytic cell egress, respectively. Since both approaches focus on events happening outside the host cell, these drugs are generally less likely to interfere with the host metabolism. APD-209, a novel drug for EKC-causing adenoviruses based on their ability to engage sialic acid during cell attachment, is currently undergoing phase II clinical trials [304-306]. However, there are also cases that warrant caution. For example, Natalizumab, an antibody targeting the glycan binding site of JCPyV, increases the risk of PML by promoting recognition of another, yet unknown receptor – further highlighting the importance of the full understanding of viral receptor interactions [208,307].

1.9.2. Adenovirus-Based Vectors for Therapeutic Approaches

Therapeutic vectors based on several viruses are being developed, including AdVs, retroviruses, and adeno-associated viruses (AAVs). These vectors are versatile tools for the specific delivery of effectors to pathological tissues that can be directed against pathogens, cancer cells, or with the aim of addressing genetic disorders. Usually, the vectors still carry some of the functions of the viral template, while they have been stripped of pathogenesis key factors and/or re-supplemented with other factors that influence their effect on the infected cells (presented exemplary in **Figure 12**). Depending on the task, viral vectors can be apathogenic or conditionally pathogenic, integrating or non-integrating, elicit or avoid an immune response, possess a broad or very specific target range, etc. HAdVs have been among the first viruses to be used as vectors, since they combine a relatively large packing volume and a 36 kbp dsDNA genome with a broad tissue tropism and a modular genomic organization that can be engineered with relative ease. To date, about 21% of all vectors that are used in gene therapy clinical trials are based on HAdVs (n=526) [308]. One very promising therapeutic approach is the use of

conditionally replicating HAdVs which specifically lyse p53- or Rb-deficient (tumor) cells (see **Chapter 3.2.1**). These 'oncolytic' viruses kill target cells and multiply within the tumor, and in the process they locally increase their concentration (**Figure 12C**) [309]. Interestingly, newer studies suggest that a large part of the efficacy of oncolytic viruses can be attributed to an 'adjuvant effect': the activation of antigen-presenting cells, mostly dendritic cells (DCs), by triggering an interferon response dependent on signaling from pattern recognition receptors, mainly toll-like receptor 9 (TLR9) [310-312]. The activated DCs can then present tumor-specific antigen epitopes to initiate a T cell response.

On the contrary, other mechanisms of innate and pre-existing adaptive immunity are generally thought to limit vector efficacy and to mediate side-effects. Uptake of viral particles by immune cells such as Kupffer cells is mediated by scavenger receptors and removes virus particles from the circulatory system. This uptake triggers the release of a number of proinflammatory cytokines such as interleukin-1α (IL-1α), IL-1β, IL6, and MIP-1α through various pathways [313,314]. Similarly, binding of the hexon by coagulation factor X shields the viral particle from the antibodies and the complement system, but increases the cytokine response after uptake by immune cells in liver and spleen (reviewed in [315]). In addition, pre-existing antibody responses can rapidly clear the applied vector. Lastly, the use of ubiquitously expressed receptor molecules can lead to off-target effects, as e.g. CAR is expressed on erythrocytes and platelets [316]. As a result, the adaptive and innate immune response to systemically administered HAdVs is much stronger than observed in naturally occurring infections. While being beneficial for an antitumor immune response, gene therapy approaches generally rely on systemic delivery and are impaired by immune responses triggered by the vector [317]. Even more so, this rapid and heavy immune elicitation can lead to hepatic or even systemic toxicity of the vector. In 1999, early clinical studies using Ad-based gene therapy for the treatment of ornithine transcarbamylase deficiency resulted in the death of a patient. This fatal development was caused by an invasion of multiple organs followed by a massive immune response triggered by the vector, ultimately leading to multiple organ failure and brain death [318]. In 2002, another patient developed a fatal case of leukemia after being treated with an integrating retrovirus vector in an attempt to cure his severe combined immune deficiency (SCID) [319]. Since then, researchers and governmental institutions have recognized the need for a profound understanding and control of the processes that influence vector efficacy and safety. These processes concern both vector behavior and effector functions, to equal parts. Much progress has been made in the development of safe effector systems, culminating in the approval of clinical trials for a CRISPR-Cas9-based T cell cancer therapy by the NIH in June 2016, an approach that harbors great potential also for *in vivo* therapies [320]. In 2012, Glybera, an AAV-based vector for the treatment of lipoprotein lipase deficiency, was the first gene therapy product to be approved for the German market [321], and the herpes simplex virus (HSV)-based oncolytic vector Imlygic (talimogene laherparepvec, T-Vec) was approved in 2015 by the US food and drug administration for the treatment of melanoma [322]. Newer approaches to improve delivery and safety on the HAdV-vector side include local application, the use of 'gutless' vectors that lack immunogenic genes, chemical shielding of capsid components to minimize antibody and coagulation factor binding, and the use of rare serotypes to circumvent pre-existing immune responses [316]. Yet, there is still only a very basic understanding of the underlying forces of fundamental processes that govern vector behavior, such as the selection of target tissues and host range. For HAdV-based approaches, most currently used vectors are based on the types C02 and C05, which have high seroprevalence and immune clearance rates and use highly abundant receptors for cell entry. While some applications such as gene therapy require a very broad target tissue range, especially those vectors targeted against tumors usually necessitate a highly specific vector tropism. To this end, developing new vectors with a more specific and controlled receptor profile is an important and fundamental task.

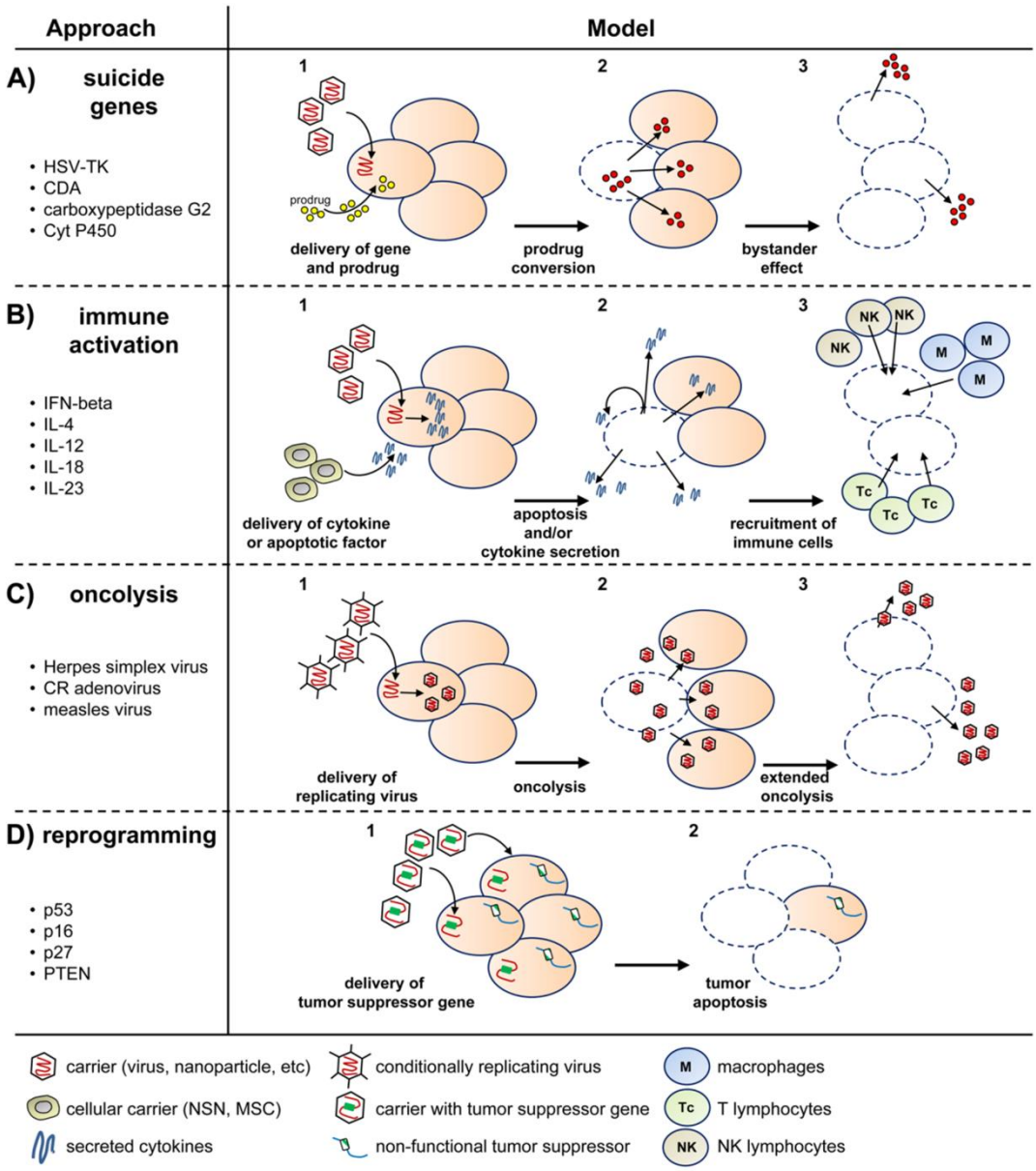
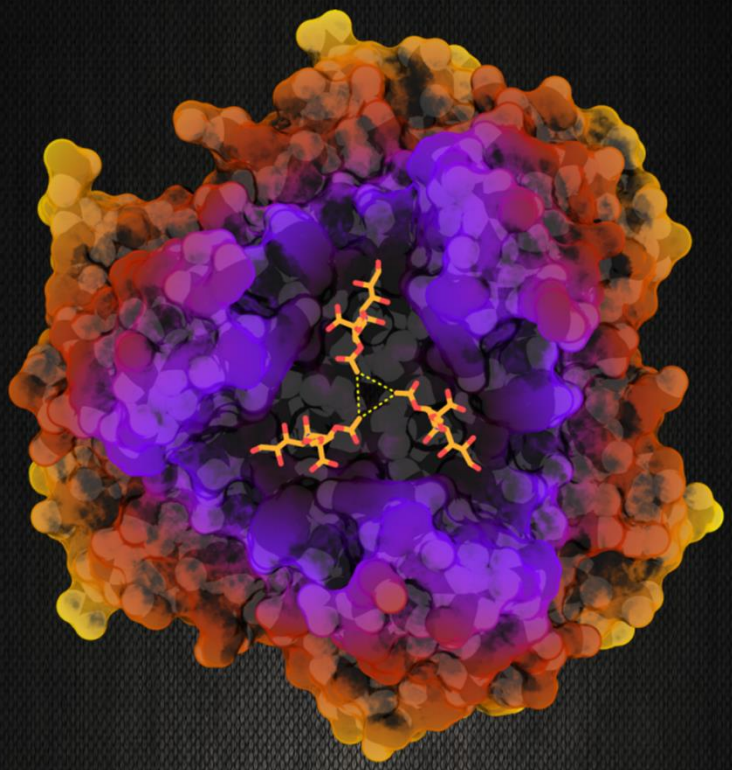


Figure 12 | *Different strategies for gene therapy of Glioblastoma multiforme.* *Four different strategies are graphically represented.* ***A*** *A commonly used strategy relies on the tissue-specific delivery of genes that convert a non-toxic prodrug into a toxin that kills the target and surrounding cells. A prominent example is the delivery of herpes simplex virus thymidylate kinase (HSV-TK), which causes the termination of DNA chain polymerization and cell death in proliferating cells by converting the pro-drug ganciclovir to its active tri-phoshorylated form.* ***B*** *Another approach relies on the tissue-specific delivery of immunostimulatory genes or apoptotic factors that trigger an immune response.* ***C*** *Oncolytic vectors conditionally replicate within proliferating cells in a self-proliferative way. This approach also mainly relies on the immune stimulation by cell debris. CR = conditionally replicating.* ***D*** *In a classic gene therapy approach, a tumor suppressor gene that may have been mutated by the tumor is resupplemented, leading to tumor apoptosis. All these approaches benefit greatly from the selective delivery of the cargo to tumor cells. Figure adapted from [323]. © 2013, Multidisciplinary Digital Publishing Institute.*

CHAPTER

2



2. AIMS OF THIS DISSERTATION

The introduction to this dissertation repeatedly highlighted the importance of the precise understanding of basic aspects of virus biology with regard to controlling virus outbreaks and the development of safe virus-based vectors for therapeutic approaches. The data collected in this dissertation contribute to three major aspects of DNA virus biology:

1) **General understanding of the key factors and dynamics of virus-host interactions.**

- **Chapter 3.1** exemplifies how minimal changes in the binding pocket of MuPyV drastically alter its virulence and tropism. This study has implications on how viruses establish a complex equilibrium in their receptor profile.
- **Chapter 3.2.1** reveals the attachment receptors of HAdV-G52, a rare HAdV type that is more closely related to SAdVs than to HAdVs, and whose two fibers recognize two different receptors.
- **Chapter 3.2.2** investigates the entry receptor profile of HAdV-D36, a virus that possesses a unique tropism and host range among HAdVs. We present the first data that correlate this behavior with the usage of a specific sialic acid variant that only occurs in animals. Additionally, this study examines the evolutionary processes that govern glycan recognition by species D HAdVs.
- **Chapter 4.1** provides groundwork for a structure of the complex between the soluble, pentameric HAdV-D09 penton base and the $\alpha_v\beta_3$ integrin. Structural information about this interaction would provide useful data about the binding stoichiometry and the weakening of the capsid by induced flexibility of the penton base protein.
- **Chapter 4.2** focuses on the structural basis of the interference of the adenoviral early gene product E4ORF1p with host cell proteins. The E4ORF1 protein has been identified as an oncogenic AdV factor and interferes with various cellular pathways. Additionally, the E4ORF1 protein is a potential 'druggable' target and is itself being tested as an anti-diabetes drug.

2) Structure-based development of antiviral drugs.

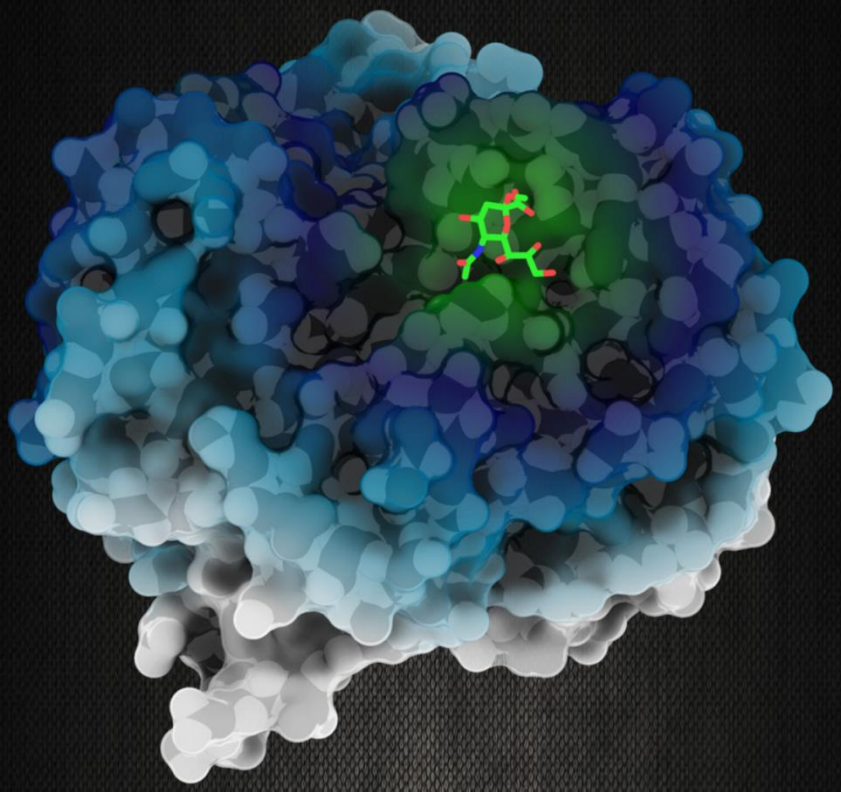
- **Chapter 3.2.3** demonstrates the development of highly efficient second-generation sialic acid-based trivalent inhibitors of HAdV-D37. The design of the inhibitors was derived from the crystal structure of the HAdV-D37 fiber knob and its physiological receptor, the GD1a glycan. Notably, these inhibitors are also able to bind the HAdV-D36 fiber knob.

3) Development of sophisticated adenovirus-based vectors.

- **Chapter 3.2.1** reports that HAdV-G52 uses the tumor antigen polysialic acid (PSia) as a preferred cell attachment receptor, and provides groundwork for the development of HAdV-G52-based vectors for targeting of PSia-expressing tumors.

CHAPTER

3



3. RESULTS AND DISCUSSION

3.1. Changes in the Receptor Binding Pocket Influence Entry Receptor Usage of Three Murine Polyomavirus Strains

MuPyV is a very well-studied DNA virus that causes tumors in newborn mice and hamsters [30,31]. In particular, the three MuPyV strains RA, PTA, and LID, which differ significantly in their tissue tropism and pathogenicity due to single amino acid exchanges in their receptor binding pocket, have sparked the interest of the field (see **Chapter 1.2**) [41-45]. This obvious relation between receptor engagement and both tropism and pathogenicity is very rare, and prompted us to structurally characterize the receptor binding properties of all three strains using X-Ray crystallography.

Gangliosides, the main attachment and entry receptors for many polyomaviruses, are usually engaged by the viral VP1 proteins at their glycan portions and directly mediate cell entry. In the case of MuPyV, the gangliosides GD1a and GT1b were the only established infectious receptors prior to this study [185] – however, other naturally occurring glycans are known to be recognized by the virus *in vitro* and are thought to influence the phenotype in yet undiscovered ways. Importantly, all of them possess a common disaccharide 'core motif': a Neu5Ac moiety that is linked to a galactose by an [α-2,3] glycosidic linkage ($Neu5Ac_a$ and Gal_a in **Figure 13**) [200,201]. This motif is crucial for the recognition and engaged at the canonical glycan binding site of PyV VP1 (see **Figure 6**). The structural basis for the recognition of glycans containing the 'core motif' by the RA strain has been established nearly two decades ago by Stehle and colleagues and led to the current notion on how the phenotypic differences between the three strains might emerge [105]. The authors used a synthetic glycan called DSLNT that additionally contains a second Neu5Ac moiety ($Neu5Ac_c$) which is attached to the core GlcNAc by an [α-2,6] glycosidic linkage (**Figure 13D**). The second Neu5Ac is engaged at an additional binding interface on VP1 that is separated from the canonical pocket by the residue at position 91. Based on this finding, the authors hypothesized that the RA strain can sustain this branched binding mode due to the presence of a glycine at position 91. In contrast, the two other strains would produce clashes as they harbor a more bulky glutamate side chain at this position and consequently have no access to this kind of receptor. The resulting narrower receptor repertoire is thought to allow PTA and LID to spread more effectively within the host.

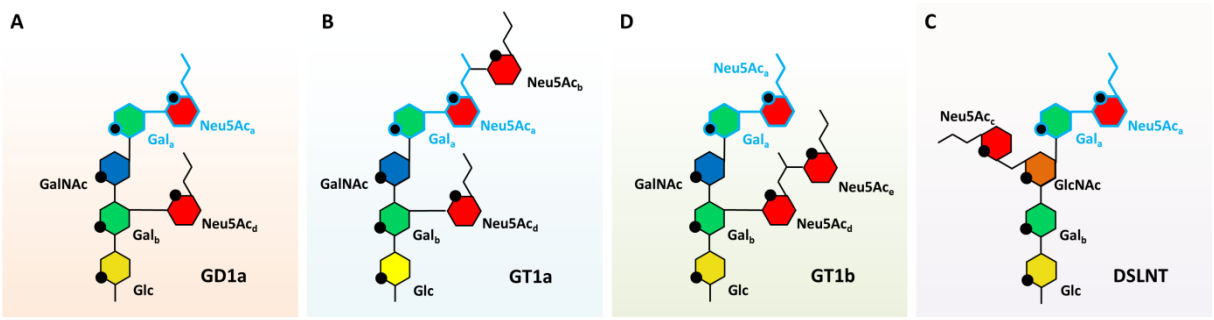


Figure 13 | Glycan portions of MuPyV receptor ganglioside candidates. *The gangliodies GD1a (**A**), GT1a (**B**), and GT1b(**C**) belong to the ganglio-series and consist of a Glc-Galb-GalNAc-Gala core decorated with alternating patterns of Neu5Ac. DSLNT (**D**) possesses a GlcNAc moiety instead of a GalNAc. The six-membered pyranoses are numbered counterclockwise starting from the bottom (C1, or C2 in Neu5Ac), and the ring oxygen is depicted as black dot. All linkages except those involving Neu5Ac are in the β conformation. The common core motif Gal_a-[α-2,3]-$Neu5Ac_a$is highlighted in blue. The figure is based on Figure 1 from [163].*

However, the complex structure of RA VP1 and DSLNT was the only structural information available prior to this work. Consequently, how the phenotypic difference between the two strains PTA and LID emerges is still unclear at this point, as an extensive structural analysis of their receptor profile has not yet been performed. We employed a comprehensive approach by testing the ability of all three strains to engage glycans containing additional, differently linked Neu5Ac moieties that are representative for most naturally occurring glycans containing a 'core motif' using X-Ray crystallography. To this end, we selected the glycan portions of GD1a (mainly containing the 'core motif' and a branching $Neu5Ac_d$), GT1a (containing an elongated $Neu5Ac_b$-[α-2,8]-$Neu5Ac_a$-[α-2,3]-Gal_a 'core motif'), and DSLNT as probes (**Figure 13A,B,D**). We chose to work with readily purifiable truncated VP1 constructs that do not assemble into VLPs and allow the generation of high-resolution structures that allow for a detailed assessment of the binding modes [105].

To our surprise, all of the glycans can be engaged by every strain, and all of them in a highly similar fashion. As expected, the GD1a glycan is engaged by the 'core motif' in a way very similar to that reported by Stehle and Harrison [324], and the additional $Neu5Ac_d$ does not contribute any additional contacts. Only two minor differences are observable among the strains: a weak van-der-Waals contact between a second Gal moiety (Gal_b) in the stem region of the glycan and the sidechain of E91 in PTA and LID that cannot form in RA, and a hydrophobic contact between V296 and C3 of $Neu5Ac_a$ that is not found in LID.

Interestingly, all three strains are capable of recognizing the GT1a glycan (**Figure 13B**), also by the 'core motif'. This interaction has not been reported to date. Moreover, the additional $Neu5Ac_b$ moiety of GT1a forms several additional polar contacts, for example between its carboxyl and glycerol functions and the key residues Y72 and R77. The glycan is stabilized by internal contacts and adopts a horseshoe-like conformation that is highly similar in all strains. We reasoned that this augmented binding interface might lead to increased receptor potency. Indeed, we were able to establish the GT1a ganglioside as a novel and highly potent physiological MuPyV receptor by supplementing ganglioside-deficient mouse embryo fibroblasts (MEFs) with GT1a and showing a significant increase in infection. In fact, GT1a seems to be much more effective at rescuing infection than the previously known receptors. However, there are virtually no differences in the way the three strains engage the glycan.

Furthermore, the bulky glutamate at position 91 does not seem to prevent binding of branched glycan motifs, as (in contrast to the earlier hypothesis) both LID and PTA are capable of engaging DSLNT in a fashion similar to GD1a. The branching [α-2,6]-linked $Neu5Ac_c$ does not contribute any contacts and is unordered in both complex structures. While the 'core motif' is positioned in a very similar way in each case, the carbohydrates forming the glycan stem and the branching $Neu5Ac_c$ display some moderate rearrangements. The slightly more spacious binding pocket of LID allows for some conformational freedom in DSLNT, which results in less well-ordered electron density at the stem. This difference is not visible in the more rigid glycans GD1a and GT1a. However, these conformational differences do not manifest themselves through altered binding contacts.

Although we did not have commercial access to the glycan portion of the established receptor GT1b at the time (which contains a disialic acid motif like GT1a, but at a branching position, see **Figure 13C**), our structures suggest that this glycan is engaged in a similar mode as GD1a, although its branching disialic acid motif might be long enough to form additional contacts with the protein. Overall, all three strains seem to be capable of engaging all tested glycans, and the difference in tropism and tumorigenicity cannot be explained by the use of different receptor subsets as was previously thought.

We reasoned that the differences in receptor usage must stem from different affinities rather than from differing receptor binding modes or selectivities. Unfortunately, the affinities between polyomavirus VP1 proteins and their glycan receptors typically reside in the low millimolar range, which makes them very hard to assess using standard techniques such as surface plasmon resonance (SPR). To overcome these difficulties, we established a method based on electron density integration of the 'core motif' in highly isomorphous crystals. Similar approaches have been successfully employed for adeno-, noro-, and other polyomaviruses before [136] or have been adapted based on our developments [204,325]. Using this approach, we were able to assess and compare the relative affinities of each strain for the respective glycans. Crystals were soaked with increasing concentrations of the respective

glycans for 1h, and the resulting electron density was integrated and plotted against the glycan concentration. In concordance with the complex structure, GT1a showed by far the best binding to all three strains. GD1a showed a weaker binding, and the relative affinity was slightly higher in PTA and LID than in RA. For the latter, the relative affinities for GD1a and DSLNT were comparable. DSLNT bound with the lowest overall affinity in all three strains. Surprisingly, there was no difference in DSLNT affinity between the strains, despite the moderate differences in DSLNT binding.

In summary, we observed only very subtle differences in binding affinities between the three strains, which seemingly contradicts the very distinct outcomes of infections. However, it appears conceivable that such minor alterations in the engagement of individual glycans translate into large differences considering the substantial avidity effects occurring in the MuPyV capsid and the dynamic behavior of the glycan receptors. As mentioned in **Chapter 1.6**, gangliosides are able to cluster within lipid rafts and can produce very high local concentrations [177,178]. Entry modes of this type are not unprecedented, as e.g. influenza viruses dramatically alter their cell binding properties upon small changes of their binding sites, which also display affinities in the mM range [140]. In the case of MuPyV, our data do not allow us to quantify how exactly the observed subtle differences in receptor engagement translate to the altered behavior of the virus particle, and future studies need to address these points in a capsid context. Unfortunately, working with MuPyV VLPs is a complicated task, since they result in poor yields and are very labor-intensive to purify and crystallize. In addition, suitable models for receptor crowding that also allow quantitative analyses are currently not availbale and very hard to establish.

As a further complicating factor, additional parameters may influence the viral pathogenicity and tropism. For one, the redundancy of glycan motifs on different glycoconjugates (glycoproteins, gangliosides, etc.) suggests the presence of non-productive 'decoy receptors'. Our analyses show that MuPyV is surprisingly versatile when it comes to engaging different glycan structures, as long as they harbor the 'core motif' and do not produce clashes. In this light, it seems plausible that additional glycans found on gangliosides and surface glycoproteins are bound by the virus. This is not an entirely new concept – however, in contrast to the original model established in the 90's, which suggested the use of different glycan reservoirs, new data imply that the same glycan motifs found on different macromolecular entities have differential effects on the viruses: In our companion manuscript, You *et al.* report that even in the absence of gangliosides, MuPyV is internalized by surface glycoproteins through an unknown alternative route [143]. However, this uptake route does not lead to an infection of the cell, but instead triggers an immune response that manifests itself in the initiation of a mitogenic response and the release of interleukin-12, a mediator of innate immunity. Based on these findings, we propose an updated model for the determinants of MuPyV cell entry: The ratio between productive and non-productive receptors in the viral microenvironment, combined with differential affinities of the virus for specific subsets of the receptor pool, balances and determines the productivity of viral entry through different routes. In addition, it cannot be excluded that effects occurring later in the life cycle are also mediated by the VP1 binding pocket. Virions are transported to the ER in complex with their receptor gangliosides, and these processes are still largely unexplored.

Altogether, the engagement of entry receptors by MuPyV seems to be much more complex and intricate than previously thought, probably involving a dynamic network of different receptor pools and decoy entry receptors, the usage of which can be fine-tuned by the virus. The virus needs to control these factors in order to create an equilibrium between efficient cell entry and cell egress after the completion of the lytic life cycle, while maintaining a sufficient selectivity for its infectious receptors. On a grander scale, there also needs to be an equilibrium in the consequences of viral infections for the host population, as frequent lethal infections also impair efficient viral spread. The fact that the more extreme strains RA and LID are only found in laboratory settings only emphasizes this point [43]. The way MuPyV interacts with its surprisingly big receptor pool critically depends on receptor availability in different situations. To date, our understanding of the receptor landscape is hampered by our relatively poor knowledge about the spatiotemporal distribution of glycan receptors and decoy receptors that does not allow for a stringent analysis.

Publication linked to this work:

Structural and Functional Analysis of Murine Polyomavirus Capsid Proteins Establish the Determinants of Ligand Recognition and Pathogenicity

Michael H.C. Buch*, **A. Manuel Liaci***, Samantha D. O'Hara, Robert L. Garcea, Ursula Neu, and Thilo Stehle

*These authors contributed equally to this work.

PLoS Pathog. 2015 Oct 16;11(10):e1005104. doi: 10.1371/journal.ppat.

Contributions: AML cloned LID VP1 by site-directed mutagenesis and purified, crystallized, and solved the native and complex structures of PTA and LID VP1. He developed the density integration procedure together with MHCB and UN. Furthermore, he wrote the paper together with MHCP, UN, SDO, RLG, and TS. He was involved in the planning of all experiments. Parts of this work are described in the Diploma Thesis of AML (Tuebingen, 2012).

3.2. Glycans as Primary Attachment Factors for Human Adenoviruses

3.2.1. Human Adenovirus 52 Uses Polysialic Acid and the Coxsackie- and Adenovirus Receptor for Binding to Target Cells

HAdV-G52 is a unique type that is only distantly related to other HAdVs [91]. Instead, its closest relatives are the SAdV types G01 and G07. HAdV-G52 is one of only three HAdV types equipped with two fiber proteins of differing length that are incorporated into the capsid in equal amounts. In this work, we identified the attachment factors for both of the fibers. We show that CHO cell lines that are treated with *V. cholera* neuraminidase are unable to mediate HAdV-G52 virion attachment, unless they heterologously overexpress CAR, suggesting that both CAR and sialic acid-containing glycans can act as attachment factors. Pretreatment of cells with different broad-spectrum proteases or inhibitors of different glycosylation pathways and ganglioside biosynthesis revealed that the glycans are most likely a part of mucin-type *O*-glycosylated proteins on the cell surface. Cell attachment assays with recombinant fiber knob domains of both the long and the short fiber revealed a very clear distribution of receptor engagement: The attachment of the long fiber knob (HAdV-G52 LFK, termed 52LFK in the publication) to CHO cells was only observed if they overexpress CAR, and was completely insensitive to neuraminidase treatment. SPR experiments showed that HAdV-G52 LFK engages the complete extracellular portion and the distal domain of CAR with a K_D of about 5 nM and 2.6 nM, respectively. With a length of 560 amino acids, the length of the long fiber is similar to that of HAdV-C05, which is known to use CAR as a physiological attachment factor. In sharp contrast, the short fiber knob (HAdV-G52 SFK, termed 52SFK in the publication) did not bind significantly better to CAR-expressing CHO cells, and was instead very sensitive to neuraminidase treatment. Furthermore, the HAdV-G52 SFK could not bind to sialic acid-deficient Lec2 cells, and it showed significant and selective binding to sialylated versions of fetuin in an ELISA assay. Although the binding affinity of soluble sialic acid-containing glycans to immobilized HAdV-G52 SFK resides in the low mM range (as shown by SPR), sialylated fetuin showed an affinity of 37µM, highlighting the importance of avidity for the binding. Thus, the long fiber appears to have an excellent specificity for CAR, while the short fiber binds to sialic acid-containing glycans decorating *O*-linked glycoproteins. We assessed the relative importance of CAR and sialic acid in a virus attachment inhibition assay using the human lung epithelium cell line A549, which presents both attachment factors on its surface. While the pre-incubation of virions with CAR, or of cells with a monoclonal anti-CAR antibody, inhibited virion attachment by 20% and 25%, respectively, the incubation of virions with soluble sialic acid or the pretreatment of cells with neuraminidase resulted in a respective reduction of 75% and 80%. Thus, sialic acid seems to have a more important role for cell attachment than CAR, at least on A549 cells, despite the much lower binding affinity.

We employed glycan array screening in order to identify the sialic acid-containing attachment factors. Interestingly, the array showed a drastically elevated signal for a group of linear [α2,8]-linked oligosialic acids that together mimic the naturally occurring glycan PSia (see **Chapter 1.6**). As already discussed, PSia is mainly expressed as a post-translational modification of the neural cell adhesion molecule NCAM-1 during embryogenesis and in some regions of the adult brain that need constant remodeling [187-189]. Apart from this, it is close to absent from the healthy human body (reviewed in [190]). Instead, the genes conferring polysialylation are upregulated by some tumors such as glioblastoma multiforme (GBM) and SCLC [191,194-198]. Therefore, PSia is generally considered an excellent tumor marker and currently ranked 31st among the most important cancer vaccine target antigens (and as the most important non-protein target) by the National Cancer Institute [326]. Due to the physiological functions of PSia in deregulating cell polarity and attachment to neighboring cells, PSia expression is correlated with metastasis and a poor prognosis [193,199]. According to the array, a length of more than three Neu5Ac moieties was required for a drastically elevated signal. A flow cytometry-based cell attachment inhibition assay with PSia of differing degrees of polymerization (DP) showed a more continuous result and indicated that a DP of three is already enough to significantly reduce the binding of HAdV-G52 SFK to A549 cells. These slight differences can be explained by the steric restrictions in the glycan array that are not present in FACS assays. We

tested the physiological relevance of PSia binding by monitoring the attachment of HAdV-G52 SFK binding to human neuroblastoma cells that either present or lack PSia on their cell surface. HAdV-G52 SFK bound significantly better to the PSia-expressing neuroblastoma cell line SH-SY5Y than to the control cell line, while no such effect was observed for the fiber knobs of HAdV-D37, C05, and G52 LFK. The binding could be reduced by pre-incubating HAdV-G52 SFK (but not HAdV-D37 FK) with soluble pentasialic acid.

HAdV-G52 SFK possesses a novel sialic acid binding site that is located at the sides of the knob trimer (**Figure 14A**). The three residues R316, G317, and N318 form a prominent 'RGN' motif that mediates the key contacts formed with the sugar's carboxyl, 4-hydroxy, and N-acetyl groups (**Figure 14B**). As for other viral lectins, these contacts are formed with only one side of Neu5Ac. The importance of these interactions was verified by mutational studies that completely abrogated HAdV-G52 SFK binding to different cell lines. Interestingly, oligosialic acids of DP3-5 are also engaged at this canonical pocket, forming directed polar contacts only with their non-reducing Neu5Ac moiety, as was verified by X-ray crystallography and STD-NMR. The three pockets on the trimer are too far apart from each other and not optimally oriented to allow for a bridged binding of long PSia chains, and the structures clearly show that sialic acids in 'endo' positions are not engaged. The contribution of the more distal carbohydrate portions stems from long-range electrostatic and transient polar interactions with a prominent electropositive rim located around the Neu5Ac binding site of HAdV-G52 SFK (**Figure 14C**), as became apparent from an MD simulation with PSia of DP8. Due to its [α2-8] linkage pattern, the poly-anionic PSia is conformationally flexible and is hovering above this 'steering rim', forming transient polar interactions with more distal residues such as R321, K349, and Q320 (**Figure 14D**). Again, these findings were verified by mutational studies that showed a sharp reduction in binding to SH-SY5Y cells for all selected 'steering rim' mutations.

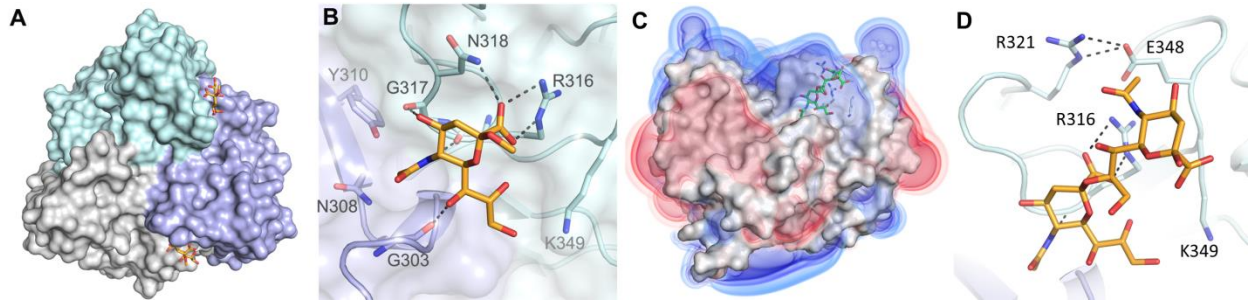


Figure 14 | Sialic acid and PSia engagement by HAdV-G52. ***A*** *The Sialic acid binding site I slocated at the side of the knob.* ***B*** *Details of the Sialic acid binding pocket.* ***C*** *The positively charged steering rim (blue) is located around the binding site. The electrostatic isosurfaces are shown at ±1; ±0.75; ±0.5 kT/e.* ***D*** *PSia is engaged at the non-reducing end, with the more distal Neu5Ac moieties hovering above steering rim residues such as R321, R316, and K349. Panel A and B are taken from [327]. © 2015, Lenman et al.*

Among HAdVs, the 'RGN' motif and 'steering rim' residues are unique to the HAdV-G52 SFK. Notably, they are also not present on the short fiber knobs of HAdV-F40 and F41, the only HAdV types besides HAdV-G52 that also possess two fibers. HAdV-D37 FK, which was used as a control for experiments addressing sialic acid engagement, was able to bind PSia to a certain extent, but was not sensitive to changes in glycan length. However, both key motifs of HAdV-G52 SFK are at least functionally conserved in the short fiber knobs of a range of simian AdVs of species G, including SAdVs G01, G02, G07, and G11 as well as SAdV-C19, all of which also possess two fibers. We tested the ability of SFKs from all these viruses to attach to SH-SY5Y cells – and with the exception of SAdV-G02, all of the knobs bound better to the PSia-expressing neuroblastoma cell line than to the control. However, the clearest relative change was observed for HAdV-G52 SFK. Thus, we conclude that PSia binding is a specific feature of a small and closely related subset of AdVs.

Although HAdV-G52 is associated with gastroenteritis, the low number of isolations does not allow for a stringent assessment of viral tropism and pathogenicity. This makes it hard to establish a clear relation between receptor availability and the tissues targeted by the virus [77,91]. Yet, apart from the fact that PSia synthesis is transiently upregulated in maturing immunogenic DCs [328], possibly aiding in DC migration, the virus is unlikely to come into direct contact with PSia-expressing human cells. Why, however, would a virus evolve to specifically recognize a receptor that it most likely does not encounter in healthy hosts? To date, the physiological relevance of PSia as HAdV-G52 attachment factor is still unclear. Due to its obvious similarity to SAdVs, HAdV-G52 might as well be a simian virus that gained the ability to infect humans (Kajon A, personal communication). Whether PSia is more prominently expressed in healthy simian tissues is not known, but this might be a possible explanantion for the observed PSia-specificity of HAdV-G52. Interestingly, PSia is also synthesized by some pathogenic bacteria such as *E. coli* K1 and *N. meningitidis,* which have evolved to express PSia through convergent evolution, most probably as a molecular mimicry measure that protects these pathogens from the host immune defense (reviewed in [329]). In this light, it appears plausible for HAdV-G52 to attach to enteric bacteria of humans and simians and use them for transport between hosts or to passively mount an infection, which would also provide a link to the presumed gastroenteric tropism of HAdV-G52. Although such a behavior has never been reported for adenoviruses, several reports establish a link between noro-, reo-, and poliovirus infections and commensal gut bacteria [148,149]. We are currently conducting experiments to show whether HAdV-G52 is capable of attaching to PSia-expressing gut bacteria. In addition, it is noteworthy that HAdV-G52 SFK can also use more abundant monosialylated glycans for cell attachment, albeit with a much lower efficiency than PSia. The alternative use of mono-and polysialylated glycans might explain why the virus has evolved to specifically recognize the non-reducing end of PSia instead of an 'internal' epitope that might produce a higher affinity. Since sialic acids usually are the capping parts of glycans, they are scarcely found at 'internal' positions in most abundant human glycans. By recognizing the non-reducing end and involving more distal Neu5Ac moieties through the 'steering rim', however, HAdV-G52 might increase its PSia affinity while maintaining the ablity to bind to other glycans.

Another question raised by our findings is why HAdV-G52 possesses two functionally distinct fibers at the same time. Although our results demonstrate that there is a very clear division of tasks between the two fibers, both of which recognize different attachment factors, this organization appears to be redundant at first sight. However, both fibers might fulfill complementary functions in several ways. For one, the spatiotemporal distribution of CAR, sialylated glycans, and PSia is highly divergent, and recognizing different receptors might help the virus to infect different tissues. Secondly, HAdVs are known to overexpress CAR- and DSG-2-binding fibers or penton dodecahedra that open epithelial junctions and facilitate virus egress and spread [134,288]. In a similar fashion, HAdV-G52 might use sialic acid-containing glycans to mount an initial infection from the apical side, and use its CAR-binding long fiber to spread within epithelial tissues or to reach beyond the epithelium. Thirdly, the long fiber of HAdV-F41, an enteric virus, was shown to not withstand gastric conditions, which led to a sequential model suggesting that initial infections in the gut are mediated by the short fiber knob that binds a yet unidentified receptor, while the long fiber is used by progeny virus within the infected tissue (Mangroo CS and Brown M, personal communication [330]). Additional insight would be gained by monitoring how the two fibers are incorporated into the virus capsid. The expression of both fibers is mediated by the same promoter, and both share a highly homologous N-terminus including the universal fiber motif FNPVYPY that was shown to mediate fiber attachment to the penton base [117]. If the incorporation of the different fibers is not just empirical, but regulated by a viral or host factor (e.g. on the translation level), the virus could adapt its capsid to a changing environment.

As discussed in **Chapter 1.9.2**, HAdVs are widely used as platforms for vector applications. Currently, most HAdV-based vectors are derived from the most well-studied type, HAdV-C05. However, the applicability of these vectors is impaired by increased hepatotoxicity due to coagulation-factor mediated liver tropism or rapid immune clearance of the virus by an existing immunological memory [331,332]. To this end, rare HAdV types such as HAdV-D26 and D48 are currently developed as alternative vectors [333]. In addition, depending on the application, vector safety and efficiency can be hampered by the lack of control over the vector's *in vivo* tropism, leading to off-target effects. Especially in the case of oncolytic viruses, therapy would greatly benefit from increased specificity

for tumors and metastases. In this light, HAdV-G52 virions with deleted long fibers would be an excellent vector candidate, since the virus does not bind coagulation factors and combines low seroprevalence and an expected low antibody cross-reactivity with a strong natural preference for the tumor antigen PSia. In fact, several vector systems that target PSia are currently under development, either involving bispecific adapters or replacing the whole fiber domains with a PSia-specific phage endosialidase called EndoNF ([334], Martin N, Gerardy-Schahn R, and Kuehnel F, personal communication).

We were able to enhance the PSia specificity of HAdV-G52 SFK by mutating residues in the periphery of the steering rim. For example, the replacement of K349 with an arginine enhanced HAdV-G52 SFK binding to SH-SY5Y cells in comparison to the wild type. We are currently in the process of constructing a PSia-optimized version of HAdV-G52 featuring a deletion of the long fiber gene (HAdV-G52v) that could be used as a platform for vector development.

In light of the dual mechanism by which oncolytic viruses fight cancer cells (see **Chapter 1.9.2**), the oncolytic potential of such a vector and the extent to which it inhibits an immune response will have to be addressed experimentally. The extent of immune activation appears to be type-specific and is not related to seroprevalence, as has been shown for the rare types HAdV-D26 and D48 [335]. A large portion of the immunogenic potential can be attributed to the relative frequency of TLR9-stimulating 'CpG-S' hexa-deoxynucleotide motifs and inhibitory 'CpG-N' motifs in the viral genome [336]. The relative CpG-S/CpG-N frequency in the HAdV-G52 genome is comparable to that of the immunostimulatory type HAdV-A12, and higher than that of e.g. HAdV-C05 (data not shown). Furthermore, although HAdV-G52 is genetically distant from other HAdV types and an assessment of the *in vivo* immunogenicity of potential T cell epitopes is difficult due to the low number of infections [77], several described HLA epitopes are conserved within its genome [337]. Future experiments with HAdV-G52v will include testing of PSia-dependent infection and lysis of our neuroblastoma cell lines and in vivo experiments using xenograft mouse models of PSia-expressing cancers such as GBM or SCLC as well as humanized mice with ectopic HLA expression [338].

In summary, we have discovered the two physiological primary attachment factors of HAdV-G52, sialic acid and CAR, both of which are specifically engaged by only one of its two fibers. Although the affinity for sialic acid is much lower than that for CAR, the former is the dominant attachment factor on A549 cells. We show that the virus uses PSia as a preferred cellular attachment factor using its short fiber knob and provide detailed and sensible structural insights into how this preference has emerged. The interaction between HAdV-G52 SFK and PSia is unusual and involves transient interactions and long-range electrostatic effects that are combined with sophisticated short-range contacts at the non-reducing ende of the glycan. The interaction is not observed in other HAdVs, and instead confined to a small subset of non-human species G HAdVs. Based on the various advantages of HAdV-G52 over conventional HAdV-based vectors that are combined with its ability to specifically interact with a glycan that is selectively expressed in high amounts on cancer various cancer types, we believe that HAdV-G52v is an excellent choice as an oncolytic vector basis for the specific treatment of PSia-expressing somatic cancers such as SCLC.

Publication linked to this work:

Human adenovirus 52 uses sialic acid-containing glycoproteins and the coxsackie and adenovirus receptor for binding to target cells.

Annasara Lenman, **A. Manuel Liaci**, Yan Liu, Carin Årdahl, Anandi Rajan, Emma Nilsson, Will Bradford, Lisa Kaeshammer, Morris S. Jones, Lars Frängsmyr, Ten Feizi, Thilo Stehle, and Niklas Arnberg.

PLoS Pathog. 2015 Feb 12;11(2):e1004657. doi: 10.1371/journal.ppat.1004657.

Contributions: AML purified, crystallized and solved the native and complex structures of HAdV-G52 SFK together with WB and LCK. He performed the structural analyses, and he designed the binding pocket mutants together with AL. In addition, he wrote the manuscript together with the other authors. Furthermore, he was involved in the planning of all experiments.

Unpublished Manuscript linked to this work:

HAdV-52 Short Fiber Knob Binds to Polysialic Acid

Annasara Lenman*, **A. Manuel Liaci***, Lars Frängsmyr , Yan Liu, Bärbel S. Blaum, Iva I. Podgorski, Balazs Harrach, Mária Benkő, Ten Feizi, Thilo Stehle, and Niklas Arnberg

*These authors contributed equally to this work.

Contributions: The contributions are described in detail in the manuscript (see **Chapter 6**).

Bachelor Thesis linked to this work:

Structural Investigations of the Short Fiber Knob from Human Adenovirus 52 and its Possible Ligands

Bachelor Thesis of Lisa Christine Käshammer. Supervisors: **A. Manuel Liaci**, Thilo Stehle.

3.2.2. HAdV-D36 and its Unique Receptor Profile

HAdV-D36 is a member of species D, the largest HAdV species containing 47 types to date [58]. Despite having closer relatives among HAdVs than HAdV-G52, the virus exhibits several surprisingly unique features. As such, it possesses a low cross-reactivity to antibodies directed against other HAdV types and a broad pathogenicity profile [85]. The virus is associated with gastroenteritis and, most distinctively, is the only human virus to be associated with human obesity. In addition, HAdV-D36 DNA is found in non-epithelial adipose tissue samples [86,339,340]. Like other HAdV types, HAdV-D36 infects a wide range of animals in experimental settings, in many of which it induces signs of adiposity and adipocyte differentiation [85,90,341,342]. However, the virus is also able to spread horizontally between chickens through natural routes, and anti-HAdV-D36 antibodies have been found in previously unchallenged rats and rhesus monkeys that had come in contact with humans before. While the adipogenic potential of the virus has been ascribed to the action of its early gene product E4ORF1p (see **Chapter 4.2**) [343], its tissue- and host tropism are believed to be at least partially controlled by the identity and distribution of host factors conferring attachment and entry. In this study, we set out to characterize the attachment factors of HAdV-D36.

In an approach similar to that employed for HAdV-G52, we probed for the chemical nature of possible HAdV-D36 attachment factors by pretreating suitable cell lines (A549) with different proteases and inhibitors of glycan incorporation, and by measuring the ability of HAdV-D36 virions to attach to different CHO cell lines that overexpress known or candidate HAdV attachment factors. Unfortunately, HAdV-D36 virions exhibited much less clear tendencies than HAdV-G52. As such, virion attachment was sensitive to treatment with the proteases ficin and proteinase K, as well as inhibitors of *O*- and *N*-linked protein glycosylation – however, comparative attachment assays with CHO cells that overexpress the known receptors CAR and CD46 as well as CD21, ICAM-1, and CD55 did not point out any specific molecule as a main attachment factor. Clear tendencies emerged only when we repeated the assays using recombinant fiber knobs instead of whole virions. These assays pointed out a clear preference for cells overexpressing either CAR or sialic acid. Indeed, the attachment of HAdV-D36 virions to A549 cells could be partially inhibited by soluble CAR, although less efficiently than that of CAR-binding HAdV-C05. Similarly, treatment of cells with benzyl-GalNAc (an inhibitor of *O*-glycosylation) impaired HAdV-D36 attachment to A549 cells by about 50%, and tunicamycin (an inhibitor of *N*-glycosylation) by about 15%. Similarly, pre-incubation of virions with soluble sialic acid inhibited virion attachment to A549 cells with an IC_{50} of about 20-25 mM, albeit never to 100%. In comparison, pre-incubation of virions with heparin had a slight enhancing effect for virion attachment, and there was no significant loss in binding to GAG-defective CHO cells. Similarly, inhibitors of ganglioside and GAG synthesis did not impair virion attachment. These findings suggest that both CAR and sialic acid-containing glycoproteins can serve as primary attachment factors for HAdV-D36. In contrast to our findings for HAdV-G52, both factors are simultaneously recognized by the HADV-D36 FK. At the moment, we cannot explain the differences observed when using the virions and fiber knobs. However, species D HAdVs (especially HAdV-D09) are known to efficiently engage α_v integrins by their penton base (see **Chapter 4.1**) - usually the secondary receptor - to an extent that they can use them for attachment [344]. The penton bases of HAdV-D09 and D36 have RGD- and variable loops of similar length, and it appears conceivable that the differential presence of integrins might influence the results obtained for virions. Similarly, it is possible that a yet unidentified novel attachment factor is present on the cells. This hypothesis is supported by the finding that blocking virions with both CAR and sialic acid at the same time had a synergistic effect, but never increased the binding by more than 80% on both human (A549) and rodent (CHO) cells.

We have analyzed the sequence and solved the crystal structure of the HAdV-D36 FK in order to gain insights into the conservation of the receptor binding interfaces. In comparison to the most closely related structurally characterized knob, HAdV-D37 FK, the HAdV-D36 FK possesses several distinctive features. For one, the whole knob domain possesses a significantly lower electropositive surface charge than HAdV-D37 FK. Further, the trimer interface in HAdV-D36 is less spacious due to a slight shift in the orientation of the monomers with respect to each other. The (quite literally) most outstanding feature of the knob, however, is a considerably elongated DG loop that projects far from the side of the knob. This loop is eight amino acids longer than that of the HAdV-D37 FK and

is partially attached to the counterclockwise neighboring monomer through polar and hydrophobic contacts. Interestingly, the DG loop is located in the proximity of the canonical binding interfaces of both CAR and CD46 (see **Chapter 1.8.1**). Superposition of the HAdV-D36 FK with the complex structure of HAdV-A12 and the distal domain of CAR indicate that this loop would interfere with potential CAR binding, and complex formation would either necessitate a rearrangement of the loop or a novel binding mode. We were able to form a stable complex of recombinant soluble CAR and HAdV-D36 FK in size exclusion chromatography experiments, however we have not yet conducted any structural studies on the interaction. Notably, the short shaft of the HAdV-D36 fiber indicates that even if CAR can be engaged, it is likely not able to cooperate with integrins to confer productive internalization, similar to HAdV-D37 (see **Chapter 1.8.1**).

In contrast to the canonical protein receptor interfaces, the sialic acid binding interface found in HAdV-D37 is highly conserved in HAdV-D36. Indeed, the complex structure of the HAdV-D36 FK with α-2-O-methyl-Neu5Ac reveals that sialic acid is engaged in the canonical binding site located at the center of the knob near the threefold axis, in a way that is highly similar to the HAdV-D37 binding mode (**Figure 9C**). All polar key contacts are conserved, and the main difference lies in the less spacious trimer interface arrangement that allows HAdV-D36 to form direct polar interactions with both faces of the sugar. However, cell-based assays, *in vitro* assays, and co-crystallization trials consistently indicate that HAdV-D36 engages Neu5Ac only about half as efficiently as HAdV-D37 does (IC_{50} values of 20-25 mM and 12 mM, respectively). The most probable explanation for this discrepancy is the lower electropositivity of HAdV-D36 FK, which was shown to be an important factor e.g. for PSia engagement of the HAdV-G52 SFK.

We next sought to identify which Neu5Ac-bearing glycoconjugates support HAdV-D36 binding by means of glycan array screening. Interestingly, two independent glycan arrays revealed that HAdV-D36 has a strong preference for a 3'sialyllactose variant carrying a 4-O-acetylated Neu5Ac (termed 4-O-Ac-3'SL in the manuscript), and that this effect is not seen in HAdV-D37. We solved the co-crystal structure of the HAdV-D36 FK with 4-O-Ac-3'SL and found that contacts are only mediated with the Neu4,5Ac_2 portion of the glycan (**Figure 15A,B**). Therefore, we resorted to the more readily accessible α-2-O-methyl-Neu4,5Ac_2 for further studies. We are currently testing these findings in a cellular context. To this end, we are carrying out comparative attachment inhibition assays where we test the inhibitory potency of α-2-O-methyl versions of Neu5Ac and Neu4,5Ac_2 on virion and FK attachment to various cell types. Additionally, we will test the ability of HAdV-D36 to infect several Neu4,5Ac_2-positive cell lines (including horse and salmon cell lines) and test if the process depends on Neu4,5Ac_2 by treating the cells with neuraminidase or 4-O-acetylation-specific viral esterases such as ISAV hemagglutinin-esterase (see **Chapter 1.6**). Furthermore, we will specifically screen for Neu4,5Ac_2 on cells of the adipose lineage, and test whether HAdV-D36 can infect adipose cells.

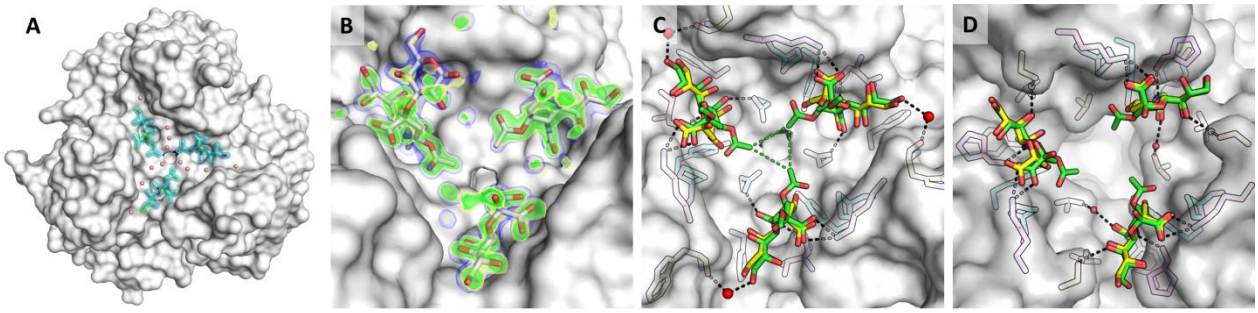


Figure 15 | Neu5Ac, 4-O-Ac-3'SL, and Neu4,5Ac_2 engagement by HAdV-D36 (A-C) and D37 (D). ***A*** *Overview of the HAdV-D36 binding pocket. The carbohydrate colored by B Factor and well-ordered water molecules are displayed. Waters in the center of the cavity are displaced.* ***B*** *4-O-Ac-3'SL is engaged only by its Neu4,5Ac2 moiety. F_o-F_c omit electron density (3σ, green; 2.5σ, yellow) and $2F_o$-F_c density (1σ, blue) after refinement are depicted.* ***C*** *Neu5Ac (yellow) and Neu4,5Ac2 (green) binding by HAdV-D36.* ***D*** *Neu5Ac (yellow) and Neu4,5Ac2 (green) binding by HAdV-D37.*

Using X-Ray crystallography, we were able to solve the structures of both HAdV-D36 and D37 in complex with α-2-O-methyl-Neu5Ac as well as α-2-O-methyl-Neu4,5Ac$_2$ using a large excess of ligand (**Figure 15C,D**). All complexes gave high-resolution structures that allow for an interpretation of the observed differences. The HAdV-D36 FK engages Neu4,5Ac$_2$ in a similar orientation as Neu5Ac, but the steric repulsion between the 4-O-acetyl group and the two binding site residues Y315 and P320 pushes the ligand towards the central threefold axis of the knob trimer. This twist is observed in all three sugar binding pockets and causes the three 4-O-acetyl functions to approach each other and engage in a triangular hydrophobic contact with an average length of 4.2 Å, thereby disposing ordered water molecules from the center of the pocket. Together with residues 313, 315, and 320, two 4-O-acetyl groups form a hydrophobic binding cavity for the third ligand. It is tempting to speculate that this arrangement leads to cooperative binding effects that are able to make up for the lower *ab initio* affinity of the HAdV-D36 FK - especially since the effect is not observable in HAdV-D37, whose wider central cavity does not bring the three 4-O-acetyl functions in close enough proximity. When comparing the Neu4,5Ac$_2$ binding modes of both viruses, it becomes apparent that the differences originate mainly from the altered dimensions of the central cavity. Through close inspection of the fiber knob structures, we identified three residues of the G strand that are responsible for a large part of the trimer interface and control the relative inclination of the monomers towards the threefold axis. While HAdV-D36 FK possesses a 'VSN' tripeptide motif at positions 311-313, HAdV-D37 FK bears a 'YGT' motif at the analogous positions 308-310. Interestingly, the species D HAdV fiber knobs cluster into two distinct phylogenetic clades when aligned, and HAdV-D36 and HAdV-D37 locate in opposing clades (**Figure 8A**). Members of the 'HAdV-D36-like clade' possess a DG loop of similar size, and most of them carry a 'VSN' trimer interface configuration. It is noteworthy that none of the types belonging to this clade has a reported protein receptor to date. Members of the 'HAdV-D37-like' clade, in turn, possess much shorted DG (or FG) loops and a conserved 'YGT' or 'YGN' interface. This clade contains all EKC-causing HAdVs [58]. Mutating the 'YGT' motif in HAdV-D37 to 'VGN' reduces the size of the central cavity and brings the three Neu4,5Ac$_2$ moieties into a position very similar to that of HAdV-D36. The HAdV-D36 'YGT' mutant, in turn, lost the ability to bind Neu4,5Ac$_2$ due to a distortion of its N-acetyl binding cavity. Despite this pattern of conservation, engagement of Neu4,5Ac$_2$ in a cooperative fashion is most likely not a common feature of HAdVs of the 'HAdV-D36-like' clade, as was shown by the complex structures of HAdV-D48 (HAdV-D36-like) and D26 (HAdV-D37-like). In these cases, minor mutations outside the binding pocket cause a complete or substantial loss of Neu5Ac binding, and none of them is likely to support Neu4,5Ac$_2$ binding. Hence, we postulate that Neu4,5Ac$_2$ binding is a unique feature of HAdV-D36 or a small number of species D viruses closely related to HAdV-D36. Complex structures and molecular modelling experiments with 7-, 8-, and 9-*O*-acetylated Neu5Ac variants suggest that 4-*O*-acetylation is the only modification of this sort that is able to create the described cooperativity effect.

Since the high IC$_{50}$ values strongly indicate a low affinity of the interactions, *in vitro* affinity studies to compare different protein mutants and acetylation patterns are challenging and cost-intensive. The method developed for MuPyV is not applicable to HAdV-D36, since the complex changes its space group and there is a low isomorphism between crystals. We are currently exploring the possibility to compare the chemical shift perturbations caused by ligands in a TROSY-NMR protein spectrum. This method is very sensitive and might allow at least the qualitative comparison of different ligand affinities. Despite its size, the trimeric HAdV-D36 FK protein gives a well-interpretable TROSY spectrum with detectable shift perturbations upon ligand addition in preliminary test runs (data not shown). Notably, STD-NMR studies with HAdV-D36 FK and 4-O-Ac-3'SL did not give any detectable spin saturation transfer differences, although the interaction has been shown in the various experiments described above. One reason for missing STD signals is a very slow off-rate that influences the spin relaxation time of the bound ligand. It seems plausible that a cooperative binding mode of Neu4,5Ac$_2$ might also lead to altered binding kinetics.

To date, the physiological relevance of the interaction between the HAdV-D36 FK and Neu4,5Ac$_2$ remains enigmatic. Current tissue profiling data suggest that Neu4,5Ac$_2$ is the rarest O-acetylated Neu5Ac variant in vertebrates [169,345]. The gene conferring 4-O-acetylation is unknown to date, but Neu4,5Ac$_2$ has never been found in healthy human tissue samples and is generally considered to be absent from humans. In contrast to this, it frequently occurs in domestic or livestock animals [345], many of which (chickens, mice, rats) have been shown

to be susceptible to HAdV-D36 infection. To date, HAdVs are generally believed to be highly species-specific. However, the unparalleled preference of HAdV-D36 for an animal carbohydrate, whose species distribution fits very well with the virus' unique ability to infect certain animals, makes it very tempting to speculate that HAdV-D36 might possess an animal reservoir. In fact, binding for O-acetylated Neu5Ac variants is a common strategy among animal viruses - and even human viruses such as Influenza virus C [176]. One general principle is that sialic acid variants often have limited expression profiles and might thus be used by pathogens to populate specific evolutionary niches. In particular, several viruses infecting mice, rats, and salmon have evolved to use Neu4,5Ac_2 as the main attachment factor, and some have evolved a 4-O-esterase function to facilitate viral release (reviewed in [164]). Another example is HPyV9, which may use nutritionally acquired Neu5Gc to infect a subset of tissues [147]. In this light, it seems plausible that HAdV-D36 adopted a similar strategy in order to expand its host range by means of small changes of its sialic acid binding site, and potentially by using a different protein receptor. The more restricted expression profile of sialic acid variants might additionally help the viruses to avoid excessive binding to non-productive 'decoy'-receptors and thus increase the efficiency of infection in specific tissues. It would be enlightening to screen for HAdV-D36 seropositivity among Neu4,5Ac_2-synthesizing livestock animals that are in frequent contact with humans, e.g. poultry or horses. At the same time, several enteric bacteria such as *N. meningitidis* also synthesize Neu4,5Ac_2 [168], and some bacteria such as *S. aureus* are in fact found to infect and persist in adipose-like 3T3-L1 cells [346]. Thus, another hypothesis would be that HAdV-D36 is able to adsorb to bacteria of this sort to mount infections of the gut, and possibly even adipose tissue, in fashion similar to that proposed for HAdV-G52.

In summary, we present initial evidence suggesting that HAdV-D36 uses CAR and Neu4,5Ac_2-containing glycotopes, presumably on glycoproteins, as preferred attachment factor, and we provide a sensible structural basis for this preference. It seems likely that HAdV virions also recognize a third, yet unidentified surface molecule that partially contributes to the attachment process, potentially α_v integrins. These findings will be evaluated in a cellular context and might have broad implications for our understanding of the unique features of HAdV-D36 and species D HAdVs in general.

Unpublished Manuscript linked to this work:

Primary Attachment Receptors of Human Adenovirus 36

A. Manuel Liaci, Naresh Chandra, Sharvani Munender, Yan Liu, Vanessa Pfenning, Paul Bachmann, Rémi Caraballo, Wengang Chai, Emil Johansson, Karolina Cupelli, Timm Hassemer, Bärbel S. Blaum, Mikael Elofsson, Ten Feizi, Niklas Arnberg, and Thilo Stehle

Contributions: The contributions are described in detail in the manuscript (see **Chapter 6**).

Bachelor Theses linked to this work:

Structural Investigations of the Human Adenovirus 36 Fiber Knob and its Candidate Receptors

Bachelor Thesis of Vanessa Pfenning. Supervisors: **A. Manuel Liaci**, Thilo Stehle.

Structural Insights into the Sialic Acid Binding Modes of Human Adenovirus D-36 and D-37 Fiber Knob Domains

Bachelor Thesis of Paul Bachmann. Supervisors: **A. Manuel Liaci**, Thilo Stehle

3.2.3. Triazole Linker-Based Trivalent Sialic Acid Inhibitors of HAdV-D37 Infection of Human Corneal Epithelial Cells and Potential New Inhibitors for HAdV-D36

Epidemic keratoconjunctivitis (EKC) is a severe and well-defined ocular infectious disease mainly caused by a subgroup of species D HAdVs comprised of HAdVs-D37, D08, D53, D54, D56, and D64 (formerly D19) [80,347-350]. Millions of cases are reported every year, most commonly in highly populated areas in Asia [351]. Although self-limiting, EKC is painful, highly contagious, lasts for several weeks and can cause corneal opacities that lead to vision impairment for a period of up to two years [352]. Treatment options are currently limited to symptomatic measures, as systemic treatment with the acyclc nucleoside analogue cidofovir can cause nephrotoxicicty and blockage of the lacrimal canal [353]. As mentioned in **Chapter 3.2.2**, the EKC-causing HAdVs cluster into a closely related 'EKC clade' when aligned by the sequences of their fiber knobs [58]. One of the main EKC-causing agents, HAdV-D37, has been shown to use the glycan GD1a on cell surface proteins as a physiological attachment factor that is found on corneal epithelial cells, and other EKC-causing HAdVs are thought to be able to engage sialic acid-containing glycans through similar binding sites [76]. Interestingly, sialic acid-binding HAdVs seem to cause more severe symptoms than non-binding types, accompanied by a higher incidence of corneal opacities [354]. In the case of HAdV-D37, the main factor that distinguishes the high-affinity attachment factor GD1a (IC_{50} 16 µM) from sialic acid (12 mM) is the valency of the ligand. As discussed in **Chapter 3.1**, GD1a contains two terminal [α-2,3]-linked Neu5Ac moieties, and the topology of the HAdV-D37 FK binding pocket allows for a bidentate binding mode that occupies two of the three available Neu5Ac binding sites (**Figure 9C**). Given the absence of receptor-cleaving enzymes in adenoviruses, this finding inspired the development of several antiviral agents exploiting avidity effects, e.g. by fusing sialic acid to serum albumin or artificial membranes [304,355,356]. Among these agents was a set of novel Neu5Ac-based trivalent small-molecule inhibitors whose distances between terminal Neu5Ac compounds mimicked those found in GD1a and that showed excellent inhibitory potential [306]. These compounds contained three terminal Neu5Ac moieties that were joined to a trimeric core (e.g. tris(2-aminoethyl)amine) by means of a linker of varying flexibility (**Figure 16A**). Although the best compound, ME0322, had an *in vitro*-affinity comparable to GD1a (K_D values 14 µM and 19 µM, respectively), ME0332 was about four orders of magnitude better than GD1a at inhibiting HAdV-D37-infection of human ocular cells. These findings served as a basis for the development of the sialic acid-based antiviral drug APD-209 by Adenovir Pharma AB, which is currently evaluated in phase II clinical trials [305].

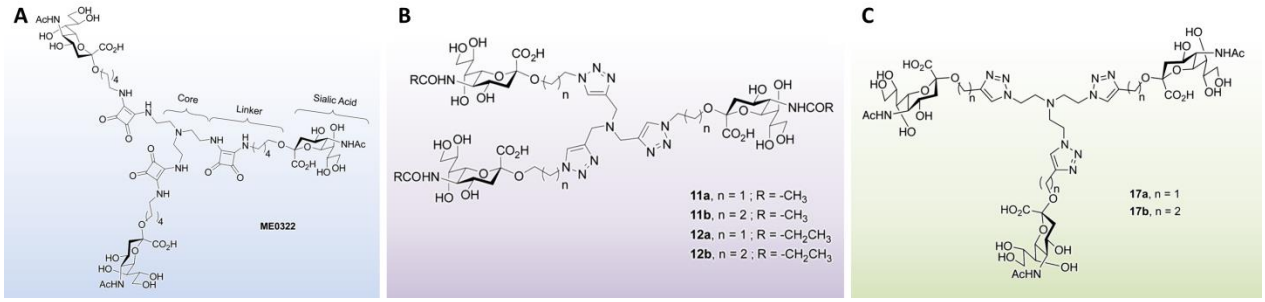


Figure 16 | Different generations of trivalent HAdV-D37 inhibitors. ***A*** *Squaric caid-based inhibitor ME0322.* ***B*** *First-generation triazole linkers.* ***C*** *Second-generation triazole linkers. Figure modified from [357]. © 2015, Royal Society of Chemistry.*

The study presented here describes the development and evaluation of a set of second-generation trivalent inhibitors. In contrast to the previously reported inhibitors that were synthesized using squaric acid chemistry, these new compounds could be conveniently trimerized using 'click' chemistry. We probed for different orientations of the resulting triazole ring, different lengths by alternating the methyl and ethyl spacers of the core and linker parts, and the replacement of the N-acetyl group of Neu5Ac by an N-propyl function. The compounds

were evaluated by means of SPR experiments, human corneal epithelium (HCE) cell attachment inhibition, inhibition of infection, structural investigations using X-ray crystallography, and ophthalmic toxicity in rabbits. The synthesis of the new first- and second-generation triazole compounds could be achieved in eight steps from commercially available educts. The two sets of inhibitors differ in the orientation and relative positioning of their triazole groups. The first-generation compounds 11a and 11b were prepared with the azide group on the Neu5Ac side, and are thus connected to the core by one of their carbon atoms and to the lower part of the linker by a ring nitrogen (**Figure 16B**). Both compounds differ in the length of the alkane linker that connects the rings to the Neu5Ac parts. Compound 11a contains an ethyl linker and showed IC_{50} values of 107 nM and 172 nM for virion attachment and infection, respectively. Compound 11b contained a propyl linker and was slightly more effective (40 nM and 54 nM, respectively). In contrast, the replacement of the N-acetyl function by an N-propyl group resulted in considerably reduced efficacies in the micromolar range. The rationale behind this modification was that an increased lipophilicity at this position might be beneficial for the binding. However, crystallographic analyses revealed that these compounds are slightly pushed out of the N-acetyl binding pocket, and that this altered arrangement leads to a slight distortion of the binding interactions that is pronounced enough to cause a loss in engagement efficacy. The second-generation inhibitors contain inverted triazole rings that are moved closer towards the Neu5Ac moieties (**Figure 16C**). The linkers that join these groups to the terminal carbohydrate parts are constituted by CH_2 or ethyl groups for compounds 17a and 17b, respectively. In this case, the shorter compound, 17a, showed an excellent potency as it inhibited virion attachment and infection with IC_{50} values of 1.4 nM for attachment and 2.9 nM for infection - and was about 200-fold more effective than the longer compound 17b. Also, 17a was about three orders of magnitude more potent at inhibiting cell attachment and two orders of magnitude better at inhibiting infection than the squaric acid-based original compound ME0322.

We tested the ophthalmic toxicity of compound 17a in rabbits in order to test their physiological tolerability. To this end, male New Zealand white rabbits were subjected to 48 administrations of 40 µl 1 mg/mL 17a per eye for a period of seven days, and relevant values such as weight, intraocular pressure and corneal thickness were recorded and compared to a control group. Administration of compound 17a did not cause irritation of ocular tissues or changes in the monitored values, giving a pre-clinical basis for the evaluation of the compounds in future clinical trials.

In summary, our triazole-based inhibitors represent a significant improvement of the first generation squaric acid-based compounds both in terms of efficacy and synthetic ease. All of the compounds were engaged by the HAdV-D37 FK by their Neu5Ac parts in the canonical sialic acid binding pocket, and the best compound, 17a, showed an extraordinary inhibitory potential in the low nM range. Although the linkers connecting the three Neu5Ac moieties are non-charged, exhibit only small differences in terms of polarity, and do not form any polar contacts with the fiber knob, the efficacy of attachment and infection inhibition of the different compounds varied considerably (1.4-376 nM for attachment, 2.9-681 nM for inhibition). A linear correlation between linker length and inhibitory potential could not be established. Crystallographic analyses showed that compound 17a had a more flexible linker than the longer 17b, which possesses a high degree of internal order causing it to adopt a relatively rigid bell-like shape with a slight tilt in the Neu5Ac moieties. In fact, 17a seemed to be the most flexible of the compounds, and presumably this flexibility accounts for its increased inhibitory potency by reducing the entropic cost of the binding. Similar entropic contributions have been demonstrated for the carbohydrate engagement of bacterial lectins [166]. Due to their high potency and tolerability, we believe that our compounds harbor great potential as drugs for the topical treatment of EKC.

Using a similar approach as that used for HAdV-D37, we tested the ability of compound 17a to inhibit HAdV-D36 attachment to A549 cells. As verified by X-Ray crystallography, 17a does indeed bind to HAdV-D36 FK. In contrast, HAdV-D26 only bound the inhibitor ME0322 at excessively high concentrations (>100 mM), showing the specificity of the interaction. The compound showed a roughly 1000-fold lower efficacy for HAdV-D36 than for HAdV-D37 binding. This seems plausible given the lower *a priori* affinity of HAdV-D36 FK for Neu5Ac that is likely multiplied by the avidity effect of the trivalent inhibitors. To this end, we are currently synthesizing a new generation of inhibitors for HAdV-D36 binding that connect the three sialic acid moieties through the O4 position, mimicking the

binding mode of Neu4,5Ac_2. Due to stability issues, the ester functions will be replaced with amides, and the three carbohydrate moieties will be linked by a minimal trivalent nitrilotripropionic acid core. In contrast to the current generation of compounds, this linker will likely contribute additional contacts to the binding.

Publication linked to this work:

Triazole linker-based trivalent sialic acid inhibitors of adenovirus type 37 infection of human corneal epithelial cells

Rémi Caraballo, Michael Saleeb, Johannes Bauer, **A. Manuel Liaci**, Naresh Chandra, Rickard J. Storm, Lars Frängsmyr, Weixing Qian, Thilo Stehle, Niklas Arnberg, and Mikael Elofsson. Organic and Molecular Biochemistry

Org Biomol Chem. 2015 Sep 21;*13*(*35*):*9194-205*. doi: 10.1039/c5ob01025j.

This article is part of the themed collection '2015 Hot Articles in Organic and Biomolecular Chemistry' and was featured on the cover of Issue 35 (2015).

Contributions: AML purified HAdV-D37 FK together with JB and crystallized and solved the complex structures of compounds 17a and 17b. He wrote the paper together with the other authors.

Unpublished Manuscript linked to this work:

Primary Attachment Receptors of Human Adenovirus 36

A. Manuel Liaci, Naresh Chandra, Sharvani Munender, Yan Liu, Vanessa Pfenning, Paul Bachmann, Rémi Caraballo, Wengang Chai, Emil Johansson, Karolina Cupelli, Timm Hassemer, Bärbel S. Blaum, Mikael Elofsson, Ten Feizi, Niklas Arnberg, and Thilo Stehle

Contributions: The contributions are described in detail in the manuscript (see **Chapter 6**).

3.3. General Considerations and Implications for the Field

3.3.1. Rules of Engangement

Our various protein-carbohydrate complex structures confirm and extend existing knowledge about viral glycan binding sites. As many other viral sialic acid binding sites, they are mostly shallow and form contacts only with the α-face of the sugar [139]. Most of the key contacts involve the distinctive functional groups of sialic acids: the carboxyl groups that engage in salt bridges, the N- (and O-) acetyl groups that contribute mostly through hydrophobic contacts, and sometimes direct or water-mediated hydrogen bonds formed by the hydroxyl groups of the glycerol chain. The presence of several acetyl groups dictates a high level of surface complementarity that is easily disturbed by single mutations. In the cases of MuPyV and HAdV-D36 and D37, key amino acids form contacts with several amino acids at the same time. From an evolutionary point of view, a restriction of the interaction to single residues or short residue stretches might help the viruses to use the interacting residues efficiently and limit the need for structural conservation to uphold infectivity. Additionally, residue motifs of this kind might be used for efficient receptor switches or differential receptor usage, as was shown for the VSN motif in HAdV-D36 and the RGN motif of HAdV-G52 that is only present in species G HAdVs. In the course of our studies, it became clear that additional effects have a critical impact on the usefulness of glycans as attachment factors for some viruses. For one, transient polar and long-range hydrophobic and electrostatic interactions as well as steric and entropic effects and the flexibility of ligands have been demonstrated here to be decisive factors for preferential ligand and drug binding in the cases of HAdV-G52 and D37. Similarly, the internal order of the GT1a and GD1a glycans has subtle implications for MuPyV receptor binding in comparison to the more flexible DSLNT. Although these are not novel principles (see **Chapter 1.6**), they are difficult to assess experimentally and therefore often neglected when analyzing crystal structures. Secondly, small rearrangements of the quaternary structure of lectins can have similar effects. These phenomena often go unnoticed when comparing structures to homologues e.g. from related virus types, especially when relying on sequence information only while doing so. In the case of HAdV-D36, thorough analysis of structural data delivered information that would not have been accessible otherwise. However, the analysis of small structural changes warrants caution with respect to over-interpretation, and it is imperative to monitor the model quality with respect to coordinate errors or reliability of the electron density, and to validate any conclusions in experimental settings.

We and others have demonstrated that HAdVs G52, D36, and D37 all recognize both CAR and sialic acid-containing carbohydrates as attachment factors [232]. However, HAdV-G52 has distributed these two tasks onto two fibers, while HAdV-D36 and D37 use only one fiber to recognize both molecules simultaneously. The reasons for this discrepancy are currently unknown. Both HAdV-D36 and D37 possess short fibers that are thought to not allow using CAR for the infection of cells in combination with integrins, and CAR binding might chiefly serve to dissolve tight junctions by excess fibers after the lytic cycle [125,225,316]. In this case, the virus still needs to recognize an attachment factor that supports integrin binding. In this light, it would seem more efficient for the viruses to possess two different fibers for two different tasks. However, each additional molecule inevitably leads to the presence of additional epitopes that might be recognized by the immune system. Additionally, the genome size of HAdVs is generally limited by genome packaging into the virus particles, and adding one gene generally comes at the cost of deleting another one. Both processes might limit the evolutionary incidence of incorporating more than one fiber, which apparently only occurred once within the simian lineage.

The sialic acid binding side of species G AdVs is much more surface-exposed than the canonical site found in species D HAdVs which is hidden in the central cavity of the protein. According to the so-called ‘canyon hypothesis’, hiding of the receptor binding site is a mechanism to avoid immune clearance by simple competitive receptor blockade [358]. This suggests a clear necessity for species G Advs to bind their receptor at such an exposed location. The presence of the ‘steering rim’ in these AdV types explains this necessity, as this binding mode is not possible e.g. at the central cavity.

3.3.2. Methods of Assessing Virus-Carbohydrate Interactions

Structural studies with virion particles are usually cumbersome and challenging, and they seldom result in the high-quality data that are required to assess details of carbohydrate engagement. Therefore, working with recombinant isolated capsid proteins has great benefits and has enabled countless insights that would not have been possible otherwise. Especially in cases such as reo- or adenovirus fiber head domains that project far from the capsid, these insights are likely to hold up in a virion context, as well. However, working with the single, recombinantly expressed proteins brings about several intrinsic simplifications that should not be neglected and that potentially pose an obstacle to the direct correlation of receptor engagement and physiological effects on cell attachment and infection. For one, virions may possess additional epitopes for the same or other attachment factors that additionally influence the virus' behavior. Secondly, the presence of a viral genome is known to alter receptor engagement properties, e.g. by establishing a cross-capsid electrostatic potential, or by altering the pressure inside the capsid [121,122]. As discussed in the introduction, avidity and cooperativity effects play essential roles in virus glycan engagement [141,142]. The list of examples has been extended by MuPyV, HAdV-G52, HAdV-D36, and perhaps most impressively by the potency of our trivalent HAdV-D37 inhibitors, whose efficacy was increased by a factor of 10^6 in comparison to monovalent sialic acid. The case of MuPyV VP1 receptor engagement shows how minimal alterations of binding affinity (that can hardly be assessed experimentally when investigating single interactions of the VP1 binding pocket) translate to large differences in a virion context.

Dissecting (and correlating) these effects is complicated by a number of inherent methodological difficulties. On a technical note, assessing thermodynamic or kinetic data of glycan engagement is often very challenging, since the avidity effects result in very weak individual protein-carbohydrate interactions residing in the mM range. Many conventional methods such as SPR or grating-coupled interferometry have proven to not be sensitive enough to assess minute changes in the course of this work, while others such as ITC require massive protein concentrations and ligand amounts that are prohibitively expensive and produce estimated errors that are too high to give reliable values (data not shown). Even very sensitive techniques such as TROSY-NMR titration and assessment of chemical shift perturbations are limited by the fact that saturation is hardly reached and that the homo-oligomeric capsid proteins are comparably large. While being very useful in the detection of potential receptor carbohydrates, glycan arrays or ELISA-based methods generally do not allow for quantitative analyses due to steric constraints. We have established a crystallography-based technique that allows a qualitative comparison of binding events among different MuPyV strains through integrating electron density in differentially soaked crystals. However, this method is quite labor-intensive and requires highly isomorphous and well-diffracting crystals. For our studies with MuPyV, we have solved more than 50 differentially soaked crystal structures for the process, many of which had to be discarded. In the case of HAdV-D36, this strategy was not applicable due to the low isomorphism of the different complex crystals. Although working with virus-like particles or even complete viruses might overcome these limitations, the absence of robust models for local glycan crowding and mobility make it hard to validate the results. The development of more robust techniques for affinity assessment as well as more true-to-life membrane models is needed in order to correlate single binding events and their implications for virion attachment.

3.3.3. Physiological Roles of Glycan Binding

The development of new and more sophisticated glycan arrays continues to reveal new details of glycan engagement by HAdVs. The discoveries made in this work implicate that the use of non-human glycans might be more frequent than anticipated, and might point at novel survival strategies of HAdVs, including the hijacking of bacteria or the use of animal reservoirs. Several newer studies suggest that many RNA and DNA viruses interact with the human microbiota, and addressing this topic may add to our understanding of how other HAdV types reach and infect epithelial tissues, as well. These findings might substantially impact our understanding of the adenoviral life cycle and open new ways to interfere with diseases.

Furthermore, although some viruses, especially polyomaviruses, show a remarkable specificity for single glycan types, our studies suggest that some viruses are more promiscuous in terms of receptor engagement. It seems likely that viruses of this sort also recognize non-productive glycan motifs that may play a role for antiviral immune responses and for the epidemiological equilibrium within the host. As such, the presence of a ubiquitous decoy receptor would require the production of more virus progeny, and it might lead to virus particles being present in tissues that they cannot infect. Consequently, receptor promiscuity might be an explanation for seemingly paradoxical results *in vivo*. In this light, it seems important to correlate the mere presence of a particular virus in the respective tissues with clear indicators of infection.

The use of recombinantly expressed viral lectins to investigate monovalent interactions with glycans is very useful to gain detailed insight into glycan recognition and to establish hypotheses of the origins of differential receptor usage. However, correlation of glycan binding data collected *in vitro* with *in vivo* tropism and selectivity remains challenging. The bulk of the experimental difficulties come from a lack of experimental tools to robustly resolve the dynamics of specific glycan presentation in tissues or to recreate these effects *in vitro*. Even the profiling of different sialic acid variants is challenging to date. As stated above, primary attachment factors are generally necessary for infectability of a cell. However, they are not the only factors that govern virus tropism, and the interplay of these factors is still not well-understood. Future developments in the fields of virology and glycobiology will therefore be instrumental in obtaining new insights into how the early steps of the viral life cycle of glycan-binding viruses influence their tropism and pathogenicity.

3.3.4. Vector and Drug Development

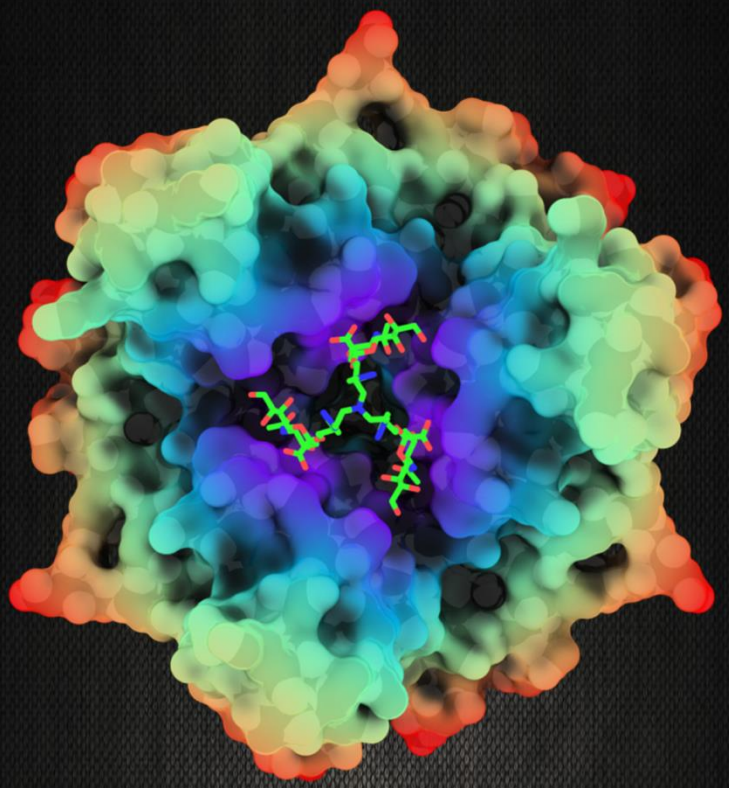
Currently, the treatment of many viral infections is limited to preventive or symptomatic measures. The process of drug development is quite tedious, in part owing to insufficient cell permeability or toxicity of drug candidates. Drugs with extracellular targets circumvent this problem in an elegant fashion. Our HAdV-D37 inhibitors can be topically administered to the eye, thereby preventing problems of serum clearance or secondary toxicity. With the new compounds, we hope to clear spreading virion particles during the acute phase, which typically lasts for two to three weeks. Addressing the early phase of the disease might minimize the severe long-term symptoms caused by sialic acid-binding HAdV types, and might restrict the disease to only one eye. The functional parts of these inhibitors are physiologically occurring carbohydrates, which should further minimize side effects. Similar approaches are used in different fields as well, and homoglycoclusters are widely used as antiadhesive agents, immune modulators, or diagnostic tools [359]. The development of a trivalent drug based on physiological receptor interactions demonstrates once more how naturally occurring principles of receptor recognition can inspire the directed and efficient development of new drugs.

The use of adenoviral vectors based on HAdV-C05 has proven to be beneficial for some applications due to their strong elicitation of immune responses and the use of widely distributed receptors that allow an efficient systemic application. However, among the manifold applications of viruses as vectors are some that warrant extraordinary tissue specificity. The understanding of details in receptor recognition is crucial especially for therapeutically administered and (conditionally) replication-incompetent viruses, where receptor usage is expected to be the main determinant of cargo delivery. In this light, the observation that some HAdV types can bind several attachment factors with the same knob domain is a very important piece of information. Some cancers such as glioma express low levels of the most prominent HAdV-C05 primary attachment factor, CAR, while others such as SCLC do not show homogeneous CAR profiles and do not permit uniform transduction [360]. Consequently, there have been numerous attempts to retarget HAdV knobs to recognize tumor antigens. Redirecting is currently done by designing adapters, fusing short peptides into surface loops of the fiber knob, or by replacing knob domains with different proteins [334,361-364]. However, these approaches often do not allow for efficient virus replication, which would be desirable in some cases. We propose a new vector system that overcomes common problems of existing vectors and combines these benefits with the specific ability of targeting to PSia-expressing tumors such as

SCLC or, with some restrictions, GBM. The expression of PSia is involved with depolarization and metastasis, which can usually not be specifically addressed by surgery. To this end, vectors generated against PSia-expressing cells might prove very useful for the treatment of metastases, even in combination with surgery.

CHAPTER

4



4. ONGOING RESEARCH

4.1. Interactions between the Adenovirus 9 Penton Base and Human Integrin $\alpha_v\beta_3$

For most HAdV species, the second step of the entry process is conferred by the interaction of RGD-binding surface integrins with the adenoviral penton base protein (see **Chapter 1.8.1**) [365]. The only HAdV types that do not possess an RGD motif are types F40, F41, G52, and - perhaps surprisingly - D60 [366]. As described in the introduction, the interaction between penton base and integrin alters the conformation of both proteins. On the integrin side, the interaction triggers several signaling processes leading to clathrin-mediated endocytosis, while the untwisting of the penton base protein is thought to destabilize the viral particle for later uncoating [122,235]. Several structural studies have been conducted on this interaction, mostly using cryo-EM techniques [230,231,367,368]. A 2009 structure reporting the interaction of the HAdV-A12 virion with $\alpha_v\beta_5$ integrin suffered from technical difficulties arising from the icosahedral shape of the capsid [231]. Although an exact stoichiometry has not been robustly determined, the authors conclude that integrins are too large to bind to the pentameric penton base in a 5:5 ratio, which essentially precludes studies with icosahedral particles due to the need to impose pentameric symmetry (**Figure 9E**). Further light has been shed on the plasticity of the interaction in a recent study with a monomeric fragment of the HAdV-D09 penton base (HAdV-D09pb) published by our laboratory [230]. Type D09 was chosen for two reasons: due to its relatively short RGD loop (**Appendix 6.2.3**), which is thought to reduce the flexibility of the binding and facilitate high-resolution structural analyses, and because it interacts strongly with α_v integrins [344,369]. The extraordinary strength of the interaction is believed to originate from a triple-threonine (TTT) stretch directly N-terminal of the RGD motif that is also found in physiological interaction partners such as fibronectin or vitronectin [344]. The authors of the newer EM study discovered four distinct interaction modes occurring between the monomeric HAdV-D09pb fragment and $\alpha_v\beta_3$ integrin (**Figure 9D**). Interestingly, some of these modes suggest additional binding events that are independent of RGD loop recognition. However, it has remained unclear whether these interactions occur simultaneously while several integrins are crowding at the virus vertices, or if they rather present distinct steps of a sequential mechanism that perhaps leads to structural changes in the proteins. A precise description of the binding dynamics and interface would give valuable insights into the mechanisms behind virus internalization, the initiation of virus uncoating, and the initiation of integrin-dependent antiviral signaling. In addition, such a structure would present a platform for the development of novel antiviral drugs or adenoviral vectors that do not target integrins.

While the purification of the extracellular portion of $\alpha_v\beta_3$ integrin has been reported elsewhere [230], we initially aimed at establishing a protocol for the heterologous expression and purification of HAdV-D09pb in *E. coli*. Usually, penton bases are expressed in baculovirus systems, and expression in *E. coli* would greatly facilitate the production. HAdV-D09 is one of several types that has been reported to form minimal T=1 icosahedral penton dodecahedra (**Chapter 1.4.2**) [370]. While dodecahedra might be physiologically relevant and useful for some applications [134,371]; they also possess an icosahedral symmetry and are thus obstructive to our cryo-EM studies. The structural basis for dodecahedron formation has been extensively studied and pinned to two loops at the base part of the penton molecule. Both loops make up the inter-pentamer interface and were shown to be a prerequisite for dodecahedron formation of HAdV-B03. While several amino acids contribute to the interface, special importance has been ascribed to one conserved residue on each loop, an asparagine and an arginine (D100 and R425 in HAdV-B03). Two copies of these residues, located on neighboring pentamers, each form salt bridge at the interface. Furthermore, a swapping of the N-terminal 63 amino acids between dodecahedra is thought to substantially contribute to the formation and integrity of these particles [116]. Several publications have reported that the recombinant expression of penton base in *E. coli* results in insoluble inclusion bodies [372,373]. To this end, we have initially designed a dodecahedron-defective variant of the protein with a cleavable N-terminal GST

tag, which was reported to facilitate the solubilization of HAdV-B01pb [374]. The construct contains extensive deletions of the loops found to mediate dodecahedron formation, and a partially truncated N-terminus. This fusion construct is soluble and expressed in large amounts in *E. coli*. However, it produces a fraction of C-terminally truncated proteins that are incorporated into the oligomers. There are two main truncations occurring at discrete points, and one of them presumably affects the last β-strand in the sequence. In addition, a small fraction of the proteins is proteolytically cleaved within the ligand-binding RGD loop during protein expression, which would likely interfere with structural analyses and might lead to some incorrectly assigned subsets in single-particle analyses. Unfortunately, extensive trials of expression conditions involving a large set of parameters (e.g. different strains, expression times, promotors, induction times, expression temperatures, etc.) did not serve to improve the quality of protein expression in *E.coli*. Negative-stain transmission electron microscopy and dynamic light scattering further revealed that the particles are indeed oligomeric and monodisperse, but they are inhomogeneous and co-purify with GroEL chaperones. Only few particles showed observable pentameric features. The C-terminal truncations and co-purification of GroEL have been present in other penton base preparations [374] and are generally a sign of poor folding quality. We therefore conclude that - contrary to previous publications and although penton bases neither contain disulfide bridges nor glycosylations - *E.coli* is generally not a suitable expression system for HAdV penton bases. In addition, the extensive mutations in two loop regions may impair protein folding.

In order to overcome these limitations, we have designed a less invasive construct of HAdV-D09pb that is optimized for heterologous expression in *S. frugiperda* cells (Sf-9 or High Five). These constructs contain more radical truncations of the N-terminus, and only a single point mutation of R369 (corresponding to R425 in HAdV-B03). Since protein solubility is not expected to be an issue in insect cells, we replaced the GST tag with a cleavable N-terminal His-tag. In addition to HAdV-D09pb, we designed a similar construct of HAdV-A12pb, which is not known to form dodecahedra and possesses the shortest RGD and variable loops of all HAdVs, although it is expected to bind less well to α_v integrins [344,375]. We did not include any mutations into this second construct, but decided to also shorten the C-terminus which is known to be susceptible to proteolytic cleavage [117]. We are currently in the process of cloning these new constructs. Upcoming experiments will assess the feasibility of pentamer production and establish a purification protocol based on that established for *E.coli* expression. Long-term goals include the formation of a complex between purified HAdV-D09pb and $\alpha_v\beta_3$ integrin in order to produce a homogeneous sample for single particle EM and small-angle X-Ray scattering (SAXS) analysis. Furthermore, size exclusion chromatography will provide quantitative information about the preferred stoichiometry of the complex.

Unpublished Manuscript linked to this work:

Purification of pentameric HAdV-D09 penton base

A. Manuel Liaci, Karolina Cupelli, Paul Bachmann, York Stierhof, and Thilo Stehle

Contributions: The contributions are described in detail in the manuscript (see **Chapter 6**).

4.2. The Adenoviral E4ORF1 Protein - Structural and Biophysical Studies on a Viral Powerhouse Protein

Despite its small size of only about 125 amino acids, the HAdV E4ORF1 protein (E4ORF1p) exerts manifold functions that stimulate the host cell metabolism and deregulate cell polarity (see **Chapter 1.8.3**). The protein is distantly related to the homotrimeric human cellular dUTPase with a sequence identity of about 20%. Interestingly, E4ORF1p has been reported to carry out its different functions in two different oligomerization states: As a monomer, E4ORF1p of HAdV-D09 associates with the PDZ domains of tight-junction associated scaffolding proteins such as MAGI-1 and MUPP-1, while the trimeric form of most types forms a ternary complex with Dlg-1 and PI3K, which is recruited to the plasma membrane and stimulates PIP_3 signaling processes that for example stimulate MYC [376]. Although E4ORF1p is a cytosolic protein, the gene product of HAdV-C05 is thought to also enter the nucleus and interact directly with MYC [271]. Due to these effects, E4ORF1p is a potential druggable target, and it is currently under investigation as a potential diabetes drug itself [377]. However, its structure and interaction interfaces are largely unexplored so far. The remarkable functional versatility and its potential clinical and therapeutical relevance have prompted us to investigate the biophysical and structural properties of E4ORF1p from different HAdV species.

With the exception of species F, all HAdVs possess E4ORF1 genes. Interestingly, the E4ORF1p sequence is highly homologous within each species (83-100% sequence identity), but differs remarkably between different species (38-70%) (see **Appendix 6.2.4**). To investigate species-specific differences, we have selected E4ORF1ps from HAdV types A12, B03, C05, D36, E04, and G52 as representatives for each species. Most of the proteins possess a remarkably high percentage of hydrophobic residues, and grand average of hydropathy (GRAVY) indices that are paralleled only by membrane proteins [378]. Because of the small size of the proteins, we have initially produced full-length E4ORF1p constructs with an N-terminal His tag and a TEV cleavage site. These constructs are generally poorly expressed by *E. coli* regardless of the expression strain we tested. The best solubilization was achieved for HAdV-C05 E4ORF1p using a high-pH buffer containing 100 mM triethanolamine (pH 8.5) and 50 mM LiCl.

In order to increase the expression level and solubility of the protein, we designed HAdV-C05 E4ORF1p constructs with the larger N-terminal tags SUMO and maltose binding protein (MBP), respectively. Both constructs gave significantly higher expression levels after overnight expression, but the recombinantly expressed proteins were almost completely insoluble using the established ethanolamine buffer. However, adding low concentrations of N-lauroylsarcosine (NLS, 0.2% w/v) helped to solubilize about 50% of the protein when applied with high dilutions of about 50 mL per mg cell pellet. In contrast, other detergents such as Triton X-100, Tween 20, or deoxycholate did not solubilize the protein. The MBP-tagged protein exhibited poorly reproducible immobilization behavior. SUMO-tagged HAdV-C05 E4ORF1, in contrast, is readily purifiable by nickel affinity chromatography in the presence of 0.2% NLS, which is well below the critical micelle concentration. The detergent could be effectively removed by several dialysis steps and extensive washing of immobilized E4ORF1p on a nickel column, as assessed by LC-MS analysis. However, analytical gel filtration runs and dynamic light scattering experiments consistently indicated that the SUMO-tagged protein forms large assemblies of about 500-600 kDa, which corresponds to about 16-20 E4ORF1 proteins. Negative-stain electron microscopy revealed that these assembled particles indeed are of uniform size, but display heterogeneous shapes and no recognizable superstructure. We were able to partially cleave off the SUMO tag using the Ulp1 protease, which recognizes the tertiary structure of correctly folded SUMO. From these experiments, we concluded that SUMO-tagged HADV-C05 E4ORF1p is poorly folded and prone to form aggregates which are held in solution by a coat of folded SUMO tags, thereby forming the observed heterogeneous assemblies, and that these assemblies likely do not represent a physiological function of the protein.

Therefore, the rest of this study focuses on the His-tagged HAdV-C05 E4ORF1p construct, despite the much lower expression levels. Initially, the bulk of the proteins was found in insoluble inclusion bodies, and the yields of

purification were very low (<0.5 mg from 6L culture). We were able to optimize the cell lysis and to establish a purification protocol consisting of an initial nickel affinity chromatography (His-Trap) step with low imidazole concentrations followed by a digest with TEV protease at a concentration ratio of 1:10 (protease : protein), a second His-Trap to remove the tag and protease, and a final gel filtration run. By scaling up the expression, we were able to minimize unspecific losses at every purification step. The final purification protocol resulted in a yield of 8 mg protein from 18L bacterial cell culture.

HAdV-C05 E4ORF1p eluted at about 50-120 kDa in SEC experiments depending on the column, suggesting that it forms stable oligomers. The size range is consistent with dynamic light scattering experiments. Differential scanning fluorimetry experiments indicated a surprisingly high thermal stablility of the complex (T_m 73-76°C) and a single melting point with a long melting phase starting at about 60°C. This is indicative for a naturally stable quaternary arrangement that cannot be readily dissolved to form monomers. Other publications use stringent detergent concentrations to solubilize the protein, and suggest that deoxycholine (DOC) might mimic a cellular cofactor that confers monomerization [376]. However, our results with SUMO- and MBP-tagged protein suggest that the presence of these detergents might actually result in unfolding of E4ORF1p.

Crystallization trials showed that E4ORF1p readily crystallizes at a wide concentration range (1.4-10 mg/mL). Nineteen out of 96 conditions of an initial JCSG+ grid screen (Qiagen) gave crystals of highly similar morphology that diffracted to about 3.5 Å at a synchrotron beamline (X06DA, Swiss Light Source) when illuminated for about 10s per degree at full beam transmission. We are currently evaluating the reproducibility of these crystals and will improve their diffraction properties in the near future. Further experiments will explore whether or not the structure of HAdV-C05 E4ORF1 can be solved with molecular replacement using several dUTPase homologues as models. An assessment of the β-sheet content by circular dichroism (CD) spectroscopy will serve as a basis to judge the suitability of such an approach. Alternatively, we will try to take advantage of the high percentage of methionines in the protein (7 out of 128 residues or 5.5%, plus one cysteine), which meets the criteria for solving the phase problem by experimental techniques using either selenomethionine (about one residue is needed per 75-100 amino acids) or sulfur single anomalous diffraction (SAD, about 5% cysteine or methionine content), or a combination thereof if the achievable resolution allows for it [379,380]. In parallel, we will assess the binding stoichiometry in solution by SAXS experiments.

In summary, we present initial data indicating that HAdV-C05 E4ORF1 forms stable oligomers and is unlikely to persist as a monomer. We further present a working protocol for the purification and crystallization of the protein. The crystal structure will serve as a basis to assess the oligomerization state and the energy needed to disrupt the oligomer interfaces. Further experiments will establish the purification and crystallization of E4ORF1 from other HAdV species, and address the complex formation of the protein with its reported ligands in both biophysical and structural experiments.

Unpublished Manuscript linked to this work:

The Adenoviral E4ORF1 Protein - Structural Studies on a Viral Powerhouse Protein

A. Manuel Liaci, Laura Hehl, Antonia Lott, York Stierhof, and Thilo Stehle

Contributions: The contributions are described in detail in the manuscript (see **Chapter 6**).

Bachelor Theses linked to this work:

Cloning, Expression and Purification of E4ORF1 gene constructs from human Adenoviruses

Bachelor Thesis of Antonia Lott. Supervisors: **A. Manuel Liaci**, Thilo Stehle.

Characterization of the Human Adenovirus protein E4-ORF1

Bachelor Thesis of Laura Hehl. Supervisors: **A. Manuel Liaci**, Thilo Stehle.

5. REFERENCES

1. Svennerholm L (1963) Chromatographic Separation of Human Brain Gangliosides. J Neurochem 10: 613-623.
2. Flint SJ, Enquist, L. W., Racianello, V. R., Skalka, A. M. (2004) Foundations. Principles of Virology: Molecular Biology, Pathogenesis, and Control of Animal Viruses. 2 ed. Washington: ASM Press. pp. 3-23.
3. de Parseval N, Lazar V, Casella JF, Benit L, Heidmann T (2003) Survey of human genes of retroviral origin: identification and transcriptome of the genes with coding capacity for complete envelope proteins. J Virol 77: 10414-10422.
4. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409: 860-921.
5. Stagl V, Sattmann H, Horweg C (2010) [Dreaded miasmas: Austrian scientists on board the frigate Novara pursue the secret of malaria]. Wien Klin Wochenschr 122 Suppl 3: 6-9.
6. Gross CP, Sepkowitz KA (1998) The myth of the medical breakthrough: smallpox, vaccination, and Jenner reconsidered. Int J Infect Dis 3: 54-60.
7. Willis NJ (1997) Edward Jenner and the eradication of smallpox. Scott Med J 42: 118-121.
8. Riedel S (2005) Edward Jenner and the history of smallpox and vaccination. Proc (Bayl Univ Med Cent) 18: 21-25.
9. Lombard M, Pastoret PP, Moulin AM (2007) A brief history of vaccines and vaccination. Rev Sci Tech 26: 29-48.
10. Sakula A (1983) Robert koch: centenary of the discovery of the tubercle bacillus, 1882. Can Vet J 24: 127-131.
11. International Committe on Taxonomy of Viruses (2016) Virus Taxonomy: 2015 Release.
12. Baltimore D (1971) Expression of animal virus genomes. Bacteriol Rev 35: 235-241.
13. Treisman R, Cowie A, Favaloro J, Jat P, Kamen R (1981) The structures of the spliced mRNAs encoding polyoma virus early region proteins. J Mol Appl Genet 1: 83-92.
14. Buck CB, Van Doorslaer K, Peretti A, Geoghegan EM, Tisza MJ, et al. (2016) The Ancient Evolutionary History of Polyomaviruses. PLoS Pathog 12: e1005574.
15. Johne R, Buck CB, Allander T, Atwood WJ, Garcea RL, et al. (2011) Taxonomical developments in the family Polyomaviridae. Arch Virol 156: 1627-1634.
16. Padgett BL, Walker DL, ZuRhein GM, Eckroade RJ, Dessel BH (1971) Cultivation of papova-like virus from human brain with progressive multifocal leucoencephalopathy. Lancet 1: 1257-1260.
17. Gardner SD, Field AM, Coleman DV, Hulme B (1971) New human papovavirus (B.K.) isolated from urine after renal transplantation. Lancet 1: 1253-1257.
18. Sweet BH, Hilleman MR (1960) The vacuolating virus, S.V. 40. Proc Soc Exp Biol Med 105: 420-427.
19. Shah K, Nathanson N (1976) Human exposure to SV40: review and comment. Am J Epidemiol 103: 1-12.
20. Allander T, Andreasson K, Gupta S, Bjerkner A, Bogdanovic G, et al. (2007) Identification of a third human polyomavirus. J Virol 81: 4130-4136.
21. Gaynor AM, Nissen MD, Whiley DM, Mackay IM, Lambert SB, et al. (2007) Identification of a novel polyomavirus from patients with acute respiratory tract infections. PLoS Pathog 3: e64.
22. Feng H, Shuda M, Chang Y, Moore PS (2008) Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science 319: 1096-1100.
23. Schowalter RM, Pastrana DV, Pumphrey KA, Moyer AL, Buck CB (2010) Merkel cell polyomavirus and two previously unknown polyomaviruses are chronically shed from human skin. Cell Host Microbe 7: 509-515.
24. van der Meijden E, Janssens RW, Lauber C, Bouwes Bavinck JN, Gorbalenya AE, et al. (2010) Discovery of a new human polyomavirus associated with trichodysplasia spinulosa in an immunocompromized patient. PLoS Pathog 6: e1001024.
25. Scuda N, Hofmann J, Calvignac-Spencer S, Ruprecht K, Liman P, et al. (2011) A novel human polyomavirus closely related to the african green monkey-derived lymphotropic polyomavirus. J Virol 85: 4586-4590.
26. Siebrasse EA, Reyes A, Lim ES, Zhao G, Mkakosya RS, et al. (2012) Identification of MW polyomavirus, a novel polyomavirus in human stool. J Virol 86: 10321-10326.
27. Lim ES, Reyes A, Antonio M, Saha D, Ikumapayi UN, et al. (2013) Discovery of STL polyomavirus, a polyomavirus of ancestral recombinant origin that encodes a unique T antigen by alternative splicing. Virology 436: 295-303.
28. Korup S, Rietscher J, Calvignac-Spencer S, Trusch F, Hofmann J, et al. (2013) Identification of a novel human polyomavirus in organs of the gastrointestinal tract. PLoS One 8: e58021.
29. Mishra N, Pereira M, Rhodes RH, An P, Pipas JM, et al. (2014) Identification of a novel polyomavirus in a pancreatic transplant recipient with retinal blindness and vasculitic myopathy. J Infect Dis 210: 1595-1599.
30. Eddy BE, Stewart SE, Young R, Mider GB (1958) Neoplasms in hamsters induced by mouse tumor agent passed in tissue culture. J Natl Cancer Inst 20: 747-761.
31. Stewart SE, Eddy BE, Borgese N (1958) Neoplasms in mice inoculated with a tumor agent carried in tissue culture. J Natl Cancer Inst 20: 1223-1243.
32. Knowles WA (2006) Discovery and epidemiology of the human polyomaviruses BK virus (BKV) and JC virus (JCV). Adv Exp Med Biol 577: 19-45.
33. Bogdanovic G, Priftakis P, Giraud G, Kuzniar M, Ferraldeschi R, et al. (2004) Association between a high BK virus load in urine samples of patients with graft-versus-host disease and development of hemorrhagic cystitis after hematopoietic stem cell transplantation. J Clin Microbiol 42: 5394-5396.
34. Randhawa P, Vats A, Shapiro R (2006) The pathobiology of polyomavirus infection in man. Adv Exp Med Biol 577: 148-159.

35. Shishido-Hara Y (2010) Progressive multifocal leukoencephalopathy and promyelocytic leukemia nuclear bodies: a review of clinical, neuropathological, and virological aspects of JC virus-induced demyelinating disease. Acta Neuropathol 120: 403-417.
36. Kazem S, van der Meijden E, Kooijman S, Rosenberg AS, Hughey LC, et al. (2012) Trichodysplasia spinulosa is characterized by active polyomavirus infection. J Clin Virol 53: 225-230.
37. Gjoerup O, Chang Y (2010) Update on human polyomaviruses and cancer. Adv Cancer Res 106: 1-51.
38. Feltkamp MC, Kazem S, van der Meijden E, Lauber C, Gorbalenya AE (2013) From Stockholm to Malawi: recent developments in studying human polyomaviruses. J Gen Virol 94: 482-496.
39. Dalianis T, Hirsch HH (2013) Human polyomaviruses in disease and cancer. Virology 437: 63-72.
40. 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.
41. Freund R, Mandel G, Carmichael GG, Barncastle JP, Dawe CJ, et al. (1987) Polyomavirus tumor induction in mice: influences of viral coding and noncoding sequences on tumor profiles. J Virol 61: 2232-2239.
42. Feunteun J, Sompayrac L, Fluck M, Benjamin T (1976) Localization of gene functions in polyoma virus DNA. Proc Natl Acad Sci U S A 73: 4169-4173.
43. Carroll J, Dey D, Kreisman L, Velupillai P, Dahl J, et al. (2007) Receptor-binding and oncogenic properties of polyoma viruses isolated from feral mice. PLoS Pathog 3: e179.
44. Dawe CJ, Law LW, Dunn TB (1959) Studies of parotid-tumor agent in cultures of leukemic tissues of mice. J Natl Cancer Inst 23: 717-797.
45. Gross L (1951) "Spontaneous" leukemia developing in C3H mice following inoculation in infancy, with AK-leukemic extracts, or AK-embrvos. Proc Soc Exp Biol Med 76: 27-32.
46. Main JH, Dawe CJ (1966) Tumor induction in transplanted tooth buds infected with polyoma virus. J Natl Cancer Inst 36: 1121-1136.
47. Rowe WP, Hartley JW, Estes JD, Huebner RJ (1959) Studies of mouse polyoma virus infection. 1. Procedures for quantitation and detection of virus. J Exp Med 109: 379-391.
48. Bolen JB, Fisher SE, Chowdhury K, Shan TC, Williams JE, et al. (1985) A determinant of polyomavirus virulence enhances virus growth in cells of renal origin. J Virol 53: 335-339.
49. Bauer PH, Cui C, Liu WR, Stehle T, Harrison SC, et al. (1999) Discrimination between sialic acid-containing receptors and pseudoreceptors regulates polyomavirus spread in the mouse. J Virol 73: 5826-5832.
50. Freund R, Garcea RL, Sahli R, Benjamin TL (1991) A single-amino-acid substitution in polyomavirus VP1 correlates with plaque size and hemagglutination behavior. J Virol 65: 350-355.
51. Dawe CJ, Freund R, Mandel G, Ballmer-Hofer K, Talmage DA, et al. (1987) Variations in polyoma virus genotype in relation to tumor induction in mice. Characterization of wild type strains with widely differing tumor profiles. Am J Pathol 127: 243-261.
52. Bauer PH, Bronson RT, Fung SC, Freund R, Stehle T, et al. (1995) Genetic and structural analysis of a virulence determinant in polyomavirus VP1. J Virol 69: 7925-7931.
53. Freund R, Calderone A, Dawe CJ, Benjamin TL (1991) Polyomavirus tumor induction in mice: effects of polymorphisms of VP1 and large T antigen. J Virol 65: 335-341.
54. Berk AJ (2013) Adenoviridae. In: Knipe DM, Howley, P. M., editor. Fields Virology. 6 ed. Philadelphia: LIPINCOTT WILLIAMS & WILKINS. pp. 1704-1731.
55. Davison AJ, Benko M, Harrach B (2003) Genetic content and evolution of adenoviruses. J Gen Virol 84: 2895-2908.
56. Wiethoff CM, Nemerow GR (2015) Adenovirus membrane penetration: Tickling the tail of a sleeping dragon. Virology 479-480: 591-599.
57. Kremer EJ, Nemerow GR (2015) Adenovirus tales: from the cell surface to the nuclear pore complex. PLoS Pathog 11: e1004821.
58. Ismail AM, Lee J, Dyer DW, Seto D, Rajaiya J, et al. (2016) Selection pressure in the human adenovirus fiber knob drives cell specificity in epidemic keratoconjunctivitis. J Virol.
59. Aoki K, Benko M, Davison AJ, Echavarria M, Erdman DD, et al. (2011) Toward an integrated human adenovirus designation system that utilizes molecular and serological data and serves both clinical and fundamental virology. J Virol 85: 5703-5704.
60. de Jong JC, Osterhaus AD, Jones MS, Harrach B (2008) Human adenovirus type 52: a type 41 in disguise? J Virol 82: 3809; author reply 3809-3810.
61. Seto D, Chodosh J, Brister JR, Jones MS, Members of the Adenovirus Research C (2011) Using the whole-genome sequence to characterize and name human adenoviruses. J Virol 85: 5701-5702.
62. Human Adenovirus Working Group Criteria for a New HAdV Type.
63. Musher DM (2003) How contagious are common respiratory tract infections? N Engl J Med 348: 1256-1266.
64. Flomenberg P (2009) Adenovirus infections. Medicine 37: 676-678.
65. Lion T (2014) Adenovirus infections in immunocompetent and immunocompromised patients. Clin Microbiol Rev 27: 441-462.
66. Wold WSM, Horwitz MS (2013) Adenoviruses. In: Knipe DM, Howley, P. M., editor. Fields Virology. Philadelphia: LIPINCOTT WILLIAMS & WILKINS. pp. 1732-1767.
67. Thomas MA, Spencer JF, Wold WS (2007) Use of the Syrian hamster as an animal model for oncolytic adenovirus vectors. Methods Mol Med 130: 169-183.
68. Cormier Z (2011) Respiratory virus jumps from monkeys to humans. Nature Publishing Group.

69. Kramer A, Schwebke I, Kampf G (2006) How long do nosocomial pathogens persist on inanimate surfaces? A systematic review. BMC Infect Dis 6: 130.
70. Prince HN, Prince, D. L., Prince, R. N. (1991) Principles of Viral Control and Transmission. In: Block SS, editor. Disinfection, Sterilization, and Preservation. Malvern, PA: Lea & Febiger. pp. 411-444.
71. Kowalski RP, Romanowski EG, Waikhom B, Gordon YJ (1998) The survival of adenovirus in multidose bottles of topical fluorescein. Am J Ophthalmol 126: 835-836.
72. Trentin JJ (1962) Immunological tolerance--fact or fiction? South Med J 55: 878-881.
73. Trentin JJ, Yabe Y, Taylor G (1962) The quest for human cancer viruses. Science 137: 835-841.
74. Yu B, Zhou Y, Wu H, Wang Z, Zhan Y, et al. (2012) Seroprevalence of neutralizing antibodies to human adenovirus type 5 in healthy adults in China. J Med Virol 84: 1408-1414.
75. Ponterio E, Gnessi L (2015) Adenovirus 36 and Obesity: An Overview. Viruses 7: 3719-3740.
76. Nilsson EC, Storm RJ, Bauer J, Johansson SM, Lookene A, et al. (2011) The GD1a glycan is a cellular receptor for adenoviruses causing epidemic keratoconjunctivitis. Nat Med 17: 105-109.
77. Banyai K, Martella V, Meleg E, Kisfali P, Peterfi Z, et al. (2009) Searching for HAdV-52, the putative gastroenteritis-associated human adenovirus serotype in Southern Hungary. New Microbiol 32: 185-188.
78. Vogels R, Zuijdgeest D, van Rijnsoever R, Hartkoorn E, Damen I, et al. (2003) Replication-deficient human adenovirus type 35 vectors for gene transfer and vaccination: efficient human cell infection and bypass of preexisting adenovirus immunity. J Virol 77: 8263-8271.
79. Turner MA, Middha S, Hofherr SE, Barry MA (2015) Comparison of the Life Cycles of Genetically Distant Species C and Species D Human Adenoviruses Ad6 and Ad26 in Human Cells. J Virol 89: 12401-12417.
80. Zhou X, Robinson CM, Rajaiya J, Dehghan S, Seto D, et al. (2012) Analysis of human adenovirus type 19 associated with epidemic keratoconjunctivitis and its reclassification as adenovirus type 64. Invest Ophthalmol Vis Sci 53: 2804-2811.
81. Bialasiewicz A (2007) Adenoviral keratoconjunctivitis. Sultan Qaboos Univ Med J 7: 15-23.
82. Gonzalez-Lopez JJ, Morcillo-Laiz R, Munoz-Negrete FJ (2013) Adenoviral keratoconjunctivitis: an update. Arch Soc Esp Oftalmol 88: 108-115.
83. Jhanji V, Chan TC, Li EY, Agarwal K, Vajpayee RB (2015) Adenoviral keratoconjunctivitis. Surv Ophthalmol 60: 435-443.
84. Wigand R, Gelderblom H, Wadell G (1980) New human adenovirus (candidate adenovirus 36), a novel member of subgroup D. Arch Virol 64: 225-233.
85. Dhurandhar NV, Israel BA, Kolesar JM, Mayhew GF, Cook ME, et al. (2000) Increased adiposity in animals due to a human virus. Int J Obes Relat Metab Disord 24: 989-996.
86. Ponterio E, Cangemi R, Mariani S, Casella G, De Cesare A, et al. (2015) Adenovirus 36 DNA in human adipose tissue. Int J Obes (Lond) 39: 1761-1764.
87. Parra-Rojas I, Del Moral-Hernandez O, Salgado-Bernabe AB, Guzman-Guzman IP, Salgado-Goytia L, et al. (2013) Adenovirus-36 seropositivity and its relation with obesity and metabolic profile in children. Int J Endocrinol 2013: 463194.
88. So PW, Herlihy AH, Bell JD (2005) Adiposity induced by adenovirus 5 inoculation. Int J Obes (Lond) 29: 603-606.
89. Whigham LD, Israel BA, Atkinson RL (2006) Adipogenic potential of multiple human adenoviruses in vivo and in vitro in animals. Am J Physiol Regul Integr Comp Physiol 290: R190-194.
90. Dhurandhar NV, Whigham LD, Abbott DH, Schultz-Darken NJ, Israel BA, et al. (2002) Human adenovirus Ad-36 promotes weight gain in male rhesus and marmoset monkeys. J Nutr 132: 3155-3160.
91. Jones MS, 2nd, Harrach B, Ganac RD, Gozum MM, Dela Cruz WP, et al. (2007) New adenovirus species found in a patient presenting with gastroenteritis. J Virol 81: 5978-5984.
92. Flint SJ, Enquist, L. W., Racianello, V. R., Skalka, A. M. (2004) Structure. Principles of Virology: Molecular Biology, Pathogenesis, and Control of Animal Viruses. Washington: ASM Press. pp. 83-124.
93. Marsh M, Helenius A (2006) Virus entry: open sesame. Cell 124: 729-740.
94. Nermut MV, Hockley DJ, Jowett JB, Jones IM, Garreau M, et al. (1994) Fullerene-like organization of HIV gag-protein shell in virus-like particles produced by recombinant baculovirus. Virology 198: 288-296.
95. Ganser BK, Li S, Klishko VY, Finch JT, Sundquist WI (1999) Assembly and analysis of conical models for the HIV-1 core. Science 283: 80-83.
96. Mattei S, Glass B, Hagen WJ, Krausslich HG, Briggs JA (2016) The structure and flexibility of conical HIV-1 capsids determined within intact virions. Science 354: 1434-1437.
97. Caspar DL, Klug A (1962) Physical principles in the construction of regular viruses. Cold Spring Harb Symp Quant Biol 27: 1-24.
98. Carrillo-Tripp M, Shepherd CM, Borelli IA, Venkataraman S, Lander G, et al. (2009) VIPERdb2: an enhanced and web API enabled relational database for structural virology. Nucleic Acids Res 37: D436-442.
99. Krupovic M, Bamford DH (2008) Virus evolution: how far does the double beta-barrel viral lineage extend? Nat Rev Microbiol 6: 941-948.
100. Benson SD, Bamford JK, Bamford DH, Burnett RM (1999) Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures. Cell 98: 825-833.
101. Hendrix RW (1999) Evolution: the long evolutionary reach of viruses. Curr Biol 9: R914-917.
102. Stehle T, Yan Y, Benjamin TL, Harrison SC (1994) Structure of murine polyomavirus complexed with an oligosaccharide receptor fragment. Nature 369: 160-163.
103. Liddington RC, Yan Y, Moulai J, Sahli R, Benjamin TL, et al. (1991) Structure of simian virus 40 at 3.8-A resolution. Nature 354: 278-284.

104. Li Z, Yan X, Yu H, Wang D, Song S, et al. (2016) The C-Terminal Arm of the Human Papillomavirus Major Capsid Protein Is Immunogenic and Involved in Virus-Host Interaction. Structure 24: 874-885.
105. Stehle T, Harrison SC (1997) High-resolution structure of a polyomavirus VP1-oligosaccharide complex: implications for assembly and receptor binding. EMBO J 16: 5139-5148.
106. Barouch DH, Harrison SC (1994) Interactions among the major and minor coat proteins of polyomavirus. J Virol 68: 3982-3989.
107. Chen XS, Stehle T, Harrison SC (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.
108. Stewart PL, Burnett RM, Cyrklaff M, Fuller SD (1991) Image reconstruction reveals the complex molecular organization of adenovirus. Cell 67: 145-154.
109. Nemerow GR, Stewart PL, Reddy VS (2012) Structure of human adenovirus. Current opinion in virology 2: 115-121.
110. San Martin C (2012) Latest insights on adenovirus structure and assembly. Viruses 4: 847-877.
111. Liu H, Jin L, Koh SB, Atanasov I, Schein S, et al. (2010) Atomic structure of human adenovirus by cryo-EM reveals interactions among protein networks. Science 329: 1038-1043.
112. Fabry CM, Rosa-Calatrava M, Conway JF, Zubieta C, Cusack S, et al. (2005) A quasi-atomic model of human adenovirus type 5 capsid. The EMBO journal 24: 1645-1654.
113. Roberts MM, White JL, Grutter MG, Burnett RM (1986) Three-dimensional structure of the adenovirus major coat protein hexon. Science 232: 1148-1151.
114. Athappilly FK, Murali R, Rux JJ, Cai Z, Burnett RM (1994) The refined crystal structure of hexon, the major coat protein of adenovirus type 2, at 2.9 A resolution. J Mol Biol 242: 430-455.
115. Rux JJ, Kuser PR, Burnett RM (2003) Structural and phylogenetic analysis of adenovirus hexons by use of high-resolution x-ray crystallographic, molecular modeling, and sequence-based methods. J Virol 77: 9553-9566.
116. Szolajska E, Burmeister WP, Zochowska M, Nerlo B, Andreev I, et al. (2012) The structural basis for the integrity of adenovirus Ad3 dodecahedron. PLoS One 7: e46075.
117. Zubieta C, Schoehn G, Chroboczek J, Cusack S (2005) The structure of the human adenovirus 2 penton. Molecular cell 17: 121-135.
118. Reddy VS, Nemerow GR (2014) Structures and organization of adenovirus cement proteins provide insights into the role of capsid maturation in virus entry and infection. Proc Natl Acad Sci U S A 111: 11715-11720.
119. Mangel WF, San Martin C (2014) Structure, function and dynamics in adenovirus maturation. Viruses 6: 4536-4570.
120. Zandi R, Reguera D (2005) Mechanical properties of viral capsids. Phys Rev E Stat Nonlin Soft Matter Phys 72: 021917.
121. Perez-Berna AJ, Ortega-Esteban A, Menendez-Conejero R, Winkler DC, Menendez M, et al. (2012) The role of capsid maturation on adenovirus priming for sequential uncoating. J Biol Chem 287: 31582-31595.
122. Snijder J, Reddy VS, May ER, Roos WH, Nemerow GR, et al. (2013) Integrin and defensin modulate the mechanical properties of adenovirus. Journal of virology 87: 2756-2766.
123. van Raaij MJ, Schoehn G, Burda MR, Miller S (2001) Crystal structure of a heat and protease-stable part of the bacteriophage T4 short tail fibre. J Mol Biol 314: 1137-1146.
124. Wu E, Pache L, Von Seggern DJ, Mullen TM, Mikyas Y, et al. (2003) Flexibility of the adenovirus fiber is required for efficient receptor interaction. J Virol 77: 7225-7235.
125. Shayakhmetov DM, Lieber A (2000) Dependence of adenovirus infectivity on length of the fiber shaft domain. J Virol 74: 10274-10286.
126. Signas C, Akusjarvi G, Pettersson U (1985) Adenovirus 3 fiber polypeptide gene: implications for the structure of the fiber protein. J Virol 53: 672-678.
127. Green NM, Wrigley NG, Russell WC, Martin SR, McLachlan AD (1983) Evidence for a repeating cross-beta sheet structure in the adenovirus fibre. EMBO J 2: 1357-1365.
128. Kidd AH, Chroboczek J, Cusack S, Ruigrok RW (1993) Adenovirus type 40 virions contain two distinct fibers. Virology 192: 73-84.
129. Liu H, Wu L, Zhou ZH (2011) Model of the trimeric fiber and its interactions with the pentameric penton base of human adenovirus by cryo-electron microscopy. J Mol Biol 406: 764-774.
130. Xia D, Henry LJ, Gerard RD, Deisenhofer J (1994) Crystal structure of the receptor-binding domain of adenovirus type 5 fiber protein at 1.7 A resolution. Structure 2: 1259-1270.
131. Nilsson J, Miyazaki N, Xing L, Wu B, Hammar L, et al. (2005) Structure and assembly of a T=1 virus-like particle in BK polyomavirus. J Virol 79: 5337-5345.
132. Norrby E (1966) The relationship between the soluble antigens and the virion of adenovirus type 3. II. Identification and characterization of an incomplete hemagglutinin. Virology 30: 608-617.
133. Fender P, Boussaid A, Mezin P, Chroboczek J (2005) Synthesis, cellular localization, and quantification of penton-dodecahedron in serotype 3 adenovirus-infected cells. Virology 340: 167-173.
134. Lu ZZ, Wang H, Zhang Y, Cao H, Li Z, et al. (2013) Penton-dodecahedral particles trigger opening of intercellular junctions and facilitate viral spread during adenovirus serotype 3 infection of epithelial cells. PLoS Pathog 9: e1003718.
135. Flint SJ, Enquist, L. W., Racianello, V. R., Skalka, A. M. (2004) Attachment and Entry. Principles of Virology: Molecular Biology, Pathogenesis, and Control of Animal Viruses. Washington: ASM Press. pp. 127-180.
136. Burmeister WP, Guilligay D, Cusack S, Wadell G, Arnberg N (2004) Crystal structure of species D adenovirus fiber knobs and their sialic acid binding sites. J Virol 78: 7727-7736.
137. Seiradake E, Henaff D, Wodrich H, Billet O, Perreau M, et al. (2009) The cell adhesion molecule "CAR" and sialic acid on human erythrocytes influence adenovirus in vivo biodistribution. PLoS Pathog 5: e1000277.

138. Marttila M, Persson D, Gustafsson D, Liszewski MK, Atkinson JP, et al. (2005) CD46 is a cellular receptor for all species B adenoviruses except types 3 and 7. J Virol 79: 14429-14436.
139. Neu U, Bauer J, Stehle T (2011) Viruses and sialic acids: rules of engagement. Curr Opin Struct Biol 21: 610-618.
140. Xiong X, Coombs PJ, Martin SR, Liu J, Xiao H, et al. (2013) Receptor binding by a ferret-transmissible H5 avian influenza virus. Nature 497: 392-396.
141. Lee RT, Lee YC (2000) Affinity enhancement by multivalent lectin-carbohydrate interaction. Glycoconj J 17: 543-551.
142. Dam TK, Brewer CF (2008) Effects of clustered epitopes in multivalent ligand-receptor interactions. Biochemistry 47: 8470-8476.
143. You J, O'Hara SD, Velupillai P, Castle S, Levery S, et al. (2015) Ganglioside and Non-ganglioside Mediated Host Responses to the Mouse Polyomavirus. PLoS Pathog 11: e1005175.
144. Lutschg V, Boucke K, Hemmi S, Greber UF (2011) Chemotactic antiviral cytokines promote infectious apical entry of human adenovirus into polarized epithelial cells. Nat Commun 2: 391.
145. Coyne CB, Bergelson JM (2006) Virus-induced Abl and Fyn kinase signals permit coxsackievirus entry through epithelial tight junctions. Cell 124: 119-131.
146. Chandrasekaran A, Srinivasan A, Raman R, Viswanathan K, Raguram S, et al. (2008) Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin. Nat Biotechnol 26: 107-113.
147. Khan ZM, Liu Y, Neu U, Gilbert M, Ehlers B, et al. (2014) Crystallographic and glycan microarray analysis of human polyomavirus 9 VP1 identifies N-glycolyl neuraminic acid as a receptor candidate. J Virol 88: 6100-6111.
148. Kuss SK, Best GT, Etheredge CA, Pruijssers AJ, Frierson JM, et al. (2011) Intestinal microbiota promote enteric virus replication and systemic pathogenesis. Science 334: 249-252.
149. Jones MK, Watanabe M, Zhu S, Graves CL, Keyes LR, et al. (2014) Enteric bacteria promote human and mouse norovirus infection of B cells. Science 346: 755-759.
150. Ebert DH, Deussing J, Peters C, Dermody TS (2002) Cathepsin L and cathepsin B mediate reovirus disassembly in murine fibroblast cells. J Biol Chem 277: 24609-24617.
151. Simmons G, Zmora P, Gierer S, Heurich A, Pohlmann S (2013) Proteolytic activation of the SARS-coronavirus spike protein: cutting enzymes at the cutting edge of antiviral research. Antiviral Res 100: 605-614.
152. Zaitseva E, Yang ST, Melikov K, Pourmal S, Chernomordik LV (2010) Dengue virus ensures its fusion in late endosomes using compartment-specific lipids. PLoS Pathog 6: e1001131.
153. Hayer A, Stoeber M, Ritz D, Engel S, Meyer HH, et al. (2010) Caveolin-1 is ubiquitinated and targeted to intralumenal vesicles in endolysosomes for degradation. J Cell Biol 191: 615-629.
154. Ewers H, Romer W, Smith AE, Bacia K, Dmitrieff S, et al. (2010) GM1 structure determines SV40-induced membrane invagination and infection. Nat Cell Biol 12: 11-18; sup pp 11-12.
155. Lakadamyali M, Rust MJ, Zhuang X (2004) Endocytosis of influenza viruses. Microbes Infect 6: 929-936.
156. Wolfrum N, Greber UF (2013) Adenovirus signalling in entry. Cell Microbiol 15: 53-62.
157. Grove J, Marsh M (2011) The cell biology of receptor-mediated virus entry. J Cell Biol 195: 1071-1082.
158. Stanley P, Schachter H, Taniguchi N (2009) N-Glycans. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P et al., editors. Essentials of Glycobiology. 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.
159. Brockhausen I, Schachter H, Stanley P (2009) O-GalNAc Glycans. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P et al., editors. Essentials of Glycobiology. 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.
160. Varki A, Lowe JB (2009) Biological Roles of Glycans. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P et al., editors. Essentials of Glycobiology. 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.
161. Varki A, Gagneux P (2012) Multifarious roles of sialic acids in immunity. Ann N Y Acad Sci 1253: 16-36.
162. Vimr E, Lichtensteiger C (2002) To sialylate, or not to sialylate: that is the question. Trends Microbiol 10: 254-257.
163. Buch MH, Liaci AM, O'Hara SD, Garcea RL, Neu U, et al. (2015) Structural and Functional Analysis of Murine Polyomavirus Capsid Proteins Establish the Determinants of Ligand Recognition and Pathogenicity. PLoS Pathog 11: e1005104.
164. Wasik BR, Barnard KN, Parrish CR (2016) Effects of Sialic Acid Modifications on Virus Binding and Infection. Trends Microbiol.
165. Ströh LJ, Stehle T (2014) Glycan Engagement by Viruses: Receptor Switches and Specificity. Annu Rev Virol 1: 285-306.
166. Fiege B, Rabbani S, Preston RC, Jakob RP, Zihlmann P, et al. (2015) The tyrosine gate of the bacterial lectin FimH: a conformational analysis by NMR spectroscopy and X-ray crystallography. Chembiochem 16: 1235-1246.
167. Tangvoranuntakul P, Gagneux P, Diaz S, Bardor M, Varki N, et al. (2003) Human uptake and incorporation of an immunogenic nonhuman dietary sialic acid. Proc Natl Acad Sci U S A 100: 12045-12050.
168. Mandal C, Schwartz-Albiez R, Vlasak R (2015) Functions and Biosynthesis of O-Acetylated Sialic Acids. Top Curr Chem 366: 1-30.
169. Langereis MA, Bakkers MJ, Deng L, Padler-Karavani V, Vervoort SJ, et al. (2015) Complexity and Diversity of the Mammalian Sialome Revealed by Nidovirus Virolectins. Cell Rep 11: 1966-1978.
170. Hellebo A, Vilas U, Falk K, Vlasak R (2004) Infectious salmon anemia virus specifically binds to and hydrolyzes 4-O-acetylated sialic acids. J Virol 78: 3055-3062.
171. Baumann AM, Bakkers MJ, Buettner FF, Hartmann M, Grove M, et al. (2015) 9-O-Acetylation of sialic acids is catalysed by CASD1 via a covalent acetyl-enzyme intermediate. Nat Commun 6: 7673.
172. Varki A (2007) Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. Nature 446: 1023-1029.
173. Schauer R (2000) Achievements and challenges of sialic acid research. Glycoconj J 17: 485-499.
174. Varki A, Schauer R (2009) Sialic Acids. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P et al., editors. Essentials of Glycobiology. 2nd ed. Cold Spring Harbor (NY).

175. Stehle T, Khan ZM (2014) Rules and exceptions: sialic acid variants and their role in determining viral tropism. J Virol 88: 7696-7699.
176. Rogers GN, Herrler G, Paulson JC, Klenk HD (1986) Influenza C virus uses 9-O-acetyl-N-acetylneuraminic acid as a high affinity receptor determinant for attachment to cells. J Biol Chem 261: 5947-5951.
177. Brown DA, London E (2000) Structure and function of sphingolipid- and cholesterol-rich membrane rafts. J Biol Chem 275: 17221-17224.
178. Sonnino S, Mauri L, Chigorno V, Prinetti A (2007) Gangliosides as components of lipid membrane domains. Glycobiology 17: 1R-13R.
179. Hakomori S, Handa K, Iwabuchi K, Yamamura S, Prinetti A (1998) New insights in glycosphingolipid function: "glycosignaling domain," a cell surface assembly of glycosphingolipids with signal transducer molecules,involved in cell adhesion coupled with signaling. Glycobiology 8: xi-xix.
180. Anderson RG (1998) The caveolae membrane system. Annu Rev Biochem 67: 199-225.
181. Simons K, Toomre D (2000) Lipid rafts and signal transduction. Nat Rev Mol Cell Biol 1: 31-39.
182. Yu RK, Nakatani Y, Yanagisawa M (2009) The role of glycosphingolipid metabolism in the developing brain. J Lipid Res 50 Suppl: S440-445.
183. Schnaar RL, Suzuki A, Stanley P (2009) Glycosphingolipids. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P et al., editors. Essentials of Glycobiology. 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.
184. Yu RK, Tsai YT, Ariga T, Yanagisawa M (2011) Structures, biosynthesis, and functions of gangliosides--an overview. J Oleo Sci 60: 537-544.
185. Tsai B, Gilbert JM, Stehle T, Lencer W, Benjamin TL, et al. (2003) Gangliosides are receptors for murine polyoma virus and SV40. EMBO J 22: 4346-4355.
186. Varki NM, Varki A (2007) Diversity in cell surface sialic acid presentations: implications for biology and disease. Lab Invest 87: 851-857.
187. Rothbard JB, Brackenbury R, Cunningham BA, Edelman GM (1982) Differences in the carbohydrate structures of neural cell-adhesion molecules from adult and embryonic chicken brains. J Biol Chem 257: 11064-11069.
188. Rutishauser U, Landmesser L (1996) Polysialic acid in the vertebrate nervous system: a promoter of plasticity in cell-cell interactions. Trends Neurosci 19: 422-427.
189. Angata K, Fukuda M (2003) Polysialyltransferases: major players in polysialic acid synthesis on the neural cell adhesion molecule. Biochimie 85: 195-206.
190. Muhlenhoff M, Rollenhagen M, Werneburg S, Gerardy-Schahn R, Hildebrandt H (2013) Polysialic acid: versatile modification of NCAM, SynCAM 1 and neuropilin-2. Neurochem Res 38: 1134-1143.
191. Amoureux MC, Coulibaly B, Chinot O, Loundou A, Metellus P, et al. (2010) Polysialic acid neural cell adhesion molecule (PSA-NCAM) is an adverse prognosis factor in glioblastoma, and regulates olig2 expression in glioma cell lines. BMC Cancer 10: 91.
192. Tanaka F, Otake Y, Nakagawa T, Kawano Y, Miyahara R, et al. (2000) Expression of polysialic acid and STX, a human polysialyltransferase, is correlated with tumor progression in non-small cell lung cancer. Cancer Res 60: 3072-3080.
193. Tanaka F, Otake Y, Nakagawa T, Kawano Y, Miyahara R, et al. (2001) Prognostic significance of polysialic acid expression in resected non-small cell lung cancer. Cancer Res 61: 1666-1670.
194. Suzuki M, Suzuki M, Nakayama J, Suzuki A, Angata K, et al. (2005) Polysialic acid facilitates tumor invasion by glioma cells. Glycobiology 15: 887-894.
195. Petridis AK, Wedderkopp H, Hugo HH, Maximilian Mehdorn H (2009) Polysialic acid overexpression in malignant astrocytomas. Acta Neurochir (Wien) 151: 601-603; discussion 603-604.
196. Lantuejoul S, Moro D, Michalides RJ, Brambilla C, Brambilla E (1998) Neural cell adhesion molecules (NCAM) and NCAM-PSA expression in neuroendocrine lung tumors. Am J Surg Pathol 22: 1267-1276.
197. Gluer S, Schelp C, Gerardy-Schahn R, von Schweinitz D (1998) Polysialylated neural cell adhesion molecule as a marker for differential diagnosis in pediatric tumors. J Pediatr Surg 33: 1516-1520.
198. Figarella-Branger DF, Durbec PL, Rougon GN (1990) Differential spectrum of expression of neural cell adhesion molecule isoforms and L1 adhesion molecules on human neuroectodermal tumors. Cancer Res 50: 6364-6370.
199. Falconer RA, Errington RJ, Shnyder SD, Smith PJ, Patterson LH (2012) Polysialyltransferase: a new target in metastatic cancer. Curr Cancer Drug Targets 12: 925-939.
200. Cahan LD, Singh R, Paulson JC (1983) Sialyloligosaccharide receptors of binding variants of polyoma virus. Virology 130: 281-289.
201. Fried H, Cahan LD, Paulson JC (1981) Polyoma virus recognizes specific sialyligosaccharide receptors on host cells. Virology 109: 188-192.
202. Neu U, Maginnis MS, Palma AS, Ströh LJ, Nelson CD, et al. (2010) Structure-function analysis of the human JC polyomavirus establishes the LSTc pentasaccharide as a functional receptor motif. Cell Host Microbe 8: 309-319.
203. Elphick GF, Querbes W, Jordan JA, Gee GV, Eash S, et al. (2004) The human polyomavirus, JCV, uses serotonin receptors to infect cells. Science 306: 1380-1383.
204. Ströh LJ, Maginnis MS, Blaum BS, Nelson CD, Neu U, et al. (2015) The Greater Affinity of JC Polyomavirus Capsid for alpha2,6-Linked Lactoseries Tetrasaccharide c than for Other Sialylated Glycans Is a Major Determinant of Infectivity. J Virol 89: 6364-6375.
205. Ströh LJ, Gee GV, Blaum BS, Dugan AS, Feltkamp MC, et al. (2015) Trichodysplasia spinulosa-Associated Polyomavirus Uses a Displaced Binding Site on VP1 to Engage Sialylated Glycolipids. PLoS Pathog 11: e1005112.

206. Maginnis MS, Ströh LJ, Gee GV, O'Hara BA, Derdowski A, et al. (2013) Progressive multifocal leukoencephalopathy-associated mutations in the JC polyomavirus capsid disrupt lactoseries tetrasaccharide c binding. MBio 4: e00247-00213.
207. Haley SA, Atwood WJ (2014) An animal model for progressive multifocal leukoencephalopathy. J Clin Invest 124: 5103-5106.
208. Ray U, Cinque P, Gerevini S, Longo V, Lazzarin A, et al. (2015) JC polyomavirus mutants escape antibody-mediated neutralization. Sci Transl Med 7: 306ra151.
209. Jelcic I, Combaluzier B, Jelcic I, Faigle W, Senn L, et al. (2015) Broadly neutralizing human monoclonal JC polyomavirus VP1-specific antibodies as candidate therapeutics for progressive multifocal leukoencephalopathy. Sci Transl Med 7: 306ra150.
210. Gilbert J, Benjamin T (2004) Uptake pathway of polyomavirus via ganglioside GD1a. J Virol 78: 12259-12267.
211. Gilbert JM, Benjamin TL (2000) Early steps of polyomavirus entry into cells. J Virol 74: 8582-8588.
212. Richterova Z, Liebl D, Horak M, Palkova Z, Stokrova J, et al. (2001) Caveolae are involved in the trafficking of mouse polyomavirus virions and artificial VP1 pseudocapsids toward cell nuclei. J Virol 75: 10880-10891.
213. Caruso M, Belloni L, Sthandier O, Amati P, Garcia MI (2003) Alpha4beta1 integrin acts as a cell receptor for murine polyomavirus at the postattachment level. J Virol 77: 3913-3921.
214. Ashok A, Atwood WJ (2003) Contrasting roles of endosomal pH and the cytoskeleton in infection of human glial cells by JC virus and simian virus 40. J Virol 77: 1347-1356.
215. Streuli CH, Griffin BE (1987) Myristic acid is coupled to a structural protein of polyoma virus and SV40. Nature 326: 619-622.
216. Sahli R, Freund R, Dubensky T, Garcea R, Bronson R, et al. (1993) Defect in entry and altered pathogenicity of a polyoma virus mutant blocked in VP2 myristylation. Virology 192: 142-153.
217. Pelkmans L, Kartenbeck J, 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.
218. Magnuson B, Rainey EK, Benjamin T, Baryshev M, Mkrtchian S, et al. (2005) ERp29 triggers a conformational change in polyomavirus to stimulate membrane binding. Mol Cell 20: 289-300.
219. Lilley BN, Gilbert JM, Ploegh HL, Benjamin TL (2006) Murine polyomavirus requires the endoplasmic reticulum protein Derlin-2 to initiate infection. J Virol 80: 8739-8744.
220. Moreland RB, Garcea RL (1991) Characterization of a nuclear localization sequence in the polyomavirus capsid protein VP1. Virology 185: 513-518.
221. Chang D, Haynes JI, Jr., Brady JN, Consigli RA (1993) Identification of amino acid sequences in the polyomavirus capsid proteins that serve as nuclear localization signals. Trans Kans Acad Sci 96: 35-39.
222. Buch MHC (2016) Structural Studies of Polyomavirus and Herpesvirus Attachment Processes [Dissertation]. Tübingen: University of Tübingen.
223. Neu U, Stehle T, Atwood WJ (2009) The Polyomaviridae: Contributions of virus structure to our understanding of virus receptors and infectious entry. Virology 384: 389-399.
224. Cupelli K, Stehle T (2011) Viral attachment strategies: the many faces of adenoviruses. Curr Opin Virol 1: 84-91.
225. Arnberg N (2009) Adenovirus receptors: implications for tropism, treatment and targeting. Rev Med Virol 19: 165-178.
226. Goodsell DS, Dutta S, Zardecki C, Voigt M, Berman HM, et al. (2015) The RCSB PDB "Molecule of the Month": Inspiring a Molecular View of Biology. PLoS Biol 13: e1002140.
227. Bergelson JM, Cunningham JA, Droguett G, Kurt-Jones EA, Krithivas A, et al. (1997) Isolation of a common receptor for Coxsackie B viruses and adenoviruses 2 and 5. Science 275: 1320-1323.
228. Roelvink PW, Mi Lee G, Einfeld DA, Kovesdi I, Wickham TJ (1999) Identification of a conserved receptor-binding site on the fiber proteins of CAR-recognizing adenoviridae. Science 286: 1568-1571.
229. Tamanini A, Nicolis E, Bonizzato A, Bezzerri V, Melotti P, et al. (2006) Interaction of adenovirus type 5 fiber with the coxsackievirus and adenovirus receptor activates inflammatory response in human respiratory cells. J Virol 80: 11241-11254.
230. Veesler D, Cupelli K, Burger M, Graber P, Stehle T, et al. (2014) Single-particle EM reveals plasticity of interactions between the adenovirus penton base and integrin alphaVbeta3. Proc Natl Acad Sci U S A 111: 8815-8819.
231. Lindert S, Silvestry M, Mullen TM, Nemerow GR, Stewart PL (2009) Cryo-electron microscopy structure of an adenovirus-integrin complex indicates conformational changes in both penton base and integrin. Journal of virology 83: 11491-11501.
232. Seiradake E, Lortat-Jacob H, Billet O, Kremer EJ, Cusack S (2006) Structural and mutational analysis of human Ad37 and canine adenovirus 2 fiber heads in complex with the D1 domain of coxsackie and adenovirus receptor. J Biol Chem 281: 33704-33716.
233. Bewley MC, Springer K, Zhang YB, Freimuth P, Flanagan JM (1999) Structural analysis of the mechanism of adenovirus binding to its human cellular receptor, CAR. Science 286: 1579-1583.
234. Excoffon KJ, Gansemer ND, Mobily ME, Karp PH, Parekh KR, et al. (2010) Isoform-specific regulation and localization of the coxsackie and adenovirus receptor in human airway epithelia. PLoS One 5: e9909.
235. Meier O, Greber UF (2004) Adenovirus endocytosis. J Gene Med 6 Suppl 1: S152-163.
236. Kotha PL, Sharma P, Kolawole AO, Yan R, Alghamri MS, et al. (2015) Adenovirus entry from the apical surface of polarized epithelia is facilitated by the host innate immune response. PLoS Pathog 11: e1004696.
237. Segerman A, Atkinson JP, Marttila M, Dennerquist V, Wadell G, et al. (2003) Adenovirus type 11 uses CD46 as a cellular receptor. J Virol 77: 9183-9191.
238. Gaggar A, Shayakhmetov DM, Lieber A (2003) CD46 is a cellular receptor for group B adenoviruses. Nat Med 9: 1408-1412.

239. Wang H, Li ZY, Liu Y, Persson J, Beyer I, et al. (2011) Desmoglein 2 is a receptor for adenovirus serotypes 3, 7, 11 and 14. Nat Med 17: 96-104.
240. Arnberg N, Edlund K, Kidd AH, Wadell G (2000) Adenovirus type 37 uses sialic acid as a cellular receptor. J Virol 74: 42-48.
241. Arnberg N, Kidd AH, Edlund K, Nilsson J, Pring-Akerblom P, et al. (2002) Adenovirus type 37 binds to cell surface sialic acid through a charge-dependent interaction. Virology 302: 33-43.
242. Arnberg N, Pring-Akerblom P, Wadell G (2002) Adenovirus type 37 uses sialic acid as a cellular receptor on Chang C cells. J Virol 76: 8834-8841.
243. Nicol CG, Graham D, Miller WH, White SJ, Smith TA, et al. (2004) Effect of adenovirus serotype 5 fiber and penton modifications on in vivo tropism in rats. Mol Ther 10: 344-354.
244. Smith TA, Idamakanti N, Marshall-Neff J, Rollence ML, Wright P, et al. (2003) Receptor interactions involved in adenoviral-mediated gene delivery after systemic administration in non-human primates. Hum Gene Ther 14: 1595-1604.
245. Smith TA, Idamakanti N, Rollence ML, Marshall-Neff J, Kim J, et al. (2003) Adenovirus serotype 5 fiber shaft influences in vivo gene transfer in mice. Hum Gene Ther 14: 777-787.
246. Bayo-Puxan N, Cascallo M, Gros A, Huch M, Fillat C, et al. (2006) Role of the putative heparan sulfate glycosaminoglycan-binding site of the adenovirus type 5 fiber shaft on liver detargeting and knob-mediated retargeting. J Gen Virol 87: 2487-2495.
247. Xu Z, Qiu Q, Tian J, Smith JS, Conenello GM, et al. (2013) Coagulation factor X shields adenovirus type 5 from attack by natural antibodies and complement. Nat Med 19: 452-457.
248. Johansson C, Jonsson M, Marttila M, Persson D, Fan XL, et al. (2007) Adenoviruses use lactoferrin as a bridge for CAR-independent binding to and infection of epithelial cells. J Virol 81: 954-963.
249. Wickham TJ, Mathias P, Cheresh DA, Nemerow GR (1993) Integrins alpha v beta 3 and alpha v beta 5 promote adenovirus internalization but not virus attachment. Cell 73: 309-319.
250. Helmuth JA, Burckhardt CJ, Koumoutsakos P, Greber UF, Sbalzarini IF (2007) A novel supervised trajectory segmentation algorithm identifies distinct types of human adenovirus motion in host cells. J Struct Biol 159: 347-358.
251. Burckhardt CJ, Suomalainen M, Schoenenberger P, Boucke K, Hemmi S, et al. (2011) Drifting motions of the adenovirus receptor CAR and immobile integrins initiate virus uncoating and membrane lytic protein exposure. Cell Host Microbe 10: 105-117.
252. Hynes RO (2004) The emergence of integrins: a personal and historical perspective. Matrix Biol 23: 333-340.
253. Fu G, Wang W, Luo BH (2012) Overview: structural biology of integrins. Methods in molecular biology 757: 81-99.
254. Zhu J, Luo BH, Xiao T, Zhang C, Nishida N, et al. (2008) Structure of a complete integrin ectodomain in a physiologic resting state and activation and deactivation by applied forces. Mol Cell 32: 849-861.
255. Nakano MY, Boucke K, Suomalainen M, Stidwill RP, Greber UF (2000) The first step of adenovirus type 2 disassembly occurs at the cell surface, independently of endocytosis and escape to the cytosol. J Virol 74: 7085-7095.
256. Clark EA, King WG, Brugge JS, Symons M, Hynes RO (1998) Integrin-mediated signals regulated by members of the rho family of GTPases. J Cell Biol 142: 573-586.
257. Kalin S, Amstutz B, Gastaldelli M, Wolfrum N, Boucke K, et al. (2010) Macropinocytotic uptake and infection of human epithelial cells with species B2 adenovirus type 35. J Virol 84: 5336-5350.
258. Wiethoff CM, Wodrich H, Gerace L, Nemerow GR (2005) Adenovirus protein VI mediates membrane disruption following capsid disassembly. J Virol 79: 1992-2000.
259. Luisoni S, Suomalainen M, Boucke K, Tanner LB, Wenk MR, et al. (2015) Co-option of Membrane Wounding Enables Virus Penetration into Cells. Cell Host Microbe 18: 75-85.
260. Suomalainen M, Luisoni S, Boucke K, Bianchi S, Engel DA, et al. (2013) A direct and versatile assay measuring membrane penetration of adenovirus in single cells. J Virol 87: 12367-12379.
261. Zeng X, Carlin CR (2013) Host cell autophagy modulates early stages of adenovirus infections in airway epithelial cells. J Virol 87: 2307-2319.
262. Greber UF, Webster P, Weber J, Helenius A (1996) The role of the adenovirus protease on virus entry into cells. EMBO J 15: 1766-1777.
263. Smith JG, Silvestry M, Lindert S, Lu W, Nemerow GR, et al. (2010) Insight into the mechanisms of adenovirus capsid disassembly from studies of defensin neutralization. PLoS Pathog 6: e1000959.
264. Smith JG, Cassany A, Gerace L, Ralston R, Nemerow GR (2008) Neutralizing antibody blocks adenovirus infection by arresting microtubule-dependent cytoplasmic transport. J Virol 82: 6492-6500.
265. Swiss Institute of Bioinformatics Adenoviridae.
266. O'Shea C, Klupsch K, Choi S, Bagus B, Soria C, et al. (2005) Adenoviral proteins mimic nutrient/growth signals to activate the mTOR pathway for viral replication. EMBO J 24: 1211-1221.
267. Frese KK, Lee SS, Thomas DL, Latorre IJ, Weiss RS, et al. (2003) Selective PDZ protein-dependent stimulation of phosphatidylinositol 3-kinase by the adenovirus E4-ORF1 oncoprotein. Oncogene 22: 710-721.
268. Kong K, Kumar M, Taruishi M, Javier RT (2015) Adenovirus E4-ORF1 Dysregulates Epidermal Growth Factor and Insulin/Insulin-Like Growth Factor Receptors To Mediate Constitutive Myc Expression. J Virol 89: 10774-10785.
269. Kong K, Kumar M, Taruishi M, Javier RT (2014) The human adenovirus E4-ORF1 protein subverts discs large 1 to mediate membrane recruitment and dysregulation of phosphatidylinositol 3-kinase. PLoS Pathog 10: e1004102.
270. Miyake-Stoner SJ, O'Shea CC (2014) Metabolism goes viral. Cell Metab 19: 549-550.
271. Thai M, Graham NA, Braas D, Nehil M, Komisopoulou E, et al. (2014) Adenovirus E4ORF1-induced MYC activation promotes host cell anabolic glucose metabolism and virus replication. Cell Metab 19: 694-701.

272. Latorre IJ, Roh MH, Frese KK, Weiss RS, Margolis B, et al. (2005) Viral oncoprotein-induced mislocalization of select PDZ proteins disrupts tight junctions and causes polarity defects in epithelial cells. J Cell Sci 118: 4283-4293.
273. Grand RJ, Owen D, Rookes SM, Gallimore PH (1996) Control of p53 expression by adenovirus 12 early region 1A and early region 1B 54K proteins. Virology 218: 23-34.
274. Schwartz RA, Lakdawala SS, Eshleman HD, Russell MR, Carson CT, et al. (2008) Distinct requirements of adenovirus E1b55K protein for degradation of cellular substrates. J Virol 82: 9043-9055.
275. Cuconati A, White E (2002) Viral homologs of BCL-2: role of apoptosis in the regulation of virus infection. Genes Dev 16: 2465-2478.
276. Polster BM, Pevsner J, Hardwick JM (2004) Viral Bcl-2 homologs and their role in virus replication and associated diseases. Biochim Biophys Acta 1644: 211-227.
277. Querido E, Blanchette P, Yan Q, Kamura T, Morrison M, et al. (2001) Degradation of p53 by adenovirus E4orf6 and E1B55K proteins occurs via a novel mechanism involving a Cullin-containing complex. Genes Dev 15: 3104-3117.
278. Toth K, Dhar D, Wold WS (2010) Oncolytic (replication-competent) adenoviruses as anticancer agents. Expert Opin Biol Ther 10: 353-368.
279. Buijs PR, Verhagen JH, van Eijck CH, van den Hoogen BG (2015) Oncolytic viruses: From bench to bedside with a focus on safety. Hum Vaccin Immunother: 0.
280. Jiang H, Gomez-Manzano C, Rivera-Molina Y, Lang FF, Conrad CA, et al. (2015) Oncolytic adenovirus research evolution: from cell-cycle checkpoints to immune checkpoints. Curr Opin Virol 13: 33-39.
281. Doronin K, Shayakhmetov DM (2012) Construction of targeted and armed oncolytic adenoviruses. Methods Mol Biol 797: 35-52.
282. Fu J, Bouvier M (2011) Determinants of the endoplasmic reticulum (ER) lumenal-domain of the adenovirus serotype 2 E3-19K protein for association with and ER-retention of major histocompatibility complex class I molecules. Mol Immunol 48: 532-538.
283. Fu J, Li L, Bouvier M (2011) Adenovirus E3-19K proteins of different serotypes and subgroups have similar, yet distinct, immunomodulatory functions toward major histocompatibility class I molecules. J Biol Chem 286: 17631-17639.
284. Sester M, Ruszics Z, Mackley E, Burgert HG (2013) The transmembrane domain of the adenovirus E3/19K protein acts as an endoplasmic reticulum retention signal and contributes to intracellular sequestration of major histocompatibility complex class I molecules. J Virol 87: 6104-6117.
285. McSharry BP, Burgert HG, Owen DP, Stanton RJ, Prod'homme V, et al. (2008) Adenovirus E3/19K promotes evasion of NK cell recognition by intracellular sequestration of the NKG2D ligands major histocompatibility complex class I chain-related proteins A and B. J Virol 82: 4585-4594.
286. Martinez-Martin N, Ramani SR, Hackney JA, Tom I, Wranik BJ, et al. (2016) The extracellular interactome of the human adenovirus family reveals diverse strategies for immunomodulation. Nat Commun 7: 11473.
287. Windheim M, Southcombe JH, Kremmer E, Chaplin L, Urlaub D, et al. (2013) A unique secreted adenovirus E3 protein binds to the leukocyte common antigen CD45 and modulates leukocyte functions. Proc Natl Acad Sci U S A 110: E4884-4893.
288. Walters RW, Freimuth P, Moninger TO, Ganske I, Zabner J, et al. (2002) Adenovirus fiber disrupts CAR-mediated intercellular adhesion allowing virus escape. Cell 110: 789-799.
289. Sharp PM, Hahn BH (2010) The evolution of HIV-1 and the origin of AIDS. Philos Trans R Soc Lond B Biol Sci 365: 2487-2494.
290. World Health Organization (2016) Ebola Situation Report - 30 March 2016.
291. World Health Organization (2016) Zika situation report.
292. Sospedra M, Schippling S, Yousef S, Jelcic I, Bofill-Mas S, et al. (2014) Treating progressive multifocal leukoencephalopathy with interleukin 7 and vaccination with JC virus capsid protein VP1. Clin Infect Dis 59: 1588-1592.
293. Niederer A (2016) Immun-Booster soll Hirn sauber halten. Neue Zürcher Zeitung.
294. Potter RN, Cantrell JA, Mallak CT, Gaydos JC (2012) Adenovirus-associated deaths in US military during postvaccination period, 1999-2010. Emerg Infect Dis 18: 507-509.
295. Choudhry A, Mathena J, Albano JD, Yacovone M, Collins L (2016) Safety evaluation of adenovirus type 4 and type 7 vaccine live, oral in military recruits. Vaccine.
296. Flint SJ, Enquist, L. W., Racianello, V. R., Skalka, A. M. (2004) Prevention and Control of Viral Diseases. Principles of Virology: Molecular Biology, Pathogenesis, and Control of Animal Viruses. Washington: ASM Press. pp. 703-757.
297. Sanofi Pasteur (2015) Dengvaxia First Dengue Vaccine Approved in Brazil
298. Tatsis N, Ertl HC (2004) Adenoviruses as vaccine vectors. Mol Ther 10: 616-629.
299. Chen H, Xiang ZQ, Li Y, Kurupati RK, Jia B, et al. (2010) Adenovirus-based vaccines: comparison of vectors from three species of adenoviridae. J Virol 84: 10522-10532.
300. Patterson LJ (2011) The "STEP-wise" future of adenovirus-based HIV vaccines. Curr Med Chem 18: 3981-3986.
301. Jefferys R The Promise and Pitfalls of Adenoviruses as Vaccine Vectors.
302. Gunther PS, Peper JK, Faist B, Kayser S, Hartl L, et al. (2015) Identification of a Novel Immunodominant HLA-B*07: 02-restricted Adenoviral Peptide Epitope and Its Potential in Adoptive Transfer Immunotherapy. J Immunother 38: 267-275.
303. Wold WS, Toth K (2015) New drug on the horizon for treating adenovirus. Expert Opin Pharmacother 16: 2095-2099.
304. Aplander K, Marttila M, Manner S, Arnberg N, Sterner O, et al. (2011) Molecular wipes: application to epidemic keratoconjuctivitis. J Med Chem 54: 6670-6675.
305. Dellgren B (2016) Adenovir Pharma brings in MSEK 5.5 in preparation for the exit process. Adenovir Pharma.

306. Spjut S, Qian W, Bauer J, Storm R, Frangsmyr L, et al. (2011) A potent trivalent sialic acid inhibitor of adenovirus type 37 infection of human corneal cells. Angew Chem Int Ed Engl 50: 6519-6521.
307. Gorelik L, Reid C, Testa M, Brickelmaier M, Bossolasco S, et al. (2011) Progressive multifocal leukoencephalopathy (PML) development is associated with mutations in JC virus capsid protein VP1 that change its receptor specificity. J Infect Dis 204: 103-114.
308. Med JG (2016) Gene Therapy Clinical Trials Worldwide. John Wiley and Sons Ltd.
309. Jiang H, Gomez-Manzano C, Lang FF, Alemany R, Fueyo J (2009) Oncolytic adenovirus: preclinical and clinical studies in patients with human malignant gliomas. Curr Gene Ther 9: 422-427.
310. Cerullo V, Diaconu I, Romano V, Hirvinen M, Ugolini M, et al. (2012) An oncolytic adenovirus enhanced for toll-like receptor 9 stimulation increases antitumor immune responses and tumor clearance. Mol Ther 20: 2076-2086.
311. Hendrickx R, Stichling N, Koelen J, Kuryk L, Lipiec A, et al. (2014) Innate immunity to adenovirus. Hum Gene Ther 25: 265-284.
312. Kaufman HL, Kohlhapp FJ, Zloza A (2015) Oncolytic viruses: a new class of immunotherapy drugs. Nat Rev Drug Discov 14: 642-662.
313. Di Paolo NC, Miao EA, Iwakura Y, Murali-Krishna K, Aderem A, et al. (2009) Virus binding to a plasma membrane receptor triggers interleukin-1 alpha-mediated proinflammatory macrophage response in vivo. Immunity 31: 110-121.
314. Shayakhmetov DM, Di Paolo NC, Mossman KL (2010) Recognition of virus infection and innate host responses to viral gene therapy vectors. Mol Ther 18: 1422-1429.
315. Lopez-Gordo E, Denby L, Nicklin SA, Baker AH (2014) The importance of coagulation factors binding to adenovirus: historical perspectives and implications for gene delivery. Expert Opin Drug Deliv 11: 1795-1813.
316. Arnberg N (2012) Adenovirus receptors: implications for targeting of viral vectors. Trends Pharmacol Sci 33: 442-448.
317. Thaci B, Ulasov IV, Wainwright DA, Lesniak MS (2011) The challenge for gene therapy: innate immune response to adenoviruses. Oncotarget 2: 113-121.
318. Marshall E (1999) Gene therapy death prompts review of adenovirus vector. Science 286: 2244-2245.
319. Check E (2002) A tragic setback. Nature 420: 116-118.
320. Reardon S (2016) First CRISPR clinical trial gets green light from US panel. Nature Publishing Group.
321. Yla-Herttuala S (2012) Endgame: glybera finally recommended for approval as the first gene therapy drug in the European union. Mol Ther 20: 1831-1832.
322. (FDA) USFaDA (2015) FDA approves first-of-its-kind product for the treatment of melanoma.
323. Kwiatkowska A, Nandhu MS, Behera P, Chiocca EA, Viapiano MS (2013) Strategies in gene therapy for glioblastoma. Cancers (Basel) 5: 1271-1305.
324. Stehle T, Harrison SC (1996) Crystal structures of murine polyomavirus in complex with straight-chain and branched-chain sialyloligosaccharide receptor fragments. Structure 4: 183-194.
325. Koromyslova AD, Leuthold MM, Bowler MW, Hansman GS (2015) The sweet quartet: Binding of fucose to the norovirus capsid. Virology 483: 203-208.
326. Cheever MA, Allison JP, Ferris AS, Finn OJ, Hastings BM, et al. (2009) The prioritization of cancer antigens: a national cancer institute pilot project for the acceleration of translational research. Clin Cancer Res 15: 5323-5337.
327. Lenman A, Liaci AM, Liu Y, Årdahl C, Rajan A, et al. (2015) Human adenovirus 52 uses sialic acid-containing glycoproteins and the coxsackie and adenovirus receptor for binding to target cells. PLoS Pathog 11: e1004657.
328. Curreli S, Arany Z, Gerardy-Schahn R, Mann D, Stamatos NM (2007) Polysialylated neuropilin-2 is expressed on the surface of human dendritic cells and modulates dendritic cell-T lymphocyte interactions. J Biol Chem 282: 30346-30356.
329. Ferrero MA, Aparicio LR (2010) Biosynthesis and production of polysialic acids in bacteria. Appl Microbiol Biotechnol 86: 1621-1635.
330. Mangroo CS, Brown M (2016) Structural Features of Human Enteric Adenovirus 41 as Determinants Affecting Entry into Host Cells. 12th International Adenovirus Meeting. Barsinghausen, Germany.
331. Zak DE, Andersen-Nissen E, Peterson ER, Sato A, Hamilton MK, et al. (2012) Merck Ad5/HIV induces broad innate immune activation that predicts CD8(+) T-cell responses but is attenuated by preexisting Ad5 immunity. Proc Natl Acad Sci U S A 109: E3503-3512.
332. Waddington SN, McVey JH, Bhella D, Parker AL, Barker K, et al. (2008) Adenovirus serotype 5 hexon mediates liver gene transfer. Cell 132: 397-409.
333. Abbink P, Lemckert AA, Ewald BA, Lynch DM, Denholtz M, et al. (2007) Comparative seroprevalence and immunogenicity of six rare serotype recombinant adenovirus vaccine vectors from subgroups B and D. J Virol 81: 4654-4663.
334. Kloos A, Woller N, Gurlevik E, Ureche CI, Niemann J, et al. (2015) PolySia-Specific Retargeting of Oncolytic Viruses Triggers Tumor-Specific Immune Responses and Facilitates Therapy of Disseminated Lung Cancer. Cancer Immunol Res 3: 751-763.
335. Liu J, Ewald BA, Lynch DM, Denholtz M, Abbink P, et al. (2008) Magnitude and phenotype of cellular immune responses elicited by recombinant adenovirus vectors and heterologous prime-boost regimens in rhesus monkeys. J Virol 82: 4844-4852.
336. Krieg AM, Wu T, Weeratna R, Efler SM, Love-Homan L, et al. (1998) Sequence motifs in adenoviral DNA block immune activation by stimulatory CpG motifs. Proc Natl Acad Sci U S A 95: 12631-12636.
337. Chakupurakal G, Onion D, Bonney S, Cobbold M, Mautner V, et al. (2013) HLA-peptide multimer selection of adenovirus-specific T cells for adoptive T-cell therapy. J Immunother 36: 423-431.
338. Billerbeck E, Horwitz JA, Labitt RN, Donovan BM, Vega K, et al. (2013) Characterization of human antiviral adaptive immune responses during hepatotropic virus infection in HLA-transgenic human immune system mice. J Immunol 191: 1753-1764.

339. Pasarica M, Mashtalir N, McAllister EJ, Kilroy GE, Koska J, et al. (2008) Adipogenic human adenovirus Ad-36 induces commitment, differentiation, and lipid accumulation in human adipose-derived stem cells. Stem Cells 26: 969-978.
340. Salehian B, Forman SJ, Kandeel FR, Bruner DE, He J, et al. (2010) Adenovirus 36 DNA in adipose tissue of patient with unusual visceral obesity. Emerg Infect Dis 16: 850-852.
341. Pasarica M, Shin AC, Yu M, Ou Yang HM, Rathod M, et al. (2006) Human adenovirus 36 induces adiposity, increases insulin sensitivity, and alters hypothalamic monoamines in rats. Obesity (Silver Spring) 14: 1905-1913.
342. Dhurandhar NV, Israel BA, Kolesar JM, Mayhew G, Cook ME, et al. (2001) Transmissibility of adenovirus-induced adiposity in a chicken model. Int J Obes Relat Metab Disord 25: 990-996.
343. Rogers PM, Fusinski KA, Rathod MA, Loiler SA, Pasarica M, et al. (2008) Human adenovirus Ad-36 induces adipogenesis via its E4 orf-1 gene. Int J Obes (Lond) 32: 397-406.
344. Arnberg N, Kidd AH, Edlund K, Olfat F, Wadell G (2000) Initial interactions of subgenus D adenoviruses with A549 cellular receptors: sialic acid versus alpha(v) integrins. J Virol 74: 7691-7693.
345. Aamelfot M, Dale OB, Weli SC, Koppang EO, Falk K (2014) The in situ distribution of glycoprotein-bound 4-O-Acetylated sialic acids in vertebrates. Glycoconj J 31: 327-335.
346. Hanses F, Kopp A, Bala M, Buechler C, Falk W, et al. (2011) Intracellular survival of Staphylococcus aureus in adipocyte-like differentiated 3T3-L1 cells is glucose dependent and alters cytokine, chemokine, and adipokine secretion. Endocrinology 152: 4148-4157.
347. Aoki K, Tagawa Y (2002) A twenty-one year surveillance of adenoviral conjunctivitis in Sapporo, Japan. Int Ophthalmol Clin 42: 49-54.
348. Ishiko H, Shimada Y, Konno T, Hayashi A, Ohguchi T, et al. (2008) Novel human adenovirus causing nosocomial epidemic keratoconjunctivitis. J Clin Microbiol 46: 2002-2008.
349. Kaneko H, Aoki K, Ishida S, Ohno S, Kitaichi N, et al. (2011) Recombination analysis of intermediate human adenovirus type 53 in Japan by complete genome sequence. J Gen Virol 92: 1251-1259.
350. Huang G, Yao W, Yu W, Mao L, Sun H, et al. (2014) Outbreak of epidemic keratoconjunctivitis caused by human adenovirus type 56, China, 2012. PLoS One 9: e110781.
351. Ford E, Nelson KE, Warren D (1987) Epidemiology of epidemic keratoconjunctivitis. Epidemiol Rev 9: 244-261.
352. Bawazeer A, Hodge WG (2015) Epidemic Keratoconjunctivitis. In: Hampton R, Law SK, Rapuano CJ, editors. Medscape Drugs & Diseases: WebMD LLC.
353. Kinchington PR, Romanowski EG, Jerold Gordon Y (2005) Prospects for adenovirus antivirals. J Antimicrob Chemother 55: 424-429.
354. Dahl M, Gerhardsson E, Lafolie P, Allard A, Laurell CG (2016) More symptoms with sialic receptor-positive adenovirus in epidemic keratoconjunctivitis. Acta Ophthalmol 94: e375-376.
355. Johansson SM, Arnberg N, Elofsson M, Wadell G, Kihlberg J (2005) Multivalent HSA conjugates of 3'-sialyllactose are potent inhibitors of adenoviral cell attachment and infection. Chembiochem 6: 358-364.
356. Johansson SM, Nilsson EC, Elofsson M, Ahlskog N, Kihlberg J, et al. (2007) Multivalent sialic acid conjugates inhibit adenovirus type 37 from binding to and infecting human corneal epithelial cells. Antiviral Res 73: 92-100.
357. Caraballo R, Saleeb M, Bauer J, Liaci AM, Chandra N, et al. (2015) Triazole linker-based trivalent sialic acid inhibitors of adenovirus type 37 infection of human corneal epithelial cells. Org Biomol Chem 13: 9194-9205.
358. Rossmann MG (1989) The canyon hypothesis. Hiding the host cell receptor attachment site on a viral surface from immune surveillance. J Biol Chem 264: 14587-14590.
359. Renaudet O, Spinelli N (2011) Synthesis and Biological Applications of Glycoconjugates: Bentham Books.
360. Qin M, Chen S, Yu T, Escuadro B, Sharma S, et al. (2003) Coxsackievirus adenovirus receptor expression predicts the efficiency of adenoviral gene transfer into non-small cell lung cancer xenografts. Clin Cancer Res 9: 4992-4999.
361. Curiel DT (1999) Strategies to adapt adenoviral vectors for targeted delivery. Ann N Y Acad Sci 886: 158-171.
362. Zheng S, Ulasov IV, Han Y, Tyler MA, Zhu ZB, et al. (2007) Fiber-knob modifications enhance adenoviral tropism and gene transfer in malignant glioma. J Gene Med 9: 151-160.
363. Van Houdt WJ, Wu H, Glasgow JN, Lamfers ML, Dirven CM, et al. (2007) Gene delivery into malignant glioma by infectivity-enhanced adenovirus: in vivo versus in vitro models. Neuro Oncol 9: 280-290.
364. Fueyo J, Alemany R, Gomez-Manzano C, Fuller GN, Khan A, et al. (2003) Preclinical characterization of the antiglioma activity of a tropism-enhanced adenovirus targeted to the retinoblastoma pathway. J Natl Cancer Inst 95: 652-660.
365. Mathias P, Wickham T, Moore M, Nemerow G (1994) Multiple adenovirus serotypes use alpha v integrins for infection. J Virol 68: 6811-6814.
366. Robinson CM, Zhou X, Rajaiya J, Yousuf MA, Singh G, et al. (2013) Predicting the next eye pathogen: analysis of a novel adenovirus. MBio 4: e00595-00512.
367. Chiu CY, Mathias P, Nemerow GR, Stewart PL (1999) Structure of adenovirus complexed with its internalization receptor, alphavbeta5 integrin. J Virol 73: 6759-6768.
368. Xiong JP, Stehle T, Zhang R, Joachimiak A, Frech M, et al. (2002) Crystal structure of the extracellular segment of integrin alpha Vbeta3 in complex with an Arg-Gly-Asp ligand. Science 296: 151-155.
369. Roelvink PW, Kovesdi I, Wickham TJ (1996) Comparative analysis of adenovirus fiber-cell interaction: adenovirus type 2 (Ad2) and Ad9 utilize the same cellular fiber receptor but use different binding strategies for attachment. J Virol 70: 7614-7621.
370. Norrby E (1969) The structural and functional diversity of Adenovirus capsid components. J Gen Virol 5: 221-236.
371. Fender P, Ruigrok RW, Gout E, Buffet S, Chroboczek J (1997) Adenovirus dodecahedron, a new vector for human gene transfer. Nat Biotechnol 15: 52-56.

372. Bai M, Campisi L, Freimuth P (1994) Vitronectin receptor antibodies inhibit infection of HeLa and A549 cells by adenovirus type 12 but not by adenovirus type 2. J Virol 68: 5925-5932.
373. Bai M, Harfe B, Freimuth P (1993) Mutations that alter an Arg-Gly-Asp (RGD) sequence in the adenovirus type 2 penton base protein abolish its cell-rounding activity and delay virus reproduction in flat cells. J Virol 67: 5198-5205.
374. Bal HP, Chroboczek J, Schoehn G, Ruigrok RW, Dewhurst S (2000) Adenovirus type 7 penton purification of soluble pentamers from Escherichia coli and development of an integrin-dependent gene delivery system. Eur J Biochem 267: 6074-6081.
375. Mathias P, Galleno M, Nemerow GR (1998) Interactions of soluble recombinant integrin alphav beta5 with human adenoviruses. J Virol 72: 8669-8675.
376. Chung SH, Weiss RS, Frese KK, Prasad BV, Javier RT (2008) Functionally distinct monomers and trimers produced by a viral oncoprotein. Oncogene 27: 1412-1420.
377. Hegde V, Na HN, Dubuisson O, Burke SJ, Collier JJ, et al. (2016) An adenovirus-derived protein: A novel candidate for anti-diabetic drug development. Biochimie 121: 140-150.
378. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, et al. (1999) Protein identification and analysis tools in the ExPASy server. Methods Mol Biol 112: 531-552.
379. Hendrickson WA, Ogata CM (1997) [28] Phase determination from multiwavelength anomalous diffraction measurements. Methods Enzymol 276: 494-523.
380. Doutch J, Hough MA, Hasnain SS, Strange RW (2012) Challenges of sulfur SAD phasing as a routine method in macromolecular crystallography. J Synchrotron Radiat 19: 19-29.
381. Human Adenovirus Working Group (2016) Serotyping Tool.

6. APPENDIX

6.1. Acknowledgments

‚No Man Is an Island' – however often this phrase might have been used already, it could not be truer. Many people have influenced my work and my personal life over the last five years, and I am thankful for all the reassuring words, challenging ideas and discussions, and the wonderful experiences they shared with me – this is my chance to say 'Thank you!'.

First off, I would like to thank **Thilo** not only for giving me the projects I loved working on so much and for being a great boss who puts a lot of trust in his group members, but also for being a true role model who is always there when it counts, no matter how little time there may be.

Further, I would like to thank **Niklas** for the very close and fruitful cooperation we have had over the last years, and for taking interest in my PhD without the need to do so.

Michael – you started out as my supervisor, became my colleague, and ended up as my friend. Thank you for the last five years and for your honesty and your support. I wish you the best of luck for the future!

I am also thankful to the members of Niklas' group, especially **Annasara** and **Naresh**, whom it was a pleasure to work with.

My students **Lisa**, **Anna**, **Timm**, **Vanessa**, **Anna-Lena**, **Gregor**, **Antonia**, **Paul,** and **Laura** have taught me (at least) equally as much as I taught them. I would also like to thank them for their dedication to the projects and for helping me to generate invaluable results, many of which have ended up here in this dissertation**.**

Many thanks go to **Georg** for being the go-to guy for almost everything in the lab, and for always seeing the best in people – good luck in the learning process of how to say no sometimes ;)

Also, I would like to thank **Christoph** for being the go-to guy for everything wetlab-related and for always being willing to share his expertise.

Felix – we spent the last decade together as classmates, colleagues, and friends. Thank you for your loyalty, your input, and your friendship!

Special thanks go to **Joana** for quickly becoming one of my closest friends. Everything becomes easier if you can share it with a person you truly trust. I consider myself lucky to have a friend like you!

I would like to thank all **other members** of the Stehle lab for the helpful discussions, the favors and the nice working atmosphere we have.

Furthermore, my thanks go to the other collaborators who made much of my work possible: **Rémi**, **Emil**, **Mikael**, **David**, **Martin**, **Yan**, **Ten**, **Bärbel**, **Hubert**, **York, Sam, and Bob.**

Further, I would like to thank Ancilla and **Remco** for helpful advice on NMR and ITC experiments, **Judith** and **Prof. Harald Groß** for letting me use their HPLC on a weekend, the **Kalbacher Lab** for MS measurements, and **Benjamin** and **Prof. Klaus Schultze-Osthoff** for help with setting up their light cycler for DSF experiments.

Ich möchte mich herzlich bei der **Sängerschaft Hohentübingen** bedanken, deren Aktive mir in einer schwierigen Situation ohne zu zögern Obhut gewährt haben. Vielen Dank!

Ein besonderes Dankeschön geht an meine Freundin **Susi**, die immer für mich da ist, sei es durch tatkräftige Unterstützung oder guten Rat, und die mir die zahlereiche Abende nachgesehen hat, die ich im Labor statt zu Hause verbracht habe. Du bist das Beste, das mir passieren konnte. Ich liebe Dich!

Auch Susis Eltern **Sabine** und **Tillo** möchte ich für ihre Unterstützung herzlich danken.

Es ist schwierig, die richtigen Worte für **meine Familie** zu finden. Oft sind es letztlich diejenigen, die einem den Erfolg am meisten wünschen und gönnen, die dann am meisten unter den Folgen zu leiden haben - finanziell und aufgrund der Distanz. Letztendlich aber seid ihr das Fundament, auf dem mein gesamter Werdegang aufgebaut ist, und der Fels zu dem ich immer wieder zurückkehren werde. Vielen Dank.

6.2. Sequence Alignments of Relevant HAdV Gene Products

The sequence alignments used for analyses throughout this dissertation are listed below. All sequences were based on the annotation from the HAdV Working Group Serotyping Tool [381]

6.2.1. Fiber Knob Domains (HAdV Types C01 – D70)

```
HAdV-B14       -ntlwtgvnpteancqmmdssesndckliltlvktgalvtafvyvigvsnnfnmlttyrn
HAdV-B55       -ntlwtginpteancqmmdssesndckliltlvktgalvtafvyvigvsnnfnmlttyrn
HAdV-B07       -ntlwtgvnptrancqimassesndckliltlvktgalvtafvyvigvsndfnmltthkn
HAdV-B11       --nlwtgvnpteancqimnssesndckliltlvktgalvtafvyvigvsnnfnmltthrn
HAdV-B34       -ntlwtginpp-pncqiventntndgkltlvlvkngglvngyvslvgvsdtvnqmftqkt
HAdV-B35       -ntlwtginpp-pncqiventntndgkltlvlvkngglvngyvslvgvsdtvnqmftqkt
HAdV-B21       -ntlwtgikpp-pncqiventdtndgkltlvlvkngglvngyvslvgvsdtvnqmftqks
HAdV-B50       -ntlwtgikpp-pncqiventdtndgkltlvlvkngglvngyvslvgvsdtvnqmftqks
HAdV-B03       -ntlwtgpkpe-anciieygkenpdskltlilvknggivngyvtlmgasdyvntlfknkn
hadv-b66       -ntlwtgpkpe-anciieygkenpdskltlilvknggivngyvtlmgasdyvntlfknkn
HAdV-B16       -ntlwtgakps-ancvikegedspdckltlvlvknggllngyitlmgaseytntlfknnq
hadv-b68       -ntlwtgakps-ancvikegedspdckltlvlvknggllngyitlmgaseytntlfknnq
HAdV-G52S      -qtlwtpptsn-pnctvyt---esdsllslcltkcgahvlgsvsltgvagtmtnmae---
HAdV-F40S      lttiwsi-spt-pncsiye---tqdanlflcltkngahvlgtitikglkgalremnd---
HAdV-F41S      -ttiwsi-spt-pncsiye---tqdanlflcltkngahvlgtitikglkgalremhd---
HAdV-G52L      -dtlwttadps-pncsiyt---dldakmwlslvkqggmvhgsvalkalkgtllspte---
HAdV-F40L      pttlwttadps-pnatfye---sldakvwlvlvkcngmvngtisikaqkgtllkpta---
HAdV-F41L      -ttlwttadps-pnatfye---sldakvwlvlvkcngmvngtisikaqkgillrpta---
HAdV-A18       -ltlwttpdpl-pncslle---dldtkltlclskngaivhgvvsfvpvkgallnlqnp--
HAdV-A12       -ltlwttpdpp-pncsliq---eldakltlcltkngsivngivslvgvkgnllniqstt-
HAdV-A31       -ltlwttpdpp-pnctlrq---eldakltlcltknesivngivsligvkgdllhiqptt-
HAdV-A61       -ltlwttpdpp-pnctlrq---eldakltlcltknesivngivsligvkgdllhiqptt-
HAdV-C02       -ltlwttpdps-pncrihs---dndckftlvltkcgsqvlatvaalavsgdlssmtgtv-
HAdV-E4        -ltlwttpdps-pncrihs---dndckftlvltkcgsqvlatvaalavsgdlssmtgtv-
HAdV-C06       -ltlwttpdps-pncrias---dkdckltlaltkcgsqilgtvsalavsgnmasingtl-
HAdV-C57       -ltlwttpdps-pncrias---dkdckltlaltkcgsqilgtvsalavsgnmasingtl-
HAdV-C01       -ltlwttpdps-pncqihs---ekdakltlvltkcgsqvlatvsalavrgslapisgti-
HAdV-C05       -ltlwttpaps-pncrlna---ekdakltlvltkcgsqilatvsvlavkgslapisgtv-
HAdV-D45       -rtlwttpdts-pnckist---ekdskltlvltkcgsqilasvsllavagsylnmtastq
HAdV-D28       -rtlwttpdts-pnckmse---vkdskltliltkcgsqilgsvsllavkgeyqnmtastn
HAdV-D43       -rtlwttpdts-pnckmse---akdskltliltkcgsqilgsvsllavkgeyqnmtantk
HAdV-D62       -rtlwttpdts-pnckmsk---ekdskltltltkcgsqilgsvsllavsgeylnmttntn
HAdV-D26       -rtlwttpdts-pnckmst---ekdskltltltkcgsqvlgnvsllavtgeyhqmtattk
HAdV-D27       -rtlwttpdts-pnckmlt---kkdskltltltkcgsqilgnvsllavsgkylnmtkd-e
HAdV-D22       -rtlwttpihl-ptaqyqk---i-kiqnnliltkcgsqilasfsllvvkgtyatvdkntt
HAdV-D42       -rtlwttpdps-pncrvse---dkdskltliltkcgsqilasfsllvvkgtyttvdkntt
HAdV-D15       -rtlwttpdps-pncrvse---dkdskltliltkcgsqilasfsllvvkgtyatvdkntt
HAdV-D69       -rtlwttpdps-pncrvse---dkdskltliltkcgsqilasfsllvvkgtyatvdkntt
HAdV-D29       -rtlwttldps-pnckidi---ekdskltlvltkcgsqilanvsliivngkfkilnnktd
HAdV-D63       -rtlwttldps-pnckidi---ekdskltlvltkcgsqilanvsliivngkfkilnnktd
HAdV-D70       -rtlwttldps-pnckidi---ekdskltlvltkcgsqilanvsliivngkfkilnnktd
HAdV-D58       -rtlwttldps-pnckidi---ekdskltlvltkcgsqilanvsliivngkfkilnnktd
HAdV-D25       -rtlwttldps-pncridv---dkdskltlvltkcgsqilanvsllvvkgrfqnlnyktn
HAdV-D67       -rtlwttldps-pncridv---dkdskltlvltkcgsqilanvsllvvkgrfqnlnyktn
HAdV-D17       -rtlwttpdts-pncridk---ekdskltlvltkcgsqilanvslivvsgkyqyidhatn
HAdV-D10       -rtlwttpdts-pnctiaq---dkdskltlvltkcgsqilanvslivvagkyhiinnknn
HAdV-D19       -rtlwttpdts-pnctiaq---dkdskltlvltkcgsqilanvslivvagkyhiinnktn
HAdV-D37       -rtlwttpdts-pnctiaq---dkdskltlvltkcgsqilanvslivvagkyhiinnktn
HAdV-D64       -rtlwttpdts-pnctiaq---dkdskltlvltkcgsqilanvslivvagkyhiinnktn
HAdV-D65       -rtlwttpdts-pnckidq---dkdskltlvltkcgsqilanvslivvagkykiinnntq
HAdV-D56       -rtlwttpdts-pnckidq---dkdskltlvltkcgsqilanvslivvagkykiinnntq
HAdV-D59       -rtlwttpdts-pnckidq---dkdskltlvltkcgsqilanvslivvagkykiinnntq
HAdV-D09       -rtlwttpdts-pncridq---dkdskltlvltkcgsqilanvslivvagrykiinnntq
HAdV-D08       -rtlwttpdts-pncridq---dkdskltlvltkcgsqilanvslivvagrykiinnntn
HAdV-D53       -rtlwttpdts-pncridq---dkdskltlvltkcgsqilanvslivvagrykiinnntn
HAdV-D54       -rtlwttpdts-pncridq---dkdskltlvltkcgsqilanvslivvagrykiinnntq
HAdV-D32       -rtlwttpdps-pnctide---erdskltlvltkcgsqilanvsllvvkgkfsninnntn
HAdV-D33       -rtlwttpdps-pnctide---erdskltlvltkcgsqilanvsllvvkgkfsninnntn
HAdV-D24       -rtlwttpdps-pnctidq---erdskltlvltkcgsqilanvsllvvkgkfsninnntn
HAdV-D46       -rtlwttpdps-pnctidq---erdskltlvltkcgsqilanvsllvvkgkfsninnnan
HAdV-D44       -rtlwttpdps-pnckidq---dkdskltfvltkcgsqilanmsllvvkgkfsminnkvn
HAdV-D48       -rtlwttpdps-pnckidq---dkdskltfvltkcgsqilanmsllvvkgkfsminnkvn
HAdV-D30       -rtlwttpdps-pnckvse---ekdskltlvltkcgsqilasvsllvvkgkfaninnetn
HAdV-D49       -rtlwttpdps-pnckvse---ekdskltlvltkcgsqilasvsllvvkgkfaninnktn
HAdV-D36       -rtlwttpdps-pnckvet---ardskltlaltkcgsqilatvsllvvtgkyaiisdtvn
HAdV-D60       -rtlwttpdps-pncriev---akdakltlvltkcgsqilasvsiivlkgtyeyakkets
HAdV-D13       -rtlwttpdps-pnckaet---ekdskltlvltkcgsqilatvsiivlkgkyefvkkete
HAdV-D38       -rtlwttpdps-pnckaet---ekdskltlvltkcgsqilatvsiivlkgkyefvkkete
HAdV-D39       -rtlwttpdps-pnckaet---ekdskltlvltkcgsqilatvsiivlkgkyefvkketd
HAdV-D51       -rtlwttpdps-pnckvse---akdskltlvltkcgsqilasvallivkgkyqtisesti
HAdV-D23       -rtlwttpdps-pnckvie---akdskltlvltkcgsqilanmsllilkgtyeyisnaia
HAdV-D20       -rtlwttpdps-pnckiee---vkdskltlvltkcgsqilatmafqvvkgtyenisknta
HAdV-D47       -rtlwttpdps-pnckieq---dkdskltlvltkcgsqilatmafqvvkgtyenisknta
                 .:*:      ..        .    : * *  . : . .
```

```
HAdV-B14        -----inftaelffdsagnlltsl--------sslktplnhksgqnmatgaitnaksfmp
HAdV-B55        -----inftaelffdsagnlltsl--------sslktplnhksgqnmatgaitnaksfmp
HAdV-B07        -----inftaelffdstgnlltsl--------sslktplnhksgqnmatgaltnakgfmp
HAdV-B11        -----inftaelffdstgnlltrl--------sslktplnhksgqnmatgaitnakgfmp
HAdV-B34        -----aniqlrlyfdssgnlltde--------sdlkiplknkss-tatsetvasskafmp
HAdV-B35        -----aniqlrlyfdssgnlltee--------sdlkiplknkss-tatsetvasskafmp
HAdV-B21        -----atiqlrlyfdssgnlltde--------snlkiplknkss-tatseaatsskafmp
HAdV-B50        -----atiqlrlyfdssgnlltde--------snlkiplknkss-tatseaatsskafmp
HAdV-B03        -----vsinvelyfdatghilpdl--------sslktdlqlkykqt----thfsargfmp
hadv-b66        -----vsinvelyfdatghilpdl--------sslktdlqlkykqt----thfsargfmp
HAdV-B16        -----vtidvnlafdntgqiityl--------sslksnlnfkdnqnmatgtitsakgfmp
hadv-b68        -----vtidvnlafdntgqiityl--------sslksnlnfkdnqnmttgtitsakgfmp
HAdV-G52S       -----tslaieftfddtgkllhsp---------lvnntfsirqgdspasnptynalafmp
HAdV-F40S       -----nalsvklpfdnqgnllnca---------lesstwryqetnav----asnaltfmp
HAdV-F41S       -----nalslklpfdnqgnllnca---------lesstwryqetnav----asnaltfmp
HAdV-G52L       -----saiviilhfdnygvrilnyptlgtqgtlgnnatwgyrqgesad-tnvlnalafmp
HAdV-F40L       -----sfisfvmyfysdgtwrknypvfdnegilansatwgyrqgqsan-tnvsnavefmp
HAdV-F41L       -----sfisfvmyfysdgtwrknypvfdnegilansatwgyrqgqsan-tnvsnavefmp
HAdV-A18        -----tsiglhilfdqhgvlqtgtpsa-----iapqaywgyrqgqsisptpvtnalkfmp
HAdV-A12        -----ttvgvhlvfdeqgrlitstpta-----lvpqaswgyrqgqsvstntvtnglgfmp
HAdV-A31        -----ttvglhlvfdrqgrlvtttpta-----lvpqaswgykqgqsvsssavanalgfmp
HAdV-A61        -----ttvglhlvfdrqgrlvtttpta-----lvpqaswgykqgqsvsssavanalgfmp
HAdV-C02        -----asvsiflrfdqngvlmens--------slkkhywnfrngnstnanpytnavgfmp
HAdV-E4         -----asvsiflrfdqngvlmens--------slkkhywnfrngnstnanpytnavgfmp
HAdV-C06        -----ssvnlvlrfddngvlmsns--------sldkqywnfrngdstngqpytyavgfmp
HAdV-C57        -----ssvnlvlrfddngvlmsns--------sldkqywnfrngdstngqpytyavgfmp
HAdV-C01        -----ssahiilrfnehgvlmnhs--------sldpqywnfrkgdltnataytnavgfmp
HAdV-C05        -----qsahliirfdengvllnns--------fldpeywnfrngdltegtaytnavgfmp
HAdV-D45        -----ksikvslmfdskgllmtts--------sidkgywnyrnknsvvgtayenaipfmp
HAdV-D28        -----knvkitllfdangvllegs--------sldkeywnfrnndstvsgkyenavpfmp
HAdV-D43        -----knvkitllfdangvllags--------sldkeywnfrsndstvsgnyenavgfmp
HAdV-D62        -----rtitikllfdakgvlltss--------sisgdywnfrnnnstvsnkyenavafmp
HAdV-D26        -----kdvkisllfdengillpss--------slskdywnyrsddsivsqkynnavpfmp
HAdV-D27        -----tgvkiillfdrngvlmqes--------sldkeywnyrndnnvigtpyenavgfmp
HAdV-D22        ----kkefsikllfdangklkses--------nls-gywnyrsdnsvvgtaynnavpfmp
HAdV-D42        ----nkqfsikllfdangklkses--------nls-gywnyrsdnsvvstpydnavpfmp
HAdV-D15        ----nkqfsikllfdangklkses--------nls-gywnyrsdnsvvstpydnavpfmp
HAdV-D69        ----nkqfsikllfdangklkses--------nls-gywnyrsdnsvvstpydnavpfmp
HAdV-D29        psl-pksfnikllfdqngvllens--------niekqylnfrsgdsilpepyknaigfmp
HAdV-D63        psl-pksfnikllfdqngvllens--------niekqylnfrsgdsilpepyknaigfmp
HAdV-D70        psl-pksfnikllfdqngvllens--------niekqylnfrsgdsilpepyknaigfmp
HAdV-D58        psl-pksfnikllfdqngvllens--------niekqylnfrsgdsilpepyknaigfmp
HAdV-D25        pnl-pktftikllfdengilkdss--------nldknywnyrngnsilaeqyknavgfmp
HAdV-D67        pnl-pkafaikllfdengilkdss--------nldknywnyrsgnsilaeqyknavgfmp
HAdV-D17        ptl--ksfkikllfdnkgvllpss--------nldstywnfrsdnltvseayknavefmp
HAdV-D10        pei--ksftikllfdkngvlldns--------nlgktywnfrsgdsnvstayekaigfmp
HAdV-D19        pki--ksftikllfnkngvlldns--------nlgkaywnfrsgnsnvstayekaigfmp
HAdV-D37        pki--ksftikllfnkngvlldns--------nlgkaywnfrsgnsnvstayekaigfmp
HAdV-D64        pki--ksftikllfnkngvlldns--------nlgkaywnfrsgnsnvstayekaigfmp
HAdV-D65        psl--kgftikllfdqngvlmess--------nlgksywnfrnensimstayekaigfmp
HAdV-D56        pal--kgftikllfdengvlmess--------nlgksywnfrnensimstayekaigfmp
HAdV-D59        pal--kgftikllfdengvlmess--------nlgksywnfrnensimstayekaigfmp
HAdV-D09        pal--kgftikllfdkneslwnss--------nlgksswnfrnensimstayekaigfmp
HAdV-D08        pal--kgftikllfdkngvlmess--------nlgksywnfrnqnsimstayekaigfmp
HAdV-D53        pal--kgftikllfdkngvlmess--------nlgksywnfrnqnsimstayekaigfmp
HAdV-D54        pal--kgftikllfdkngvlmess--------nlsksywnfrnensimstayekaigfmp
HAdV-D32        ptd--kkitvkllfnekgvlmdss--------slkkeywnyrndnstvsqaydnavpfmp
HAdV-D33        ptd--kkitvkllfnekgvlmdss--------slkkeywnyrndnstvsqaydnavpfmp
HAdV-D24        ptd--kkitvkllfnekgvlmdss--------tlkkeywnyrndnstvsqaydnavpfmp
HAdV-D46        ptd--kkitvkllfnekgvlmdss--------tlkkeywnyrndnstvsqaydnavpfmp
HAdV-D44        gtddykkftikllfdekgvllkds--------sldkeywnyrsnnnnvgsayeeavgfmp
HAdV-D48        gtddykkftikllfdekgvllkds--------sldkeywnyrsnnnnvgsayeeavgfmp
HAdV-D30        pgedykkfsvkllfdangklltgs--------sldgnywnyknkdsvigspyenavpfmp
HAdV-D49        pgedykkfsvkllfdangklltgs--------sldgnywnyknkdsvigspyenavpfmp
HAdV-D36        ----pkqfsikllfndkgvllsds--------nldgtywnyrsnnnnigtpykeavgfmp
HAdV-D60        ----vkefsikllfdkngvllpes--------nldkdywnyrsddltiakpyenavpfmp
HAdV-D13        ----pksfdvkllfdskgvllpts--------nlskeywnyrsydnnigtpyenavpfmp
HAdV-D38        ----pksfdvkllfdskgvllpts--------nlskeywnyrsydnnigtpyenavpfmp
HAdV-D39        ----pksfdvkllfdskgvllpts--------nlskeywnyrsydnnigtpyenavpfmp
HAdV-D51        pkd-qrnfsvklmfdekgklldks--------sldkeywnfrsndsvvgtaydnavpfmp
HAdV-D23        ----nksftikllfndkgvlmdgs--------sldkdywnyksddsvmskayenavpfmp
HAdV-D20        ----knsfsikllfddngkllegs--------sldkdywnfrsddsiipnqydnavpfmp
HAdV-D47        ----kksfsikllfddngkllegs--------sldkdywnfrnddsimpnqydnavpfmp
                          : *                       :          .  ***
```

```
HAdV-B14        sttaypfnnn--s--------r--enyiygtchyta-sdhtafpidisvmlnqra----i
HAdV-B55        sttaypfnnn--s--------rekenyiygtchyta-sdhtafpidisvmlnqra----i
HAdV-B07        sttaypfnvn--s--------rekenyiygtcyyta-sdhtafpidisvmlnqra----l
HAdV-B11        sttaypfndn--s--------rekenyiygtcyyta-sdrtafpidisvmlnrra----i
HAdV-B34        sttaypfntt--t--------rdsenyihgicyymtsydrslfplnisimlnsrm----i
HAdV-B35        sttaypfntt--t--------rdsenyihgicyymtsydrslfplnisimlnsrm----i
HAdV-B21        sttaypfntt--t--------rdsenyihgicyymtsydrslvplnisimlnsht----i
HAdV-B50        sttaypfntt--t--------rdsenyihgicyymtsydrslvplnisimlnsrt----i
HAdV-B03        sttaypfvlpnag--------tdnenyifgqcyyka-sdgalfplevtvtlnkrl----p
hadv-b66        sttaypfvlpnag--------tdnenyifgqcyyka-sdgalfplevtvtlnkrl----p
HAdV-B16        sttaypfityate--------tlnedyiygecyyks-tngtlfplkvtvtlnrrm----l
hadv-b68        sttaypfityate--------tlnedyiygecyyks-tngtlfplkvtvtlnrrm----s
HAdV-G52S       nstlyarggs-----------geprnnyyvqtylrg---nvqrpitltvtfnsaa-----
HAdV-F40S       nstvyprnkt-----------adpgnmli----------qispnitfsvvynein-----
HAdV-F41S       nstvyprnkt-----------ahpgnmli----------qispnitfsvvynein-----
HAdV-G52L       sskryprgrg-----------sevqnqtvgytciqg---dlsmpvpyqiqynygp-----
HAdV-F40L       sskrypnekg-----------sevqnmaltytflqg---dpnmaisfqsiynhai-----
HAdV-F41L       sskrypnqkg-----------sevqnmaltytflqg---dpnmaisfqsiynhal-----
HAdV-A18        nssayprang-----------sepksqtftstymqg---diakpmtmkvsfnaas-----
HAdV-A12        nvsayprpna-----------seaksqmvsltylqg---dtskpitmkvafngit-----
HAdV-A31        nvsayprpna-----------geaksqmlsqtylqg---dttkpitmkvvfngna-----
HAdV-A61        nvsayprpna-----------geaksqmlsqtylqg---dttkpitmkvvfngna-----
HAdV-C02        nllaypktqs-----------qtaknnivsqvylhg---dktkpmiltitlngtsestet
HAdV-E4         nllaypktqs-----------qtaknnivsqvylhg---dktkpmiltitlngtsestet
HAdV-C06        nlkaypktqs-----------ktaksnivsqvylng---dkskplhftitlngtde---t
HAdV-C57        nlkaypktqs-----------ktaksnivsqvylng---dkskplhftitlngtde---t
HAdV-C01        nlkaypktqs-----------rtaksnivsqvylng---ekekpmtltitlngtdenqt-
HAdV-C05        nlsaypkshg-----------ktaksnivsqvylng---dktkpvtltitlngtqetgd-
HAdV-D45        nlvayprpntpdsk-------iyarskivgnvylag---layqpivitvsfnqek-----
HAdV-D28        nitay-kpvn--sk-------syarshifgnvyida---kpynpvvikisfnqet-----
HAdV-D43        nitay-kptn--sk-------syarsvifgnvyida---kpynpvvikisfnqet-----
HAdV-D62        nltaypkptt--tk-------syarsyiygnvylga---lsyqpviikisfnqek-----
HAdV-D26        nltaypkpsaqnak-------nysrtkiisnvylga---ltyqpviitiafnqet-----
HAdV-D27        nlvaypkptsadak-------nysrskiisnvylkg---liyqpviiiasfnqet-----
HAdV-D22        nttaypkiidstt--npadkkssakkiivgnvylgg---npgqpvavaisfnkei-----
HAdV-D42        nttaypkiinntt--dpenkkssaknnidgnvyleg---nagqpvavaisfnket-----
HAdV-D15        nttayp-----------enkkssakktivgnvyleg---nagqpvavaisfnket-----
HAdV-D69        nttayp-----------enkkssakktivgnvyleg---nagqpvavaisfnket-----
HAdV-D29        nllayakattdqsk-------iyarntiygniyldn---qpynpvvikitfnnea-----
HAdV-D63        nllayakattdqsk-------iyarntiygniyldn---qpynpvvikitfnnea-----
HAdV-D70        nllayakattdqsk-------iyarntiygniyldn---qpynpvvikitfnnea-----
HAdV-D58        nllayakattdqsk-------iyarntiygniyldn---qpynpvvikitfnnea-----
HAdV-D25        nlaaypkstttqsk-------lyarntifgniylds---qaynpvvikitfnqea-----
HAdV-D67        nlaaypkstttqsk-------lyarntifgniylds---qaynpvvikitfnqea-----
HAdV-D17        nlvaypkptt-gsk-------kyardivygniylgg---layqpvvikvtfneea-----
HAdV-D10        nlvaypkps--nsk-------kyardivygtiylgg---kpdqpavikttfnqet-----
HAdV-D19        nlvaypkps--nsk-------kyardivygtiylgg---kpdqpavikttfnqet-----
HAdV-D37        nlvaypkps--nsk-------kyardivygtiylgg---kpdqpavikttfnqet-----
HAdV-D64        nlvaypkps--nsk-------kyardivygtiylgg---kpdqpavikttfnqet-----
HAdV-D65        nlvaypkpta-gsk-------kyardivygniylgg---kpdqpvtikttfnqet-----
HAdV-D56        nlvaypkpta-gsk-------kyardivygniylgg---kpdqpvtikttfnqet-----
HAdV-D59        nlvaypkpta-gsk-------kyardivygniylgg---kpdqpvtikttfnqet-----
HAdV-D09        nlvaypkptt-gsk-------kyardivygniylgg---kphqpatikttfnqet-----
HAdV-D08        nlvaypkptt-gsk-------kyardivygniylgg---kphqpvtikttfnqet-----
HAdV-D53        nlvaypkptt-gsk-------kyardivygniylgg---kphqpvtikttfnqet-----
HAdV-D54        nlvaypkptt-gsk-------kyardivygniylgg---kphqpatikttfnqet-----
HAdV-D32        nikaypkpttdts-akpedkksaakryivsnvyigg---lpdktvvitiklnae------
HAdV-D33        nikaypkpttdts-akpedkksaakryivsnvyigg---lpdktvvitiklnae------
HAdV-D24        nikaypkpttdts-akpedkksaakryivsnvyigg---lpdktvvitikfnae------
HAdV-D46        nikaypkpstdts-akpedkksaakryivsnvyigg---lpdktvvitikfnae------
HAdV-D44        sttaypkpptpptnpttpleksqaknkyvsnvylgg---qagnpvattvsfnke------
HAdV-D48        sttaypkpptpptnpttpleksqaknkyvsnvylgg---qagnpvattvsfnke------
HAdV-D30        nstaypkiinngtan-pedkksaakktivtnvylgg---dagqpvattvsfnket-----
HAdV-D49        nstaypkiinngtan-pedkksaakktivtnvylgg---daakpvattisfnket-----
HAdV-D36        sttaypkptnntst-dpdkkvsqgknkivsniylgg---evyqpgfivvkfnqet-----
HAdV-D60        nlkaypkpdtttqt-tpgdkkssgknkivsnvyfgg---evyqpgvivvafnqek-----
HAdV-D13        nlkaypkptktasd-kaenkissaknkivsnfyfgg---qayqpgtiiikfneei-----
HAdV-D38        nlkaypkptktasd-kaenkissaknkivsnfyfgg---qayqpgtiiikfneei-----
HAdV-D39        nlkaypkptktasd-kaenkvssaknkivsnfyfgg---qtyqpgtiiikfneei-----
HAdV-D51        nlkaypknttt-sst-npddkisagkknivsnvyleg---rvyqpvaltvkfnsen-----
HAdV-D23        nlkaypnpttsttn-pstdkksngknaivsnvyleg---rayqpvaititfnket-----
HAdV-D20        nlkaypkp--stvl-pstdknsngkntivsnlyleg---kayqpvavtitfnkei-----
HAdV-D47        nlkaypnpktstvl-pstdkksngkntivsnlyleg---kayqpvavtitfnket-----
                .   *                                           *
```

```
HAdV-B14        radtsyciritwswntgdapegqtsattlvtspftfyyired------
HAdV-B55        radtsyciritwswntgdapegqtsattlvtspftfyyiredd-----
HAdV-B07        nnetsycirvtwswntgvapevqtsattlvtspftfyyired------
HAdV-B11        ndetsyciritwswntgdapevqtsattlvtspftfyyired------
HAdV-B34        ssnvayaiqfewnlnasespes--niatlttspfffsyited------
HAdV-B35        ssnvayaiqfewnlnasespes--niatlttspfffsyited------
HAdV-B21        ssnvayaiqfewnlnakes-----------------------------
HAdV-B50        ssnvayaiqfewnlnakespes--niatlttspfffsyired------
HAdV-B03        dsrtsyvmtflwslnaglapet--tqatlitspftfsyired------
hadv-b66        dsrtsyvmtflwslnaglapet--tqatlitspftfsyitedd-----
HAdV-B16        asgmayamnfswslnaeeapet--tevtlitspfffsyired------
hadv-b68        asgmayamnfswslnaeeapet--tevtlitspfffsyiredd-----
HAdV-G52S       ---tgyslsfkwtavv---------rekfaapatsfcyiteq------
HAdV-F40S       ---sgyaftfkwsaep---------gkpfhpptavfcyiteq------
HAdV-F41S       ---sgyaftfkwsaep---------gkpfhpptavfcyiteq------
HAdV-G52L       ---tgysfkfiwrtvs---------rqpfdipccffsyitee------
HAdV-F40L       ---egyslkftwrvrn---------nerfdipccsfsyvteq------
HAdV-F41L       ---egyslkftwrvrn---------nerfdipccsfsyvteq------
HAdV-A18        -tvtgyslsftwtgis--nyrn----qafscpscsfsylaqe------
HAdV-A12        -slngysltfmwsgls--nyin----qpfstpscsfsyitqe------
HAdV-A31        -tvdgysltfmwtgvs--nyln----qqfstpscsfsyiaqe------
HAdV-A61        -tvdgysltfiwtgvs--nyln----qqfstpscsfsyiaqe------
HAdV-C02        sevstysmsftwswes-gkytt----etfatnsytfsyiaqe------
HAdV-E4         sevstysmsftwswes-gkytt----etfatnsytfsyiaqe------
HAdV-C06        nqvskysisfswswns-gqytn----dkfatnsytfsyiaqe------
HAdV-C57        nqvskysisfswswns-gqytn----dkfatnsytfsyisqe------
HAdV-C01        tpastysisfswswpsnqtyig----qtfatnsytfsyiaqe------
HAdV-C05        ttpsaysmsfswdwsg-hnyin----eifatssytfsyiaqe------
HAdV-D45        dascaysitfefawnk--dyvg-----qfdttsftfsyiaqe------
HAdV-D28        qnncvysisfdytcsk--eytg----mqfdvtsftfsyiaqe------
HAdV-D43        qnncvysisfdytlsk--dypn----mqfdvtsftfsyiaqe------
HAdV-D62        dvncaysitfeytctk--dyan----qqfdvssftfsyiaqe------
HAdV-D26        engcaysitftftwqk--dysa----qqfdvtsftfsyltqe------
HAdV-D27        tngcvysisfdftcsk--dytg----qqfdvtsftfsyiaqe------
HAdV-D22        --llifn-nirfawgk--ayet---pvpfdtssmtfsyiaqe------
HAdV-D42        --tadysitfdfawsk--ayet---pvpfdtssmtfsyiaqe------
HAdV-D15        --tadysitfdfawsk--ayet---pvpfdtssmtfsyiaqe------
HAdV-D69        --tadysitfdfawsk--ayet---pvpfdtssmtfsyiaqenqdkge
HAdV-D29        --dsaysitfnyswtk--dydn----ipfdstsftfsyiaqe------
HAdV-D63        --dsaysitfnyswtk--dydn----ipfdstsftfsyiaqe------
HAdV-D70        --dsaysitfnyswtk--dydn----ipfdstsftfsyiaqe------
HAdV-D58        --nsaysitfnyswtk--dydn----vpfdstsftfsyiaqe------
HAdV-D25        --dsaysitlnyswgk--dyen----ipfdstsftfsyiaqe------
HAdV-D67        --dsaysitlnyswgk--dyen----ipfdstsftfsyiaqe------
HAdV-D17        --dsaysitfefvwnk--eyar----vefettsftfsyiaqq------
HAdV-D10        --gceysitfdfswsk--tyen----vefettsftfsyiaqq------
HAdV-D19        --gceysitfnfswsk--tyen----vefettsftfsyiaqe------
HAdV-D37        --gceysitfnfswsk--tyen----vefettsftfsyiaqe------
HAdV-D64        --gceysitfnfswsk--tyen----vefettsftfsyiaqe------
HAdV-D65        --gceysitfdfswak--tyvn----vefettsftfsyiaqe------
HAdV-D56        --gceysitfdfswak--tyvn----vefettsftfsyiaqe------
HAdV-D59        --gceysitfdfswak--tyvd----vefettsftfsyiaqe------
HAdV-D09        --gceysitfdfswak--tyvn----vefettsftfsyiaqe------
HAdV-D08        --gceysitfdfswak--tyvn----vefettsftfsyiaqe------
HAdV-D53        --gceysitfdfswak--tyvn----vefettsftfsyiaqe------
HAdV-D54        --gceysitfdfswak--tyvn----vefettsftfsyiaqe------
HAdV-D32        -tesaysmtfeftwak--tfen----lqfdsssftfsyiaqe------
HAdV-D33        -tesaysmtfeftwak--tfen----lqfdsssftfsyiaqe------
HAdV-D24        -tecaysitfeftwak--tfed----vqfdsssftfsyiaqe------
HAdV-D46        -tecaysitfeftwak--tfed----vqfdsssftfsyiaqe------
HAdV-D44        -tgctysitfdfawnk--tyen----vqfdssfltfsyiaqe------
HAdV-D48        -tgctysitfdfawnk--tyen----vqfdssfltfsyiaqe------
HAdV-D30        esncvysitfdfawnk--tykn----vpfdsssltfsyiaqd------
HAdV-D49        esncvysitfdfawnk--tykn----vpfdsssltfsyiaqd------
HAdV-D36        dancaysitfdfgwgk--vykd---pipydtssftfsyiaqe------
HAdV-D60        eancaysitlkfgwgk--tyet---pipfdtssftfsyiaqenedkeq
HAdV-D13        detcaysitfnfgwgk--vydn---pfpfdttsftfsyiaqenedkd-
HAdV-D38        detcaysitfnfgwgk--vydn---pfpfdttsftfsyiaqe------
HAdV-D39        ddtcaysitfnfgwgk--tydn---pfpfdttsftfsyiaqe------
HAdV-D51        --dcaysitfdfvwsk--tyes---pvafdsssftfsyiaqe------
HAdV-D23        --gctysmtfdfgwsk--vynd---pipfdtssltfsyiaqe------
HAdV-D20        --gctysitfdfgwak--tydv---pipfdsssftfsyiaqe------
HAdV-D47        --gctysitfefgwak--tydv---pipfdsssftfsyiaqe------
```

6.2.2. Fiber Protein Total Lengths (Types C01 – D70)

HadV-No	Species	Length
3	B	319
66	B	319
21	B	323
34	B	323
35	B	323
50	B	323
7	B	325
11	B	325
55	B	325
16	B	353
68	B	353
28	D	359
43	D	359
62	D	360
9	D	361
54	D	361
8	D	362
14	B	362
53	D	362
56	D	362
69	D	362
65	D	362
59	D	362
52S	G	363
19	D	365
37	D	365
64	D	365
17	D	366
10	D	367
15	D	367
27	D	367
45	D	369
25	D	371
29	D	371
36	D	371
42	D	371
58	D	371

HadV-No	Species	Length
67	D	371
63	D	371
70	D	371
44	D	373
48	D	373
26	D	374
60	D	376
20	D	379
13	D	381
22	D	381
23	D	381
24	D	381
38	D	381
46	D	381
32	D	382
33	D	382
39	D	382
47	D	382
51	D	383
30	D	385
49	D	385
40S	F	387
41S	F	387
4	E	426
6	C	528
57	C	528
40L	F	547
41L	F	547
61	A	555
31	A	556
52L	G	560
5	C	581
1	C	582
2	C	582
18	A	586
12	A	587

6.2.3. Penton Base (HAdV Types C01 – D70)

```
HAdV-C05       ------mrraamyeegpppsyesvvsaapvaaalgspfdapldppfvpprylrptggrns 54
HAdV-C02       ------mqraamyeegpppsyesvvsaapvaaalgspfdapldppfvpprylrptggrns 54
HAdV-C57       ------mrraamyeegpppsyesvvsaapvaaalgspfdapldppfvpprylrptggrns 54
HAdV-C01       ------mrraamyeegpppsyesvvsaapvaaalgspfdapldppfvpprylrptggrns 54
HAdV-C06       ------mrraamyeegpppsyesvvsaapvaaalgspfdapldppfvpprylrptggrns 54
HAdV-D13       -mr------rav--vspppsyesvmaqa------------tlevpfvpprymaptegrns 39
HAdV-D37       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D53       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D69       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D60*      -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D20       -mr------ramvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D28       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D49       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D32       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D44       -mr------rav--vspppsyesvmaqa------------tlevpfvpprymaptegrns 39
HAdV-D65       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D08       -mr------ra-vvsspppsyesvmaqa------------tlevpfvpprymaptegrns 40
HAdV-D58       -mr------r--avvspppsyesvmaqa------------tlevpfvpprymaptegrns 39
HAdV-D38       -mr------rav--vspppsyesvmaqa------------tlevpfvpprymaptegrns 39
HAdV-D70       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D54       -mr------ra-vvsspppsyesvmaqa------------tlevpfvpprymaptegrns 40
HAdV-D67       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D45       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D17       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D64       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D22       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D42       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D59       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D39       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D09       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D10       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D56       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D26       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D47       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D25       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D15       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D46       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D36       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D29       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D43       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D33       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D51       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D30       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D63       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D19       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D23       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D48       -mr------ravv--spppsyesvmaqa------------tlevpfvpprymaptegrns 39
HAdV-D24       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D27       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-D62       -mr------ravvsssppppsyesvmaqa------------tlevpfvpprymaptegrns 41
HAdV-F40*      mrravgvppvmayaegpppsyesvmeta--------dlpatlqalhvpprylgptegrns 52
HAdV-F41*      mrravgvppvmayaegpppsyesvmgsa--------dspatlealyvpprylgptegrns 52
HAdV-A18       -mrravelqpvafaeapppsyetvmaa----a---aaqpstleapyvpprylgptegrns 52
HAdV-A12       -mrravelqtvafpetpppsyetvmaa---------a------ppyvpprylgptegrns 44
HAdV-A31       -mrravelqtvafpeapppsyetvmaa---------aqtsaleapyvpprylaptegrns 50
HAdV-A61       -mrravelqtvafpeapppsyetvmaa---------aqtsaleapyvpprylaptegrns 50
HadV-B16       mrrravlggavvypegpppsyesvmqqq---a---amiqppleapfvpprylaptegrns 54
HAdV-B68       mrrravlggavvypegpppsyesvmqqq---a---amiqppleapfvpprylaptegrns 54
HAdV-B03       mrrravlggavvypegpppsyesvmqqq---a---amiqppleapfvpprylaptegrns 54
HAdV-B66       mrrravlggavvypegpppsyesvmqqq---a---amlqppleapfvpprylaptegrns 54
HAdV-B07       mrrravlggavvypegpppsyesvmqqq---a---amiqpplevpfvpprylaptegrns 54
HAdV-E04       mm-------rraypegpppsyesvmqqamaaa---aamqppleapyvpprylaptegrns 50
HAdV-B21       mmrrtvlggavvypegpppsyesvmqqa-a-a---atmqppleapfvpprylaptegrns 55
HAdV-B50       -mrrtvlggavvypegpppsyesvmqqa-a-a---aamqppleapfvpprylaptegrns 54
HAdV-B34       -mrrvvlggavvypegpppsyesvmqqqqa-t---avmqspleapfvpprylaptegrns 55
HadV-B11       -mrrvvlggavvypegpppsyesvmqqqqa-t---avmqspleapfvpprylaptegrns 55
HAdV-B35       -mrrvvlggavvypegpppsyesvmqqqqa-t---avmqspleapfvpprylaptegrns 55
HAdV-B14       -mrrvvlggavvypegpppsyesvmqqqqa-t---avmqspleapfvpprylaptegrns 55
HAdV-B55       -mrrvvlggavvypegpppsyesvmqq-qa-t---avmqspleapfvpprylaptegrns 54
                              ******:*:                    .*****: ** ****
```

```
HAdV-C05        iryselaplfdttrvylvdnkstdvaslnyqndhsnflttviqnndyspgeastqtinld 114
HAdV-C02        iryselaplfdttrvylvdnkstdvaslnyqndhsnflttviqnndyspgeastqtinld 114
HAdV-C57        iryselaplfdttrvylvdnkstdvaslnyqndhsnflttviqnndyspgeastqtinld 114
HAdV-C01        iryselaplfdttrvylvdnkstdvaslnyqndhsnflttviqnndyspgeastqtinld 114
HAdV-C06        iryselaplfdttrvylvdnkstdvaslnyqndhsnflttviqnndyspgeastqtinld 114
HAdV-D13        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 99
HAdV-D37        iryselaplydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D53        iryselaplydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D69        iryselaplydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D60*       iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D20        iryselaplydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D28        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D49        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D32        iryselaplydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D44        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 99
HAdV-D65        iryselaplydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D08        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 100
HAdV-D58        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 99
HAdV-D38        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 99
HAdV-D70        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D54        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndfttaeastqtinfd 100
HAdV-D67        iryselaplydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D45        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D17        iryselaplydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D64        iryselaplydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D22        iryselaplydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D42        iryselaplydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D59        iryselaplydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D39        iryselaplydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D09        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D10        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D56        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D26        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D47        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D25        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D15        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D46        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D36        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D29        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D43        iryselaplydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D33        iryselaplydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D51        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D30        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D63        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D19        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D23        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D48        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 99
HAdV-D24        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D27        iryselapqydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-D62        iryselaplydttrvylvdnksadiaslnyqndhsnflttvvqnndftpaeastqtinfd 101
HAdV-F40*       iryselaplydttrvylvdnksadiaslnyqndhsnfqttvvqnndftpteagtqtinfd 112
HAdV-F41*       iryselaplydttrvylvdnksadiaslnyqndhsnfqttvvqnndftpaeagtqtinfd 112
HAdV-A18        iryselsplydttrvylvdnkssdiaslnyqndhsnflttvvqnndyspieagtqtinfd 112
HAdV-A12        iryselsplydttrvylvdnkssdiaslnyqndhsnflttvvqnndyspieagtqtinfd 104
HAdV-A31        iryselsplydttrvylvdnkssdiaslnyqndhsnflttvvqnndyspieagtqtinfd 110
HAdV-A61        iryselsplydttrvylvdnkssdiaslnyqndhsnflttvvqnndyspieagtqtinfd 110
HadV-B16        iryselspqydttklylvdnksadiaslnyqndhsnflttvvqnndftpteastqtinfd 114
HAdV-B68        iryselspqydttklylvdnksadiaslnyqndhsnflttvvqnndftpteastqtinfd 114
HAdV-B03        irysdvsplydttklylvdnksadiaslnyqndhsnflttvvqnndftpteastqtinfd 114
HAdV-B66        iryselsplydttklylvdnksadiaslnyqndhsnflttvvqnndftpteastqtinfd 114
HAdV-B07        iryselsplydttklylvdnksadiaslnyqndhsnflttvvqnndftpteastqtinfd 114
HAdV-E04        iryselaplydttrlylvdnksadiaslnyqndhsnflttvvqnndftpteastqtinfd 110
HAdV-B21        iryselaplydttrlylvdnksadiaslnyqndhsnflttvvqnndftpteastqtinfd 115
HAdV-B50        iryselaplydttrlylvdnksadiaslnyqndhsnflttvvqnndftpteastqtinfd 114
HAdV-B34        iryselapqydttrlylvdnksadiaslnyqndhsnflttvvqnndftpteastqtinfd 115
HadV-B11        iryselapqydttrlylvdnksadiaslnyqndhsnflttvvqnndftpteastqtinfd 115
HAdV-B35        iryselapqydttrlylvdnksadiaslnyqndhsnflttvvqnndftpteastqtinfd 115
HAdV-B14        iryselapqydttrlylvdnksadiaslnyqndhsnflttvvqnndftpteastqtinfd 115
HAdV-B55        iryselapqydttrlylvdnksadiaslnyqndhsnflttvvqnndftpteastqtinfd 114
                ****:::* :***::*******:*:************ ***:****::  **.*****:*
```

```
HAdV-C05        drshwggdlktilhtnmpnvnefmftnkfkarvmvsrlptk---------dnqvelkyew 165
HAdV-C02        drshwggdlktilhtnmpnvnefmftnkfkarvmvsrsltk---------dkqvelkyew 165
HAdV-C57        drshwggdlktilhtnmpnvnefmftnkfkarvmvsrsltk---------dkqvelkyew 165
HAdV-C01        drshwggdlktilhtnmpnvnefmftnkfkarvmvsrsltk---------dkqvelkyew 165
HAdV-C06        drshwggdlktilhtnmpnvnefmftnkfkarvmvsrsltk---------dkqvelkyew 165
HAdV-D13        ersrwggdlktilhtnmpnvneymftskfkarvmvarkhpegve---atdlskdileydw 156
HAdV-D37        ersrwggdlktilhtnmpnvneymftskfkarvmvarkkaegad---andrskdileyqw 158
HAdV-D53        ersrwggdlktilhtnmpnvneymftskfkarvmvarkkaegad---andrskdileyqw 158
HAdV-D69        ersrwggdlktilhtnmpnvneymftskfkarvmvarkkaegad---andrskdileyqw 158
HAdV-D60*       ersrwggdlktilhtnmpnvneymftskfkarvmvarkkaegad---andrskdileyqw 158
HAdV-D20        ersrwggdlktilhtnmpnvneymftskfkarvmvarkhpdgvd---kddlskdileyew 158
HAdV-D28        ersrwggdlktilhtnmpnvneymftskfkarvmvarkhpkdvd---asdlskdileydw 158
HAdV-D49        ersrwggdlktilhtnmpnvneymftskfkarvmvsrkrpega-----tdasqdilkyew 156
HAdV-D32        ersrwggdlktilhtnmpnvneymftskfkarvmvarkhpkdvq--dpndrskdileykw 159
HAdV-D44        ersrwggdlktilhtnmpnvneymftskfkarvmvarkhpkdvq--dpndrskdileykw 157
HAdV-D65        ersrwggdlktilhtnmpnvneymftskfkarvmvarkkaega---dandrskdileyew 158
HAdV-D08        ersrwggdlktilhtnmpnvneymftskfkarvmvarkhpqgv---aetdlsqdkleyew 157
HAdV-D58        ersrwggdlktilhtnmpnvneymftskfkarvmvarkhpegv---vetdlsqdkleyqw 156
HAdV-D38        ersrwggdlktilhtnmpnvneymftskfkarvmvsrkkaegad---andpskdileyqw 156
HAdV-D70        ersrwggdlktilhtnmpnvneymftskfkarvmvarkkaegad---andrskdileyqw 158
HAdV-D54        ersrwggdlktilhtnmpnvneymftskfkarvmvarkhpegvd---ntdlsqdkleyqw 157
HAdV-D67        ersrwggdlktilhtnmpnvneymftskfkarvmvarkhpenva---kedlsqdileykw 158
HAdV-D45        ersrwggdlktilhtnmpnvneymftskfkarvmvsrkv---vd---endrskdileyqw 155
HAdV-D17        ersrwggdlktilhtnmpnvneymftskfkarvmvarkhpqgve---atdlskdileyew 158
HAdV-D64        ersrwggdlktilhtnmpnvneymftskfkarvmvarkhpqgve---atdlskdileyqw 158
HAdV-D22        ersrwggdlktilhtnmpnvneymftskfkarvmvarkkaegad---andrskdileyqw 158
HAdV-D42        ersrwggdlktilhtnmpnvneymftskfkarvmvarkkaegad---andrskdileyqw 158
HAdV-D59        ersrwggdlktilhtnmpnvneymftskfkarvmvarkhpkdvp---vndlskdileyew 158
HAdV-D39        ersrwggdlktilhtnmpnvneymftskfkarvmvarkhpdgvg---kddlskdileyew 158
HAdV-D09        ersrwggdlktilhtnmpnvneymftskfkarvmvarkhpegvv---etdlsqdkleyew 158
HAdV-D10        ersrwggdlktilhtnmpnvneymftskfkarvmvarkhpegvv---etdlsqdkleyew 158
HAdV-D56        ersrwggdlktilhtnmpnvneymftskfkarvmvarkhpegvv---etdlsqdkleyew 158
HAdV-D26        ersrwggdlktilhtnmpnvneymftskfkarvmvsrkhpegvv---etdlsqdkleyew 158
HAdV-D47        ersrwggdlktilhtnmpnvnefmftskfkarvmvarkhpegvd---ktdlsqdkleyew 158
HAdV-D25        ersrwggdlktilhtnmpnvneymftskfkarvmvarkhpenvd---ktdlsqdkleyew 158
HAdV-D15        ersrwggdlktilhtnmpnvneymftskfkarvmvarkhpenva---kedlsqdileykw 158
HAdV-D46        ersrwggdlktilhtnmpnvneymftskfkarvmvarkhpenva---kedlsqdileykw 158
HAdV-D36        ersrwggdlktilhtnmpnvneymftskfkarvmvarkhpkdvd---asdlskdileykw 158
HAdV-D29        ersrwggdlktilhtnmpnvneymftskfkarvmvarkhpqgve---atdlskdileyew 158
HAdV-D43        ersrwggdlktilhtnmpnvneymftskfkarvmvarkhpqgve---atdlskdileykw 158
HAdV-D33        ersrwggdlktilhtnmpnvneymftskfkarvmvarkhpkdvd---asdlskdileydw 158
HAdV-D51        ersrwggdlktilhtnmpnvneymftskfkarvmvarkhpkdvq---dddhskdileydw 158
HAdV-D30        ersrwggdlktilhtnmpnvneymftskfkarvmvsrkkaegad---endrskdileyqw 158
HAdV-D63        ersrwggdlktilhtnmpnvneymftskfkarvmvsrkkaegad---endrskdileyqw 158
HAdV-D19        ersrwggdlktilhtnmpnvneymftskfkarvmvsrkrpegat-----dasqdilkyew 156
HAdV-D23        ersrwggdlktilhtnmpnvneymftskfkarvmvsrkrpegat-----dasqdileykw 156
HAdV-D48        ersrwggdlktilhtnmpnvneymftskfkarvmvsrkrpegat-----dasqdilkyew 154
HAdV-D24        ersrwggdlktilhtnmpnvneymftskfkarvmvarkvv---d---qndrskdelkyew 155
HAdV-D27        ersrwggdlktilhtnmpnvneymftskfkarvmvarkvv---d---qndrskdelkyew 155
HAdV-D62        ersrwggdlktilhtnmpnvneymftskfkarvmvarkkaegad---addrskdileyew 158
HAdV-F40*       drsrwggdlktilrtnmpninefmstnkfrarvmvekvn-----------rktnapryew 161
HAdV-F41*       ersrwgadlktilrtnmpninefmstnkfkarlmvekkn-----------ketglpryew 161
HAdV-A18        ersrwggdlktilhtnmpnvndfmfttkfkakvmvarekn-------kedpskdilkyew 165
HAdV-A12        ersrwggdlktilhtnmpnvndfmfttkfkarvmvarktn-------ne--gqtileyew 155
HAdV-A31        ersrwggdlktilhtnmpnvndfmfttkfkarvmvsrkkd-------sedqskdilkyew 163
HAdV-A61        ersrwggdlktilhtnmpnvndfmfttkfkarvmvsrkkd-------sedqskdilkyew 163
HadV-B16        ersrwggqlktimhtnmpnvneymfsnkfkarvmvsrkapegvtvnd---hkddilkyew 171
HAdV-B68        ersrwggqlktimhtnmpnvneymfsnkfkarvmvsrkapegvtvnd---hkddilkyew 171
HAdV-B03        ersrwggqlktimhtnmpnvneymfsnkfkarvmvsrkapegvtvndtydhkedilkyew 174
HAdV-B66        ersrwggqlktimhtnmpnvneymfsnkfkarvmvsrkapegvivndtydhkedilkyew 174
HAdV-B07        ersrwgghlktimhtnmpnvneymfsnkfkarvmvsrkapegvtvndtydhkedilkyew 174
HAdV-E04        ersrwggqlktimhtnmpnvnefmysnkfkarvmvsrktpngvtvgddydgsqdelkyew 170
HAdV-B21        ersrwggqlktimhtnmpnvneymfsnkfkarvmvsrkapegvtvddnydhkqdileyew 175
HAdV-B50        ersrwggqlktimhtnmpnvneymfsnkfkarvmvsrkapegvtvddnydhkqdileyew 174
HAdV-B34        ersrwggqlktimhtnmpnvneymfsnkfkarvmvsrkppegvrvddnydhkqdilkyew 175
HadV-B11        ersrwggqlktimhtnmpnvneymfsnkfkarvmvsrkppdgaavgdtydhkqdilkyew 175
HAdV-B35        ersrwggqlktimhtnmpnvneymfsnkfkarvmvsrkppdgaavgdtydhkqdileyew 175
HAdV-B14        ersrwggqlktimhtnmpnvneymfsnnfkarvmvsrkppegaavgdtydhkqdileyew 175
HAdV-B55        ersrwggqlktimhtnmpnvneymfsnnfkarvmvsrkppegaavgdtydhkqdileyew 174
                :**:**..****::*****:*::* :.:*:*::** :                 .*.*
```

```
HAdV-C05        veftlpegnysetmtidlmnnaivehylkvgrqngvlesdigvkfdtrnfrlgfdpvtgl 225
HAdV-C02        veftlpegnysetmtidlmnnaivehylkvgrqngvlesdigvkfdtrnfrlgfdpvtgl 225
HAdV-C57        veftlpegnysetmtidlmnnaivehylkvgrqngvlesdigvkfdtrnfrlgfdpvtgl 225
HAdV-C01        veftlpegnysetmtidlmnnaivehylkvgrqngvlesdigvkfdtrnfrlgfdpvtgl 225
HAdV-C06        veftlpegnysetmtidlmnnaivehylkvgrqngvlesdigvkfdtrnfrlgfdpvtgl 225
HAdV-D13        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 216
HAdV-D37        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D53        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D69        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D60*       feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D20        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D28        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D49        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfrlgwdpetkl 216
HAdV-D32        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfrlgwdpetkl 219
HAdV-D44        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfrlgwdpetkl 217
HAdV-D65        feftlpegnfsetmtidlmnnailenylqvgrqhgvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D08        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 217
HAdV-D58        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 216
HAdV-D38        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfrlgwdpetkl 216
HAdV-D70        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D54        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 217
HAdV-D67        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D45        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 215
HAdV-D17        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D64        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D22        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D42        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D59        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D39        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D09        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D10        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D56        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D26        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D47        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D25        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D15        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D46        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D36        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfrlgwdpvtkl 218
HAdV-D29        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D43        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D33        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D51        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D30        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D63        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-D19        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 216
HAdV-D23        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 216
HAdV-D48        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 214
HAdV-D24        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 215
HAdV-D27        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 215
HAdV-D62        feftlpegnfsetmtidlmnnailenylqvgrqngvlesdigvkfdsrnfklgwdpvtkl 218
HAdV-F40*       feftlpegnysetmtidlmnnaivdnylavgrqngvlesdigvkfdtrnfrlgwdpvtkl 221
HAdV-F41*       feftlpegnysetmtidlmnnaivdnylevgrqngvlesdigvkfdtrnfrlgwdpvtkl 221
HAdV-A18        aefvlpegnysetmtidlmnnaiiehyllvgrqngvlesdigvkfdtrnfrlgwdpetkl 225
HAdV-A12        aefvlpegnysetmtidlmnnaiiehylrvgrqhgvlesdigvkfdtrnfrlgwdpetql 215
HAdV-A31        aefvlpegnysetmtidlmnnaiiehylrvgrqngvlesdigvkfdtrnfrlgwdpetkl 223
HAdV-A61        aefvlpegnysetmtidlmnnaiiehylrvgrqngvlesdigvkfdtrnfrlgwdpetkl 223
HadV-B16        feftlpegnfsatmtidlmnnaiidnylkigrqngvlesdigvkfdtrnfrlgwdpetkl 231
HAdV-B68        feftlpegnfsatmtidlmnnaiidnylkigrqngvlesdigvkfdtrnfrlgwdpetkl 231
HAdV-B03        fefilpegnfsatmtidlmnnaiidnyleigrqngvlesdigvkfdtrnfrlgwdpetkl 234
HAdV-B66        feftlpegnfsatmtidlmnnaiidnyleigrqngvlesdigvkfdtrnfrlgwdpetkl 234
HAdV-B07        fefilpegnfsatmtidlmnnaiidnyleigrqngvlesdigvkfdtrnfrlgwdpetkl 234
HAdV-E04        vefelpegnfsvtmtidlmnnaiidnylavgrqngvlesdigvkfdtrnfrlgwdpvtel 230
HAdV-B21        feftlpegnfsatmtidlmnnaiidnylevgrqngvlesdigvkfdtrnfrlgwdpetkl 235
HAdV-B50        feftlpegnfsatmtidlmnnaiidnylevgrqngvlesdigvkfdtrnfrlgwdpetkl 234
HAdV-B34        feftlpegnfsvtmtidlmnnaiidnylkvgrqngvlesdigvkfdtrnfklgwdpetkl 235
HadV-B11        feftlpegnfsvtmtidlmnnaiidnylkvgrqngvlesdigvkfdtrnfklgwdpetkl 235
HAdV-B35        feftlpegnfsvtmtidlmnnaiidnylkvgrqngvlesdigvkfdtrnfklgwdpetkl 235
HAdV-B14        feftlpegnfsvtmtidlmnnaiidnylkvgrqngvlesdigvkfdtrnfklgwdpetkl 235
HAdV-B55        feftlpegnfsvtmtidlmnnaiidnylkvgrqngvlesdigvkfdtrnfklgwdpetkl 234
                 ** *****:* ***********:::** :***:************:***:**:** * *
```

```
HAdV-C05        vmpgvytneafhpdiillpgcgvdfthsrlsnllgirkrqpfqegfrityddleggnipa 285
HAdV-C02        vmpgvytneafhpdiillpgcgvdfthsrlsnllgirkrqpfqegfrityddleggnipa 285
HAdV-C57        vmpgvytneafhpdiillpgcgvdfthsrlsnllgirkrqpfqegfrityddleggnipa 285
HAdV-C01        vmpgvytneafhpdiillpgcgvdfthsrlsnllgirkrqpfqegfrityddleggnipa 285
HAdV-C06        vmpgvytneafhpdiillpgcgvdfthsrlsnllgirkrqpfqegfrityddleggnipa 285
HAdV-D13        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedlvggnipa 276
HAdV-D37        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedlvggnipa 278
HAdV-D53        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedlvggnipa 278
HAdV-D69        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedlvggnipa 278
HAdV-D60*       vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 278
HAdV-D20        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedlvggnipa 278
HAdV-D28        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedlvggnipa 278
HAdV-D49        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 276
HAdV-D32        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 279
HAdV-D44        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 277
HAdV-D65        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 278
HAdV-D08        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 277
HAdV-D58        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 276
HAdV-D38        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 276
HAdV-D70        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedlvggnipa 278
HAdV-D54        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 277
HAdV-D67        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 278
HAdV-D45        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 275
HAdV-D17        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 278
HAdV-D64        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 278
HAdV-D22        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 278
HAdV-D42        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 278
HAdV-D59        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 278
HAdV-D39        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 278
HAdV-D09        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 278
HAdV-D10        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 278
HAdV-D56        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 278
HAdV-D26        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 278
HAdV-D47        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 278
HAdV-D25        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 278
HAdV-D15        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 278
HAdV-D46        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfkimyedleggnipa 278
HAdV-D36        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 278
HAdV-D29        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 278
HAdV-D43        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 278
HAdV-D33        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 278
HAdV-D51        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 278
HAdV-D30        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 278
HAdV-D63        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 278
HAdV-D19        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 276
HAdV-D23        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 276
HAdV-D48        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 274
HAdV-D24        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 275
HAdV-D27        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 275
HAdV-D62        vmpgvytyeafhpdvvllpgcgvdftesrlsnllgirkkqpfqegfrimyedleggnipa 278
HAdV-F40*       vmpgvytneafhpdivllpgcgvdftqsrlnnllgirkrmpfqkgfqimyedleggnipa 281
HAdV-F41*       vmpgvytneafhpdivllpgcgvdftqsrlsnllgirkrlpfqegfqimyedleggnipa 281
HAdV-A18        itpgvytneafhpdivllpgcgvdftesrlsnvlgirkrqpfqegfiimyedleggnipa 285
HAdV-A12        vtpgvytneafhpdivllpgcgvdftesrlsnilgirkrqpfqegfvimyehleggnipa 275
HAdV-A31        vtpgvytneafhpdivllpgcgvdftesrlsnmlgirkrqpfqegfvimyedleggnipa 283
HAdV-A61        vtpgvytneafhpdivllpgcgvdftesrlsnmlgirkrqpfqegfvimyedleggnipa 283
HadV-B16        impgvytyeafhpdivllpgcgvdftesrlsnllgirkrhpfqegfkimyedleggnipa 291
HAdV-B68        impgvytyeafhpdivllpgcgvdftesrlsnllgirkrhpfqegfkimyedleggnipa 291
HAdV-B03        impgvytyeafhpdivllpgcgvdftesrlsnllgirkrhpfqegfkimyedleggnipa 294
HAdV-B66        impgvytyeafhpdivllpgcgvdftesrlsnllgirkrhpfqegfkimyedleggnipa 294
HAdV-B07        impgvytyeafhpdivllpgcgvdftesrlsnllgirkrhpfqegfkimyedleggnipa 294
HAdV-E04        vmpgvytneafhpdivllpgcgvdftesrlsnmlgirkrqpfqegfqimyedldggnipa 290
HAdV-B21        impgvytyeafhpdivllpgcgvdftesrlsnllgirkrhpfqegfkilyedleggnipa 295
HAdV-B50        impgvytyeafhpdivllpgcgvdftesrlsnllgirkrhpfqegfkilyedleggnipa 294
HAdV-B34        impgvytyeafhpdivllpgcgvdftesrlsnllgirkkhpfqegfkilyedleggnipa 295
HadV-B11        impgvytyeafhpdivllpgcgvdftesrlsnllgirkkqpfqegfkilyedleggnipa 295
HAdV-B35        impgvytyeafhpdivllpgcgvdftesrlsnllgirkkqpfqegfkilyedleggnipa 295
HAdV-B14        impgvytyeafhpdivllpgcgvdftesrlsnllgirkkqpfqegfkilyedleggnipa 295
HAdV-B55        impgvytyeafhpdivllpgcgvdftesrlsnllgirkkqpfqegfkilyedleggnipa 294
                : ***** ******::**********.***.*:*****: ***:** * *:.* ******
```

```
HAdV-C05        lldvdayqaslkddteqggggaggsnssgsgaeensnaaaaamqpvedmndhairgdtfa 345
HAdV-C02        lldvdayqaslkddteqggdgagggnnsgsgaeensnaaaaamqpvedmndhairgdtfa 345
HAdV-C57        lldvdayqaslkddteqggggagggnnsgsgaeensnaaaaamqpvddmndhairgdtfa 345
HAdV-C01        lldvdayqaslkddteqggggagggnnsgsgaeensnaaaaamqpvedmndhairgdtfa 345
HAdV-C06        lldvdayqaslkddteqggggagggnnsgsgaeensnaaaaamqpvedmndhairgdtfa 345
HAdV-D13        lldvkkylkdkeeagtt--dant----------------------ikaqndavprgdnfa 312
HAdV-D37        llnvkeylkdkeeagka--dant----------------------ikaqndavprgdnya 314
HAdV-D53        llnvkeylkdkeeagka--dant----------------------ikaqndavprgdnya 314
HAdV-D69        llnvkeylkdkeeagta--dant----------------------ikaqndavprgdnya 314
HAdV-D60*       llnvkeylkdkeeagt------------------------------------------- 294
HAdV-D20        llnvkeylkdkeeagta--dant----------------------ikaqndavprgdnya 314
HAdV-D28        llnvkeylkdkeeagta--dant----------------------ikaqndavprgdnya 314
HAdV-D49        lldveaylkskndleeatknanr----------------------aaangggetrgdtfl 314
HAdV-D32        lldvdaylkskndleeatknanr----------------------aaangggetrgdtfl 317
HAdV-D44        lldvdaylkskndleeatknanr----------------------aaangggetrgdtfl 315
HAdV-D65        lldveaylnskndkeeatknanr----------------------dadngggetrgdtfl 316
HAdV-D08        lldveaylkskndreeatqnanr----------------------vaangggeirgdtfl 315
HAdV-D58        lldveaylnskndmeeatknanr----------------------aadngggetrgdtfl 314
HAdV-D38        lldvpkyleskkkleegakeagn----------------------t----kapirgdtya 310
HAdV-D70        lldvpkyleskkkleegakeagn----------------------t----kapirgdtya 312
HAdV-D54        llnvtkyleskktlqkavenaak----------------------v----napargdssv 311
HAdV-D67        lldvakyleskkkveeaikkaae----------------------t----ngtprgdsdv 312
HAdV-D45        lldvakyleskkkleeavenaak----------------------a----ngpargdssv 309
HAdV-D17        lldvpkyleskkkleealenaak----------------------a----ngpargdssv 312
HAdV-D64        llnvtkyleskkklee---naak----------------------a----ngpargdssv 309
HAdV-D22        llnvtkyleskkkleeavenaak----------------------a----ngpargdssv 312
HAdV-D42        llnvtkyleskkkleeavenaak----------------------a----ngpargdssv 312
HAdV-D59        llnvtkyleskkkleeavenaak----------------------a----ngpargdssv 312
HAdV-D39        lldvpkyleskkkvedetknaaa----------------------at-ad-tttrgdtfa 314
HAdV-D09        lldvpkyleskkkvedetknaaa----------------------at-ad-tttrgdtfa 314
HAdV-D10        lldvpkyleskkkvedetknaaa----------------------at-ad-tttrgdtfa 314
HAdV-D56        lldvpkyleskkkvedetknaaa----------------------at-ad-tttrgdtfa 314
HAdV-D26        lldvpkyleskkkvedetknaaa----------------------at-ad-tttrgdtfa 314
HAdV-D47        llnvtkyleskkqmedaakeaak----------------------an----apirgdsdv 312
HAdV-D25        lldtkkyldskkeledaakeaak----------------------qq-gdgavtrgdthl 315
HAdV-D15        lldtkkyldskkeledaakeaak----------------------qq-gdgavtrgdthl 315
HAdV-D46        lldtkkyleskkeledaakeaak----------------------qq-gdgavtrgdthl 315
HAdV-D36        lldtkkyldskkeledaakeaak----------------------qq-gdgavtrgdthl 315
HAdV-D29        lldtkkyldskkeledaakeaak----------------------hq-gdgavtrgdthl 315
HAdV-D43        lldtkkyldskkeledaakeaak----------------------qq-gdgavtrgdthl 315
HAdV-D33        lldtkkyldskkeiedakqkaaq----------------------a----ggeirgdsad 312
HAdV-D51        lldtkkyldskkdiedakqkaaq----------------------p----ggeirgdsad 312
HAdV-D30        lldtkkyldskkdiedakqkaaq----------------------a----ggeirgdsad 312
HAdV-D63        lldtkkyldskkdiedakqkaaq----------------------a----ggeirgdsad 312
HAdV-D19        lldtkkyldskkeiedakqkaaq----------------------p----ggeirgdsad 310
HAdV-D23        lldtkkyldskkeiedakqkaaq----------------------a----ggeirgdsad 310
HAdV-D48        lldvkqyldskkkleeatqnatr----------------------a----agdirgdshi 308
HAdV-D24        lldvkqyldskkkleeatqnatr----------------------a----agdirgdshi 309
HAdV-D27        lldvkqyldskkkleeatqnatr----------------------a----agdirgdshi 309
HAdV-D62        lldvkqyldskkkleeatqnatr----------------------a----agdirgdtyv 312
HAdV-F40*       lldvekyeasikea-------------------------------------qeirgadfk 304
HAdV-F41*       lldvakyeasiqkakeeg---------------------------------keigddtfa 308
HAdV-A18        lldvkkyedslknes------------------------------------airgdnft 308
HAdV-A12        lldvkkyenslqdqn------------------------------------tvrgdnfi 298
HAdV-A31        lldvkkyedslqpdn------------------------------------tvrgdnfi 306
HAdV-A61        lldvkkyedslqpdg------------------------------------tvrgdnfi 306
HadV-B16        lldvtayeeskkdtttetgekavv------------ktttvavaeetseddnitrgdtyi 339
HAdV-B68        lldvtayeeskkdtttetgekavv------------ktttvavaeetseddnitrgdtyi 339
HAdV-B03        lldvtayeeskkdtttet--------------------ttlavaeetsedditrgdtyi 334
HAdV-B66        lldvtayeeskkdtttet--------------------ttlavaeetseddnitrgdtyi 334
HAdV-B07        lldvtayeeskkdtttet--------------------ttlavaeetseddnitrgdtyi 334
HAdV-E04        lldveayekskeesvaastaav-----------------------atastevrgdnfa 325
HAdV-B21        lldveayenskkeqeakteaakaaa---------iakanivvsdpvrvanaeevrgdnyt 346
HAdV-B50        lldveayknskkereakteaakaaa---------iakanivvsdpvrvanaeevrgdnyt 345
HAdV-B34        lldvdayenskkdqkakieaa---a---------eakanivandpvrvanaseirgdsfa 343
HadV-B11        lldvdayenskkeqkakieaataaa---------eakanivasdstrvanagevrgdnfa 346
HAdV-B35        lldvdayenskkeqkakieaataaa---------eakanivasdstrvanagevrgdnfa 346
HAdV-B14        lldvdayenskkeqkakieaa---a---------eakanivasdstrvanagevrgdnfa 343
HAdV-B55        lldvdayenskkeqkakieaa---a---------eakanivasdstrvanagevrgdnfa 342
                **:.  *  . :
```

```
HAdV-C05        traeekraeaeaaaeaaap---aaqpevekpqkkpvikpltedskkrsynlisndstftq 402
HAdV-C02        traeekraeaeaaaeaaap---aaqpevekpqkkpvikpltedskkrsynlisndstftq 402
HAdV-C57        traeekraeaeaaaeaaapaaeaaqpevekpqkkpvikpltedskkrsynlisndstftq 405
HAdV-C01        traeekraeaeaaaeaaapaaeaaqpevekpqkkpvikpltedskkrsynlisndstftq 405
HAdV-C06        traeekraeaeaaaeaaapaaetaqpevekpqkkpvikpltedskkrsynlisndstftq 405
HAdV-D13        saadak----------------------aagkeielkailkddsgrsynvieg-ttdtl 348
HAdV-D37        saaeak----------------------aagkeielkailkddsdrsynvieg-ttdtl 350
HAdV-D53        saaeak----------------------aagkeielkailkddsdrsynvieg-ttdtl 350
HAdV-D69        saaeak----------------------aagkeielkailkddsnrsynvieg-ttdtl 350
HAdV-D60*       ----------------------------aagkeielkailkddsdrsynvieg-ttdtl 324
HAdV-D20        saaeak----------------------aagkeielkailkddsnrsynvieg-ttdtl 350
HAdV-D28        saaeak----------------------aagkeielkailkddsnrsynvieg-ttdtl 350
HAdV-D49        tteqlr----------------------aagkelvikpikedaskrsynvigd-thdtl 350
HAdV-D32        tteqlr----------------------aagkelvikpikedaskrsynvidd-thdtl 353
HAdV-D44        tteqlr----------------------aagkelvikpikedaskrsynvidd-thdtl 351
HAdV-D65        tteqlr----------------------aagkelvikpikedaskrsynvidg-thdtl 352
HAdV-D08        tteqlr----------------------aadkelvikpikedaskrsynvigd-thdtl 351
HAdV-D58        tteqlr----------------------aagkelvikpikedaskrsynvidg-thdtl 350
HAdV-D38        ttaeee----------------------aakkelvilpvtedeskrsynlieg-ttdtl 346
HAdV-D70        ttaeee----------------------aakkelvilpvtedeskrsynlieg-ttdtl 348
HAdV-D54        pravek----------------------aaekeleivpiekddsnrsyniipg-ttdtl 347
HAdV-D67        arevek----------------------aaqtqlviepikqddskrsynlieg-tmdtl 348
HAdV-D45        prdvek----------------------aaekelviepikqddskrsynlieg-tmdtl 345
HAdV-D17        srevek----------------------aaekelviepikqddskrsynlieg-tmdtl 348
HAdV-D64        srevek----------------------aaekelviepikqddskrsynlieg-thdtl 345
HAdV-D22        srevek----------------------aaekelviepikqddskrsynlieg-thdtl 348
HAdV-D42        srevek----------------------aaekelviepikqddskrsynlieg-thdtl 348
HAdV-D59        srevek----------------------aaekelviepikqddskrsynlieg-thdtl 348
HAdV-D39        tpaqek----------------------aadqkvevlpiekdesgrsynlipg-thdtl 350
HAdV-D09        tpaqet----------------------aadkkvevlpiekdesgrsynliqg-thdtl 350
HAdV-D10        tpaqet----------------------aadkkvevlpiekdesgrsynliqg-thdtl 350
HAdV-D56        tpaqet----------------------aadkkvevlpiekdesgrsynliqg-thdtl 350
HAdV-D26        tpaqet----------------------aadkkvevlpiekdesgrsynliqg-thdtl 350
HAdV-D47        sraaek----------------------aaekqlviepidkdesgrsynlikd-thdtl 348
HAdV-D25        tvaqek----------------------aaekelvivpiekdesnrsynlikd-thdtm 351
HAdV-D15        tvaqek----------------------aaqkelvivpiekdesnrsynlikd-thdtl 351
HAdV-D46        tvaqek----------------------aaekelvivpiekdesgrsynlikd-thdtl 351
HAdV-D36        tvaqek----------------------aaekelvivpiekdesnrsynlikd-thdtl 351
HAdV-D29        tvaqek----------------------aaekelvivpiekdesnrsynlikd-thdtl 351
HAdV-D43        tvaqek----------------------aaekelvivpiekdesnrsynlikd-thdtl 351
HAdV-D33        traaeq----------------------atekelviepieqdeskrsynvikg-thdtl 348
HAdV-D51        tkaveq----------------------aaekelviepieqdeskrsynvikg-thdtl 348
HAdV-D30        traaek----------------------aaekelviepieqdeskrsynvikg-thdtl 348
HAdV-D63        traaek----------------------aaekelviepieqdeskrsynvikg-thdtl 348
HAdV-D19        tkaaeq----------------------aaekelviepieqdeskrsynvikg-thdtl 346
HAdV-D23        traaeq----------------------aaekelviepieqdeskrsynvikg-thdtl 346
HAdV-D48        praveq----------------------aaekdlvivpvtqdeskrsynvidg-thdtl 344
HAdV-D24        praveq----------------------aaekdlvivpvtqdeskrsynvidg-thdtl 345
HAdV-D27        praveq----------------------aaekdlvivpvtqdeskrsynvidg-thdtl 345
HAdV-D62        praveq----------------------aaekdlvivpvtqdeskrsynvidg-thdtl 348
HAdV-F40*       pn------------------------------pqdleivpvekdskersynllegdknnta 335
HAdV-F41*       tr------------------------------pqdlviepvakdsknrsynllpndqnnta 339
HAdV-A18        al------------------------------mkvatiepvaadskgrsynllpd-kknte 338
HAdV-A12        al------------------------------nkaariepvetdpkgrsynllpd-kkntk 328
HAdV-A31        al------------------------------nkaakieaveadskgrsynllpd-kkntk 336
HAdV-A61        al------------------------------nkaakieaveadskgrsynllpd-kkntk 336
HadV-B16        tekqkreaaaael------------llmsevkkelkiqplekdsknrsynvled-kinta 386
HAdV-B68        tekkkreaaaael------------llmsevkkelkiqplekdsknrsynvled-kinta 386
HAdV-B03        tekqkreaaaa-------------------evkkelkiqplekdsksrsynvled-kinta 375
HAdV-B66        tekqkreaaaa-------------------evkkelkiqplekdsksrsynvled-kinta 375
HAdV-B07        tekhkreaaaa-------------------evkkelkiqplekdsksrsynvled-kinta 375
HAdV-E04        saa------avaa------------akadeteskiviqpvekdskdrsynvlsd-kknta 366
HAdV-B21        assvatde-slla------------avaettetkltikpvekdsksrsynvled-kvnta 392
HAdV-B50        assvatee-slla------------avaettetkltikpvekdsksrsynvled-kvnta 391
HAdV-B34        atsvptke-slld------------dvsqnielkltikpvekdgknrsynvled-kinta 389
HadV-B11        ptpvptae-slla------------dvsegtdvkltiqpvekdsknrsynvled-kinta 392
HAdV-B35        ptpvptae-slla------------dvsegtdvkltiqpvekdsknrsynvled-kinta 392
HAdV-B14        ptpvptae-slla------------dvtggtnvkltiqpvekdsknrsynvled-kinta 389
HAdV-B55        ptpvptae-slla------------dvsggtdvkltiqpvekdsknrsynvled-kinta 388
                                           :  :  * . ****:: .    *
```

```
HAdV-C05        yrswylaynygdpqtgirswtllctpdvtcgseqvywslpdmmqdpvtfrstrqisnfpv 462
HAdV-C02        yrswylaynygdpqtgirswtllctpdvtcgseqvywslpdmmqdpvtfrstsqisnfpv 462
HAdV-C57        yrswylaynygdpqagirswtllctpdvtcgseqvywslpdmmqdpvtfrstsqisnfpv 465
HAdV-C01        yrswylaynygdpqagirswtllctpdvtcgseqvywslpdmmqdpvtfrstsqisnfpv 465
HAdV-C06        yrswylaynygdpqagirswtllctpdvtcgseqvywslpdmmqdpvtfrstsqisnfpv 465
HAdV-D13        yrswylsytygdpekgvqtwtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 408
HAdV-D37        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 410
HAdV-D53        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 410
HAdV-D69        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 410
HAdV-D60*       yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 384
HAdV-D20        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 410
HAdV-D28        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 410
HAdV-D49        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 410
HAdV-D32        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 413
HAdV-D44        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 411
HAdV-D65        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 412
HAdV-D08        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 411
HAdV-D58        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 410
HAdV-D38        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 406
HAdV-D70        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 408
HAdV-D54        yrswylsytygdpkkgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 407
HAdV-D67        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 408
HAdV-D45        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 405
HAdV-D17        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 408
HAdV-D64        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 405
HAdV-D22        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 408
HAdV-D42        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 408
HAdV-D59        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 408
HAdV-D39        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 410
HAdV-D09        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 410
HAdV-D10        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 410
HAdV-D56        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 410
HAdV-D26        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 410
HAdV-D47        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 408
HAdV-D25        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 411
HAdV-D15        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 411
HAdV-D46        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 411
HAdV-D36        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 411
HAdV-D29        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 411
HAdV-D43        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 411
HAdV-D33        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 408
HAdV-D51        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 408
HAdV-D30        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 408
HAdV-D63        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 408
HAdV-D19        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 406
HAdV-D23        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 406
HAdV-D48        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 404
HAdV-D24        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 405
HAdV-D27        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 405
HAdV-D62        yrswylsytygdpekgvqswtllttpdvtcgaeqvywslpdlmqdpvtfrstqqvsnypv 408
HAdV-F40*       yrswflaynygdaekgvkswtllttdvtcgsqqvywslpdmmqdpvtfrpstqvsnypv 395
HAdV-F41*       yrswflaynygdpkkgvqswtllttadvtcgsqqvywslpdmmqdpvtfrpstqvsnypv 399
HAdV-A18        yrswylaynygdsekgvrswtllttpdvtggseqvywslpdmmqdpvtfrssrqvsnypv 398
HAdV-A12        yrswylaynygdpekgvrswtllttpdvtggseqvywslpdmmqdpvtfrssrqvsnypv 388
HAdV-A31        yrswylaynygdpekgvrswtllttpdvtggseqvywslpdmmqdpvtfrssrqvsnypv 396
HAdV-A61        yrswylaynygdpekgvrswtllttpdvtggseqvywslpdmmqdpvtfrssrqvsnypv 396
HadV-B16        yrswylsynygnpekgirswtllttsdvtcgaeqvywslpdmmqdpitfrssrqvnnypv 446
HAdV-B68        yrswylsynygnpekgirswtllttsdvtcgaeqvywslpdmmqdpitfrssrqvnnypv 446
HAdV-B03        yrswylsynygnpekgirswtllttsdvtcgaeqvywslpdmmqdpvtfrstrqvnnypv 435
HAdV-B66        yrswylsynygnpekgirswtllttsdvtcgaeqvywslpdmmqdpvtfrstrqvnnypv 435
HAdV-B07        yrswylsynygnpkkgirswtllttsdvtcgaeqvywslpdmmqdpvtfrstrqvnnypv 435
HAdV-E04        yrswylaynygdhekgvrswtllttsdvtcgveqvywslpdmmqdpvtfrstrqvsnypv 426
HAdV-B21        yrswylsynygdpekgvrswtllttsdvtcgaeqvywslpdmmqdpvtfrstrqvsnypv 452
HAdV-B50        yrswylsynygdpekgvrswtllttsdvtcgaeqvywslpdmmqdpvtfrstrqvsnypv 451
HAdV-B34        yrswylsynygdpekgvrswtllttsdvtcgaeqvywslpdmmqdpvtfrstrqvsnypv 449
HadV-B11        yrswylsynygdpekgvrswtllttsdvtcgaeqvywslpdmmkdpvtfrstrqvsnypv 452
HAdV-B35        yrswylsynygdpekgvrswtllttsdvtcgaeqvywslpdmmkdpvtfrstrqvsnypv 452
HAdV-B14        yrswylsynygdpekgvrswtllttsdvtcgaeqvywslpdmmqdpvtfrstrqvsnypv 449
HAdV-B55        yrswylsynygdpekgvrswtllttsdvtcgaeqvywslpdmmqdpvtfrstrqvsnypv 448
                ****:*:*.**: : *:::**** * *** * :********:*:**:*** : *:.*:**
```

```
HAdV-C05        vgaellpvhsksfyndqavysqlirqftslthvfnrfpenqilarppaptittvsenvpa 522
HAdV-C02        vgaellpvhsksfyndqavysqlirqftslthvfnrfpenqilarppaptittvsenvpa 522
HAdV-C57        vgaellpvhsksfyndqavysqlirqftslthvfnrfpenqilarppaptittvsenvpa 525
HAdV-C01        vgaellpvhsksfyndqavysqlirqftslthvfnrfpenqilarppaptittvsenvpa 525
HAdV-C06        vgaellpvhsksfyndqavysqlirqftslthvfnrfpenqilarppaptittvsenvpa 525
HAdV-D13        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 468
HAdV-D37        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 470
HAdV-D53        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 470
HAdV-D69        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 470
HAdV-D60*       vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 444
HAdV-D20        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 470
HAdV-D28        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 470
HAdV-D49        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 470
HAdV-D32        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 473
HAdV-D44        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 471
HAdV-D65        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 472
HAdV-D08        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 471
HAdV-D58        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 470
HAdV-D38        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 466
HAdV-D70        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 468
HAdV-D54        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 467
HAdV-D67        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 468
HAdV-D45        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 465
HAdV-D17        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 468
HAdV-D64        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 465
HAdV-D22        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 468
HAdV-D42        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 468
HAdV-D59        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 468
HAdV-D39        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 470
HAdV-D09        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 470
HAdV-D10        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 470
HAdV-D56        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 470
HAdV-D26        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 470
HAdV-D47        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 468
HAdV-D25        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 471
HAdV-D15        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 471
HAdV-D46        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 471
HAdV-D36        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 471
HAdV-D29        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 471
HAdV-D43        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 471
HAdV-D33        vgaelmpfraksfyndlavysqlirsytsltyvfnrfpdnqilcrppaptittvsenvpa 468
HAdV-D51        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 468
HAdV-D30        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 468
HAdV-D63        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 468
HAdV-D19        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 466
HAdV-D23        vgaelmpfraksfyndlavysqlirsytsltyvfnrfpdnqilcrppaptittvsenvpa 466
HAdV-D48        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 464
HAdV-D24        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 465
HAdV-D27        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 465
HAdV-D62        vgaelmpfraksfyndlavysqlirsytslthvfnrfpdnqilcrppaptittvsenvpa 468
HAdV-F40*       vgvellpvhaksfyneqavysqlirqstalthifnrfpenqilvrppaptittvsenvpa 455
HAdV-F41*       vgvellpvhaksfyneqavysqlirqstalthvfnrfpenqilvrppaptittvsenvpa 459
HAdV-A18        vaaellpvhaksfyndqavysqlirqstalthvfnrfpenqilvrppaatittvsenvpa 458
HAdV-A12        vaaellpvhaksfyneqavysqlirqstaltrvfnrfpenqilvrppaatittvsenvpa 448
HAdV-A31        vaaelmpvhaksfyneqavysqlirqstalthvfnrfpenqilvrppaatvttvsenvpa 456
HAdV-A61        vaaelmpvhaksfyneqavysqlirqstalthvfnrfpenqilvrppaatittvsenvpa 456
HadV-B16        vgaelmpvfsksfyneqavysqqlrqstslthvfnrfpenqilirppaptittisenvpa 506
HAdV-B68        vgaelmpvfsksfyneqavysqqlrqstslthvfnrfpenqilirppaptittisenvpa 506
HAdV-B03        vgaelmpvfsksfyneqavysqqlrqatslthvfnrfpenqilirppaptittvsenvpa 495
HAdV-B66        vgaelmpvfsksfyneqavysqqlrqatslthvfnrfpenqilirppaptittvsenvpa 495
HAdV-B07        vgaelmpvfsksfyneqavysqqlrqatslthvfnrfpenqilirppaptittvsenvpa 495
HAdV-E04        vgaelmpvysksffneqavysqqlraftslthvfnrfpenqilvrppaptittvsenvpa 486
HAdV-B21        vgaelmpvfsksfyneqavysqqlrqstslthvfnrfpenqilirppaptittvsenvpa 512
HAdV-B50        vgaelmpvfsksfyneqavysqqlrqstslthvfnrfpenqilirppaptittvsenvpa 511
HAdV-B34        vgaelmpvfsksfyneqavysqqlrqstslthvfnrfpenqilirppaptittvsenvpa 509
HadV-B11        vgaelmpvfsksfyneqavysqqlrqstslthvfnrfpenqilirppaptittvsenvpa 512
HAdV-B35        vgaelmpvfsksfyneqavysqqlrqstslthvfnrfpenqilirppaptittvsenvpa 512
HAdV-B14        vgaelmpvfsksfyneqavysqqlrqstslthvfnrfpenqilirppaptittvsenvpa 509
HAdV-B55        vgaelmpvfsksfyneqavysqqlrqstslthvfnrfpenqilirppaptittvsenvpa 508
                *..**:*. :***:*: ***** :*  *:** :*****:**** **** *:**:******
```

```
HAdV-C05        ltdhgtlplrnsiggvqrvtitdarrrtcpyvykalgivsprvlssrtf   571
HAdV-C02        ltdhgtlplrnsiggvqrvtitdarrrtcpyvykalgivsprvlssrtf   571
HAdV-C57        ltdhgtlplrnsiggvqrvtitdarrrtcpyvykalgivsprvlssrtf   574
HAdV-C01        ltdhgtlplrnsiggvqrvtitdarrrtcpyvykalgivsprvlssrtf   574
HAdV-C06        ltdhgtlplrnsiggvqrvtitdarrrtcpyvykalgivsprvlssrtf   574
HAdV-D13        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   517
HAdV-D37        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   519
HAdV-D53        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   519
HAdV-D69        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   519
HAdV-D60*       ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   493
HAdV-D20        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   519
HAdV-D28        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   519
HAdV-D49        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   519
HAdV-D32        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   522
HAdV-D44        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   520
HAdV-D65        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   521
HAdV-D08        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   520
HAdV-D58        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   519
HAdV-D38        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   515
HAdV-D70        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   517
HAdV-D54        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   516
HAdV-D67        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   517
HAdV-D45        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   514
HAdV-D17        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   517
HAdV-D64        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   514
HAdV-D22        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   517
HAdV-D42        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   517
HAdV-D59        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   517
HAdV-D39        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   519
HAdV-D09        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   519
HAdV-D10        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   519
HAdV-D56        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   519
HAdV-D26        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   519
HAdV-D47        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   517
HAdV-D25        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   520
HAdV-D15        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   520
HAdV-D46        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   520
HAdV-D36        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   520
HAdV-D29        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   520
HAdV-D43        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   520
HAdV-D33        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   517
HAdV-D51        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   517
HAdV-D30        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   517
HAdV-D63        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   517
HAdV-D19        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   515
HAdV-D23        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   515
HAdV-D48        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   513
HAdV-D24        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   514
HAdV-D27        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   514
HAdV-D62        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   517
HAdV-F40*       ltdhgtlplrssisgvqrvtitdarrrtcpyvhkalgivapkvlssrtf   504
HAdV-F41*       ltdhgtlplrssisgvqrvtitdarrrtcpyvhkalgivapkvlssrtf   508
HAdV-A18        ltdhgtlplrssisgvqrvtitdarrrtcpyvykalgivsprvlssrtf   507
HAdV-A12        ltdhgtlplrssisgvqrvtitdarrrtcpyvykalgivsprvlssrtf   497
HAdV-A31        ltdhgtlplrssisgvqrvtitdarrrtcpyvykalgivsprvlssrtf   505
HAdV-A61        ltdhgtlplrssisgvqrvtitdarrrtcpyvykalgivsprvlssrtf   505
HadV-B16        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   555
HAdV-B68        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   555
HAdV-B03        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   544
HAdV-B66        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   544
HAdV-B07        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   544
HAdV-E04        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   535
HAdV-B21        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   561
HAdV-B50        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   560
HAdV-B34        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   558
HadV-B11        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   561
HAdV-B35        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   561
HAdV-B14        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   558
HAdV-B55        ltdhgtlplrssirgvqrvtvtdarrrtcpyvykalgivaprvlssrtf   557
                **********.** ******:***********:******:*:*******
```

6.2.4. E4ORF1p (HAdV Types C01 – D70)

The genomes of HAdV types **D19p** (JQ326209), **D37** (AB448778), and **D64** (EF121005) all show a deletion of a single adenine base from a stretch of eight at the same position which introduces a frameshift after residue 27. HAdV-**D28** contains a point mutation at position 85 that introduces a stop codon. It is unclear whether these changes are physiological or due to sequencing errors. The respective sequences were left out of the alignment. HAdVs of species F do not possess the E4ORF1p gene.

```
HAdV-D54      ---makclyafidspggiapvqegasnkytffcpqsfyifphgvvlvylkvsvlvptgyq
HAdV-D08      ---maeclyafidspggmapvqegasnkytffcpesfyifphgvvlvylkvsvliptgyq
HAdV-D49      ---maeclyafidspggmapvqegasnkytffcpesfyifphgvvlvylkvsvlvptgyq
HAdV-D53      ---maeclyafidspggmapvqegasnkytffcpesfyifphgvvlvylkvsvlvptgyq
HAdV-D56      ---maeslyafidspggiasvqegtsnrytffcpesfhilphgvvllhlkvsvlvptgyq
HAdV-D59      ---maeslyafidspggiapvqegtsnrynffcpesfhipphgvvllylkvsvlvptgyq
HAdV-D42      ---maeslyafidspggiapvqegtsnrynffcpesfhipphgvvllhlkvsvlvpngyq
HAdV-D22      ---maeslyafidspggiapvqegtsnrynffcpksfhipphgvvllhlkvsvlvptgyq
HAdV-D20      ---maeslyafidspggiapvqegtsnrynffcpqsfhipphgvvllhlkvsvlvptgyq
HAdV-D17      ---maeslyafidspggiapvqegasnrynffcpesfhipphgvvllhlkvsvlvptgyq
HAdV-D09      ---maeslyafidspggiapvqegtsnrytffcpesfhipphgvvllhlkvsvlvptgyq
HAdV-D10      ---maeslyafidspggiapvqegtsnrytffcpesfhipphgvvllhlkvsvlvptgyq
HAdV-D15      ---maeslyafidspggiapvqegtsnrynffcpesfhipphgvvllhlkvsvlvptgyq
HAdV-D24      ---maeslyafidspggiapvqegtsnrynffcpesfhipphgvvllhlkvsvlvptgyq
HAdV-D46      ---maeslyafidspggiapvqegtsnrynffcpesfhipphgvvllhlkvsvlvptgyq
HAdV-D69      ---maeslyafidspggiapvqegtsnrynffcpesfhipphgvvllhlkvsvlvptgyq
HAdV-D70      ---maeslyafidspggiapvqegtsnrynffcpesfhipphgvvllhlkvsvlvptgyq
HAdV-D62      ---maeslyafidspggiapvqegasnryiffcpesfhipphgvillhlrvsvmvptgyq
HAdV-D45      ---maeslyafidspggiapvqegasnrytffcpeffhipphgvillhlrvsvlvptgyq
HAdV-D36      ---maeslyafidspggiapvqegasnryiffcpesfhipphgvillhlrvsvlvptgyq
HAdV-D30      ---maeslyafidspggiapvqegasnrytffcpesfhipphgvillhlrvsvlvptgyq
HAdV-D43      ---maeslyafidspggiapvqegasnrytffcpesfhipphgvillhlrvsvlvptgyq
HAdV-D47      ---maeslyafidspggiapvqegasnrytffcpesfhipphgvillhlrvsvlvptgyq
HAdV-D51      ---maeslyafidspggiapvqegasnrytffcpesfhipphgvillhlrvsvlvptgyq
HAdV-D39      ---maeslyafidspgglapvqegasnrynffcpesfhipphgvvllhlrvsvliptgyq
HAdV-D38      ---maeslyafidspggiapvqegtsnrynffcpesfhipphgvvlvhlrvsvliptgyq
HAdV-D44      ---maeslyafidspggiapvqegtsnrydffcpesfhipphgvvllhlrvsvliptgyq
HAdV-D65      ---maeslyafidspggiapvqegssnrynffcpesfhipphgvvllhlrvsvliptgyq
HAdV-D13      ---maeslyafidspggiapvqegtsnrytffcpesfhipphgvvllhlrvsvlipngyq
HAdV-D60      ---maeslyafidspggiapvqegtsnrytffcpesfhipphgvvllhlrvsvlipngyq
HAdV-D23      ---maeslyafidspggiapvqegssnrynffcpesfhipphgvvllhlrvsvliptgyq
HAdV-D26      ---maeslyafidspggiapvqegssnrynffcpesfhipphgvvlvhlrvsvliptgyq
HAdV-D33      ---maeslyafidspggiapvqegssnrynffcpesfhipphgvvlvhlrvsvliptgyq
HAdV-D67      ---maeslyafidspggiapvqegtsnrynffcpesfhipphgvvlvhlrvsvlirngyq
HAdV-D32      ---maeslyafidspggiapvqegtsnrynffcpesfhipphgvvlvhlrvsvliptgyq
HAdV-D58      ---maeslyafidspggiapvqegtsnrynffcpesfhipphgvvlvhlrvsvlipngyq
HAdV-D25      ---maeslyafidspggiapvqegtsnrytffcpesfhipphgvvlvhlrvsvlipngyq
HAdV-D27      ---maeslyafidspggiapvqegtsnrynffcpesfhipphgvvlvhlrvsvlipngyq
HAdV-D29      ---maeslyafidspggiapvqegtsnrynffcpesfhipphgvvlvhlrvsvlipngyq
HAdV-D48      ---maeslyafidspggiapvqegtsnrynffcpesfhipphgvvlvhlrvsvlipngyq
HAdV-D63      ---maeslyafidspggiapvqegtsnrynffcpesfhipphgvvlvhlrvsvlipngyq
HAdV-E04      --mdaqvlyvflegagallpvq--kgsnyifyapanfvlhphgvallelrlsivvpqgfi
HAdV-B16      --madealyvyfrgpgatlpeqqqqrnnyifyspvpftlyprgvallylrlsiiiprgyv
HAdV-B14      --madealyvylegpgatlpeqq-qrnnyifyspvpftlyprgvallylrlsiiiprgyi
HAdV-B55      --madealyvylegpgatlpeqq-qrnnyifyspvpftlyprgvallylrlsiiiprgyi
HAdV-B35      --madealyvylegpgatlpkqq-qrnnyifyspvpftlyprgvallylrlsiiipkgyv
HAdV-B03      --madealyvylegpgatlpeqq-qrnnyifyspvpftlyprgvallylklsiiiprgyv
HAdV-B66      --manealyvylegpgatlpeqq-qrnnyifyspvpftlyprgvallylrlsiiiprgyv
HAdV-B34      --madealyvylegpgatlpeqq-qrnnyifyspvpftlyprgvallylrlsiiipkgyv
HAdV-B11      --madealyvylegpgatlpeqq-qrnnyifyspvpftlyprgvallylrlsiiipkgyv
HAdV-B68      --madealyvylegpgatlpeqq-qrnnyifyspvpftlyprgvallylrlsiiiprgyv
HAdV-B21      --madealyvyldgpgatlpeqr-qrnnyifyspvpftlyprgvallylrlsiiiprgyv
HAdV-B07      --madealyvylegpgatlpeqq-qrnnyifyspvpftlyprgvallylrlsiiiprgyv
HAdV-B50      --madealyvyldgpgatlpeqq-qrnnyifyspvpftlyprgvallylrlsiiiprgyv
HAdV-C01      maaavealyvvleregailprqegfsgvyvffspinfvippmgavmlslrlrvcippgyf
HAdV-C02      maaavealyvvleregailprqegfsgvyvffspinfvippmgavmlslrlrvcippgyf
HAdV-C05      maaavealyvvleregailprqegfsgvyvffspinfvippmgavmlslrlrvcippgyf
HAdV-C06      maaavealyvvleregailprqegfsgvyvffspinfvippmgavmlslrlrvcippgyf
HAdV-C57      maaavealyvvleregailprqegfsgvyvffspinfvippmgavmlslrlrvcippgyf
HAdV-G52      --madqhiyvhllgrraflpqqqgysnmyvlfspedfvlaprgiillslqlsldiptgyl
HAdV-A31      -mavfeavyvyfkgpgamlpeqegysnayvlfspanfvipphgvvllylniavdlppgyv
HAdV-A61      -mavfeavyvyikgpgamlpeqegysnayvlfspanfvipphgvvllylniavdippgyv
HAdV-A12      -maafetlyvyftgpgamlpkqegdsnayvlfspanfvipphgvvllylhiavdippgyl
HAdV-A18      -maalqalyvyfkgpgamlpeqegysnayvlfspanfvipphgvvllylhiavdippgyl
                   : :*. :    .    *    . * ::.*  * : * *  :: *.: : :  *:
```

```
HAdV-D54      gifmalndyhatgiltqsdvifagrrqeltvllfnhtdqflyvrkghpvgtlllkrvifp
HAdV-D08      gtfmalndyhargiltqsdvifagrrqelavllfnhtdqflyvrkghpvgtlllkkvifp
HAdV-D49      gtfmalndyhargiltqsdvifagrrqelavllfnhtdqflyvrkghpvgtlllkkvifp
HAdV-D53      gtfmalndyhargiltqsdvifagrrqelavllfnhtdqflyvrkghpvgtlllkkvifp
HAdV-D56      grfmalndyhardiltqsdvifagrrqeltvllfnhtdrflyvrkghpvgtlllervifp
HAdV-D59      grfmalndyhardiltqsdvifagrrqeltvllfnhtdrflyvrkghpvgtlllervifp
HAdV-D42      grfmalndyhardiltqsdvifagrrqeltvllfnhtdrflyvrkghpvgtlllervifp
HAdV-D22      grfmalndyhardiltqsdvifagrrqeltvllfnhtdrflyvrkghpvgtlllervifp
HAdV-D20      grfmalndyhardiltqsdvifagrrqeltvllfnhtdrflyvrkghpvgtlllervifp
HAdV-D17      grfmalndyhardiltqsdvifagrrqeltvllfnhtdrflyvrkghpvgtlllervifp
HAdV-D09      grfmalndyhardiltqsdvifagrrqeltvllfnhtdrflyvrkghpvgtlllervifp
HAdV-D10      grfmalndyhardiltqsdvifagrrqeltvllfnhtdrflyvrkghpvgtlllervifp
HAdV-D15      grfmalndyhardiltqsdvifagrrqeltvllfnhtdrflyvrkghpvgtlllervifp
HAdV-D24      grfmalndyhardiltqsdvifagrrqeltvllfnhtdrflyvrkghpvgtlllervifp
HAdV-D46      grfmalndyhardiltqsdvifagrrqeltvllfnhtdrflyvrkghpvgtlllervifp
HAdV-D69      grfmalndyhardiltqsdvifagrrqeltvllfnhtdrflyvrkghpvgtlllervifp
HAdV-D70      grfmalndyhardiltqsdvifagrrqeltvllfnhtdrflyvrkghpvgtlllervifp
HAdV-D62      grfmalndyhargiltqsdvifagrrhdlsvllfnhtdrflyvreghpvgtlllervifp
HAdV-D45      grfmalndyhargiltqsdvifagrrhdlsvllfnhtdrflyvreghpvgtlllervifp
HAdV-D36      grfmalndyhargiltqsdvifagrrhdlsvllfnhtdrflyvreghpvgtlllervifp
HAdV-D30      grfmalndyhargiltqsdvifagrrhdlsvllfnhtdrflyvreghpvgtlllervifp
HAdV-D43      grfmalndyhargiltqsdvifagrrhdlsvllfnhtdrflyvreghpvgtlllervifp
HAdV-D47      grfmalndyhargiltqsdvifagrrhdlsvllfnhtdrflyvreghpvgtlllervifp
HAdV-D51      grfmalndyhargiltqsdvifagrrhdlsvllfnhtdrflyvreghpvgtlllervifp
HAdV-D39      grfmalndyhargiltqsdvifagrrqelsvllfnhtdrflyvreghpvgilllervifp
HAdV-D38      grfmalndyhargiltqsdvifagrrhelsvllfnhtdrflyvreghpvgilllekvifp
HAdV-D44      grfmalndyhargiltqsdvifagrrqelsvllfnhtdrflyvreghpvgilllekvifp
HAdV-D65      grfmalndyhargiltqsdvifagrrqelsvllfnhtdrflyvreghpvgilllekvifp
HAdV-D13      grfmalndyhtsgiltqsdvifagrrqelavllfnhtdrflyvreghpvgtlllqrvifp
HAdV-D60      grfmalndyhtsgiltqsdvifagrrqelavllfnhtvrflyvreghpvgtlllqrvifp
HAdV-D23      grfmalndyhargiltqsdvifagrrheltvllfnhtdrflyvreghpvgtlllervifp
HAdV-D26      grflalndfhargiltqsdvifagrrheltvllfnhtdrflyvreghpvgtlllervifp
HAdV-D33      grflalndyhargiltqsdvifagrrheltvllfnhtdrflyvreghpvgtlllervifp
HAdV-D67      grfmalndyhargiltqsdvifagrrheltvllfnhtdrflyvreghpvgtlllervifp
HAdV-D32      grfmalndyhargiltqsdvifagrrqeltvllfnhtdrflyvreghpvgtlllervifp
HAdV-D58      grfmalndyhsrgiltqsdvifagrrqeltvllfnhtdrflyvreghpvgtlllervifp
HAdV-D25      grfmalndyhsrgiltqsdvifagrrqeltvllfnhtdrflyvreghpvgtlllervifp
HAdV-D27      grfmalndyhsrgiltqsdvifagrrqeltvllfnhtdrflyvreghpvgtlllervifp
HAdV-D29      grfmalndyhsrgiltqsdvifagrrqeltvllfnhtdrflyvreghpvgtlllervifp
HAdV-D48      grfmalndyhsrgiltqsdvifagrrqeltvllfnhtdrflyvreghpvgtlllervifp
HAdV-D63      grfmalndyhsrgiltqsdvifagrrqeltvllfnhtdrflyvreghpvgtlllervifp
HAdV-E04      grffsltdanvpgvyassriihaghreglsvmlfnhnvsfyngragdpvaclvlerviyp
HAdV-B16      gcffsltdanmsglyassriihaghreelsvllfnhddrfyegragdpvaclvmerliyp
HAdV-B14      gcflsltdanmfglyassriihaghreelsvllfnhddrfyegragdpvaclvmerliyp
HAdV-B55      gcflsltdanmfglyassriihaghreelsvllfnhddrfyegragdpvaclvmerliyp
HAdV-B35      gcffsltdanmsglyassrlihaghreelsvllfnhddrfyegragdpvaclvmerliyp
HAdV-B03      gcffsltdanmsglyassriihaghreelsvllfnhndrfyegragdpvaclvmerlifp
HAdV-B66      gcffsltdanmsglyassriihaghreelsvllfnhddrfyegragdpvaclvmerlifp
HAdV-B34      gcffsltdanmsglyassriihaghreelsvllfnhhdrfyegragdpvaclvmerliyp
HAdV-B11      gcffsltdanmsglyassriihaghreelsvllfnhddrfyegragdpvaclvmerliyp
HAdV-B68      gcffsltdanmsglyassriihaghreelsvllfnhddrfyegqagdpvaclvmerliyp
HAdV-B21      gcffsltdanmsglyassriihaghreelsvllfnhddrfyegragdpvaclvmerliyp
HAdV-B07      gcffsltdanmsglyassriihaghreelsvllfnhddrfyegragdpvaclvmerliyp
HAdV-B50      gcffsltdanmsglyassriihaghreelsvllfnhddrfyegragdpvaclvmerliyp
HAdV-C01      grflaltdvnqpdvftesyimtpdmteelsvvlfnhgdqffyghagmavvrlmlirvvfp
HAdV-C02      grflaltdvnqpdvftesyimtpdmteelsvvlfnhgdqffyghagmavvrlmlirvvfp
HAdV-C05      grflaltdvnqpdvftesyimtpdmteelsvvlfnhgdqffyghagmavvrlmlirvvfp
HAdV-C06      grflaltdvnqpdvftesyimtpdmteelsvvlfnhgdqffyghagmavvrlmlirvvfp
HAdV-C57      grflaltdvnqpdvftesyimtpdmteelsvvlfnhgdqffyghagmavvrlmlirvvfp
HAdV-G52      grffsvadmnvrgvllcaqeiqpstwwevsvvlfnhsdefyrgsrgqpvacllleryiyp
HAdV-A31      gtlfslsgmnargvfvgaetlypgsrmelsvllfnhsdvfcdvrakqpvarlvlnrvifp
HAdV-A61      gtlfslsgmnargvfvgaetlypgsrmelsvllfnhsdvfcdvrakqpvarlvlnrvifp
HAdV-A12      gtlfslcdmnargvfvgaetlypgsrmelsvllfnhsdvfcdvrakqpvarlllsrvvfp
HAdV-A18      gtlfslsdmnargvfvgaetlypgsrmelsvllfnhsdvfcdvrakqpvarlllsrvifp
              * :::: . :  .:   :  :  .    ::*:****   *        *  *:: ::::*
```

```
HAdV-D54      svkiatlv
HAdV-D08      svkiatlv
HAdV-D49      svkiatlv
HAdV-D53      svkiatlv
HAdV-D56      svkiatlv
HAdV-D59      svkiatlv
HAdV-D42      svkiatlv
HAdV-D22      svkiatlv
HAdV-D20      svkiatlv
HAdV-D17      svkiatlv
HAdV-D09      svkiatlv
HAdV-D10      svkiatlv
HAdV-D15      svkiatlv
HAdV-D24      svkiatlv
HAdV-D46      svkiatlv
HAdV-D69      svkiatlv
HAdV-D70      svkiatlv
HAdV-D62      svriatlv
HAdV-D45      svriatlv
HAdV-D36      svriatlv
HAdV-D30      svrlatlv
HAdV-D43      svriatlv
HAdV-D47      svriatlv
HAdV-D51      svriatlv
HAdV-D39      svrlatlv
HAdV-D38      svrlatlv
HAdV-D44      svrlatlv
HAdV-D65      svrlatlv
HAdV-D13      svrlatlv
HAdV-D60      svrlatlv
HAdV-D23      svrlatlv
HAdV-D26      svrlatlv
HAdV-D33      svrlatlv
HAdV-D67      svrlatlv
HAdV-D32      svrlatlv
HAdV-D58      svrlttlv
HAdV-D25      svrlatlv
HAdV-D27      svrlatlv
HAdV-D29      svrlatlv
HAdV-D48      svrlatlv
HAdV-D63      svrlatlv
HAdV-E04      pvrqasmv
HAdV-B16      pvrqatmi
HAdV-B14      pvrqatli
HAdV-B55      pvrqatli
HAdV-B35      pvrqatmi
HAdV-B03      pvrqatmi
HAdV-B66      pvrqatmi
HAdV-B34      pvrqatmi
HAdV-B11      svrqatmi
HAdV-B68      pvrqatmi
HAdV-B21      pvrqatmi
HAdV-B07      pvrqatmi
HAdV-B50      pvrqatmi
HAdV-C01      vvrqasnv
HAdV-C02      vvrqasnv
HAdV-C05      vvrqasnv
HAdV-C06      vvrqasnv
HAdV-C57      vvrqasnv
HAdV-G52      tvrqaslv
HAdV-A31      pirqatll
HAdV-A61      pirqatll
HAdV-A12      pvcqasli
HAdV-A18      pvrqasll
               :  :: :
```

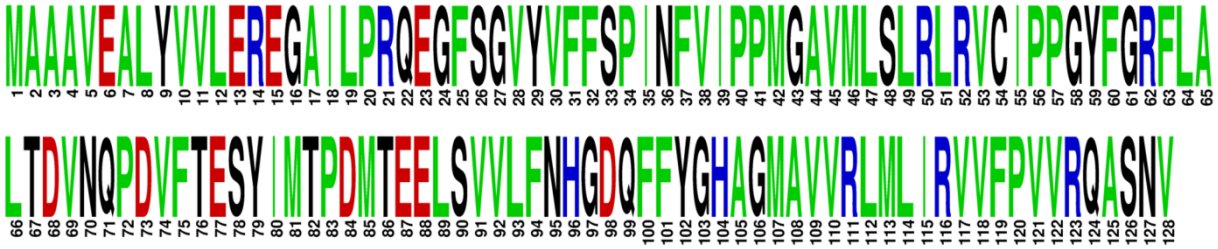
Conservation Logos:

Apolar: green; Polar: Black; Positive Charge: blue; Negative Charge: Red

HAdV-A:

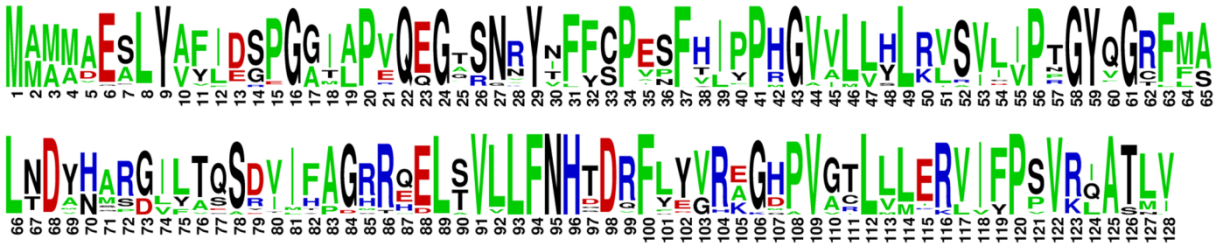
HAdV-B:

HAdV-C:



HAdV-D:

HAdV-A-G:



6.3. Licenses for Figures from Other Publications

Some figures in this dissertation were taken from original presentations. All original publications are referenced in the respective figure subtext and are listed here. If the journals are not licensed under the freely distributable Creative Commons Attribution (CC BY) license, or do not provide free access to their content for dissertations, licenses were obtained from the s100.copyright.com website of RightsLink™.

- **Cover Ch. 1; Figure 8: RCSB PDB-101 Molecule of the Month, Issues December 2010 and February 2011**
  Published by David S. Goodsell and the Research Collaboratory for Structural Bioinformatics Protein Data Bank. Molecule of the Month illustrations are available under a CC-BY-4.0 license.

- **Figure 1: Bauer *et al.* 1999: Jounal of Virology, modified**
  Published by the American Society for Microbiology. ASM authorizes an advanced degree candidate to republish the requested material in his/her doctoral thesis or dissertation.

- **Figure 4A: Nemerow *et al.*, 2012, Curr. Opin. Virol.**

Supplier	Elsevier Limited The Boulevard, Langford Lane Kidlington, Oxford, OX5 1GB, IK
Customer Name	Antonio Manuel Liaci
License Number	**3965450155690**
License date	Oct 10, 2016
Licensed Content Publisher	Elsevier
Licensed Content Publication	Current Opinion in Virology
Licensed Content Title	Structure of human adenovirus
Licensed Content Author	Glen R Nemerow,Phoebe L Stewart,Vijay S Reddy
Licensed Content Date	April 2012
Original figure numbers	Figure 2

- **Figure 4B: Zubieta *et al.*, 2005, Molecular Cell**

Supplier	Elsevier Limited The Boulevard, Langford Lane Kidlington, Oxford, OX5 1GB, IK
Customer Name	Antonio Manuel Liaci
License Number	**3965450889114**
License date	Oct 10, 2016
Licensed Content Publisher	Elsevier
Licensed Content Publication	Molecular Cell
Licensed Content Title	The Structure of the Human Adenovirus 2 Penton
Licensed Content Author	Chloe Zubieta,Guy Schoehn,Jadwiga Chroboczek,Stephen Cusack
Licensed Content Date	7 January 2005
Original figure numbers	Figure 4A

- **Figure 5A,C:** **Wasik *et al.*, 2016, Trends in Microbiology**

Supplier	Elsevier Limited The Boulevard, Langford Lane Kidlington, Oxford, OX5 1GB, IK
Customer Name	Antonio Manuel Liaci
License Number	**3965460040960**
License date	Oct 10, 2016
Licensed Content Publisher	Elsevier
Licensed Content Publication	Trends in Microbiology
Licensed Content Title	Effects of Sialic Acid Modifications on Virus Binding and Infection
Licensed Content Author	Brian R. Wasik,Karen N. Barnard,Colin R. Parrish
Licensed Content Date	Available online 1 August 2016
Original figure numbers	Figure 1

- **Figure 5B:** **Ströh *et al.*, 2014, Annual Review of Virology**

Published by Annual Reviews. Material may be republished in a thesis / dissertation without obtaining additional permission from Annual Reviews, providing that the author and the original source of publication are fully acknowledged.

- **Figures 5D, 6B:** **Neu *et al.*, 2011, Current Opinion in Structural Biology**

Supplier	Elsevier Limited The Boulevard, Langford Lane Kidlington, Oxford, OX5 1GB, IK
Customer Name	Antonio Manuel Liaci
License Number	**3965451297117**
License date	Oct 10, 2016
Licensed Content Publisher	Elsevier
Licensed Content Publication	Current Opinion in Structural Biology
Licensed Content Title	Viruses and sialic acids: rules of engagement
Licensed Content Author	Ursula Neu,Johannes Bauer,Thilo Stehle
Licensed Content Date	October 2011
Original figure numbers	Figure 1A,B

- **Figure 6C:** **Ströh *et al.*, 2015, PLoS Pathogens**

Published by the Public Library of Science. PLoS articles are licensed under the Creative Commons Attribution (CC BY) license.

- **Figure 7:** **Michael H. C. Buch, Dissertation, modified with permission**

Published by the University of Tübingen, Faculty of Science, 2016

- **Figure 9D:** **Veesler *et al.*, 2014, Proceedings of the National Academy of Sciences of the United States of America**

Anyone may, without requesting permission, use original figures or tables published in PNAS for noncommercial and educational use (i.e., in a review article, in a book that is not for sale), provided that the full journal reference is cited.

- **Figure 9E:** **Lindert *et al.*, 2009, Journal of Virology**

Published by the American Society for Microbiology. ASM authorizes an advanced degree candidate to republish the requested material in his/her doctoral thesis or dissertation.

- **Figure 12: Kwiatkowska *et al.*, 2013, Cancers (Basel)**
  Published by MDPI (Multidisciplinary Digital Publishing Institute). Everyone is free to re-use MDPI published material if proper accreditation/citation of the original publication is given.

- **Figure 14 A,B: Lenman *et al.*, 2015, PLoS Pathogens**
  Published by the Public Library of Science. PLoS articles are licensed under the Creative Commons Attribution (CC BY) license.

- **Figure 16: Caraballo *et al.*, 2015, Organic and Biomolecular Chemistry**
  Published by The Royal Society of Chemistry. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence. Material from this article can be used in other publications provided that the correct acknowledgement is given with the reproduced material and it is not used for commercial purposes.

6.4. Conference & Meeting contributions

- 26th International Carbohydrate symposium. 2012/07/22-27, Madrid, Spain
  Attended

- 4th Adenovirus Workshop 2013. 2013/02/20-22, Barsinghausen, Germany
  Attended

- IFIB PhD retreat 2013. 2013/09/20-22, Blaubeuren, Germany
  Presentation: **Investigating Multimeric Proteins**

- 27th Rhine-Knee Regional Meeting on Biocrystallography. 2013/09/25-27, Schluchsee, Germany
  Attended

- 33rd Annual Meeting of the American Society for Virology. 2014/06/21-25, Fort Collins, USA
  Poster: **Human Adenovirus 52 uses Sialic-Acid Containing Glycans for Binding to Target Cells**

- SFB 685 Retreat 2015. 2015/07/06-08, Blaubeuren, Germany
  Poster: **Human Adenovirus 52 Uses Sialic-Acid Containing Glycans for Binding to Target Cells**

- 29th Rhine Knee Meeting, 2015/09/, Engelberg, Switzerland
  Presentation: **Human Adenovirus 52 uses Polysialic Acid for Cell Attachment**

- CCP4 Study Weekend 2016. 2016/01/08-10, Nottingham, United Kingdom
  Attended

- 2nd Meeting of the Glycovirology Team, 2015/11/03, Arlanda, Sweden
  Presentation: **Primary Attachment Factors of HAdV-D36, D37, and G52**

- IFB PhD Retreat 2016. 2016/04/21-23, Freudenstadt, Germany
  Poster: **Structural Biology of Glycan-Binding Adenoviruses**

- Frontiers in Sialic Acid Research. 2016/04/23-25, Bad Lauterberg, Germany
  Poster: **Structural Biology of Glycan-Binding Adenoviruses**

- 12th International Adenovirus Meeting. 2016/08/16-20, Barsinghausen, Germany
  Poster: **Primary Attachment Receptors of Human Adenovirus 36**
  This poster was awarded with the poster price.

- 30th Rhine-Knee Regional Meeting on Biocrystallography. 2016/09/28-30, Schöntal, Germany
  Attendance & Organization

6.5. Supervisions

- Anna-Lisa Fuchs, Diploma Rotation, Feb 2013
  "An alternative crystal form of hPyV10 VP1 for soaking experiments"

- Lisa Christine Käshammer, Bachelor Rotation, Jun 2013
  "*Crystallization of the Short Fiber Knob of the Human Adenovirus 52*"

- Lisa Christine Käshammer, Bachelor Thesis, Sep 2013
  "*Structural Investigations of the Short Fiber Knob from Human Adenovirus 52 and its Possible Ligands*"

- Vanessa Pfenning, Bachelor Thesis, Aug 2014
  "*Structural Investigations of the Human Adenovirus 36 Fiber Knob and its Candidate Receptors*"

- Gregor Wiese, Master Rotation, Dec 2014
  "*Structural Investigations of Carbohydrate Ligands for Human Adenoviruses 52 and 36*"

- Antonia Lott, Bachelor Rotation, Apr 2015
  "Working towards the structural characterization of the E4orf1 gene product of human Adenovirus HAdV 52 - cloning, expression, first purification -"

- Antonia Lott, Bachelor Thesis, July 2015
  "*Cloning, Expression and Purification of E4ORF1 Gene Constructs from Human Adenoviruses*"

- Paul Bachmann, Bachelor Rotation, Jul 2016
  "*Expression, Purification and Crystallization of Human Adenovirus D-36 and D-37 Fiber Knob Mutants*"

- Paul Bachmann, Bachelor Thesis, Sep 2016
  "*Structural Insights into the Sialic Acid Binding Modes of Human Adenovirus D-36 and D-37 Fiber Knob Domains*"

- Laura Hehl, Bachelor Rotation, Aug 2016
  "*Working towards the Characterization of human Adenovirus protein E4-ORF1*"

- Laura Hehl, Bachelor thesis, Dec 2016
  "*Characterization of the Human Adenovirus protein E4-ORF1*"

Unpublished data from these works are incorporated into this Thesis only with the consent of the supervised students. The contributions are appointed as such in the accompanying manuscripts.

7. PUBLICATIONS AND UNPUBLISHED MANUSCRIPTS

7.5. Contributions to Manuscripts Included in this work

Buch MHC*, **Liaci AM***, O'Hara SD, Garcea RL, Neu U, Stehle T (2015) Structural and Functional Analysis of Murine Polyomavirus Capsid Proteins Establish the Determinants of Ligand Recognition and Pathogenicity. PLoS Pathogens 11:e1005104.

*These Authors contributed equally.

Contributions: AML cloned LID VP1 by site-directed mutagenesis and purified, crystallized, and solved the native and complex structures of PTA and LID VP1. He developed the density integration procedure together with MHCB and UN. Furthermore, he wrote the paper together with MHCP, UN, SDO, RLG, and TS. He was involved in the planning of all experiments. Parts of this work are described in the Diploma Thesis of AML (Tuebingen, 2012).

Lenman A, **Liaci AM**, Liu Y, Årdahl C, Rajan A, Nilsson E, Bradford W, Kaeshammer L, Jones MS, Frängsmyr L, Feizi T, Stehle T, Arnberg N (2015) Human Adenovirus 52 Uses Sialic Acid-containing Glycoproteins and the Coxsackie and Adenovirus Receptor for Binding to Target Cells. PLoS Pathogens 11:e1004657.

Contributions: AML purified, crystallized and solved the native and complex structures of HAdV-G52 SFK together with WB and LCK. He performed the structural analyses, and he designed the binding pocket mutants together with AL. In addition, he wrote the manuscript together with the other authors. Furthermore, he was involved in the planning of all experiments. Parts of this work are described in the Bachelor Thesis of LCK (Tuebingen, 2013).

Caraballo R, Saleeb M, Bauer J, **Liaci AM**, Chandra N, Storm RJ, Frängsmyr L, Quiean W, Stehle T, Arnberg N, Elofsson M (2015) Triazole linker-based trivalent sialic acid inhibitors of adenovirus type 37 infection of human corneal epithelial cells. Organic and Molecular Biochemistry 13:9194-205. doi: 10.1039/c5ob01025j

This article is part of the themed collection '2015 Hot Articles in Organic and Biomolecular Chemistry' and was featured on the cover of Issue 35 (2015).

Contributions: AML purified HAdV-D37 FK together with JB and crystallized and solved the complex structures of compounds 17a and 17b. He wrote the paper together with the other authors.

The contributions to unpublished manuscripts are listed there in detail.

7.6. Publications and Unpublished Manuscripts

PLOS | PATHOGENS

RESEARCH ARTICLE

Structural and Functional Analysis of Murine Polyomavirus Capsid Proteins Establish the Determinants of Ligand Recognition and Pathogenicity

Michael H. C. Buch[1☯], A. Manuel Liaci[1☯], Samantha D. O'Hara[2], Robert L. Garcea[2], Ursula Neu[1¤]*, Thilo Stehle[1,3]*

1 Interfaculty Institute of Biochemistry, University of Tübingen, Tübingen, Germany, **2** Department of Molecular, Cellular, and Developmental Biology, and the BioFrontiers Institute, University of Colorado, Boulder, Colorado, United States of America, **3** Department of Pediatrics, Vanderbilt University School of Medicine, Nashville, Tennessee, United States of America

☯ These authors contributed equally to this work.
¤ Current Address: MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London, UK
* uneu@nimr.mrc.ac.uk (UN); thilo.stehle@uni-tuebingen.de (TS)

OPEN ACCESS

Citation: Buch MHC, Liaci AM, O'Hara SD, Garcea RL, Neu U, Stehle T (2015) Structural and Functional Analysis of Murine Polyomavirus Capsid Proteins Establish the Determinants of Ligand Recognition and Pathogenicity. PLoS Pathog 11(10): e1005104. doi:10.1371/journal.ppat.1005104

Editor: Craig Meyers, Penn State University School of Medicine, UNITED STATES

Received: March 18, 2015

Accepted: July 22, 2015

Published: October 16, 2015

Data Availability Statement: All relevant data are within the paper and its Supporting Information files except for the protein sequences of MuPyV RA VP1 and PTA VP1, which are available from UniProt under the accession numbers P49302 and Q76TX8, respectively, as well as the PDB entries of the constructs used in this publication (PDB IDs 5CPU, 5CPW, 5CPX, 5CPY, 5CPZ, and 5CQ0). Please also refer to Table 2 in the manuscript.

Funding: MHCB, AML, UN, and TS are grateful for financial support from the Deutsche Forschungsgemeinschaft SFB 685 (www.sfb685.de).

Abstract

Murine polyomavirus (MuPyV) causes tumors of various origins in newborn mice and hamsters. Infection is initiated by attachment of the virus to ganglioside receptors at the cell surface. Single amino acid exchanges in the receptor-binding pocket of the major capsid protein VP1 are known to drastically alter tumorigenicity and spread in closely related MuPyV strains. The virus represents a rare example of differential receptor recognition directly influencing viral pathogenicity, although the factors underlying these differences remain unclear. We performed structural and functional analyses of three MuPyV strains with strikingly different pathogenicities: the low-tumorigenicity strain RA, the high-pathogenicity strain PTA, and the rapidly growing, lethal laboratory isolate strain LID. Using ganglioside deficient mouse embryo fibroblasts, we show that addition of specific gangliosides restores infectability for all strains, and we uncover a complex relationship between virus attachment and infection. We identify a new infectious ganglioside receptor that carries an additional linear [α-2,8]-linked sialic acid. Crystal structures of all three strains complexed with representative oligosaccharides from the three main pathways of ganglioside biosynthesis provide the molecular basis of receptor recognition. All strains bind to a range of sialylated glycans featuring the central [α-2,3]-linked sialic acid present in the established receptors GD1a and GT1b, but the presence of additional sialic acids modulates binding. An extra [α-2,8]-linked sialic acid engages a protein pocket that is conserved among the three strains, while another, [α-2,6]-linked branching sialic acid lies near the strain-defining amino acids but can be accommodated by all strains. By comparing electron density of the oligosaccharides within the binding pockets at various concentrations, we show that the [α-2,8]-linked sialic acid increases the strength of binding. Moreover, the amino acid

SDO and RLG acknowledge the National Institutes of Health R01-CA37667 (www.nih.gov). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

exchanges have subtle effects on their affinity for the validated receptor GD1a. Our results indicate that both receptor specificity and affinity influence MuPyV pathogenesis.

Author Summary

Viruses are obligate intracellular pathogens, and all of them share one crucial step in their life cycle—the attachment to their host cell via cellular receptors, which are usually proteins or carbohydrates. This step is decisive for the selection of target cells and virus entry. In this study, we investigated murine polyomavirus (MuPyV), which attaches to host gangliosides with its major capsid protein, VP1. We have solved the crystal structures of VP1 in complex with previously known interaction partners as well as with the ganglioside GT1a, which we have identified as a novel functional receptor for MuPyV. Earlier studies have shown that different strains with singular amino acid exchanges in the receptor binding pocket of VP1 display altered pathogenicity and viral spread. Our investigations show that, while these exchanges do not abolish binding or significantly alter interaction modes to our investigated carbohydrates, they have subtle effects on glycan affinity. The combination of receptor specificity, abundance, and affinity reveals a much more intricate regulation of pathogenicity than previously believed. Our results exemplify how delicate changes to the receptor binding pocket of MuPyV VP1 are able to drastically alter virus behavior. This system provides a unique example to study how the first step in the life cycle of a virus can dictate its biological properties.

Introduction

The engagement of one or several host cell receptors is the first step in the infectious cycle of a virus. A large number of viruses, including many human pathogens, depend on carbohydrate recognition for initial attachment to the cell surface. Viral tropism and the internalization pathway are usually determined by the specificity and affinity of the receptor interaction as well as the glycan distribution on different cell surfaces (reviewed in [1]). Many viruses use glycoproteins, glycolipids, or both as receptors for cell entry [2]. Gangliosides are ubiquitous glycolipids on the outer leaflet of mammalian cell membranes that serve as receptors for a number of viruses. They are composed of a membrane-embedded ceramide moiety linked to a complex carbohydrate structure that projects away from the cell. Gangliosides almost always contain α-5-*N*-acetyl-neuraminic acid (sialic acid, Neu5Ac) that can be attached to the core of the molecule with [α-2,3], [α-2,6], or [α-2,8] linkages (Fig 1). Gangliosides exist on cell surfaces in complex and poorly understood patterns that are cell type-, age-, and tissue-dependent ([3,4], reviewed in [5]).

Murine Polyomavirus (MuPyV) is a double-stranded DNA virus that can induce tumors in newborn animals. It was long known to engage glycan receptors that contain a minimal motif of sialic acid [α-2,3]-linked to galactose [6,7], and more recently gangliosides GD1a and GT1b were identified as MuPyV receptors [8]. Viral attachment is mediated by the major capsid protein, VP1, which forms pentameric capsomers that assemble into the T = 7*d* icosahedral capsid of the virus [9,10]. Sialylated oligosaccharide receptors are engaged in a shallow groove on top of VP1 formed by loop structures on the protein surface [11–13], similar to other polyomaviruses [1].

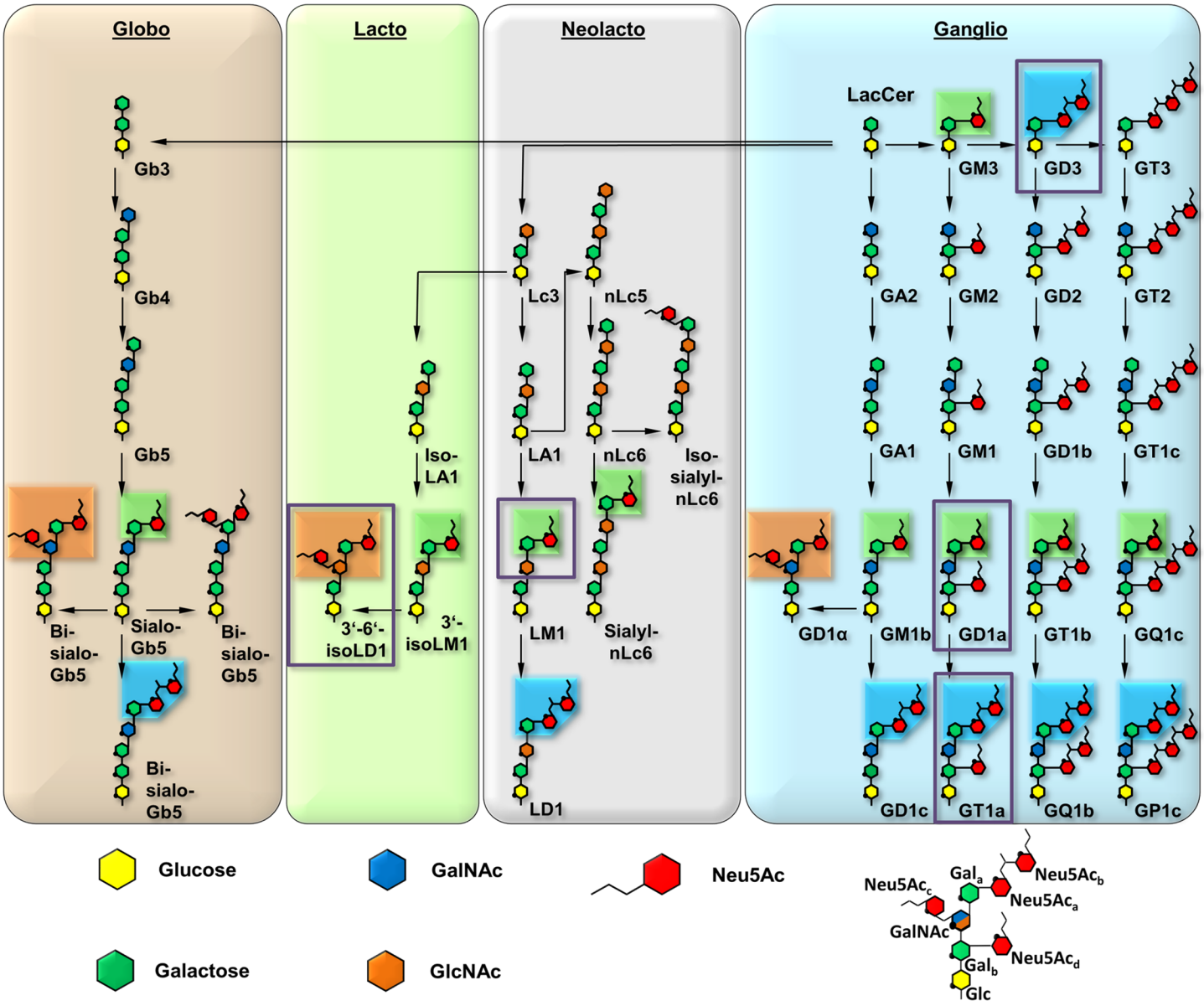


Fig 1. Overview and biosynthetic pathway of the four most prominent ganglioside series. The glycan parts of important members are shown for each series. The downstream biosynthetic steps are identical for all members of a row, although they may vary in linkage orientation. The six-membered pyranose rings are numbered counterclockwise, starting from the bottom (C1, except for C2 in Neu5Ac), and the ring oxygen is symbolized with a black dot. Neu5Ac moieties are rearranged for clarity, and all linkages are mediated by O2 or O8. Most of the gangliosides (e.g. LM1) can be further modified, e.g. by fucosylation. Linkages involving Neu5Ac are present in the α conformation, all other linkages are in the β conformation. Boxes represent three distinguishable sialoglycotopes that contain linkages found in GT1a (blue, representative for [α-2,8]), GD1a (green, [α-2,3]), and 3'-6'-iso-LD1 (also referred to as DSLNT, orange, [α-2,6]). The naming is according to the corresponding gangliosides; if possible, the Svennerholm shorthand is used [64–66] All biosynthetic routes were verified using the KEGG metabolic pathway database [67]. A prototype glycan that exemplifies the different positions of Gal and Neu5Ac moieties is depicted on the lower right. The glycan portions investigated in this study are highlighted by purple boxes.

doi:10.1371/journal.ppat.1005104.g001

MuPyV displays striking differences in pathogenicity and spread among three closely related prototype strains upon infection of newborn virus-free mice. The laboratory-derived RA strain [14] shows limited spread and induces few tumors of strictly mesenchymal origin after a long latency period, while the naturally occurring PTA strain [15,16] has disseminated infection and causes multiple tumors of epithelial and mesenchymal origin within a short time.

Table 1. Description of the investigated MuPyV strains.

	RA Strain	PTA Strain	LID Strain
Distinctive amino acids	G91, V296	E91, V296	E91, A296
Pathogenicity	No or only singular tumors, mesenchymal origin.	High tumor density of epithelial and mesenchymal origin.	Virulent. Damage of host tissues, early death due to brain hemorrhages and kidney failure.
Latency	Long	Short	Very short

doi:10.1371/journal.ppat.1005104.t001

LID [17,18], another laboratory isolate MuPyV strain, spreads most rapidly, causing early death by damaging host tissues, leading to brain hemorrhages and kidney failure [19]. The differences among the three strains have been mapped to amino acid variations at two positions, 91 and 296, within the receptor-binding region of VP1 [20–24]. While RA bears a glycine residue at position 91, this residue is replaced with a glutamate in both PTA and LID. An additional valine-to-alanine exchange at position 296 is present in LID (Table 1). The pathogenicity profile of one strain can be introduced into the other strains by mutating these two residues, confirming that these substitutions are necessary and sufficient to generate a specific phenotype [25]. The same substitutions have also been observed for other strains of MuPyV [21,22]. MuPyV found in feral mice has the VP1 sequence of PTA [26], but the virus is controlled by an intact immune system. As studies of viral spread can be conducted *in vivo* and virus infectivity can be tested *in vitro* using ganglioside deficient mouse cells, MuPyV represents an attractive and rare model system to define the relationships between receptor binding and viral spread and tropism.

Crystal structures of the low pathogenicity strain RA have shown how this virus engages 3'-sialyllactose, a short, linear trisaccharide terminating in [α-2,3]-linked sialic acid, as well as an oligosaccharide that additionally contains a second, branching [α-2,6]-linked sialic acid [11,12]. These structures also identified the location of residues 91 and 296 in the carbohydrate-binding region, suggesting that they might modulate interactions of VP1 with its receptors in the higher pathogenicity strains PTA and LID. Modelling suggested that a glutamate side chain at position 91 would lead to electrostatic repulsion of the [α-2,6]-branched sialic acid, thereby preventing binding of such branched structures by either LID or PTA. Branched sugars carrying an [α-2,6]-linked sialic acid could thus act as pseudoreceptors that will not facilitate productive infection but hamper the spread of RA within the host, in contrast to PTA and LID [8,12]. In line with this hypothesis, gangliosides GD1a and GT1b, which do not contain an [α-2,6]-branched sialic acid, have been identified as entry receptors for the PTA [8,16] and RA strain [27] of MuPyV. However, the molecular determinants of GD1a or GT1b receptor interactions with PTA and LID are not understood, because all structural information is limited to date to RA MuPyV.

To define the interactions of the three MuPyV strains with receptors on the cell surface, we have solved high-resolution structures of RA, PTA, and LID VP1 pentamers in complex with three ganglioside glycans that represent common motifs found in members of the four most prominent ganglioside biosynthesis series and that feature [α-2,3]-, [α-2,6]-, and [α-2,8]-linked sialic acids (for carbohydrate structures, nomenclature, and annotations see Fig 1). We have also conducted crystallographic soaking experiments at different ligand concentrations to compare the relative affinities of each of the three strains for their interaction partners. We find that expanding the well-characterized Neu5Ac-[α-2,3]-Gal epitope with a linear [α-2,8]-linked sialic acid (as found for example in GT1a vs. GD1a) leads to additional interactions between carbohydrate and VP1 in all three strains. Consequently, we identify ganglioside GT1a as an infectious receptor for all three strains. Moreover, the branching [α-2,6]-linked sialic acid is

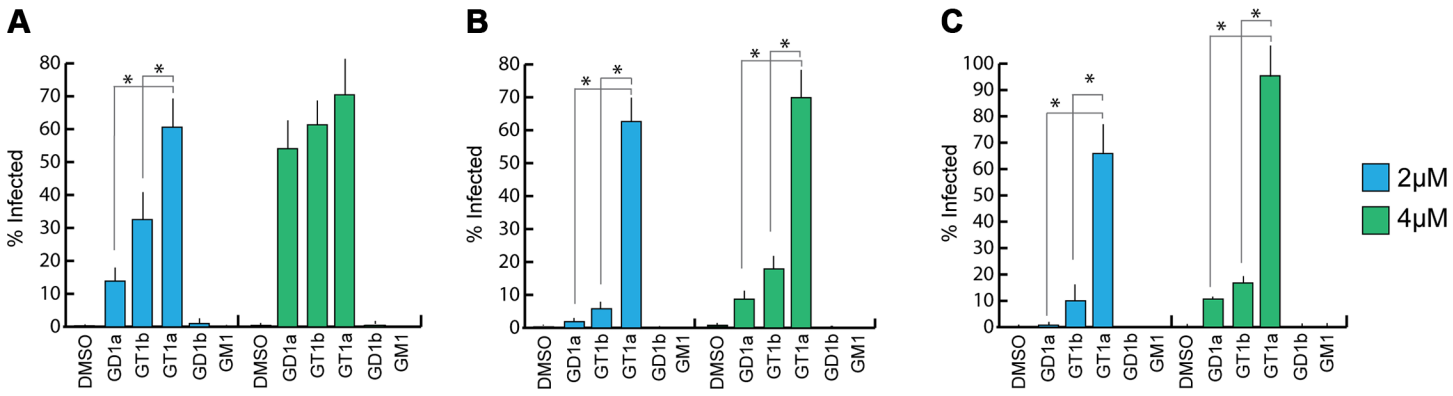


Fig 2. GT1a, GT1b, and GD1a supplementation rescues MuPyV infection of Gang-/- MEFs. Ganglioside knock-out (Gang-/-) MEFs were completely resistant to infection of all strains of MuPyV as shown by the absence of T-antigen positive nuclei at 24 hours post infection (DMSO control). GD1a, GT1b, and GT1a ganglioside supplementation of Gang-/- MEFs restored RA (**A**), PTA (**B**), and LID MuPyV (**C**) infection, while GD1b and GM1 supplementation resulted in little to no infection by any virus strain. Infection levels were quantified at both 2 μM and 4 μM ganglioside supplementation (blue and green bars, respectively). Infection levels are normalized to MuPyV infection of WT MEFs, and error bars correspond to standard error.

doi:10.1371/journal.ppat.1005104.g002

close to the strain-defining amino acids, but can be accommodated by all strains, in contrast to the earlier model. However, the amino acid exchanges defining each strain have subtle effects on their affinity for the validated receptor GD1a. Our results exemplify the effect of minimal changes in a binding pocket on the receptor binding properties of a virus.

Results

GT1a, GD1a, and GT1b gangliosides are infectious receptors for MuPyV

Previous efforts to identify receptors for MuPyV used immortalized cell lines, such as Vero or C6 glioma cells that were supplemented with candidate gangliosides before infection [8,28]. We utilized a mouse embryo knock-out fibroblast cell line (Gang-/- MEFs) specifically deficient in ganglioside synthesis and completely resistant to MuPyV infection (S1A Fig and [29]) to test the ability of ganglioside receptors to rescue infection by different strains of MuPyV. Gang-/- MEFs were supplemented with individual gangliosides followed by infection with RA, PTA, and LID MuPyV (Fig 2). Importantly, it should be noted that the three MuPyV strains we used do not have the same particle to PFU ratio. The viruses have been normalized to similar MOIs, but they cannot be quantitatively compared to one another. However, each strain has been normalized to its own infection rate of WT MEFs; therefore, infection rates upon supplementation of gangliosides can be compared within a strain. The previously identified ganglioside receptors GD1a and GT1b [8] rescued RA, PTA, and LID infection of Gang-/- MEFs in a dose-responsive manner. We also analyzed the GT1a ganglioside that had not been previously investigated as a candidate infectious receptor for MuPyV. We found that GT1a, a member of the ganglio-series synthesized from GD1a (Fig 1), also rescued RA, PTA, and LID infection in a dose responsive manner (Fig 2). Moreover, GT1a supplementation of Gang-/- MEFs conferred higher levels of RA, PTA, and LID MuPyV infection than the previously identified receptors GD1a and GT1b. Finally, we tested the ability of the gangliosides GD1b and GM1 to rescue MuPyV infection of Gang-/- MEFs. GD1b and GT1b supplementation has previously been shown to restore BK polyomavirus infection of ganglioside deficient cells [30]; however, GD1b restored little to no MuPyV infection of Gang-/- MEFs. GM1 supplementation has previously been shown to restore infection by SV40 [8]; however, GM1 did not rescue MuPyV infection of Gang-/- MEFs. These data confirm that GT1a is an infectious receptor for all strains of MuPyV.

We also investigated whether MuPyV cell surface binding to infectious or non-infectious ganglioside receptors correlated with infection. To this end, we measured the levels of free (unbound) virus in each ganglioside supplemented sample at 4 hours post infection. We did not detect significant differences in MuPyV cell surface binding to different ganglioside receptors or WT MEFs, indicating that cell surface binding alone does not determine infection (S1B Fig). Instead, a considerable amount of virus binds to Gang-/- MEFs even in the absence of ganglioside supplementation (S1A Fig). MuPyV is also endocytosed in Gang-/- MEFs, which however does not lead to infection [29]. Taken together, these data confirm that gangliosides are not required for cell surface binding. They are, however, required for infection, and GT1a appears to be more efficient than GD1a and GT1b.

Structure of MuPyV VP1 bound to GT1a

In order to define the mode of recognition of GT1a, particularly to the naturally occurring PTA strain of MuPyV, we have soaked VP1 crystals with the glycan portion of GT1a and solved the structure of the complex (Table 2). While the receptor interaction pocket of RA VP1 has been described [11–13], no structural information for the pathogenicity-defining amino acids at positions 91 and 296 in the pockets of PTA and LID has been available. PTA and LID both carry a glutamate at position 91, and this side chain is being held in a characteristic position with the carboxyl group facing away from the glycan receptor due to a salt bridge formed with K186 (Fig 3), as previously predicted [12]. The GT1a glycan is a branched structure with a long disialylated arm, which has the sequence Neu5Ac$_b$-[α-2,8]-Neu5Ac$_a$-[α-2,3]-Gal$_a$-[β-1,3]-GalNAc, and a second short arm, which consists of a single Neu5Ac$_d$ [α-2,3]-linked to Gal$_b$ (for carbohydrate structures, nomenclature, and moiety indexing see Fig 1). The disialylated arm of GT1a is clearly visible in the crystal structure of PTA VP1; it is well defined by electron density and makes extensive contacts with the protein (Fig 4B–4D). Overall, the GT1a glycan adopts a twisted horseshoe-like shape, with Neu5Ac$_a$ and Neu5Ac$_b$ wrapping around the side chains of Y72 and R77 of VP1. Its longer, disialylated arm contains a Neu5Ac$_a$-[α-2,3]-Gal$_a$ sequence that is also present in GD1a and simpler compounds such as 3'-sialyllactose (3SL), and the interactions of this motif with VP1 are essentially identical to those seen in previous structures [11–13]. However, our structure visualizes an additional network of contacts made by the terminal [α-2,8]-linked Neu5Ac$_b$ (Fig 4C and 4D). Its carboxyl group engages Y72 and forms water-mediated hydrogen bonds with Q71, Y72, as well as D85 of the neighboring monomer (D85*). In addition, the *N*-acetyl nitrogen of Neu5Ac$_b$ forms a hydrogen bond with the backbone carbonyl of T67, and O8 and O9 in the glycerol chain of the sugar are hydrogen-bonded with the R77 side chain. The carboxyl groups of Neu5Ac$_a$ and Neu5Ac$_b$ are about 4 Å apart, and the positively charged side chain of R77 counteracts their negative charges (Fig 4C and 4D). Neu5Ac$_a$ and Neu5Ac$_b$ contribute binding interfaces of approximately 160 Å^2 and 190 Å^2, respectively (calculated using the PISA server [31]). The remaining Gal$_a$-GalNAc-Gal$_b$ stem of GT1a forms fewer contacts with the protein, which include a hydrogen bond between G78 and the Gal$_a$ O4 hydroxyl group (Fig 4) as well as several van der Waals interactions. Notably, the C$_\beta$ and C$_\gamma$ atoms of E91 are within van-der-Waals range of O6 and C6 of Gal$_a$, and the E91 carboxylate group is close to C6 of GalNAc. The total contact surface for this portion of the glycan is 142 Å^2.

Because the differences in tumorigenicity and host spread among strains have been mapped to the glycan binding pocket of VP1, and because GT1a appears to be particularly efficient in facilitating productive infection, we set out to determine how the three strains engage GT1a. By solving the crystal structures of RA and LID VP1 complexed with GT1a using the identical strategy used for the PTA-GT1a complex, we found that the overall binding mode of GT1a is

Table 2. Data collection and refinement statistics.

	PTA VP1 Native	PTA VP1 + GT1a	PTA VP1 + GD1a	PTA VP1 + DSLNT	RA VP1 + GT1a	RA VP1 + GD1a
Data Collection						
Beamline	SLS, X06DA	SLS, X06DA	SLS, X06DA	ESRF, ID 14–1	SLS, X06DA	SLS, X06DA
Space Group	$P3_121$	$P3_121$	$P3_121$	$P3_121$	$P3_121$	$P3_121$
Cell Dimensions						
a = b, c [Å]	219.61, 99.82	219.60, 99.74	220.45, 99.71	219.73, 100.00	219.55, 99.60	219.06, 99.40
α = β, γ [°]	90, 120	90, 120	90, 120	90, 120	90, 120	90, 120
Resolution [Å]	50–1.64 (1.68–1.64)	50–1.75 (1.79–1.75)	50–1.93 (1.98–1.93)	50–1.87 (1.92–1.87)	50–1.71 (1.75–1.71)	50–1.90 (1.95–1.90)
R_{meas} [%]	10.5 (68.7)	13.3 (68.3)	11.3 (68.8)	15.2 (69.1)	7.4 (74.1)	11.1 (71.7)
I/σ(I)	10.5 (2.3)	7.08 (1.51)	12.8 (3.0)	7.6 (3.0)	15.71 (2.17)	11.3 (1.9)
Completeness [%]	99.9 (99.9)	97.6 (96.6)	96.0 (98.1)	99.8 (99.8)	99.8 (99.9)	99.0 (99.1)
Redundancy	5.0 (4.8)	3.4 (3.2)	5.4 (5.5)	3.7 (3.7)	4.3 (3.9)	2.9 (2.8)
Wilson B-Factor [$Å^2$]	23.1	25.3	25.7	23.1	26.0	23.3
Refinement						
Resolution [Å]	48.2–1.64	47.6–1.75	47.8–1.93	48.4–1.83	50–1.71	48.1–1.90
No. of Reflections	324,802	261,253	192,327	220,105	285,887	205,733
R_{work} / R_{free} [%]	15.85 / 17.30	16.0 / 18.13	15.27 / 17.42	15.38 / 17.56	15.27 / 16.98	15.46 / 17.84
No. of Atoms						
Protein	11,117	11,088	11,150	10,996	11,323	11,225
Solvent	1,840	1,827	1,884	1,632	2,059	1,971
Carbohydrate	-	425	285	202	385	285
B-Factors [$Å^2$]						
Protein	19.3	20.7	21.4	20.1	21.1	19.9
Solvent	29.6	30.9	31.5	30.5	32.6	31.1
Carbohydrate	-	30.3	35.1	34.2	36.0	38.2
R. m. s. Deviations						
Bond Lengths [Å]	0.007	0.007	0.006	0.008	0.007	0.008
Bond Angles [°]	1.16	1.17	1.06	1.20	1.13	1.19
Ramachandran Plot						
Favored	1,340 (97.2%)	1,336 (97.0%)	1,335 (96.9%)	1,334 (96.9%)	1,342 (97.0%)	1,340 (97.0%)
Allowed	38 (2.8%)	41 (3.0%)	42 (3.2%)	42 (3.1%)	42 (3.0%)	42 (3.0%)
Disallowed	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
PDB ID	5CPU	5CPW	5CPY	5CPX	5CPZ	5CQ0

doi:10.1371/journal.ppat.1005104.t002

very similar across the three strains (Fig 5A), with a conserved binding mode of the [α-2,8]-linked $Neu5Ac_b$. Although the replacement of glutamate with glycine at position 91 leads to a contact area decrease of 33 $Å^2$ in RA, the orientation of GT1a in this strain is not altered (compare Fig 5B and 5C). Likewise, the substitution of valine with alanine at position 296 in LID removes a hydrophobic contact but does not affect the conformation of GT1a (Fig 5E; S2 Fig).

The $Neu5Ac_a$-Gal_a-GalNAc linkages in the long arm of GT1a adopt conformations that have been reported in numerous structures (for example [32–34]). While the [α-2,3] linkage between $Neu5Ac_a$ and Gal_a adopts the conformation that has been reported for DSLNT and 3SL, the branching $Neu5Ac_d$-[α-2,3]-Gal_b linkage adopts a different conformation, which has been reported for structures containing O-4-substituted galactoses (as in [35,36]). While a higher variability is observed for Neu5Ac-[α-2,8]-Neu5Ac linkages (S2E Fig), this linkage adopts torsion angles that are in agreement with other, related structures such as in the

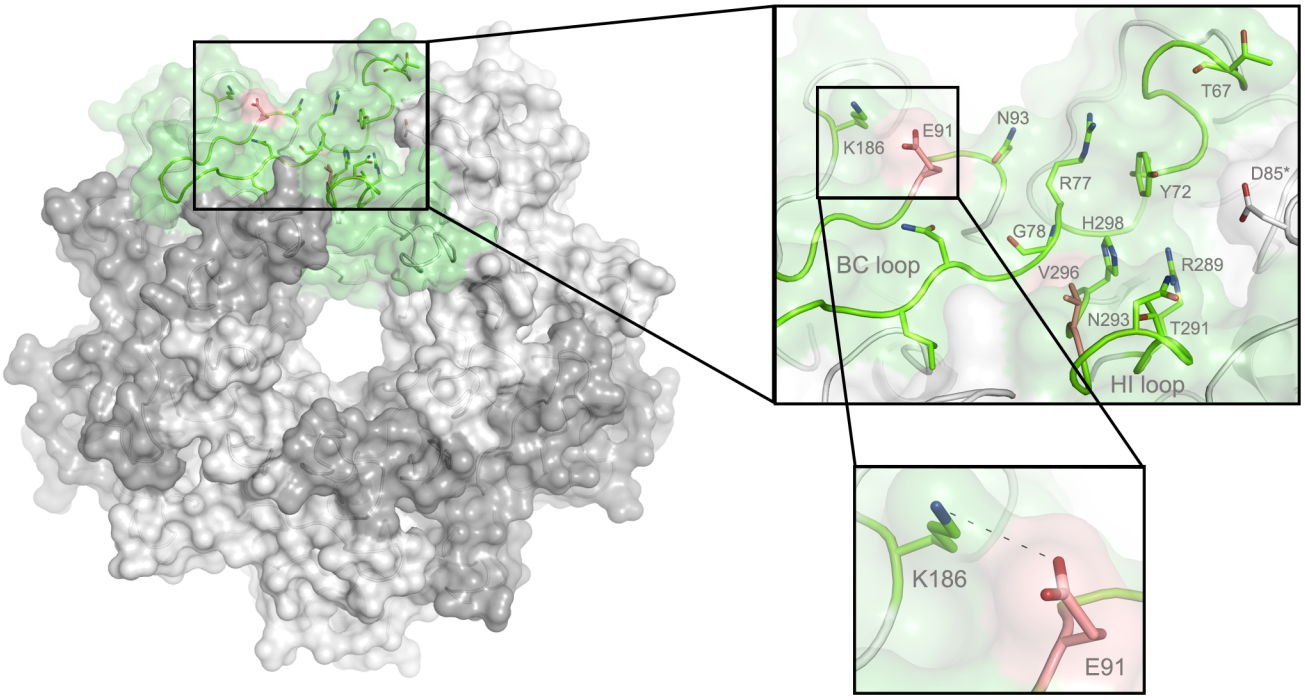


Fig 3. The MuPyV binding pocket. Top view on the receptor-binding region of PTA, which is shown with E91 and V296 highlighted in salmon. Residues that are known to participate in receptor binding are contributed by the BC and HI loops and are highlighted as stick models. One monomer is shaded in green and the other monomers are alternatingly shaded light and dark grey for better distinction.

doi:10.1371/journal.ppat.1005104.g003

structure of human liver fructose-1,6-bisphosphatase in complex with an allosteric inhibitor [37] or in the complex of tetanus toxin with a GT1b analog [38]. The overall structure is in good agreement with a molecular dynamics simulation using an AMBER force field in an aqueous environment [39]. A well-defined set of water molecules mediates bridged hydrogen bonds between the pyranose moieties, especially between $Neu5Ac_b$ and $Neu5Ac_d$ (S3 Fig). Due to these steric constraints, the GT1a complexes feature well-defined electron density not only for the binding epitope, but also for the non-binding, branching $NeuNAc_d$ in its preferred solution conformation [40], which brings this moiety to about 5 Å near the end of the long arm and gives the glycan the characteristic, horseshoe-like topology that is observable in all complex structures.

Structures of MuPyV VP1 strains bound to other sialylated glycans

As RA, PTA, and LID VP1 all bind GT1a in a highly similar conformation, we hypothesized that the differences in pathogenicity and spread among the three strains might be due to the recognition of additional carbohydrates by only a subset of MuPyV strains. As shown in Fig 1, the many different gangliosides share a relatively small set of common sialoglycotopes. We therefore investigated the ability of all three VP1 proteins to bind other glycan structures that are representative for these epitopes. We solved structures of VP1 bound to the glycan portions of two of these gangliosides: The GD1a glycan is an established infectious receptor and essentially a truncated version of GT1a lacking the [α-2,8]-linked $Neu5Ac_b$ in the long arm. The human milk hexasaccharide DSLNT is the glycan portion of the lacto-series ganglioside 3'-6'-isoLD1 (Fig 1) [41], which is overexpressed in the central nervous system. In contrast to GT1a and GD1a, DSLNT does not contain an [α-2,3]-linked $Neu5Ac_d$ as a short arm but instead a

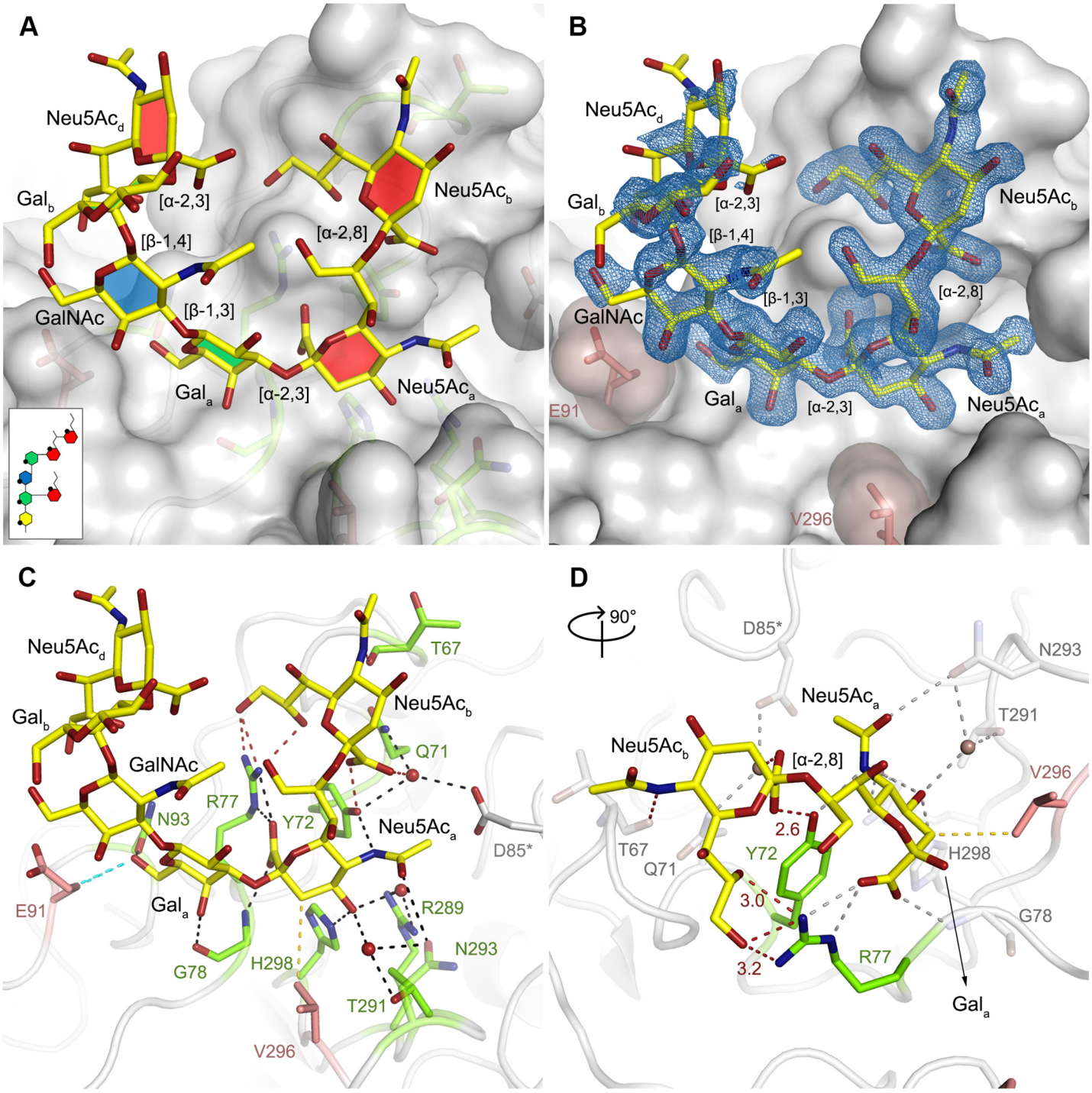


Fig 4. Binding of GT1a to PTA. A The PTA binding pocket and the GT1a conformation upon binding are shown from an angle parallel to the fivefold axis. A scheme of the glycan is shown in the inset, and the sugar rings are filled according to the coloring scheme from Fig 1. **B** Simulated annealing F_{obs}-F_{calc} omit map (resolution 1.71 Å, calculated at 3.5 σ, carved 2.3 Å around the glycan). **C** Possible binding interactions of GT1a and PTA. E91 and V296 are highlighted in salmon. Hydrogen bonds are shown in black, the hydrophobic contact mediated by V296 in gold, and the van-der-Waals contacts of E91 are shown in cyan. Waters that mediate key hydrogen bonds are shown as red spheres. Unique interactions mediated by the novel GT1a-like binding motif are shown in red. **D** Zoomed view of the binding to the two terminal Neu5Ac moieties. The rest of the glycan is omitted for clarity. Residues except Y72 and R77 as well as waters involved in contacts with these two glycan moieties are pale grey and salmon, respectively.

doi:10.1371/journal.ppat.1005104.g004

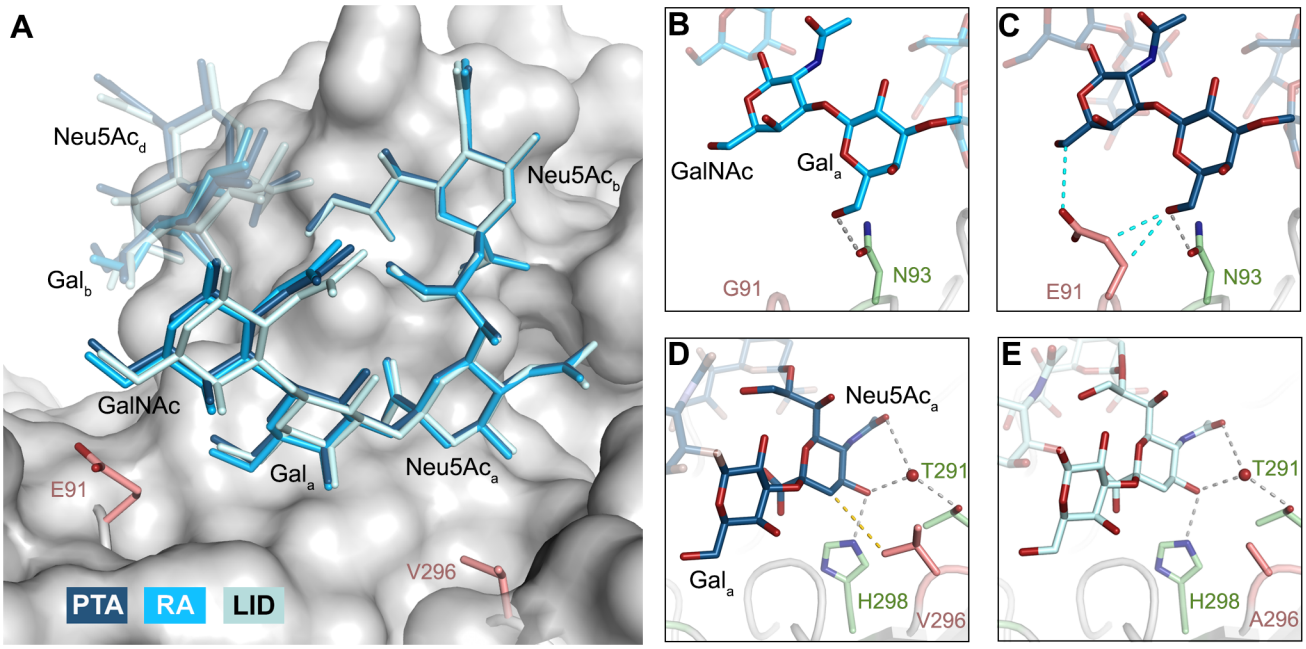


Fig 5. Binding modes of GT1a to the different MuPyV strains. A Superposition of the GT1a-binding mode of RA (GT1a in sky blue), PTA (dark blue), and LID (pale blue). The Neu5Ac$_b$-[α-2,8]-Neu5Ac$_a$-[α-2,3]-Gal$_a$ motif is shown in solid sticks, together with the adjacent GalNAc moiety. All superpositionings were carried out in PyMOL [61] using 'align' for the protein chains only. Surface, E91 and V296 are from PTA/GT1a. All 'align' rmsd values are below 0.16 Å. **B & C** Close view of the van-der-Waals contacts introduced by the E91 side chain present in PTA and LID (C), but not in RA (B). Hydrogen bonds are shown in grey, van-der-Waals contacts in cyan. **D & E** Close view of the hydrophobic contact mediated by V296 in RA and PTA (D), but not by A296 in LID (E). The 4.0 Å hydrophobic contact is not present in the LID strain, whose pocket is opened to the right. Hydrogen bonds are shown in grey, hydrophobic contacts are shown in gold.

doi:10.1371/journal.ppat.1005104.g005

branching [α-2,6]-linked Neu5Ac$_c$. This structure is similar to a very common epitope on O-linked glycoproteins [42–44]. DSLNT was used in previous studies of MuPyV as a model "pseudoreceptor" [12] and was investigated here to help rationalize these earlier data, to facilitate a comparison among strains, and to establish a binding profile for glycans containing an [α-2,6]-linked sialic acid.

GD1a. The previously identified MuPyV receptor GD1a is similar to a truncated GT1a structure, containing only a Neu5Ac$_a$-[α-2,3]-Gal$_a$ motif instead of Neu5Ac$_b$-[α-2,8]-Neu5Ac$_a$-[α-2,3]-Gal. The disaccharide engages all three strains in a very similar manner (Fig 6A). Neither the longer E91 side chain (in PTA and LID) nor the shorter A296 side chain (in LID) result in an altered conformation of the ligand.

DSLNT. The DSLNT glycan terminates in a Neu5Ac$_a$-[α-2,3]-Gal$_a$ motif, which is the part of the molecule best defined by electron density in all complexes. DSLNT also contains an additional [α-2,6]-linked, branched Neu5Ac$_c$ residue, which is not present in either GT1a or GD1a. There is weak electron density for Neu5Ac$_c$ in one of the five binding pockets of the RA strain, but it only engages in few interactions [12]. While PTA and LID do bind DSLNT, the complex structures do not show any electron density for the Neu5Ac$_c$, indicating that this sugar is conformationally flexible and does not contribute contacts. When bound to the PTA strain, the stem of DSLNT is moderately rearranged (Fig 6B). In comparison to GalNAc in GT1a, the GlcNAc moiety is slightly tilted away from E91 due to a ~20° rotation of the psi angle in the Gal$_a$-[β-1,3]-GlcNAc linkage (Fig 6B and 6C, assessed using CARP) that propagates throughout the sugar. In addition, there is no visible electron density for the GlcNAc O6 that is engaged in the [α-2,6]-branching as well as an increased B-Factor variance within the

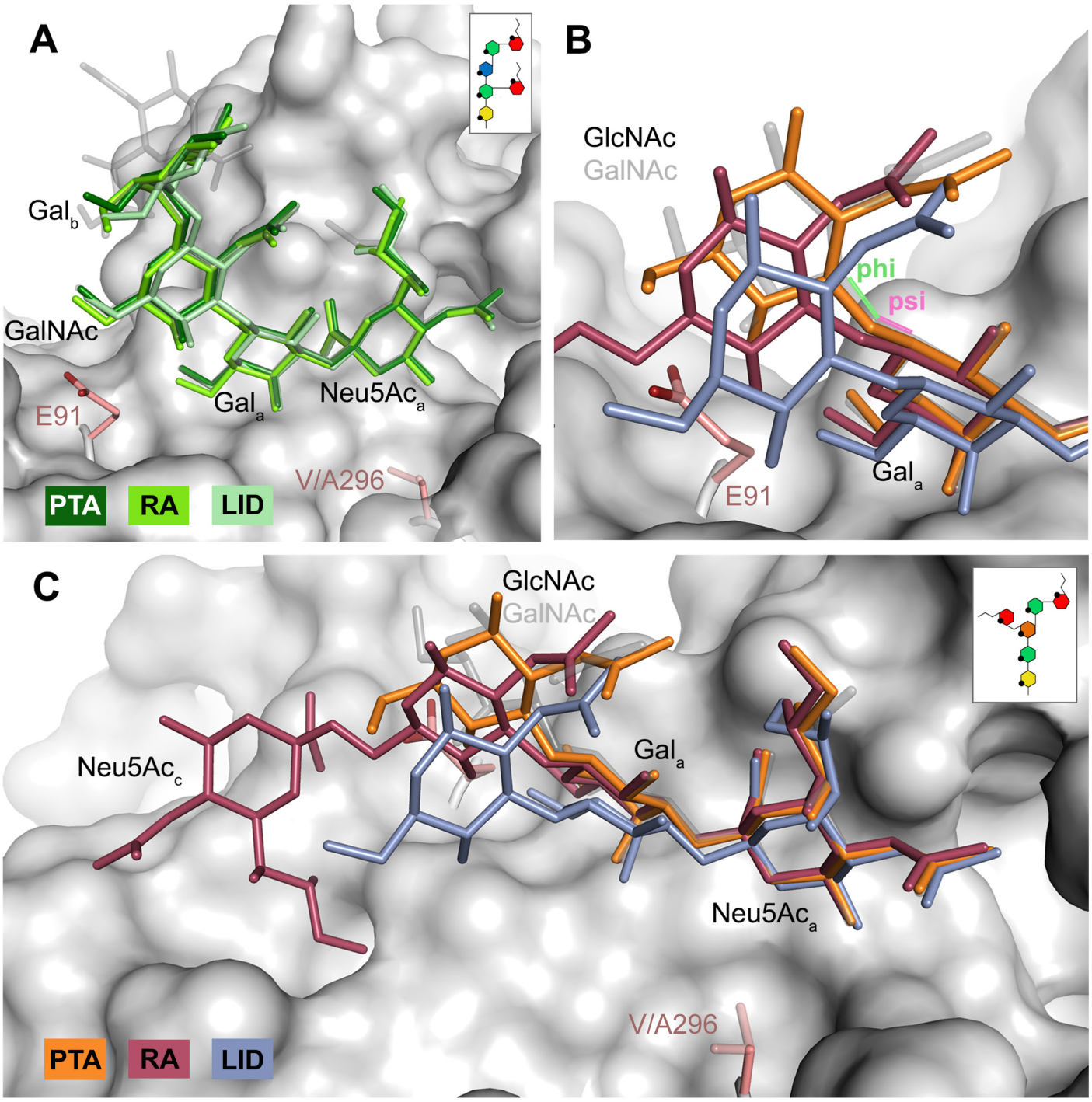


Fig 6. Differences in receptor binding patterns across glycans. A Superposition of the binding modes of GD1a to RA (light green), PTA (dark green), and LID (pale green). The sequence of GD1a is shown in the inset. The Neu5Ac$_a$-[α-2,3]-Gal$_a$ motif is shown in solid sticks, together with the adjacent GalNAc moiety. In all figures, GT1a bound to PTA is overlaid as a grey ghost for comparison, with Neu5Ac$_b$ omitted for clarity. Deviations exceeding the atomic error of the structure and alignment rmsd values are only found in the stem region of the sugar, starting at Gal$_b$. All superpositionings were carried out in PyMOL [61] using 'align' for the protein chains only. Surface, E91 and V296 are from PTA/GT1a. All 'align' rmsd values are below 0.16 Å. **B & C** Comparison of the DSLNT binding modes to RA (red), PTA (orange), and LID (violet). In PTA-DSLNT, [α-2,6]-branching causes a 15–20° psi angle shift of the GlcNAc moiety compared to GD1a and GT1a, resulting in a 1 Å sideward twist movement of the stem. In RA-DSLNT, combination of this shift with a 15° shift in the phi angle results in a downward movement of GlcNAc and its branching Neu5Ac$_c$ compared to PTA-DSLNT. In LID, the shift is already observable for Gal$_a$ and results in the loss of ordered density for GlcNAc. All angles were analyzed with CARP. The sequence of DSLNT is shown in the inset of panel C.

doi:10.1371/journal.ppat.1005104.g006

glycan (S4 Fig). The reason for the sideward twist and the missing electron density for $Neu5Ac_c$ observed in PTA is probably the electrostatic repulsion between the carboxyl groups of $Neu5Ac_c$ and E91. While the charge of E91 is compensated by K186 (Fig 3), as was hypothesized before [12], the two carboxylate groups would come within about 2 Å of one another if DSLNT bound to PTA in the same way as observed in RA. This hypothesis is confirmed by a PTA E91Q mutant that rescued binding of $Neu5Ac_c$ (S5 Fig). In turn, when bound to RA, DSLNT exhibits a stronger conformational rearrangement (Fig 6B and 6C). Due to the missing side chain at position 91, the psi angle rotation between Gal_a and GlcNAc is accompanied by an additional 15° rotation of the phi angle, bringing the GlcNAc moiety and the attached $Neu5Ac_c$ closer to the protein surface [12]. In LID, the valine to alanine mutation at position 296 reduces its van-der-Waals radius. This change results in a broader binding pocket compared to the other strains and the loss of a hydrophobic interaction between position 296 in VP1 and C3 of $Neu5Ac_a$ for all glycans. This gives room for a stark alteration in the binding mode of DSLNT that starts with a slight tilt of $Neu5Ac_a$ and propagates through the sugar (Fig 6C), ultimately resulting in a prominent sideward shift of the whole glycan stem. The resulting increased conformational freedom of DSLNT is reflected by a lack of electron density in its stem region as well as by an elevated temperature factor variance (S4 Fig). This alteration of the binding mode in LID is likely to be observed for other glycans that are not conformationally restrained by the [α-2,3]-linked $Neu5Ac_d$.

Relative affinities of MuPyV strains for sialylated glycans

Since all three MuPyV strains are able to engage the three different glycan structures in a largely identical manner, we reasoned that the differences in pathogenicity and spread might be attributable to subtle differences in affinity, rather than specificity, among the strains. The affinities of RA VP1 for 3'-sialyllactose and DSLNT were estimated to be in the low mM range [11]. Coupled with the high costs of glycans and the high amount required due to their low binding affinity, weak binding poses technical challenges for classical affinity measurements. We therefore utilized a crystallographic approach to quantitatively compare ligand binding. We crystallized all three VP1 pentamers in the same condition, and soaked each with the oligosaccharide portions of GT1a, GD1a, and DSLNT at different concentrations in parallel. X-ray data of all crystals were collected in the same manner, and the data sets were processed using the same protocol and integrated as described previously [45]. All data sets were processed in the same unit cell, scaled, and the bias-reduced difference electron density around the central $Neu5Ac_a$-[α-2,3]-Gal_a motif was quantified for each data set (see the Methods section for details). Our crystallization condition contains a high amount of ammonium sulfate, which competes with the carboxyl group of $Neu5Ac_a$ and has to be displaced by the carbohydrates. Therefore, our observed binding is weaker than in a physiological setting. However, while not yielding dissociation constants in the traditional sense, this method enables us to compare relative levels of binding across our three different strains and three different glycans.

The GT1a glycan exhibits the strongest binding in all three VP1 variants compared with DSLNT or GD1a (Fig 7A–7C), with no detectable difference between the strains (Fig 7D). This finding is in accord with our ganglioside add-back experiments in cell culture (Fig 2), which consistently showed higher levels of infection mediated by GT1a compared to GD1a. The stronger overall binding of GT1a can be attributed to the additional [α-2,8]-linked sialic acid present in GT1a ($Neu5Ac_b$), which contributes several interactions and an increased buried surface area. These contacts seem to outweigh the differences in van der Waals contacts with the side chains of E91 or V296, at least to the extent discernable in our assay.

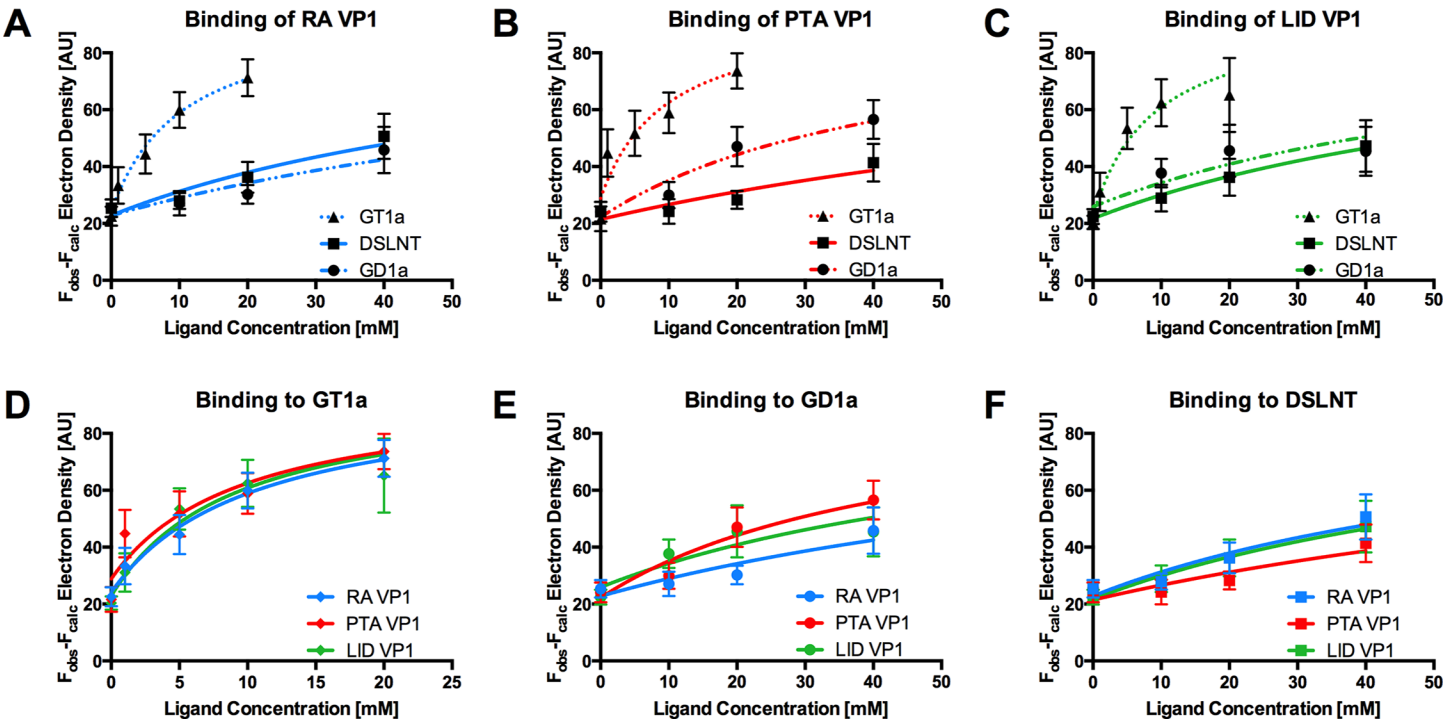


Fig 7. Binding of ligands to MuPyV VP1. The average simulated annealing F_{obs}-F_{calc} electron density for the Neu5Ac$_a$-[α-2,3]-Gal$_a$ in GT1a, GD1a, and DSLNT is plotted against ligand concentration. RA VP1 is colored blue, PTA VP1, is colored red, and LID VP1 is colored green. GT1a is displayed in a dotted line with triangles, GD1a in a dashed line with circles, and DSLNT in a solid line with squares. The error bars correspond to the standard deviation of the mean electron density observed in the five chains of VP1. **A** Electron density of GT1a, GD1a, and DSLNT in RA VP1. **B** Same as in **A**, but for PTA VP1. **C** Same as in **A**, but for LID VP1. **D** Comparison of GT1a-derived electron density in RA VP1, PTA VP1, and LID VP1. **E** Same as in **D**, but for GD1a. **F** Same as in **D**, but for DSLNT.

doi:10.1371/journal.ppat.1005104.g007

GD1a binds less well to all strains compared to GT1a. In addition, there are differences in binding strength among the three strains. PTA and LID VP1 appear to bind GD1a at the same level and better compared with RA (Fig 7A–7C and 7E) because these two strains gain additional interaction surface and van-der-Waals contacts from their E91 side chain. This effect is more pronounced than in GT1a, because in GD1a it cannot be masked by the additional contacts of the [α-2,8]-linked Neu5Ac$_b$.

DSLNT displays the lowest overall affinity to all strains, with levels comparable to GD1a in RA for all three strains (Fig 7A–7C) despite the DSLNT conformation being slightly different in each VP1 complex (Fig 7F). Neither the blocking of Neu5Ac$_c$ binding by E91, nor the increased conformational freedom in LID appears to alter binding affinity. It is possible that Neu5Ac$_c$ in RA adopts a conformation that might not be favorable and therefore not heavily contribute to affinity, in spite of the added contact surface. Combined with the fact that electron density for Neu5Ac$_c$ could only be observed in one binding pocket of RA VP1 [12], we believe that this conformation is possible but not probable in solution. Instead, an increased number of conformational options might make up for a loss of binding contacts.

Discussion

Many viruses engage cell-surface glycans to mount an infection, and subtle differences in the recognition of such receptors can be linked with altered tropism and pathogenicity. Examples include the canine parvovirus and feline panleukopenia virus [46,47], the human BK polyomavirus [48], B-lymphotropic polyomavirus [49,50] as well as avian and human influenzaviruses

[51,52]. However, MuPyV is a rare example of a virus in which drastic differences in pathogenicity directly correlate with single amino acid substitutions in the viral capsid.

In order to provide a structural basis for understanding the profoundly different pathogenicities of the three MuPyV strains RA, PTA and LID, we have solved structures of their VP1 proteins and characterized their receptor-binding properties. We show that the ganglioside GT1a serves as a MuPyV receptor and promotes infection with higher potency than the previously identified receptors GD1a and GT1b. Structurally, the increased potency of GT1a can be directly explained by a set of additional contacts involving the [α-2,8]-linked $Neu5Ac_b$ that is only present in this glycan and that gives it a characteristic horseshoe-like shape. It had previously been suggested that the G91E mutation present in PTA and LID abolishes binding to branched glycans containing [α-2,6]-linked Neu5Ac and thus allows the virus to spread more efficiently in the host [8,11]. However, our analyses show that the presence of a glutamate at position 91 still allows binding of the branched oligosaccharides GT1a, GD1a, or DSLNT to all three strains, albeit with subtle differences in binding affinity. While all three strains bind GT1a with comparable affinity, PTA and LID bind GD1a better than RA. The DSLNT glycan binds similarly to all three strains, with the lowest overall affinity. This is again in line with the structures, which show that the branched $Neu5Ac_c$ of DSLNT does not engage in any specific contacts. The limited contacts between $Neu5Ac_c$ and RA observed in an earlier structure [12] have to be considered a crystallization artifact as they were only observed in one out of five binding sites, and this visible $Neu5Ac_c$ moiety was located near a crystal contact.

The ligand binding promiscuity of MuPyV is surprisingly high. Binding mostly requires a ubiquitous minimal Neu5Ac-[α-2,3]-Gal motif, in agreement with earlier findings [6,7]. It therefore seems plausible that the virus also recognizes other glycans bearing this motif, resulting in differences in pathogenicity and spread. Preliminary studies show that glycans with an N-acetyllactose core (Neu5Ac-[α-2,3]-Gal-[β-1,4]-GlcNAc), as found in neolacto gangliosides such as the predominant ganglioside of peripheral nerve cells, LM1 [53,54], can also be bound in a manner similar to DSLNT and with higher flexibility than GT1a or GD1a (S6 Fig).

Based on our structures, certain requirements that contribute to receptor specificity can be established. For example, branches at Gal-O4 within the minimal motif produce clashes and cannot be tolerated. Therefore, although the GD1a glycan possesses two Neu5Ac-[α-2,3]-Gal motifs, it prefers the one on its longer arm for complex formation. For the same reason, glycans such as GM1 or GM2 that only possess such a branched Neu5Ac-[α-2,3]-Gal epitope cannot engage MuPyV productively. In support of this, the GM1 ganglioside is not able to rescue MuPyV infection of Gang-/- MEFs (Fig 2, [29]), although low-level and probably non-specific interactions with cells can be detected (S1B Fig). GT1b possesses a disialylated arm at Gal_b and is monosialylated at Gal_a. We predict that GT1b engages VP1 with its monosialylated arm. The second, disialylated arm is likely to be accommodated in such a binding mode, and the [α-2,8]-linked sialic acid might contribute additional contacts. Binding via the monosialylated arm is in line with our findings that supplementation of Gang-/- cells with GT1b rescues infection at a level between GD1a and GT1a. Some gangliosides whose glycan epitopes are capable of engaging VP1 *in vitro* might not be infectious receptors *in vivo*, mainly because of steric complications in the context of the cell membrane. For example, while the crystal structure of PTA with the glycan portion of GD3 shows an identical binding mechanism to GT1a (S7 Fig), supplementation of Gang-/- MEFs with GD3 does not restore infectivity [29]. We reason that the glycan stem of GD3 (and of gangliosides with a similar length such as GM3) is too short to allow efficient attachment of the MuPyV capsid to the cell membrane.

The discrepancy in pathogenicity in MuPyV strains that differ only at one single position is stark. In sharp contrast, the differences among receptor binding between the three strains investigated here are subtle, and a correlation of the structural data with the observed

pathogenicity profiles remains challenging. One reason for this is that avidity effects in the virus capsid, which can engage many ligands simultaneously, multiply subtle changes in the affinity of capsomers for single glycans. It was shown for influenza viruses that small changes between millimolar binding affinities of single binding sites can result in dramatically altered viral binding properties [52]. As discussed above, we found the main difference between RA and PTA/LID to be a differing affinity for GD1a, which appears to bind better to the latter strains due to the larger E91 side chain. This might facilitate attachment and productive infection by these strains to cells that display GD1a, and may thus give them an advantage over RA. While we could not show differences between the PTA and LID strain in terms of glycan affinity to isolated VP1 pentamers, it is unclear how this translates to avidity effects. As such, it is possible that capsid avidities differ enough to explain the more limited spread of PTA. Although direct correlations cannot be made, it becomes increasingly clear that the virus needs to uphold a delicate equilibrium between efficient infection and release from infected and lysed cells as well as selective affinity for productive receptors. The absence of the RA and LID strains outside the laboratory [26] emphasizes that this equilibrium is affected by minute changes in the receptor binding properties.

The MuPyV receptor pocket can clearly accommodate several related but distinct glycan structures (Figs 1 and 4–6). These structures also decorate glycoproteins on many cell surfaces. It therefore seems likely that MuPyV can also engage glycans that are not attached to gangliosides. For instance, the glycan stem of GD1α, which is very similar to DSLNT and prominent on glycoproteins [42–44], is a likely receptor candidate. The different cell-surface distribution patterns of glycoproteins and gangliosides may likewise influence MuPyV spread [8]. Glycoprotein receptors with unknown identity have in fact been shown to promote non-productive internalization of MuPyV, which in turn elicits innate immune responses by the host [29]. Along these lines, our results suggest that virus particles adhere to and enter ganglioside deficient MEFs to levels that are not significantly lower than for wild-type and ganglioside supplemented Gang-/- cells, although without detectable infection. Although not representative for other cell types, these results suggest that the amount of non-productive "pseudoreceptors" on the MEF cell surface is much higher than anticipated.

Our data demonstrate that varying affinities for different gangliosides are the key determinants of a successful MuPyV infection, in line with earlier reports [6–8]. Perhaps unexpectedly, we also find that (even non-specific) attachment of the virus to a host cell can lead to successful internalization, but that this does not necessarily lead to an infection. Thus, we propose that the ratio between productive (ganglioside bound) and non-productive (ganglioside and glycoprotein bound) glycotopes on the host cell itself or in its microenvironment helps to determine the productivity of infection through diverging entry routes, and that differential affinities to these receptors dictate this equilibrium. The nature of these diverging routes, their underlying driving forces, and potential biological consequences other than immune stimulation [29] remain unknown–as does the point at which they diverge. We cannot exclude the possibility that the distribution and binding properties of (pseudo-)receptors are of importance mostly for the post-entry stage rather than for events taking place at the cell surface. A better understanding of the distribution patterns and densities of glycans on specific cells is clearly needed to fully appreciate the many aspects of pathogenesis and tropism of MuPyV as well as many other glycan-binding viruses.

Methods

Ganglioside supplementation and quantification of MuPyV infection

WT and Gang-/- MEFs were seeded onto 96-well Costar 3906 imaging plates in Dulbecco's Modified Eagle's Medium supplemented with 10% fetal bovine serum (FBS). WT (B4+/+St8

+/+) and Gang-/- MEFs (B4-/-St8-/-) were provided by Thomas Benjamin at Harvard Medical School. Gangliosides were purchased from Matreya LLC and resuspended in DMSO upon arrival, aliquoted, and stored at -20°C until use. Cells were incubated overnight in serum free media prior to infection. For ganglioside supplemented Gang-/- MEFs, cells were starved in serum free media containing the indicated concentration of ganglioside. Gangliosides were then removed, and cells were washed with serum free media to remove any free ganglioside. Cells were then infected with NG59RA, PTA, and LID MuPyV (MOI ~10–30). At 24 hours post infection cells were washed in phosphate buffered saline and fixed with 4% paraformaldehyde at room temperature for 10 minutes. Cells were then permeabilized with 0.1% Triton X-100, blocked in 10% FBS in PBS, and then stained for the viral protein, T-antigen (E1). Samples were then incubated with Alexa Fluor labeled secondary antibodies (546). Plates were imaged with the Molecular Devices ImageXpress Micro XL High-Content Screener. The percent infected was calculated for each well (5 images were taken per well). Three wells were quantified per sample and the average percent infected, standard error, and standard deviation were calculated for each sample. To quantify infection, T-antigen staining was measured per each DAPI labeled nucleus. For image analysis, the DAPI channel on each image was thresholded, and nuclei were counted using ImageJ (Analyze Particles). These particles were marked as "Regions of Interest" (ROI), and then the average pixel intensity of T-antigen staining was measured for each nucleus (ROI). These were then binned into T-antigen positive or T-antigen negative nuclei to create % infected.

VP1 immunofluorescence staining

WT and Gang-/- MEFs were seeded onto glass coverslips in Dulbecco's Modified Eagle's Medium supplemented with 10% (FBS). Cells were incubated overnight in serum free media prior to infection. For ganglioside supplementation, Gang-/- MEFs were starved in serum free media containing the indicated concentration of ganglioside. Gangliosides were then removed and cells were washed with serum free media to remove any free ganglioside. Cells were then infected with NG59RA. At indicated times post infection the cells were fixed with 4% paraformaldehyde at room temperature. Cells were blocked in 10% FBS in PBS and then stained for GD1a using mAb MAB5606 (Millipore). Cells were then permeabilized with 0.1% Triton X-100 and stained for the viral proteins, VP1 (I58 antibody) and T-antigen (E1 antibody). Samples were washed and then incubated with Alexa Fluor labeled secondary antibodies (488, 546, 647). Slides were then mounted using DAPI prolong gold mounting media. Slides were imaged with a Nikon A1R confocal microscope. All images were taken as a 9 to 13 step (.25μm) z-stacks on a laser scanning confocal microscope. Each z-stack was aligned and compressed into a max intensity Z projection image.

Virus binding to ganglioside supplemented Gang-/- MEFs

WT and Gang-/- MEFs were seeded onto a 24 well dish in Dulbecco's Modified Eagle's Medium supplemented with 10% (FBS). Cells were incubated overnight in serum free media prior to infection. For ganglioside supplemented Gang-/- MEFs, cells were starved in serum free media containing the indicated concentration of ganglioside. Gangliosides were then removed and cells were washed with serum free media to remove any free ganglioside. Cells were then infected with either NG59RA, PTA, or LID at an MOI ~10–30 (250 μL/well). At 4 hours post infection 150 μL of virus supernatant was removed and placed into a microcentrifuge tube. This virus supernatant was then used to infect WT MEFs seeded onto a 96-well plate (50 μL/well). The amount of free virus was then quantified as percent of infection of the 96-well reinfection plate. At 24 hours post virus addition the plate was washed in PBS and

fixed with 4% PFA at RT for 10 minutes. Cells were then permeabilized with 0.1% Triton X-100, blocked in 10% FBS in PBS, and then stained for the viral protein, T-antigen (E1). Samples were then incubated with Alexa Fluor labeled secondary antibodies (546). Plates were imaged with the Molecular Devices ImageXpress Micro XL High-Content Screener. The percent infected was calculated for each well (5 images were taken per well) as indicated by T-antigen positive nuclei. Three wells were quantified per sample and the average percent infected, standard error, and standard deviation were calculated for each sample. For image analysis, the DAPI channel on each image was thresholded and nuclei were counted using ImageJ (Analyze Particles). These particles were marked as "Regions of Interest" (ROI) and then the average pixel intensity of T-antigen staining was measured for each nucleus (ROI). These were then binned into T-antigen positive or T-antigen negative nuclei to create % infected.

Expression and purification of VP1 pentamers

DNA encoding residues 33–316 of RA (GenBank # M34958.1) or PTA VP1 (GenBank # PSU27812) was cloned into the expression vector pET15b (Novagen) in frame with an N-terminal hexahistidine tag (His-tag) and a thrombin cleavage site. DNA for LID VP1 (GenBank # PSU27813) was generated by site-directed mutagenesis of PTA VP1 residue 296. VP1 pentamers were overexpressed in *E. coli* (BL21) after IPTG induction, and purified by nickel affinity chromatography. The His-tag was removed by thrombin cleavage on column for 72 hours (leaving the non-native residues GSHM at the N-terminus), followed by size exclusion chromatography on a Superdex-200 column.

Crystallization and crystal soaking

Pure VP1 pentamers were supplemented with 20 mM DTT, concentrated to 7.5–8 mg/mL (RA VP1) or 8.5–9 mg/mL (PTA and LID VP1), and crystallized by sitting-drop vapor diffusion. RA VP1 was crystallized at 20°C against reservoir solutions containing a range of 1.25–1.8 M ammonium sulfate and 1–10% (*v/v*) isopropanol. PTA and LID were crystallized at 4°C against reservoir solutions containing 0.1 M HEPES pH 7–8.5 and 1.6–1.8 M K-Na phosphate. For complex formation, the crystals were soaked in the reservoir solution supplemented with the glycan. The detailed crystallization and soaking procedures are listed in S1 Table. The GT1a and GD1a glycans were purchased from Elicityl SA (France), and the DSLNT glycan was purchased from Carbosynth (United Kingdom).

For concentration-dependent soaking VP1 pentamers of all three strains were crystallized at 20°C against a mother liquor containing 1.5 M ammonium sulfate and 6% (*v/v*) isopropanol. These crystals were soaked in drops of mother liquor containing the appropriate concentration of glycan for 30 minutes.

All crystals were cryoprotected by incubation in mother liquor supplemented with the appropriate concentration of glycan and 25% (*v/v*) glycerol. They were then flash-frozen in liquid nitrogen.

Structure determination and electron density quantification

Data reduction was carried out in XDS [55], and the structure of native RA VP1 was solved in Molrep [56] using a model generated from the previously solved structure of P16 VP1 (PDB code 1VPN [12]). Other structures were solved by molecular replacement using the RA VP1 structure in Phenix [57]. All structures were completed by alternating rounds of manual model building in Coot [58], followed by restrained coordinate and isomorphous B-factor refinement including TLS refinement and five-fold non-crystallographic symmetry restraints in Refmac5 [59]. TLS parameters were obtained from the TLSMD server [60]. All models agree well with

the experimental data and have good geometry (Table 2). The PDB accession codes for the structures are listed in Table 2. Structural figures were prepared in PyMOL [61].

Data collected for concentration-dependent soaking experiments was processed as described above. The unit cell parameters for all datasets were treated as equal for all datasets and isomorphous to the dataset "RA Nat" (S2 Table). They were scaled against "RA Nat" in Scaleit [62] and then subjected to B-factor refinement and simulated annealing in Phenix against models of RA, PTA, or LID VP1, which lacked atoms of all solvent molecules in the receptor binding pocket as well as those of tryptophan residues 98 and 227 as controls. The resulting bias-reduced F_{obs}-F_{calc} electron density for Neu5Ac$_a$-[α-2,3]-Gal$_a$ and the two marker tryptophans was calculated as a summation of values of the grid points in a mask generated 1 Å around these groups using the program Mapman [63]. The overall binding of a sugar at different concentrations influences the electron density of the Neu5Ac$_a$-[α-2,3]-Gal$_a$ portion, which is included in GT1a, GD1a, and DSLNT. In contrast, it has no effect on the electron density of the marker tryptophan residues, which do not differ significantly for all data points. For each data point, the average density of the five chains was plotted against ligand concentration and submitted to a non-linear least squares fit using the equation

$$Y = \frac{X}{(K_D + X)} \cdot (B_{max} - B_0) + B_0 \quad (1)$$

where B_{max} was the highest observed electron density value overall (constrained to 95.03 AU) and B_0 the electron density in the binding pocket at 0 mM ligand concentration. Plotting and fitting was done using the program Prism 6 (GraphPad Software, Inc., La Jolla, California, USA).

Supporting Information

S1 Fig. Gangliosides are required for MuPyV infection, but not for cell surface binding. (A) WT, Gang-/- MEFs, and Gang-/- MEFs supplemented with GD1a were infected with NG59RA MuPyV. The MuPyV ganglioside receptor GD1a can be detected on the WT MEFs and GD1a-supplemented Gang-/- MEFs (green), but is absent in Gang-/- MEFs. Virus binds WT, Gang-/-, and GD1a-supplemented Gang-/- MEFs as shown by VP1 staining (red) on the cell surface at 1 hour post infection. At 24 hours post infection, WT and GD1a-supplemented Gang-/- MEFs show robust infection as indicated by nuclear T-antigen staining (magenta). Despite high levels of virus binding, Gang-/- MEFs are completely resistant to infection as shown by lack of T-antigen staining at 24 hours post infection. **(B)** Gang-/- MEFs were supplemented with 2μM GD1a, GT1b, GT1a, GD1b, and GM1 followed by infection with RA, PTA, and LID MuPyV. At 4 hours post infection, virus supernatant was removed and the amount of free virus was quantified for each sample by re-infection of WT MEFs. Virus bound to all cells at similar levels, and there were no significant differences in virus binding to infectious *versus* non-infectious ganglioside receptors. Error bars are standard error, and virus binding to WT MEFs is normalized to one.
(TIF)

S2 Fig. CARP Plots of GT1a bound to the PTA VP1 pentamer. The observed phi and psi torsion angles for the linkages occurring in the PTA-GT1a complex have been plotted and compared to other linkages found in the PDB using CARP with the crystallographic definition of torsion angles. The observed linkages are: Neu5Ac-[α-2,3]-Gal (**A,B**), Gal-[β-1,3]-GalNAc (**C**), GalNAc-[β-1,4]-Gal (**D**), and Neu5Ac-[α-2,8]-Neu5Ac (**E**). The inlay on the lower right shows the schematic and observed structure of GT1a. The linkages are named according to the panels;

the coloring of the glycan rings was adopted from Fig 1.
(TIF)

S3 Fig. Ordered water molecules between the two branches of GT1a. Possible hydrogen bonds between the glycan and ordered water molecules are depicted in grey. In GD1a, the glycerol-like tail of $Neu5Ac_a$ could principally also stabilize the glycan, but preferentially adopts a conformation that does not participate in intramolecular contacts.
(TIF)

S4 Fig. B-factor variance across different strains and glycans. The glycans are colored by B factor on an absolute scale between 0 (dark blue) and 80 (deep red). GT1a/PTA is shown as a grey ghost for comparison. For GD1a and DSLNT, the intramolecular B-factor variance when bound to LID is considerably higher, while values for GT1a are comparable to the other strains.
(TIF)

S5 Fig. Mutation to a glutamine at position 91 of PTA VP1 restores the DSLNT binding mode of RA VP1. Shown are the superimposed DSLNT complex structures of RA (PDB-ID 1VPS [12], transparent red) and PTA E91Q (yellow, r.m.s.d. value for the superposition in PyMOL: 0.159 Å). An F_{obs}-F_{calc} omit map (2σ, carved 1.6 Å around the ligand) is shown for the PTA E91Q complex. On the lower right, DSLNT is represented schematically. As for RA VP1, visible electron density for $Neu5Ac_c$ in PTA E91Q can be seen in one of the five chains.
(TIF)

S6 Fig. The reduced van-der-Waals radius at position 296 in LID allows for a more versatile glycan binding. The van-der-Waals radius of 3.5 Å is indicated as dotted sphere for V296 (PTA, yellow) and A296 (LID, blue). The mutation opens the pocket to one side and allows for a more flexible binding mode of glycans without internal stabilization (DSLNT and 3'-N-Acetyl-sialyllactosamine (3'-SLN), the prototype glycan of the LM1 ganglioside). O4 of Gal_a is pointing directly towards the surface in all complexes. In this binding mode, no branching at this point (as is the case e.g. for GD1b) can be tolerated. Glycans that adopt a binding mode similar to the rigid GT1a are colored in green tones, glycans that exhibit shifts of their moieties are colored in shades of red. GT1a bound to PTA is shown as a grey ghost for comparison.
(TIF)

S7 Fig. GD3 binding to VP1. The complex structures of PTA VP1 with GT1a (dark blue) and GD3 (light pink) are superimposed in PyMOL. On the upper right, the structure of GD3 is represented schematically.
(TIF)

S1 Table. Crystallization and soaking conditions.
(DOCX)

S2 Table. Data set statistics for concentration-dependent soaking experiments.
(DOCX)

Acknowledgments

SDO and RLG would like to thank the BioFrontiers Advanced Light Microscopy Core for microscope facility use and imaging support. We thank Luisa Ströh, Georg Zocher, and Bärbel Blaum for helpful discussions. We are also grateful to the staff at the Swiss Light Source (SLS, Villigen, Switzerland), the European Synchrotron Radiation Facility (ESRF, Grenoble, France),

and the Deutsches Elektronen-Synchrotron (DESY, Hamburg, Germany) for beam time and beam line assistance.

Author Contributions

Conceived and designed the experiments: MHCB AML UN TS SDO RLG. Performed the experiments: MHCB AML SDO. Analyzed the data: MHCB AML UN TS SDO RLG. Wrote the paper: MHCB AML UN TS SDO RLG.

References

1. Ströh LJ, Stehle T (2014) Glycan Engagement by Viruses: Receptor Switches and Specificity. Annu Rev Virol 1: 285–306.
2. Matrosovich M, Herrler G, Klenk HD (2015) Sialic Acid Receptors of Viruses. Top Curr Chem. 367:1–28 doi: 10.1007/128_2013_466 PMID: 23873408
3. Ngamukote S, Yanagisawa M, Ariga T, Ando S, Yu RK (2007) Developmental changes of glycosphingolipids and expression of glycogenes in mouse brains. J Neurochem 103: 2327–2341. PMID: 17883393
4. Yu RK, Macala LJ, Taki T, Weinfield HM, Yu FS (1988) Developmental changes in ganglioside composition and synthesis in embryonic rat brain. J Neurochem 50: 1825–1829. PMID: 3131485
5. Yu RK, Tsai YT, Ariga T, Yanagisawa M (2011) Structures, biosynthesis, and functions of gangliosides—an overview. J Oleo Sci 60: 537–544. PMID: 21937853
6. Cahan LD, Singh R, Paulson JC (1983) Sialyloligosaccharide receptors of binding variants of polyoma virus. Virology 130: 281–289. PMID: 6316632
7. Fried H, Cahan LD, Paulson JC (1981) Polyoma virus recognizes specific sialyligosaccharide receptors on host cells. Virology 109: 188–192. PMID: 6258307
8. Tsai B, Gilbert JM, Stehle T, Lencer W, Benjamin TL, et al. (2003) Gangliosides are receptors for murine polyoma virus and SV40. EMBO J 22: 4346–4355. PMID: 12941687
9. Liddington RC, Yan Y, Moulai J, Sahli R, Benjamin TL, et al. (1991) Structure of simian virus 40 at 3.8-A resolution. Nature 354: 278–284. PMID: 1659663
10. Stehle T, Gamblin SJ, Yan Y, Harrison SC (1996) The structure of simian virus 40 refined at 3.1 A resolution. Structure 4: 165–182. PMID: 8805523
11. Stehle T, Harrison SC (1996) Crystal structures of murine polyomavirus in complex with straight-chain and branched-chain sialyloligosaccharide receptor fragments. Structure 4: 183–194. PMID: 8805524
12. Stehle T, Harrison SC (1997) High-resolution structure of a polyomavirus VP1-oligosaccharide complex: implications for assembly and receptor binding. EMBO J 16: 5139–5148. PMID: 9305654
13. Stehle T, Yan Y, Benjamin TL, Harrison SC (1994) Structure of murine polyomavirus complexed with an oligosaccharide receptor fragment. Nature 369: 160–163. PMID: 8177322
14. Feunteun J, Sompayrac L, Fluck M, Benjamin T (1976) Localization of gene functions in polyoma virus DNA. Proc Natl Acad Sci U S A 73: 4169–4173. PMID: 186787
15. Dawe CJ, Law LW, Dunn TB (1959) Studies of parotid-tumor agent in cultures of leukemic tissues of mice. J Natl Cancer Inst 23: 717–797. PMID: 13814555
16. Gross L (1951) "Spontaneous" leukemia developing in C3H mice following inoculation in infancy, with AK-leukemic extracts, or AK-embrvos. Proc Soc Exp Biol Med 76: 27–32. PMID: 14816382
17. Main JH, Dawe CJ (1966) Tumor induction in transplanted tooth buds infected with polyoma virus. J Natl Cancer Inst 36: 1121–1136. PMID: 4287623
18. Rowe WP, Hartley JW, Estes JD, Huebner RJ (1959) Studies of mouse polyoma virus infection. 1. Procedures for quantitation and detection of virus. J Exp Med 109: 379–391. PMID: 13641563
19. Bolen JB, Fisher SE, Chowdhury K, Shan TC, Williams JE, et al. (1985) A determinant of polyomavirus virulence enhances virus growth in cells of renal origin. J Virol 53: 335–339. PMID: 2981359
20. Bauer PH, Bronson RT, Fung SC, Freund R, Stehle T, et al. (1995) Genetic and structural analysis of a virulence determinant in polyomavirus VP1. J Virol 69: 7925–7931. PMID: 7494305
21. Dawe CJ, Freund R, Mandel G, Ballmer-Hofer K, Talmage DA, et al. (1987) Variations in polyoma virus genotype in relation to tumor induction in mice. Characterization of wild type strains with widely differing tumor profiles. The American journal of pathology 127: 243–261. PMID: 2437801
22. Freund R, Calderone A, Dawe CJ, Benjamin TL (1991) Polyomavirus tumor induction in mice: effects of polymorphisms of VP1 and large T antigen. J Virol 65: 335–341. PMID: 1845894

23. Freund R, Dawe CJ, Benjamin TL (1988) The middle T proteins of high and low tumor strains of polyomavirus function equivalently in tumor induction. Virology 167: 657–659. PMID: 2849243

24. Freund R, Garcea RL, Sahli R, Benjamin TL (1991) A single-amino-acid substitution in polyomavirus VP1 correlates with plaque size and hemagglutination behavior. Journal of virology 65: 350–355. PMID: 1845896

25. Bauer PH, Cui C, Liu WR, Stehle T, Harrison SC, et al. (1999) Discrimination between sialic acid-containing receptors and pseudoreceptors regulates polyomavirus spread in the mouse. J Virol 73: 5826–5832. PMID: 10364334

26. Carroll J, Dey D, Kreisman L, Velupillai P, Dahl J, et al. (2007) Receptor-binding and oncogenic properties of polyoma viruses isolated from feral mice. PLoS Path 3: e179.

27. Gilbert J, Benjamin T (2004) Uptake pathway of polyomavirus via ganglioside GD1a. J Virol 78: 12259–12267. PMID: 15507613

28. O'Hara SD, Stehle T, Garcea R (2014) Glycan receptors of the Polyomaviridae: structure, function, and pathogenesis. Curr Opin Virol 7: 73–78. doi: 10.1016/j.coviro.2014.05.004 PMID: 24983512

29. You J, O'Hara SD, Velupillai P, Castle S, Levery S, et al. (2015) Ganglioside and non-Ganglioside Receptors for the Mouse Polyomavirus. PLoS Path 11: e1005175.

30. Low JA, Magnuson B, Tsai B, Imperiale MJ (2006) Identification of gangliosides GD1b and GT1b as receptors for BK virus. J Virol 80: 1361–1366. PMID: 16415013

31. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol 372: 774–797. PMID: 17681537

32. Blaum BS, Hannan JP, Herbert AP, Kavanagh D, Uhrin D, et al. (2015) Structural basis for sialic acid-mediated self-recognition by complement factor H. Nat Chem Biol 11: 77–82. doi: 10.1038/nchembio.1696 PMID: 25402769

33. Chen Y, Tan M, Xia M, Hao N, Zhang XC, et al. (2011) Crystallography of a Lewis-binding norovirus, elucidation of strain-specificity to the polymorphic human histo-blood group antigens. PLoS Path 7: e1002152.

34. Neu U, Hengel H, Blaum BS, Schowalter RM, Macejak D, et al. (2012) Structures of Merkel cell polyomavirus VP1 complexes define a sialic acid binding site required for infection. PLoS Path 8: e1002738.

35. Merritt EA, Kuhn P, Sarfaty S, Erbe JL, Holmes RK, et al. (1998) The 1.25 A resolution refinement of the cholera toxin B-pentamer: evidence of peptide backbone strain at the receptor-binding site. J Mol Biol 282: 1043–1059. PMID: 9753553

36. Neu U, Woellner K, Gauglitz G, Stehle T (2008) Structural basis of GM1 ganglioside recognition by simian virus 40. Proc Natl Acad Sci U S A 105: 5219–5224. doi: 10.1073/pnas.0710301105 PMID: 18353982

37. Cubrilovic D, Haap W, Barylyuk K, Ruf A, Badertscher M, et al. (2014) Determination of protein-ligand binding constants of a cooperatively regulated tetrameric enzyme using electrospray mass spectrometry. ACS Chem Biol 9: 218–226. doi: 10.1021/cb4007002 PMID: 24128068

38. Fotinou C, Emsley P, Black I, Ando H, Ishida H, et al. (2001) The crystal structure of tetanus toxin Hc fragment complexed with a synthetic GT1b analogue suggests cross-linking between ganglioside receptors and the toxin. J Biol Chem 276: 32274–32281. PMID: 11418600

39. Sharmila DJ, Veluraja K (2006) Conformations of higher gangliosides and their binding with cholera toxin—investigation by molecular modeling, molecular mechanics, and molecular dynamics. J Biomol Struct Dyn 23: 641–656. PMID: 16615810

40. Zhang Y, Yamamoto S, Yamaguchi T, Kato K (2012) Application of paramagnetic NMR-validated molecular dynamics simulation to the analysis of a conformational ensemble of a branched oligosaccharide. Molecules 17: 6658–6671. doi: 10.3390/molecules17066658 PMID: 22728360

41. Kuan CT, Chang J, Mansson JE, Li J, Pegram C, et al. (2010) Multiple phenotypic changes in mice after knockout of the B3gnt5 gene, encoding Lc3 synthase—a key enzyme in lacto-neolacto ganglioside synthesis. BMC Dev Biol 10: 114. doi: 10.1186/1471-213X-10-114 PMID: 21087515

42. Parkkinen J, Rogers GN, Korhonen T, Dahr W, Finne J (1986) Identification of the O-linked sialyloligosaccharides of glycophorin A as the erythrocyte receptors for S-fimbriated Escherichia coli. Infect Immun 54: 37–42. PMID: 2875951

43. Schauer R (2000) Achievements and challenges of sialic acid research. Glycoconj J 17: 485–499. PMID: 11421344

44. Thomsson KA, Holmen-Larsson JM, Angstrom J, Johansson ME, Xia L, et al. (2012) Detailed O-glycomics of the Muc2 mucin from colon of wild-type, core 1- and core 3-transferase-deficient mice highlights differences compared with human MUC2. Glycobiology 22: 1128–1139. doi: 10.1093/glycob/cws083 PMID: 22581805

45. Burmeister WP, Guilligay D, Cusack S, Wadell G, Arnberg N (2004) Crystal structure of species D adenovirus fiber knobs and their sialic acid binding sites. J Virol 78: 7727–7736. PMID: 15220447

46. Parrish CR (1990) Emergence, natural history, and variation of canine, mink, and feline parvoviruses. Adv Virus Res 38: 403–450. PMID: 2171302

47. Parrish CR (1991) Mapping specific functions in the capsid structure of canine parvovirus and feline panleukopenia virus using infectious plasmid clones. Virology 183: 195–205. PMID: 1647068

48. Neu U, Allen SA, Blaum BS, Liu Y, Frank M, et al. (2013) A structure-guided mutation in the major capsid protein retargets BK polyomavirus. PLoS Path 9: e1003688.

49. Neu U, Khan ZM, Schuch B, Palma AS, Liu Y, et al. (2013) Structures of B-lymphotropic polyomavirus VP1 in complex with oligosaccharide ligands. PLoS Path 9: e1003714.

50. Kanda T, Furuno A, Yoshiike K (1986) Mutation in the VP-1 gene is responsible for the extended host range of a monkey B-lymphotropic papovavirus mutant capable of growing in T-lymphoblastoid cells. J Virol 59: 531–534. PMID: 3488416

51. Xiong X, Martin SR, Haire LF, Wharton SA, Daniels RS, et al. (2013) Receptor binding by an H7N9 influenza virus from humans. Nature 499: 496–499. doi: 10.1038/nature12372 PMID: 23787694

52. Xiong X, Coombs PJ, Martin SR, Liu J, Xiao H, et al. (2013) Receptor binding by a ferret-transmissible H5 avian influenza virus. Nature 497: 392–396. doi: 10.1038/nature12144 PMID: 23615615

53. Ogawa-Goto K, Funamoto N, Ohta Y, Abe T, Nagashima K (1992) Myelin gangliosides of human peripheral nervous system: an enrichment of GM1 in the motor nerve myelin isolated from cauda equina. J Neurochem 59: 1844–1849. PMID: 1402926

54. Ishikawa Y, Gasa S, Minami R, Makita A (1987) Characterization of neutral glycosphingolipids from fetal human brain: evidence for stage-specific expression of the globo, ganglio, and neolacto series in the central nervous system. J Biochem 101: 1369–1375. PMID: 3667553

55. Kabsch W (2010) Integration, scaling, space-group assignment and post-refinement. Acta Crystallogr D Biol Crystallogr 66: 133–144. doi: 10.1107/S0907444909047374 PMID: 20124693

56. Vagin A, Teplyakov A (2010) Molecular replacement with MOLREP. Acta Crystallogr D Biol Crystallogr 66: 22–25. doi: 10.1107/S0907444909042589 PMID: 20057045

57. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66: 213–221. doi: 10.1107/S0907444909052925 PMID: 20124702

58. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60: 2126–2132. PMID: 15572765

59. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 53: 240–255. PMID: 15299926

60. Painter J, Merritt EA (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr D Biol Crystallogr 62: 439–450. PMID: 16552146

61. The PyMOL Molecular Graphics System, Version 1.5.0.4, Schrödinger, LLC.

62. Howell PL, Smith GD (1992) Identification of heavy-atom derivatives by normal probability methods. J Appl Cryst 25: 81–86.

63. Kleywegt GJ, Jones TA (1996) xdlMAPMAN and xdlDATAMAN—Programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets. Acta Crystallogr D Biol Crystallogr 52: 826–828. PMID: 15299647

64. Kato Y, Kuan CT, Chang J, Kaneko MK, Ayriss J, et al. (2010) GMab-1, a high-affinity anti-3'-isoLM1/3',6'-isoLD1 IgG monoclonal antibody, raised in lacto-series ganglioside-defective knockout mice. Biochem Biophys Res Commun 391: 750–755. doi: 10.1016/j.bbrc.2009.11.132 PMID: 19944071

65. Svennerholm L (1969). Comprehensive Biochemistry.

66. Taube S, Jiang M, Wobus CE (2010) Glycosphingolipids as receptors for non-enveloped viruses. Viruses 2: 1011–1049. doi: 10.3390/v2041011 PMID: 21994669

67. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28: 27–30. PMID: 10592173

PLOS | PATHOGENS

RESEARCH ARTICLE

Human Adenovirus 52 Uses Sialic Acid-containing Glycoproteins and the Coxsackie and Adenovirus Receptor for Binding to Target Cells

Annasara Lenman[1]*, A. Manuel Liaci[2], Yan Liu[3], Carin Årdahl[1], Anandi Rajan[1], Emma Nilsson[1], Will Bradford[2], Lisa Kaeshammer[2], Morris S. Jones[4], Lars Frängsmyr[1], Ten Feizi[3], Thilo Stehle[2,5]*, Niklas Arnberg[1]

1 Division of Virology, Department of Clinical Microbiology, Umeå University, Umeå, Sweden, **2** University of Tübingen, Interfaculty Institute of Biochemistry, Tübingen, Germany, **3** Glycosciences Laboratory, Department of Medicine, Imperial College London, London, United Kingdom, **4** Division of Infectious Diseases, Naval Medical Center, San Diego, California, United States of America, **5** Vanderbilt University School of Medicine, Nashville, Tennessee, United States of America

* thilo.stehle@uni-tuebingen.de (TS, for structural biology); annasara.lenman@climi.umu.se (AL, for virology)

OPEN ACCESS

Citation: Lenman A, Liaci AM, Liu Y, Årdahl C, Rajan A, Nilsson E, et al. (2015) Human Adenovirus 52 Uses Sialic Acid-containing Glycoproteins and the Coxsackie and Adenovirus Receptor for Binding to Target Cells. PLoS Pathog 10(2): e1004657. doi:10.1371/journal.ppat.1004657

Editor: Patrick Hearing, Stony Brook University, UNITED STATES

Received: August 14, 2014

Accepted: January 5, 2015

Published: February 12, 2015

Data Availability Statement: All relevant data are within the paper and its Supporting Information files except for the gene sequence of native HAdV-52 fiber-1 which is available from Genbank under the accession number DQ923122.2, the Protein sequence database entry for native fiber-1 (Accession code ABK35058.1), as well as the PDB entry of the construct used in this publication (PDB ID 4XL8).

Funding: This work was founded by grants from the Swedish Research Council (2013-2753 and 2013-8616, NA), Knut & Alice Wallenberg Foundation

Abstract

Most adenoviruses attach to host cells by means of the protruding fiber protein that binds to host cells via the coxsackievirus and adenovirus receptor (CAR) protein. Human adenovirus type 52 (HAdV-52) is one of only three gastroenteritis-causing HAdVs that are equipped with two different fiber proteins, one long and one short. Here we show, by means of virion-cell binding and infection experiments, that HAdV-52 can also attach to host cells via CAR, but most of the binding depends on sialylated glycoproteins. Glycan microarray, flow cytometry, surface plasmon resonance and ELISA analyses reveal that the terminal knob domain of the long fiber (52LFK) binds to CAR, and the knob domain of the short fiber (52SFK) binds to sialylated glycoproteins. X-ray crystallographic analysis of 52SFK in complex with 2-*O*-methylated sialic acid combined with functional studies of knob mutants revealed a new sialic acid binding site compared to other, known adenovirus:glycan interactions. Our findings shed light on adenovirus biology and may help to improve targeting of adenovirus-based vectors for gene therapy.

Author Summary

HAdVs are common pathogens in humans, causing disease mainly in eyes, airways and gastrointestinal tract. Most HAdVs are equipped with twelve protruding fiber proteins that mediate attachment to host cell receptor molecules. Recently, a new human gastroenteritis-associated adenovirus (HAdV-52) was identified and classified as the first member of a novel species (HAdV-G). Unlike most other HAdVs, this virus contains two different fiber proteins, a long and a short one, a feature shared only with the two members of

(KAW 2013.0019, NA), The Swedish Cancer Society (CAN 2011/340, NA), United Kingdom Research Council Basic Technology Initiative Glycoarrays (GRS/79268, TF), Translational Grant (EP/G037604/1, TF), Wellcome Trust Grants (WT093378MA and WT099197MA, TF), and from The Deutsche Forschungsgemeinschaft (SFB-685, ST and AML). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

species HAdV-F (HAdV-40 and -41). To gain further insights into the mechanisms of HAdV-52 infection of human cells, we set out to identify the host cell receptors used by the long and short fibers. We find that the long fiber binds to a protein-based receptor known as the coxsackievirus and adenovirus receptor (CAR), and that the short fiber binds to glycoproteins that contain sialic acid-capped glycans. The crystal structure determination of a complex of the short fiber knob bound to sialic acid demonstrates that this interaction is unique among HAdVs, and bioinformatic analysis indicates that simian AdVs may also engage sialic acids in the manner seen in HAdV-52. The results presented here provide insights into the plasticity of adenovirus-host cell interactions.

Introduction

Human adenoviruses (HAdVs) are classified into seven species (A-G), with more than 50 different types known to date [1]. Most HAdVs cause disease in the eyes (members of species HAdV-B, -D, -E), airways (HAdV-A, -B, -C, -E) and gastrointestinal tract (HAdV-F mainly) [2]. HAdV-52 was recently identified as a novel, human pathogen associated with gastroenteritis [3], and was found to be divergent from other HAdVs placing this virus in a new species (HAdV-G). HAdVs from species HAdV-A and HAdV-C through HAdV-F use the coxsackievirus and adenovirus receptor (CAR) as a primary adhesion receptor [4–6]. Members of species HAdV-B that cause ocular, respiratory and/or urinary tract infections utilize CD46 and/or desmoglein-2 as cellular receptors [7–10]. Specific members of species HAdV-D cause a more severe ocular infection, epidemic keratoconjunctivitis, and engage glycoproteins that carry glycans mimicking those in the GD1a ganglioside: Neu5Acα(2–3)Galβ(1–3)GalNAcβ(1–4)(Neu5Acα(2–3))Galβ(1–4)Glc as receptors [11–13]. In addition to these AdVs, canine AdV-2 (species CAdV-A) is another glycan-binding AdV, engaging Neu5Acα(2–3)[6S]Galβ(1–4)GlcNAc-containing glycans at a different location on the knob[14]. The locations of the two known glycan binding sites are distinct from the regions that allow some knobs to engage CAR [15] or CD46 [16]. In the case of CAR, the length and flexibility of the fiber shaft also seem to play a role in infection, as short and sturdy fibers cannot bend to bind CAR on a cell surface [17]. HAdVs can also enter cells through interactions with coagulation factors that mediate indirect binding to heparan sulfate proteoglycans on target cells [18,19]. With a few exceptions, HAdVs are equipped with a single type of capsid fiber protein that interacts with receptors via its knob domain. HAdV-40, -41, and -52 on the other hand are equipped with two different fibers, one long and one short [20,21]. The long fibers of HAdV-40 and -41 bind to CAR [6], but no function has been described for any of the short fibers. Phylogenetic analyses have shown that the closest human relatives to the knobs of the HAdV-52 long and short fibers are the knobs of the long and short fibers of species HAdV-F (HAdV-40 and -41) [22].

Adenoviruses are frequently used as vectors for diverse applications including vaccination [23–25], treatment of cancer [26] and hereditary disorders [27], cardiovascular applications [28], and stem cell research [29]. Three of the main challenges for the most commonly used HAdV-5 (species HAdV-C) based vectors are: i) pre-existing, neutralizing antibodies [30], ii) poor access to CAR [31], and, iii) coagulation factor-dependent off-target transduction of the liver [18]. Potential solutions to these obstacles have been to use vector candidates based on less common HAdV types [24], and HAdV types that use receptors alternative to CAR [32,33]. Inefficient targeting have been addressed by ablating CAR- and/or coagulation factor-interactions, and/or by retargeting to receptors that are overexpressed on target cells [34]. Ideally, a

multi-purpose vector would therefore be based on a rare type that efficiently target host cells by means of specific receptor interactions and low or absent off-target transduction.

The seroprevalence for HAdV-40 and -41 is relatively high (40–50%) in the human population [35–37]. The seroprevalence for HAdV-52 in humans has not been investigated, but the close relationship with simian AdVs and the low frequency of detection in humans [3,38] suggest that the seroprevalence in humans is low. In combination with its uncommon capsid organization this prompted us to gain more insight into HAdV-52 interactions with host cells and more specifically to identify cellular receptors used by HAdV-52 for attachment to host cells.

Results and Discussion

HAdV-52 binds both to CAR and sialic acid-containing glycans on target cells

To investigate whether CAR, CD46 or sialic acid-containing glycans can function as receptors for HAdV-52, we first analyzed ^{35}S-labelled HAdV-52 virion binding to CHO cells expressing or lacking these receptors. HAdV-52 bound with similar efficiency to sialic acid-expressing control CHO (Pro-5) cells, CD46-expressing CHO cells and CHO MOCK (with respect to CAR), but with increased efficiency to CAR-expressing CHO cells and with decreased efficiency to sialic acid-lacking Lec2 cells (derived from Pro-5) as compared to the other cells (**Fig. 1A**). Pretreatment of cells with sialic acid-cleaving *V. cholerae* neuraminidase reduced HAdV-52 binding to background levels for all cells except to CHO-CAR. To test if this neuraminidase removed sialic acids with equal efficiency from all cells, we treated the cells with *V. cholerae* neuraminidase and quantified MAL-II lectin binding. This treatment reduced MAL-II binding to background levels (**S1 Fig.**) and we therefore concluded that HAdV-52 could bind to CHO-CAR independently of sialic acid. As HAdV-52 bound with equal efficiency to Pro-5, CHO-MOCK, and CHO-CD46, and as neuraminidase treatment of CHO-CD46 cells reduced HAdV-52 binding efficiently, these results indicate that CD46 is probably of no or low importance as a receptor for HAdV-52. HAdV-52 also infected Pro-5 cells more efficiently than Lec2 cells, and pretreatment of Pro-5 cells with neuraminidase abolished infection (**Fig. 1B**). HAdV-52 is associated with gastroenteritis, but the number of human cases described is limited and the cellular tropism of the virus is unclear. We therefore investigated the relative contributions of sialic acid and CAR using respiratory A549 cells, which support productive infection of most HAdVs and express both sialic acid and CAR at the cell surface. HAdV-52 binding to these cells was reduced by 20% and 25%, respectively, when preincubating HAdV-52 virions with soluble CAR-D1 (consisting of the N-terminal, most membrane-distal immunoglobulin-like domain), or when preincubating cells with monoclonal anti-CAR antibodies (clone RmcB) prior to virion binding (**Fig. 1C**). CAR-D1 and anti-CAR antibodies reduced HAdV-5 binding with 50% and 75%, respectively (**S2 Fig.**), thus demonstrating their function. On the other hand, HAdV-52 binding was reduced by 75% and 80% after preincubating virions with sialic acid or when pretreating cells with neuraminidase, respectively, prior to virion binding. Pretreatments with CAR-D1 or anti-CAR antibodies in combination with either sialic acid or neuraminidase reduced binding to background levels. The involvement of sialic acid-containing glycans as functional human cell receptors for HAdV-52 was confirmed by neuraminidase pretreatment of A549 cells, which reduced HAdV-52 infection by at least 80% (**Fig. 1D**). Finally, preincubation of virions with coagulation factor IX and X efficiently enhanced HAdV-5 binding to and infection of A549 cells but had no or limited effect on HAdV-52 (**Fig. 2A,B**). These results show that HAdV-52 does not use FIX, FX, or CD46 for attachment to A549 cells. We conclude that HAdV-52 binds to A549 cells mainly via sialic acid-containing glycans, and that the role of CAR is dwarfed by that of the sialylated receptors. However, we cannot exclude

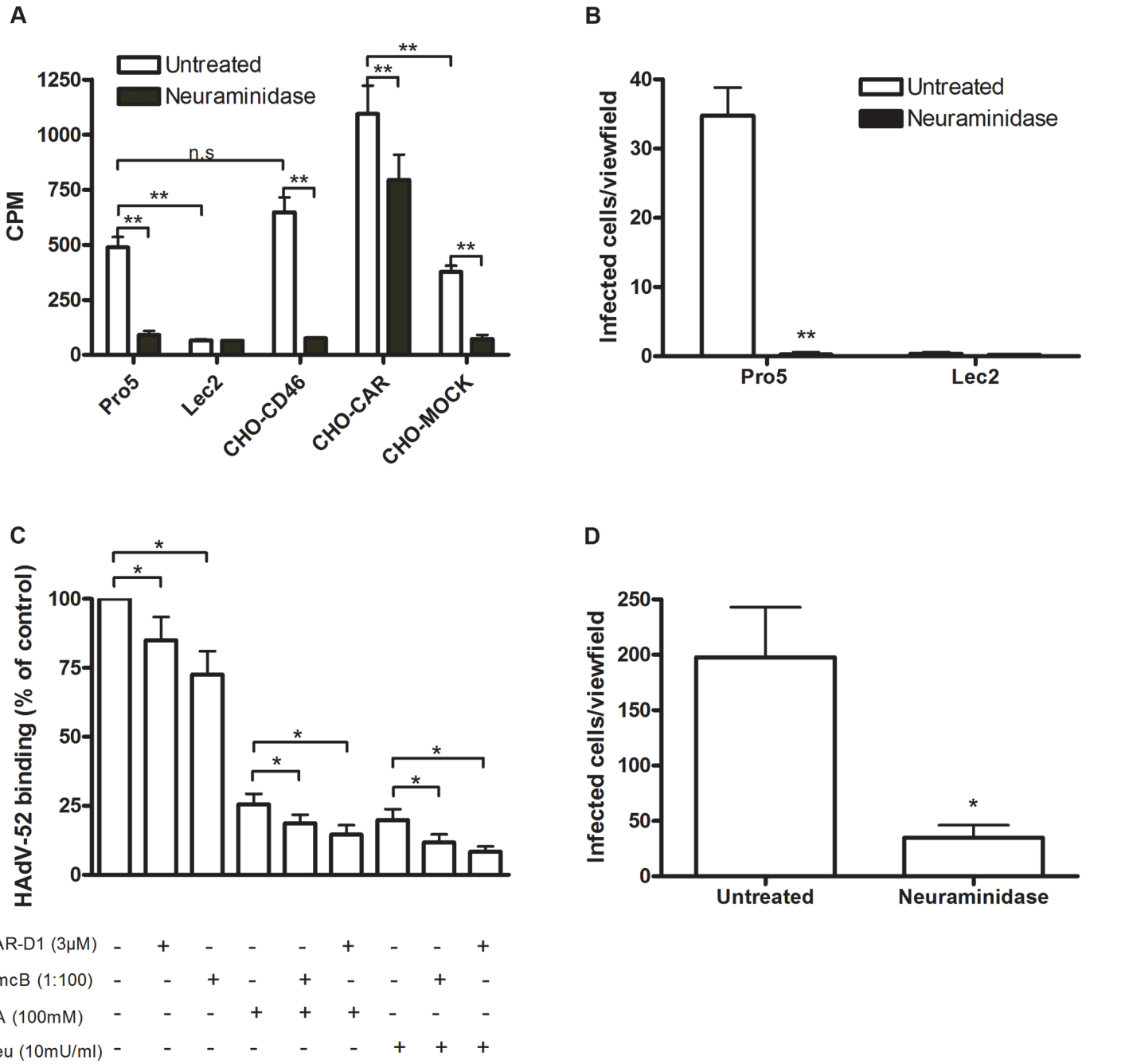


Fig 1. HAdV-52 uses sialic acid and CAR for binding to and infection of cells. (A) ^{35}S-labeled HAdV-52 virion binding to CHO cells expressing or lacking known HAdV receptors. Pro-5 is a sialic acid-positive, reference cell line and parental cell line to sialic acid-negative Lec2 cells. CHO-CD46 and CHO-CAR cells express human CD46 and CAR, respectively. CHO-MOCK is mock transfected with respect to CHO-CAR. Black bars show HAdV-52 binding to cells after pretreatment with *V. cholerae* neuraminidase. Binding was quantified by liquid scintillation counting and shown as counts per minute (CPM). (B) HAdV-52 infection of Pro-5 and Lec2 pretreated with (black bars) or without (white bars) *V. cholerae* neuraminidase. The number of infected cells was quantified by immunofluorescence. (C) ^{35}S-labeled HAdV-52 virion binding to A549 cells. Virions were preincubated with or without soluble CAR-D1 or sialic acid monosaccharides, and cells were preincubated with or without mouse anti-CAR mab (clone RmcB) or *V. cholerae* neuraminidase as indicated, prior to binding. (D) HAdV-52 infection of A549 cells pretreated with or without *V. cholerae* neuraminidase. All experiments were performed three times with duplicate samples in each experiment. Error bars represent mean ± SD. n.s = not significant, * P of < 0.05 and ** P of < 0.01.

doi:10.1371/journal.ppat.1004657.g001

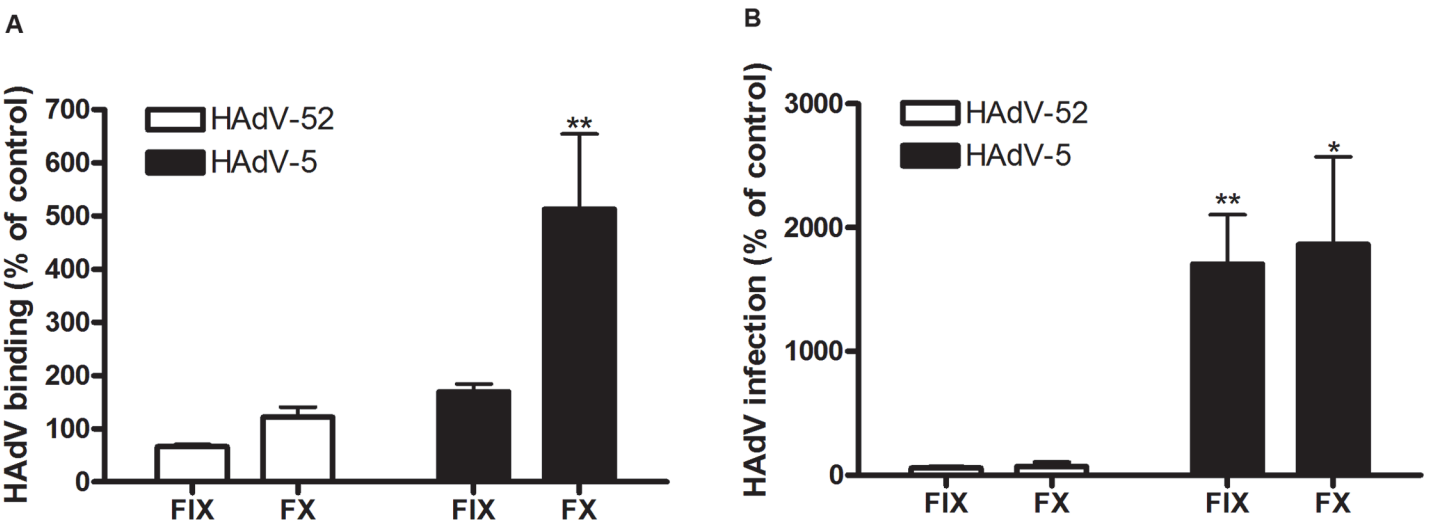


Fig 2. HAdV-52 does not use coagulation factors for binding and infection of A549 cells. (A) ^{35}S-labeled HAdV-52 virion binding to A549 cells after virion preincubation with physiological concentrations of coagulation factor IX and X (FIX: 5μg/ml and FX: 10μg/ml). (B) HAdV-52 infection of A549 cells after virion preincubation with physiological concentrations of coagulation factors. All experiments were performed three times with duplicate samples in each experiment. Error bars represent mean ± SD. * P of < 0.05 and ** P of < 0.01 versus control.

doi:10.1371/journal.ppat.1004657.g002

that the role of CAR as an attachment receptor for HAdV-52 may be more pronounced on other cell types than on A549 cells.

The short fiber of HAdV-52 binds to sialic acid and the long fiber binds to CAR

To characterize the nature of the sialic acid-containing glycans as receptors and the mechanism of interaction, we next quantified binding of HAdV-52 virions and HAdV-52 long and short fiber knobs (52LFK and 52SFK) to A549 cells pretreated with enzymes, lectins or metabolic inhibitors that alter the expression levels of cell surface molecules. Whereas inhibitors of glycolipid biosynthesis (P4) and *N*- (via Asp) linked glycosylation (tunicamycin) did not reduce virion binding to A549 cells significantly (**Figs.** 3A, S3A,B), benzyl *N*-acetyl-α-D-galactosaminide (benzyl-α-GalNAc, an inhibitor of *O*-linked glycosylation, via Ser or Thr) reduced binding of both the virions and 52SFK, but not of 52LFK (Fig. 3A,B). Protease (ficin, proteinase K, and bromelain) treatments of the same cells reduced binding of both 52SFK and 52LFK by 55–85% (Fig. 3C,D). These results suggest that on A549 cells, in contrast with the 52LFK which engages proteins directly without involvement of glycans, mucin type *O*-linked glycans are the dominant receptors for 52SFK and glycolipids and *N*-linked glycans appear not to play a major role.

To determine the relative contribution of each fiber to cell attachment and infectivity, we first performed western blot analysis to characterize the relative fiber content in virus particles. Unlike HAdV-41 virions, which contain short and long fibers in a 6:1 ratio [39], HAdV-52 virions contained equal amounts of long and short fibers according to western blot analysis using a monoclonal antibody, which recognizes an epitope that is conserved in all HAdVs (Fig. 4A). This suggests that the apparent key role of sialic acid cannot be accounted for by the short fiber being more abundant in the HAdV-52 virion. We also found by flow cytometry analysis that A549 cells expressed higher levels of CAR compared with another epithelial cell line (human corneal epithelial cells; HCE) (Fig. 4B), suggesting that the modest function of CAR during HAdV-52 binding to A549 cells was not due to low expression levels on these cells. Homology alignment of the long and short fiber knob sequences with corresponding sequences of sialic

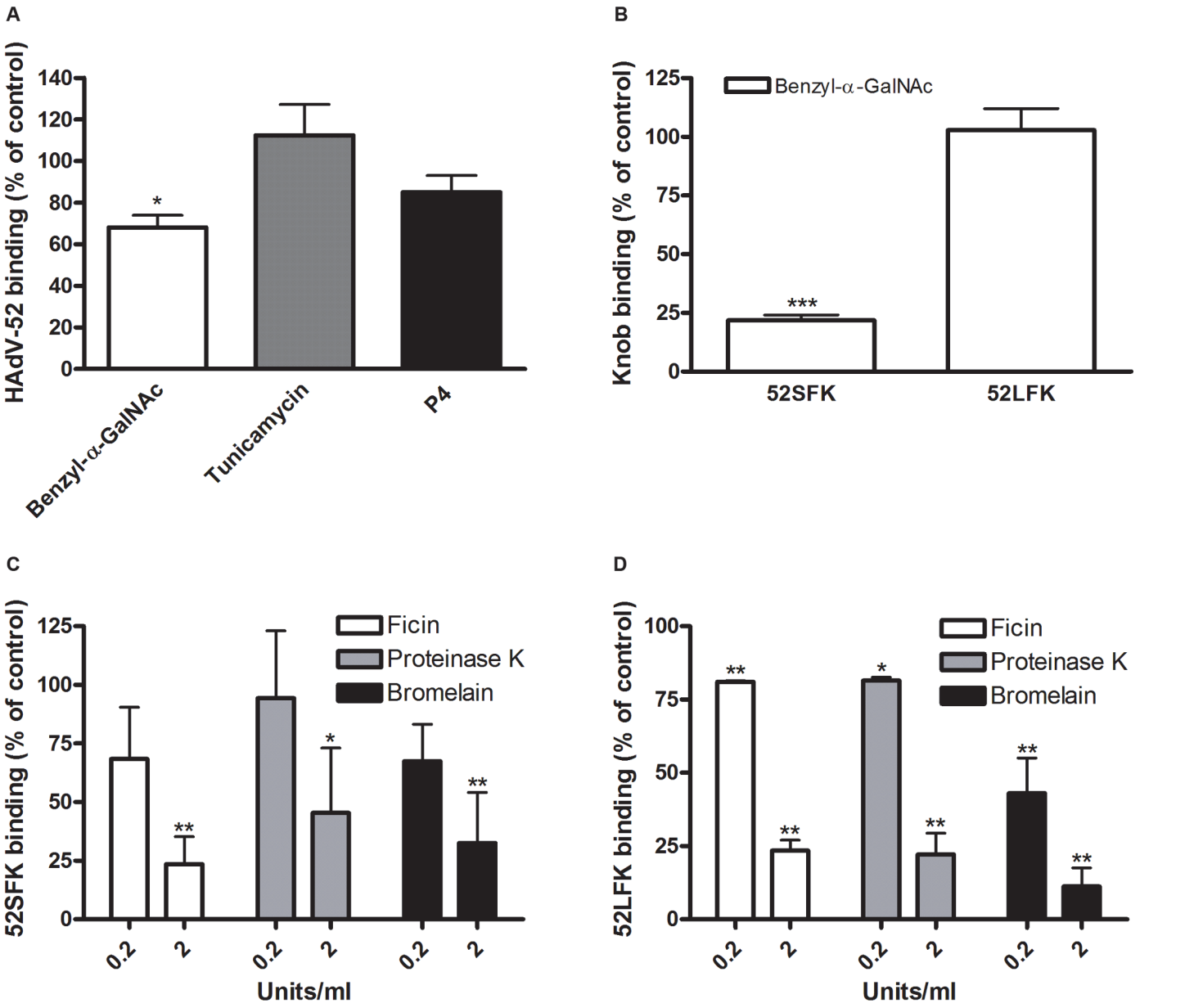


Fig 3. HAdV-52 short fiber knob binds to *O*-linked glycoproteins on A549 cells. (A) ^{35}S-labeled HAdV-52 virion binding to A549 cells pretreated with benzyl-α-GalNAc, tunicamycin or P4 (inhibitors of *O*-linked glycan synthesis, *N*-linked glycan synthesis, and glycolipid synthesis, respectively). (B) HAdV-52 short fiber knobs (52SFK) and HAdV-52 long fiber knobs (52LFK) binding to A549 cells pretreated with benzyl-α-GalNAc (inhibitor of *O*-linked glycan synthesis). (C) 52SFK and (D) 52LFK binding to A549 cells pretreated with ficin, proteinase K or bromelain proteases at indicated concentrations. All experiments were performed three times with duplicate samples in each experiment. Error bars represent mean ± SD. * P of < 0.05, ** P of < 0.01 and *** P of < 0.001 versus control.

doi:10.1371/journal.ppat.1004657.g003

acid-interacting HAdV-37 (ocular tropism) and CAR-interacting HAdV-5 (respiratory tropism) and HAdV-12 (respiratory and intestinal tropism) revealed that, while the majority of the known CAR-interacting residues [15,40] are conserved in 52LFK, only a few of these residues are conserved in 52SFK (**S4 Fig.**). Furthermore, when examining the potential for interactions with sialic acid based on the structure of the HAdV-37 knob bound to sialic acid [41], only two out of the seven sialic acid-contacting residues are conserved in 52SFK and none of

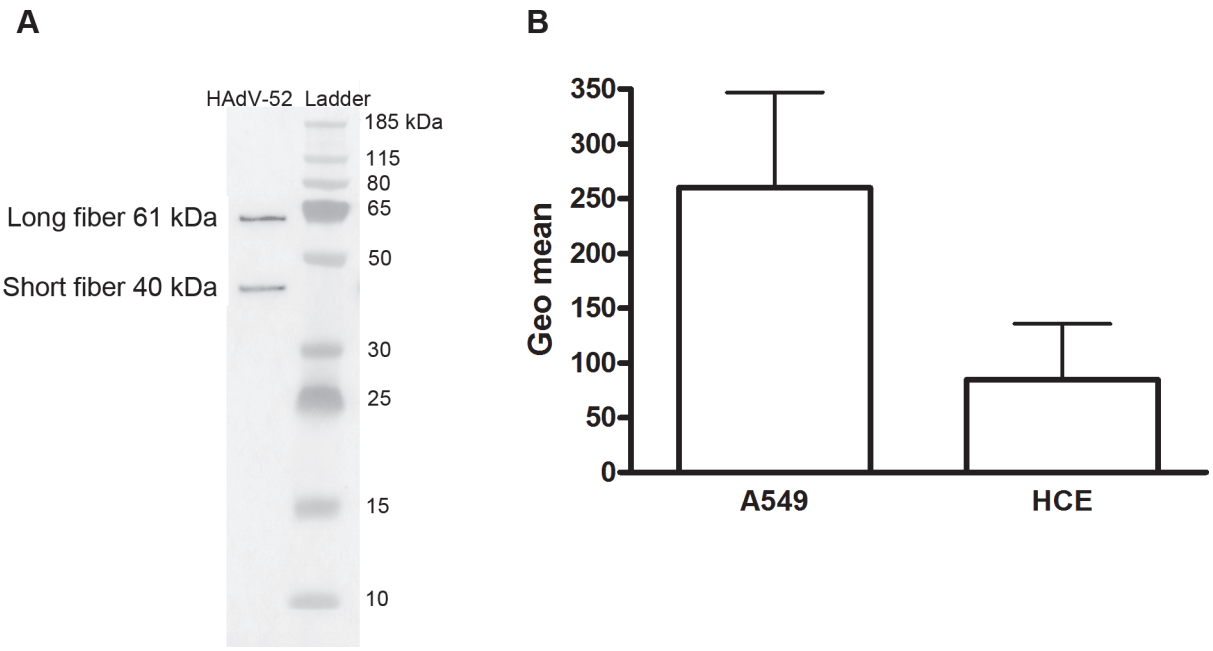


Fig 4. Virion composition and relative expression of CAR and sialic acid on human epithelial cells. A) Western blot analysis of HAdV-52 virion fiber content using a mouse mab (clone 4D2) recognizing an epitope (MKRARPSEDTFNPVYPY) conserved in the tail domain of all HAdVs. The experiment was performed three times (with three different virus preparations) and the figure shows one representative set of results. B) Flow cytometry analysis of CAR expression on A549 and human corneal epithelial (HCE) cells using an anti CAR mouse mab (clone E1-1). Data are shown as geometrical mean (geo mean) and the experiment was performed three times with duplicate samples in each experiment. Error bars represent means ± SD.

doi:10.1371/journal.ppat.1004657.g004

these are conserved in 52LFK. Flow cytometry analysis confirmed that 52LFK can only bind to CAR-expressing cells (Fig. 5A). 52SFK bound with similar efficiency to all cells (including CAR-expressing cells) but not to sialic acid-deficient Lec2 cells (Fig. 5B). Neuraminidase treatment of A549 cells reduced binding of 52SFK to A549 cells but not of 52LFK (Fig. 5C), confirming that 52SFK binds to sialic acid-containing receptors on human target cells. ELISA experiments showed that 52SFK (in solution) bound efficiently to sialylated fetuin glycoprotein (immobilized) but not to two desialylated variants of fetuin (Fig. 5D), this was also confirmed with surface plasmon resonance (SPR) where fetuin bound to immobilized 52SFK with an affinity of 37 μM, while for the desialylated fetuin type II a K_D could not be determined (S5 Fig.). 52LFK did not bind to any of these proteins. SPR analysis demonstrated that the 52LFK:CAR-D1D2 (full length extracellular domain) and 52LFK:CAR-D1 interactions were of high affinity (5 and 2.6 nM, respectively; **Figs. 5E and S6**), which is in the same range as of other CAR:HAdV-knob interactions [42]. According to SPR analysis, 52SFK did not interact with CAR at all (S7 Fig.). As the sialic acid-containing glycan(s) used by 52SFK for binding to A549 cells are not known, we can only speculate that such monovalent interactions would probably be of low affinities, as most other protein:glycan interactions, and thereby lower than the affinity of the LFK:CAR interaction. We conclude from these results that the HAdV-52 long fiber binds to CAR and that the short fiber binds to sialic acid-containing glycans. It has been shown that cells infected with HAdV-2 (species HAdV-C) secrete an excess of fibers that unlocks junctional, intercellular CAR-CAR homodimers, resulting in increased extracellular space and improved intercellular transport of subsequently released virions [43], and similar effects have been shown for HAdV-3 (species B) penton dodecahedra [44]. It is therefore tempting to speculate that a possible function of the HAdV-52 short fiber is to mediate virion attachment to non-infected cells whereas excess of long fibers are secreted from infected cells and facilitate

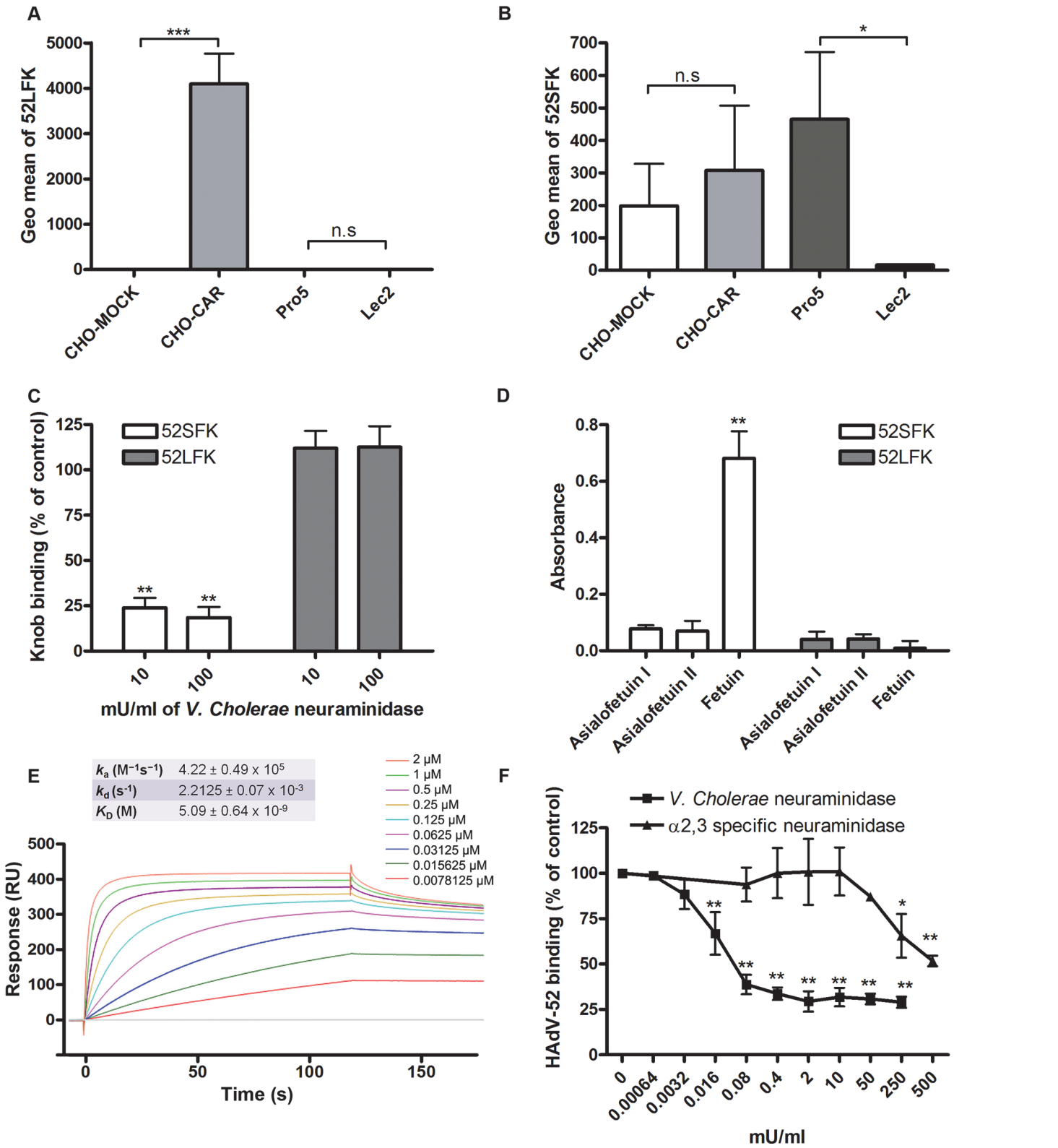


Fig 5. HAdV-52 short fiber knob binds to sialic acid and long fiber knob binds to CAR. (A) HAdV-52 long fiber knob (52LFK) and (B) HAdV-52 short fiber knob (52SFK) binding to CHO-cells lacking human CAR (all cells except CHO-CAR), lacking sialic acid (only Lec2) and expressing human CAR (CHO-CAR). (C) 52SFK and 52LFK binding to A549 cells pretreated with *V. cholerae* neuraminidase. (D) ELISA analysis of 52SFK and 52LFK (in solution) binding to immobilized, sialylated fetuin and asialofetuin type I (chemically prepared) and II (enzymatically prepared). Relative absorbance is shown. (E) Surface

plasmon resonance analysis of 52LFK (in solution) binding to CAR (immobilized). A twofold dilution series of 52LFK is shown, ranging from 2 μM to 8 nM. Results are shown as response units (RU). (F) ^{35}S-labeled HAdV-52 virion binding to A549 cells pretreated with *V. cholerae* neuraminidase (removes α2,3-, α2,6- and α2,8-linked sialic acid) or α2,3-specific neuraminidase at indicated concentrations. All experiments were performed three times with duplicate samples in each experiment. Figure E show one representative set of results. Error bars represent mean ± SD, n.s = not significant, * P of < 0.05, ** P of < 0.01 and *** P of < 0.001 versus control.

doi:10.1371/journal.ppat.1004657.g005

transmission of subsequently released virions within a tissue, or between tissues. Further support for this hypothesis is provided in that sialic acid-containing, *O*-linked glycans are abundant on the apical side of polarized epithelial cells *in vivo*, whereas CAR is mainly expressed laterally and basolaterally [45]. Thus it is plausible that virions approaching non-infected cells from the apical side have access to sialylated glycans, but not to CAR.

HAdV-37 has been shown to interact primarily with sialic acids linked via α2,3-glycosidic bonds to galactose (Siaα2,3Gal). We found here that Siaα2,3Gal-binding *M. amurensis* type II (MAL-II) lectins and/or Siaα2,6Gal-binding *S. nigra* (SNA) lectins did not compete with HAdV-52 virion binding to A549 cells (**S8 Fig.**). We noted that α2,3-specific neuraminidase inhibited HAdV-52 virion binding to A549 cells (**Fig. 5F**), but only at 100-fold higher concentrations than what has been observed for inhibition of HAdV-37 virion binding [11]. Pretreatment of A549 cells with neuraminidase from *V. cholerae*, which cleaves α2,3/6/8-linked sialic acids with similar efficiencies, inhibited HAdV-52 binding at much lower concentrations. By means of glycan microarray screening we identified a number of α2,3-sialylated probes that are bound by 52SFK, whereas no binding was detected with probes that contain exclusively α2,6-sialyl linkage (**Fig. 6**). The probe most strongly bound was a synthetic glycolipid with type II (Galβ-4GlcNAc) backbone sequence (GSC-273). In contrast, no binding was detected to the type I (Galβ-3GlcNAc) analog (GSC-272). Weaker binding was observed to three of the four sulfated sialyl analogs with or without 3-linked fucose. There was also weak binding to a neoglycolipid derived from GD1a glycan, the previously described ligand for HAdV-37 [35]. No binding was detected to GD1a glycosylceramide. It should be mentioned that, although five of the ligand-positive sialyl probes in the array were glycolipids, their glycan sequences are common to glycoproteins. Among the *N*-glycan probes analysed, there was binding to probes with α2,3-linked terminal sialic acids. Collectively, we cannot exclude α2,3-linked sialic acid-containing *N*-glycans from contributing to HAdV-52 binding to A549 cells, but it is likely that other types of sialic acid-containing glycans also contribute. Our results suggest that the short fiber is capable of binding to sialic acids on *O*-glycosylated proteins on A549 cells, but that on other cell types binding to sialyl-*N*-glycans may also occur.

Structure of the complex of HAdV-52 SFK bound to sialic acid

In order to define the interactions of the HAdV-52 fiber with sialic acid, we solved the crystal structure of 52SFK in complex with 2-*O*-methyl-sialic acid (a stereochemically uniform analogue of sialic acid) at a resolution of 1.65 Å. Similar to all other known AdV fiber knob structures [13,46–48] the 52SFK has a nine-stranded antiparallel β-sandwich fold and forms a stable trimer in solution. The three shallow sialic acid binding sites of 52SFK are formed at the contact site of two neighboring monomers by the EG and GH loops at the side of the short fiber knob domain (**Fig. 7A-C**). Two of the three binding sites are partially blocked by crystal contacts, and therefore the structure contains only one fully occupied sialic acid, while a second sialic acid is visible with partial occupancy in one of the two partially blocked binding sites. The location of this binding site is distinct from those of the other structurally characterized sialic acid-binding fiber knobs, HAdV-37 and canine adenovirus type 2 (CAdV-2; included here since it is the only known sialic acid-interacting AdV besides those of species HAdV-D)

Probe [a]	Sequence [b]	Fluorescence intensities [c]
GSC-272	**NeuAcα-3**Galβ-3GlcNAcβ-3Galβ-4Glcβ-C30	- [d]
GSC-273	**NeuAcα-3**Galβ-4GlcNAcβ-3Galβ-4Glcβ-C30	8,436
LSTc	**NeuAcα-6**Galβ-4GlcNAcβ-3Galβ-4Glc-DH	-
GSC-268	SU-6 | **NeuAcα-3**Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36 | Fucα-3	1,392
GSC-269	SU-6 | **NeuAcα-3**Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36 | Fucα-3	-
GSC-270	SU-6 SU-6 | | **NeuAcα-3**Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36 | Fucα-3	3,219
C4U	**NeuAcα-3**Galβ-4GlcNAcβ-3Galβ-3GlcNAc-DH | | | SU-6 SU-6 SU-6	2,732
GD1a	**NeuAcα-3**Galβ-3GalNAcβ-4Galβ-4Glcβ-Cer | **NeuAcα-3**	-
GD1a-hexa	**NeuAcα-3**Galβ-3GalNAcβ-4Galβ-4Glc-DH | **NeuAcα-3**	2,356
DSLNT	**NeuAcα-3**Galβ-3GlcNAcβ-3Galβ-4Glc-DH | **NeuAcα-6**	3,213
DST-AO	**NeuAcα-3**Galβ-3GalNAc-AO | **NeuAcα-6**	3,027
NA2	Galβ-4GlcNAcβ-2Manα-6 | Manβ-4GlcNAcβ-4GlcNAc-DH | Galβ-4GlcNAcβ-2Manα-3	113
A2F(2-3)	**NeuAcα-3**Galβ-4GlcNAcβ-2Manα-6 Fucα-6 | | Manβ-4GlcNAcβ-4GlcNAc-DH | **NeuAcα-3**Galβ-4GlcNAcβ-2Manα-3	3,575
A2(2-6)	**NeuAcα-6**Galβ-4GlcNAcβ-2Manα-6 | Manβ-4GlcNAcβ-4GlcNAc-DH | **NeuAcα-6**Galβ-4GlcNAcβ-2Manα-3	66
A3	**NeuAcα-3**Galβ-4GlcNAcβ-2Manα-6 | Manβ-4GlcNAcβ-4GlcNAc-DH | **NeuAcα-3**Galβ-4GlcNAcβ-4Manα-3 | **NeuAcα-6**Galβ-4GlcNAcβ-2	3,473

Fig 6. Features of 52SFK binding to selected sialyl sequences in the microarray. [a] The oligosaccharide probes are all lipid-linked, neoglycolipids (NGLs) or glycosylceramides and are from the collection assembled in the course of research in the Glycosciences Laboratory. [b] The selected α2-3-linked and α2-6-linked sialyl sequences are marked in **bold**. For definition of the lipid moieties of the probes, please see https://glycosciences.med.ic.ac.uk/docs/lipids.pdf [c] Numerical scores for the binding signals are shown as means of duplicate spots at 5 fmol per spot. [d]-, signal less than 1.

doi:10.1371/journal.ppat.1004657.g006

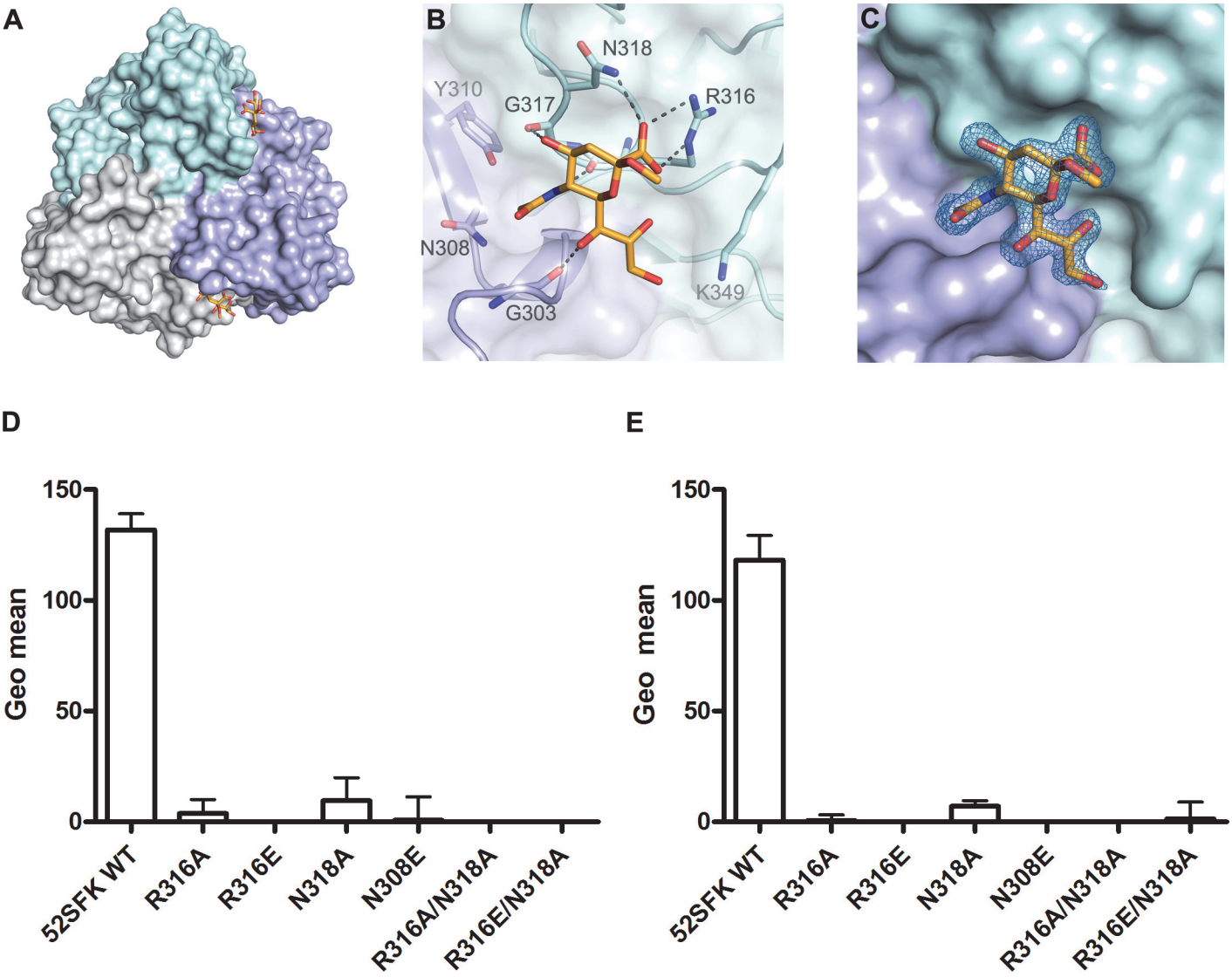


Fig 7. HAdV-52 interaction with sialic acid. (A) Surface representation of the HAdV-52 short fiber knob structure viewed from the top along the three-fold symmetry axis. The three monomers are shown in green, blue and grey. The 2-*O*-methyl sialic acid bound to two of three binding sites is shown as orange and red stick model. (B) Detailed view of the interactions in the ligand binding site. The side chain of R316 forms a bidentate salt bridge with the carboxyl group of sialic acid. Residues G303, R316 and G317 form backbone hydrogen bonds with the glycerol, amide and O4 groups of the sialic acid, respectively, while the side chain of N318 engages in a hydrogen bond with the sialic acid carboxylate. The methyl group is not involved in binding contacts. (C) Simulated annealing Fo-Fc omit map for the sialic acid. The map was calculated at 3.0 σ and is displayed with a radius of 1.7 Å around the ligand. 52SFK WT (wild type) and mutant binding to Pro5 (D) and A549 (E) cells. All experiments were performed three times with duplicate samples in each experiment. Error bars represent mean ± SD. * P of < 0.05, ** P of < 0.01 and *** P of < 0.001 versus control.

doi:10.1371/journal.ppat.1004657.g007

(Fig. 8; PDB_IDs 1UXA and 2WBV). The bound 2-*O*-methyl-sialic acid is well defined by electron density (Fig. 7C) and engages the 52SFK mainly through contacts between the sugar's carboxylate group and the side chains of R316 and N318 (Fig. 7B). The bidentate salt bridge formed by R316 is a prominent binding motif among glycan binding viruses [49,50] (S9 Fig.). In addition, the backbone carbonyl oxygens of R316, G317, and G303 form hydrogen bonds with the sialic acid O4, N-acetyl and glycerol-like functions, respectively (Fig. 7B). The tripeptide R316-G317-N318 located on the GH loop forms a hook-shaped motif (RGN motif) that is contributing most of the interactions, and that therefore largely defines the specificity of SFK52

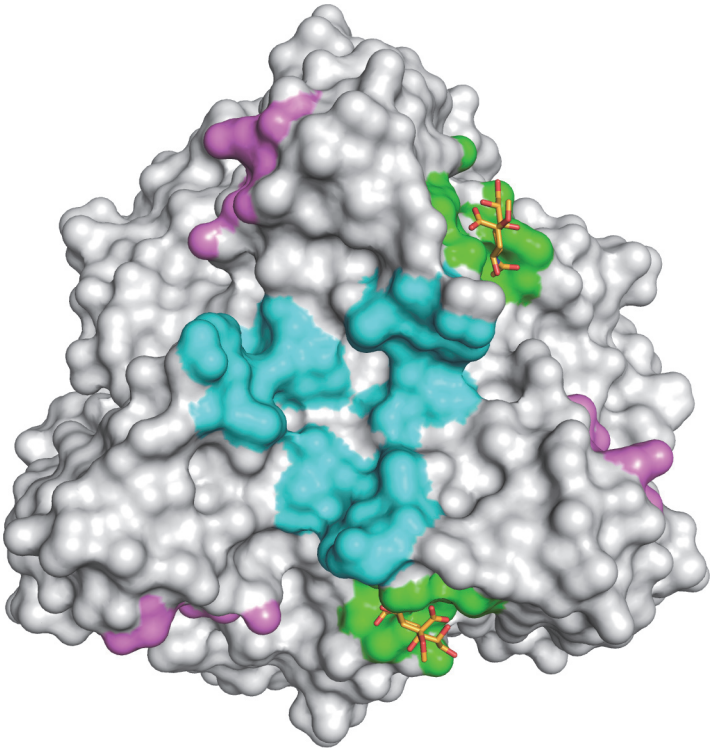


Fig 8. HAdV-52 has a unique sialic acid binding site compared to other sialic acid binding HAdVs. Comparison of the sialic acid binding site in HAdV-52 (green with orange sticks) with those of HAdV-37 (cyan) and CAdV-2 (magenta). The protein chains were superposed using the PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC) align tool, and the contact surface of each ligand was mapped onto HAdV-52 (calculated 4.5 Å around sialic acid).

doi:10.1371/journal.ppat.1004657.g008

for sialic acid. The pattern of polar contacts formed by this motif is highly similar to HAdV-37 (**S9 Fig.**) [50]. Mutating either R316 or N318 to alanine, replacing the R316 side chain with a negatively charged glutamate, or introducing a steric clash (and a polar clash, introducing a charge) at position 308 (N308E) all abolished the attachment of 52SFK to Pro-5 and A549 cells (**Fig. 7D,E**). The sugar's O2 function, to which additional sugars would be attached in a glycan chain, is pointing away from the protein and towards the tip of the knob, suggesting that more complex glycan receptors that bind the knob with their sialic acid caps would have to face towards the capsid in order to be bound by the virus. The methyl group attached to this oxygen in the compound used for structural analysis (2-*O*-methyl-sialic acid) does not participate in interactions with the knob. Alignment of multiple knob sequences suggests that the RGN motif is conserved in the knob domain of the short fibers of other members of species HAdV-G: simian AdV-1 and -7 (**S10 Fig.**), and we therefore predict that the ability to engage sialylated receptors is shared by these HAdVs. The RGN motif is not conserved in any other known human and non-human AdV knob sequences, including the short fiber knobs of HAdV-40 and -41.

In conclusion, we have identified two types of cellular receptors used by HAdV-52, the only human member of species HAdV-G. By analogy with HAdV-40 and -41, we identified CAR as a receptor for the HAdV-52 long fiber. We also present evidence that *O*-glycosylated proteins carrying sialic acid-containing glycans serve as receptors on A549 cells for HAdV-52 short fibers. The 52LFK:CAR interaction is probably of higher affinity than the 52SFK:sialic acid interaction, however the relative importance of cellular receptors is not only determined by the

affinity but also to a high extent on the abundance of the receptors. It has been shown for example by us and others that HAdV-37 binds with lower affinity to CAR (20 nM) than other HAdVs, but with even lower affinity to sialic acid-containing glycans (19 μM) [41,51]. Still, HAdV-37 uses sialic acid-containing glycoproteins as the main receptor [11,52]. Accessibility may also influence receptor usage. *In vivo*, CAR localizes to the lateral and basolateral side of polarized epithelial cells, which are not easily accessible for HAdVs, while sialic acid is abundant on the apical surface and may therefore be more available for interaction. Thus we suggest that sialylated proteins rather than CAR function as primary receptors for HAdV-52 virions on A549 cells. HAdV-52 might have retained the ability to bind to CAR as secretion of CAR-interacting fibers can disrupt CAR-CAR homodimers in the tight junctions and thereby facilitate virion escape and transmission within a tissue, or between tissues. The mode of interaction between 52SFK and its sialylated receptors is fundamentally different, both in location on the protein and in contacts formed to the ligand, from the known interactions between HAdV-37 fiber knob and sialic acid. As the sialic acid-binding RGN motif of 52SFK is not conserved in any other HAdV fiber, it appears that HAdV-52 and other members of species HAdV-G employ a unique strategy for engaging sialic acid. This is the first information presented about the receptors used by viruses in species HAdV-G, and the first describing a receptor recognized by an AdV short fiber. These findings shed light on AdV biology and tropism and may be useful for development of vectors based on members of species HAdV-G.

Materials and Methods

Cells, viruses and antibodies

A549 cells (gift from Dr. Alistair Kidd) were grown in Dulbecco´s modified Eagle medium (Sigma-Aldrich) supplemented with 5% fetal bovine serum (FBS: Invitrogen), 20 mM HEPES (Sigma-Aldrich) and 20 U/ml penicillin + 20 μg/ml streptomycin (Invitrogen), human corneal epithelial (HCE) cells (gift from Dr. Araki-Sasaki) were grown as previously described [53]. Pro-5 and Lec2 cells [54,55] (both purchased from LGC Promochem), Chinese hamster ovary (CHO)-CAR, CHO-MOCK (gift from Dr. Jeffrey Bergelson) [4], and CHO-CD46 (isoform BC1; gift from Dr John P. Atkinson) [56] were grown as described.

Species G HAdV-52 (strain TB3-2243)[3] and species C HAdV-5 (Ad75; source ATCC) virions were produced with or without ^{35}S-labeling in A549 cells as described previously [57], with the exception that the virions were eluted in sterile phosphate buffered saline (PBS) when desalting on a NAP column (GE Healthcare). Serotype-specific rabbit polyclonal antisera to each HAdV was a gift from Dr Göran Wadell [58]. Antiserum produced against HAdV-41 virions was used for detection of HAdV-52 antigens in infection experiments.

Virion binding experiments

Cells were detached with PBS containing 0.05% EDTA, reactivated in growth medium for one hour at 37°C (in solution), pelleted in 96 well plates ($2x10^5$ cells/well) and washed with binding buffer (BB: Dulbecco´s modified Eagle medium supplemented with 20 mM HEPES, 20 U/ml penicillin + 20°g/ml streptomycin and 1% bovine serum albumin).^{35}S-labeled virions ($2x10^9$ virions diluted in BB, 100 μl/sample) were added to the cells and incubated for 1 h on ice. Unbound virions were washed away with BB and the cell associated radioactivity was measured in a Wallac 1409 liquid scintillation counter (Perkin-Elmer). This experiment was performed with the following additions/variations:

i) Cells were pretreated with increasing concentrations of *Vibrio cholerae* neuraminidase (Sigma-Aldrich) or α2,3-sialidase (TaKaRa Bio Inc) for 1 h at 37°C, or, with a 1:100 dilution of

anti-CAR antibody (clone RmcB, Upstate, Millipore) or with 50 μg/ml *M. Aamurensis* type II (MAL-II) or *S. Nigra* (SNA) lectins (Vector Laboratories) for 1 h on ice before addition of virions. The effect of *V. cholerae* neuraminidase treatment of CHO cells was examined by flow cytometry. Cells were incubated with 1μg/ml of biotinylated MAL-II lectin for 30 min on ice, followed by a 30 min incubation with a 1:100 diluted streptavidin-FITC (on ice).

ii) Virions were preincubated for 1 h on ice with a) 100 mM sialic acid (or concentrations as indicated; N-acetylneuraminic acid, Dextra Laboratories), b) CAR-D1 (a kind gift from Dr. Paul Freimuth), c) physiological concentrations of coagulation factor IX (FIX 5 μg/ml, resulting in a 1 virion:2700 FIX ratio, Calbiochem) or coagulation factor X (FX 10 μg/ml, resulting in a 1 virion:5100 FX ratio, Haematologic Technologies Inc.) before incubation with A549 cells.

iii) To affect both CAR and sialic acid binding, A) virions were preincubated with a combination of sialic acid and CAR-D1 1h on ice before addition to A549 cells, B) cells were preincubated with a 1:100 dilution of anti-CAR antibody on ice, and virions were preincubated with sialic acid, C) cells were pretreated with *Vibrio cholerae* neuraminidase for 1 h at 37°C and virions were preincubated with CAR-D1 for 1 h on ice before they were mixed together, or D) cells were pretreated with *Vibrio cholerae* neuraminidase for 1 h at 37°C followed by incubation with a 1:100 dilution of anti-CAR antibody on ice before addition of virions. HAdV-5 was used as a CAR-binding control, to check the efficiency of soluble CAR-D1 and anti-CAR mab.iv) Cells were pretreated with or without 2.5 μM P4 [(1*R*,2*R*)-1-phenyl-2-hexadecanoylamino-3-pyrrolidino-1-propanol] or 2.5 μM inactive enantiomer of P4 (1*S*,2*S*, both kindly provided by Dr Roland L. Schnaar) for 5 days at 37°C to inhibit glycolipid biosynthesis via the glycosylceramide synthase enzyme [59]. Media was changed after 3 days when new P4 was added. The effect of P4 on the cells was analyzed by performing flow cytometry using ganglioside G_{M1}-binding AF488-conjugated cholera toxin subunit B (Invitrogen, Molecular probes). v) Cells were pretreated with 0.3 μg/ml tunicamycin (Sigma Aldrich) for 24 h at 37°C to inhibit *N*-glycosylation, or with 3 mM benzyl-α-GalNAc (Sigma-Aldrich) for 48 h at 37°C to inhibit *O*-glycosylation. To determine the effect of tunicamycin, cells were incubated with FITC-conjugated Phaseolus vulgaris erythroagglutinating lectin (PHA-E; Vector Laboratories) for 1 h on ice, washed and analyzed with flow cytometry. Cell viability was verified with trypan blue staining prior to binding experiments

Infection experiments

Pro-5, Lec2, or A549 cells, grown as monolayers on glass slides in 24-well plates, were washed three times with serum-free medium and treated with or without 10 mU/well of *Vibrio cholerae* neuraminidase for 1 h at 37°C. Virions were added to the cells and incubated for 1 h on ice. After incubation, the wells were washed three times with serum-free medium in order to remove unbound virions. Cell culture medium containing 1% FBS was added and the plates were incubated for 44 h at 37°C. Thereafter the glass slides were washed with PBS (pH 7.4) once, fixed with methanol and stained with polyclonal rabbit anti-HAdV diluted 1:200 for 1 h at room temperature. The slides were washed twice with PBS and incubated for an additional hour with a FITC-conjugated swine anti-rabbit IgG antibody (DakoCytomation) diluted 1:100 in PBS. After washing, the slides were mounted and examined in a fluorescence microscope using 20 X magnification (Axioskop2, Carl Zeiss). Ten pictures was taken of each well and the number of infected cells was calculated using ImageJ [60]. In one experiment, virions were preincubated with or without physiological concentrations of FIX (5 μg/ml, equal to 1:60000 virion:FIX ratio) or FX (10 μg/ml, equal to 1:110000 virion:FX ratio) for 1 h on ice before addition to cells.

Cloning and purification of fiber knobs

DNA isolation from HAdV-52 virions was performed by using the Blood & Cell Culture DNA Mini kit (Qiagen Nordic). DNA fragments encoding HAdV-52 long fiber knob (52LFK) and HAdV-52 short fiber knob (52SFK) were amplified by polymerase chain reaction (PCR) using KOD Hot Start (Novagen, Merck) and the following primers (DNA Technology): 52LFK forward (5´-aaaaggatccggaaacatagctgtttctcct), reverse (5´-aaaacccgggcggaggaagccttactgtgcgtgt), 52SFK forward (5´-aaaaggatccaggtttaacagcagt-ggagcc), reverse (5´-aaaacccgggagggttt tattgttcggtaatgtagca). Fragments were then cloned into a pQE30Xa expression vector encoding an N-terminal His-tag (Qiagen) using restriction sites for BamHI and XmaI (Fermentas, ThermoFisher Scientific). All constructs were confirmed by sequencing (Eurofins MWG Operon). Proteins were expressed in *Escherichia coli* (strain M15) and purified with Ni-NTA agarose beads according to protocol from the supplier (Qiagen). Proteins were analyzed by denaturing gel (NuPAGE Bis-Tris, Invitrogen, Life Technologies) and western blot with monoclonal antibodies directed against the His-tag (Qiagen).

Fiber knob mutants

Six different 52SFK mutants were created using a QuikChange mutagenesis kit (Agilent) according to their protocol. The following mutants were created: 1) R316A,2) R316E, 3) N318A, 4) N308E, 5) R316A/N318A, and 6) R316E/N318A. Correct trimerization of all proteins was confirmed with gas-phase electrophoretic mobility molecular analysis (GEMMA)[61] (**S11 Fig.**, showing 52SFK wt and one representative mutant; R316A), which is a protein oligomer measurement technique where the protein solution is converted into gas phase by a charged reduced electrospray process. The particles are separated according to size in a differential mobility analyzer and quantified by a particle counter. All mutant and wt fiber knobs were analyzed in the same manner: G-25 columns (GE Healthcare) were used for buffer exchange to 20mM ammonium acetate buffer, pH 7.8 containing 0.005% (v/v) Tween 20. Buffer exchange was done to remove the non-volatile salts from the protein solution. The concentrations for 52SFK wt and mutants 1, 2 and 3 were 0.05 mg/ml while for mutants 4, 5 and 6 it was 0.06 mg/ml. Three to five scans were taken with the GEMMA system (TSI Corp.) for each sample with a capillary pressure of either 1.7 or 3.7 psi. These parameters depended on the protein sample and the stability of the signal. Each sample was scanned for 120 seconds per scan at the size range of 2.55–255 nm. For molecular mass calculations, a particle density of 0.58 g/cm^3 was used.

Fiber knob binding experiments

Cells were detached with PBS-EDTA, reactivated in growth medium for one hour at 37°C, pelleted in 96 well plates (2x10^5 cells/well) and washed once with BB. The cells were then incubated with 10 μg/ml of 52SFK or 52LFK in 100 μl BB for one hour on ice. Unbound fiber knobs were washed away with PFN (PBS containing 2% FBS and 0.01% NaN_3) and the cells were then incubated with an anti RGS-His mouse monoclonal antibody (Qiagen; diluted 1:200 in PFN) for 30 min. Followed by one wash with PFN, the cells were incubated with polyclonal rabbit-anti-mouse FITC antibodies (Dako Cytomation; diluted 1:20 in PFN) for 30 min on ice. Thereafter the cells were washed with PFN and analyzed with flow cytometry using FACSLSRII instrument (Becton Dickinson). Results were analyzed using FACSDiva software (Becton Dickinson). This experiment was performed with the following additions/variations: The cells were i) grown in the presence or absence of benzyl-α-GalNAc (as described above), ii) preincubated with or without different concentrations of proteases (ficin, proteinase K and bromelain; all from Sigma-Aldrich) for 30 min at 37°C before incubation with fiber knobs, and iii) treated

with or without *Vibrio cholerae* neuraminidase for 1 h at 37°C before incubation with fiber knobs.

Western blot

Purified HAdV-52 virions were resolved on 10% Bis-Tris denaturing gels (NuPAGE, Invitrogen, Life Technologies) and transferred to Trans-Blot nitrocellulose membranes (Bio-Rad Laboratories, Solna, Sweden) by electroblotting. The membrane was blocked with 5% milk in PBS-T (PBS supplemented with 0.05% Tween20). Staining was carried out using 1:5000 dilution of a monoclonal anti-adenovirus fiber antibody (epitope region suggested by the manufacturer: MKRARPSEDTFNPVYPY, clone 4D2, ab3233, Abcam) in PBS-T with 2.5% milk, followed by a 1:1000 dilution of a HRP-conjugated rabbit anti-mouse IgG antibody (Dako Cytomation) in PBS-T with 2.5% milk. The fibers were then detected by chemiluminescence using super signal west pico or femto (Thermo Scientific) and visualized using the multipurpose CCD camera system FujiFilm LAS-4000. Pictures were taken every 10s and the relative abundance of the two fibers were evaluated using ImageJ.

Cellular expression of CAR

A549 and HCE cells were detached with PBS-EDTA, reactivated in growth medium for one hour at 37°C, pelleted in 96 well plates ($2x10^5$ cells/well) and washed once with PFN. The cells were then incubated with a mouse monoclonal antibody directed against CAR (E1-1, Merck Millipore) for 30 min on ice followed by one wash with PFN. A polyclonal rabbit-anti-mouse FITC antibody (Dako Cytomation) was added to the cells and incubated for 30 min on ice followed by one wash with PFN before flow cytometry analysis.

ELISA

96-well plates (Nunc maxisorp, Thermo Scientific) were coated with 1 μg/ml of fetuin or asialofetuin type I or II (Sigma-Aldrich) for 2 h at room temperature (RT) in coating buffer (bicarbonate/carbonate coating buffer 100 mM, pH 9.6). Meanwhile fiber knobs (0.4 μg/ml) were preincubated with monoclonal anti RGS-His antibodies (Qiagen; dilution 1:1000) in PBS-T for 1 h at RT. The wells were then washed four times with PBS-T and incubated with the fiber knob mixtures for 1 h at RT. After washing, the plate was incubated with a HRP-conjugated rabbit anti-mouse IgG antibody (Dako Cytomation; diluted 1:2000 in PBS-T) for 1 h at RT. The wells were washed again and incubated with 100 μl enhanced K-Blue TMB substrate (Neogen Europe) for 15 min and the reaction was then stopped by addition of 100 μl 1 M H_2SO_4. The absorbance was measured at 450 nm using Tecan infinite F2000 Pro (Tecan Nordic AB).

Determination of binding parameters by use of surface plasmon resonance (SPR)

All SPR experiments were performed at 25°C with a Biacore T100 instrument and a data collection rate of 1 Hz. For CAR interaction studies: CM5 sensor chips, amine-coupling kit, and HBS-EP+ buffer (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.005% [vol/vol] surfactant P20, pH 7.4) were all purchased from GE Healthcare. Recombinant human CAR (CXADR Fc chimera; R&D Systems; full length extracellular D1D2 domain), or CAR-D1 was coupled to the CM5 sensor chip by using the amine coupling reaction according to the manufacturer's instructions, resulting in an immobilization density of 900–1100 RU. The surface of the upstream flow cell was subjected to the same coupling reaction in the absence of protein and used as reference. All binding assays were carried out at 25°C, and HBS-EP+ buffer was used as running

buffer. The analytes (52LFK and 52SFK) were serially diluted in running buffer to prepare a two-fold concentration series ranging from 8 nM to 2 μM, and then injected in series over the reference and experimental biosensor surfaces for 120 s at a flow rate of 30 μl/min. Blank samples containing only running buffer were also injected under the same conditions to allow for double referencing. After each cycle, the biosensor surface was regenerated with a 60 s pulse of 10 mM Tris-Glycine pH 1.5 at a flow rate of 30 μl/min. For 52SFK interaction studies: Ni-NTA sensor chips, and HBS-EP+ buffer were purchased from GE Healthcare. 52SFK was diluted in running buffer (HBS-EP+) to a concentration of 0.03μM and captured on the Ni-NTA sensor chip according to the manufacturer's instructions, resulting in an immobilization density of 700 RU. In short: an automated program cycle of the following sequence: (1) activation of the sensor chip with Ni(II), (2) capture of 52SFK (3) analyte injection, (4) regeneration of the surface with 0.3 M EDTA, and (5) rinse with HBS-EP+ without EDTA. All steps were performed at a flow rate of 30 μl/min. All binding assays were carried out at 25°C, and HBS-EP+ buffer was used as running buffer. The analytes (fetuin and asialofetuin type II) were serially diluted in running buffer to prepare a two-fold concentration series ranging from 125 to 1 μM, and then injected in series over the reference and experimental biosensor surfaces for 180 s and a dissociation time of 100 s. Blank samples containing only running buffer were also injected under the same conditions to allow for double referencing.

Glycan microarray

Microarrays were composed of lipid-linked oligosaccharide probes robotically printed on nitrocellulose-coated glass slides at 2 and 5 fmol per spot in duplicate using a non-contact instrument as described previously [62]. Binding signals were probe-dose dependent. For 52SFK binding, the results of 15 oligosaccharide probes at 5 fmol per spot are shown in Fig. 6. The full microarray data will be described elsewhere. The microarray binding assay of the recombinant His-tagged 52SFK was performed at 20°C essentially as described [63]. In brief, the arrayed slide was blocked for 1 h with 5 mM HEPES pH 7.4, 150 mM NaCl, 5mM CaCl2, 0.3% (v/v) Blocker Casein (Pierce), 0.3% (w/v) bovine serum albumin (Sigma) (0.3% casein/0.3% BSA). 52SFK was 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). The 52SFK-antibody pre-complexes were prepared by pre-incubating Ab1 with Ab2 for 15 min at ambient temperature, followed by addition of 52SFK and incubation for an additional 15 min on ice. The VP1-antibody complexes were diluted in 0.3% casein/0.3% BSA, to give a final 52SFK concentration of 150 μg/ml, and overlaid onto the arrays at 20°C for 2 h. Binding was detected with Alexa Fluor-647-labelled streptavidin (Molecular Probes); imaging and data analysis was as described [62].

Structural analysis of HAdV-52:glycan interactions

Expression and purification. Residues 183–363 of the HAdV-52 short fiber (accession # DQ923122.2) were cloned into a pQE-30Xa vector (Qiagen) for expression of SFK52 (as described above). This construct was expressed in E. coli BL-21 DE3 at 20°C for approximately 16h after induction with 0.5 mM IPTG. Cells were harvested, resuspended in pellet buffer A_{His} [50 mM Tris (pH 7.5), 250 mM NaCl, 5% (V/V) glycerol, 10 mM imidazole] supplied with PMSF [1 mM] and lysozyme [1 mg/ml], and incubated at 4°C with agitation for 30 minutes. Cells were sonicated at maximal microtip setting with four two-minute cycles with 0.5 seconds pulses at a rate of 1 Hz. The resulting solution was centrifuged at 43,000 x g at 4°C for one hour, and the supernatant was filtered through 0.45 and 0.20 μM nitrocellulose filters. HAdV-52 was loaded onto a HisTrap Nickel IMAC column (GE Healthcare) using the Äkta Prime

FPLC system and washed with buffer A_{His} until the UV absorbance returned to baseline. After a washing step with a 10% solution of B_{His} [20 mM Tris (pH 7.5), 250 mM NaCl, 5% glycerol, 500 mM imidazole] the protein was eluted by a gradient of 50–500 mM imidazole (10–100% B_{His}). 20 mM DTT was added to the samples, followed by concentration and purification on a Superdex HiLoad 16/60 SD200 column (GE Life Sciences). During this step, the buffer was changed to gel filtration buffer [30 mM Tris (pH 7.5), 150 mM NaCl]. Purity was checked by SDS-PAGE analysis. Prior to crystallization, protein was concentrated to 8.5 mg/mL using a Millipore concentrator.

Crystallization, soaking and structure determination. For the high resolution structure, crystals of SFK52 were grown by hanging drop vapor diffusion in a reservoir solution of 15% (w/V) PEG_{1000}, 12.5% (w/V) PEG_{3350}, 12.5% (w/V) MPD, 20 μM of each Na L-glutamate, DL-alanine, glycine, DL-lysine HCl, and DL-serine, 0.1 M Tris/Bicine (pH = 8.5) using a combination of micro-and macroseeding. No cryoprotection was necessary. For complex formation, crystals were soaked with mother liquor supplied with 20 mM 2-*O*-methyl sialic acid for 1 hour. Data were collected at the PXIII beam line (SLS Villigen, Switzerland) at a wavelength of 1.000 Å using a Pilatus 2M detector. Initial phases were obtained by molecular replacement with Phaser [64] using a CHAINSAW model derived from the HAdV-12 fiber knob structure (PDB ID: 1NOB). Refinement was carried out using Phenix [65] and Buster [66] using three-fold NCS averaging, TLS ans isotropic B factor refinement. Data statistics are given in **S12 Fig**. Structural figures were prepared with PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC).

Alignment

Homology sequence alignment was performed using the Clustal W algorithm and the online server: http://www.ebi.ac.uk/Tools/clustalw2/index.html

Statistical analysis

All experiments were performed three times with duplicate samples in each experiment. The results are expressed as means ± standard deviations and either t-test or one-way ANOVA with Dunnett's post test was performed using GraphPad Prism version 4.00 for Windows, GraphPad Software, San Diego California USA. P-values < 0.05 were considered statistically significant.

Supporting Information

S1 Fig. MAL-II binding to CHO-cells. Binding of the sialic acid binding MAL-II lectin to CHO cells expressing or lacking known HAdV receptors. Pro-5 is a sialic acid-positive, reference cell line, CHO-CD46 and CHO-CAR cells express human CD46 and CAR, respectively. CHO-MOCK is mock transfected with respect to CHO-CAR. Black bars show MAL-II binding to cells after pretreatment with *V*. *cholerae* neuraminidase.
(TIF)

S2 Fig. Effect of RmcB and soluble CAR-D1 on HAdV-5 binding to A549 cells. ^{35}S-labeled HAdV-5 virion binding to A549 cells after virion preincubation with or without soluble CAR-D1 or cell preincubation with or without mouse anti-CAR mab (clone RmcB). The experiment was performed three times with duplicate samples in each experiment. Error bars represent mean ± SD. ∗∗ P of < 0.01.
(TIF)

S3 Fig. Control of P4 active/inactive and tunicamycin function. (A) Ganglioside G_{M1}-binding AF488-conjugated cholera toxin subunit B (CT-B) binding to A549 cells pretreated with active or inactive forms of P4 (inhibitor of glycolipid synthesis). CT-B binding was analyzed using flow cytometry. (B) *N*-linked glycan-binding Phaseolus vulgaris erythroagglutinating lectin (PHA-E; FITC-conjugated) binding to A549 cells pretreated with tunicamycin (inhibitor of *N*-linked glycan synthesis. Note that PHA-E can only bind to a specific subset of N-linked glycans). PHA-E binding was analyzed using flow cytometry. All experiments were performed three times with duplicate samples in each experiment. Error bars represent mean ± SD. ∗ P of < 0.05, ∗∗ P of < 0.01 and ∗∗∗ P of < 0.001 versus control.
(TIF)

S4 Fig. Potential CAR- and sialic acid-interacting residues in AdV fiber knobs. CAR-engaging residues in the knobs of HAdV-5[40] and -12[15] are shown on yellow background. Residues in direct contact with CAR are Asp191, Leu202, Lys205 (HAdV-12) and in indirect contact (via water) are Pro193, Pro194, Val226, Lys227, Gln263, Gln270, Ser273, Val274, Asn296, and Glu299 (HAdV-12). CAR-interacting residues of HAdV-5 (identified by mutagenesis): Ser193, Pro194, Lys201, Lys205, Asp259, Pro260, Glu261, Tyr262 and Tyr276. Sialic acid-engaging residues in the knob of HAdV-37 [41,50] and CAdV-2 [51] are shown on blue background. Residues in direct contact with sialic acid are Tyr312, Pro317, and Lys345 (all HAdV-37) and Ser237, Gln238, Ser240, Asn256 and Arg336 (all CAdV-2) and residues in indirect contact are Tyr 308, Thr310, Val322 and Ser344 (all HAdV-37). Potentially conserved CAR- and sialic acid-interacting residues in 52LFK and 52SFK are shown on yellow and blue backgrounds, respectively. Secondary-structure beta strands elements of HAdV-5 are indicated with arrows.
(TIF)

S5 Fig. Surface plasmon resonance analysis of fetuin and asialofetuin type II (in solution) binding to 52SFK (immobilized). A twofold dilution series of fetuin and asialofetuin is shown, ranging from 125 μM to 1 μM. The affinity of the 52SFK:fetuin interaction was calculated to 37 μM, whereas no affinity could be calculated for the 52SFK:asialofetuin interaction. Results are shown as response units (RU).
(TIF)

S6 Fig. Affinity and kinetics of HAdV-52 long fiber interaction with CAR-D1. Surface plasmon resonance analysis of 52LFK (in solution) binding to CAR-D1 (immobilized). A twofold dilution series of 52LFK is shown, ranging from 2 μM to 8 nM. Results are shown as response units (RU).
(TIF)

S7 Fig. Surface plasmon resonance analysis of 52SFK (in solution) binding to CAR-D1D2 (immobilized). No affinity could be calculated for the HAdV-52 short fiber knob interaction with CAR-D1D2. Results are shown as response units (RU). The experiment was performed three times and the figure shows one representative set of results.
(TIF)

S8 Fig. HAdV-52 binding is not affected by preincubation of cells with MAL-II and or SNA lectins. ^{35}S-labeled HAdV-52 virion binding to A549 cells preincubated with *M. Amurensis* lectin type II (MAL-II; preferentially binds to α2,3-linked sialic acid), *S. Nigra* lectin (SNA; preferentially binds to α2,6-linked sialic acid), or both. Virion binding was quantified by liquid scintillation counting. The experiment was performed three times with duplicate samples in

each experiment. Error bars represent mean ± SD.
(TIF)

S9 Fig. Comparison of the polar contacts of HAdV-52 and HAdV-37 bound to sialic acid. The sialic acid moieties of HAdV-52 and HAdV-37[50] (PDB-ID: 1UXA) were superimposed using the "align" function in PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC). The sialic acid moiety of HAdV-52 is shown in orange, and the sialic acid bound to HAdV-37 is overlaid as a ghost. Polar contacts formed with HAdV-52 and the respective residues are colored orange, contacting residues of HAdV-37 and the respective bonds are colored light blue. Although the binding pocket of HAdV-37 is located in an entirely different part of the knob and the interacting amino acids are not conserved, the polar contacts formed are highly similar to those of the RGN motif. The salt bridge contributed by R316 in HAdV-52 is formed by K345 in HAdV-37. The hydrogen bonds formed with the sugar's O4 and N-acetyl group are also retained.
(TIF)

S10 Fig. Conservation of sialic acid-interacting residues in 52SFK. Sialic acid-interacting residues in 52SFK are shown on red background, together with similar, potential sialic acid-interacting residues of other AdV:s. Representative types have been selected from human species A-G and canine species A (HA-G and CA, respectively). Secondary-structure elements (beta strands) of HAdV-5 are indicated by arrows.
(TIF)

S11 Fig. Trimerization of HAdV-52 knobs. A) 52SFK wt and B) 52SFK R316A mutant were separated according to size using gas-phase electrophoretic mobility molecular analysis (GEMMA). Results are shown as number of molecules (particles) in respect to molecular weight (kDa).
(TIF)

S12 Fig. X-ray data collection and refinement statistics. Values for the highest resolution shell are shown in parenthesis.
(TIF)

Acknowledgments

We thank Kristina Lindman for technical support. The microarrays contain many saccharides provided by collaborators whom we thank as well as members of the Glycosciences Laboratory for their collaboration in the establishment of the neoglycolipid-based microarray system.

Author Contributions

Conceived and designed the experiments: AL AML YL WB LF TF TS NA. Performed the experiments: AL AML YL CÅ EN WB LF AR LK. Analyzed the data: AL AML YL WB LF TF TS NA AR. Contributed reagents/materials/analysis tools: AL AML YL CÅ EN WB MSJ LF AR LK. Wrote the paper: AL AML YL TF TS NA.

References

1. Harrach B, Benkö M, Both GW, Brown M, Davison AJ, et al. (2011) Family Adenoviridae. In: King AMQ, Adams MJ, Carstens EB, LE J, editors. Virus Taxonomy Ninth Report of the International Committee on Taxonomy of Viruses. San Diego: Elsevier Academic Press. pp. 125–141.
2. Wold WSM, Horwitz MS (2007) Adenoviruses. In: Knipe DM, Howley PM, editors. Fields Virology. 5 ed. Philadelphia: Lippincott Williams & Wilkins. pp. 2395–2436.

3. Jones MS, Harrach B, Ganac RD, Gozum MM, Dela Cruz WP, et al. (2007) New adenovirus species found in a patient presenting with gastroenteritis. J Virol 81: 5978–5984. PMID: 17360747

4. Bergelson JM, Cunningham JA, Droguett G, Kurt-Jones EA, Krithivas A, et al. (1997) Isolation of a common receptor for Coxsackie B viruses and adenoviruses 2 and 5. Science 275: 1320–1323. PMID: 9036860

5. Tomko RP, Xu R, Philipson L (1997) HCAR and MCAR: the human and mouse cellular receptors for subgroup C adenoviruses and group B coxsackieviruses. Proc Natl Acad Sci U S A 94: 3352–3356. PMID: 9096397

6. Roelvink PW, Lizonova A, Lee JG, Li Y, Bergelson JM, et al. (1998) The coxsackievirus-adenovirus receptor protein can function as a cellular attachment protein for adenovirus serotypes from subgroups A, C, D, E, and F. J Virol 72: 7909–7915. PMID: 9733828

7. Gaggar A, Shayakhmetov DM, Lieber A (2003) CD46 is a cellular receptor for group B adenoviruses. Nat Med 9: 1408–1412. PMID: 14566335

8. Segerman A, Atkinson JP, Marttila M, Dennerquist V, Wadell G, et al. (2003) Adenovirus type 11 uses CD46 as a cellular receptor. J Virol 77: 9183–9191. PMID: 12915534

9. Marttila M, Persson D, Gustafsson D, Liszewski MK, Atkinson JP, et al. (2005) CD46 is a cellular receptor for all species B adenoviruses except types 3 and 7. J Virol 79: 14429–14436. PMID: 16254377

10. Wang H, Li ZY, Liu Y, Persson J, Beyer I, et al. (2011) Desmoglein 2 is a receptor for adenovirus serotypes 3, 7, 11 and 14. Nat Med 17: 96–104. doi: 10.1038/nm.2270 PMID: 21151137

11. Arnberg N, Edlund K, Kidd AH, Wadell G (2000) Adenovirus type 37 uses sialic acid as a cellular receptor. J Virol 74: 42–48. PMID: 10590089

12. Arnberg N, Kidd AH, Edlund K, Olfat F, Wadell G (2000) Initial interactions of subgenus D adenoviruses with A549 cellular receptors: sialic acid versus alpha(v) integrins. J Virol 74: 7691–7693. PMID: 10906228

13. Nilsson E (2011) Cellular receptors for viruses with ocular tropism. Umeå: Umeå university. doi: 10.1080/17437199.2011.587961 PMID: 25473706

14. Rademacher C, Bru T, McBride R, Robison E, Nycholat CM, et al. (2012) A Siglec-like sialic-acid-binding motif revealed in an adenovirus capsid protein. Glycobiology 22: 1086–1091. doi: 10.1093/glycob/cws073 PMID: 22522600

15. Bewley MC, Springer K, Zhang YB, Freimuth P, Flanagan JM (1999) Structural analysis of the mechanism of adenovirus binding to its human cellular receptor, CAR. Science 286: 1579–1583. PMID: 10567268

16. Persson BD, Müller S, Reiter DM, Schmitt BBT, Marttila M, et al. (2009) An arginine switch in the species B adenovirus knob determines high-affinity engagement of the cellular receptor CD46. J Virol 83: 673–686. doi: 10.1128/JVI.01967-08 PMID: 18987134

17. Wu E, Pache L, Von Seggern DJ, Mullen TM, Mikyas Y, et al. (2003) Flexibility of the adenovirus fiber is required for efficient receptor interaction. J Virol 77: 7225–7235. PMID: 12805421

18. Waddington SN, McVey JH, Bhella D, Parker AL, Barker K, et al. (2008) Adenovirus serotype 5 hexon mediates liver gene transfer. Cell 132: 397–409. doi: 10.1016/j.cell.2008.01.016 PMID: 18267072

19. Lenman A, S. M, Nygren MI, Frängsmyr L, Stehle T, et al. (2011) Coagulation factor IX mediates serotype-specific binding of species A adenoviruses to host cells. J Virol 85: 13420–13431. doi: 10.1128/JVI.06088-11 PMID: 21976659

20. Kidd AH, Chroboczek J, Cusack S, Ruigrok RW (1993) Adenovirus type 40 virions contain two distinct fibers. Virology 192: 73–84. PMID: 8517033

21. Yeh HY, Pieniazek N, Pieniazek D, Gelderblom H, Luftig RB (1994) Human adenovirus type 41 contains two fibers. Virus Res 33: 179–198. PMID: 7975882

22. Roy S, Sandhu A, Medina A, Clawson DS, M. WJ (2012) Adenoviruses in fecal samples from asymptomatic rhesus macaques, United States. Emerg Infect Dis 18: 1081–1088. doi: 10.3201/eid1807.111665 PMID: 22709783

23. Wei CJ, Boyington JC, McTamney PM, Kong WP, Pearce MB, et al. (2010) Induction of broadly neutralizing H1N1 influenza antibodies by vaccination. Science 329: 1060–1064. doi: 10.1126/science.1192517 PMID: 20647428

24. Barouch DH, Liu J, Li H, Maxfield LF, Abbink P, et al. (2012) Vaccine protection against acquisition of neutralization-resistant SIV challenges in rhesus monkeys. Nature 482: 89–93. doi: 10.1038/nature10766 PMID: 22217938

25. Roberts DM, Nanda A, Havenga MJ, Abbink P, Lynch DM, et al. (2006) Hexon-chimaeric adenovirus serotype 5 vectors circumvent pre-existing anti-vector immunity. Nature 441: 239–243. PMID: 16625206

26. Heise C, Hermiston T, Johnson L, Brooks G, Sampson-Johannes A, et al. (2000) An adenovirus E1A mutant that demonstrates potent and selective systemic anti-tumoral efficacy. Nat Med 6. PMID: 11100133

27. Kozarsky KE, Jooss K, Donahee M, F. SJ, M. WJ (1996) Effective treatment of familial hypercholesterolaemia in the mouse model using adenovirus-mediated transfer of the VLDL receptor gene. Nat Gen 13: 54–62. PMID: 8673104

28. George SJ, Wan S, Hu J, MacDonald R, Johnson JL, et al. (2011) Sustained reduction of vein graft neointima formation by ex vivo TIMP-3 gene therapy. Circulation 124: S135–142. doi: 10.1161/CIRCULATIONAHA.110.012732 PMID: 21911803

29. Stadtfeld M, Nagaya M, Utikal J, Weir G, Hochedlinger K (2008) Induced pluripotent stem cells generated without viral integration. Science 322: 945–949. doi: 10.1126/science.1162494 PMID: 18818365

30. Zak DE, Andersen-Nissen E, Peterson ER, Sato A, Hamilton MK, et al. (2012) Merck Ad5/HIV induces broad innate immune activation that predicts CD8? T-cell responses but is attenuated by preexisting Ad5 immunity. Proc Natl Acad Sci U S A 109: 3503–3512.

31. Li Y, Pong RC, Bergelson JM, Hall MC, Sagalowsky AI, et al. (1999) Loss of adenoviral receptor expression in human bladder cancer cells: a potential impact on the efficacy of gene therapy. Cancer Res 59: 325–330. PMID: 9927041

32. Li H, Rhee EG, Masek-Hammerman K, Teigler JE, Abbink P, et al. (2012) Adenovirus serotype 26 utilizes CD46 as a primary cellular receptor and only transiently activates T lymphocytes following vaccination of rhesus monkeys. J Virol 86: 10862–10865. doi: 10.1128/JVI.00928-12 PMID: 22811531

33. Thirion C, Lochmüller H, Ruzsics Z, Boelhauve M, König C, et al. (2006) Adenovirus vectors based on human adenovirus type 19a have high potential for human muscle-directed gene therapy. Hum Gene Ther 17: 193–205. PMID: 16454653

34. Arnberg N (2012) Adenovirus receptors: implications for targeting of viral vectors. Trends Pharmacol Sci 33: 442–448. doi: 10.1016/j.tips.2012.04.005 PMID: 22621975

35. Jarecki-Khan K, Unicomb LE (1992 Oct;) Seroprevalence of enteric and nonenteric adenoviruses in Bangladesh. J Clin Microbiol 30: 2733–2734. PMID: 1400977

36. Saderi H, Roustai MH, Sabahi F (2000) Antibodies to enteric adenoviruses (Ad40 and Ad41) in sera from Iranian children. J Clin Virol 16: 145–147. PMID: 10720819

37. Shinozaki T, Araki K, Ushijima H, Fujii R (1987) Antibody response to enteric adenovirus types 40 and 41 in sera from people in various age groups. J Clin Microbiol 25: 1679–1682. PMID: 3654940

38. Bányai K, Martella V, Meleg E, Kisfali P, Péterfi Z, et al. (2009) Searching for HAdV-52, the putative gastroenteritis-associated human adenovirus serotype in Southern Hungary. New Microbiol 32: 185–188. PMID: 19579697

39. Song JD, Liu XL, Chen DL, Zou XH, Wang M, et al. (2012) Human adenovirus type 41 possesses different amount of short and long fibers in the virion. Virology 432: 336–342. doi: 10.1016/j.virol.2012.05.020 PMID: 22727834

40. Roelvink PW, Mi Lee G, Einfeld DA, Kovesdi I, Wickham TJ (1999) Identification of a conserved receptor-binding site on the fiber proteins of CAR-recognizing adenoviridae. Science 286: 1568–1571. PMID: 10567265

41. Nilsson E, Storm RJ, Bauer J, Johansson SM, Lookene A, et al. (2011) The GD1a glycan is a cellular receptor for adenoviruses causing epidemic keratoconjunctivitis. Nat Med 17: 105–109. doi: 10.1038/nm.2267 PMID: 21151139

42. Kirby I, Lord R, Davison E, Wickham TJ, Roelwink PW, et al. (2001) Adenovirus type 9 fiber knob binds to the coxsackie B virus-adenovirus receptor (CAR) with lower affinity than fiber knobs of other CAR-binding adenovirus serotypes. J Virol 75: 7210–7214. PMID: 11435605

43. Walters RW, Freimuth P, Moninger TO, Ganske I, Zabner J, et al. (2002) Adenovirus fiber disrupts CAR-mediated intercellular adhesion allowing virus escape. Cell 110: 789–799. PMID: 12297051

44. Lu ZZ, Wang H, Zhang Y, Cao H, Li Z, et al. (2013) Penton-dodecahedral particles trigger opening of intercellular junctions and facilitate viral spread during adenovirus serotype 3 infection of epithelial cells. PLoS Pathog Oct; 9(10):e1003718. doi: 10.1371/journal.ppat.1003718 PMID: 24204268

45. Walters RW, Grunst T, Bergelson JM, Finberg RW, Welsh MJ, et al. (1999) Basolateral localization of fiber receptors limits adenovirus infection from the apical surface of airway epithelia. J Biol Chem 274: 10219–10226. PMID: 10187807

46. Xia D, Henry LJ, Gerard RD, Diesenhofer J (1995) Crystal structure of the receptor-bining domain of adenovirus type 5 fiber protein at 1.7 Å resolution. Sructure 2: 7907–7915.

47. Persson BD, Reiter DM, Marttila M, Mei YF, Casasnovas JM, et al. (2007) Adenovirus type 11 binding alters the conformation of its receptor CD46. Nat Struct Mol Biol 14: 164–166. PMID: 17220899

48. Seiradake E, Cusack S (2005) Crystal structure of enteric adenovirus serotype 41 short fiber head. J Virol 79: 14088–14094. PMID: 16254343

49. Neu U, Bauer J, Stehle T (2011) Viruses and sialic acids: rules of engagement. Curr Opin Struct Biol 21: 610–618. doi: 10.1016/j.sbi.2011.08.009 PMID: 21917445

50. Burmeister WP, Guilligay D, S. C, Wadell G, Arnberg N (2004) Crystal structure of species D adenovirus fiber knobs and their sialic acid binding sites. J Virol 78: 7727–7736. PMID: 15220447

51. Seiradake E, Henaff D, Wodrich H, O. B, Perreau M, et al. (2009) The cell adhesion molecule "CAR" and sialic acid on human erythrocytes influence adenovirus in vivo biodistribution. PLoS Pathog 5: e1000277. doi: 10.1371/journal.ppat.1000277 PMID: 19119424

52. Seiradake E, Lortat-Jacob H, Billet O, Kremer EJ, Cusack S (2006) Structural and mutational analysis of human Ad37 and canine adenovirus 2 fiber heads in complex with the D1 domain of coxsackie and adenovirus receptor. J Biol Chem 281: 33704–33716. PMID: 16923808

53. Araki-Sasaki K, Ohasi KY, Sasabe T, Hayashi K, Watanabe H, et al. (1995) An SV-40-immortalized human corneal epithelial cell line and its characterization. Investig Ophtahlmol 36: 614–621.

54. Stanley P, Caillibot V, Siminovitch L (1975) Selection and characterization of eight phenotypically distinct lines of lectin-resistant Chinese hamster ovary cell. Cell 6: 121–128. PMID: 1182798

55. Deutscher SL, Nuwayhid N, Stanley P, Briles EI, Hirschberg CB (1984) Translocation across Golgi vesicle membranes: a CHO glycosylation mutant deficient in CMP-sialic acid transport. Cell 39: 295–299. PMID: 6498937

56. Liszewski MK, Atkinson JP (1996) Membrane cofactor protein (MCP; CD46). Isoforms differ in protection against the classical pathway of complement. J Immunol 156: 4415–4421. PMID: 8666815

57. Johansson SM, Nilsson EC, Elofsson M, Ahlskog N, Kihlberg J, et al. (2007) Multivalent sialic acid conjugates inhibit adenovirus type 37 from binding to and infecting human corneal epithelial cells. Antiviral Res 73: 92–100. PMID: 17014916

58. Wadell G, Allard A, Hierholzer JC (1999) Adenoviruses. In: Murray PR, Baron EJ, Pfaller MA, Tenover FC, Yolken RH, editors. Manual of Clinical Microbiology. 7th ed. Washington: ASM Press. pp. 970–982.

59. Mistry N, Inoue H, Jamshidi F, Storm RJ, Oberste MS, et al. (2011) Coxsackievirus A24 variant uses sialic acid-containing O-linked glycoconjugates as cellular receptors on human ocular cells. J Virol 85: 11283–11290. doi: 10.1128/JVI.05597-11 PMID: 21880775

60. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9: 671–675. PMID: 22930834

61. Rofougaran R, Vodnala M, Hofer A (2006) Enzymatically active mammalian ribonucleotide reductase exists primarily as an alpha6beta2 octamer. J Biol Chem 281: 27705–27711. PMID: 16861739

62. Liu Y, Childs RA, Palma AS, Campanero-Rhodes MA, Stoll MS, et al. (2012) Neoglycolipid-based oligosaccharide microarray system: preparation of NGLs and their noncovalent immobilization on nitrocellulose-coated glass slides for microarray analyses. Methods Mol Biol 808: 117–136. doi: 10.1007/978-1-61779-373-8_8 PMID: 22057521

63. Neu U, Khan ZM, Schuch B, Palma AS, Liu Y, et al. (2013) Structures of B-Lymphotropic Polyomavirus VP1 in complex with oligosaccharide ligands. PLoS Pathog 9: e1003714. doi: 10.1371/journal.ppat.1003714 PMID: 24204265

64. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, et al. (2007) Phaser crystallographic software. J Appl Crystallogr 40: 658–674. PMID: 19461840

65. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, et al. (2002) PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr 58: 1948–1954. PMID: 12393927

66. Bricogne G. BE, Brandl M., Flensburg C., Keller P., Paciorek W., Roversi P, Sharff A., Smart O.S., Vonrhein C., Womack T.O. (2011) BUSTER version 2.10.0. Cambridge, United Kingdom: Global Phasing Ltd. doi: 10.1080/17437199.2011.587961 PMID: 25473706

Human adenovirus 52 short fiber binds to polysialic acid

Annasara Lenman[1*], A. Manuel Liaci[2*], Yan Liu[3], Lars Frängsmyr[1] , Bärbel S. Blaum[2], Iva I. Podgorski[4], Balazs Harrach[4], Mária Benkő[4], Ten Feizi[3], Thilo Stehle[2,5] and Niklas Arnberg[1]

[1]Division of Virology, Department of Clinical Microbiology, and Laboratory for Molecular Infection Medicine Sweden, Umeå University, SE-90185 Umeå, Sweden, [2]University of Tübingen, Interfaculty Institute of Biochemistry, D-72076 Tübingen, Germany, [3]Glycosciences Laboratory, Department of Medicine, Imperial College London, W12 0NN, UK, [4]Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian Academy of Sciences, Budapest, Hungary [5]Vanderbilt University School of Medicine, Nashville, TN 37232, USA

*These authors contributed equally to this work.

Human adenovirus 52 (HAdV-52) is one of only three HAdVs equipped with two different fiber proteins, one long and one short. The knob domain of the HAdV-52 short fiber (52SFK) binds to previously unknown sialic acid-containing glycans on the cellular surface of target cells. Using glycan array screening we show here that 52SFK displays a selectivity for long chains of [α-2,8]-linked (poly) sialic acids (PSia). Cell based knob attachment assays confirmed a strong preference for 52SFK binding to PSia- expressing cells compared to PSia-deficient cells. This binding was efficiently reduced by pre-incubating 52SFK with soluble oligo-PSias. X-ray crystallography analysis of 52SFK in complex with oligo-PSia revealed engagement at the non-reducing end of oligo-PSia to the canonical sialic acid binding site, but also suggested the presence of a 'steering rim' consisting of positively charged amino acids contributing to the interaction by long-range electrostatic interactions. Here we discuss the role of PSia as a receptor for HAdV-52 and the potential impact of using PSia-targeted HAdV52-based vectors for treatment of PSia-overexpressing cancer forms.

Human adenoviruses (HAdVs) are frequently explored as vectors for multiple applications, including vaccination (*1-3*), cardiovascular applications (*4*) and treatment of cancer (*5, 6*). HAdV-based vectors can be manipulated to conditionally replicate in and kill cancer cells, while leaving healthy cells unaffected (*7*). Besides lysing cancer cells as a direct effect of viral replication, these vectors can also cause cell death by expression of cytotoxic proteins (*8, 9*), activation of anti-tumor immune response (*10, 11*) and sensitizing tumor cells to the cytotoxic effect of radiotherapy and chemotherapy (*12*). This new form of cancer therapy has the potential to be of importance especially for certain types of cancer that are not responding to conventional treatments. Glioblastoma multiforme (GBM) is the most common primary malignant brain tumor, and even with state-of-the-art treatment patients usually survive only 12-15 months after diagnosis (*13*). To combat GBM, HAdV-based vectors have been developed that carry conditional cytotoxic genes such as herpes simplex virus type 1-thymidine kinase (*7, 14*). This enzyme is involved in the conversion of the pro-drug ganciclovir to its active tri-phosphorylated form, causing termination of DNA chain polymerization and cell death in proliferating cancer cells (*8*).

However, most HAdV vectors explored in clinical trials are still based on HAdV-5, despite the fact that decades of research have revealed a number of problems such as pre-existence of neutralizing antibodies in the patient (*15*); poor access to their primary attachment receptor CAR (coxsackie- and adenovirus receptor) on target cells (*16*); and coagulation factor-dependent off-target effects, which reduce bioavailability of the vector and

decrease transduction of target cells (*17*). Several strategies have been developed to overcome these problems, including usage of vector candidates based on less common HAdV types to reduce the amount of pre-existing antibodies (*3*); creation of chimeric vectors that retarget vectors to alternative receptors (*18-20*); and/or abolish interactions with coagulation factors to reduce off-target effects (*21*). For many applications, an ideal vector would harbor all these features by nature.

HAdV-52 was isolated in 2003 from a small outbreak of gastroenteritis (*22*). It was shown to be divergent from other HAdVs and was therefore classified into a new species, HAdV-G. HAdVs are normally equipped with one fiber protein but HAdV-52, along with the two more common species F types HAdV-40 and -41, differ from all other HAdVs by having two different fiber proteins, one long and one short (*22-24*). We showed recently that the knob domain of the HAdV-52 long fiber (52LFK) binds to CAR and that the knob domain of the short fiber (52SFK) binds to sialic acid (Sia)-containing glycoproteins on target cells (*25*). The close relationship of HAdV-52 to simian AdVs (SAdVs) and the low frequency of detection in humans suggest a low seroprevalence (*22, 26*). Together with its inability to interact with coagulation factors (*25*) we believe that HAdV-52 would constitute a good vector candidate.

Here we show that HAdV-52 binds to a class of glycans known as polysialic acid through an interaction involving the short fiber knob. In humans, the *N*-linked attachment of PSia to its target proteins is tightly regulated both spatially and temporally through the expression of the two sialyltransferases ST8Sia-II and -IV. The expression of both enzymes is largely restricted to the developing brain during embryogenesis, and their major target protein is the neural cell adhesion molecule (NCAM) (*27*). Apart from being present in the gut microbiota (*28*), brain regions requiring persistent neuronal plasticity (*29*), and under specific conditions on dendritic cells (*30*)**,** PSia is usually not detected in healthy adults. However, several types of cancer including glioma (*31-33*), neuroblastoma (*34, 35*) and lung cancer (*36, 37*) reactivate the synthetic machinery of polysialylation, which is strongly associated with a poor prognosis and an aggressive and invasive disease development (*29, 33, 38*). Therefore, PSia has been recognized as a highly promising tumor marker, and researchers have been trying to redirect Ad5-based oncolytic adenovirus vectors to target PSia (*39*). An adenovirus vector with a natural preference for PSia-expressing cancer cells might thus prove useful for treatment of several cancer types.

Here we set out to characterize the interaction between the short fiber knob of HAdV-52 and its cellular receptor PSia.

MATERIALS AND METHODS

Cells

Human respiratory epithelial A549 cells (gift from Dr. Alistair Kidd) were grown in Dulbecco´s modified Eagle medium (DMEM, Sigma-Aldrich) supplemented with 5% fetal bovine serum (FBS: Invitrogen), 20 mM HEPES (Sigma-Aldrich) and 20 U/ml penicillin + 20 µg/ml streptomycin (Invitrogen). Human neuroblastoma SK-N-SH cells (purchased from LGC Promochem) were grown in DMEM supplemented with 10% FBS, 20 mM HEPES, 20 U/ml penicillin, 20 µg/ml streptomycin. Human neuroblastoma SH-SY5Y cells were purchased from LGC Promochem and grown in DMEM: Ham´s-F12 (Sigma-Aldrich) 1:1, with the same supplements as the parental SH-N-SH cell line.

Production of fiber knobs

52SFK and 52LFK were produced as described previously (*25*). HAdV-37 FK (amino acids 172-365), HAdV-5 FK (387-581) were produced in the same way. SAdV-1 SFK (183-363), SAdV-2 SFK (132-312), SAdV-7 SFK (167-347), SAdV-11 SFK (183-364) and SAdV-19 (101-286) were all cloned with the same RGS-hexa-his tag as the fiber knobs described above. SAdV fiber knobs were expressed in *E. coli* (strain Rosetta) and purified with Ni-NTA agarose beads followed by anion exchange (Q-sepharose). Fiber

knob mutants of 52SFK were created using a QuikChange mutagenesis kit (Agilent Technologies). The following mutants were created: 1) K349A, 2) R321Q, 3) R321Q/E348Q, 4) R321I, 5) R321V, 6) R321L, 7) K349A/R321Q and 8) K349R. The mutants were produced as 52SFK, described above. All fiber knobs were analysed by denaturing gel (NuPAGE Bis-Tris, LifeTechnologies) and western blot with monoclonal antibodies directed against the His-tag (Qiagen).

Glycan microarray

Microarrays were composed of lipid-linked oligosaccharide probes robotically printed on nitrocellulose-coated glass slides at 2 and 5 fmol per spot in duplicate using a non-contact instrument as described previously (*40*). Binding signals were probe-dose dependent. The microarray binding assay of the recombinant His-tagged 52SFK was performed at 20°C essentially as described previously (*41*). In brief, the arrayed slide was blocked for 1 h with 5 mM HEPES pH 7.4, 150 mM NaCl, 5mM $CaCl_2$, 0.3% (v/v) Blocker Casein (Pierce), 0.3% (w/v) bovine serum albumin (Sigma) (0.3% casein/0.3% BSA). 52SFK was 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). The 52SFK-antibody pre-complexes were prepared by pre-incubating Ab1 with Ab2 for 15 min at ambient temperature, followed by addition of 52SFK and incubation for an additional 15 min on ice. The VP1-antibody complexes were diluted in 0.3% casein/0.3% BSA, to give a final 52SFK concentration of 150 µg/ml, and overlaid onto the arrays at 20 °C for 2 h. Binding was detected with Alexa Fluor-647-labelled streptavidin (Molecular Probes); imaging and data analysis was performed as described previously (*40*). The oligosaccharide probes are all lipid-linked, neoglycolipids (NGLs) or glycosylceramides and are from the collection assembled in the course of research in the Glycosciences Laboratory. For definition of the lipid moieties of the probes, please see https://glycosciences.med.ic.ac.uk/docs/lipids

ELISA

96-well plates (Nunc maxisorp, Thermo Scientific) were coated with 1 µg/ml of colominic acid (Sigma-Aldrich) for 2 h at room temperature (RT) in coating buffer (100 mM bicarbonate/carbonate, pH 9.6). The plate was blocked with asialofetuin type II (Sigma-Aldrich) 1 mg/ml in PBS-T (phosphate-buffered saline, 140 mM NaCl, 2.7 mM KCl, 10 mM phosphate buffer pH 7.4, supplemented with 0.05% Tween-20) for 1 h at RT and then washed three times with PBS-T. Meanwhile, fiber knobs (10 µg/ml) were incubated with monoclonal anti RGS-His antibodies (Qiagen; dilution 1:1000) in PBS-T for 1 h at RT. The wells were then washed three times with PBS-T and incubated with fiber knob:antibody mixtures for 1 h at RT. After washing, the plate was incubated with a HRP-conjugated rabbit anti-mouse IgG antibody (Dako Cytomation; diluted 1:2000 in PBS-T) for 1 h at RT. The wells were washed again and incubated with 100 µl enhanced K-Blue TMB substrate (Neogen Europe) for 15 min and the reaction was then stopped by addition of 100 µl 1 M H_2SO_4. The absorbance was measured at 450 nm using Tecan infinite F2000 Pro (Tecan Nordic AB).

Flow cytometry

Cells were detached with PBS-EDTA (PBS supplemented with 0.05% EDTA), reactivated in growth medium for one hour at 37˚C, pelleted in 96-well plates ($2x10^5$ cells/well) and washed once with binding buffer (BB: Dulbecco´s modified Eagle medium supplemented with 20 mM HEPES, 20 U/ml penicillin + 20 µg/ml streptomycin and 1% bovine serum albumin). Fiber knobs were added (10 µg/ml in BB) to the cells and incubated for 1h on ice. Unbound fiber knobs were washed away with PF buffer (PBS supplemented with 2% FBS) and the cells were then incubated with an anti RGS-His mouse monoclonal antibody (Qiagen; diluted 1:200 in PF) for 30 min. Followed by one wash with PF, the cells were incubated with polyclonal rabbit-anti-mouse FITC antibodies (Dako Cytomation; diluted 1:40 in PF) for 30 min on ice. Thereafter the cells were washed once with PF and analysed with flow cytometry using

FACSLSRII instrument (Becton Dickinson). Results were analysed using FACSDiva software (Becton Dickinson).

The experiment was performed with the following variations: i) fiber knobs were pre-incubated with different concentrations of oligo [α-2,8]-linked Sia for 1h hour on ice before addition to cells; ii) cells were incubated on ice for 30 min with 4 µg/ml of biotinylated *a) M. Aamurensis* type I or II (MAL-I or II) lectins, b) *S. Nigra* (SNA) lectins, c) wheat germ agglutinin (WGA; all from Vector Laboratories), or d) monoclonal mouse-anti PSia antibody (mab 735, kind gift from Rita Gerardy-Schahn, 1:500) diluted in PF-buffer. Following a-c, the cells were incubated for 30 min on ice with a 1:100 diluted streptavidin-FITC for MAL-I & II, SNA and WGA, or (following d), incubated for 30 min on ice with 1:40 diluted polyclonal rabbit-anti-mouse FITC antibodies.

Crystallization, data collection, and refinement of complex structures

52SFK crystals were prepared as described previously (*25*). PSia complex crystals were generated by soaking in 17,5% (w/v) PEG 1000, 12,5% (v/v) PEG 3350, 12,5% MPD, 100 mM Bicine/Tris pH 8.5 supplemented with 50 mM PSia DP3-5 (GERBU) for 18 to 36 h. The GD3 complex structure was prepared by soaking with 20 mM GD3 (Elicityl) for 1.5 h. No cryoprotection was necessary for crystal freezing. Data collection was done at the X06DA beam line of the Swiss Light Source (Villigen) at a wavelength of 1 or 0.92 Å using a Pilatus 2M detector. Structures were indexed with XDS (*42*) and initial phases were obtained by molecular replacement with Molrep (*43*). The structures were refined using phenix.refine (*44*) and Refmac5 (*45*) from the PHENIX and CCP4 software suites, respectively, using threefold NCS restraints. Figures were prepared with PyMOL (The PyMOL Molecular Graphics System, Version 1.8, Schrödinger, LLC). Data collection and refinement statistics can be found in **Tables S1** and **S2**. Poisson-Boltzmann electrostatics were calculated using the PDB2PQR and APBS plugins in PyMOL (*46, 47*).

Saturation transfer difference NMR

NMR spectra were recorded at 285 K using 3 mm tubes (200 µL sample volume) and a Bruker AVIII-600 spectrometer equipped with a room temperature probe head and processed with TOPSPIN 3.0 (Bruker). Samples contained 1 mM of either oligosialic acid DP3 or DP5 (GERBU) and 20 µM of 52SFK WT or R316A mutant protein (monomeric concentration). The proteins were buffer-exchanged prior to NMR experiments to 20 mM potassium phosphate pH 7.4, 150 mM NaCl in D_2O and the glycans were subsequently added from concentrated stock solutions in D_2O. Off- and on-resonance irradiation frequencies were set to -30.0 ppm and 7.0 ppm, respectively. The irradiation power of the selective pulses was 57 Hz, the saturation time was 2 s, and the total relaxation delay was 3 s. A 50 ms continuous-wave spin-lock pulse with a strength of 3.2 kHz was employed to suppress residual protein signals. A total number of 512 scans and 10,000 points were collected, and spectra were multiplied with a Gaussian window function prior to Fourier transformation.

Surface plasmon resonance

Ni-NTA sensor chips, and HBS-EP+ buffer were purchased from GE Healthcare. 52SFK was diluted in running buffer (HBS-EP+) to a concentration of around 0.03 µM (0,022-0,035 µM) and captured on the Ni-NTA sensor chip according to the manufacturer's instructions, resulting in an immobilization density between 700- 900 RU. In short: an automated program cycle of the following sequence: (1) activation of the sensor chip with Ni(II), (2) capture of 52SFK (3) analyte injection, (4) regeneration of the surface with 0.35 M EDTA, and (5) rinse with HBS-EP+ without EDTA. All steps were performed at a flow rate of 30 µl/min. All binding assays were carried out at 25°C, and HBS-EP+ buffer was used as running buffer. The analytes (oligo-PSia acid DP3-5 and colominic acid) were serially diluted in running buffer to prepare a two-fold concentration series ranging from 0.1 to 8 mM (with small variations depending on the analyte), and then

injected in series over the reference and experimental biosensor surfaces for 120 s and a dissociation time of 120 s. Blank samples containing only running buffer were also injected under the same conditions to allow for double referencing.

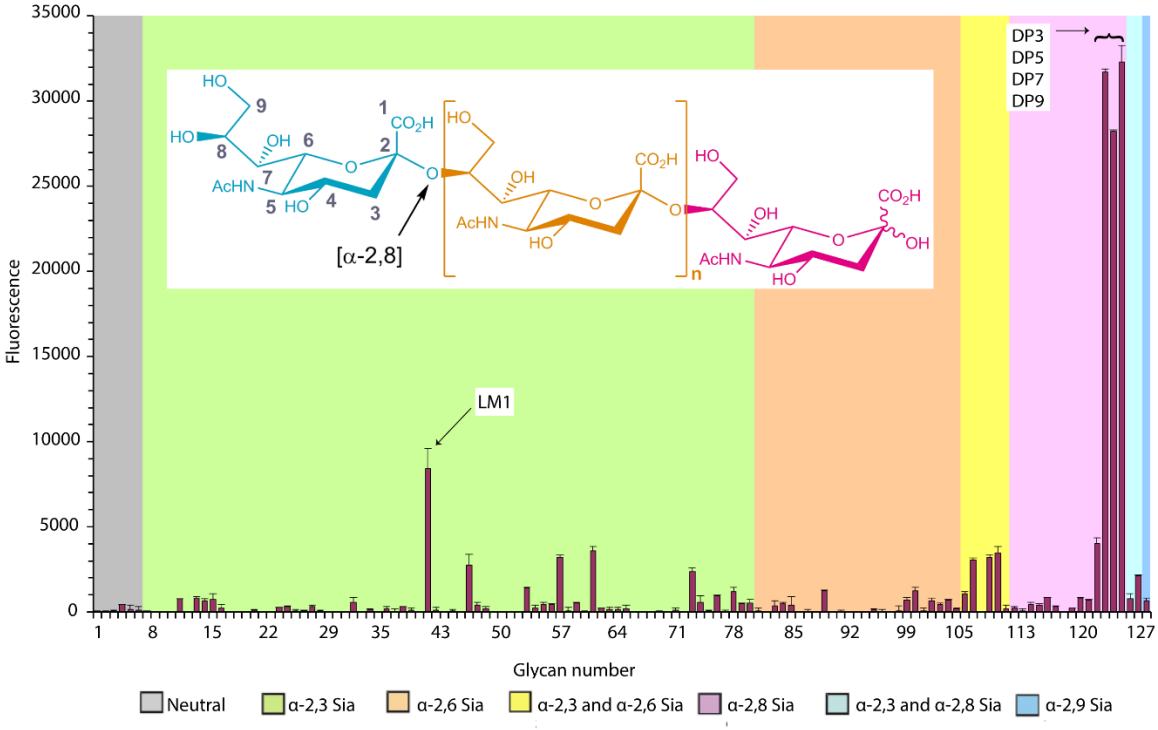


Figure 1. Glycan array of HAdV-52 short fiber knob interactions with sialylated glycans. 52SFK binding to sialylated glycans is shown as fluorescence intensity. Higher peaks represents stronger binding.
LM1= sialyl-3-paragloboside (Neu5NAc-[α-2,3]-Gal-[β-1,4]-GlcNAc-[β-1,3]-Gal-[β-1,4]-Glc), DP3-DP9= [α-2,8]-linked sialic acids with a degree of polymerization (DP) between 3-9. Inlay: general structure of PSia. Up to ~100 sialic acid moieties are linearly connected via an [α-2,8]-linkage. Blue: non-reducing end (nr); Pink: reducing end. PSias can vary considerably in length.

RESULTS AND DISCUSSION

HAdV-52 binds to polysialic acid

We showed previously that HAdV-52 mainly uses sialic acid-containing glycans as cellular receptors for binding and infection of human epithelial cells (*25*). To date, the precise structure of the glycans that can be engaged by the short fiber knob remains unknown. We employed a glycan array to screen for naturally occurring glycans that might serve as HAdV-52 receptor candidates. Analysis of 52SFK binding to 128 different sialylated glycans showed only moderate binding to [α-2,3]- and [α-2,6]-linked sialic acids. However, very strong binding was observed for a group of linear [α-2,8]-linked oligosialic acids, mimicking the naturally occurring PSia (**Fig. 1 & Table S3**). A drastic increase in binding was seen at a degree of polymerization (DP) of more than 3 monosaccharides. To confirm the ability of 52SFK to interact with PSia and evaluate the specificity of this interaction we developed an ELISA with immobilized, *E.coli*-derived PSia (colominic acid; MW=24-38 kDa) and fiber knobs (FKs) from sialic acid-binding HAdV-37 and CAR-binding HAdV-5 and 52LFK. 52SFK bound efficiently to PSia, while CAR-binding FKs did not show any binding at all (**Fig. 2**). 37FK, which binds with relatively high affinity to the branched, disialylated GD1a glycan (*48*), bound less efficiently to PSia than 52SFK. We therefore conclude that HAdV-

52 is able to interact specifically with PSia via the knob domain of its short fiber.

52SFK binds to polysialic acid on neuroblastoma cells

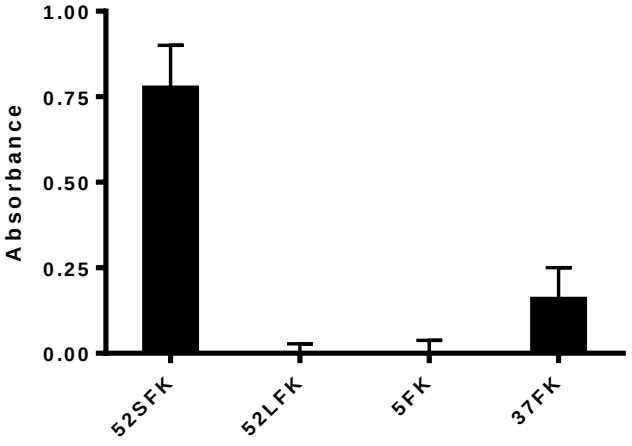


Figure 2. HAdV fiber knob binding to immobilized polysialic acid. ELISA-based experiment with immobilized *E.coli*-derived PSia (colominic acid, MW=24-38 kDa) and HAdV fiber knobs. Relative absorbance is shown. The experiment was performed three times with duplicate samples in each experiment. Error bars represent mean ± SD.

To test the relevance of PSia recognition by HAdV-52 in a cellular context, we used human PSia-expressing neuroblastoma cells, SH-SY5Y, and its PSia-lacking parental cell line SK-N-SH (*49*). The levels of PSia on these cells were confirmed by flow cytometry (**Fig. S1**). 52SFK bound five times better to PSia-expressing SH-SY5Y compared to the control cell line, whereas none of the control knobs, including 37FK, bound with a pronounced difference to either cell line (**Fig. 3A**). In order to exclude the possibility that the increased 52SFK binding to SH-SY5Y was due to a higher level of glycans with terminal monosialic acids on these cells rather than a specific binding to PSia, we used monosialic acid-binding lectins to determine the relative levels of glycans with terminal sialic acids on both cell lines. On the contrary, MAL-I & -II (bind to [α-2,3]-linked sialic acid), SNA (binds to [α-2,6]-linked sialic acid) and WGA (binds to terminal sialic acid regardless of glycosidic bond as well as to N-acetyl-D-glucosamine) all showed higher binding to SK-N-SH cells than SH-SY5Y cells (**Fig. 3B**). Furthermore, pre-incubation of 52SFK with soluble PSia (DP5) reduced 52SFK binding to SH-SY5Y cells with up to 75%, while no effect was observed in the case of 37FK (**Fig. 3C**). From these results we conclude that 52SFK shows a clear preference for PSia-expressing cells over cells lacking PSia. We also conclude that PSia is the major receptor molecule for 52SFK on SH-SY5Y neuroblastoma cells, and that this feature is not shared by sialic acid- or CAR-binding knobs from other HAdVs.

PSia is engaged at the non-reducing end, similarly to mono- and di-sialylated glycans

We previously identified a sialic acid binding site on the upper side of 52SFK. This site is physically distinct from the binding site of 37FK, which engages sialic acid within its central cavity. Using 2-O-methyl-sialic acid as a ligand, we found the major determinant of the 52SFK binding site to be a stretch of three adjacent residues that together formed a prominent RGN motif (*25*). In a PSia context, several questions regarding the binding mode remain, e.g. whether the glycan is engaged via its non-reducing end or via a more promiscuous internal binding mode, or how the increased binding potency arises on a structural level. To address these questions, we solved the complex structures of 52SFK with three oligosialic acids (DP3-5) as well as the GD3 glycan (Neu5NAc-[α-2,8]-Neu5NAc-[α-2,3]-Gal-[β-1,4]-Glc, representing a disialic acid motif). All complex structures produced highly similar results, shown exemplary for DP3 in **Fig. 4**. In all cases, unambiguous electron density for a sialic acid moiety was found in the canonical binding pocket. The electron density around O8 and its direction relative to the protein clearly indicate that it is the non-reducing end of the glycans that is engaged at this position, and the observed binding mode is highly similar to what has been observed for monosialic acid. In all cases except GD3, there was additional electron density for a second sialic acid

moiety projecting from the pocket towards the solvent. The overall density for this moiety was much weaker, deteriorating from the glycerol group to the pyranose ring and thereby indicating an increased flexibility. All structures showed similar angles for the [α-2,8]-glycosidic linkage, which agree with what has been found in other crystal structures (**Fig. S3**) (*50*). Interestingly, the second sialic acid moiety did not seem to contribute any directed contacts to the overall interaction, except for a van-der-Waals contact between its N-Acetyl group and E328. This contact seemed to cause a local decrease of electron density and a slight rotation of the N-Acetyl group. The third sialic acid could not be unambiguously traced in any of the structures. In order to verify this finding in solution, we employed saturation transfer difference NMR (STD-NMR) to screen for glycan protons of DP3 and 5 that are placed within 5-6 Å of the protein in solution (shown exemplary for DP3 in **Fig. 4B** and **C**). The spectrum of the glycan alone compared well to the literature (*51, 52*). Since all of the sialic acid repeats were in a highly similar chemical environment in solution, the respective peaks overlap - with the exception of the non-reducing end, which experienced an upfield shift. The experiment showed a saturation transfer occurring almost exclusively at the non-reducing end, while the other moieties only received a very moderate spin saturation occurring exclusively in the N-Acetyl group region, putatively from the second sialic acid. In the case of the R316A mutant (which destroys the canonical RGN motif and prevents 52SFK attachment to sialic acid on A549 cells) (*25*), the saturation transfer was completely abrogated. Together, these results demonstrate that 52SFK engages PSia exclusively via its canonical sialic acid binding site, without any additional epitopes present on the protein.

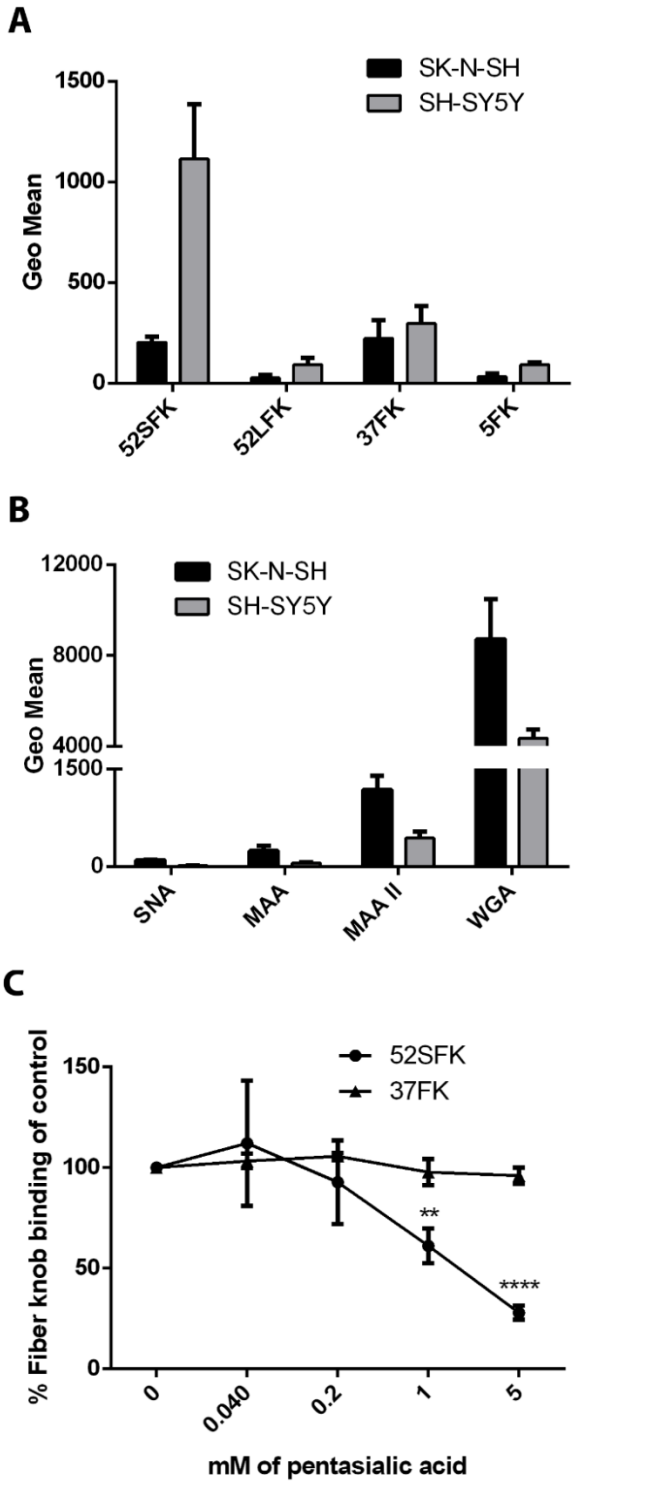


Figure 3. HAdV-52 short fiber knob binds strongly to polysialic acid on human cells. A) Flow cytometry-based quantification of HAdV fiber knob binding to human neuroblastoma cells expressing (SH-SY5Y) or lacking (SK-N-SH) PSia. B) Flow cytometry-based quantification of sialic acid on SK-N-SH and SH-SY5Y cells using sialic acid-binding

Long-range effects are major determinants of 52SFK:PSia interactions

Despite the lack of directed contacts, a length of more than DP3 appeared to be beneficial for oligosialic acid binding as seen in our glycan array data **(Fig. 1)**. According to the cell attachment inhibition data, a DP of more than 2 was already sufficient to substantially decrease 52SFK binding (**Fig. 5A**). Similar results were acquired from SPR experiments, where the biggest increase in affinity was shown between DP2 and 3

lectins. C) Flow cytometry-based quantification of 52SFK and 37FK binding to SH-SY5Y cells after fiber knob pre-incubation with increasing concentrations of pentasialic acid. All experiments were performed three times with duplicate samples in each experiment. Error bars represent mean ± SD.

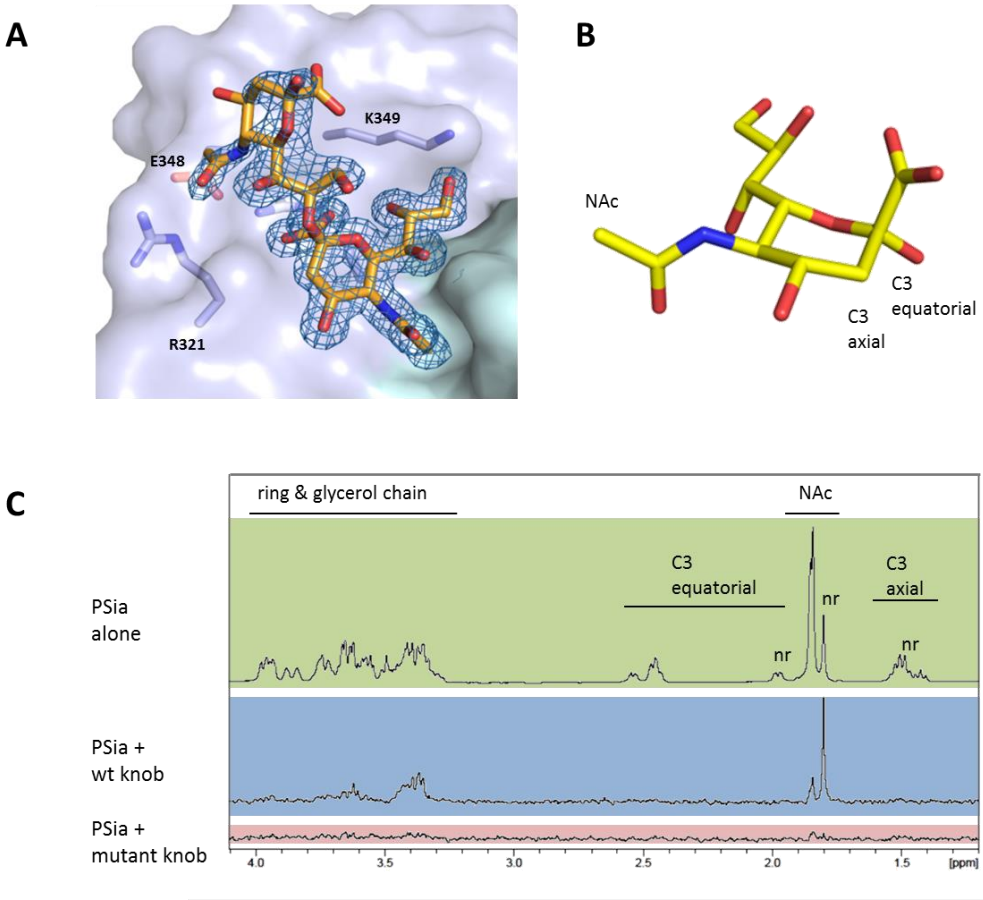


Figure 4. [α-2,8]-linked oligosialic acids are engaged in the canonical binding pocket of HAdV-52 short fiber knob via their reducing end. A) Complex structure of 52SFK and DP3. Shown is a $2F_o$-F_c map calculated at 1.1 σ after refinement. The third sialic acid moiety could not be resolved. B) Schematic representation of sialic acid in the α-conformation. The positions of distinctive protons for NMR are indicated. C) STD-NMR of 52SFK and DP3. Green box: PSia alone; blue box: Saturation transfer difference spectrum of the 52SFK:DP3 complex; red box: Saturation transfer difference spectrum of the R316A-52SFK:DP3 complex, nr= non-reducing end.

(**Fig. 5B-F**). In combination with the structural data, these findings suggest that long-range effects account for increased binding affinity rather than directed short-range contacts. We hypothesize that these effects could be electrostatic interactions or entropy-related contributions.

According to in-solution NMR studies, the poly-anionic PSia seemed to at least transiently adopt a

left-handed helical conformation, with the negatively charged carboxyl groups facing towards the non-reducing end (*53*). However, PSia is expected to be rather flexible in solution due to the lack of branching and the fact that the individual sialic acid moieties are

linked via an [α-2,8]-linkage, which provides a much larger conformational freedom than more common glycosidic linkages. Recent molecular dynamics (MD) simulations showed that helical structures can be observed within a timeframe of 1 µs. Even though these are not stable (M. Frank, data not shown),it is tempting to speculate that such an arrangement might lead to an electrostatic momentum that is beneficial for the binding. In line with this, an inspection of the electrostatic potential of the 52SFK revealed a positively charged rim located around the sialic acid binding site, which we coined the 'steering rim'. The rim is mainly formed by residues R321, R316, and K349

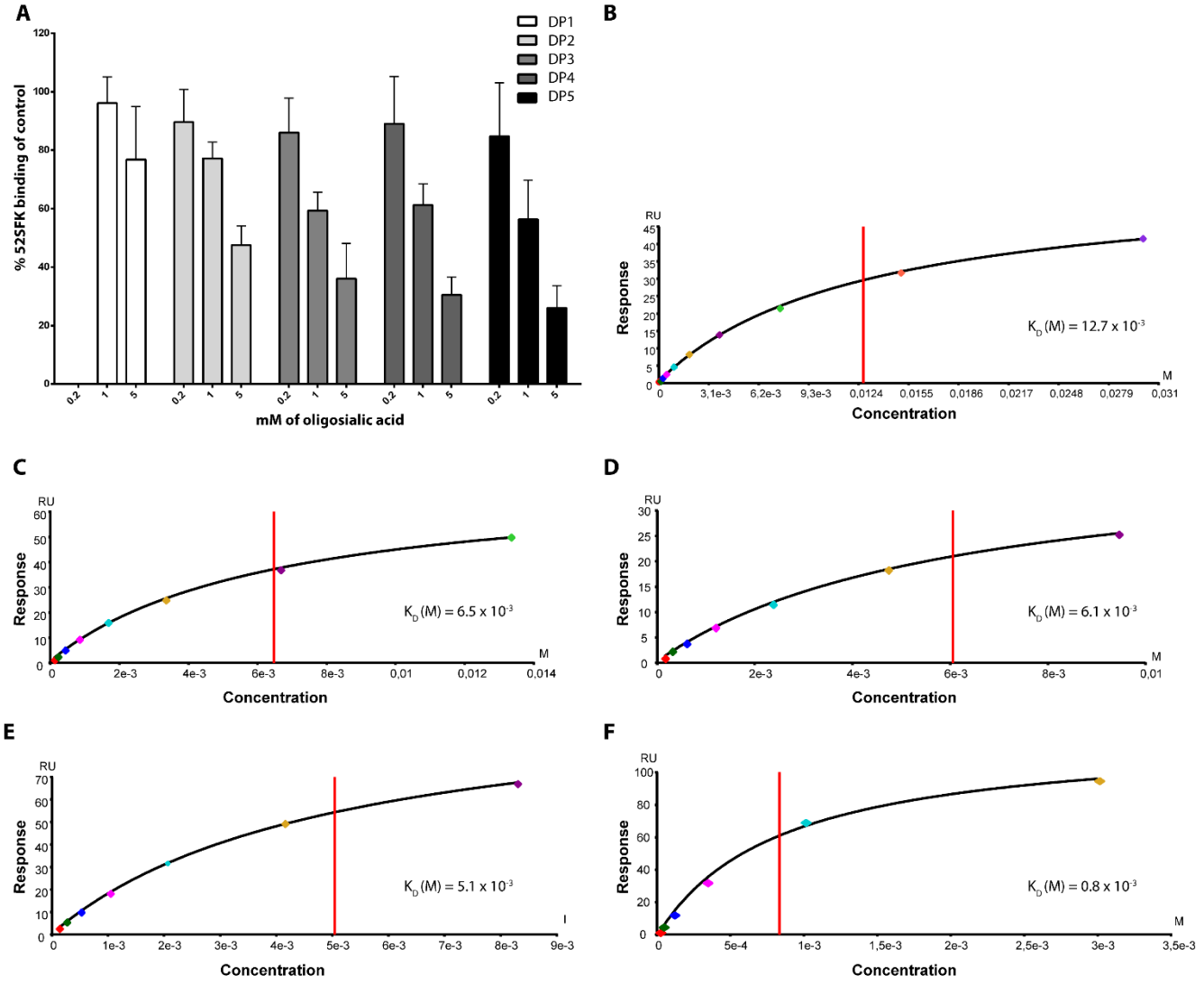


Figure 5. A degree of polymerization greater than 2 is sufficient for interactions with 52FSK. A) Flow cytometry-based quantification of 52SFK binding to A549 cells after fiber knob pre-incubation with increasing concentrations of oligosialic acid. The experiments were performed three times with duplicate samples in each experiment. Error bars represent mean ± SD. Surface plasmon resonance analysis of 52SFK (immobilized) binding to B) disialic acid, DP2, C) trisialic acid, DP3, D) tetrasialic acid, DP4, E) pentasialic acid, DP5, and F) *E.coli*-derived PSia, DP≈80-100 (in solution).

(**Fig. 6**). In the DP3 complex structure, the second sialic acid moiety is placed over the ε-amino group of K349. If the PSia glycan roughly follows the proposed left-handed helical arrangement with energy-minimal glycosidic angles similar to those observed between the first two moieties (**Fig. S2A,B**), the third or fourth moiety are placed above R321 (indicated in **Fig. 6D**), with all carboxyl groups facing towards the protein surface. The non-directional nature of electrostatic interactions might allow for a certain degree of flexibility in this arrangement. Preliminary MD simulations with 52SFK and DP5 confirmed these assumptions and indicate that the second sialic acid is indeed constantly placed over K349, while the third and fourth moiety are more mobile, but transiently under the influence of R321 in a helix-like arrangement (M. Frank, data not shown).

To further test this hypothesis, we produced fiber

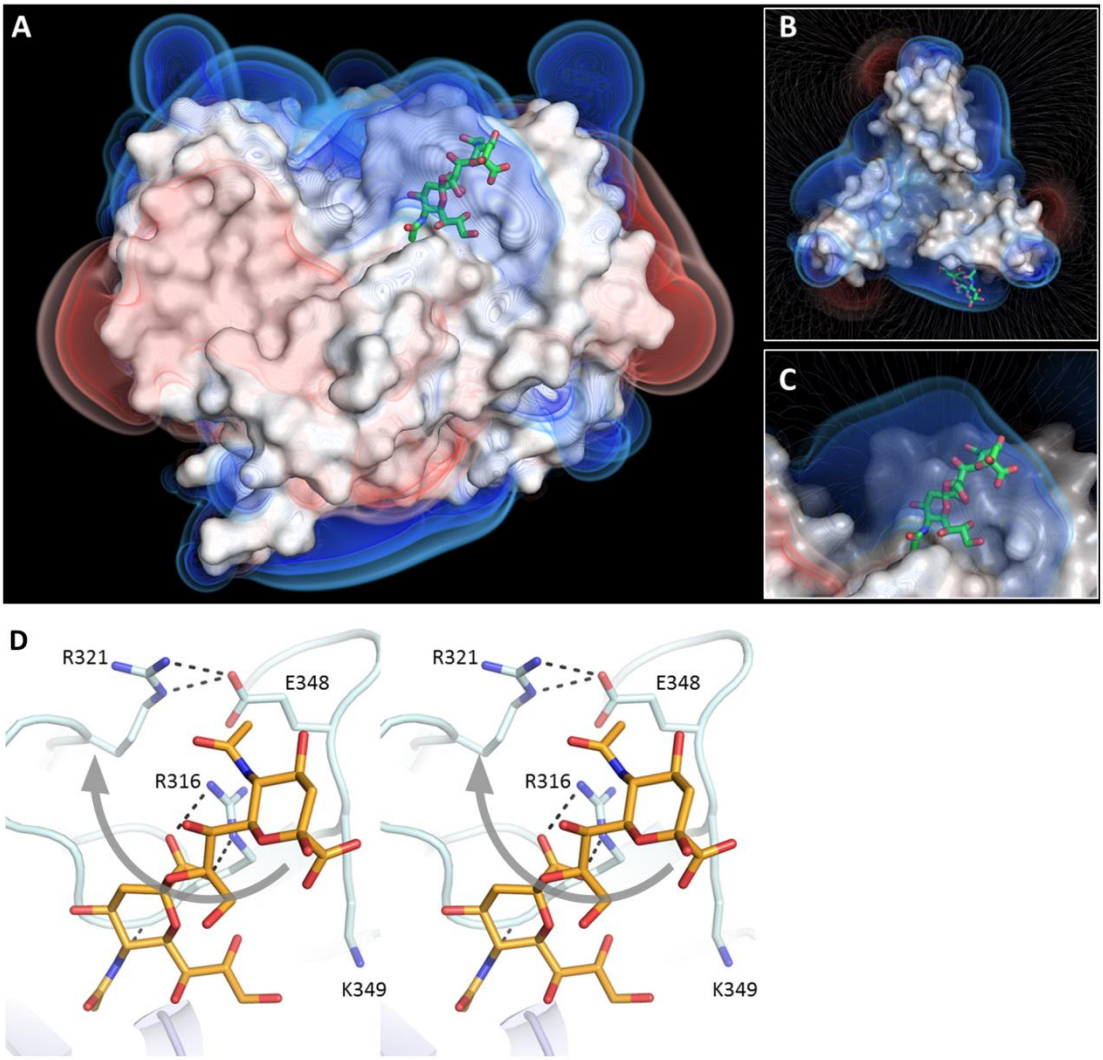


Figure 6. Representation of the HAdV-52 short fiber knob 'steering rim'. Poisson-Boltzmann electrostatic potential isosurfaces and field lines for the protein were calculated at ±1; ±0.75; ±0.5 kT/e. The positively charged rim can be seen in blue. Bound DP3 is shown as sticks. A) Side view B) Top view including field lines C) Detailed view of the binding pocket including field lines. D) Stereo figure showing the relative placement of glycan and 'steering rim' residues. Residues influencing the 'steering rim' are highlighted as sticks. R321 and E348 are forming a salt bridge, as do R316 and the carboxyl group at the non-reducing end of DP3. The expected rough orientation of the glycan is indicated as a grey arrow.

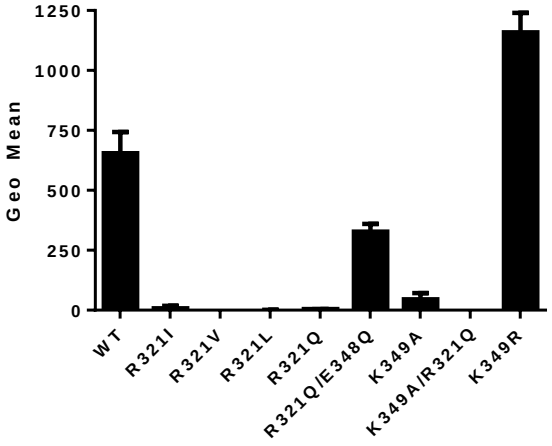


Figure 7. HAdV-52 short fiber knob mutant binding to polysialic acid-expressing SH-SY5Y cells. Flow cytometry-based analysis of wild type and mutant 52SKF binding to SH-SY5Ycells. The experiment was performed three times with duplicate samples in each experiment. Error bars represent mean ± SD.

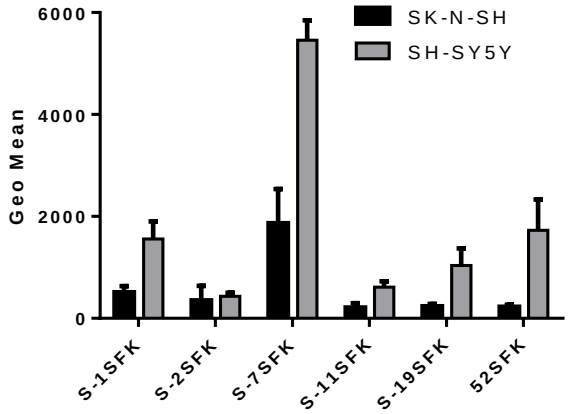


Figure 8. HAdV short fiber knob binding to polysialic acid-expressing/-lacking cells. Flow cytometry-based quantification of simian (S) and human short fiber knobs-binding to human neuroblastoma cells expressing (SH-SY5Y) or lacking (SK-N-SH) PSia. The experiment was performed three times with duplicate samples in each experiment. Error bars represent mean ± SD.

knobs with mutations in the 'steering rim', and analyzed knob binding to PSia-expressing SH-SY5Y cells. The K349A mutant almost completely lost its cell binding capacity, and similar effects were observed for the R321Q and analogous mutants (**Fig. 7**). When mutated, R321 could no longer counteract the closely situated E348, which then would likely repel the polyanionic PSia, and might thus contribute to an unexpectedly strong loss in binding. Indeed, if E348 was also mutated to a non-charged residue, the effect of the R321Q mutation was partially reversed (**Fig. 7**). This implies that R321 interacts more weakly with PSia than the other two residues, which fits well with the assumption of a flexible 'pseudo-helical' arrangement. In addition to interacting with the more distant sialic acid moieties of PSia, both R321 and K349 might have a supportive direct effect on the non-reducing end.

The PSia binding site and the 'steering rim' are conserved in closely related simian adenoviruses

The PSia-binding RGN motif is conserved in the short fibers of other closely related members of species HAdV-G: SAdV-1, -2, -7 and -11, as well as SAdV-19 (SAdV-C), but it is not found in any other known non-human or human AdVs, including the short fiber knobs of HAdV-40 and -41 (**Fig. S3A**). Interestingly, the three positively charged residues forming the 'steering rim' are also functionally conserved in these SAdV types, but in different permutations (RRK, RKK, RRR) **(Fig. S3A)**. No other HAdV fiber knobs with known structures exhibits a comparable 'steering rim' (**Fig. S3C**). This further supports our hypothesis that PSia binding is a specific ability limited to a small subset of AdVs. We assayed the PSia specificity and binding capacity of fiber knobs belonging to this subset in a cell attachment assay with cells expressing or lacking PSia. All knobs except S-2SFK bound better to PSia-expressing cells than the control cell line (**Fig. 8**). One possible explanation for the inability of S-2SFK to bind PSia, despite a conserved 'steering rim', could be that S-2SFK harbor a fiber sequence more distantly related to 52SFK (**Fig. S3B**) which, in theory,

could result in a different overall arrangement of the residues. However, 52SFK still displayed the strongest discrepancy between SH-SY5Y and SK-N-SH cells, indicating a more specific interaction of 52SFK with PSia rather than a general high binding to both cell lines as seen for S-7SFK.

In humans, PSia is present in high amounts on NCAM during embryogenesis but almost absent in healthy adults. It is currently unclear why a subset of human and simian adenoviruses might have developed a specificity for a glycan which is rarely present in their main host. However, such a finding is not unprecedented and may confer a currently unrecognized evolutionary advantage. As an example, human polyomavirus 9 and the Shiga-toxigenic *E.coli* subtilase cytotoxin have evolved a preference for 5-N-glycolylsialic acid, despite the fact that this sugar is only incorporated into human cells through nutritional uptake (*54, 55*). Secondly, there is increasing evidence for an interplay between commensal bacteria and viruses (e.g. noro-, reo- and polioviruses) to mount infections in the gut (*56, 57*), which seems especially interesting given the observed gastroenteric tropism of HAdV-52 and the presence of PSia on commensal bacteria.

Although PSia is close to absent on cells in healthy adults it is often overexpressed on many tumors, where it modulates cell adhesion, migration and invasion (*31-37*). PSia expression is therefore often associated with higher tumor aggressiveness and invasiveness resulting in poor clinical prognosis (*29, 33, 38*). These types of cancers are also often recurrent and non-responsive to conventional treatments (*33*). Large attention has therefore been drawn to novel therapeutic approaches, including AdV vectors for gene delivery and use of modified oncolytic AdVs. HAdV-52 is a naturally occurring virus type that combines low seroprevalence rates, reduced liver tropism, and a tumor-specific natural receptor profile. In light of the many drawbacks of HAdV-5 based vectors and the ability of HAdV-52 to specifically interact with a receptor selectively expressed in high amounts on cancer cells, we believe that HAdV-52 is an excellent choice as an oncolytic vector for treatment of cancer with elevated PSia-expression.

REFERENCES

1. F. C. Zhu *et al.*, Safety and immunogenicity of a novel recombinant adenovirus type-5 vector-based Ebola vaccine in healthy adults in China: preliminary report of a randomised, double-blind, placebo-controlled, phase 1 trial. *Lancet* **385**, 2272 (Jun 6, 2015).
2. C. J. Wei *et al.*, Induction of broadly neutralizing H1N1 influenza antibodies by vaccination. *Science* **329**, 1060 (Aug 27, 2010).
3. D. H. Barouch *et al.*, Vaccine protection against acquisition of neutralization-resistant SIV challenges in rhesus monkeys. *Nature* **482**, 89 (Feb 2, 2012).
4. S. J. George *et al.*, Sustained reduction of vein graft neointima formation by ex vivo TIMP-3 gene therapy. *Circulation* **124**, S135 (Sep 13, 2011).
5. I. Ganly *et al.*, A phase I study of Onyx-015, an E1B attenuated adenovirus, administered intratumorally to patients with recurrent head and neck cancer. *Clin Cancer Res* **6**, 798 (Mar, 2000).
6. C. Heise *et al.*, An adenovirus E1A mutant that demonstrates potent and selective systemic anti-tumoral efficacy. *Nat Med* **6**, 1134 (Oct, 2000).
7. M. G. Castro *et al.*, Adenoviral vector-mediated gene therapy for gliomas: coming of age. *Expert Opin Biol Ther* **14**, 1241 (Sep, 2014).
8. F. L. Moolten, Tumor chemosensitivity conferred by inserted herpes thymidine kinase genes: paradigm for a prospective cancer control strategy. *Cancer Res* **46**, 5276 (Oct, 1986).
9. C. J. Springer, I. Niculescu-Duvaz, Prodrug-activating systems in suicide gene therapy. *J Clin Invest* **105**, 1161 (May, 2000).
10. I. Liikanen *et al.*, Oncolytic adenovirus with temozolomide induces autophagy and antitumor immune responses in

cancer patients. *Mol Ther* **21**, 1212 (Jun, 2013).
11. A. Koski *et al.*, Treatment of cancer patients with a serotype 5/3 chimeric oncolytic adenovirus expressing GMCSF. *Mol Ther* **18**, 1874 (Oct, 2010).
12. K. Valerie *et al.*, Improved radiosensitization of rat glioma cells with adenovirus-expressed mutant herpes simplex virus-thymidine kinase in combination with acyclovir. *Cancer Gene Ther* **7**, 879 (Jun, 2000).
13. J. L. Fisher, J. A. Schwartzbaum, M. Wrensch, J. L. Wiemels, Epidemiology of brain tumors. *Neurol Clin* **25**, 867 (Nov, 2007).
14. A. M. Maatta, H. Samaranayake, J. Pikkarainen, T. Wirth, S. Yla-Herttuala, Adenovirus mediated herpes simplex virus-thymidine kinase/ganciclovir gene therapy for resectable malignant glioma. *Curr Gene Ther* **9**, 356 (Oct, 2009).
15. D. E. Zak *et al.*, Merck Ad5/HIV induces broad innate immune activation that predicts CD8(+) T-cell responses but is attenuated by preexisting Ad5 immunity. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E3503 (Dec 11, 2012).
16. Y. Li *et al.*, Loss of adenoviral receptor expression in human bladder cancer cells: a potential impact on the efficacy of gene therapy. *Cancer Res* **59**, 325 (Jan 15, 1999).
17. S. N. Waddington *et al.*, Adenovirus serotype 5 hexon mediates liver gene transfer. *Cell* **132**, 397 (Feb 8, 2008).
18. V. N. Krasnykh, G. V. Mikheeva, J. T. Douglas, D. T. Curiel, Generation of recombinant adenovirus vectors with modified fibers for altering viral tropism. *J Virol* **70**, 6839 (Oct, 1996).
19. S. Lecollinet *et al.*, Improved gene delivery to intestinal mucosa by adenoviral vectors bearing subgroup B and d fibers. *J Virol* **80**, 2747 (Mar, 2006).
20. I. V. Ulasov *et al.*, Tamoxifen improves cytopathic effect of oncolytic adenovirus in primary glioblastoma cells mediated through autophagy. *Oncotarget* **6**, 3977 (Feb 28, 2015).
21. J. J. Short *et al.*, Substitution of adenovirus serotype 3 hexon onto a serotype 5 oncolytic adenovirus reduces factor X binding, decreases liver tropism, and improves antitumor efficacy. *Mol Cancer Ther* **9**, 2536 (Sep, 2010).
22. M. S. Jones, 2nd *et al.*, New adenovirus species found in a patient presenting with gastroenteritis. *Journal of virology* **81**, 5978 (Jun, 2007).
23. A. H. Kidd, J. Chroboczek, S. Cusack, R. W. Ruigrok, Adenovirus type 40 virions contain two distinct fibers. *Virology* **192**, 73 (Jan, 1993).
24. H. Y. Yeh, N. Pieniazek, D. Pieniazek, H. Gelderblom, R. B. Luftig, Human adenovirus type 41 contains two fibers. *Virus Res* **33**, 179 (Aug, 1994).
25. A. Lenman *et al.*, Human adenovirus 52 uses sialic acid-containing glycoproteins and the coxsackie and adenovirus receptor for binding to target cells. *PLoS Pathog* **11**, e1004657 (Feb, 2015).
26. K. Banyai *et al.*, Searching for HAdV-52, the putative gastroenteritis-associated human adenovirus serotype in Southern Hungary. *The new microbiologica* **32**, 185 (Apr, 2009).
27. K. Angata, M. Fukuda, Polysialyltransferases: major players in polysialic acid synthesis on the neural cell adhesion molecule. *Biochimie* **85**, 195 (Jan-Feb, 2003).
28. D. Scholl, S. Adhya, C. Merril, Escherichia coli K1's capsule is a barrier to bacteriophage T7. *Appl Environ Microbiol* **71**, 4872 (Aug, 2005).
29. R. A. Falconer, R. J. Errington, S. D. Shnyder, P. J. Smith, L. H. Patterson, Polysialyltransferase: A New Target in Metastatic Cancer. *Current Cancer Drug Targets* **12**, 925 (2012).
30. S. Curreli, Z. Arany, R. Gerardy-Schahn, D. Mann, N. M. Stamatos, Polysialylated neuropilin-2 is expressed on the surface of

human dendritic cells and modulates dendritic cell-T lymphocyte interactions. *The Journal of biological chemistry* **282**, 30346 (Oct 19, 2007).
31. M. C. Amoureux *et al.*, Polysialic acid neural cell adhesion molecule (PSA-NCAM) is an adverse prognosis factor in glioblastoma, and regulates olig2 expression in glioma cell lines. *BMC cancer* **10**, 91 (2010).
32. A. K. Petridis, H. Wedderkopp, H. H. Hugo, H. Maximilian Mehdorn, Polysialic acid overexpression in malignant astrocytomas. *Acta neurochirurgica* **151**, 601 (Jun, 2009).
33. M. Suzuki *et al.*, Polysialic acid facilitates tumor invasion by glioma cells. *Glycobiology* **15**, 887 (Sep, 2005).
34. S. Gluer, C. Schelp, R. Gerardy-Schahn, D. von Schweinitz, Polysialylated neural cell adhesion molecule as a marker for differential diagnosis in pediatric tumors. *Journal of pediatric surgery* **33**, 1516 (Oct, 1998).
35. D. F. Figarella-Branger, P. L. Durbec, G. N. Rougon, Differential spectrum of expression of neural cell adhesion molecule isoforms and L1 adhesion molecules on human neuroectodermal tumors. *Cancer research* **50**, 6364 (Oct 1, 1990).
36. F. Tanaka *et al.*, Expression of polysialic acid and STX, a human polysialyltransferase, is correlated with tumor progression in non-small cell lung cancer. *Cancer research* **60**, 3072 (Jun 1, 2000).
37. S. Lantuejoul, D. Moro, R. J. Michalides, C. Brambilla, E. Brambilla, Neural cell adhesion molecules (NCAM) and NCAM-PSA expression in neuroendocrine lung tumors. *The American journal of surgical pathology* **22**, 1267 (Oct, 1998).
38. F. Tanaka *et al.*, Prognostic significance of polysialic acid expression in resected non-small cell lung cancer. *Cancer research* **61**, 1666 (Feb 15, 2001).
39. A. Kloos, N. Woller, R. Gerardy-Schahn, F. Kuhnel, Retargeted oncolytic viruses provoke tumor-directed T-cell responses. *Oncoimmunology* **4**, e1052933 (Dec, 2015).
40. Y. Liu *et al.*, Neoglycolipid-based oligosaccharide microarray system: preparation of NGLs and their noncovalent immobilization on nitrocellulose-coated glass slides for microarray analyses. *Methods Mol Biol.* **808**, 117 (2012).
41. U. Neu *et al.*, Structures of B-Lymphotropic Polyomavirus VP1 in complex with oligosaccharide ligands. *PLoS Pathog.* **9**, e1003714 (2013).
42. W. Kabsch, Xds. *Acta crystallographica. Section D, Biological crystallography* **66**, 125 (Feb, 2010).
43. A. Vagin, A. Teplyakov, MOLREP: an automated program for molecular replacement. *J Appl Crystallogr* **30**, 1022 (Dec 1, 1997).
44. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D* **66**, 213 (Feb, 2010).
45. G. N. Murshudov, A. A. Vagin, E. J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* **53**, 240 (May 1, 1997).
46. N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon, Electrostatics of nanosystems: Application to microtubules and the ribosome. *P Natl Acad Sci USA* **98**, 10037 (Aug 28, 2001).
47. T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, N. A. Baker, PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Research* **32**, W665 (Jul 1, 2004).
48. E. C. Nilsson *et al.*, The GD1a glycan is a cellular receptor for adenoviruses causing epidemic keratoconjunctivitis. *Nature medicine* **17**, 105 (Jan, 2011).

49. U. Valentiner, M. Muhlenhoff, U. Lehmann, H. Hildebrandt, U. Schumacher, Expression of the neural cell adhesion molecule and polysialic acid in human neuroblastoma cell lines. *Int J Oncol* **39**, 417 (Aug, 2011).
50. M. H. Buch *et al.*, Structural and Functional Analysis of Murine Polyomavirus Capsid Proteins Establish the Determinants of Ligand Recognition and Pathogenicity. *PLoS Pathog* **11**, e1005104 (Oct, 2015).
51. G. J. Ray *et al.*, Complete structural elucidation of an oxidized polysialic acid drug intermediate by nuclear magnetic resonance spectroscopy. *Bioconjug Chem* **25**, 665 (Apr 16, 2014).
52. J. R. Brisson, H. Baumann, A. Imberty, S. Perez, H. J. Jennings, Helical epitope of the group B meningococcal alpha(2-8)-linked sialic acid polysaccharide. *Biochemistry* **31**, 4996 (Jun 2, 1992).
53. M. D. Battistel, M. Shangold, L. Trinh, J. Shiloach, D. I. Freedberg, Evidence for helical structure in a tetramer of alpha2-8 sialic acid: unveiling a structural antigen. *J Am Chem Soc* **134**, 10717 (Jul 4, 2012).
54. Z. M. Khan *et al.*, Crystallographic and glycan microarray analysis of human polyomavirus 9 VP1 identifies N-glycolyl neuraminic acid as a receptor candidate. *J Virol* **88**, 6100 (Jun, 2014).
55. E. Byres *et al.*, Incorporation of a non-human glycan mediates human susceptibility to a bacterial toxin. *Nature* **456**, 648 (Dec 4, 2008).
56. M. K. Jones *et al.*, Enteric bacteria promote human and mouse norovirus infection of B cells. *Science* **346**, 755 (Nov 7, 2014).
57. S. K. Kuss *et al.*, Intestinal microbiota promote enteric virus replication and systemic pathogenesis. *Science* **334**, 249 (Oct 14, 2011).
58. S. V. Vasudevan, P. V. Balaji, Molecular dynamics simulations of alpha2 --> 8-linked disialoside: conformational analysis and implications for binding to proteins. *Biopolymers* **63**, 168 (Mar, 2002).
59. E. Seiradake, S. Cusack, Crystal structure of enteric adenovirus serotype 41 short fiber head. *J Virol* **79**, 14088 (Nov, 2005).

SUPPORTING INFORMATION

Table S1. Data collection statistics of 52SFK complex structures.

	GD3	PSia DP3	PSia DP4	Psia DP5
Beamline	X06DA (PXIII)	X06DA (PXIII)	X06DA (PXIII)	X06DA (PXIII)
Detector	Pilatus 2M	Pilatus 2M	Pilatus 2M	Pilatus 2M
Angle increment [°]	0.1	0.1	0.1	0.1
Total rotation angle [°]	360	360	360	360
Wavelength [Å]	1.00	1.00	0.92	0.92
Data collection				
Resolution [Å]	50.0 - 1.17 (1.25 - 1.17)	50 - 1.50 (1.59 – 1.50)	50 - 1.35 (1.43 – 1.35)	50 – 1.48 (1.57 – 1.48)
Completeness [%]	99 (88)	97.7 (94.7)	99.5 (97.0)	99.5 (97.1)
Observed	2,103,614 (289,758)	1,041,202 (150,022)	1,412,870 (216,260)	1,078,654 (154,372)
Redundancy	13.3 (12.3)	13.2 (12.3)	13.1 (12.9)	13.1 (12.1)
CC1/2* [%]	100 (75.3)	100.0 (66.8)	100.0 (72.3)	100.0 (70.1)
I/σI	22.7 (2.4)	21.3 (2.0)	23.6 (2.2)	23.5 (2.2)
Wilson B-Factor [Å^2]	17.0	16.9	14.8	15.8
FreeR [% of reflections]	5	5	5	5
Crystal properties				
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell axes [Å]	a = 63.88	a=64.80	a=63.89	a=63.92
	b=82.03	b=81.80	b=93.16	b=81.70
	c=93.31	c=93.40	c=81.77	c=93.39
Unit cell angles [°]	α=β=γ=90	α=β=γ=90	α=β=γ=90	α=β=γ=90

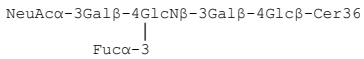
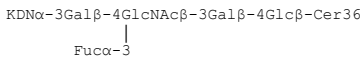
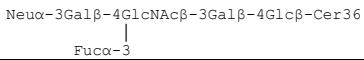
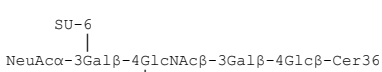
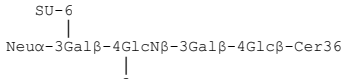
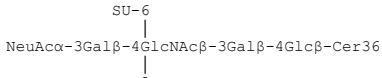
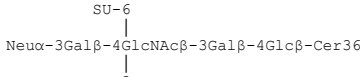
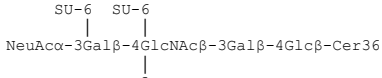
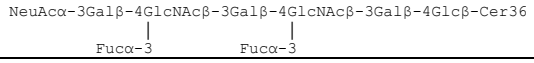
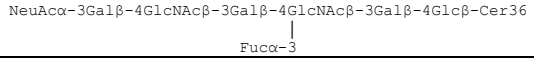
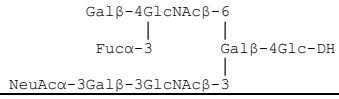
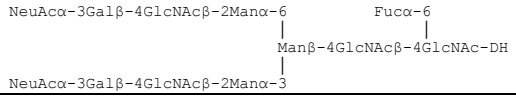
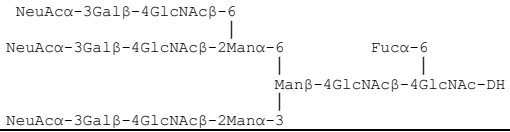
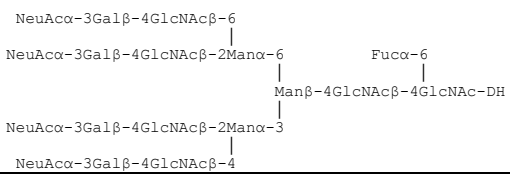
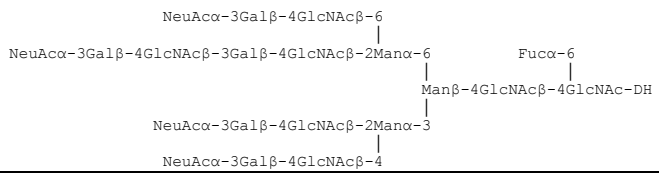
Table S2. Refinement statistics of 52SFK complex structures

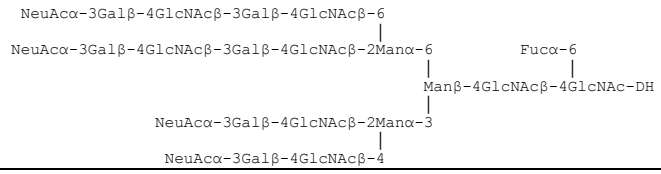
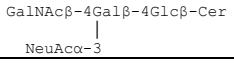
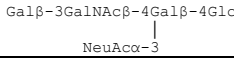
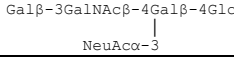
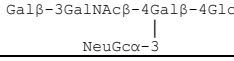
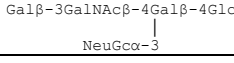
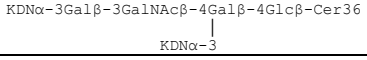
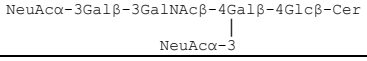
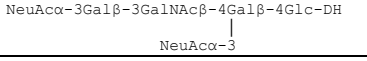
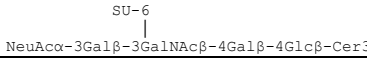
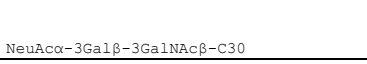
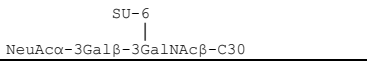
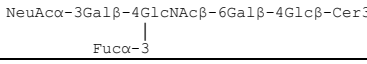
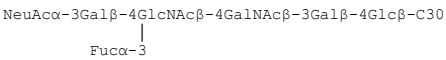
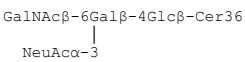
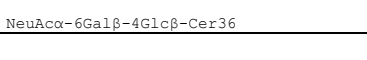
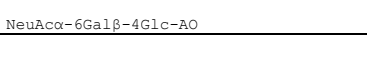
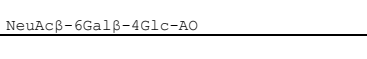
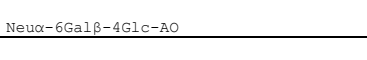
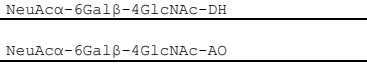
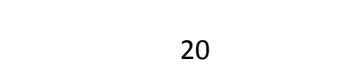
		GD3	PSia DP3	PSia DP4	Psia DP5
Rwork [%]		11.95	16.56	16.65	15.06
Rfree [%]		14.06	18.25	18.81	17.76
Rmsd bond [Å]		0.016	0.009	0.011	0.015
Rmsd angles [°]		1.729	1.353	1.426	1.624
B-Factor [Å^2]	overall	16.87	20.8	17.8	21.2
	protein	15.47	19.9	18.8	17.8
	solvent	29.80	26.0	24.6	30.3
	ligand	25.31	38.3	31.4	34.8
No of atoms					
protein		4077	3811	3817	3927
ligand		21	41	41	40
solvent		511	237	263	324
Ramachandran favored (%)		98.1	98.1	98.1	97.7
Ramachandran allowed (%)		1.9	1.9	1.9	2.1
Ramachandran outliers (%)		0	0	0	0.2

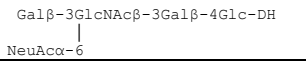
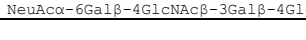
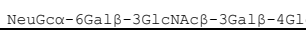
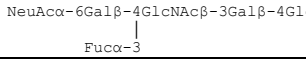
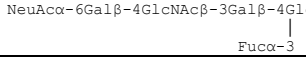
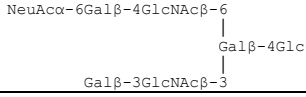
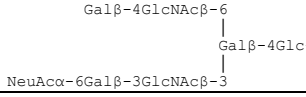
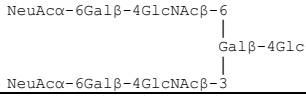
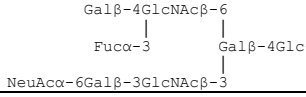
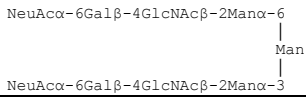
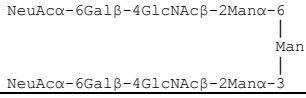
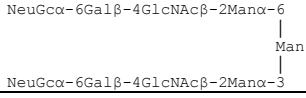
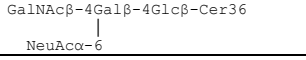
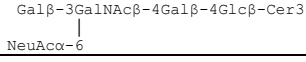
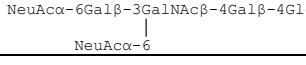
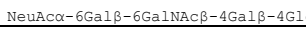
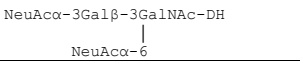
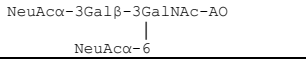
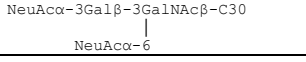
Table S3. List of sialylated glycans included in microarray screening analysis.

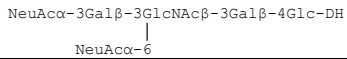
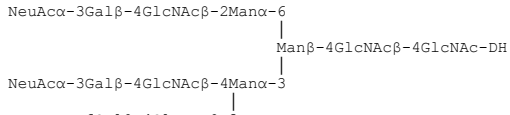
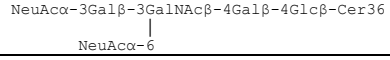
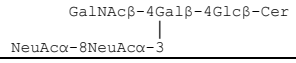
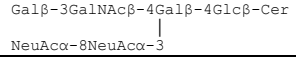
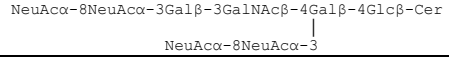
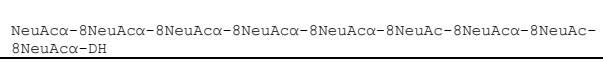
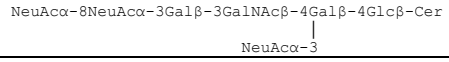
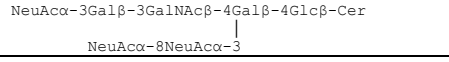
Position	Probe [a]	Sequence	Fluorescence intensities [b,c]
1	Lac-AO	Galβ-4Glc-AO	27
2	LacNAc-AO	Galβ-4GlcNAc-AO	-
3	LNT	Galβ-3GlcNAcβ-3Galβ-4Glc-DH	10
4	LNnT	Galβ-4GlcNAcβ-3Galβ-4Glc-DH	430
5	LNFP-III	Galβ-4GlcNAcβ-3Galβ-4Glc-DH | Fucα-3	154
6	NA2	Galβ-4GlcNAcβ-2Manα-6 | Manβ-4GlcNAcβ-4GlcNAc-DH | Galβ-4GlcNAcβ-2Manα-3	113
7	GM4	NeuAcα-3Galβ-Cer	-
8	GM3	NeuAcα-3Galβ-4Glcβ-Cer	-
9	GM3(Gc)	NeuGcα-3Galβ-4Glcβ-Cer	-
10	Haematoside	NeuAcα-3Galβ-4Glcβ-Cer	-
11	NeuAcα-(3')Lac-AO	NeuAcα-3Galβ-4Glc-AO	755
12	GSC-199	KDNα-3Galβ-4Glcβ-C30	-
13	NeuAcβ-(3')Lac-AO	NeuAcβ-3Galβ-4Glc-AO	832
14	Neuα-(3')Lac-AO	Neuα-3Galβ-4Glc-AO	667
15	Neu4,5Ac-(3')Lac-AO	(4-OAc)NeuAcα-3Galβ-4Glc-AO	750
16	GSC-75	(4-deoxy)NeuAcα-3Galβ-4Glcβ-Cer36	221
17	GSC-76	(7-deoxy)NeuAcα-3Galβ-4Glcβ-Cer36	-
18	GSC-77	(8-deoxy)NeuAcα-3Galβ-4Glcβ-Cer36	-
19	GSC-51	(9-deoxy)NeuAcα-3Galβ-4Glcβ-Cer36	-
20	GSC-78	(4-OMe)NeuAcα-3Galβ-4Glcβ-Cer36	84
21	GSC-79	(9-OMe)NeuAcα-3Galβ-4Glcβ-Cer36	-
22	GSC-161	NeuAcα-3Galβ-4Glcβ-C30 | Fucα-3	-
23	NeuAcα-(3')LN1-3-AO	NeuAcα-3Galβ-3GlcNAc-AO	266
24	NeuAcα-(3')LN	NeuAcα-3Galβ-4GlcNAc-DH	326
25	NeuAcα-(3')LN-AO	NeuAcα-3Galβ-4GlcNAc-AO	55

26	SA(3')-Lea-Tri-AO	NeuAcα-3Galβ-3GlcNAc-AO | Fucα-4	2
27	SA(3')-Lex-Tri-AO	NeuAcα-3Galβ-4GlcNAc-AO | Fucα-3	308
28	GSC-440	NeuAcα-3Galβ-4GlcNAcβ-C30 | Fucα-3	11
29	GSC-512	(4-OAc)NeuAcα-3Galβ-4GlcNAcβ-C30 | Fucα-3	-
30	GSC-513	(9-OAc)NeuAcα-3Galβ-3GlcNAcβ-C30 | Fucα-4	-
31	GSC-511	(9-OAc)NeuAcα-3Galβ-4GlcNAcβ-C30 | Fucα-3	-
32	GSC-479	NeuAcα-3Galβ-4GlcNAcβ-3Galβ-C30 | Fucα-3	577
33	GSC-105	NeuAcα-3Galβ-4GlcNAcβ-3Galβ-Cer36 | Fucα-3	-
34	GSC-177	NeuGcα-3Galβ-4GlcNAcβ-3Galβ-Cer36 | Fucα-3	129
35	GSC-341	KDNα-3Galβ-4GlcNAcβ-3Galβ-C30 | Fucα-3	-
36	GSC-257	NeuAcα-3(4,6-deoxy)Galβ-4GlcNAcβ-3Galβ-Cer36 | Fucα-3	201
37	GSC-175	NeuAcα-3(4-deoxy)Galβ-4GlcNAcβ-3Galβ-Cer36 | Fucα-3	-
38	GSC-176	NeuAcα-3(6-deoxy)Galβ-4GlcNAcβ-3Galβ-Cer36 | Fucα-3	326
39	LSTa	NeuAcα-3Galβ-3GlcNAcβ-3Galβ-4Glc-DH	108
40	GSC-272	NeuAcα-3Galβ-3GlcNAcβ-3Galβ-4Glcβ-C30	-
41	GSC-273	NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-C30	8436
42	GSC-396	NeuGcα-3Galβ-3GlcNAcβ-3Galβ-4Glcβ-C30	106
43	Sialylparagloboside	NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer	-
44	GSC-31	NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36	79
45	GSC-516B	Neuα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36 | SU-6	-
46	C4U	NeuAcα-3Galβ-4GlcNAcβ-3Galβ-3GlcNAc-DH | | | SU-6 SU-6 SU-6	2732
47	SA(3')-LNFP-II	NeuAcα-3Galβ-3GlcNAcβ-3Galβ-4Glc-DH | Fucα-4	396
48	SA(3')-LNFP-III	NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4Glc-DH | Fucα-3	175
49	GSC-64	NeuAcα-3Galβ-4GlcNAcβ-3Galβ-4Glcβ-Cer36 | Fucα-3	-

50	GSC-533		-
51	GSC-149		-
52	GSC-472		-
53	GSC-268		1392
54	GSC-268 deNAc		249
55	GSC-269		447
56	GSC-406		419
57	GSC-270		3219
58	GSC-220		24
59	GSC-221		527
60	MSMFLNH		1
61	A2F(2-3)		3575
62	P22-1 (GTP 3N(2,3)-3A(2,6)+F)		211
63	P6-1 (GTP 4N(2,3)-4A+F)		139
64	P7-2 (GTP 4N(2,3)-4A+1R+F)		137

65	P8-1 (GTP 4N(2,3)-4A+2R+F)		185
66	GM2		-
67	GM1		-
68	GM1-penta		-
69	GM1(Gc)		-
70	GM1(Gc)-penta		-
71	GSC-195		107
72	GD1a		-
73	GD1a-hexa		2356
74	GSC-335		547
75	GSC-488		50
76	GSC-489		963
77	GSC-154		-
78	GSC-441		1193
79	GSC-384		478
80	GSC-284		531
81	GSC-27		39
82	GSC-61		-
83	NeuAcα-(6')Lac-AO		335
84	NeuAcβ-(6')Lac-AO		476
85	Neuα-(6')Lac-AO		390
86	NeuAcα-(6')LN		-
87	NeuAcα-(6')LN-AO		-

88	Neu5,9Ac-(6')LN		-
89	LSTb		1243
90	LSTc		-
91	GSC-397		-
92	GSC-97		-
93	SA(6')-LNFP-VI		-
94	MSLNH		-
95	MSLNnH-I		138
96	DSLNnH		-
97	MFMSLNnH		-
98	A2(2-6)		66
99	AGP-Bi-Ac2		700
100	AGP-Bi-Gc2		1222
101	GSC-442		70
102	GSC-68		652
103	GSC-155		444
104	GSC-107		704
105	GSC-70		174
106	DST		1086
107	DST-AO		3027
108	GSC-490		-

109	DSLNT		3213
110	A3		3473
111	GSC-118		195
112	GSC-230		223
113	GSC-231		58
114	GSC-439		454
115	GD3		391
116	GD3-tetra-AO		873
117	GSC-229		319
118	GSC-437		-
119	GD2		228
120	GD1b		800
121	GQ1b		673
122	SA3(α8)		4026
123	SA5(α8)*		31735
124	SA7(α8)*		28224
125	SA9(α8)*		32320
126	GT1a		781
127	GT1b		2104
128	GSC-96		-

[a] the oligosaccharide probes are all lipid-linked, neoglycolipids (NGLs) or glycosylceramides and are from the collection assembled in the course of research in the Glycosciences Laboratory. For definition of the lipid moieties of the probes, please see https://glycosciences.med.ic.ac.uk/docs/lipids.pdf. [b] Numerical scores for the binding signals are shown as means of duplicate spots at 5 fmol per spot. [c] -, signal less than 1.

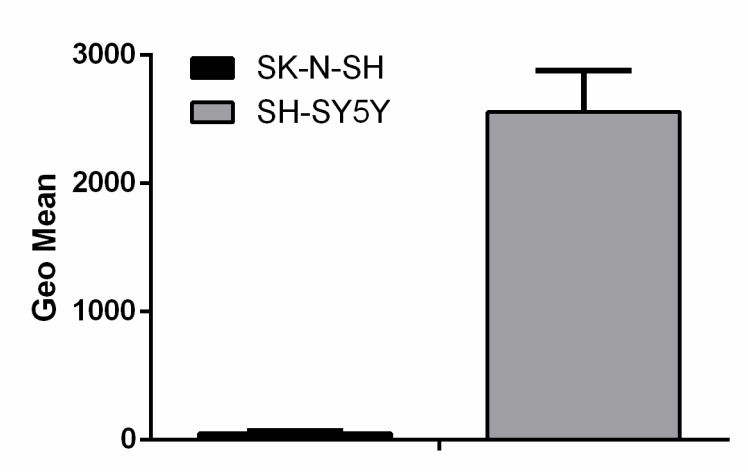


Figure S1. Levels of polysialic acid on SK-N-SH and SH-SY5Y cells. PSia expression was determined using the anti-PSia antibody (mab735). The experiment was performed three times with duplicate samples in each experiment. Error bars represent mean ± SD.

A

	ϕ C1-C2-O-C8	ψ C2-O-C8-C7	ω C9-C8-C7-C6	χ C8-C7-C6-C5
DP3	-55,45	121,38	69,79	168,74
DP4	-41,90	116,76	46,26	164,96
DP5	-45,33	128,74	40,84	165,53

B

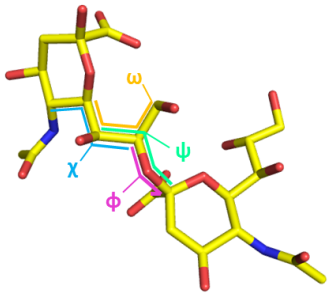


Figure S2. Observed glycosidic angles in 52SFK / PSia complex structures. A) Torsion angles observed for PSia of different DP in the complex crystal structures. All angles adopt similar conformations. B) Definition of the glycan torsion angles as in (*58*).

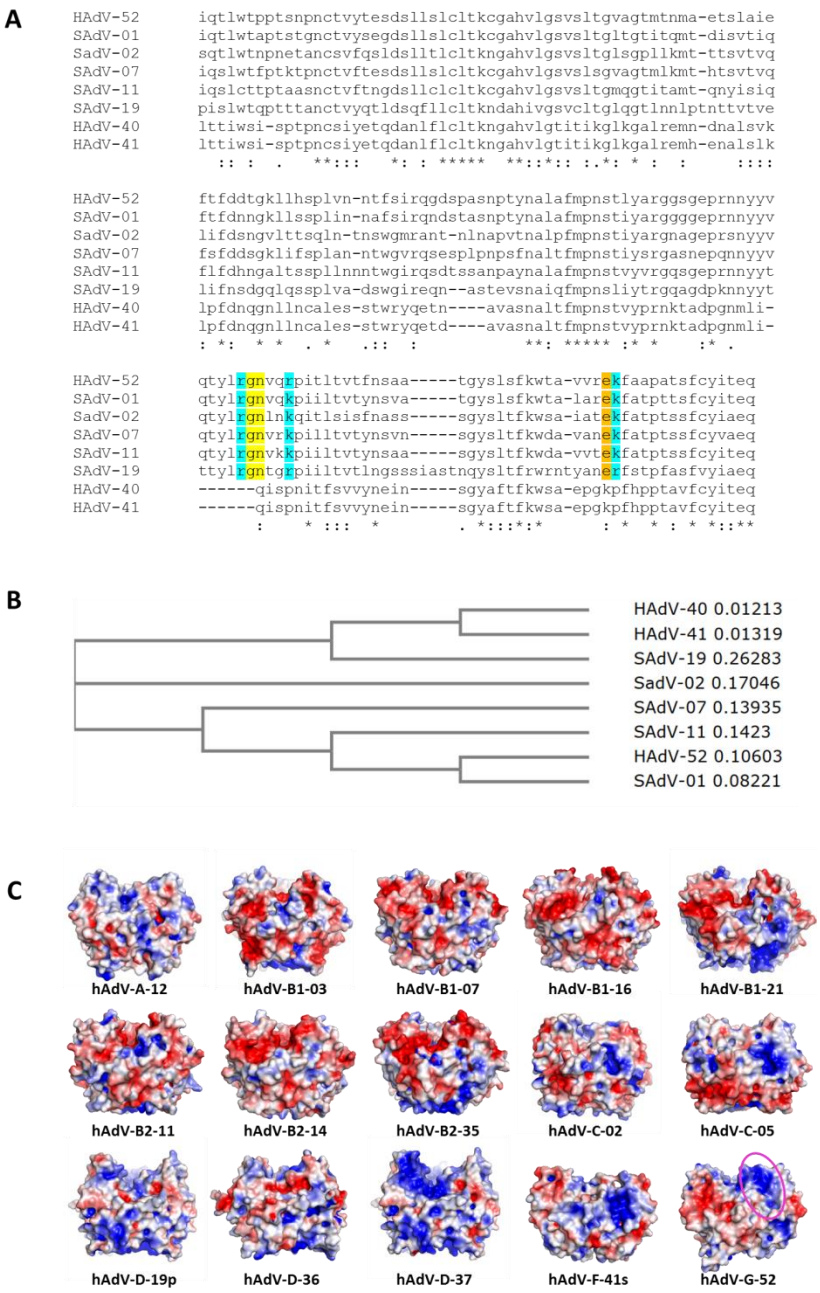


Figure S3. A) Sequence alignment of AdV knobs closely related to HAdV-52. While several simian AdV types have functionally conserved motives, the most closely related short fiber knobs of HAdV-40 and -41 largely differ in this region. Sequences are ordered by name. B) Phylogenetic cladogram of the short fiber knob sequences based on ClustalOmega alignment. Although SAdV-19 possesses both the RGN and 'steering rim' motives, it is more closely related to HAdV-41 than to other SAdV types. SAdV-2 does not cluster with the other types. C) Poisson-Boltzmann electrostatic surface potentials of known HAdV fiber knobs across all species. No other human type with a known structure possesses a sophisticated 'steering rim' on their surface. In the case of HAdV-41 short fiber knob, a severely disordered G strand is thought to obstruct the electropositive patch on the side (*59*). HAdV-5 possesses an electropositive patch, but lacks a useable sialic acid binding and has never been reported to use sialic acid as attachment receptor. Instead, it has been used as a negative control in many studies (see also Fig. 2). Electrostatic surface potentials have been calculated at ±3 kT/e.

Primary Attachment Receptors of Human Adenovirus 36

Liaci AM[1], Chandra N[2], Munender S[2], Liu Y[3], Pfenning V[1], Bachmann P[1], Caraballo R[4], Chai W[3], Johansson E[4], Cupelli K[1], Hassemer T[1], Blaum B[1], Elofsson M[4], Feizi T[3], Arnberg N[2], Stehle T[1,5]

[1]Interfaculty Institute of Biochemistry, University of Tuebingen, 72076 Tuebingen, Germany; [2] Umeå Centre for Microbial Research,,Umeå University, SE90187 Umeå, Sweden; [3]Glycosciences Laboratory, Department of Medicine, Imperial College London, London, United Kingdom; [4]Department of Chemistry, Umeå University, SE90187 Umeå, Sweden; [5]Department of Pediatrics, Vanderbilt University School of Medicine, Nashville 37232, TN, USA

Abstract

Human Adenovirus type 36 (HAdV-D36) is a member of HAdV species D and has first been isolated in the early 80's. The virus is commonly associated with gastroenteritis, but its pathogenicity profile and infectious routes are not well understood. Interestingly, HAdV-36 is found in human adipose tissue, and it has been causally linked to obesity in animals in a number of studies. In the course of these studies, HAdV-D36 has been demonstrated to infect animals such as chickens or rats and to spread from infected to healthy chickens through natural routes, which is unusual for HAdVs.

Despite the extensive research that has been conducted on the virus following this discovery, there is currently no information about its receptor profile. We investigated the attachment factor portfolio of HAdV-D36 using a combined structural biology and virology approach. The HAdV-D36 fiber knob domain (FK), which is responsible for the primary attachment of HAdVs to host cells, possesses a significantly elongated DG loop that alters known binding interfaces for established receptors such as the coxsackie- and adenovirus receptor (CAR) and CD46. Our data indicate that HAdV-D36 attaches to host cells using sialic acid.-containing glycans, CAR, and an unknown protein or glycoprotein. However, bioinformatic comparison of the fiber shaft with other HAdVs suggests that CAR is not used for productive infection. Sialic acids are recognized at the same binding site used by other HAdVs of species D such as HAdV-D37, although with lower IC_{50} values in cell attachment experiments. Using glycan microarrays, we demonstrate that HAdV-36 displays a binding preference for glycans containing a rare sialic acid variant, 4-O,5-N-diacetylneuraminic acid, over the more prominent 5-N-acetylneuraminic acid. To date, this sialic acid variant has not been detected in humans, although it can be synthesized by various animal species, including a range of domestic and livestock animals. The structural analysis of complex structures explains this preference, and shows that the interactions with the diacetylneuraminic acid are specific. Taken together, our results indicate that HAdV-D36 has evolved to recognize a specialized set of primary attachment receptors that is different from those of other known HAdV types and coincides with a unique host range and pathogenicity profile.

Introduction

Human adenovirus 36 (HAdV-D36) is a member of HAdV species D. The virus was first isolated in 1983 from a gastroenteritis patient [1]. HAdV-D36 was found to be serologically unique, which prompted Dhurandhar *et al.* to use it as an experimental replacement for SMAM-1, a chicken adenovirus that is associated with excessive fat accumulation in infected animals [2,3]. Interestingly, HAdV-D36 was shown to infect chickens, rhesus monkeys, marmosets, mice, and rats alike, causing obesity in all of them [2,4-6]. Moreover, the virus was shown to be transmitted horizontally from infected to healthy chickens through natural routes [5]. Both findings were surprising, since human adenoviruses (HAdVs) usually do not infect other species by natural routes, and predominantly infect mucous epithelial tissues. In HAdV-D36-infected chickens, however, the virus accumulated in fat tissue as opposed to skeletal muscle. In addition, the virus caused the differentiation of a murine preadipocyte cell line, while the CAR-binding HAdV-C02 required heterologous overexpression of the coxsackie- and adenovirus receptor (CAR) for efficient infection and did not induce adipogenesis [7]. After recognizing the adipogenic potential of the virus, large-scale epidemiological studies have been carried out in order to investigate a possible correlation of acute HAdV-D36 infection or seropositivity and human obesity. Indeed, significantly elevated HAdV-D36 seroprevalence was found in obese patients in multiple studies (reviewed in [8,9]), and viral DNA was isolated from human adipose tissue [10,11]. Three findings suggest a causative role of HAdV-D36 in human obesity. For one, infection of human adipose-derived stem cells (hASCs) with HAdV-D36 (but not HAdV-C02) resulted in lineage commitment, differentiation, and lipid accumulation, and natural infection of hASCs significantly increased the adipogenic potential [11]. This adipogenic effect can be attributed to the action of the HAdV-D36 E4ORF1 protein [12]. Secondly, obese patients with HAdV-D36 infections showed reduced serum lipids, while HAdV-D36 infection enhanced lipid and glucose uptake by adipose tissue explants and in human skeletal muscle cells, resulting in elevated intracellular lipid content and increased lipogenesis [13-15]. Thirdly, HAdV-D36 infections of human adipose tissues were found to be accompanied by abnormal fat deposits, e.g. in the thorax and abdomen - while sparing other body parts [16]. Since the rapid and worldwide increase of obesity since the 1980's is reminiscent of an infectious epidemic, several authors speculate that HAdV infections might be one of the factors contributing to this phenomenon [17-20]. However, the role of HAdV-D36 infection in human adiposity is still controversially discussed [21].

Despite the extensive research that has been conducted on the virus, information about the determinants of its unusual tropism and its ability to horizontally spread between animals has remained elusive to date. This prompted us to investigate the virus' primary attachment receptor portfolio, which is believed to be one of the determinants of HAdV tropism, employing a combined structural biology and virology approach. We postulate that the virus engages one or several unique receptors that might at least partially account for the observed phenomena.

Results

HAdV-D36 uses a sialic acid-containing glycoprotein and a protein receptor to attach to A549 cells

We initially screened for a cell line that allows for efficient attachment of ^{35}S-labeled HAdV-D36 virions. Due to the unusual tissue tropism of the virus, we tested a range of cell types originating from different tissues such as lung, adipose, small intestine, liver, and kidney (**Figure S1 A,B**). Among these cells, A549 (human adenocarcinoma alveolar basal epithelium cells) proved to be most suitable for HAdV-D36 attachment assays. We then sought to investigate the chemical nature of the factor(s) mediating the attachment using differently pretreated A549 cells.

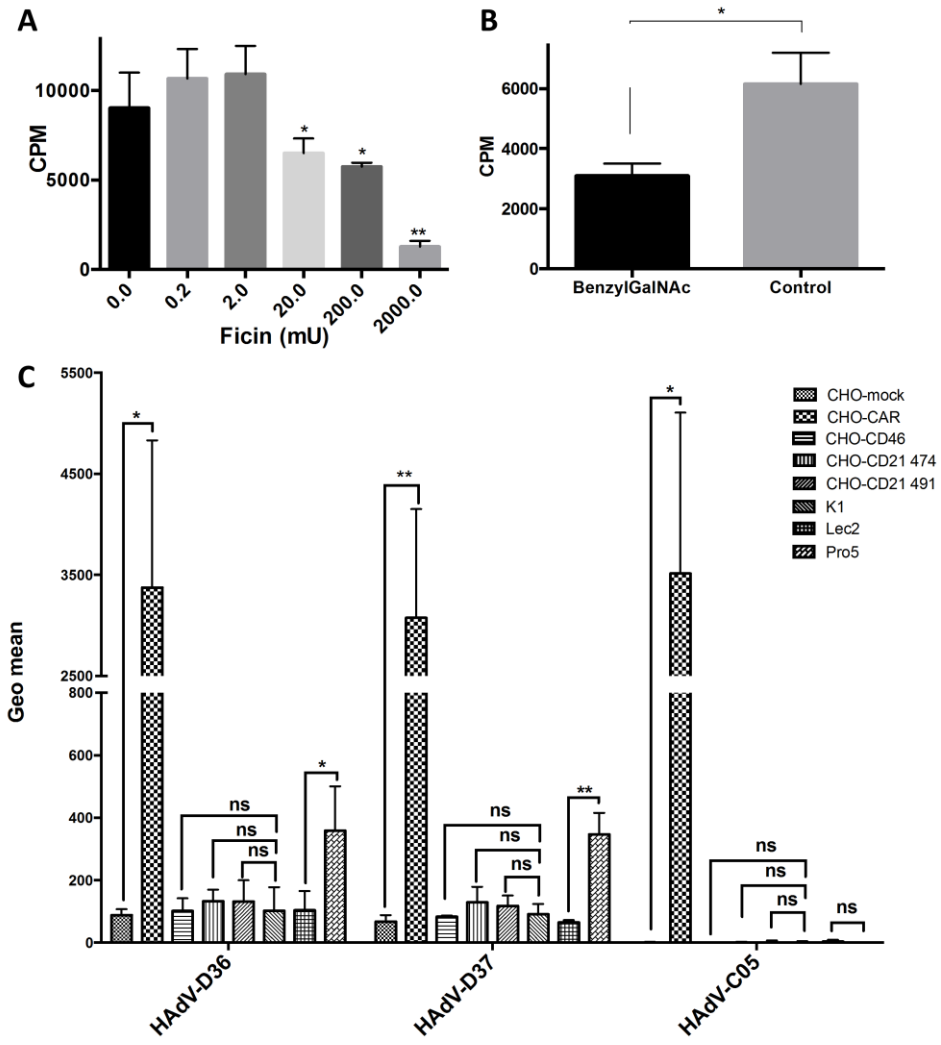


Figure 1 | HAdV-D36 uses CAR and sialic-acid-containing glycotopes to attach to A549 cells by its fiber knob. ***A*** *Binding of ^{35}S-labelled HAdV-D36 virions to A549 cells pretreated with different concentrations of the protease ficin.* ***B*** *Binding of ^{35}S-labelled HAdV-D36 virions to A549 cells cultured in the presence and absence of the metabolic inhibitor BenzylGalNAc* ***C*** *Binding of ^{35}S-labelled HAdV-D36 FK to different CHO cell lines expressing different human cell surface proteins (CAR, CD46, CD21; see legend on the right), or overexpressing sialic acid (Pro5). CHO-K1 and CHO-mock cells represent control cell lines. Y-axis shows the amount of knobs bound to cells, represented as CPM (counts per minute). All experiments were performed three times with duplicate samples. Error bars represent mean ± SD. * P of < 0.05 and ** P of < 0.*

Pretreatment of cells with proteases ficin (**Figure 1A**) and proteinase K (**Figure S2A**) significantly reduced the attachment of ^{35}S-labeled HAdV-D36 virions, whereas other proteases such as bromelain and V8 did not show an effect (**Figure S2B-C**). Additionally, we treated A549 cells with inhibitors of the *de novo* synthesis of glycolipids as well as *N*- and *O*-linked glycosylation and tested for their inhibitory potency on virion attachment. Treatment with Benzyl-GalNAc, an inhibitor of *N*-glycosylation, significantly reduced virus binding (**Figure 1B**). Tunicamycin, which inhibits *O*-glycosylation, showed a moderate reduction in virus binding which was statistically non-significant (**Figure S2D**). No effect was observed for pretreatment of cells with P4 (an inhibitor of glycolipid synthesis) or heparinase III (an enzyme which removes cell surface heparan sulfate; **Figure S2E-F**), as well as for the pre-treatment of cells with soluble heparin (**Figure S2G**).

We therefore concluded that the surface molecules mediating the attachment of HAdV-D36 virions to A549 cells must be a protein and/or a glycoprotein entity rather than glycolipids or glycosaminoglycans (GAGs). We screened for possible involvement of known HAdV attachment factors by testing virion binding to a library of CHO cell lines that overexpress the coxsackie- and adenovirus receptor (CAR), the complement regulatory protein CD46, and a number of candidate cell surface receptors such as CD21, ICAM-1, and CD55, including a control cell line (CHO-K1) and a mock-transfected CHO cell line (**Figure S3**). However, these experiments did not reveal any clear receptor candidate, as almost all cell lines displayed higher binding levels than the control. Variations were high even between control and mock-treated cells, and there was no clear tendency for functional usage of any of the overexpressed proteins. When using the FK domain instead of virions, however, we detected statistically significant attachment only for CHO cells that either ectopically express CAR or display high levels of sialic acid (Pro5) (**Figure 1C**). We thus hypothesized that HAdV-D36 uses both CAR and sialic acid-containing glycoproteins for cell attachment, and that both factors are recognized simultaneously by the FK domain.

We verified these results in a number of cell attachment inhibition assays and comparing HAdV-D36 virions to the known sialic acid binder HAdV-D37 and the CAR-binding HAdV-C05. We used HAdV-D37 as a control throughout this study, since it is functionally and structurally well-characterized and displays a more specific tropism and host range than HAdV-D36 - despite apparent similarities in its fiber knob. HAdV-D37 is a causative agent of epidemic keratoconjunctivitis in humans [22]. Similarly to HAdV-D36, the virus is known to engage sialic acid-containing glycans and uses the O-glycosylation motif GD1a as a primary attachment factor [23-25]. Treatment of A549 cells with *V. cholerae* neuraminidase significantly inhibited the attachment of HAdV-D36 viral particles, about half as efficiently as that of HAdV-D37 (**Figure 2A**). Pre-incubation of virions with soluble 5-N-acetylneuraminic acid (Neu5Ac) resulted in a similar pattern, with an IC50 of about 25 mM for HAdV-D36 and about 12 mM for HAdV-D37 (**Figure 2B**). Pre-incubation of HAdV-D36 virions with the soluble domain 1 of CAR (CAR-D1), which mediates the cell attachment of other HAdV types, also decreased attachment levels, although slightly less effectively than those of CAR-binding HAdV-C05 (**Figure 2C**). The canonical binding site for CAR and the sialic acid binding site of other species D HAdVs are located on distinct and spatially separated parts of the knob domain. Therefore, CAR and Neu5Ac engagement are unlikely to interfere with each other. Indeed, pre-incubation of A549 cells with both CAR-D1 and Neu5Ac had a synergistic inhibitory effect on virion binding (**Figure 2D**).

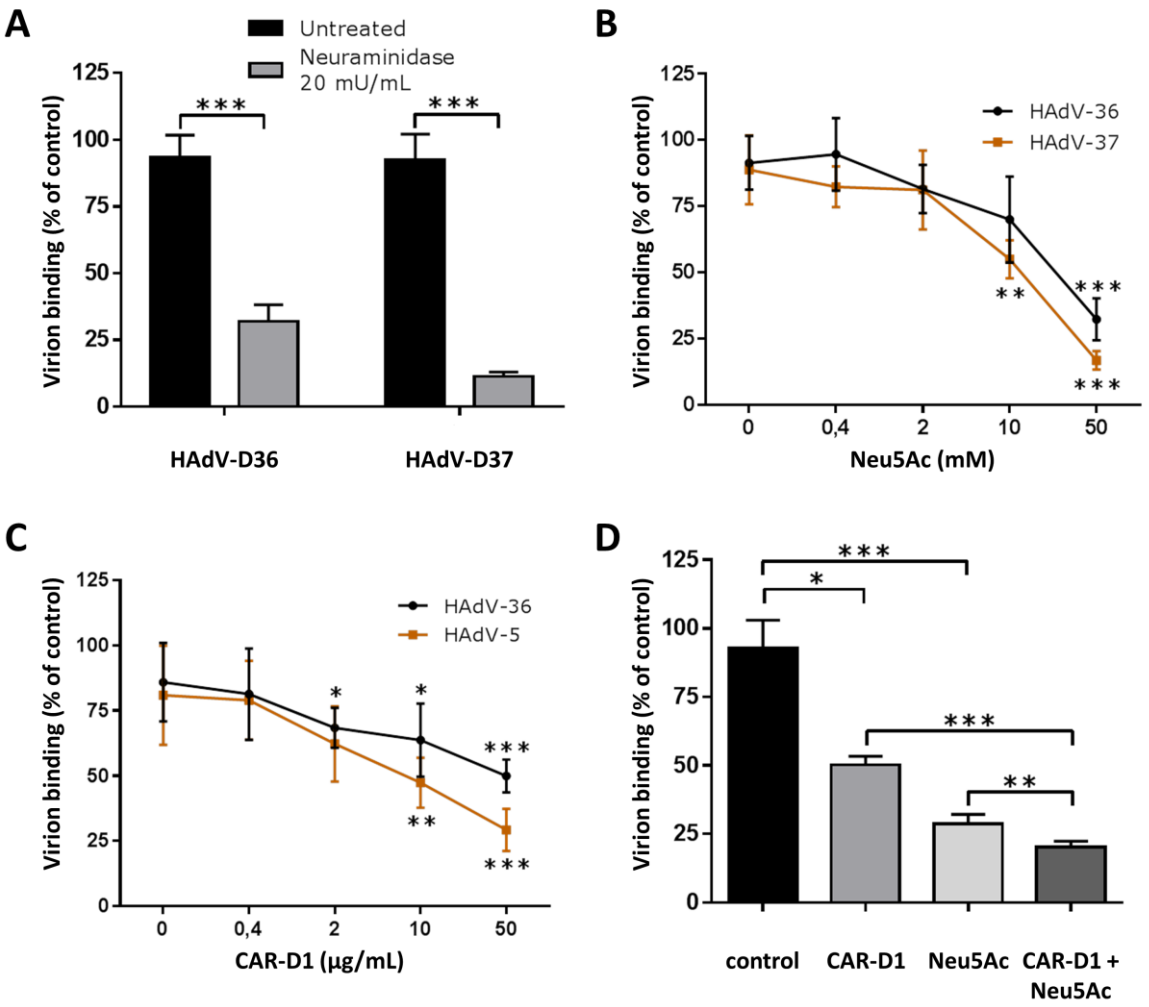


Figure 2 | CAR and Neu5Ac engagement have a synergistic effect on virion attachment. ***A*** *Neuraminidase treatment of A549 cells reduces HAdV-D36 virion attachment.* ***B*** *Pretreatment of HAdV-D36 virions with sialic acid reduced its attachment to A549 cells.* ^{35}*S-labelled HAdV-D36 and -D37 virions were pre-incubated with different concentrations of Neu5Ac.* ***C*** *Pretreatment of HAdV-D36 virions with soluble CAR-D1 reduced its attachment to A549 cells.* ^{35}*S-labelled HAdV-D36 and –C05 virions were pre-incubated with different concentrations of CAR-D1.* ***D*** *Pretreatment with both CAR-D1 and Neu5Ac has a synergistic effect on attachment inhibition. The highest concentrations of CAR-D1 (50 µg/mL) and Neu5Ac (50 mM) were used. All data is presented in % of control, where control refers to the binding of virions to cells without Neu5Ac and/or CAR-D1. All experiments were performed three times with duplicate samples. Error bars represent mean ± SD. ** P of < 0.01 and *** P of < 0.001.*

Structure of the HAdV-D36 FK and implications for CAR engagement

In order to define the regions known to mediate binding to sialic acid, CAR, and other known receptors such as CD46, we determined the crystal structure of the HAdV-D36 FK. Like all HAdV fiber knobs, the HAdV-D36 FK is a compact trimer. Each monomer is formed by a β sandwich domain, in which the β-strands of the two β-sheets (ABCJ and DIHG, respectively) are connected by long, surface-exposed loops. The E and F strands, which are present in some knobs are not formed in the case of HAdV-D36.

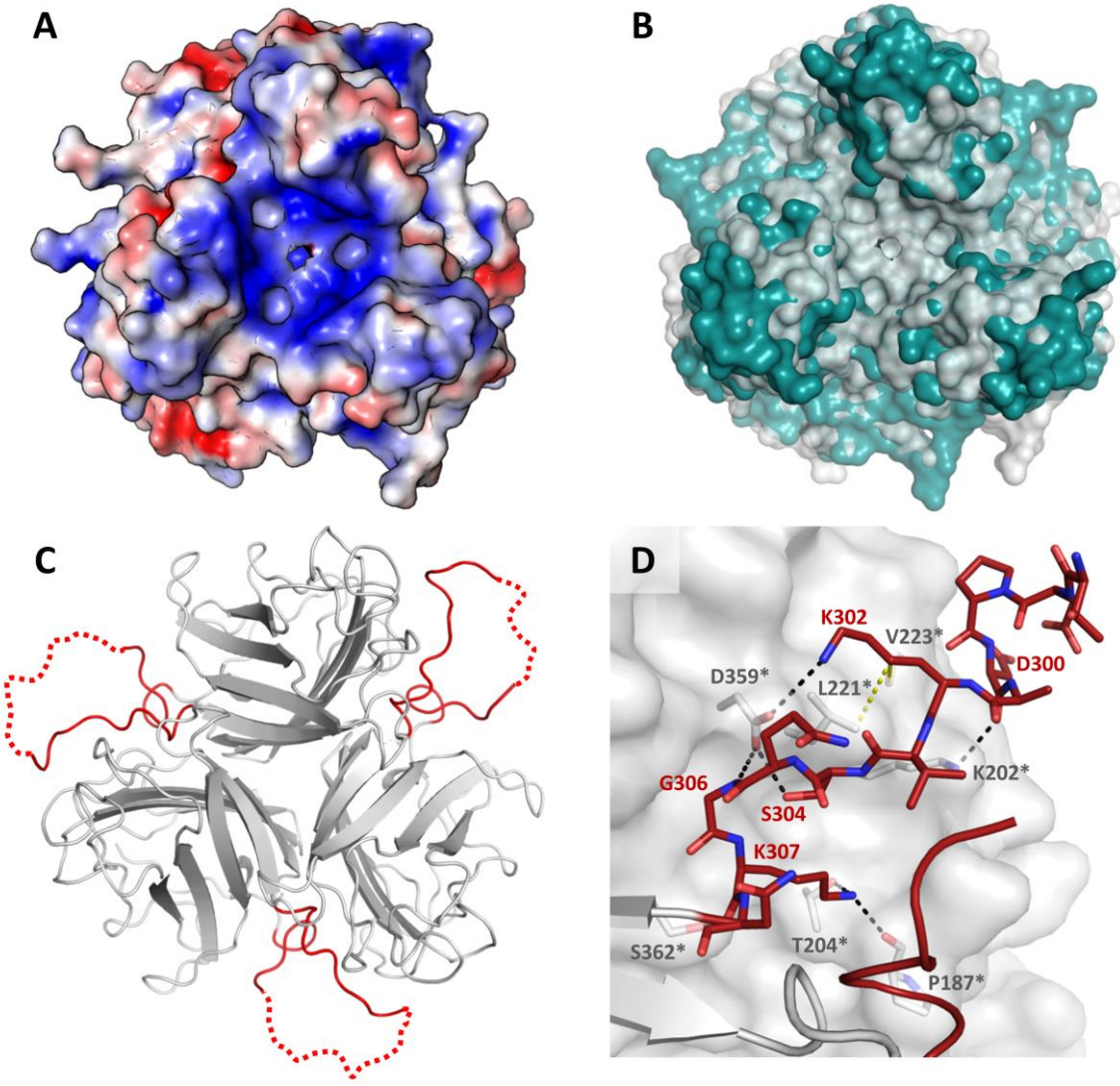


Figure 3 | Structural features of the HAdV-D36 FK. ***A*** *Poisson-Boltzmann electrostatic potential surface representation of the HAdV-D36 FK calculated at ±3 kT/e and shown from a top view along the threefold axis. The HAdV-D36 FK is less electropositive than that of HAdV-D37 [25].* ***B*** *Comparison of the quaternary arrangement of the FKs of HAdV-D36 (gray) and HAdV-D37 (green, PDB-ID 1UXE). The HAdV-D36 FK displays a narrower central cavity due to a tilt of the monomers towards the threefold axis. The superposition was calculated using all atoms of both trimers.* ***C*** *The HAdV-D36 FK possesses long DG loops (red) that are partially ordered and contact the counterclockwise neighboring monomer when viewed from the top.* ***D*** *Close-up view of the interactions between the DG loop and the counterclockwise neighboring monomer. Polar and hydrophobic interactions are displayed as black and yellow dashes, respectively.*

The knob possesses three main distinctive features when compared to other species D FKs. Firstly, it has a reduced positive charge in comparison to HAdV-D37 and D19p (**Figure 3A**) [25]. Secondly, the HAdV-D36 FK possesses a narrower central cavity compared to the two other knobs (**Figure 3B**). The reason for this is an altered relative arrangement of the monomers in the trimeric knobs (rmsd=0.79Å) rather than a significant difference between the monomers of each type (rmsd=0.48Å). Thirdly, HAdV-D36 possesses a significantly extended DG loop at located at the side (**Figure 3C**). With a total length of 21 amino acids, this loop is among the largest DG loops found in HAdV FKs. The loop is only partially flexible and forms a number of interactions with the clockwise neighboring monomer involving a stretch of nine residues (**Figure 3D**), probably contributing to the extraordinary stability of the fiber knob (**Figure S4**). Namely, there are interactions between the backbone of D300 and the side chain of K202 (AB loop). Additional interactions near the sheet are mediated by S304, G306 (both with D359) and K307 (with P187/T204, both AB loop). The bulky K302 forms a salt bridge with D359, while the alkane part of its side chain interacts with V223 and L221. Even at a pH below 3, the fiber knob still displays a remarkable stability (Tm 64.6°C), which might be one of the prerequisites for its putative enteric tropism.

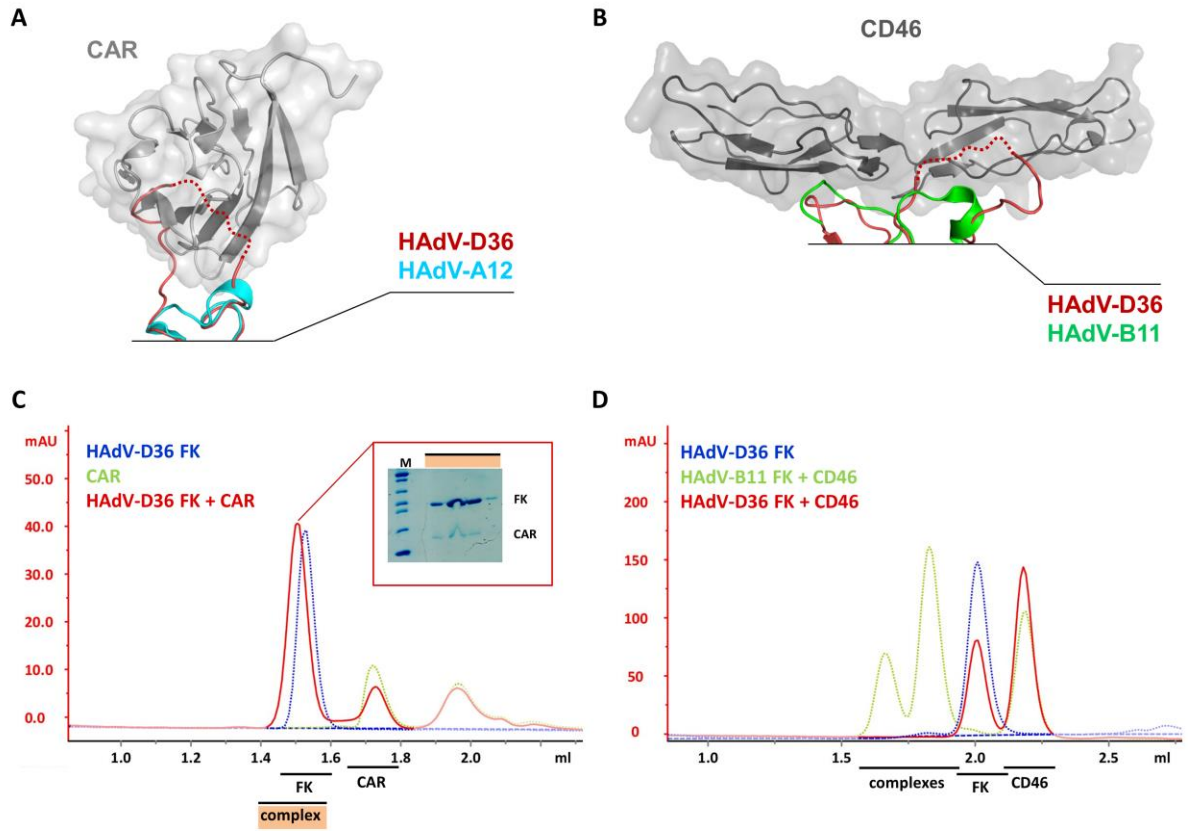


Figure 4 | Molecular modelling and SEC analysis of CAR and CD46 binding. ***A*** *When superposed onto the HAdV-A12 FK (blue) in the CAR complex structure (PDB-ID 1KAC), the HAdV-D36 FK (red) heavily clashes with CAR (light gray) through its DG loop.* ***B*** *Superposition of HAdV-D36 FK onto HAdV-B11 in the CD46 complex structure (PDB-ID 2O39). The HAdV-D36 DG loop is located proximally in front of CD46 (dark gray) and would likely interfere with CD46 binding without producing direct clashes.* ***C*** *Size exclusion chromatography of the HAdV-D36 FK / CAR complex. HAdV-D36 FK was incubated with soluble CAR-D1 for 20 minutes prior to SEC. Due to the small size of CAR-D1, the shift of the complex (red line) in comparison to uncomplexed HAdV-D36 FK (blue line) is weak. However, SDS-PAGE of the peak fraction (orange bar) confirmed the formation of a complex. Green line = uncomplexed CAR-D1. Equal amounts of FK and CAR were used for optimal comparability.* ***D*** *Size exclusion chromatography of HAdV-D36 and CD46. When compared to HAdV-B11, which readily forms complexes of different stoichiometry with the soluble extracellular portion of CD46 (CD46-D4, green line) and the elution profile of the unliganded HAdV-D36 FK, it becomes obvious that HAdV-D36 FK and CD46-D4 do not form a stable complex (red line). The proteins causing the respective peaks are noted below the chromatograms.*

Interestingly, the DG loop is located near the canonical binding interfaces of both CAR (determined in complex with HAdVs A12 and D37, PDB IDS 1KAC and 2J12) and CD46 (in complex with HAdVs B11 and B21, PDB IDs 2O39 and 3L89), and the length of this loop is one of the determinants for binding of both receptors (reviewed in [26]). Although the residues directly interacting with CAR, mainly located on the AB loop, are generally conserved in HAdV-D36 and the main part of the canonical binding site shows a high shape complementarity to CAR (assessed by the SC program in CCP4 [27], data not shown), structural superposition of the HAdV-D36 FK with those complex structures results in major clashes between the DG loop of the counterclockwise adjacent FK monomer and CAR (**Figure 4A**). These clashes could only be overcome by a major rearrangement in the loop involving the dissociation of the inter-monomer contacts, or by altering the binding region. HAdV-D36 FK readily forms a stable complex with the distal extracellular domain of CAR (CAR-D1) using a protocol similar to that used for the formation of the HAdV-D37/CAR-D1 complex (**Figure 4C**) [28]. However, the physiological relevance of CAR binding for HAdV-D36 infection is questionable. As such, HAdV-D36 (but not HAdV-C02) is capable of infecting and differentiating mouse preadipocytes that do not express human CAR [7]. The length of the HAdV-D36 fiber shaft is very similar to that of HAdV-D37. The latter also binds to CAR, but cannot use it for productive infection, likely due to the sturdiness of its fiber that prevents engagement of the secondary receptor. Similarly to HAdV-D37 and -D26, HAdV-D36 has only seven to eight predicted repeat units and is lacking a non-repeat motif between repeat three and four that is thought to be crucial for fiber bending [29].

As a confirmatory measure, we also inspected the interference of the long DG loop with potential binding of CD46, the second protein HAdV receptor with a known interface. The DG loop is expected to interfere with a potential CD46 binding site, although no direct clashes were observed in the model (**Figure 4B**). The major determinants of CD46 binding that have been determined for HAdV-B11, B07, and B14 are not conserved in HAdV-D36 [30]. In constrast to CAR, CD46 does not form a stable complex with HAdV-D36 FK (**Figure 4D**).

The alignment of the FK sequences of all species D serotypes (including types from HAdV-D08 to D56) reveals two major clades of similar size (**Figure 5A**). The 18 members of the 'HAdV-D36-like' clade possess similarly elongated DG (or FG) loops, while the 17 knobs of the 'HAdV-D37-like' clade containing all EKC-causing HAdVs possess DG/FG loops shorter by eight to eleven residues. Species D is the largest HAdV species and contains 47 types, many of which result from homologous recombination occurring upon co-infection of multiple species D HAdVs in the gut [31,32]. Many of the species D types have low seroprevalence rates and no associated disease or tissue profile [31-34]. From what is known in the literature so far, there seems to be no consistent tissue tropism within either of the two clades, except for the fact that EC-causing HAdVs are so far exclusively found in the 'HAdV-D37-like' clade [32,35]. However, to our knowledge, none of the HAdVs of the 'HAdV-D36-like clade' has a reported primary attachment receptor. Since the DG/FG residues that confer specific inter-monomer contacts are conserved within the 'HAdV-D36-like' clade, we postulate that there are similar implications for the receptor binding of these knobs as well. Similar assumptions have been made previously based on the HAdV-D36 genome sequence [36].

The amino acids of the G strand directly adjacent to the DG loop show a similar conservation pattern as the loop itself. The G strand forms a large portion of the trimeric interface and is located at the base of the central cavity, right underneath the canonical sialic acid binding site. Visual analysis of the trimer interfaces of the HAdV-D36 and D37 FKs revealed a stretch of three amino acids within this strand (*VSN*311-313 in HAdV-D36 and *YGT*308-310 in HAdV-D37) that likely determines the relative orientation of the monomers with respect to each other, thus accounting for the narrower trimer interface in HAdV-D36 (**Figure 5B,C**). This area also harbors the sialic acid binding site of HAdV-D37.

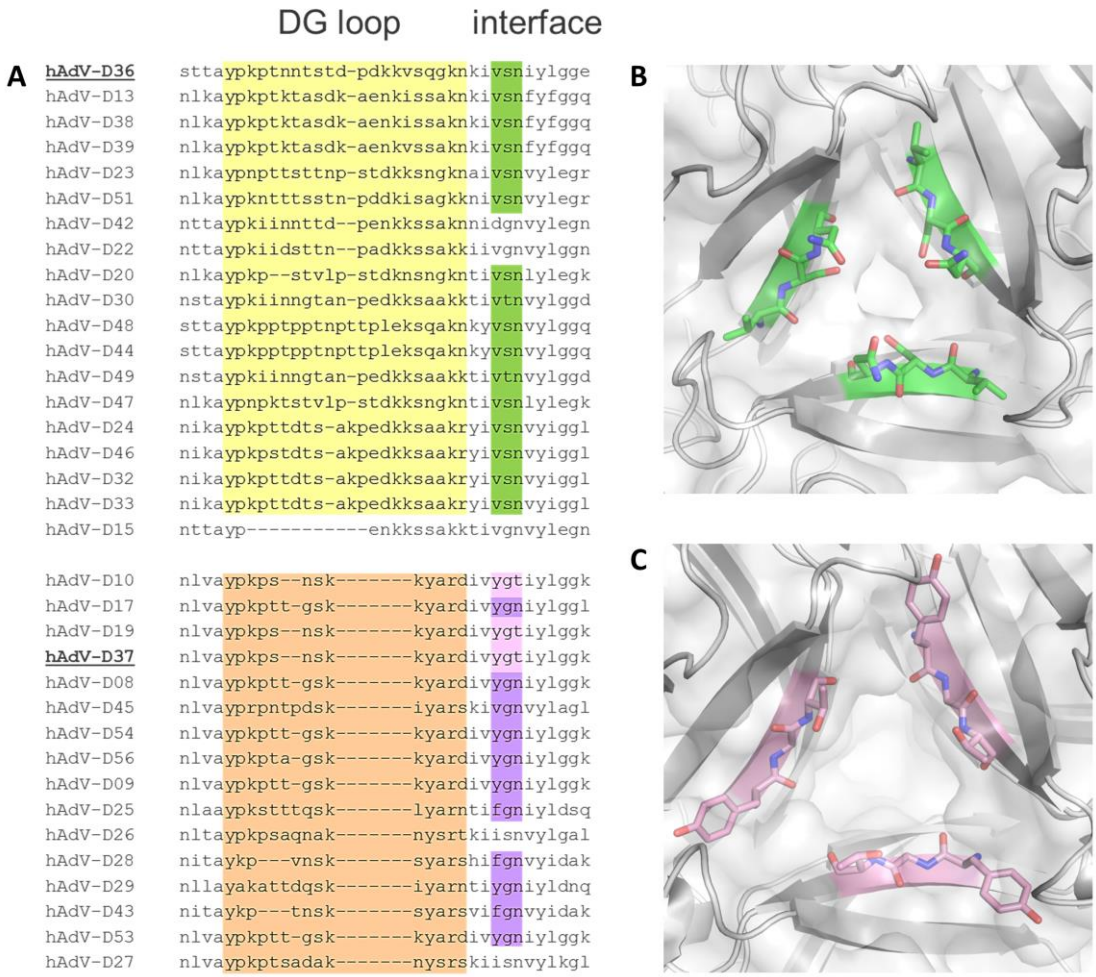


Figure 5 | Conservation of DG loop and trimer interface among HAdV species D. ***A*** *Excerpt of the alignment of all species D HAdVs from type HAdV-D08 to D56 performed with Clustal Omega [37]. The 18 HAdV-types belonging to the 'HAdV-D36-like' clade possess elongated DG loops (yellow box) and a VSN or VTN interface configuration (green). The 17 types belonging to the 'HAdV-D37-like' clade have shorter DG loops and a YGT (light pink) or YGN (purple) interface configuration. HAdV-D15 could not be assigned to any of the two clades.* ***B*** *VSN trimer interface (green) as found in the HAdV-D36 FK.* ***C*** *YGT trimer interface as observed in the HAdV-D37 FK (PDB-ID 1UXE).*

The glycan binding site of HAdV-D36 FK is similar to other species D HAdVs

The structure of the HAdV-D36 FK in complex with the synthetic Neu5Ac analogue α-2-O-methyl-Neu5Ac confirms that this compound binds to the canonical sialic acid binding site (**Figure 6A**). In the case of the HAdV-D37 fiber knob, the two key polar interactions are formed between residue K345 and the carboxyl group of Neu5Ac and between the backbone of P317 and the sugar's amide nitrogen, respectively (**Figure 6B**). Both interactions are formed with the β-face of the sugar and are also present in the HAdV-D36 complex structure, although the analogous position of P317 is occupied by a valine (V320) in HAdV-D36.

The narrower binding cavity has several implications for the way HAdV-D36 engages Neu5Ac. For one, this arrangement allows HAdV-D36 to contact the sugar from the α-face, using a second monomer to form a direct hydrogen bond between N313 and the N-acetyl group of Neu5Ac, as well as a water-mediated contact between the backbone of W348 and the sugar's glycerol chain. HAdV-D37, on the other hand, only mediates water-bridged contacts to the α-face. Secondly, the binding cavity for the N-acetyl group is rearranged and becomes less spacious in the HAdV-D36 FK (**Figures 3A and 6A-C**). The residues forming this cavity are Y315 as

well as V311 and F325 from the monomer facing the α-side of the sugar. In HadV-D37, the N-acetyl group is accommodated in a canyon-shaped cavity formed mostly by two tyrosines (Y312, Y308) from the neighboring monomer, as well as V322, which occupies the position analogous to F325. Overall, the HAdV-D36 FK quaternary structure causes a shift of the sugar in the direction of the α-face (**Figure 6C**).

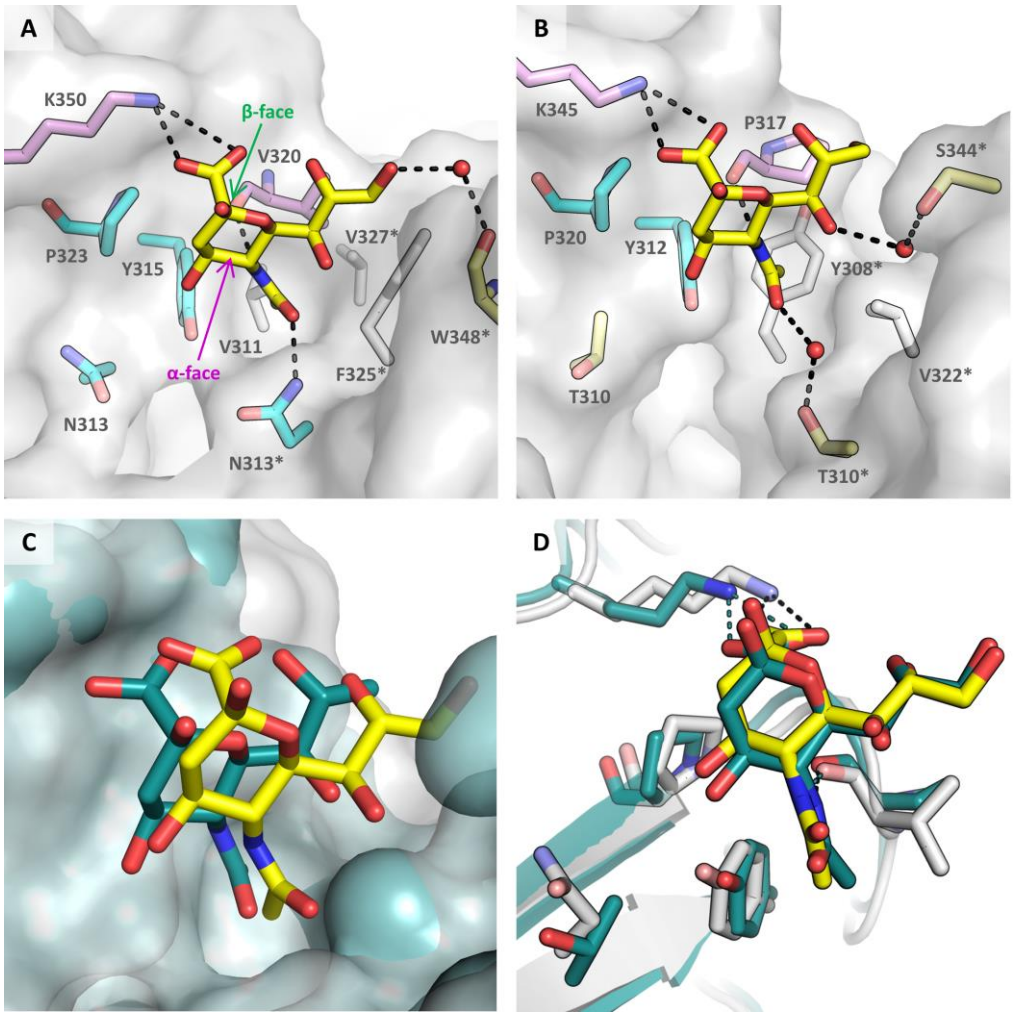


Figure 6 | Comparison of the Neu5Ac binding modes of HAdV-D36 and HAdV-D37. ***A*** *Neu5Ac (yellow) binding mode observed for the HAdV-D36 FK (gray surface). The structure was solved with α-2-O-methyl-Neu5Ac, and methyl functions have been left out for clarity. Residues marked with a '*' are contributed by the clockwise neighboring monomer. Polar key contacts are mediated by K350 and V320 (purple), while the hydrophobic binding cavity for the 5-acetyl function is formed by V311, V327*, and F352*. W348 (yellow) contacts the glycerol function by a water-bridged contact. The three residues N313, Y315, and P323 (cyan) put steric constraints onto the sugar's O4 atom, while N313* contacts the 5-acetyl group by means of a hydrogen bond from the α-face. The α- and β- faces of Neu5Ac are marked with an arrow.* ***B*** *Neu5Ac binding mode observed for HAdV-D37 (PDB-ID 1UXA). Coloring according to A. Unlike N313 in HAdV-D36, T310 (yellow) does not put steric pressure onto the Neu5Ac O4, and T310* contacts the sugar via a water-bridged polar contact.* ***C*** *Relative placement of Neu5Ac complexed to HAdV-D36 FK (yellow) and HAdV-D37 FK (teal) upon superposition of the knob trimers. The sugars show a prominent relative shift.* ***D*** *Comparison of the two binding modes upon superposition of only one monomer. The binding modes are largely similar, and the differences observed in C largely originate from an altered relative placement of the monomers within the knob.*

When aligning two monomers of the HAdV-D36 and D37 fiber knobs (**Figure 6D**), it becomes apparent that this shift is not caused by differences in the tertiary structure. Despite the overall similarity of the binding, our glycan array, attachment inhibition, and hemagglutination data consistently show that HAdV-D36 has a lower affinity for Neu5Ac than HAdV-D37 (**Figures 7 and S5, 2B, S6**). This discrepancy is likely to originate at least partially from the lower electropositive charge of the HAdV-D36 binding pocket.

HAdV-D36 preferentially binds to a rare sialic acid variant

We employed glycan array screening with a library of about 500 compounds in order to identify sialic acid-containing glycans that support efficient binding of the HAdV-D36 FK (**Figure 7A, Table S1**). The array showed a low overall fluorescence signal and no elevated signal for the GD1a glycan. Interestingly, one glycan gave the highest signal by far: a 3'sialyllactose (3'SL) probe containing a rare sialic acid variant with an additional 4-O-acetylation (Neu4,5Ac_2, the glycan will be referred to as 4-O-Ac-3'SL). Glycans containing a 3'SL motif capped with the more common Neu5Ac showed little to no binding. In contrast, HAdV-D37 showed preferential binding to 3'SL over 4-O-Ac-3'SL and also bound to several other sialylated probes in the array, in overall agreement with our earlier findings (**Figure S5**). A second array that specifically probed for the binding to differentially acetylated 3'SL variants confirmed these results (**Figure 7B,C**). The gene conferring 4-O-acetlyation of Neu5Ac in vertebrates has not been characterized to date [38,39]. Only few studies detected 4-O-acetylated sialic acid in humans [40,41], and Neu4,5Ac_2 was not among them. The sialic acid variant has not been detected in more recent screenings of human tissues using lectins from Mouse Hepatitis Virus Strain S (MHV-S) and Infectious Salmon Anemia Virus (ISAV). However, Neu4,5Ac_2 is readily detected in a wide range of domesticated animals such as chickens, mice, rats, rabbits, guinea pigs, or horses, as well as a large number of fish species including salmon, cod, herring, and trout [39,42,43]. Furthermore, it is present as one of several sialic acid modifications on certain bacteria such as *E.coli* K1 and *N. meningitidis* [44]. As described in the introduction, several of these animals have been shown to be infected by HAdV-D36. The physiological role of 4-O-acetylation and its apparent absence in some animals remain enigmatic. One current theory is that its function may lie in the inhibition of pathogen sialidases (reviewed in [45]). 4-O-Ac-3'SL has a more restricted species distribution than Neu4,5Ac_2 and has been reported only in the milk of Australian monotremes such as the short-beaked echidna [46].

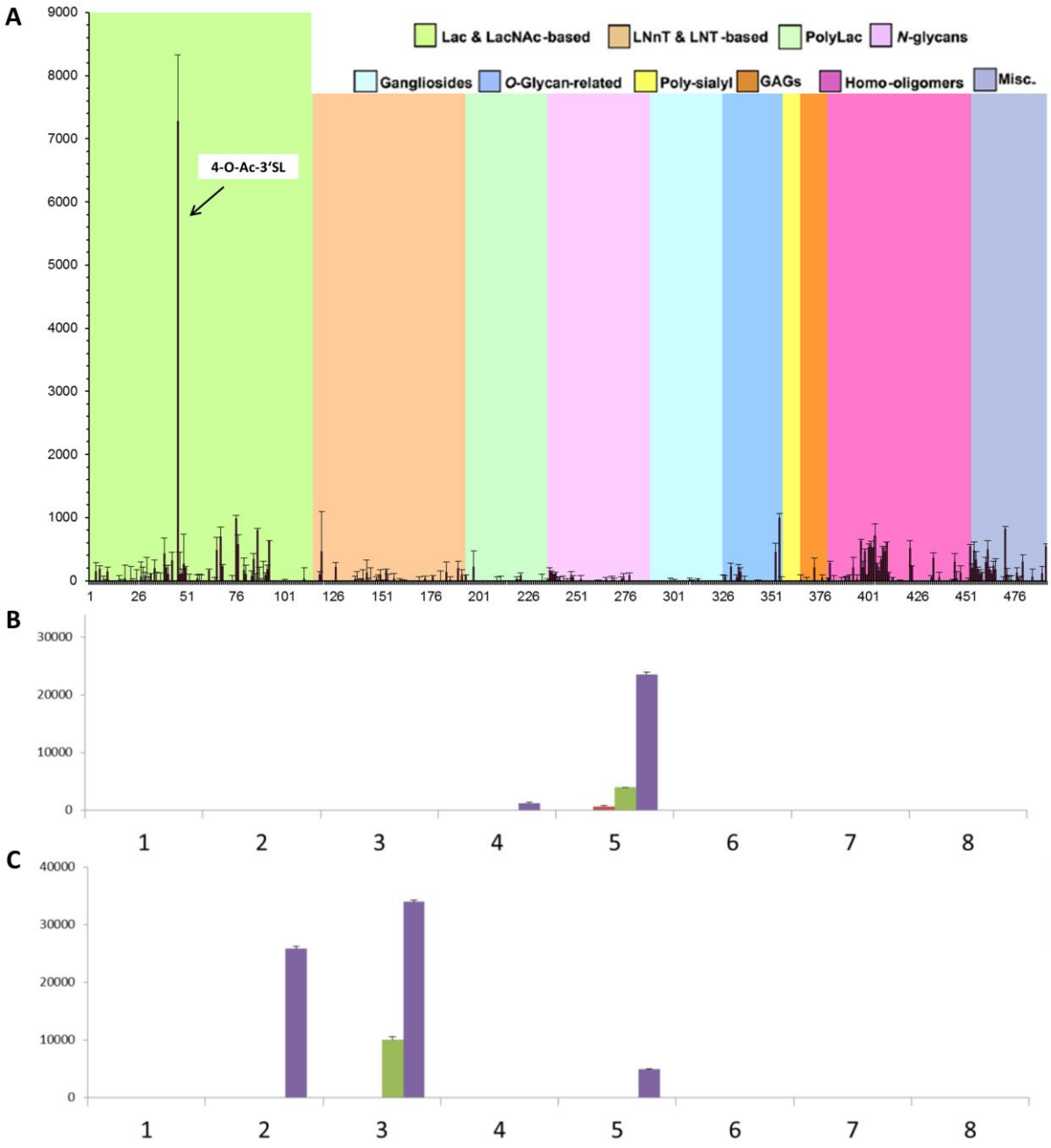


Figure 7 | Glycan array screening for the HAdV-D36 FK and comparison to the HAdV-D37 FK. ***A*** *A large glycan array containing 498 glycans (see legend on the upper right) showed a low overall signal with one outstanding signal caused by 4-O-Ac-3'SL (7,279 fluorescence units).* ***B*** *A smaller screen designed to test for HAdV-D36 FK binding to differentially acetylated 3'SL variants. 1=Lactose (Gal-[β-1,4]-Glc); 2,3=3'SL (Neu5Ac-[α2,3]-Gal-[β-1,4]-Glc); 4,5=4-O-Ac-3'SL (Neu4,5Ac$_2$-[α2,3]-Gal-[β-1,4]-Glc); 6,7=6'SL (Neu5Ac-[α2,6]-Gal-[β-1,4]-Glc), 8=6'SL with containing GlcNAc instead of Glc (Neu5Ac-[α2,3]-Gal-[β-1,4]-GlcNAc). Even numbers are attached via a DHPE linker, while uneven numbers are attached via an aminooxy (AO) functionalized linker DHPE linker. Glycan quantities used per spot: 0.3 fmol (blue bars), 0.8 fmol (red), 1.7 fmol (green), 5 fmol (purple). The HAdV-D36 FK only shows a signal for 4-O-Ac-3'SL.* ***C*** *The same array as in B was applied to the HAdV-D37 FK. The knob shows a preference for 3'SL, while 4-O-Ac-3'SL gives a much lower signal. Differences between lipid linkers are routinely observed in glycan arrays.*

We were able to solve the crystal structure of the HAdV-D36 fiber knob in complex with 4-O-Ac-3'SL isolated from echidna milk, and found that the lactose part of the sugar is not ordered in two of the three complexed glycans and does not seem to contribute to the binding (**Figure 8**). The only detectable contact mediated by the glycan stem was formed between the poorly ordered O6 atom of the galactose moiety and a backbone carbonyl function Thus, the Neu4,5Ac_2 part must be the decisive factor for the increased binding. Since 4-O-Ac-3'SL is the only Neu4,5Ac_2-containing glycan in the array, we conclude that the sialic acid variant is probably recognized in several glycan contexts, similar to what has been fund for HAdV-D37 which recognizes a range of Neu5Ac-containing glycans. For subsequent studies, we therefore replaced 4-O-Ac-3'SL with the more readily accessible synthetic α-2-O-methyl-Neu4,5Ac_2. We could demonstrate by X-Ray crystallography that the complex structures of both compounds are highly similar (**Figure 9**). The crystal structure of the HAdV-D36 FK in complex with a similar synthetic sialic acid variant, α-2-O-methyl-Neu4,5,9Ac_3, revealed that an additional acetylation at position 9 can be accommodated, but does not contribute additional interactions to the binding (**Figure S7A**). Similarly, none of the three Neu5,9Ac_2-containing probes from the glycan array displayed a binding signal (**Table S1**). Molecular modelling of sialic acid variants containing an additional O-acetylation at positions 7 and 8 suggest that both would likely induce clashes and would result in an abrogation of the binding (**Figure S7B**). We therefore conclude that 4-O-acetylation has a specific effect on HAdV-D36 recognition that is not seen with other sialic acid variants.

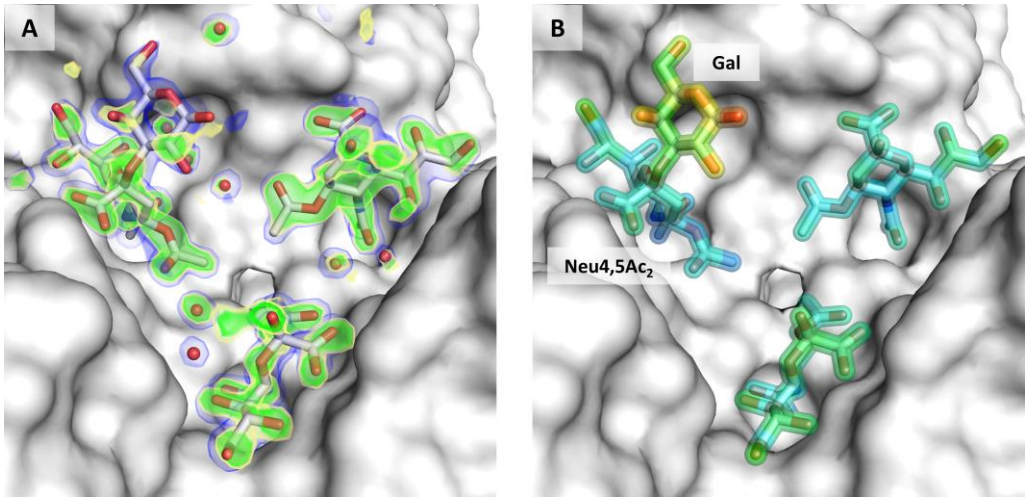


Figure 8 | Structure of the HAdV-D36 FK in complex with 4-O-Ac-3'SL. ***A*** *Simulated annealing omit map (green: 3σ; yellow: 2.5σ) and $2F_o$-F_c map after refinement (1σ, blue). The lactose stem is generally poorly ordered and only partially visible in one of three cases.* ***B*** *4-O-Ac-3'SL colored by B factors ranging from 10Å^2 (blue) to 60 Å^2 (red).*

Structural basis for the Neu4,5Ac$_2$ preference

The HAdV-D36 FK/Neu4,5Ac$_2$ binding mode is generally similar to that of Neu5Ac. The key polar interactions mediated by the carboxyl and amide functions of Neu5Ac are also found in the Neu4,5Ac$_2$ complex structure, as are the residues that make up the hydrophobic cavity for the N-acetyl group and the water-bridged hydrogen bond mediated by the glycerol function. The main distinguishing element is the relative orientation of the three Neu4,5Ac$_2$ moieties that are tilted towards the threefold axis (**Figure 9A,B**). Interestingly, the O-acetyl function does not contribute any polar contacts, and instead induces the tilting movement under the steric influence of the bulky side chains of Y315 and P323. This interaction is supposedly not favorable for the binding; however, the steric constraints bring the O-acetyl groups of the three sugar moieties into a position in which they contact each other by means of a triangular hydrophobic interaction (mean distance between methyl C atoms: 4.3 Å) that displaces water molecules in the area (**Figure 9B**). Since the interaction is mediated solely *via* the O-acetyl groups, a similar effect is not expected for Neu5Ac. The binding environment of the O-acetyl groups is made up by three amino acids: N313, Y315, and P323 (**Figure 9C**). Y315 thereby acts as a gatekeeper residue that inserts itself between the N-acetyl and O-acetyl groups of Neu4,5Ac$_2$.

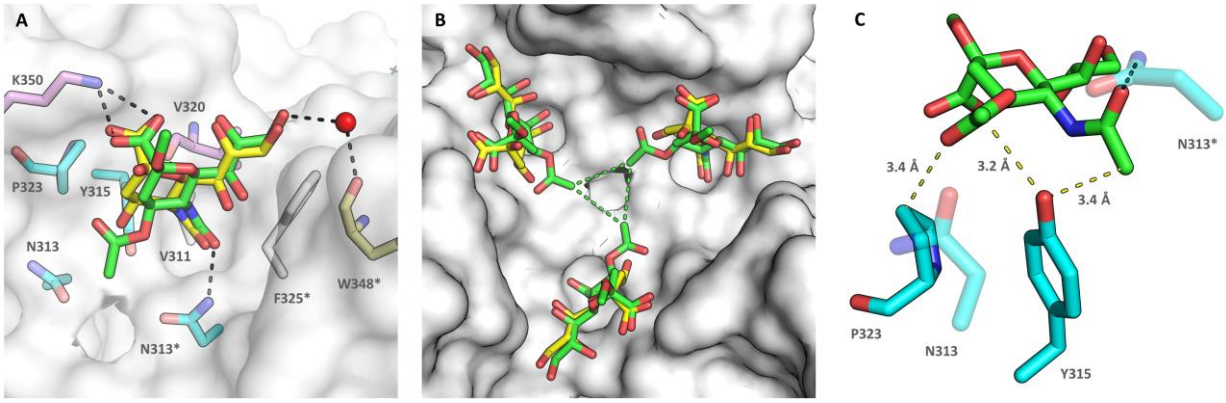


Figure 9 | Comparison of Neu5Ac (yellow) and Neu4,5Ac$_2$ (green) engagement by HAdV-D36. ***A*** *Superposition of the two complex structures, coloring analogous to Figure 6. Neu4,5Ac$_2$ is forced towards the threefold axis by the steric influence of N313, Y315, and P320, while the polar key contacts are only slightly altered.* ***B*** *This movement causes the three neighboring Neu4,5Ac$_2$ moieties to form a triangular hydrophobic contact (green dashes) that is not possible for Neu5Ac.* ***C*** *Steric influence of Y315 and P320 on Neu4,5Ac$_2$. Short-ranged van-der-Waals contacts are formed with both acetyl groups of the sugar. As a result, even slight alterations in the Y315 conformation would interfere with the binding.*

Even a subtle displacement of this residue would likely interfere with Neu4,5Ac$_2$ binding. Interestingly, all known complex structures with 4,5-acetylated sialic acid display a similar arrangement (**Figure S8**) [47,48]. N313 is one of the residues of the conserved *VSN* stretch and closes the binding pocket from below. The three N313 residues of the trimeric knob are located right at the threefold axis, and each residue contacts two Neu4,5Ac$_2$ residues simultaneously. Similarly to the Neu5Ac complex structure, N313 engages in a direct hydrogen bond with the β-face of one Neu4,5Ac$_2$ moiety (**Figure 10A**). Additionally, each N313 residue is in the van der Waals range of the O-acetyl groups from two neighboring Neu4,5Ac$_2$ compounds simultaneously (**Figure 10B**). The topology and reduced polarity of this residue allow for the formation of the triangular inter-sugar hydrophobic contact.

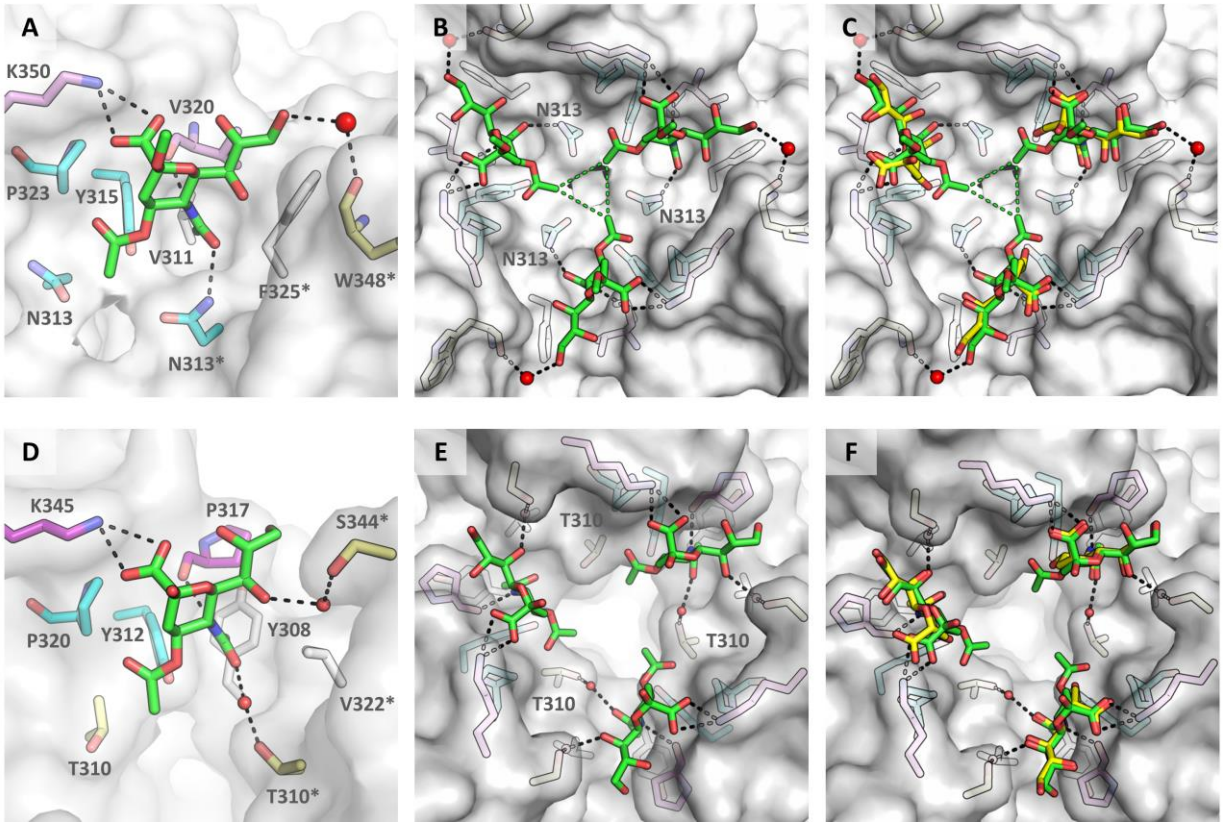


Figure 10 | Comparison of the Neu5Ac and Neu4,5 Ac2 binding between HAdV-D36 (upper panel) and HAdV-D37 (Lower panel). *Coloring of interacting residues according to Figure 9.* ***A-C*** *Neu4,5Ac$_2$ (green) binding mode observed for the HAdV-D36 FK (gray surface) analogous to Figure 9A-B.* ***D-F*** *Neu4,5Ac$_2$ (green) binding mode observed for the HAdV-D37 FK (gray surface). The triangular hydrophobic interaction is not observed in HAdV-D37. Instead Neu4,5Ac$_2$ moieties are located in the periphery of the binding cavity. The relative movement between Neu5Ac (yellow, PDB-ID 1UXA) and Neu4,5Ac$_2$ is less pronounced for HAdV-D37 than for HAdV-D36.*

In the complex structure of HAdV-D37 [25], the steric influence of the two residues Y312 and P320 is observable to the same extent, but the overall arrangement of the knob trimer does not allow for the formation of this hydrophobic interaction (mean distance between methyl C atoms: 5.2 Å) (**Figure 10E**). This finding is in line with the preference for Neu5Ac over Neu4,5Ac$_2$ observed for HAdV-D37. As a result, the observed relative twist between Neu5Ac and Neu4,5Ac$_2$ is much less pronounced (**Figure 10C,F**). It is tempting to speculate that the presence of this direct interaction between the three sugars might lead to a cooperative effect that alters the binding kinetics and accounts for the observed binding preference of HAdV-D36, especially in the absence of the strong electrostatic attraction between knob and carbohydrate observed for HADV-D37.

How does the affinity of HAdV-D36 for Neu4,5Ac$_2$ compare to Neu5Ac *in vivo* and *in vitro*?

While our available data consistently indicate that Neu4,5Ac$_2$ binding by HAdV-D36 is stronger than Neu5Ac engagement, we are evaluating the physiological importance if this interaction in ongoing experiments. To this end, we will employ several cell line-based assays that are outlined here: (I) We will probe for the relative attachment inhibition potency of both sialic acid variants in various cell types (A549 cells, human Simpson-Golabi-Behmel syndrome (SGBS) preadipocyte cells, equine dermis (EDERM) cells). HAdV-D36 virions or fiber knobs will be preincubated with increasing concentrations of both sugars and the loss in attachment will be

compared. HAdv-D37 will serve as a reference for these experiments. (II) The dependency of HAdV-D36 virion and FK attachment and infection on surface-Neu4,5Ac$_2$ will be tested with suitable Neu4,5Ac$_2$-expressing animal cell lines. The inhibitory potency of 4-O-acetyl-specific esterases, neuraminidases, and sodium salicylate, (a de-O-acetylating agent), will be used to evaluate the results.

In addition, we are currently investigating methods to assess the relative affinity of the HAdV-D36 FK to both sialic acid variants *in vitro*. Although Neu4,5Ac$_2$ has been clearly demonstrated to interact with the HAdV-D36 FK in this study, STD-NMR experiments with 4-O-Ac-3'SL did not show a detectable difference signal. An absent STD-signal can be caused by a slow ligand off-rate, which seems plausible based on our structural data. Similar findings have been made for the simian B-Lymphotropic polyomavirus (LPyV), which is known to efficiently engage 3'SL [49] but did not give an STD signal (B. S. Blaum, personal communication). Furthermore, we have started initial tests to assess relative affinities using TROSY-NMR. This method is highly sensitive and relies on the assessment of chemical shift perturbations (CSPs) of specific amino acid side chains upon interaction with a ligand and makes it possible to assess affinity data by ligand titration. To this end, we have collected initial 2-dimensional (^{15}N and ^{1}H) spectra of the HAdV-D36 fiber knob alone and with 3'SL as a readily affordable test ligand with presumed low affinity. Despite the size of the knob trimer (~70 kDa), initial tests show interpretable TROSY spectra and distinct chemical shift perturbations upon ligand addition (data not shown).

Difference to Ad37 and other HAdV-D knobs - Is the binding site unique among species D Ads?

To provide additional support for our hypothesis that the relative arrangement of the monomers is the main cause of the difference in Neu4,5Ac$_2$ binding, we replaced the central three amino acid stretch of HAdV-D36 (VSN) with that of HAdV-D37 (YGT) and *vice versa*. Indeed, the mutations assimilated the spatial arrangement at the central cavity (**Figure S9A,D**). As a result, the HAdV-D37 *VSN* mutant restored the triangular hydrophobic contact (**Figure S10**, mean distance between methyl C atoms: 4.4 Å), although the electron density for the sugar was not as pronounced as for the HAdV-D36 wild type. The HAdV-D36 FK *YGT* mutant, on the other hand, displays a widened binding pocket as compared to the wild type. The mutant knob still engages Neu5Ac, but completely lost its ability to bind Neu4,5Ac$_2$ due to the loss of the water-bridged contact with T313*, which distorts the sugar orientation and produces clashes with the gatekeeper residue Y315 as well as P323 and T313 (**Figure S9D-K**). Although the *VSN* trimer-interface configuration is highly conserved among HAdVs of the 'HAdV-D36-like' clade, they are not under general suspicion of engaging Neu4,5Ac$_2$. As an example, the low-seroprevalence type HAdV-D48 [50] exhibits an altered spatial arrangement of the lysine at the position that usually mediates the key contact to the sugar's carboxyl group (**Figure S11B**). This displacement results in a largely distorted binding site and a complete abrogation of Neu5Ac binding, as became apparent from co-crystallization trials. The HAdV-D26 FK belonging to the 'HAdV-D37'-like clade, in turn, exhibits a less displaced lysine residue at the respective position and partially maintained its ability to bind Neu5Ac (**Figure S11C**), albeit with affinities that are likely below physiological conditions (K. Cupelli, personal communication). However, this movement results in a somewhat different binding orientation that would produce a clash with the gatekeeper tyrosine in a Neu4,5Ac$_2$ context. Notably, this movement is caused by a minor rearrangement of the IJ loop and a mutation at position 327, despite the apparent presence of all the known key residues. These examples demonstrate the level of sophistication needed to maintain the three-dimensional arrangement of HAdV-D sialic acid binding sites that are sensitive to subtle alterations affecting the quaternary structure or secondary mutations that have an effect on the binding site topology without directly interfering with the binding. In this light, it is of little surprise that HAdV-D36 is the only HAdV type known to efficiently engage Neu4,5Ac$_2$ to date.

Development of Novel Trivalent Inhibitors

We recently reported the development of sialic acid-based high-affinity inhibitors of HAdV-D37 cell attachment and infection [51]. We tested the efficacy of one of these inhibitors for inhibition of HAdV-D36 attachment. Indeed, the inhibitor compound 17a efficiently inhibited HAdV-D36 binding to A549 cells with an IC_{50} value of 6-10 µM (**Figure S12**). A complex crystal structure of the HAdV-D36 Fk and compound 17a showed that the inhibitor is indeed engaged at the sialic acid binding pocket (data not shown). We are currently synthesizing a new generation of Inhibitors specifically designed for the inhibition of HAdV-D36 glycan binding that was inspired by the binding mode of $Neu4,5Ac_2$ to the HAdV-D36 FK. The usability and efficacy of these compounds critically depends on the relative physiological importance of the yet unidentified protein receptor *in vivo*.

Discussion

The systematic analysis of possible HAdV-D36 attachment factors revealed that virion attachment to various cell types depends on CAR and sialic-acid-containing glycoproteins. The attachment is sensitive to treatment of cells with proteases and neuraminidase. It is furthermore diminished by inhibitors of *O*- and, to a lesser extent, *N*- glycosylation, and can be competed for with soluble sialic acid. However, we were never able to inhibit the binding by more than 80% (**Figure 2D**). Thus, although CAR and Neu5Ac-containing glycoproteins appear to be the major factors for HAdV-D36 cell attachment, there likely are additional factors that can account for about 20% of the binding, at least on A549 cells. In light of the results we obtained for CHO-cell attachment of virions (**Figure S3**) and fiber knobs (**Figure 1C**), it seems plausible that an additional factor is present on CHO cells as well, and that this factor might be responsible for the high overall attachment seen for the virion. However, since knob attachment showed an unambiguous preference for CAR and sialic acid, the additional attachment factor is likely recognized by a different capsid protein. We are planning to test this hypothesis by assessing the combined inhibitory potency of CAR-D1 and Neu5Ac on FK attachment. We have so far been unable to unambiguously identify the additional attachment factor. However, our results indicate that it is probably not among the known set of protein attachment factors for HAdVs. Coagulation factor X, which is known to engage the HAdV hexon and mediate cell uptake, is not present in the system and has been demonstrated to not bind the HAdV-D36 particle (**Figure S2H**). Species D HAdVs, especially HAdV-D09, are known to very efficiently engage α_v integrins by their penton base, to an extent that they can be used for cell attachment [52]. The penton base sequence of HAdV-D36 is similar to that of HAdV-D09, and a similar mechanism seems possible for A549 cells. However, HAdVs are generally not capable of engaging non-human integrins as they would appear on CHO cells.

Our cell-based assays show that HAdV-D36 recognizes CAR with its FK domain, and our structural data indicate that it probably possesses an altered binding interface compared to established CAR-binding knobs such as HAdV-C05 FK. Although we could clearly demonstrate that CAR is functional in supporting HAdV-D36 attachment to different cell types, whether CAR can serve to mediate cell infection, as well, remains to be clarified. In the case of the closely related HAdV-D37, which recognizes CAR with its knob domain and can use it for cell attachment, but not for infection, the stiffness of the fiber shaft does not allow an efficient interplay between CAR and integrins on the cell surface [29]. As discussed above, HAdV-D36 possesses a similar shaft sequence and misses known bending motifs, as well. Therefore, CAR recognition of both viruses might be more important for virus spread after lytic completion of the life cycle than for the actual infection process [53]. This notion agrees well with the finding that both HAdV-D36 and –D37 FKs additionally attach to glycoproteins, which are also much more prominently displayed on the apical side of epithelia than CAR.

Glycan array screening and structural investigations indicate that the rare sialic acid variant Neu4,5Ac_2 is engaged more effectively by the knob of HAdV-D36 than the more common Neu5Ac. We have identified the factors responsible for this preference and identified residues that influence the binding by altering the knob topology. Analysis of the knob phylogeny suggests that these residues are evolutionarily conserved in about half of species D, although Neu4,5Ac_2 binding is likely a specific ability acquired only by HadV-D36 or a small subset of HAdVs of the 'HAdV-D36-like clade'. Given the strong evolutionary pressure on viral capsid proteins, especially receptor-recognizing molecules, it appears unlikely that an interaction with this level of sophistication appears at random. The specific exploitation of sialic acid variants by human or animal viruses is not unprecedented, and the presence or absence of sialic acid derivatives in different organisms determines the tropism of many sialic acid-binding viruses (reviewed in [45]). A prominent example is human influenza C virus, which uses 9-O-acetyl-Neu5Ac as a receptor and possesses a hemagglutinin with an esterase activity as receptor-destroying enzyme [54]. Some toro- and lineage A betacoronaviruses have long been known to specifically use O-acetylated sialic acid variants as receptors on the cell surface *via* their hemagglutinin-esterase (HE) proteins (reviewed in [55]). Although 9-O-acetyl-Neu5Ac is the prototypical receptor for coronaviruses, some virus strains have developed altering specificities. For example, two biotypes of mouse coronaviruses are distinguished. While members of group I use 9-O-acetyl-Neu5Ac, some viruses constituting group II evolved to specifically use Neu4,5Ac_2. As discussed above, mouse hepatitis virus strain S (MHV-S), a murine coronavirus of class II, has been demonstrated to establish this receptor switch by modest changes in both the lectin and esterase domain that allow the formation of a binding pocket optimal for the spacing of its two acetyl groups [47,48]. This arrangement resembles the pocket formed by the protein and two neighboring residues in the HAdV-D36 complex structure. For these viruses, the presence of Neu4,5Ac_2 in the colon and brain is likely a determinant of tropism [56]. Since Neu4,5Ac_2 is abundant on fish skin [42], several piscine viruses including infectious salmon anemia virus (ISAV) use this receptor as well [57]. Neu4,5Ac_2 from horse and guinea pig sera influences the tropism of some influenza A H2 and H3 strains by inhibiting their neuraminidase [58]. In fact, this mode of action is assumed to be one of the physiological functions of Neu4,5Ac_2 [45,59]. Although Neu4,5Ac_2 has not yet been detected in human tissue samples, its occurrence and distribution in humans is still not well understood. The carbohydrate is difficult to analyze histochemically due to the acid/base liability of the 4-O-ester function (reviewed in [45,59]). Suitable detection methods have been developed only recently, and tissue stainings using viral hemagglutinin-esterases have so far not covered all tissues [42,43]. Furthermore, studies in mice suggest that its synthesis is restricted to only a few tissue types [43,60]. In this light, it is possible that the sugar might be expressed in specialized tissues that have not yet been investigated in this regard. From a pathogen's perspective, specifying on sialic acid variants with a restricted expression profile may help to circumnavigate the vast amounts of sialylated non-productive 'decoy'-glycotopes that are present on virtually any tissue in high local concentrations. Bare a receptor-destroying enzyme, it appears as a feasible strategy for HAdV-D36 to bind effectively to a spatially confined epitope while having a low affinity for the more ubiquitous variant. Since no study searching for Neu4,5Ac_2–displaying tissues in humans included adipose tissue [43], we are planning to specifically screen for the presence of Neu4,5Ac_2 on a human preadipocyte cell line (SGBS) and in fat tissues obtained from liposuction samples from obese US citizens.

The preference for Neu4,5Ac_2 displayed by HAdV-D36 and the distribution of the sugar among different species overlap remarkably well with the animal range that is known to be permissive for the virus. It is therefore tempting to speculate that HAdV-D36 might be the first HAdV known to possess an animal reservoir. Following this rationale, it also appears plausible that HAdV-D36 has maintained its ability to engage Neu5Ac and additionally binds a protein receptor. It should not go unnoticed that the presence of an animal host reservoir would directly implicate a route of virus transmission through contact with domestic animals or the consumption of livestock, respectively. Many human viruses are known to infect intermediary animal hosts, and adenoviruses are known to be spread through foodborne oral-fecal transmission [61]. In particular, chickens seem to be a suitable candidate for such a reservoir due to the practice of intensive mass animal farming and the fact that they are one of the main sources of meat in many countries. However, there is no

publicly available data for HAdV-D36 seroprevalence in animals to date. The results obtained from studies of this kind will give further insights into the physiological relevance of Neu4,5Ac_2 for human HAdV-D36 infection.

Materials and Methods

Cells and viruses

Human lung carcinoma cells A549 (a gift from Dr. Alistair Kidd) were grown in Dulbecco's modified Eagle medium (DMEM; Gibco, Paisley, UK) supplemented with 10 % fetal bovine serum (FBS; Thermo Scientific, Cramlington, UK), 20 mM HEPES (Fischer Scientific, Fair Lawn, USA), and 20 U/mL penicillin + 20 µg/mL streptomycin (GE Healthcare, South Logan, USA). Chinese hamster ovary cells CHO-K1 (gift from David Fitzgerald) were grown in Ham's F12 medium (Gibco) supplemented with 10 % FBS (Thermo Scientific) and 20 U/mL penicillin + 20 µg/mL streptomycin (GE Healthcare). HAdV-D36 (strain 275) was used in the experiments. In some experiments HAdV-5 (strain Adenoid 75, source ATCC) and HAdV-37 (Strain 1477) were used as compound controls.

Animal red blood cells used for hemagglutination assays were kindly provided by Agrisera (Vännäs, Sweden). Bovine, chicken, equine, goat, porcine and rabbit blood were collected in sodium citrate. Human venous blood was collected in sodium citrate from three volunteer donors with different blood types according to the AB0 system (A, B, and 0).

Virus production

Confluent A549 cells were split in 175 cm^2 cell culture flasks and incubated overnight in 37 ^oC. The next day, the growth medium was removed and 5 mL of DMEM (Gibco) supplemented with 1 % FBS (Thermo Scientific), 20 mM HEPES (Fischer Scientific) and 20 U/mL penicillin + 20 µg/mL streptomycin (GE Healthcare) were added to each flask. The inoculation material (i.m.; previously prepared from infected A549 cells) was freeze-thawed three times in 37^oC and -80^oC, and 100 µL were added to each flask and incubated for 90 mins. The medium with i.m. was poured off to remove unbound virions and 30 mL of DMEM (Gibco) supplemented with 1 % FBS (Thermo Scientific), 20 mM HEPES (Fischer Scientific) and 20 U/mL penicillin + 20 µg/mL streptomycin (GE Healthcare) were added to each flask. About 72 h post infection, the cells were harvested, pelleted, resuspended in 6 ml of DPH (DMEM, PEST and 20 mM HEPES) and freeze-thawed three times. 1,1,1,2,3,4,4,5,5,5- Decafluoropentane (6ml, Sigma-Aldrich, St Louis, USA) was added to the cell suspension and shaken by hand for 2-3 min. After centrifugation the top layer containing the virus particles was added onto a CsCl gradient with densities 1,27 g/mL, 1,32 g/mL and 1,37 g/mL and ultracentrifuged (SW41Ti rotor, OptimaTM L-80 XP Ultracentrifuge, Beckman Coulter) at 25 000 rpm, 4^oC for 1 h 30 min. The lower band, which contained the virus particles, was harvested and desalted on a NAP column (GE Healthcare, Buckinghamshire, UK) in sterile PBS. Glycerol was added to a final concentration of 10 %, the virion concentration was measured with NanoDrop (ND-1000 Spectrophotometer, Saveen Werner), and the virion solution was aliquoted and stored in -80^oC until further use.

Production of ^{35}S-radiolabelled virions was performed as described previously [62], with the following exceptions: After 24 h of infection, cells were washed twice with sterile phosphate buffered saline (PBS) containing 0,05 % EDTA to get rid of traces of L-cystein and L-methionin and starved for 2 h in DMEM without L-cystein and L-methionin (Gibco) supplemented with 1 % FBS (Thermo Scientific), 20 mM HEPES (Fischer Scientific), 20 U/mL penicillin + 20 µg/mL streptomycin (GE Healthcare) and 4 mM L-glutamine (Gibco). The

isotope ^{35}S (Easy tag express protein labeling mix, Perkin Elmer, Boston, USA) was added thereafter at a concentration of 1 mCi/flask. L-cystein (Sigma-Aldrich) and L-methionin (Sigma-Aldrich) was added to a final concentration of 2 mM at 1 h or 4,5 h after addition of ^{35}S, respectively. L-cystein (Sigma-Aldrich) and L-methionin (Sigma-Aldrich) were added again to a final concentration of 2 mM 24 h after addition of ^{35}S.

Cell-binding assays

Cells were split one day before the experiment. On the day of the experiment cells were detached with PBS containing 0,05 % EDTA and reactivated for 1 h at 37°C in growth media. Cells ($1x10^5$ cells/well) were placed on a V-shaped bottom 96-well plate and washed once with binding buffer (BB: DMEM (Gibco) supplemented with 20 mM HEPES (Fischer Scientific), 20 U/mL penicillin + 20 µg/mL streptomycin (GE Healthcare) and 1 % bovine serum albumin (Roche, Mannheim, Germany)). ^{35}S-labelled viruses (HAdV-D36: 1x109 virions/well, HAdV5: 1x109 virions/well and HAdV-37: 5x108 virions/well) diluted in BB were added to the cells and incubated at 4°C for 1 h. Cells were then pelleted and washed twice with PBS to remove unbound virions. Samples were analyzed by measuring radioactivity in a scintillation counter (1450 Microbeta, Wallac).

The assay were performed with following variations: Binding of HAdV-D36 was analysed to (i) A549, 3T3-L1, SGBS, EKVX, HEK293, FHS, HEP-G2, CHO-K1, CHO-mock, CHO-CAR, CHO-CD46, CHO-CD21, CHO-ICAM-1, CHO-CD55, CHO-Pro5, CHO-Lec2, CHO-2241 and CHO-2241 cells (ii) A549 cells pretreated for 1 h at 37 °C with different concentrations of enzymes (Ficin, proteinase K, Bromelain, V8 protease, heparinase III (*F. heparinum*) and neuraminidase (20 mU/mL, *V. cholerae*). In some experiments the binding of HAdV-D36 virus, pre-incubated 1 h at 4°C with different concentrations of sialic acid (N-acetylneuraminic acid, Dextra Laboratories, Reading, UK), heparin (from porcine intestinal mucosa, Sigma-Aldrich) and compound 17a [51], to A549 cells was analyzed. The data represent values from two individual experiments with duplicate samples in each experiment. Error bars represent mean ± SD. Untreated cells or absence of compounds were used as controls.

Hemagglutination assay

Erythrocytes from blood samples were washed three times and diluted to a 1 % (v/v) solution in PBS. Cells were counted in a Countess II cell counter (Life Technologies). Fifty microliter PBS were added to a round bottom 96-well plate. Virions were added in a range of 1x109 to 3,2x1011 depending on virus identity, followed by 2-fold serial dilution. Fifty microliter of the 1 % cell suspension were added to each well and the plates were incubated 1 h in room temperature. The results were interpreted visually as complete agglutination when a lattice of red blood cells were formed in the well and hemagglutination negative when buttons are formed in the bottom of the wells. If buttons were not running when the plate was tilted they were judged as incomplete agglutination, running buttons were considered definitely negative. To estimate the strength of the interaction between virions and cells, the number of virus particles per cell needed to cause hemagglutination, which is one hemagglutinating unit (vp/HAU) was calculated.

Cloning, purification, and crystallization of HAdV fiber knobs and their complex structures

The HAdV-D37 was purified and crystallized as described previously. The HAdV-D36 FK amplified from HAdV-D36 strain 275 genomic DNA and cloned into a pPROEX htb vector using the following primers: 5'-gcc cat ggg aga ctt agt agc ttg-3' (forward); 5'-cgc ctc gag tca ttc ttg agc gat ata tga gaa ag-3' (reverse). The construct design was analogous to that of the HAdV-D37 FK. The YGT and VSN mutants were prepared by site-directed mutagenesis using a modified Strategene protocol. PCR was performed in 22 cycles, routinely using annealing temperatures from 55-57°C in eight 50 µL setups and the ExactRun Polymerase (Genaxxon). Each setup

contained 5 ng of template plasmid. All eight setups were pooled, incubated with 3µL DpnI (Thermo Fischer Scientific) and subjected to an ethanol precipitation for 1-3 days and resuspended in in H_2O prior to transformation. All construct sequences were verified by Sanger-sequencing (performed at MWG Eurofins) and comparison to GenBank entries GQ384080.1 (HAdV_D36 complete genome) and ABK59080.1 (HAdV-D37 fiber). Expression and purification of the HAdV-/D36D37 FKs was performed essentially as reported previously for HAdV-D37 [51]. For crystallization, the HAdV D37 FK was concentrated to 12-13 mg/ml, and the HAdV-D36 FK to 8.0-8.2 mg/ml. The HAdV-D37 FK was crystallized as described previously. The HAdV-D36 FK was crystallized by the hanging drop vapor diffusion method using 23-26% PEG 3,350, 175-200 mNH4Ac, 0.1 M Bis-Tris pH 5.5 and an initial drop size of 1+1µL at 4°C. Complex crystals were prepared by co-crystallization with 10 mM of the respective carbohydrate compound. Crystals were flash frozen in liquid nitrogen without cryoprotection. Data collection was carried out at the beamlines X06SA and X06DA at the SLS (Villigen, Switzerland). Data were processed with XDS [63,64]. For the initial structure of HAdV-D36 wt, the phase problem was solved using Molrep and a HAdV-D37 FK-based model prepared with CHAINSAW [65]. If possible, subsequent structures were solved by simple rigid body refinement and simulated annealing in phenix.refine [66]. Otherwise, the phase problem was solved with Molrep [67] or Phaser [68]. Refinement was carried out using Coot [69] for real space refinement and phenix.refine with automatically determined non-crystallographic symmetry (NCS) restraints and isotropic B-Factor refinement. At high enough resolution, translation-libration-screw (TLS) tensors (all structures in **Table S1**) or anisotropic B-Factors were refined. Ligands were unambiguously placed into F_o-F_c difference maps and refined using restraints from the CCP4 monomer library. Waters were located using an automated algorithm in Coot. Simulated annealing omit maps were calculated with phenix.refine and FFT [70].

Cloning.purification, crystallization, and refinement of HAdV-D26 and –D48 FKs have been performed as described elsewhere [71].

Superpositioning of complex Structures

Superpositions of fiber knobs and the corresponding complex structures were performed using the 'align' algorithm in PyMol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.). An important distinction has to be made between aligning all atoms in the knob trimer and alignment based on single chains.

Complexation of HAdV-D36 FK with CAR-D1 and CD46-D4

The complexation experiments were carried out as described previously [28] with minor alterations. Purified HAdV-D36 FK was mixed with CAR-D1 (laboratory stock produced by A. Thor) and CD46 D4 (purified as described previously [72]) in different molar ratios in their respective buffers and incubated for 20 min at 4°C. 30 µL were subjected to analytical size exclusion chromatography using a Superdex 200 column on an ETTAN system (GE Healthcare, Sweden) using a standard buffer containing 30 mM Tris-HCl (pH 7.5) and 150 mM NaCl. The peaks were analysed by comparison to a standard laboratory calibration curve and, if possible, by SDS-PAGE. In the case of CAR-D1, both HAdV-D36 FK and CAR-D1 alone were run at the same concentrations for comparison. In the case of CD46 D4, a HAdV-B11 FK complexed with CD46 D4 was run as a positive control instead of CD46 D4 alone.

Differential Scanning Fluorimetry (DSF)

Samples were diluted to a concentration of 0.1 mg/mL in a buffer containing 150 mM NaCl, 20 mM imidazole, and 30 mM Tris-HCl, pH 7.5. The experiments were performed using a sample volume of 10 µl per capillary. The capillaries were filled directly from respective solution. An initial fluorescence scan of loaded capillaries was performed at 20 °C, low gain sensitivity in order to ensure that samples were within the optimal concentration range. The DSF experiment was performed in a single run by heating all samples from 20°C to 95 °C with 1°C/min. To determine the melting point (T_m), the shift in native tryptophan fluorescence was monitored by plotting changes in the emission at 350 and 330 nm.

TROSY NMR spectroscopy

Wildtype HAdV-D36 FK was expressed in 0.5 L M9 minimal medium and purified as described above. Proton-nitrogen correlated TROSY spectra were recorded at 25°C on a Bruker AVIII-600 spectrometer with a room temperature probe head. The Ad36 sample was prepared in 30 mM Tris, pH 7.5 and 150 mM NaCl, which was based on 90% H2O/10% D2O at a protein concentration of 0.5 mM Ad36 monomer. In addition to a reference spectrum, spectra were recorded at a 10- and 40-fold molar excess of sialyl lactose. Before recording the data in presence of 40-fold excess of sialyl lactose, the buffer conditions were changed to 60 mM Tris pH 7.5, 150 mM NaCl to provide buffer capacity for the acidic ligand.NMR data were processed using the NMRPipe-NMRDraw software suite and figures displaying NMR spectra were produced using NMRView (www.onemoonscientific.com).

Saturation transfer difference NMR

NMR spectra were recorded at 283 K using 3 mm tubes (200 µL sample volume) and a Bruker AVIII-600 spectrometer equipped with a room temperature probe head and processed with TOPSPIN 3.0 (Bruker). The sample contained 1 mM of 4-O-Ac-3'SL and 20 µM of HAdV-D36 FK protein (monomeric concentration). The proteins were buffer-exchanged prior to NMR experiments to 20 mM potassium phosphate pH 7.4, 150 mM NaCl in D_2O and the glycans were subsequently added from concentrated stock solutions in D_2O. Off- and on-resonance irradiation frequencies were set to -30.0 ppm and 7.0 ppm, respectively. The irradiation power of the selective pulses was 57 Hz, the saturation time was 2 s, and the total relaxation delay was 3 s. A 50 ms continuous-wave spin-lock pulse with a strength of 3.2 kHz was employed to suppress residual protein signals. A total number of 512 scans and 10,000 points were collected, and spectra were multiplied with a sine-bell squared window function prior to Fourier transformation.

Glycan array screening

His-tagged HAdV-D36 FK was analzyed in two different array sets: the big screening array set containing 492 sequence-defined lipid-linked oligosaccharide probes (Glycosciences Array Set 32–39) and a focused sialyl oligosaccharide array set containing 8 oligosaccharide probes in dose response format (see **Figure 7B,C**). The probes were robotically printed in duplicate on nitrocellulose-coated glass slides at the levels indicated using a non-contact instrument [73]. The microarray analyses were performed essentially as described previously [74]. In brief, microarrays were blocked in 0.3% (v/v) Blocker Casein (Pierce), 0.3% (w/v) bovine serum albumin (Sigma A8577) in Hepes buffered saline (5 mM Hepes, pH 7.4, 150 mM NaCl, 5 mM CaCl2). The His-Ad36 fiber knob was tested as protein-antibody complexes that were prepared by preincubating fiber knob protein with mouse monoclonal antipolyhistidine and biotinylated anti-mouse IgG antibodies (both from Sigma) at a ratio of 4:2:1 (by weight) and diluted in blocking solution to provide a final fiber knob concentration of 150 µg/ml. Binding was detected using Alexa Fluor-647-labeled streptavidin from Molecular Probes (1 ug/ml). In the small

sialyl oligosaccharide array, His-Ad36 fiber knob was also tested without precomplexation for comparison. The oligosaccharide probes are all lipid-linked (neoglycolipids (NGLs) or glycosylceramides) and are from the collection assembled in the course of research in the Glycosciences Laboratory. For definition of the lipid moieties of the probes, please see https://glycosciences.med.ic.ac.uk/docs/lipids.

Purification of 4-O-acetyl-3'-sialyllactose (4-O-Ac-3'SL) from echidna milk oligosaccharides (EMOs)

4-O-Ac-3'SL was isolated from an oligosaccharide mixture derived from the milk of the Australian short-beaked echidna. EMOs [75] received from Professor Tadasu Urashima (Obihiro University) were fractionated by gel filtration chromatography on a Bio-Gel P4 column (16 x 90 cm) with elution by ammonium acetate (0.2 M) and detection by refractive index (Figure S13A). As detected by hexose assay, fractions 5-10 contain carbohydrate materials. The pooled fraction F6 was further fractionated by normal phase HPLC (Amide column) with elution by a gradient of H_2O/ACN containing 0.05 mM phosphate and detection at UV 196 nm (Figure S13B) and the subfraction 2 (F6-2) was analysed by negative-ion electrospray mass spectrometry. The spectrum (Figure S13D) indicated that F6-2 contained the trisaccharide of interest as the major component ([M-H]- at m/z 674) but also contain a minor component disialylated hexasaccharide ([M-2H]2- at m/z 664 and [M-H]- at m/z 1330). Further HPLC purification using the same system but with an amine column resolved the two component (Figure S13C) and the mass spectra identified F6-2a as the sialylated lactose with an acetyl on the 4-O-position of NeuAc (Figure S13E,G) and F6-2d as the disialylated hexsasaccharide containing a single OAc (Figure S13F). Quantitation was carried out by microscale orcinol assay as essentially as described [76].

Synthesis of differentially O-acetylated α-2-O-methyl-Neu5Ac variants

Methyl 5-acetamido-4-*O*-acetyl-3,5-dideoxy-D-glycero-a-D-galacto-2-nonulopyranosidonic acid (2-O-me-Neu4,5Ac_2) and methyl 5-acetamido-4,9-di-*O*-acetyl-3,5-dideoxy-D-glycero-a-D-galacto-2-nonulopyranosidonic acid (2-O-me-Neu4,5,9Ac_3) were synthesized essentially as described previously [77].

Acknowledgments and Funding

The authors are thankful to Dr. Ancilla Neu and Prof. Remco Sprangers for providing the instrumentation and practical advice for TROSY-NMR measurements, and to Alexandra Thor and Michael Braun for providing purified CAR-D1 and CD46-D4, respectively. We are also grateful to the staff at the Swiss Light Source (SLS, Villigen, Switzerland) for beam time and beamline assistance. This work has been funded by the SFB 685 of the German Research Foundation (TS), the Swedish Research Council (2013-2753 and 2013-8616, NA), Knut & Alice Wallenberg Foundation (KAW 2013.0019, NA), and The Swedish Cancer Society (CAN 2011/340, NA). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Wigand R, Gelderblom H, Wadell G (1980) New human adenovirus (candidate adenovirus 36), a novel member of subgroup D. Arch Virol 64: 225-233.
2. Dhurandhar NV, Israel BA, Kolesar JM, Mayhew GF, Cook ME, et al. (2000) Increased adiposity in animals due to a human virus. Int J Obes Relat Metab Disord 24: 989-996.
3. Dhurandhar NV, Kulkarni P, Ajinkya SM, Sherikar A (1992) Effect of adenovirus infection on adiposity in chicken. Vet Microbiol 31: 101-107.
4. Dhurandhar NV, Whigham LD, Abbott DH, Schultz-Darken NJ, Israel BA, et al. (2002) Human adenovirus Ad-36 promotes weight gain in male rhesus and marmoset monkeys. J Nutr 132: 3155-3160.
5. Dhurandhar NV, Israel BA, Kolesar JM, Mayhew G, Cook ME, et al. (2001) Transmissibility of adenovirus-induced adiposity in a chicken model. Int J Obes Relat Metab Disord 25: 990-996.
6. Pasarica M, Shin AC, Yu M, Ou Yang HM, Rathod M, et al. (2006) Human adenovirus 36 induces adiposity, increases insulin sensitivity, and alters hypothalamic monoamines in rats. Obesity (Silver Spring) 14: 1905-1913.
7. Vangipuram SD, Sheele J, Atkinson RL, Holland TC, Dhurandhar NV (2004) A human adenovirus enhances preadipocyte differentiation. Obes Res 12: 770-777.
8. Ponterio E, Gnessi L (2015) Adenovirus 36 and Obesity: An Overview. Viruses 7: 3719-3740.
9. Xu MY, Cao B, Wang DF, Guo JH, Chen KL, et al. (2015) Human Adenovirus 36 Infection Increased the Risk of Obesity: A Meta-Analysis Update. Medicine (Baltimore) 94: e2357.
10. Ponterio E, Cangemi R, Mariani S, Casella G, De Cesare A, et al. (2015) Adenovirus 36 DNA in human adipose tissue. Int J Obes (Lond) 39: 1761-1764.
11. Pasarica M, Mashtalir N, McAllister EJ, Kilroy GE, Koska J, et al. (2008) Adipogenic human adenovirus Ad-36 induces commitment, differentiation, and lipid accumulation in human adipose-derived stem cells. Stem Cells 26: 969-978.
12. Rogers PM, Fusinski KA, Rathod MA, Loiler SA, Pasarica M, et al. (2008) Human adenovirus Ad-36 induces adipogenesis via its E4 orf-1 gene. Int J Obes (Lond) 32: 397-406.
13. Atkinson RL, Dhurandhar NV, Allison DB, Bowen RL, Israel BA, et al. (2005) Human adenovirus-36 is associated with increased body weight and paradoxical reduction of serum lipids. Int J Obes (Lond) 29: 281-286.
14. Wang ZQ, Cefalu WT, Zhang XH, Yu Y, Qin J, et al. (2008) Human adenovirus type 36 enhances glucose uptake in diabetic and nondiabetic human skeletal muscle cells independent of insulin signaling. Diabetes 57: 1805-1813.
15. Rogers PM, Mashtalir N, Rathod MA, Dubuisson O, Wang Z, et al. (2008) Metabolically favorable remodeling of human adipose tissue by human adenovirus type 36. Diabetes 57: 2321-2331.
16. Salehian B, Forman SJ, Kandeel FR, Bruner DE, He J, et al. (2010) Adenovirus 36 DNA in adipose tissue of patient with unusual visceral obesity. Emerg Infect Dis 16: 850-852.
17. Atkinson RL (2008) Could viruses contribute to the worldwide epidemic of obesity? Int J Pediatr Obes 3 Suppl 1: 37-43.
18. van Ginneken V, Sitnyakowsky L, Jeffery JE (2009) "Infectobesity: viral infections (especially with human adenovirus-36: Ad-36) may be a cause of obesity. Med Hypotheses 72: 383-388.
19. Dhurandhar NV (2001) Infectobesity: obesity of infectious origin. J Nutr 131: 2794S-2797S.
20. Pasarica M, Dhurandhar NV (2007) Infectobesity: obesity of infectious origin. Adv Food Nutr Res 52: 61-102.
21. Yamada T, Hara K, Kadowaki T (2012) Association of adenovirus 36 infection with obesity and metabolic markers in humans: a meta-analysis of observational studies. PLoS One 7: e42031.
22. Aoki K, Tagawa Y (2002) A twenty-one year surveillance of adenoviral conjunctivitis in Sapporo, Japan. Int Ophthalmol Clin 42: 49-54.
23. Arnberg N, Kidd AH, Edlund K, Nilsson J, Pring-Akerblom P, et al. (2002) Adenovirus type 37 binds to cell surface sialic acid through a charge-dependent interaction. Virology 302: 33-43.
24. Nilsson EC, Storm RJ, Bauer J, Johansson SM, Lookene A, et al. (2011) The GD1a glycan is a cellular receptor for adenoviruses causing epidemic keratoconjunctivitis. Nat Med 17: 105-109.
25. Burmeister WP, Guilligay D, Cusack S, Wadell G, Arnberg N (2004) Crystal structure of species D adenovirus fiber knobs and their sialic acid binding sites. J Virol 78: 7727-7736.
26. Cupelli K, Stehle T (2011) Viral attachment strategies: the many faces of adenoviruses. Curr Opin Virol 1: 84-91.
27. Lawrence MC, Colman PM (1993) Shape complementarity at protein/protein interfaces. J Mol Biol 234: 946-950.

28. Seiradake E, Lortat-Jacob H, Billet O, Kremer EJ, Cusack S (2006) Structural and mutational analysis of human Ad37 and canine adenovirus 2 fiber heads in complex with the D1 domain of coxsackie and adenovirus receptor. J Biol Chem 281: 33704-33716.
29. Wu E, Pache L, Von Seggern DJ, Mullen TM, Mikyas Y, et al. (2003) Flexibility of the adenovirus fiber is required for efficient receptor interaction. J Virol 77: 7225-7235.
30. Persson BD, Muller S, Reiter DM, Schmitt BB, Marttila M, et al. (2009) An arginine switch in the species B adenovirus knob determines high-affinity engagement of cellular receptor CD46. J Virol 83: 673-686.
31. Robinson CM, Singh G, Lee JY, Dehghan S, Rajaiya J, et al. (2013) Molecular evolution of human adenoviruses. Sci Rep 3: 1812.
32. Robinson CM, Seto D, Jones MS, Dyer DW, Chodosh J (2011) Molecular evolution of human species D adenoviruses. Infect Genet Evol 11: 1208-1217.
33. Hierholzer JC, Wigand R, Anderson LJ, Adrian T, Gold JW (1988) Adenoviruses from patients with AIDS: a plethora of serotypes and a description of five new serotypes of subgenus D (types 43-47). J Infect Dis 158: 804-813.
34. Al Qurashi YM, Alkhalaf MA, Lim L, Guiver M, Cooper RJ (2012) Sequencing and phylogenetic analysis of the hexon, fiber, and penton regions of adenoviruses isolated from AIDS patients. J Med Virol 84: 1157-1165.
35. Ismail AM, Lee J, Dyer DW, Seto D, Rajaiya J, et al. (2016) Selection pressure in the human adenovirus fiber knob drives cell specificity in epidemic keratoconjunctivitis. J Virol.
36. Arnold J, Janoska M, Kajon AE, Metzgar D, Hudson NR, et al. (2010) Genomic characterization of human adenovirus 36, a putative obesity agent. Virus Res 149: 152-161.
37. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7: 539.
38. Iwersen M, Dora H, Kohla G, Gasa S, Schauer R (2003) Solubilisation and properties of the sialate-4-O-acetyltransferase from guinea pig liver. Biol Chem 384: 1035-1047.
39. Baumann AM, Bakkers MJ, Buettner FF, Hartmann M, Grove M, et al. (2015) 9-O-Acetylation of sialic acids is catalysed by CASD1 via a covalent acetyl-enzyme intermediate. Nat Commun 6: 7673.
40. Miyoshi I, Higashi H, Hirabayashi Y, Kato S, Naiki M (1986) Detection of 4-O-acetyl-N-glycolylneuraminyl lactosylceramide as one of tumor-associated antigens in human colon cancer tissues by specific antibody. Mol Immunol 23: 631-638.
41. Zanetta JP, Pons A, Iwersen M, Mariller C, Leroy Y, et al. (2001) Diversity of sialic acids revealed using gas chromatography/mass spectrometry of heptafluorobutyrate derivatives. Glycobiology 11: 663-676.
42. Aamelfot M, Dale OB, Weli SC, Koppang EO, Falk K (2014) The in situ distribution of glycoprotein-bound 4-O-Acetylated sialic acids in vertebrates. Glycoconj J 31: 327-335.
43. Langereis MA, Bakkers MJ, Deng L, Padler-Karavani V, Vervoort SJ, et al. (2015) Complexity and Diversity of the Mammalian Sialome Revealed by Nidovirus Virolectins. Cell Rep 11: 1966-1978.
44. Claus H, Borrow R, Achtman M, Morelli G, Kantelberg C, et al. (2004) Genetics of capsule O-acetylation in serogroup C, W-135 and Y meningococci. Mol Microbiol 51: 227-239.
45. Wasik BR, Barnard KN, Parrish CR (2016) Effects of Sialic Acid Modifications on Virus Binding and Infection. Trends Microbiol.
46. Urashima T, Inamori H, Fukuda K, Saito T, Messer M, et al. (2015) 4-O-Acetyl-sialic acid (Neu4,5Ac2) in acidic milk oligosaccharides of the platypus (Ornithorhynchus anatinus) and its evolutionary significance. Glycobiology 25: 683-697.
47. Bakkers MJ, Zeng Q, Feitsma LJ, Hulswit RJ, Li Z, et al. (2016) Coronavirus receptor switch explained from the stereochemistry of protein-carbohydrate interactions and a single mutation. Proc Natl Acad Sci U S A 113: E3111-3119.
48. Langereis MA, Zeng Q, Heesters BA, Huizinga EG, de Groot RJ (2012) The murine coronavirus hemagglutinin-esterase receptor-binding site: a major shift in ligand specificity through modest changes in architecture. PLoS Pathog 8: e1002492.
49. Neu U, Khan ZM, Schuch B, Palma AS, Liu Y, et al. (2013) Structures of B-lymphotropic polyomavirus VP1 in complex with oligosaccharide ligands. PLoS Pathog 9: e1003714.
50. Abbink P, Lemckert AA, Ewald BA, Lynch DM, Denholtz M, et al. (2007) Comparative seroprevalence and immunogenicity of six rare serotype recombinant adenovirus vaccine vectors from subgroups B and D. J Virol 81: 4654-4663.
51. Caraballo R, Saleeb M, Bauer J, Liaci AM, Chandra N, et al. (2015) Triazole linker-based trivalent sialic acid inhibitors of adenovirus type 37 infection of human corneal epithelial cells. Org Biomol Chem 13: 9194-9205.
52. Arnberg N, Kidd AH, Edlund K, Olfat F, Wadell G (2000) Initial interactions of subgenus D adenoviruses with A549 cellular receptors: sialic acid versus alpha(v) integrins. J Virol 74: 7691-7693.

53. Arnberg N (2012) Adenovirus receptors: implications for targeting of viral vectors. Trends Pharmacol Sci 33: 442-448.
54. Rogers GN, Herrler G, Paulson JC, Klenk HD (1986) Influenza C virus uses 9-O-acetyl-N-acetylneuraminic acid as a high affinity receptor determinant for attachment to cells. J Biol Chem 261: 5947-5951.
55. Smits SL, Gerwig GJ, van Vliet AL, Lissenberg A, Briza P, et al. (2005) Nidovirus sialate-O-acetylesterases: evolution and substrate specificity of coronaviral and toroviral receptor-destroying enzymes. J Biol Chem 280: 6933-6941.
56. Kazi L, Lissenberg A, Watson R, de Groot RJ, Weiss SR (2005) Expression of hemagglutinin esterase protein from recombinant mouse hepatitis virus enhances neurovirulence. J Virol 79: 15064-15073.
57. Hellebo A, Vilas U, Falk K, Vlasak R (2004) Infectious salmon anemia virus specifically binds to and hydrolyzes 4-O-acetylated sialic acids. J Virol 78: 3055-3062.
58. Rogers GN, Pritchett TJ, Lane JL, Paulson JC (1983) Differential sensitivity of human, avian, and equine influenza A viruses to a glycoprotein inhibitor of infection: selection of receptor specific variants. Virology 131: 394-408.
59. Mandal C, Schwartz-Albiez R, Vlasak R (2015) Functions and Biosynthesis of O-Acetylated Sialic Acids. Top Curr Chem 366: 1-30.
60. Rinninger A, Richet C, Pons A, Kohla G, Schauer R, et al. (2006) Localisation and distribution of O-acetylated N-acetylneuraminic acids, the endogenous substrates of the hemagglutinin-esterases of murine coronaviruses, in mouse tissue. Glycoconj J 23: 73-84.
61. Kingsley DH (2016) Emerging Foodborne and Agriculture-Related Viruses. Microbiol Spectr 4.
62. Johansson SM, Nilsson EC, Elofsson M, Ahlskog N, Kihlberg J, et al. (2007) Multivalent sialic acid conjugates inhibit adenovirus type 37 from binding to and infecting human corneal epithelial cells. Antiviral Res 73: 92-100.
63. Kabsch W (2010) Integration, scaling, space-group assignment and post-refinement. Acta Crystallogr D Biol Crystallogr 66: 133-144.
64. Kabsch W (2010) Xds. Acta Crystallogr D Biol Crystallogr 66: 125-132.
65. Schwarzenbacher R, Godzik A, Grzechnik SK, Jaroszewski L (2004) The importance of alignment accuracy for molecular replacement. Acta Crystallogr D Biol Crystallogr 60: 1229-1236.
66. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, et al. (2002) PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr 58: 1948-1954.
67. Vagin A, Teplyakov A (1997) MOLREP: an automated program for molecular replacement. J Appl Cryst: 1022-1025.
68. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, et al. (2007) Phaser crystallographic software. J Appl Crystallogr 40: 658-674.
69. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66: 486-501.
70. Ten Eyck LF (1985) Fast Fourier transform calculation of electron density maps. Methods Enzymol 115: 324-337.
71. Hassemer T (2013) Structural Analysis of Human Adenovirus 26 and 48 Fiber Knobs and their Interactions with Sialic Acid Ligands [Diploma Thesis]. Tübingen: University of Tübingen.
72. Persson BD, Schmitz NB, Santiago C, Zocher G, Larvie M, et al. (2010) Structure of the extracellular portion of CD46 provides insights into its interactions with complement proteins and pathogens. PLoS Pathog 6: e1001122.
73. Liu Y, Childs RA, Palma AS, Campanero-Rhodes MA, Stoll MS, et al. (2012) Neoglycolipid-based oligosaccharide microarray system: preparation of NGLs and their noncovalent immobilization on nitrocellulose-coated glass slides for microarray analyses. Methods Mol Biol 808: 117-136.
74. Neu U, Maginnis MS, Palma AS, Ströh LJ, Nelson CD, et al. (2010) Structure-function analysis of the human JC polyomavirus establishes the LSTc pentasaccharide as a functional receptor motif. Cell Host Microbe 8: 309-319.
75. Messer M, Kerry KR (1973) Milk carbohydrates of the echidna and the platypus. Science 180: 201-203.
76. Chai W, Stoll MS, Galustian C, Lawson AM, Feizi T (2003) Neoglycolipid technology: deciphering information content of glycome. Methods Enzymol 362: 160-195.
77. Knibbs RN, Osborne SE, Glick GD, Goldstein IJ (1993) Binding determinants of the sialic acid-specific lectin from the slug Limax flavus. J Biol Chem 268: 18524-18531.
78. Bachmann P (2016) Structural Insights into the Sialic Acid Binding Modes of Human Adenovirus D-36 and D-37 Fiber Knob Domains [Bachelor Thesis]. Tübingen: University of Tübingen.

79. Pfenning V (2014) Structural Investigations of the Human Adenovirus 36 Fiber Knob and its Candidate Receptors [Bachelor Thesis]. Tübingen: University of Tübingen.
80. Smart OSW, T. O.; Sharff, A.; Flensburg, C.; Keller, P.; Paciorek, W.; Vonrhein, C.; Bricogne, G. (2011) grade, version 1.102. Cambridge, United Kingdom: Global Phasing Ltd.

Supplementary Information

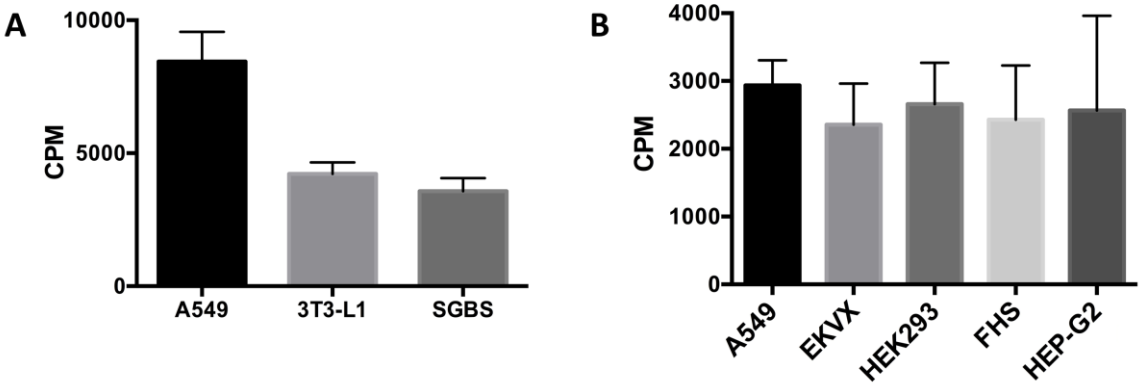


Figure S1 |HAdV-D36 binds better to A549 cells than to other cell lines. Different cell lines were tested for their ability to bind ^{35}S-labeled HAdV-D36 virions on their surface. (A) *Binding of ^{35}S-labelled HAdV-D36 virions to A549, 3T3-L1 (mouse mouse fibroblasts that can differentiate into an adipocyte-like phenotype) and SGBS (human Simpson-Golabi-Behmel syndrome preadipocyte cell line) cells. (b) Binding of ^{35}S-labelled HAdV-D36 virions to different human derived cell lines A549 (adenocarcinomic human alveolar basal epithelial cells), EKVX (human lung adenocarcinoma cell line), HEK293 (human embryonic kidney 293 cells), FHS (human small intestine cell line), HEP-G2 (human hepatocellular carcinoma cell line). All experiments were performed three times with duplicate samples. Error bars are representing mean +_SD.*

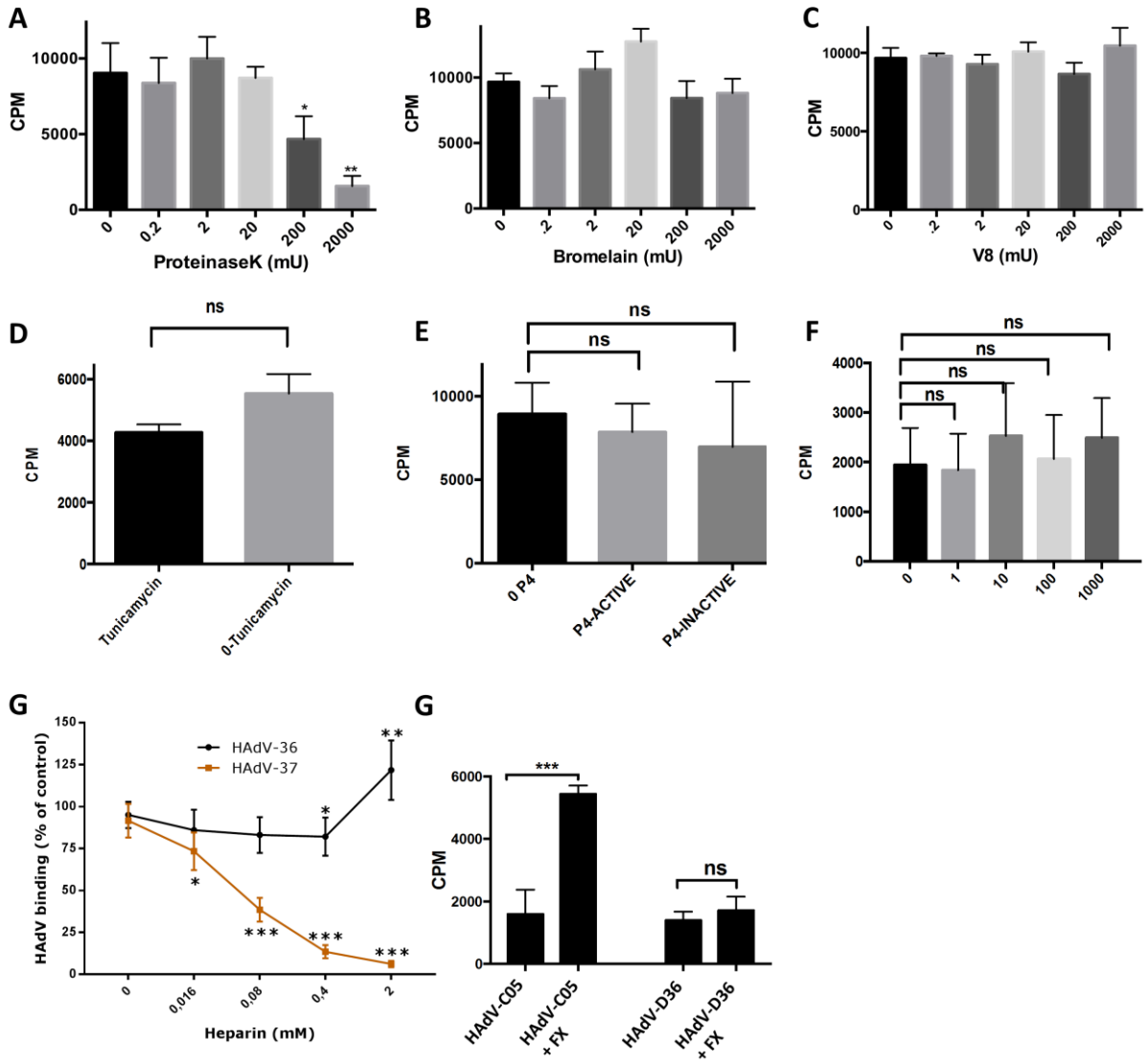


Figure S2 |Evaluation of HAdV-D36 binding to A549 cells. *Binding of ^{35}S-labelled HAdV-D36 virions to A549 cells treated with different proteases (**A**) Proteinase K (a broad-spectrum serine-protease), (**B**) Bromelain (a mixture of pineapple cysteine proteases), (**C**) V8 (an endoproteinase Glu-C serine protease cleaving after glutamic acid), (**D**) tunicamycin (an inhibitor of N-glycoslylation), (**E**), P4 (DL-threo-1-phenyl-2-palmitoylamino-3-pyrrolidino-1-propanol, an inhibitor of glycolipid synthesis) and (F) heparinase III (which cleaves heparan sulfate from cell surface). **G** Attachment inhibition assay with soluble heparin (M_w approx. 21000). HAdV-D37 was used as a positive control (N. Chandra, personal communication). (H) ^{35}S-labeled HAdV-D36 virion binding to A549 cells after virion preincubation with physiological concentrations of coagulation factor X (FX: 10 µg/ml). X-axis represents different concentrations of enzymes, treatment and concentrations of heparin. In figure A-F and H, y-axis shows the amount of virus particles bound to cells AND represented as CPM (count per minute). All experiments were performed three times with duplicate samples. Error bars are representing mean +_ SD. n.s = not significant, * P of < 0.05, ** P of < 0.01 and *** P of < 0.001.*

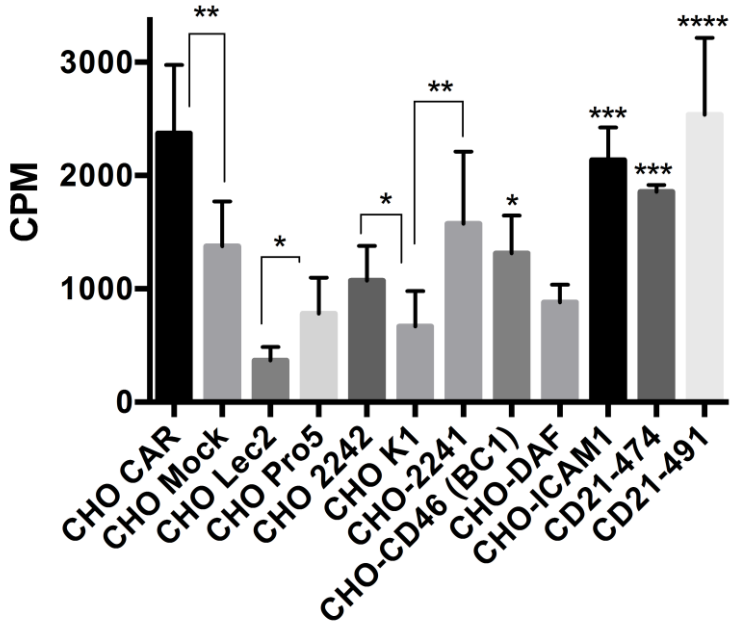


Figure S3 | Binding of ^{35}S-labelled HAdV-D36 virions to different CHO cell lines expressing different human cell surface proteins. *CHO-K1 and CHO-mock cells represent control cell lines. Y-axis shows the amount of knobs bound to cells, represented as CPM (counts per minute). All experiments were performed three times with duplicate samples. Error bars represent mean ± SD. * P of < 0.05 and ** P of < 0.*

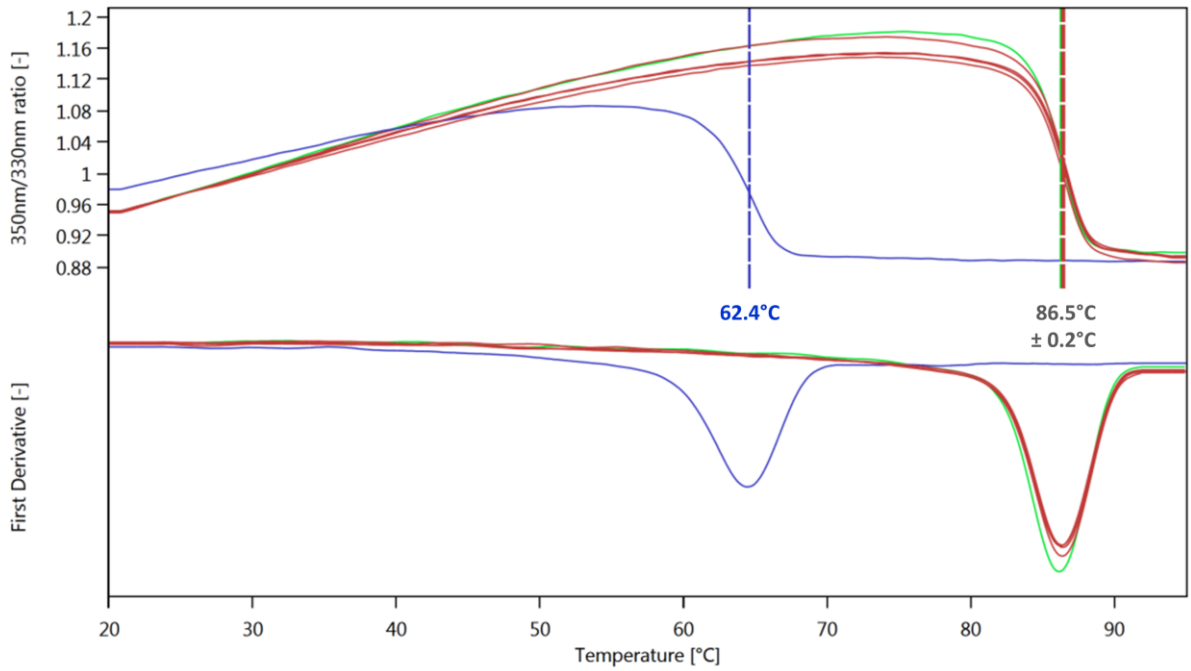


Figure S4 |Differential Scanning Fluorimetry of the HAdV-D36FK. *The scan was performed using a Prometheus NT.48. The ratio of the fluorescence at 350 nm and 330 nm is plotted against the temperature (upper figure). The first derivative calculation can be deduced to determine the Tm (lower figure). At pH 7.5 (red and green curves, n=5), a Tm of 86.3-86.5°C was observed. At a pH of 2-3 (blue), the FK stability was lowered, but still high (Tm=62.4°C). The experiment was performed in the course of a Prometheus demonstration by Dr. Fabian Zehender (NanoTemper Technologies GmbH).*

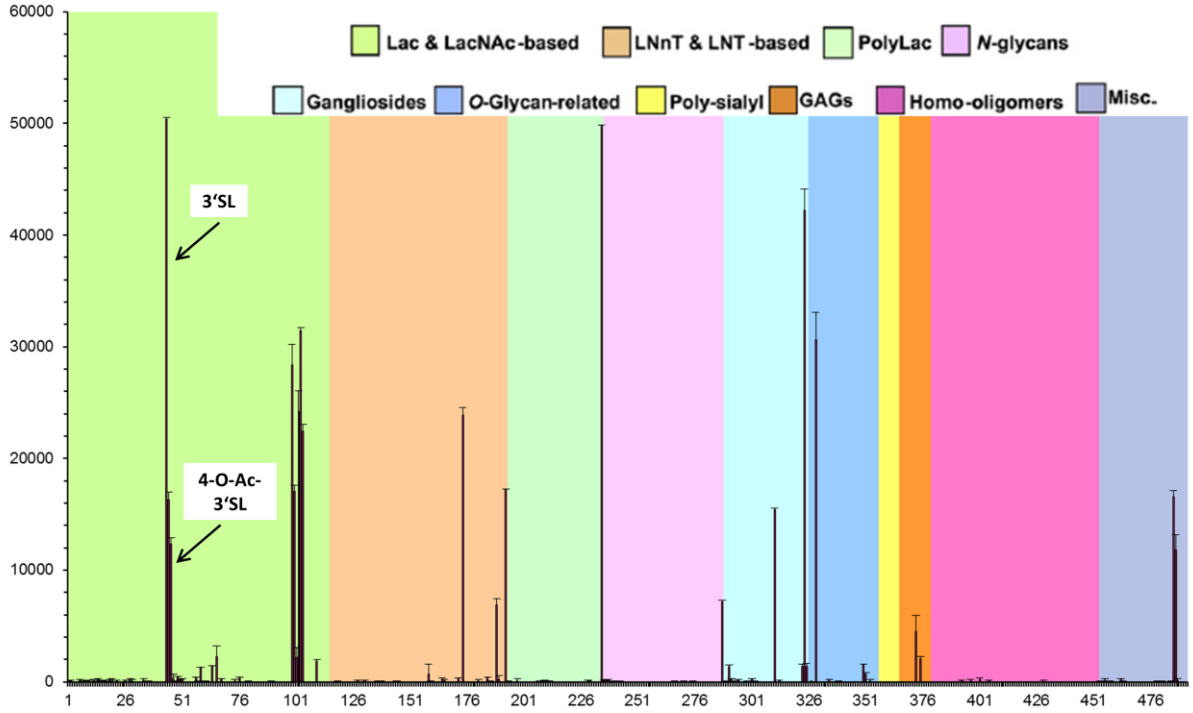
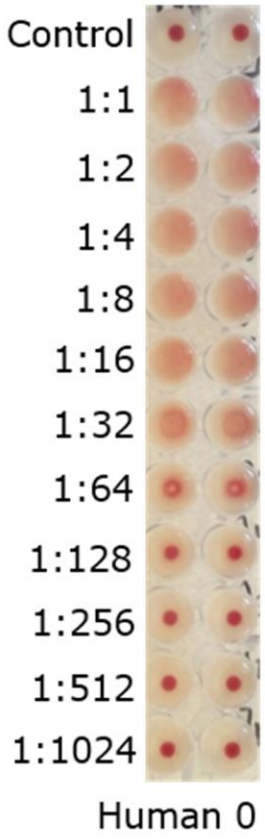


Figure S5 | Glycan array screening for the HAdV-D37. *A large glycan array containing 498 glycans (see legend on the upper right) was in good overall agreement with previous findings [24].*

Table S1 | Excerpt of the fluorescence signals obtained for large arrays of HAdV-D36 and HAdV-D37.

Glycan No.	Probe	Structure	HAdV-D36	HAdV-D37
44	NeuAcα-(3')Lac	NeuAcα-3Galβ-4Glc-DH	-	50,438
45	NeuAcα-(3')Lac-AO	NeuAcα-3Galβ-4Glc-AO	-	16,272
46	Neu4,5Ac-(3')Lac	(4-OAc)NeuAcα-3Galβ-4Glc-DH	7,279	12,366
47	Neu4,5Ac-(3')Lac-AO	(4-OAc)NeuAcα-3Galβ-4Glc-AO	107	297
107	Neu5,9Ac-(6')LN	(9-OAc)NeuAcα-6Galβ-4GlcNAc-DH	-	-
114	GSC-513	(9-OAc)NeuAcα-3Galβ-3GlcNAcβ-C30 | Fucα-4	-	-
115	GSC-511	(9-OAc)NeuAcα-3Galβ-4GlcNAcβ-C30 | Fucα-3	-	-
311	GD1a-hexa	NeuAcα-3Galβ-3GalNAcβ-4Galβ-4Glc-DH | NeuAcα-3	-	15,425

A



B

Virus	Vp/cells (human RBCs)
HAdV-D37	153
HAdV-D36	2500

Figure S6 |Hemagglutination potency of HAdV-D36 and HAdV-D37. ***A*** *Depiction of the hemagglutination experiment. Human blood was diluted as indicated in the methods.* ***B*** *Quantification of the number of virus particles per cell required for complete agglutination of human red blood cells.*

Table S1 | Data collection and refinement statistics of selected datasets. *The structures are not yet completely refined. All other structures are listed in [71], [78],[79].*

	HAdV-D36 FK + 4-O-Ac-3'SL	HAdV-D36 FK + 2-O-me-Neu4,5Ac$_2$	HAdV-D36 FK + 2-O-me-Neu4,5,9Ac$_3$	HAdV-D37 FK + 2-O-me-Neu4,5Ac$_2$
Data Collection				
Beamline	SLS, X06DA	SLS, X06DA	SLS, X06DA	SLS, X06DA
Space Group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_1$
Cell Dimensions				
a, b, c [Å]	51.13, 98.64, 99.83	59.46, 99.16, 110.99	59.59, 99.21, 111.04	56.88, 66.73, 73.95
α, β, γ [°]	90, 90, 90	90, 90, 90	90,90,90	90, 97.75, 90
Resolution [Å]	**50 - 1.57 (1.67 - 1.57)**	**50 - 1.69 (1.79 - 1.69)**	**50 – 1.90 (2.01 – 1.90)**	**73.27 - 1.65 (1.75 - 1.65)**
$CC_{1/2}$ (%)	99.9 (65.1)	99.9 (45.3)	99.7 (53.2)	99.9 (49.5)
R_{meas} [%]	9.8 (103.2)	15.7 (271.1)	23.1 (211.1)	7.0 (157.5)
I/σ(I)	14.0 (1.8)	13.1 (1.0)	10.7 (1.35)	15.8 (1.2)
Completeness [%]	98.8 (94.3)	99.0 (93.7)	99.4 (96.4)	99.7 (98.7)
Redundancy	6.8 (6.3)	13.1 (12.4)	13.4 (13.0)	6.8 (6.9)
Wilson B-Factor [Å²]	25.4	31.6	33.1	34.62
Refinement				
Resolution [Å]	45.50 - 1.57	48.43 - 1.69	48.45 – 1.90	73.27 - 1.65
No. of Reflections	133755	73884	52253	65451
R_{work} / R_{free} [%]	**16.47 / 19.50**	**17.10 / 19.70**	**18.39 / 21.16**	**16.73 / 19.17**
No. of Atoms				
Protein	4288	4206	4254	4306
Solvent	386	319	249	250
Carbohydrate	83	75	84	75
B-Factors [Å²]				
Protein	19.0	30.1	31.75	34.67
Solvent	28.1	37.2	39.27	43.0
Carbohydrate	27.5	27.1	36.50	34.2
R. m. s. deviations				
Bond Length [Å]	**0.006**	**0.006**	**0.007**	**0.006**
Bond Angle [°]	**1.186**	**0.909**	**1.003**	**0.890**
Ramachandran Plot				
Favored	520 (97.0%)	510 (97.1%)	513 (97.3 %)	514 (96.8%)
Allowed	13 (2.4%)	13 (0.4%)	12 (2.3 %)	14 (2.6%)
Disallowed	3 (0.6%)	2 (0.4%)	2 (0.4%)	3 (0.6%)

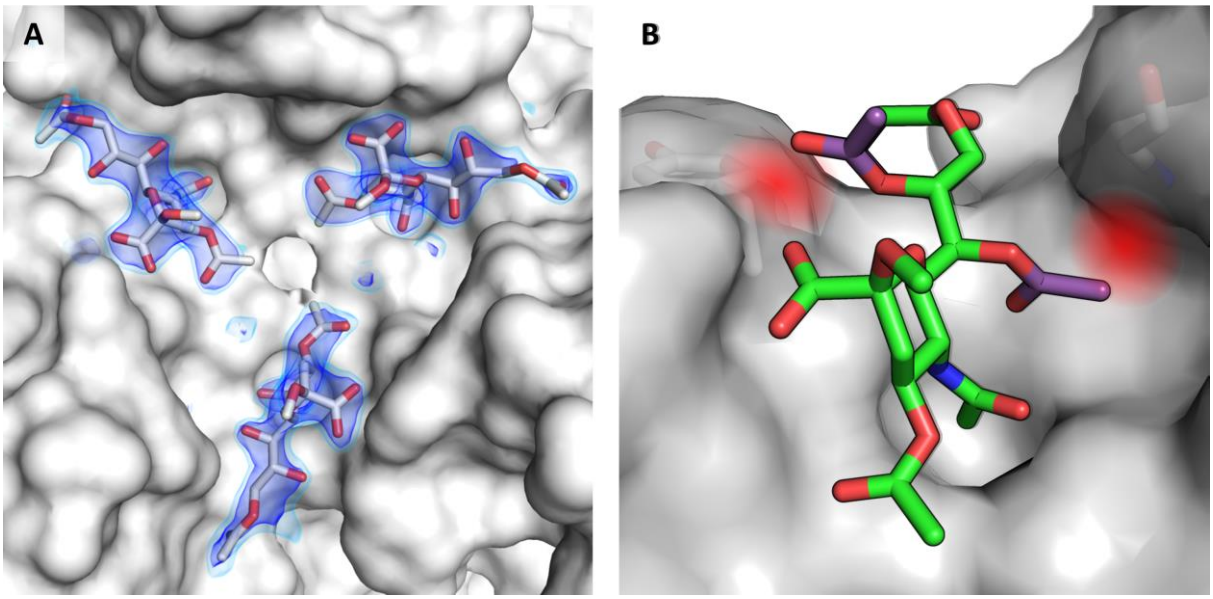


Figure S7 |Binding of differentially acetylated sialic acid variants by HAdV-D36. ***A*** *Complex structure of HAdV-D36 with synthetic α-2-O-methyl-Neu4,5,9Ac$_3$. Shown is a 2F$_o$-F$_c$ map contoured at 1.5σ (dark blue) and 1σ (light blue), respectively. The additional 9-O-acetyl group displays vavrious conformations and does not contribute to the binding.* ***B*** *Modelling of additional O-acetylations at position 7 and 8 (purple). Modelling was performed in Coot without application of a force field. In agreement with [43], the Acetyl groups were added in a sterically favorable position that retains good ligand geometry. Steric clashes are probable in both cases (red discs represent the clashing regions).*

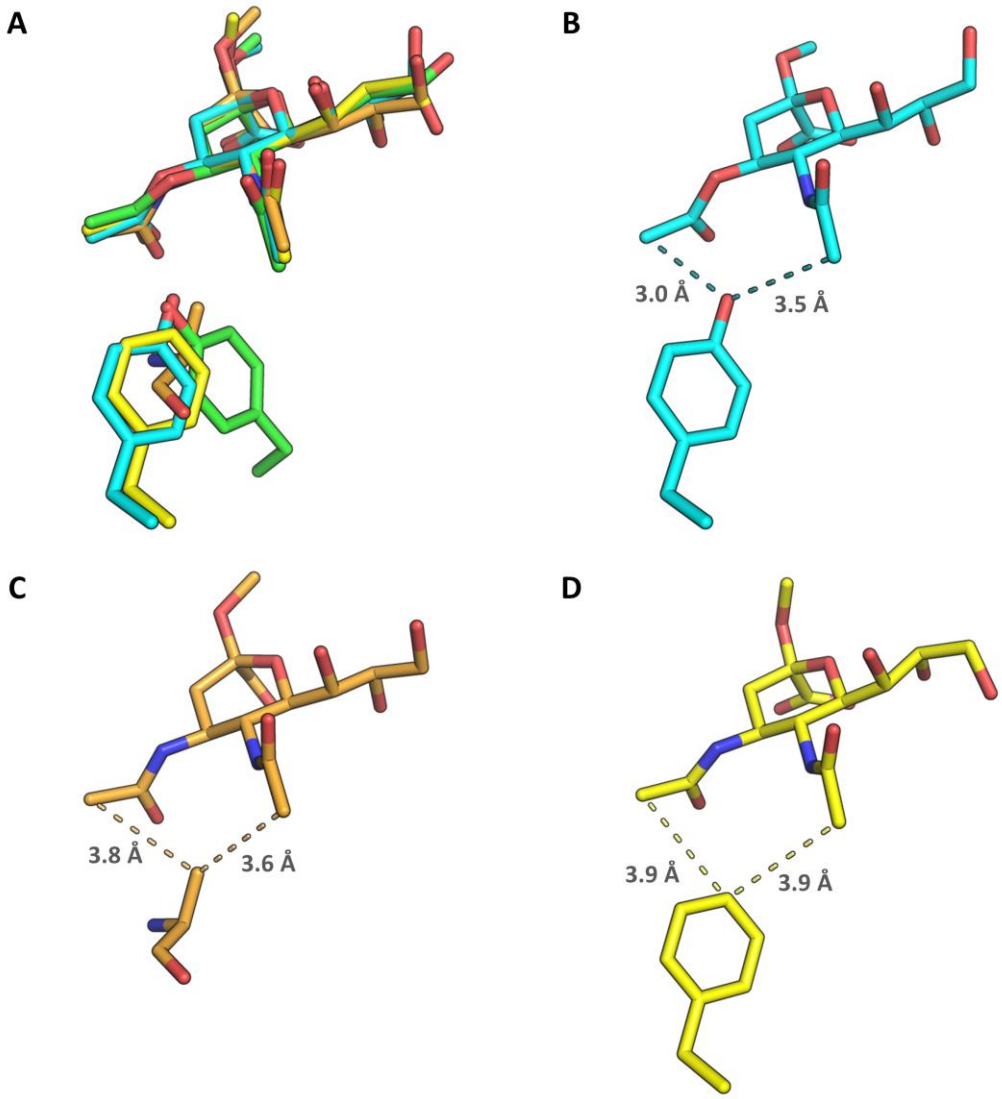


Figure S8 | Gatekeeping residues of HAdV-D36 and different nidovirus hemagglutinin-esterases. *Superpositioning was done according to the carbohydrate portions using Pymol.* ***A*** *Superposition of HAdV-D36 (green), mouse hepatitis virus strain S hemagglutinin-esterase (MHV-S HE^0, cyan, PDB ID 4C7W), and the lectin (yellow) and catalytic (orange) sites of rat coronavirus strain New-Jersey hemagglutinin-esterase (RCoV-NJ HE^0, PDB ID 5JIL) in complex with the non-hydrolysable Neu4,5Ac$_2$ analogue α-2-O-methyl-Neu4,5-di-N-acetylneuraminic acid. B Gatekeeping residue of MHV-S HE^0. C,D Gatekeeping residues of the RCoV-NJ HE^0 esterase (orange) and lectin (yellow) binding sites.*

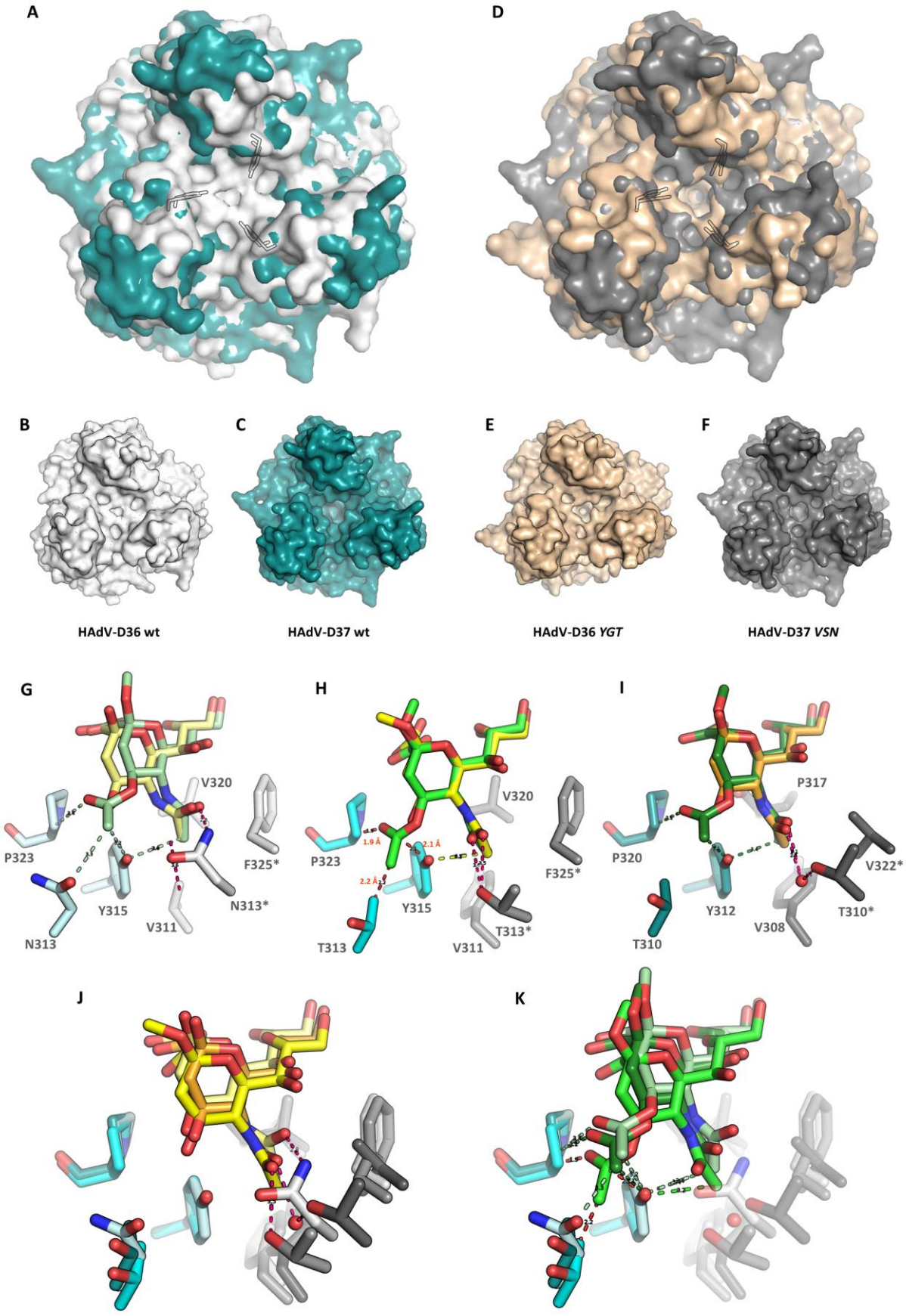


Figure S9 | Quaternary arrangements of HAdV-D36/D37 FK wild type and trimer interface mutants. *All superpositions were performed in PyMol aligning all atoms of the trimer.* ***A*** *Superposition of the HAdV-D36 wt FK (gray) with the HAdV-D37*

wt FK (teal) in analogy to Figure 4B. ***B-C*** *Separate surface representations of the two wt knobs.* ***D*** *Superposition of the HAdV-D36 YGT FK (orange) with the HAdV-D37 VSN FK (dark gray). The HAdV-D37 VSN FK shows a narrower central cavity than HAdV-D36 YGT. A clockwise movement of the Y317 side chain in HAdV-D36 YGT distorts the N-acetyl binding cavity (both analogous tyrosine residues shown as sticks).* ***E-F*** *Separate surface representations of the two mutant knobs.* ***G*** *Superposition of sialic acid complexes of HAdV-D36 wt. The interacting residues were aligned on the sidechain of the gatekeeping residue Y315. The O-acetyl group of Neu4,5Ac$_2$ (green) is in the van-der-Waals range of sidechains N313, Y315, and P323, while the N-acetyl group forms a direct hydrogen bond with N313*.* ***H*** *Modeling of Neu4,5Ac$_2$ (green) on the HAdV-D36/Neu5Ac complex structure. Neu4,5Ac$_2$ was aligned onto Neu5Ac (yellow) in Coot. The relative spacing of T313* and the N-acetyl group prevent the formation of a water-mediated contact. Instead, the N-acetyl group moves towards T313* and forms a direct, long hydrogen bond. Thereby, the N-acetyl function is slightly rotated out of its ideal position and causes a rotation of the whole ligand. As a consequence, the O-acetyl function of a possible Neu4,5Ac$_2$ complex would produce heavy clashes with T313, Y315, and P323 that cannot be overcome by a rotation towards the center.* ***I*** *Superposition of sialic acid complexes of HAdV-D37 wt. T310* is located further away from the sugar's N-acetyl group and allows for the formation of a water-mediated contact in both cases.* ***J*** *Superposition of the Neu5Ac complex structures of HAdV-D36 wt, HAdV-D36 YGT, and HAdV-D37 wt. The need to form a direct contact in HAdV-D36 YGT pulls the sugar towards the gatekeeping residue Y315. Coloring according to G-I.* ***K*** *Superposition of the Neu4,5Ac$_2$ complex structures of HAdV-D36 and D37 wt with the modelled complex of HAdV-D36 YGT. The downward motion and rotation of the sugar in the latter induce heavy clashes. Coloring according to G-I.*

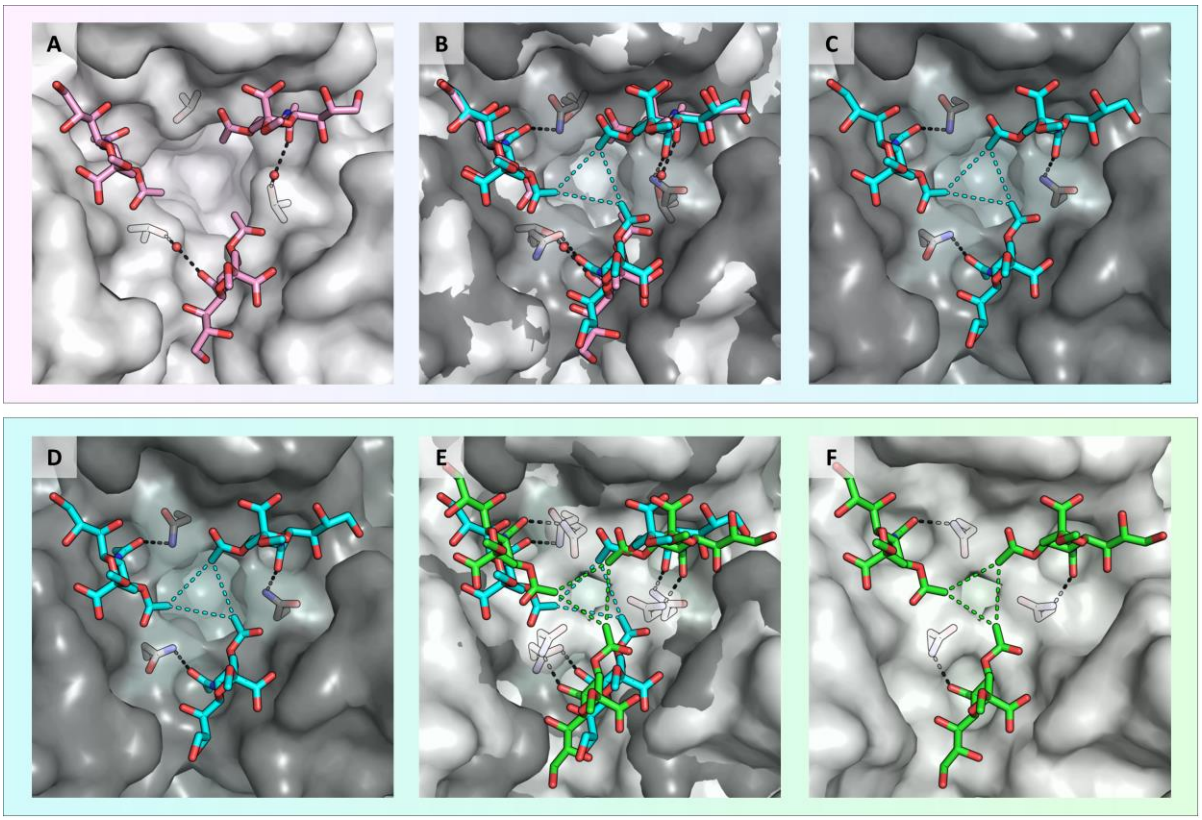


Figure S10 | Comparison of the Neu4,5Ac$_2$ binding modes of HAdV-D37 wt, HAdV-D37 *VSN*, and HAdV-D36 wt. *Residues at the analogous positions 310 (HAdV-D37) and 313 (HAdV-D36) are displayed as sticks and their contacts with the sugar as black dashed lines.* ***A*** *Binding mode of HAdV-D37 wt (pink sticks, light surface).* ***B*** *Superposition of the Neu4,5Ac$_2$ complex structured of HAdV-D37 wt and VSN (cyan sticks, dark surface). The triangular hydrophobic contact is restored in the VSN mutant.* ***C, D*** *Binding mode of HAdV-D37 VSN. The same panel is displayed twice for reasons of clarity.* ***E*** *Superposition of the Neu4,5Ac$_2$ complex structures of HAdV-D37 VSN and HAdV_d36 wt (green sticks, light surface). Despite a relative clockwise and upward positioning of the sugars in HAdV-D36, the length of the triangular hydrophobic contact is very similar.* ***F*** *Binding mode of HAdV-D37 wt analogous to Figure 10.*

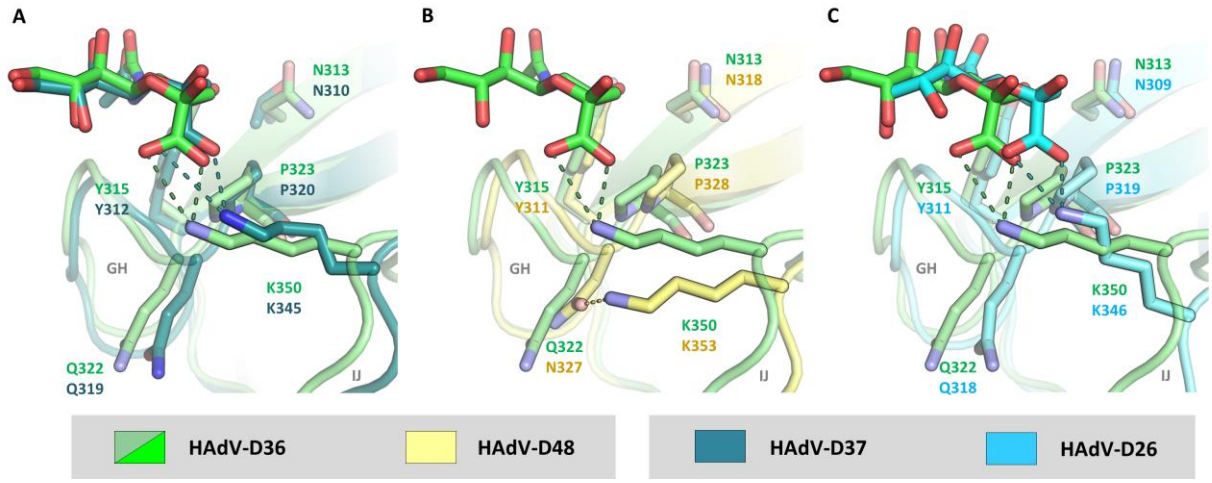


Figure S11 |Conservation of Sialic acid binding among species D HAdVs. *Only the displayed chain was used for superpositioning the structures in PyMol.* ***A*** *Comparison of HAdV-D36 and HAdV-D37, analogous to Figure 7D. The binding modes are highly similar.* ***B*** *In HAdV-D48, the position analogous to Q322 (HAdV-D36) is mutated to a slightly shorter asparagine. N327 engages K353 in a polar contact and, together with a slight shift of the IJ loop, displaces the canonical lysine residue from the binding site. As a result, HAdV-D48 has lost its ability to bind sialic acid.* ***C*** *HAdV-D26 displays a similar alteration of the IJ loop conformation, but since K246 is not engaged in a polar contact it still reaches into the binding pocket and can engage the carboxyl group of Neu5Ac, although this results in a relative shift of the sugar compared to HAdV-D36.*

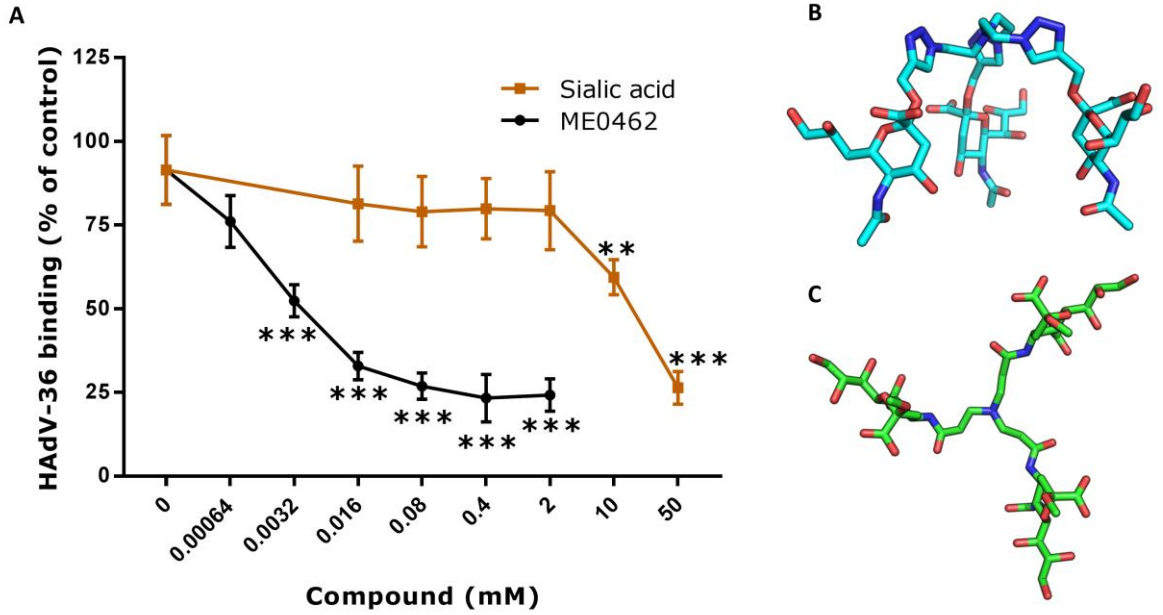


Figure S12 | Possible sialic acid-based inhibitors of HAdV-D36 attachment.A *The trivalent sialic acid-based HAdV-D37 inhibitor 17a is a potent inhibitor of HAdV-D36 attachment.* ^{35}*S-labelled HAdV-D36 virions were pre-incubated with soluble sialic acid or ME0462 (17a). 17a (*IC_{50}*: 6-10 µM) is about* 10^4 *times more potent than Neu5Ac. The data is presented in % of control (binding of virions without any preincubation).* ***B*** *Conformation of 17a bound to HAdV-D37 (PDB-ID 4XQA). C structural model of a Neu4,5*Ac_2*-based new generation inhibitor. A geometry restraint file was generated using the GRADE web server [80] and modelling was done in Coot [69] based on the HAdV-D36 FK/Neu4,5*Ac_2 *complex structure and showed optimal geometry values and no steric clashes.*

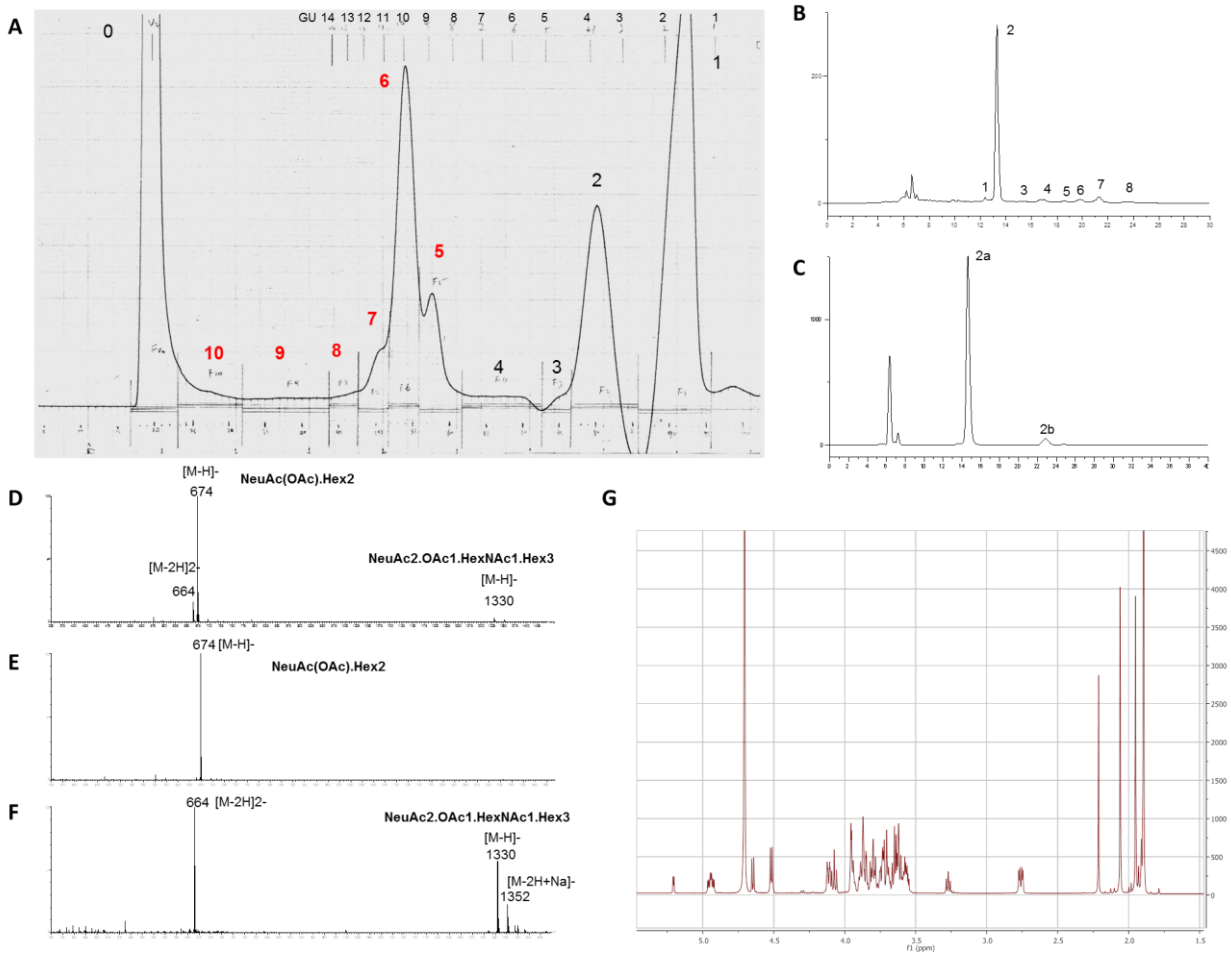


Figure S13 | Purification of 4-O-Ac-3'SL from echidna milk oligosaccharides (EMOs). ***A*** *Bio-Gel P4 Fractionation of EMOs.* ***B*** *HPLC of Bio-Gel P4 fraction F6* ***C*** *Re-HPLC of Bio-Gel P4 fraction F6-2.* ***D*** *Negative-ion mass spectra of HPLC fraction F6-2.* ***E*** *Negative-ion mass spectra of HPLC fraction F6-2a* ***F*** *Negative-ion mass spectra of HPLC fraction F6-2b.* ***G*** *^{1}H-NMR of HPLC F6-2a.*

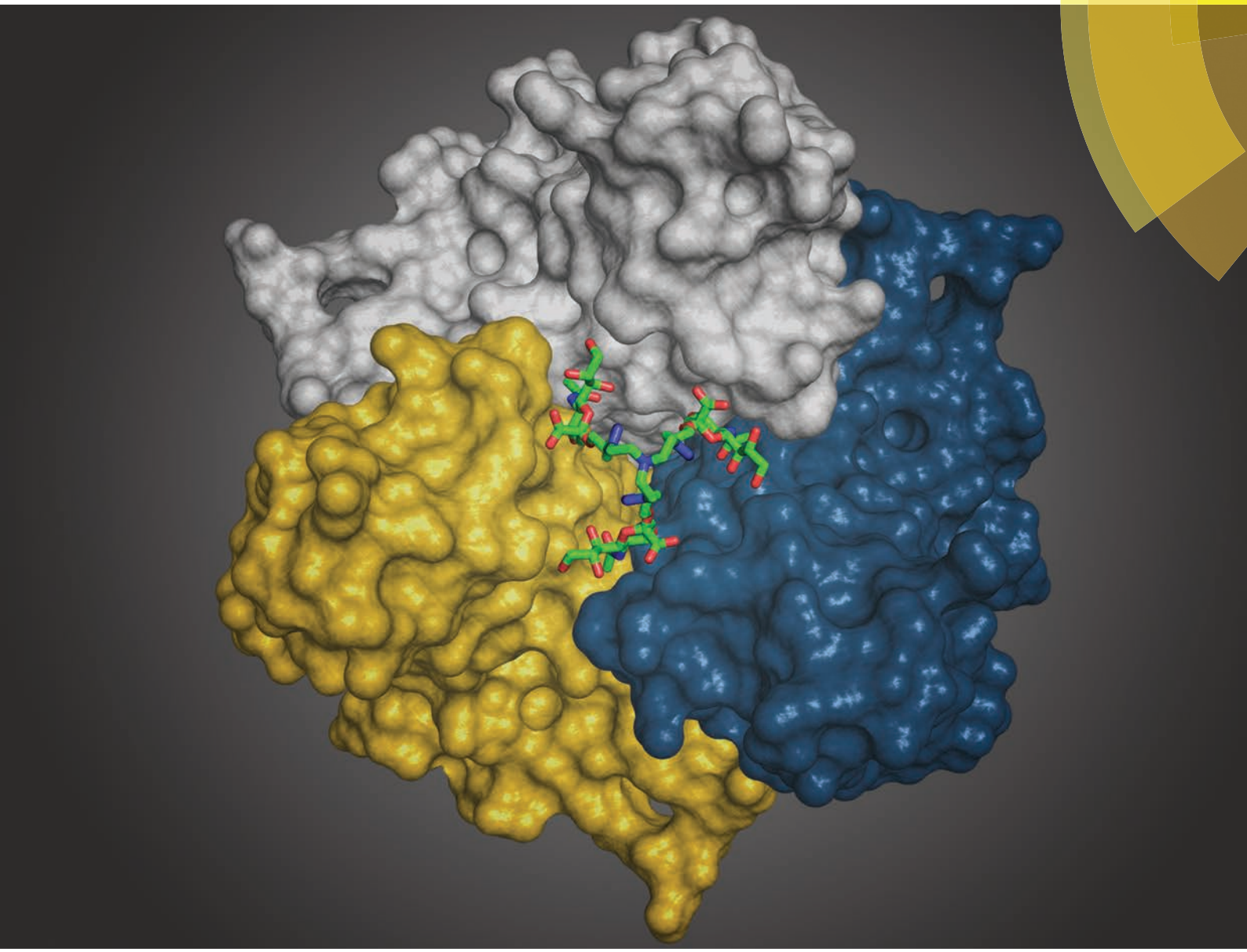
Author contributions

- Cell-based assays were performed by SM and NC, and planned by SM, NC, and AML.
- Initial cloning, purification, crystallization, and structure determination of the HAdV-D37 wt FK were performed by VP in the course of a bachelor thesis under close supervision by AML. The construct and experimental design was made by AML.
- Purification, crystallization, and structure determination of HAdV-D36 and D37 wt complexes as well as the refinement of all wt structures were performed by AML.
- Design and cloning of the VSN and YGT mutants was performed by AML. Purification, crystallization, structure determination, and refinement of the mutants were performed by PB in the course of a bachelor thesis under close tutelage by AML.
- STD-NMR experiments were planned by AML and performed and analyzed by BB.
- 4-O-Ac-3'SL from echidna milk was purified by WC.
- Synthesis of α-2-O-methyl-Neu4,5Ac_2 and α-2-O-methyl-Neu4,5,9Ac_3 was planned and performed by RC and EJ.
- Glycan array screenings were planned and performed by YL, and the protein was provided by VP and AML.
- DSF measurements were done by Dr. Fabian Zehender (NanoTemper GmbH) during a demonstration of the Prometheus NT.48. Sample preparation and handling was done by AML.
- TROSY-NMR was performed by AML and PB with the help of Dr. Ancilla Neu, MPI Tübingen.
- The Manuscript was written by AML, except for the respective methods contributions. Cell-based assay figures were prepared by NC, all other figures by AML.

Volume 13 | Number 35 | 21 September 2015 | Pages 9143–9326

Organic & Biomolecular Chemistry

www.rsc.org/obc



ISSN 1477-0520

ROYAL SOCIETY OF CHEMISTRY

PAPER

Mikael Elofsson *et al.*

Triazole linker-based trivalent sialic acid inhibitors of adenovirus type 37 infection of human corneal epithelial cells

Organic & Biomolecular Chemistry

PAPER

View Article Online
View Journal | View Issue

Cite this: *Org. Biomol. Chem.*, 2015, **13**, 9194

Triazole linker-based trivalent sialic acid inhibitors of adenovirus type 37 infection of human corneal epithelial cells†

Rémi Caraballo,[a,b] Michael Saleeb,[a,b] Johannes Bauer,‡[c] A. Manuel Liaci,[c] Naresh Chandra,[b,d,e] Rickard J. Storm,[b,d,e] Lars Frängsmyr,[b,d,e] Weixing Qian,[a] Thilo Stehle,[c,f] Niklas Arnberg[b,d,e] and Mikael Elofsson*[a,b]

Received 21st May 2015,
Accepted 9th July 2015

DOI: 10.1039/c5ob01025j

www.rsc.org/obc

Adenovirus type 37 (Ad37) is one of the principal agents responsible for epidemic keratoconjunctivitis (EKC), a severe ocular infection that remains without any available treatment. Recently, a trivalent sialic acid derivative (**ME0322**, *Angew. Chem. Int. Ed.*, 2011, **50**, 6519) was shown to function as a highly potent inhibitor of Ad37, efficiently preventing the attachment of the virion to the host cells and subsequent infection. Here, new trivalent sialic acid derivatives were designed, synthesized and their inhibitory properties against Ad37 infection of the human corneal epithelial cells were investigated. In comparison to **ME0322**, the best compound (**17a**) was found to be over three orders of magnitude more potent in a cell-attachment assay (IC_{50} = 1.4 nM) and about 140 times more potent in a cell-infection assay (IC_{50} = 2.9 nM). X-ray crystallographic analysis demonstrated a trivalent binding mode of all compounds to the Ad37 fiber knob. For the most potent compound ophthalmic toxicity in rabbits was investigated and it was concluded that repeated eye administration did not cause any adverse effects.

Introduction

Human adenoviruses (Ads), which belong to the mammalian adenovirus genus, *Mastadenovirus*, are commonly encountered infectious agents. In humans, Ads are associated with a multitude of clinical symptoms encompassing upper and lower respiratory tract infections, gastroenteritis, hemorrhagic cystitis and ocular diseases such as conjunctivitis and epidemic keratoconjunctivitis (EKC).[1] Ads are ubiquitous in nature and new types are continuing to be discovered.[2,3] Since the isolation of the first Ads about 60 years ago,[4,5] over 60 types that are grouped into seven species (A–G) have been identified. Ad infections are usually self-limited in immunocompetent patients while they can become a serious life-threatening disease in immunocompromised individuals.[6,7]

To date, there are no specific antiviral drugs available for the treatment of Ad infections.[8] Ads are obligate intracellular pathogens that are fully dependent on the cellular replication machinery. The selective inhibition of Ad replication by antiviral compounds is therefore difficult to achieve as some of the essential functions of the host cells may also be altered. The acyclic nucleoside analogue cidofovir has been shown effective against EKC-causing Ads. Unfortunately, nephrotoxicity and lacrimal canalicular blockage have been reported after systemic administration.[1,8] An alternative approach to the intracellular activity of nucleoside analogues is to block the initial interaction between the virus and host cell and thereby block cell attachment and subsequent entry.[9]

EKC is a severe and highly contagious ocular infection that is contracted by millions of individuals every year. Among the Ad types responsible for EKC, Ad8, Ad19 and Ad37 remain the principal causative agents of the infection;[10] however, new EKC-causing types such as Ad53, Ad54 and Ad56 have been recently isolated from patients.[11–13] Common symptoms are keratitis, conjunctivitis, edema, pain, lacrimation, formation of pseudomembranes and decreased vision.[14] Because these viruses are spread by contact (*e.g.* hand to eye contact),[14] EKC

[a]*Department of Chemistry, Umeå University, SE90187 Umeå, Sweden. E-mail: mikael.elofsson@chem.umu.se*
[b]*Umeå Centre for Microbial Research, Umeå University, SE90187 Umeå, Sweden*
[c]*Interfaculty Institute of Biochemistry, University of Tübingen, 72076 Tübingen, Germany*
[d]*Department of Clinical Microbiology, Division of Virology Umeå University, SE90185 Umeå, Sweden*
[e]*Laboratories for Molecular Infection Medicine Sweden, Umeå University, SE-901 87 Umeå, Sweden*
[f]*Department of Pediatrics, Vanderbilt University School of Medicine, Nashville 37232, TN, USA*

†Electronic supplementary information (ESI) available: Synthesis, characterization and copies of NMR spectra for isolated intermediates and target compounds; SPR data; X-ray crystallographic tables. See DOI: 10.1039/c5ob01025j

‡Current address: Centre for Molecular Medicine Norway, University of Oslo, PO Box 1137 Blindern, 0318 Oslo, Norway.

is frequent in densely populated areas and in medical wards with insufficient hygiene precautions. The infection commonly lasts for up to two weeks; however, some patients continue to suffer from sight impairment for several months, years or even permanently. Whereas other non-EKC causing Ads use CD46 and coxsackie-adenovirus receptor (CAR) as receptors,[15] EKC-causing Ads bind *via* their homotrimeric fiber knobs to sialic acid-containing glycans that are situated on epithelial cells in the cornea and/or conjunctiva.[16] The fiber knobs of EKC-causing Ads are located at the most distal part of each of the 12 fibers that protrude from the icosahedral virion. The knobs are highly homologous and the critical sialic acid-interacting residues are thus conserved within the different types of an Ad species.[17] Consequently, substances targeting a single type also have the potential to prevent cell attachment by the other types. Recently, glycoproteins with glycans corresponding to the carbohydrate portion of the GD1a gangliosides were shown to function as receptors for the infection of ocular cells by EKC-causing Ads.[17] The crystal structure of the Ad37-GD1a complex showed that the terminal sialic acid residues, located on each of the two branches of the GD1a glycan, are accommodated into two out of three carbohydrate recognition sites on top of the Ad37 fiber knob. Thus, inhibition of Ads with natural or synthetic sialic acid derivatives may prevent the virion to attach to, penetrate into and infect new cells. As a result, the infection and its spread would become limited. Importantly and especially in the case of EKC, the poor pharmacologic properties of carbohydrate-based drugs that include rapid serum clearance and poor cellular uptake obtained after systemic administration can be bypassed by the use of a topical mode of administration (*e.g.* cream, ointment, eye drops) and extracellular targeting of the virus particles.

Efficient sialic acid–based inhibitors of Ad37 infection of human corneal epithelial (HCE) cells have been recently reported.[18–21] In order to circumvent the relatively low efficacy of monovalent sialic acid derivatives, the authors took advantage of the trimeric binding site at the Ad37 fiber knob. The use of multivalent sialic acid derivatives or glycoconjugates that can simultaneously bind to several carbohydrate recognition domains per knob considerably improved the inhibitory potency in comparison to the sialic acid monosaccharide. For instance, **ME0322**[21] (Fig. 1a), a synthetic trivalent sialic acid derivative was reported as four orders of magnitude more potent than the natural sialic acid monosaccharide. Other successful reports of trivalent glycosides include, for example, the synthesis of glycoclusters targeting the hepatic asialoglycoprotein receptor.[22–24]

Herein, we report on the design, synthesis and evaluation of new potent trivalent sialic acid inhibitors of Ad37 infection of HCE cells where the trivalent scaffolds were conveniently accessed by "click" chemistry. We probed the beneficial effect of a more compact linker and the influence of a small modification at the *N*-acetyl moiety on efficacy. Finally, the analysis of the respective crystal structures allowed us to confirm that the trivalent binding mode is engaged by all the inhibitors to the fiber knob as well as to reason on the potency differences.

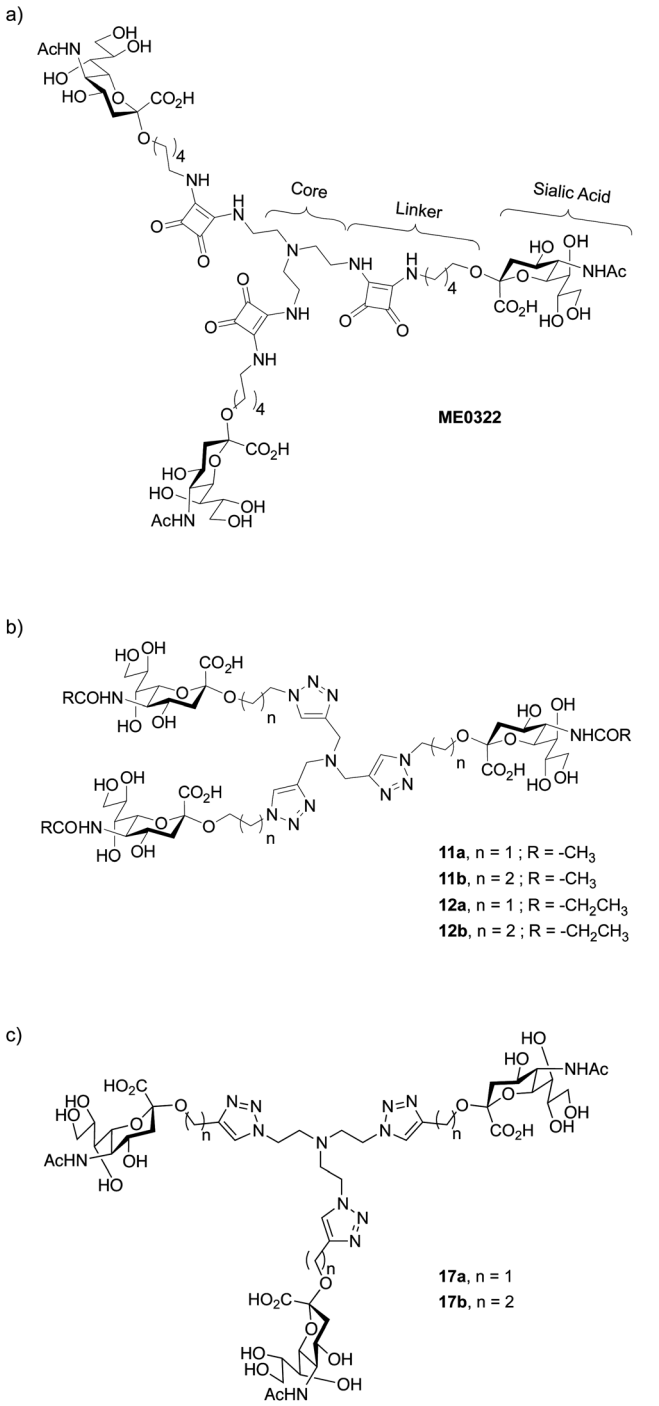


Fig. 1 Structure of (a) trivalent sialic acid **ME0322**,[21] a potent inhibitor of Ad37, (b) the first-generation triazole linker-based trivalent sialic acid derivatives and (c) the second-generation triazole linker-based trivalent sialic acid derivatives.

Results and discussion

Design strategy for the first-generation triazole linker-based compounds

The design of new trivalent sialic acid conjugates was based on the structural features of **ME0322** (Fig. 1). Thus, the com-

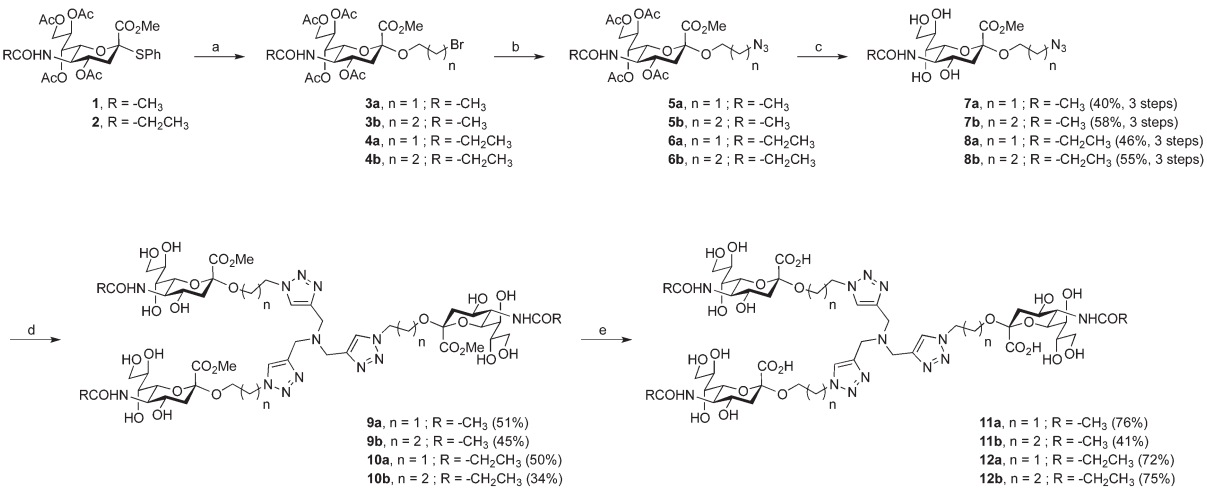
pounds were composed of three key building blocks: the *core fragment* that is necessary for the construction of a trivalent network; the *sialic acid residue* that is required for the binding to the carbohydrate recognition sites at the Ad37 fiber knob and the *linker* that is connecting the sialic acid residues to the core fragment.

The tertiary amino scaffold that constituted the core fragment in **ME0322** and the sialic acid residues that are crucial for the interactions with the target protein were conserved in the new ligands. Contrarily, the linker, which could not be visualized in the crystal structure of **ME0322** in complex with the Ad37 fiber knob[21] likely due to flexibility and lack of interactions with the protein, was revised. The design efforts were initially focused on both, shortening the original linker with the idea to bring the core fragment closer to the receptor and, introducing a heteroaromatic ring moiety to possibly create new contacts with the fiber knob (Fig. 1b). After analysis of the commercial availability of functionalized tertiary amino building blocks (core fragment) and robust chemical reactions that allow a rapid access to multivalent sialic acid containing structures our choice was triazole-based linkers.[25–30] Two linker lengths were investigated to probe an eventual size effect on the potency.

In addition to trivalent sialic acid compounds **11a** and **11b**, their corresponding *N*-acetyl modified analogues **12a** and **12b** were also included in this series (Fig. 1b). Thus, we aimed to further investigate the hydrophobic pocket to which the *N*-acetyl group of the sialic acid is directed in the Ad37-**ME0322** complex.[21] Previous studies on a set of multivalent human serum albumin (HSA) conjugates highlighted a detrimental effect of an increased lipophilicity at the *N*-acetyl group of the sialic acid on the compound potency, although crystallographic investigations suggested that these modifications should readily fit in the hydrophobic fiber knob pocket that accommodate the sialic acid *N*-acetyl group.[31] This loss in potency could eventually be scaffold- and/or linker-dependent as suggested by a better tolerance of an increased lipophilicity during the evaluation of their corresponding monovalent analogues. Therefore, the effect of such modifications on trivalent sialic acid derivatives that have a linker especially designed to better accommodate into the fiber knob cellular receptor should clarify the scaffold- and linker-dependence hypothesis.

Synthesis of the first-generation triazole linker-based compounds

The route to unmodified *N*-acyl trivalent sialic acids **11a** and **11b** proved straightforward and could be achieved in eight steps from commercially available chemicals or in five steps from key intermediate **1** (Scheme 1). The synthesis of the sialic acid thiophenyl derivative **1**, readily prepared from sialic acid, was performed according to published procedures.[32] Sialosides **3a** and **3b** were then accessed in good conversion by glycosylation of 2-bromoethanol and 3-bromopropan-1-ol respectively, with compound **1**. The reactions yielded inseparable mixtures of anomers together with the resulting elimination product that was not further purified at this stage. Bromo derivatives **3a** and **3b** were readily converted to their azido analogues **5a** and **5b**. Subsequent *O*-deacylation using standard Zemplèn conditions afforded anomerically pure **7a** and **7b** in 40% and 58% yields, respectively, over three steps. Then, compounds **7a** and **7b** were reacted with tripropargylamine in a copper-catalyzed azide–alkyne cycloaddition reaction ("click" reaction). Thus, methyl esters **9a** and **9b** were obtained in 51% and 45% yields, respectively. Subsequent saponification provided the final target compounds **11a** and **11b** in 76% and 41% yields, respectively. The route to *N*-acyl modified trivalent sialic acids **12a** and **12b** proved somewhat more challenging. The target compounds could be reached in 13 steps from



Scheme 1 Synthesis of **11a**, **11b**, **12a** and **12b**. Reagents and conditions: (a) i: molecular sieves 3 Å, 2-bromoethanol or 3-bromopropan-1-ol, CH_3CN/CH_2Cl_2 (3 : 2), rt, 2 h, ii: AgOTf, IBr, −73 °C, 4.5 h, iii: DIPEA, −73 °C, 30 min. (b) NaN_3, TBAI, DMSO, rt, 6 h. (c) i: NaOMe, MeOH, rt, 3 h, ii: H^+ ion exchange resin. (d) Tripropargylamine, $CuSO_4$, sodium ascorbate, THF/H_2O (1 : 1), 50 °C, 3 h then rt, 18 h. (e) i: LiOH, MeOH, rt, 9 h, ii: H^+ ion exchange resin.

commercially available sialic acid or in five steps from key intermediate **2** (Scheme 1). The synthesis of compound **2** was performed according to published procedures.[31] The following steps were analogous to the above-described synthetic route to compounds **11a** and **11b**. Thus, the successive glycosylation, azide formation and *O*-deacylation reactions afforded anomerically pure product **8a** and **8b** in 46% and 55% yields over three steps, respectively. Subsequent "click" reactions provided the trivalent compounds **10a** and **10b** in 50% and 34% yields, respectively. Finally, saponification of the methyl esters gave the target products **12a** and **12b** in good yields.

Biological evaluation of the first-generation triazole linker-based compounds

Cell-binding assays using ^{35}S-labeled virions were performed to investigate the efficiency of compounds **11a**, **11b**, **12a** and **12b** to prevent the attachment of Ad37 virions to HCE cells (Fig. 2a). The assays were carried out essentially as previously described but with minor modifications[16,33] and, **ME0322**, sialic acid and GD1a glycan were used as reference compounds.[17,21] In brief, ^{35}S-labeled Ad37 virions were mixed with or without the trivalent sialic acid derivatives, GD1a glycan and sialic acid at various concentrations. The mixtures were incubated with HCE cells and unbound virions were then washed away. Finally, the cell-associated radioactivity was measured by using a scintillation counter.

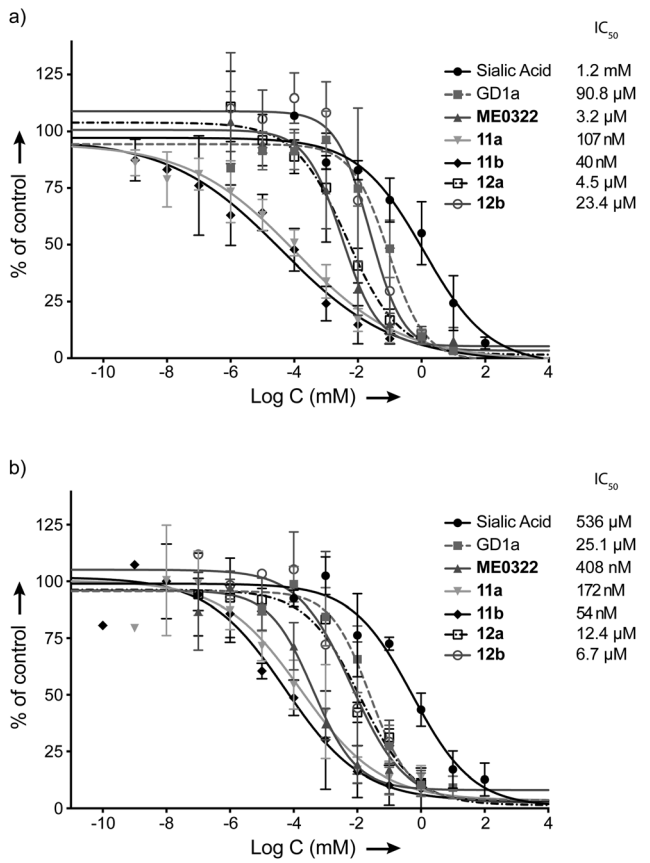


Fig. 2 Effect of the set of trivalent sialic acid derivatives **11a–b** and **12a–b** on Ad37 binding to and infection of HCE cells. (a) Virion binding in the presence of inhibitors at different concentrations. (b) Infection at different concentrations of the inhibitors. Data are presented as % of control that is the value obtained in the absence of inhibitor.

The attachment of Ad37 virions to HCE cells was dramatically hindered in the presence of the triazole linker-based compounds (Fig. 2a). Indeed, the new trivalent sialic acid derivatives were considerably more potent than the sialic acid monosaccharide (IC_{50} = 1.2 mM) and bivalent GD1a glycan (IC_{50} = 90.8 μM). Compounds **11a** and **11b** were found as the most efficient compounds to prevent Ad37 virions from HCE cells attachment with IC_{50} values of 107 nM and 40 nM, respectively. The previously described compound, **ME0322**, proved to be less potent (IC_{50} = 3.2 μM). The *N*-acyl modified derivatives **12a** and **12b** completed the series with IC_{50} values of 4.5 μM and 23.4 μM, respectively.

A correlation between the linker length and the ligand potency could not be demonstrated. Thus, reduction of the linker size in the *N*-acetyl series (**11a** *vs.* **11b**) resulted in a 2.7 times decrease in binding potency. Contrarily, the same transformation in the *N*-acyl modified series (**12a** *vs.* **12b**) proved 5.2 times more beneficial. These results suggest that some degree of variation of the linker length is tolerated. However, increasing the lipophilicity of the ligands at the *N*-acyl moiety was clearly not beneficial to the binding to Ad37 virions (**11a** *vs.* **12a** and **11b** *vs.* **12b**). Even though these data are in agreement with previous investigations;[31] herein, the drop in potency is not as pronounced. Indeed, while *N*-acyl modified multivalent HSA conjugates were very poorly effective, compounds **12a** and **12b** prevent the attachment of Ad37 virions to HCE cells rather efficiently. Therefore, the loss in binding potency cannot only be attributed to an increased lipophilicity and/or bulkiness at the *N*-acyl moiety and, the nature of the multivalent scaffold is most likely to play an important role.

Infection experiments were then performed to further evaluate our set of compounds (Fig. 2b). The assays were carried out essentially as previously described but with minor modifications.[16,33] In brief, unlabeled virions were mixed with or without the trivalent sialic acid derivatives, GD1a glycan or sialic acid at various concentrations. These mixtures were then added to HCE cells and incubated at +4 °C. Unbound virions were washed away, the resulting mixtures were incubated at +37 °C and a synchronized infection – all virions enter the cells simultaneously – was then obtained. After 44 h of infection, the cells were rinsed, fixed, incubated with rabbit polyclonal anti-Ad37 antibodies (that recognize viral capsid proteins) prior to being washed and stained. Finally, the cells were washed and the number of infected cells was quantified by immunofluorescence microscopy.

Hence, the trends from the cell-binding assays were confirmed in the infection experiments (Fig. 2a and b). Compounds **11a** and **11b** prevented infection of HCE cells by Ad37 virions most efficiently with IC_{50} values of 172 nM and 54 nM, respectively. The half maximal inhibitory concentration of **ME0322** that had previously been estimated to 380 nM[21] was

herein calculated to 408 nM. Compounds **12a** and **12b** were evaluated as the least potent trivalent sialic acid derivatives with IC_{50} values of 12.4 μM and 6.7 μM, respectively. The GD1a glycan (IC_{50} = 25.1 μM) and sialic acid (IC_{50} = 536 μM) completed the series.

Design strategy for the second-generation triazole linker-based compounds

A second round of linker engineering (Fig. 1c), where the connecting "click" strategy was reversed with respect to the first generation of ligands (Fig. 1b), was performed. The linker lengths of **11a** and **11b** were preserved in compounds **17a** and **17b**, respectively; however, the triazole ring was moved closer to the sialic acid residues. Therefore, we could not only investigate the influence of the triazole ring orientation but also probe if additional interactions with the fiber knob could be achieved.

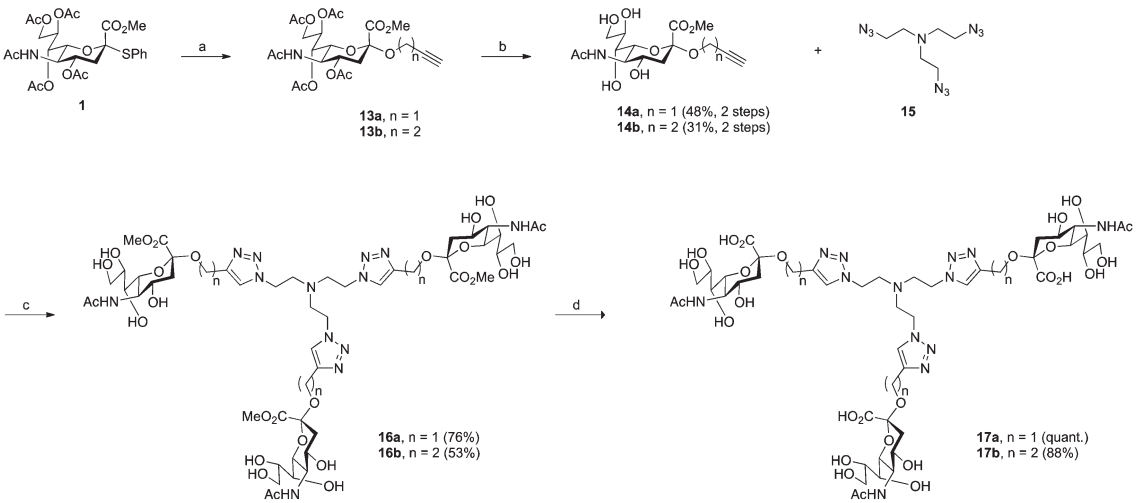
Synthesis of the second-generation triazole linker-based compounds

Compounds **17a** and **17b** were prepared in four steps from key intermediate **1** (Scheme 2). The synthesis of anomerically pure sialosides **14a** and **14b** (48% and 31% yields over two steps, respectively) followed the above-described glycosylation and *O*-deacylation procedures, albeit the use of the appropriate alcohol derivatives in the first step. Compounds **14a** and **14b** were then reacted with tris (2-azidoethyl)amine (**15**) in a "click" reaction. The core building block **15** was synthesized in two steps from commercially available triethanolamine. First, tris (2-chloroethyl)amine was obtained according to published procedure[34] and the chloro derivative was then readily converted to its azido analogue **15**. Compound **15** was judged as potentially explosive and was therefore kept in solution at all times. The methyl esters **16a** and **16b** were obtained in 76% and 53% yields, respectively. Subsequent saponification provided the final target compounds **17a** and **17b** in quantitative and 88% yields, respectively.

Biological evaluation of the second-generation triazole linker-based compounds

The compounds were subsequently investigated in cell-binding assays (Fig. 3a). The linker optimization strategy proved successful and a 76 times increase in potency was observed for **17a** (IC_{50} = 1.4 nM) with respect to **11a**. Contrarily, **17b** (IC_{50} = 376 nM) was evaluated about 9 times less efficient than **11b** to prevent the attachment of Ad37 virions to HCE cells and **17a** was over two orders of magnitude more potent than **17b**. These results contrast our earlier observations where a clear effect of the linker length on the ligand potency could not be evidenced. A tentative explanation for the potency of **17a** could be that the position of the triazole ring adjacent to the sialic acid residue might result in additional interactions with the fiber knob. These potentially favorable contacts would be harder to achieve for **11a**, **17b** and **11b**, where the triazole ring is situated further away from the sialic acid residue by one or two methylene groups, respectively. The reasons for the relatively low efficiency of **17b** in comparison to **11a** and **11b** to prevent Ad37 virions attachment to HCE cells however remain unclear.

Compounds **17a** and **17b** were then subjected to cell-infection assays. The results were in agreement with the cell-binding assays and **17a** (IC_{50} = 2.9 nM) was confirmed as the best inhibitor against Ad37 infection to HCE cells. Interestingly, the optimization strategy entirely based on the linker redesign provided an inhibitor over 140 times more potent than the initial lead compound (**17a** *vs.* **ME0322**). Compound **17b** also proved efficient against Ad37 infection to HCE cells, despite a relatively low potency in relation to **17a**.



Scheme 2 Synthesis of **17a** and **17b**. Reagents and conditions: (a) *i*: molecular sieves 3 Å, propargyl alcohol or 3-butyn-1-ol, CH_3CN/CH_2Cl_2 (3 : 2), rt, 2 h, *ii*: AgOTf, IBr, −73 °C, 4.5 h, *iii*: DIPEA, −73 °C, 30 min. (b) *i*: NaOMe, MeOH, rt, 3 h, *ii*: H^+ ion exchange resin. (c) $CuSO_4$, sodium ascorbate, THF/H_2O (1 : 1), 50 °C, 3 h then rt, 18 h. (d) *i*: LiOH, MeOH, rt, 9 h, *ii*: H^+ ion exchange resin.

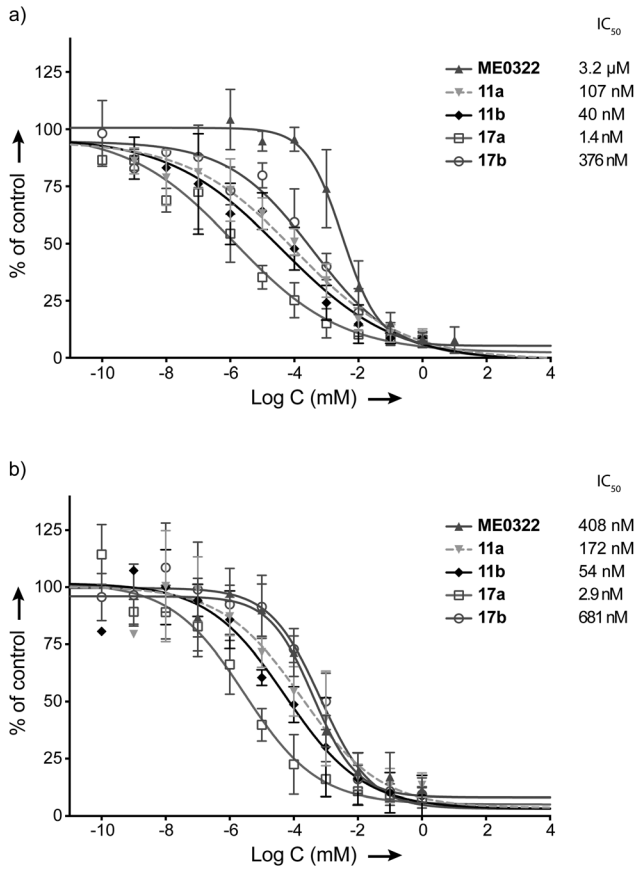


Fig. 3 Effect of the set of trivalent sialic acid derivatives (**17a–b**) on Ad37 binding to and infection of HCE cells. (a) Virion binding in the presence of inhibitors at different concentrations. (b) Infection at different concentrations of the inhibitors. Data are presented as % of control that is the value obtained in the absence of inhibitor.

Surface plasmon resonance (SPR) experiments of selected trivalent inhibitors

Finally, **11a**, **11b**, **17a**, **17b** and **ME0322** were investigated in surface plasmon resonance (SPR) experiments and their respective binding affinities (K_D) for immobilized Ad37 fiber knobs were determined (see Experimental section for further details). Due to the fast *on* and *off* rates, the association and dissociation constants could not be determined. SPR data corroborated well the trends from both cell-binding and cell-infection assays and the five compounds proved to interact with the Ad37 fiber knob. Thus, **17a** (K_D = 10.3 ± 1.4 μM, n = 3) was confirmed to best interact with the Ad37 fiber knob, followed by **11b** (K_D = 72.8 ± 3.3 μM, n = 3), **11a** (K_D = 78.6 ± 2.7 μM, n = 3), **ME0322** (K_D = 135.4 ± 17.9 μM, n = 3) and **17b** (K_D = 163.4 ± 12.3 μM, n = 3) (see ESI† for SPR curve tracing). It is also worth noting the influence of the Ad37 fiber knob construct on the K_D values. Indeed, while histidine-free fiber knobs were used during our assays, histidine-tagged proteins were previously utilized thus providing a different K_D value for **ME0322** (K_D value previously evaluated at 14 μM).[21] This clearly underlines the need of an internal reference during the SPR experiments.

Crystallography

Crystal structures of Ad37 in complex with our new trivalent sialic acid conjugates were obtained by soaking (using 2 mM **12a**) or co-crystallization (for **11a**, **11b**, **12b**, **17a** and **17b**) using previously reported methods.§[17,21] In all complex structures (Fig. 4, S11 and Table S2†), electron density for the entire conjugate was visible thus allowing for unambiguous placement of the ligands. A simultaneous binding of the three sialic acid residues from a single ligand to the same fiber knob was evidenced, and most sialic acid/Ad37 contacts observed in previous complexes[17,21,35] were retained within the different complex structures (Fig. 4 and S12†). Lys345, Pro317 and Tyr312 were confirmed as key contributors to ligand binding and directly interact with the sialic acid moiety, whereas the Ser344/sialic acid contact occurs *via* water-mediated hydrogen bonds. Additional interactions from the linker and/or core fragment with the fiber knob could however not be observed and the conformations and flexibility of the linker varied significantly from one Ad37-ligand complex to another (Fig. 4 and 5a). Although the hypothesis of favorable direct interactions between the triazole ring of **17a** and the fiber knob protein could thus not be verified, there may still exist long-range electrostatic interactions that favor binding. However, the analysis of the Ad37-**17b** complex highlighted an additional internal order of **17b** that could eventually explain its lower potency. Indeed, in the complex the sialic acid residues of **17b** were rotated sideward and the linker formed a bell-like shape, possibly due to a staggered arrangement of the triazole rings (Fig. 4f and magenta structure in Fig. 5a). This ordered and more compact conformation could then directly affect the potency of the ligand.

Crystal structures of the *N*-acyl modified compounds **12a** and **12b** showed an overall binding mode similar to their corresponding unmodified *N*-acyl analogues (Fig. 4a–d, respectively). Moving the analysis to the sialic acid binding site, the Ad37-**12a** and Ad37-**12b** complexes provided important information (Fig. 5b and c, respectively). The additional methyl group in **12a** and **12b** that is oriented away from Tyr308 and Val322 pushes the entire ligand slightly upwards. This unfavorable interaction most likely causes the drop in inhibitory potency for the *N*-acyl modified series of compounds.

Ophthalmic toxicity of compound 17a in rabbit

To examine if the most potent compound (**17a**) can be used for topical treatments of eye infections, ophthalmic toxicity was studied in rabbits. Six male New Zealand white rabbits were divided into two groups of three and to each eye 40 μL of 1 mg mL^{-1} of **17a** in 0.9% aq. NaCl or vehicle alone was administered topically. In total, each eye received 48 administrations over a period of seven days. Before the first administration, day 2, and at day 8 the animals were subjected to body weight recordings, slit lamp examinations, intraocular

§ Atomic coordinates and structure factors have been deposited with the Protein Data Bank under Accession Codes 4K6T (Ad37-**11a**), 4K6U (Ad37-**11b**), 4K6W (Ad37-**12a**), 4K6V (Ad37-**12b**), 4XQA (Ad37-**17a**) and 4XQB (Ad37-**17b**).

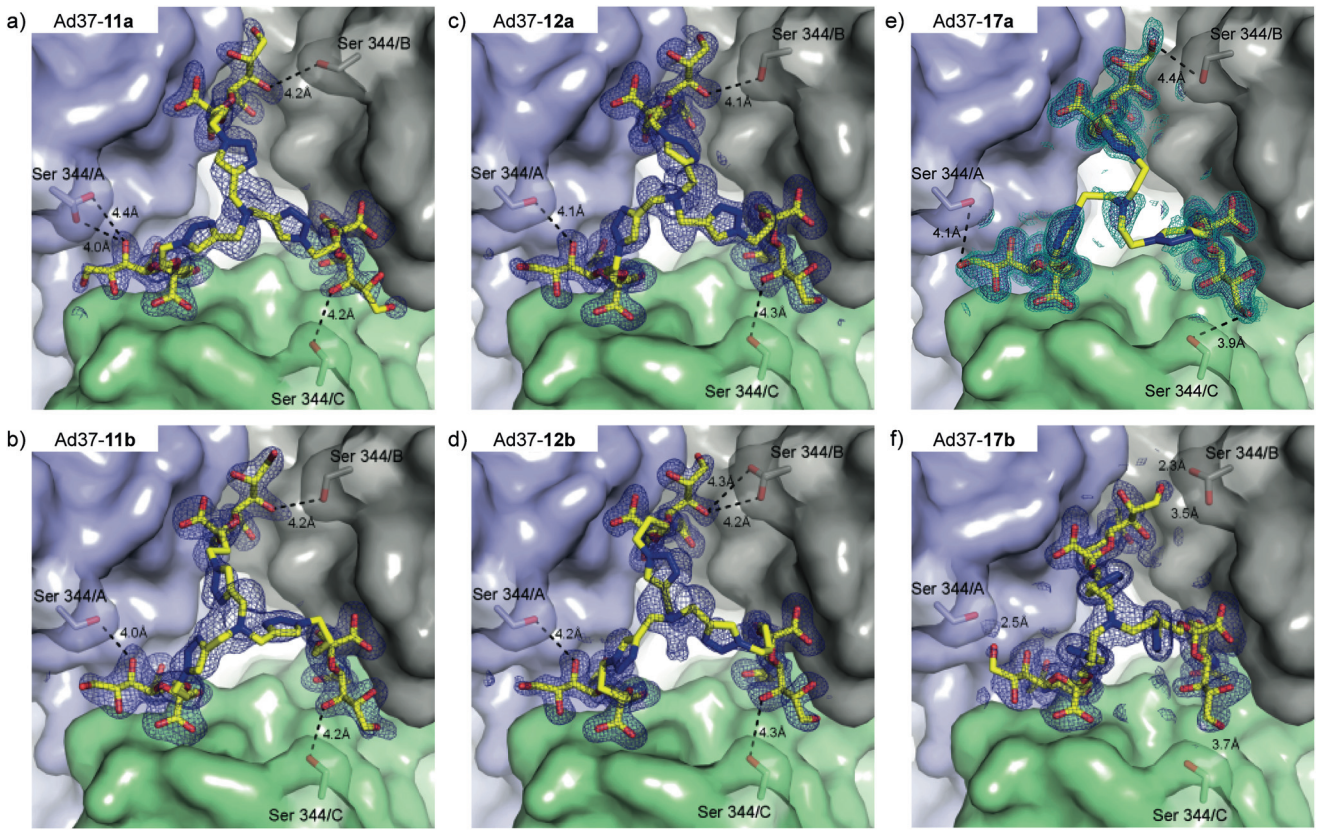


Fig. 4 Binding of **11a**, **11b**, **12a**, **12b**, **17a** and **17b** to the Ad37 fiber knob (top view, a–f). Simulated annealing omit difference electron density maps for **11a** (a, 2.0 Å resolution, PDB ID 4K6T), **11b** (b, 1.9 Å resolution, PDB ID 4K6U), **12a** (c, 1.5 Å resolution, PDB ID 4K6W), **12b** (d, 1.5 Å resolution, PBD ID 4K6V), **17a** (e, 1.4 Å resolution, 4XQA) and **17b** (f, 1.6 Å resolution, 4XQB) were contoured at 3σ and shown with a radius of 2.4 Å around the ligand (dark blue). For **17a** (e), an additional map at 2σ shows an ordered density for the ligand base (cyan). One inhibitor molecule is simultaneously bound with its three terminal sialic acid residues. Serine 344 (Ser344) is within the van-der-Waals radius of the sialic acid moieties and only **17b** (f) is capable of forming hydrogen bonds with this residue. Black dashed lines in a–e indicate distances to Ser344.

pressure, and pachymetry measurements under local anesthesia. The variations in body weights noted are common among rabbits and are probably not related to the treatment. The data showed that intraocular pressures and corneal thicknesses were within normal limits throughout the study period. The results of the slit lamp examinations show that all eyes were of normal status before start of the study. At the examinations at days 2 and 8 no signs of influence of the test or control items were noted on the cornea, the depth of the anterior chamber, the iris or the lens. No signs of irritation of the test or control item on surrounding tissues were noted. We conclude that compound **17a** did not cause irritation in ocular tissues after repeated administration during seven days. In addition, repeated administration did not cause changes in body weights, intraocular pressure, or corneal thickness that would indicate a toxic reaction to the compound.

Conclusions

Two generations of new trivalent sialic acid derivatives have been designed, synthesized and evaluated against Ad37 infection of HCE cells. The design of these Ad37 inhibitors was based on the structural features of **ME0322**, a previously characterized trivalent Ad37 inhibitor, as well as on robust chemical reactions allowing the rapid access to the target compounds. Thus, in this study we set out to improve the potency of **ME0322** by revising the linker strategy. First-generation ligands **11a** and **11b** efficiently prevented the attachment and infection of Ad37 virions to HCE cells while second-generation compound **17a** was determined as the most potent inhibitor of Ad37 infection of HCE cells. In addition, the original lead potency was greatly improved. Co-crystallization of the trivalent sialic acids in complex with the Ad37 fiber knob allowed the unambiguous placement of the ligands and therefore confirmed a one-to-one binding mode between the compounds and the Ad37 fiber knob. However, ligand–receptor interactions originating from the linker and/or core fragment were not observed.

In this study, we also explored the effect of an increased lipophilicity at the *N*-acyl moiety (**12a** and **12b**). The trivalent ligands, albeit a great inhibitory potency, were less efficient in preventing Ad37 infection of HCE cells compared to their unmodified analogues (**11a** and **11b**). The analysis of the

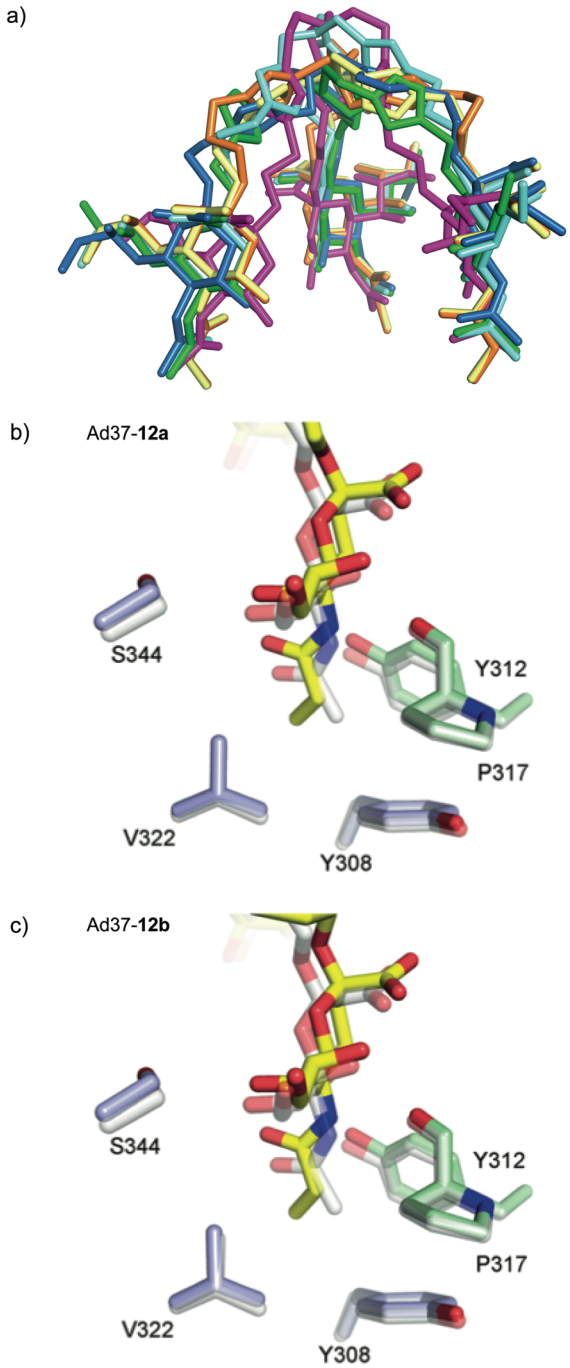


Fig. 5 (a) Superposition of the inhibitor compounds (**11a** = blue, **11b** = cyan, **12a** = yellow, **12b** = orange, **17a** = green, **17b** = magenta). The corresponding protein chains were aligned in PyMOL. **17b** possesses additional internal order, its sialic acid moieties are rotated sideward and its linker forms a bell-like structure; (b) and (c) hydrophobic interactions of **12a** (b) and **12b** (c). For the two compounds, the additional methyl group of the propionic acid group is facing away from Y308 and V322, pushing the whole ligand up.

Ad37-inhibitor structures evidenced an unfavorable interaction between the additional methyl group and the target protein that could explain the lower potency of these compounds.

In conclusion, we have synthesized highly potent inhibitors against Ad37 infection to HCE cells. Compound **17a**, the most potent inhibitor, was accessed using a straightforward synthetic route. Some advantages of such drugs are that (i) they can be used for topical treatment, which would overcome systemic treatment related challenges such as rapid serum clearance and poor cellular uptake of carbohydrate based drugs, (ii) their mechanism of action is on the extracellular level whereas most other antiviral drugs act intracellularly (minimizing side effects), and (iii) the active molecule is a normal carbohydrate (also minimizing side effects). Our investigation of ophthalmic toxicity in rabbits show that this type of trivalent sialic acids are well tolerated and thus have the potential to be developed into novel treatments of viral eye infections.

Experimental section

General chemical procedures

^{1}H NMR and ^{13}C NMR spectra were recorded with a Bruker DRX-400 spectrometer at 400 MHz and 100 MHz respectively. NMR experiments were conducted at 298 K in $CDCl_3$ (residual solvent peak = 7.26 ppm (δ_H)), CD_3OD (residual solvent peak = 3.31 ppm (δ_H) and 49.00 ppm (δ_C)) and D_2O (residual solvent peak = 4.79 ppm (δ_H)). LCMS was carried out with a Waters LC system equipped with an Xterra C18 column (50 × 19 mm, 5 μm, 125 Å), eluted with a linear gradient of CH_3CN in water, both of which contained formic acid (0.2%). A flow rate of 1.5 mL min^{-1} was used and detection was performed at 214 nm. Mass spectra were obtained on a Water micromass ZQ 2000 using positive and negative electrospray ionization. HRMS was performed using a Bruker MicroTOF II mass spectrometer with electrospray ionization (ES^+); Tune Mix ESI solution was used for the calibration. Semi-preparative HPLC separations were performed on a Gilson system HPLC, using a Nucleodur C-18 column HTEC 5 μm (VP 250/21) with a flow rate 20 mL min^{-1}, detection at 214 nm and eluent system: A. aq. 0.005% HCOOH, and B. 0.005% HCOOH in CH_3CN. Column chromatography was performed on silica gel (Merck, 60 Å, 70–230 mesh ASTM). Thin Layer Chromatography (TLC) were performed on Silica gel 60 F_{254} (Merck) with detection under UV light and/or development with 5% H_2SO_4 in EtOH and heat. Optical rotations were measured with a Perkin-Elmer 343 polarimeter at 20 °C. Organic solvents were dried using a Glass Contour Solvent Systems (SG Water USA) except CH_3CN and MeOH that were dried over molecular sieves 3 Å. All commercial reagents were used as received. **ME0322** was synthesized according to published procedure.[21] All target compounds were ≥95% pure according to HPLC UV-traces. Statistics were calculated using GraphPad Prism (GraphPad Software, Inc, La Jolla, CA).

Synthetic procedures

General method for the glycosylation reaction. Glycosyl donor **1** or **2** (1.0 equiv.) and freshly crushed molecular sieves 3 Å (1.5 g mmol^{-1}) were dissolved/suspended in a mixture of

CH_3CN/CH_2Cl_2 (3 : 2; 35 mL $mmol^{-1}$) at room temperature and under nitrogen atmosphere. 2-Bromoethanol, 3-bromopropan-1-ol, propargyl alcohol or 3-butyn-1-ol (4.5 equiv.) was added and the mixture was stirred for 2 h. The reaction was protected from light and a solution of silver triflate (2.0 equiv.) in CH_3CN was added. The mixture was cooled to −73 °C (−70 °C < t < −75 °C) and IBr (1.4 equiv., 1 M in CH_2Cl_2) was added. The reaction was allowed to proceed for 4.5 h at −73 °C. After completion, DIPEA (6.0 equiv.) was added. The reaction mixture was stirred for a further 30 min at −73 °C and then allowed to warm to room temperature. The mixture was filtered through a Celite® pad, washed with CH_2Cl_2 or CH_3CN and the solvents concentrated to dryness.

General method for the synthesis of azido derivatives. To the bromo derivatives (1.0 equiv.) dissolved in DMSO (40 mL $mmol^{-1}$) were successively added portion-wise sodium azide (6.0 equiv.) and TBAI (2.0 equiv.). The reaction was allowed to proceed for 6 h at room temperature and under nitrogen atmosphere. After completion, the mixture was diluted with CH_2Cl_2, washed with water and brine, dried over $MgSO_4$, filtered and concentrated to dryness.

General method for the *O*-deacylation of sialosides. To peracylated sialoside (1.0 equiv.) dissolved in MeOH (70 mL) was added sodium methoxide (3.9 equiv.). The reaction was allowed to proceed for 3 h at room temperature and under nitrogen atmosphere. After completion, the solution was neutralized by drop-wise addition of glacial AcOH or by Amberlyst® 15. The solvent was then concentrated to dryness.

General method for the synthesis of the first generation of trivalent sialic acid derivatives. To the azido derivative (3.7 equiv.) dissolved in THF/H_2O (1 : 1, 81 mL $mmol^{-1}$) was successively added tripropargylamine (1.0 equiv.), $CuSO_4$ (0.9 equiv.) and sodium ascorbate (0.9 equiv.). The reaction was allowed to proceed at 50 °C for 3–3.5 h and at room temperature for a further 18 h. After complete consumption of the starting azide, THF was evaporated under vacuum and the crude was freeze-dried. The crude solid was dissolved in DMSO and purified by HPLC (A: aq. 0.005% HCOOH in H_2O, B: aq. 0.005% HCOOH in CH_3CN, organic phase gradient 7% to 25%). The collected compound-containing fractions were freeze-dried to afford pure product.

General method for the synthesis of the second generation of trivalent sialic acid derivatives. To a solution of tris (2-azidoethyl)amine (1.1 equiv.) dissolved in THF/H_2O (1 : 1, 81 mL $mmol^{-1}$) was successively added alkyne derivative (4.5 equiv.), sodium ascorbate (0.9 equiv.) and copper(II) sulfate (0.9 equiv.). The reaction was allowed to proceed at 50 °C for 3–3.5 h and at room temperature for a further 18 h. After complete consumption of the starting azide, THF was evaporated under vacuum and the residue was diluted with distilled water and then purified with preparative HPLC (A: aq. 0.005% HCOOH in H_2O, B: 0.005% HCOOH in CH_3CN, organic phase gradient 5% to 20%/30 min.). The collected compound-containing fractions were freeze-dried to afford pure product.

General method for saponification of the methyl esters. To the trivalent methyl ester derivatives (1.0 equiv.) dissolved in MeOH (135 mL $mmol^{-1}$) was added an aqueous solution of LiOH (1 M, 9.0 equiv.). The mixture was allowed to proceed for 9–44 h at room temperature. After completion, the reaction mixture was neutralized with Dowex 50W8 (H^+) or amberlite IR120. After filtration, the solvent was evaporated under vacuum and the crude, dissolved in water, was eluted on a C-18 plug with H_2O. The compound-containing fractions were freeze-dried to yield pure trivalent sialic acid derivative.

Methyl 2-(prop-2-ynyloxy(5-*N*-acetamido-4,7,8,9-*tetra*-*O*-acetyl-3,5-dideoxy-D-glycero-α-D-galacto-2-nonulopyranosyl))-onate (13a). Compound **13a** was synthesized following the general method for the glycosylation reaction. Purification by column chromatography (gradient *n*-heptane/EtOAc) afforded compound **13a**, the corresponding reverse anomer and the glycal product (α/β/glycal (n.d.)). Compound **13a** was used in the next step without additional purification. ESI-MS m/z calcd for $C_{23}H_{32}NO_{13}$ $(M + H)^+$ 530.19 and $C_{23}H_{31}NNaO_{13}$ $(M + Na)^+$ 552.17; found 530.00 and 552.04, respectively.

Methyl 2-(prop-2-ynyloxy(5-*N*-acetamido-3,5-dideoxy-D-glycero-α-D-galacto-2-nonulopyranosyl))-onate (14a). Compound **14a** was synthesized following the general method for the *O*-deacylation of sialosides. Purification by HPLC (A: aq. 0.005% HCOOH in H_2O, B: aq. 0.005% HCOOH in CH_3CN, organic phase gradient 5% to 20%) afforded compound **14a** (178 mg, 48% yield over two steps). ^{1}H NMR (400 MHz, CD_3OD): δ 4.40 (dd, J = 4.3 Hz, J = 15.9 Hz, 1H), 4.33 (dd, J = 4.3 Hz, J = 15.9 Hz, 1H), 3.81–3.91 (m, 5H), 3.78 (d, J = 10.3 Hz, 1H), 3.63–3.72 (m, 2H), 3.60 (dd, J = 1.5 Hz, J = 10.4 Hz, 1H), 3.52 (dd, J = 1.4 Hz, J = 9.0 Hz, 1H), 2.86 (t, J = 2.4 Hz, 1H), 2.72 (dd, $J_{3eq,4}$ = 4.6 Hz, $J_{3eq,3ax}$ = 12.7 Hz, 1H), 2.01 (s, 3H), 1.75 (dd, $J_{3eq,3ax}$ = 12.7 Hz, $J_{3ax,4}$ = 11.8 Hz, 1H). ^{13}C NMR (100 MHz, CD_3OD): δ 175.12, 170.42, 99.44, 80.31, 75.71, 74.90, 72.17, 70.09, 68.38, 64.73, 53.65, 53.47, 52.69, 41.51, 22.71. ESI-MS m/z calcd for $C_{15}H_{23}NO_9$ $(M + H)^+$ 362.15; found 361.99.

Tris (2-azidoethyl)amine (15). A solution of triethanolamine (0.298 g, 2.0 mmol) in 0.5 mL of $CHCl_3$ was slowly added into a stirred solution of thionyl chloride (0.52 mL, 7.0 mmol) in 0.8 mL of $CHCl_3$. After addition, the reaction mixture was heated to reflux temperature for 4 h. After cooling to room temperature the white solid product was filtered and washed with CH_2Cl_2 (1.0 mL × 2) to give tris (2-chloroethyl)amine hydrochloride (0.395 g) in 82% yield after overnight drying under vacuum. Following, Tris (2-chloroethyl)amine hydrochloride (0.198 g, 0.82 mmol) and sodium azide (0.320 g, 4.92 mmol) were added to DMSO (7.0 mL). The resulting mixture was stirred at 92 °C for 22 h. After cooling the mixture was poured into distilled water (40.0 mL) and the solution was alkalized with Na_2CO_3 (10% aq.) to pH = 10, extracted with CH_2Cl_2 (15.0 mL × 3). The organic phase was washed with water (20.0 mL) and then dried over Na_2SO_4. CH_2Cl_2 was concentrated to 1 mL, and then 15.0 mL of THF was added, concentrated again to 1.0 mL, 15.0 mL of THF added and concentrated to 1.8 mL. This THF solution containing 0.8 mmol of tris (2-azidoethyl)amine was used in next step without further purification (note: compound **15** was judged as potentially explosive and was therefore kept in solution at

all times). ^{1}H NMR (400 MHz, $CDCl_3$): δ 3.33 (t, J = 6.2 Hz, 6H), 2.76 (t, J = 6.2 Hz, 6H). ESI-MS m/z calcd for $C_6H_{13}N_{10}$ $(M + H)^+$ 225.13; found 225.33.

Tris ((4-(2-*O*-(methyl (5-*N*-acetamido-3,5-dideoxy-D-glycero-α-D-galacto-2-nonulopyranosyl)-onate))-2-oxomethyl-1*H*-1,2,3-triazol-1-yl)ethyl)amine (16a). Compound **16a** was synthesized following the general method for the synthesis of the second generation of trivalent sialic acid derivatives (76% yield). ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.78 (s, 3H), 4.92 (d, J = 12.6 Hz, 3H), 4.64 (d, J = 12.1 Hz, 3H), 4.30 (t, J = 6.2 Hz, 6H), 3.79–3.93 (m, 18H), 3.60–3.74 (m, 9H), 3.47–3.58 (m, 3H), 3.03 (t, J = 5.9 Hz, 6H), 2.67 (dd, $J_{3eq,4}$ = 4.6 Hz, $J_{3eq,3ax}$ = 12.8 Hz, 3H), 2.00 (s, 9H), 1.74 (t, $J_{3eq,3ax}$ = 12.3 Hz, 3H). ^{13}C NMR (100 MHz, CD_3OD): δ 175.05, 170.75, 145.52, 126.09, 100.06, 74.98, 72.26, 70.33, 68.53, 64.93, 58.51, 55.09, 53.71, 53.59, 41.65, 22.75. HRMS m/z calcd for $C_{51}H_{81}N_{13}NaO_{27}$ $(M + Na)^+$ 1330.5263; found 1330.5186.

Tris ((4-(2-*O*-(5-*N*-acetamido-3,5-dideoxy-D-glycero-α-D-galacto-2-nonulopyranosylonic acid))-2-oxomethyl-1*H*-1,2,3-triazol-1-yl)-ethyl)amine (17a). 17a was synthesized following the general method for the saponification of methyl ester (20 mg, quant.). $[\alpha]_D^{20}$ −11.7 (c 1.0 mg mL^{-1}, H_2O). ^{1}H NMR (400 MHz, D_2O): δ 7.84 (s, 3H), 4.85 (d, J = 12.0 Hz, 3H), 4.61 (d, J = 12.0 Hz, 3H), 4.35 (t, J = 6.0 Hz, 6H), 3.77–3.94 (m, 9H), 3.49–3.76 (m, 12H), 3.04 (t, J = 6.0 Hz, 6H), 2.73 (dd, $J_{3eq,4}$ = 4.5 Hz, $J_{3eq,3ax}$ = 12.4 Hz, 3H), 2.03 (s, 9H), 1.66 (t, $J_{3eq,3ax}$ = 12.3 Hz, 3H). ^{13}C NMR (100 MHz, D_2O): δ 179.90, 174.12, 127.82, 101.35, 73.50, 72.50, 69.00, 68.91, 63.74, 63.40, 52.51, 51.39, 49.84, 40.87, 30.05, 10.36. HRMS m/z calcd for $C_{48}H_{76}N_{13}O_{27}$ $(M + H)^+$ 1266.4974 and $C_{48}H_{75}N_{13}NaO_{27}$ $(M + Na)^+$ 1288.4793; found 1266.4901 and 1288.4742, respectively.

Cell-binding assay

The assay was carried out essentially as described previously but with minor modifications.[16,33] ^{35}S-labeled Ad37 virions (5×10^8 per well) were mixed with or without the trivalent sialic acid derivatives, GD1a glycan or sialic acid at various concentrations in binding buffer (50 μL; BB: Dulbecco's modified eagle's medium containing 1% BSA (Roche AB, Stockholm, Sweden) and HEPES (20 mM, EuroClone, Milan, Italy), pH 7.5). The mixtures were then added to HCE cells prepelleted (1×10^5 per well) in a 96-well microplate. After re-suspension, the mixtures were incubated at +4 °C for 1 h. Finally, unbound virions were washed away with BB and the cell-associated radioactivity was counted by using a Wallac 1409 scintillation counter.

Infection assay

The assay was carried out essentially as described previously with minor modification.[16,33] 6×10^6 non-labeled virions were added to serum free growth media (50 μL), with or without trivalent sialic acid derivatives, GD1a glycan or sialic acid at various concentrations. The resulting mixtures were then added to monolayers of HCE cells in a 96-well plate (3×10^4 cells per well) and incubated at +4 °C. After 1 h, unbound virions were washed away with serum free growth media. Cells were then incubated with growth media containing 1% fetal bovine serum (FBS) at +37 °C. After 44 h of infection, the cells were rinsed with PBS, fixed with cold (−20 °C) 99% methanol for 10 min and incubated with rabbit polyclonal anti-Ad37 antibodies diluted 1 : 100 in PBS (pH 7.4) at room temperature. After 1 h, the cells were washed in PBS and incubated with swine anti-rabbit-IgG Alexa flour 647 secondary antibodies diluted 1 : 250 in PBS for 1 h at room temperature. Finally, the cells were washed in PBS and the number of infected cells was quantified in TROPHOS Plate RUNNER (immunofluorescence microscope).

Surface plasmon resonance (SPR)

The affinity measurements were performed using a surface plasmon resonance BIAcore T200 instrument. Ad37 knob proteins were covalently coupled to a CM5 sensorchip using the amine coupling kit (GE Healthcare), to a concentration of 14–15 ng mm^{-2} (~15 000 RU). Binding of the trivalent sialic acid conjugates **ME0322**, **11a**, **11b**, **17a** and **17b** to the immobilized knob was performed in 10 mM HEPES, 0.15 M NaCl and 0.05% P20 pH 7.4 (1× HBS-EP+, GE Healthcare). The concentrations of trivalent sialic acid used were 400, 200, 100, 50 (twice), 25, 12.5 (twice), 6.25, 3.125, 1.56 and 0.78 μM. The experiment was performed three times for **ME0322**, **11a**, **11b**, **17a** and **17b**. The binding affinities (K_Ds) were calculated using BIAcore T200 evaluation software.

Protein production and structure determination

Expression and purification of Ad37 fiber knob protein were essentially carried out as described previously.[17] For co-crystallization Ad37 fiber knob trimers were concentrated to 13.0–14.4 mg mL^{-1} and then incubated with a 1.3 fold excess of the trivalent sialic acid conjugates **11a**, **11b**, **12b**, **17a** or **17b**. Co-crystals of Ad37-inhibitor complexes were grown following the approach described previously.[21] For complex production of Ad37-**12a**, Ad37 fiber knobs were soaked for 2 h in reservoir solution containing 2 mM conjugate **12a**. Crystals were cryo-protected by using 29% (wt/vol) polyethylene glycol 8000, 50 mM zinc acetate, and 100 mM HEPES (pH 6.9–7.2), then flash frozen in liquid nitrogen, followed by data collection on beamlines X06SA and X06DA at the SLS (Villigen, Switzerland) as well as MX-14-1 at BESSY (Berlin, Germany). Diffraction data were recorded with a Pilatus 6M (for **11a**, **11b**, **17a** and **17b**) and a Pilatus 2M-F (for **12a** and **12b**) pixel detector and processed with the XDS-software.[36] The structures of Ad37 in complex with **11a**, **11b**, **12a**, and **12b** were solved first by molecular replacement using Phaser[37] in CCP4[38] and the native Ad37 knob trimer (pdb-code: 1uxe[35]) as the search model. For the structures of Ad37 in complex with **17a** and **17b**, Molrep[39] was used for the same purpose. All conjugates were unambiguously placed in $F_{\text{co-crystallized/soaked}} - F_{\text{native}}$ difference Fourier maps, incorporated into the model, and refined with restraints from either the Refmac5[40] or the PHENIX[41] monomer library and the PRODRG2 server[42] (for the ligand). Structural refinement was carried out by alternating rounds of model building in Coot[43] and restrained refinement including a combination of isotropic B-factor refinement and the transition-libration-screw method (TLS)[44] with Refmac5 or

PHENIX (for compound **17b**). In the case of **17a**, high resolution allowed for an anisotropic refinement of B-factors for all atoms in PHENIX. For TLS refinement, each protomer in the asymmetric unit was attributed to one TLS group (for Ad37-**12a** two TLS groups per protomer were used). In the case of **17b**, PHENIX automatically assigned TLS groups. Waters were located with ARP/wARP[45] solvent in CCP4. Simulated annealing was carried out with PHENIX. The final models had excellent geometry. All figures were prepared with PyMOL.[46] Statistics on data collection and refinement are given in Table S1† and the simulated annealing omit difference electron density maps for conjugates **11a**, **11b**, **12a**, **12b**, **17a** and **17b** are given in Fig. 4.

Ophthalmic toxicity in rabbit

The experiment was performed in agreement with the EMEA-guideline for local tolerance testing of medicinal products.[47] Compound **17a** was dissolved in 0.9% aq. NaCl to a final concentration of 1 mg mL^{-1}. Six male New Zealand white rabbits were obtained and were allowed to acclimatize before start of the study. The animals were divided into two groups of three. The compound (1 mg mL^{-1} in 0.9% aq. NaCl) or vehicle (0.9% aq. NaCl) was administered topically in a volume of 40 μL hourly (4 administrations during Saturday and Sunday and 8 administrations during weekdays; $5 \times 8 + 2 \times 4 = 48$ administrations in total) in both eyes during seven days. Before the first administration, day 2, and at day 8 the animals were subjected to body weight recordings, slit lamp examinations, intraocular pressure, and pachymetry measurements under local anesthesia. The study was performed at Adlego Biomedical AB (Uppsala, Sweden) with approval of the local animal ethics committee in Stockholm (N169/14).

Acknowledgements

Author contributions: The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. R. C., M. S. and W. Q. synthesized the compounds. J. B. and A. M. L. performed the crystallography experiments. N. C. and R. J. S. performed the cell-based assays. L. F. did the SPR experiments. M. E., N. A., T. S. and R. C. designed the experiments and analyzed the data together with J. B., A. M. L., N. C. and L. F.

Funding sources: Swedish Research Council: Dnr: 521-2010-3078 and 621-2010-4746; Knut och Alice Wallenbergs Stiftelse: Dnr: 2009.0009; Stiftelsen för Strategisk Forskning: Dnr: F06-0011; Torsten Söderbergs Stiftelse: Dnr: M4/11; Deutsche Forschungsgemeinschaft SFB 685.

Notes and references

1 W. S. M. Wold and M. S. Horwitz, in *Fields Virology*, ed. D. M. Knipe, P. M. Howley, D. E. Griffin, A. M. Martin, R. A. Lamb, B. Roizman and S. E. Straus, Lippincott, Williams and Wilkins, Philadelphia, PA, 5 edn, 2007, vol. 2, pp. 2395–2436.
2 H. Ishiko, Y. Shimada, T. Konno, A. Hayashi, T. Ohguchi, Y. Tagawa, K. Aoki, S. Ohno and S. Yamazaki, *J. Clin. Microbiol.*, 2008, **46**, 2002–2008.
3 M. Walsh, A. Chintakuntlawar, C. Robinson, I. Madisch, B. z. Harrach, N. Hudson, D. Schnurr, A. Heim, J. Chodosh, D. Seto and M. Jones, *PLoS One*, 2009, **4**, e5635.
4 M. R. Hilleman and J. H. Werner, *Proc. Soc. Exp. Biol. Med.*, 1954, **85**, 183–188.
5 W. P. Rowe, R. J. Huebner, L. K. Gilmore, R. H. Parrott and T. G. Ward, *Proc. Soc. Exp. Biol. Med.*, 1953, **84**, 570–573.
6 J. De Jong, A. Wermenbol, M. Verweij-Uijterwaal, K. Slaterus, P. Wertheim-Van Dillen, G. Van Doornum, S. Khoo and J. Hierholzer, *J. Clin. Microbiol.*, 1999, **37**, 3940–3945.
7 J. R. Klinger, M. P. Sanchez, L. A. Curtin and D. M. M. B, *Am. J. Respir. Crit.*, 1998, **157**, 645–649.
8 P. Kinchington, E. Romanowski and Y. Jerold Gordon, *J. Antimicrob. Chemother.*, 2005, **55**, 424–429.
9 E. De Clercq, *Nat. Rev. Drug Discovery*, 2002, **1**, 13–25.
10 K. Aoki and Y. Tagawa, *Int. Ophthalmol. Clin.*, 2002, **42**, 49–54.
11 G. Huang, W. Yao, W. Yu, L. Mao, H. Sun, W. Yao, J. Tian, L. Wang, Z. Bo, Z. Zhu, Y. Zhang, Z. Zhao and W. Xu, *PLoS One*, 2014, **9**, e110781.
12 H. Ishiko, Y. Shimada, T. Konno, A. Hayashi, T. Ohguchi, Y. Tagawa, K. Aoki, S. Ohno and S. Yamazaki, *J. Clin. Microbiol.*, 2008, **46**, 2002–2008.
13 H. Kaneko, K. Aoki, S. Ishida, S. Ohno, N. Kitaichi, H. Ishiko, F. Tsuguto, Y. Ikeda, M. Nakamura, G. Gonzalez, K. O. Koyanagi, H. Watanabe and T. Suzutani, *J. Gen. Virol.*, 2011, **92**, 1251–1259.
14 M. Azar, D. Dhaliwal, K. Bower, R. Kowalski and Y. Gordon, *Am. J. Ophthalmol.*, 1996, **121**, 711–712.
15 N. Arnberg, *Trends Pharmacol. Sci.*, 2012, **33**, 442–448.
16 N. Arnberg, K. Edlund, A. Kidd and G. Wadell, *J. Virol.*, 2000, **74**, 42–48.
17 E. Nilsson, R. Storm, J. Bauer, S. Johansson, A. Lookene, J. Ångström, M. Hedenström, T. Eriksson, L. Frängsmyr, S. Rinaldi, H. Willison, F. Pedrosa Domellöf, T. Stehle and N. Arnberg, *Nat. Med.*, 2011, **17**, 105–109.
18 K. Aplander, M. Marttila, S. Manner, N. Arnberg, O. Sterner and U. Ellervik, *J. Med. Chem.*, 2011, **54**, 6670–6675.
19 S. Johansson, N. Arnberg, M. Elofsson, G. Wadell and J. Kihlberg, *ChemBioChem*, 2005, **6**, 358–364.
20 S. Johansson, E. Nilsson, M. Elofsson, N. Ahlskog, J. Kihlberg and N. Arnberg, *Antiviral Res.*, 2007, **73**, 92–100.
21 S. Spjut, W. Qian, J. Bauer, R. Storm, L. Frängsmyr, T. Stehle, N. Arnberg and M. Elofsson, *Angew. Chem., Int. Ed.*, 2011, **50**, 6519–6521.
22 E. A. L. Biessen, D. M. Beuting, H. C. P. F. Roelen, G. A. van de Marel, J. H. Van Boom and T. J. C. Van Berkel, *J. Med. Chem.*, 1995, **38**, 1538–1546.

23 D. T. Connolly, R. R. Townsend, K. Kawaguchi, W. R. Bell and Y. C. Lee, *J. Biol. Chem.*, 1982, **257**, 939–945.
24 R. T. Lee, P. Lin and Y. C. Lee, *Biochemistry*, 1984, **23**, 4255–4261.
25 S. Andre, D. V. Jarikote, D. Yan, L. Vincenz, G.-N. Wang, H. Kaltner, P. V. Murphy and H.-J. Gabius, *Bioorg. Med. Chem. Lett.*, 2012, **22**, 313–318.
26 I. Azcune, E. Balentova, M. Sagartzazu-Aizpurua, S. J. Ignacio, J. I. Miranda, R. M. Fratila and J. M. Aizpurua, *Eur. J. Org. Chem.*, 2013, 2434–2444.
27 V. Cendret, M. Francois-Heude, A. Mendez-Ardoy, V. Moreau, F. J. M. Garcia and F. Djedaini-Pilard, *Chem. Commun.*, 2012, **48**, 3733–3735.
28 I. S. MacPherson, J. S. Temme, S. Habeshian, K. Felczak, K. Pankiewicz, L. Hedstrom and I. J. Krauss, *Angew. Chem., Int. Ed.*, 2011, **50**, 11238–11242.
29 V. Percec, P. Leowanawat, H.-J. Sun, O. Kulikov, C. D. Nusbaum, T. M. Tran, A. Bertin, D. A. Wilson, M. Peterca, S. Zhang, N. P. Kamat, K. Vargo, D. Moock, E. D. Johnston, D. A. Hammer, D. J. Pochan, Y. Chen, Y. M. Chabre, T. C. Shiao, M. Bergeron-Brlek, S. Andre, R. Roy, H.-J. Gabius and P. A. Heiney, *J. Am. Chem. Soc.*, 2013, **135**, 9055–9077.
30 F. Pertici and R. J. Pieters, *Chem. Commun.*, 2012, **48**, 4008–4010.
31 S. Johansson, E. Nilsson, W. Qian, D. Guilligay, T. Crepin, S. Cusack, N. Arnberg and M. Elofsson, *J. Med. Chem.*, 2009, **52**, 3666–3678.
32 A. Marra and P. Sinay, *Carbohydr. Res.*, 1989, **187**, 35–42.
33 N. Arnberg, A. Kidd, K. Edlund, F. Olfat and G. Wadell, *J. Virol.*, 2000, **74**, 7691–7693.
34 M. Sun, C.-Y. Hong and C.-Y. Pan, *J. Am. Chem. Soc.*, 2012, **134**, 20581–20584.
35 W. P. Burmeister, D. Guilligay, S. Cusack, G. Wadell and N. Arnberg, *J. Virol.*, 2004, **78**, 7727–7736.
36 W. Kabsch, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **66**, 125–132.
37 A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read, *J. Appl. Crystallogr.*, 2007, **40**, 658–674.
38 M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. J. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2011, **67**, 235–242.
39 A. Vagin and A. Teplyakov, *J. Appl. Crystallogr.*, 1997, **30**, 1022–1025.
40 G. N. Murshudov, P. Shubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long and A. A. Vagin, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2011, **67**, 355–367.
41 P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **66**, 213–221.
42 A. W. Schuttelkopf and D. M. F. van Aalten, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2004, **60**, 1355–1363.
43 P. Emsley, B. Lohkamp, W. G. Scott and K. D. Cowtan, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **66**, 486–501.
44 P. D. Painter and E. A. Merritt, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2006, **62**, 439–450.
45 V. S. Lamzin and K. S. Wilson, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 1993, **49**, 129–147.
46 The PyMOL Molecular Graphics System, Version 1.5.0.4, Schrödinger, LLC.
47 Committee for Proprietary Medicinal Products (CPMP). Note for guidance on non-clinical local tolerance testing of medicinal products. CPMP/SWP/2145/00. 1 March 2001. London, The European Agency for the Evaluation of Medicinal Products.

UNPUBLISHED MANUSCRIPT

Purification of Pentameric HAdV-D09 Penton Base

Liaci AM[1], Cupelli K[1], Bachmann P[1], Stierhof Y[2], Stehle T[1]

[1]Interfaculty Institute of Biochemistry, University of Tuebingen, 72076 Tuebingen, Germany;
[2] Center for Plant Molecular Biology, University of Tuebingen, 72076 Tuebingen, Germany

Abstract

Following initial attachment, Human adenoviruses (HAdVs) use their penton base protein to bind to integrins on the host cell surface and mediate cell entry. Almost all HAdVs are recognized by their integrin receptors in a similar fashion through a conserved tri-amino acid motif on a surface exposed loop, making this step of the life cycle a promising drug target. Several structural studies have therefore addressed this interaction, but are complicated by the flexibility of both proteins. Therefore, some essential features such as the binding stoichiometry are still enigmatic to date.

In this study, we set out to generate an expression system that allows studying the interactions between soluble, pentameric penton base and the extracellular domains of integrin $\alpha_v\beta_3$ by single particle cryo-electron microscopy. In this preliminary manuscript, we report the current state of the expression and purification of recombinant pentameric penton base of HAdV-D09. Heterologous expression in *E. coli* gave reasonable amounts of soluble protein that was readily purifiable, but poorly folded and therefore not suitable for structural studies. Therefore, a new generation of constructs for the expression in insect cell systems has been designed and is presented here.

Introduction

Human adenoviruses (HAdVs) use a two-step mechanism to select and enter their host cells. An initial primary attachment step is followed by a secondary interaction that is not capable of mediating attachment but promotes virus entry [1,2]. With the exception of HAdVs F40 and F41, all HAdV types are recognized by α_v integrins through a canonical arginine-glycine-aspartic acid (RGD) motif located in the surface-accessible RGD loop of their penton base (pb) [3]. The pb is a pentameric protein located at the vertices of the HAdV capsid [4]. This protein is considered the 'weak link' in the otherwise highly stable capsid and is the first protein to be released during uncoating. The process of uncoating itself is initialized at the plasma membrane, partially by the influence of integrin binding. For one, the relative movement of primary attachment factors and the more immobile integrins provide sheer stress that promotes fiber loss [5]. Secondly, the interaction between Integrin and pb promotes an untwisting of the pb pentamer that possibly promotes fiber loss and weakens the capsid [6,7]. Thirdly, clustering of integrins by several RGD motifs on the HAdV capsid produces an avidity effect and induces integrin signaling, which ultimately leads to cellular uptake through clathrin-coated pits [1]. Even within the endosome, integrin signaling likely continues to play a role for the passage of the virus-loaded endocytotic vesicles through the actin cortex [8]. Some HAdV types, among them HAdVs B03 and D09, produce excess pb proteins in their cytosol that self-assemble into fiber-equipped T=1 icosahedral particles called penton dodecahedra and are released upon host cell lysis [9-11]. These particles are too small to carry the viral genome, but HAdV-B03 penton dodecahedra have been shown to promote viral spread by engaging its receptor DSG-2 in cell-cell junctions and triggering its autocatalytic cleavage [12,13]. Notably, the engagement of DSG-2 requires the presence of at least two fibers in a precise spatial arrangement and would not be possible with just excess fibers [14].

The interaction between the pb protein and α_v integrins is a crucial step for cell entry and a potentially druggable target. Detailed information about the flexibility, stoichiometry, and the binding interface is of critical relevance for potential drug development. Several studies have addressed this interaction from a structural point of view. The interaction between the $\alpha_v\beta_3$ integrin and a circular RGD peptide was located to a binding site located between the β-propeller domain of the α chain and the I domain of the β chain in a crystal structure [15]. In a 2009 study, Lindert *et al.* solved the complex structure of a complete HAdV-A12 capsid in complex with $a_v\beta_5$ integrin [6]. However, modest resolution and the need to impose icosahedral symmetry averaging for structure solution - which was incoherent with the integrin stoichiometry - impeded the interpretability of the results. Integrin is probably bound in an extended, active conformation, with a proposed binding stoichiometry of four integrin heterodimers per pb as determined by EM and SPR studies [6,16]. Docking experiments showed that several integrins bound by the same pb pentamer are unlikely to adopt identical conformations. Instead, binding of four integrins might lead to an untwisting motion of pb. More detailed insights were provided by Veesler *et. al* in 2014, who determined the complex structure of $a_v\beta_3$ integrin with the monomeric HAdV-D09 penton base (HAdV-D09pb) insertion domain (which forms the integrin interface and contains the RGD loop) using single particle cryo-EM [17]. HAdV-D09pb was chosen because it binds exceptionally well to α_v integrins, and the interaction is strong enough to tether virions to the cell surface, which is usually not the case for HAdVs [18,19]. In addition, it contains a relatively short RDG loop compared to other HAdVs (see **Appendix 6.2.3.** of the main manuscript) and therefore forms a less flexible integrin complex. However, the complex reported in the publication still showed a remarkable level of flexibility and several preferred binding orientations. It is currently unclear whether all of these orientations are possible in a virion context, and whether they represent simultaneously occurring binding modes or steps of a sequential binding.

We sought to extend the existing data by addressing the interactions between $a_v\beta_3$ integrin and the recombinantly expressed pentameric HAdV-D09pb. We designed a pb mutant that is not capable of forming penton dodecahedra. This way, we hope to overcome the problems arising from icosahedral capsid symmetry and at the same time provide a more natural pb-integrin interface. This manuscript reports the current state of the purification of HAdV-D09pb.

Results and Discussion

Construct design

While the expression and purification of the HAdV-D09pb insertion domain is readily done in *E.coli*, classical protocols to purify full-length pb proteins usually involve the expression in insect cells. In an attempt to establish a more economic and less time-consuming purification protocol, we chose an *E.coli*-based approach for protein production. Several publications reported that the expression of pb from bacteria results in insoluble inclusion bodies [20,21]. However, the addition of an N-terminal GST tag was reported to facilitate the solubilisation of the HAdV-B07pb [22]. To this end, our initial HAdV-D09pb construct includes an N-terminal GST tag that can be cleaved off due to a HRV3C cleavage site. In addition, the construct contains a non-cleavable C-terminal hexahistidine tag (**Figure 1A**).

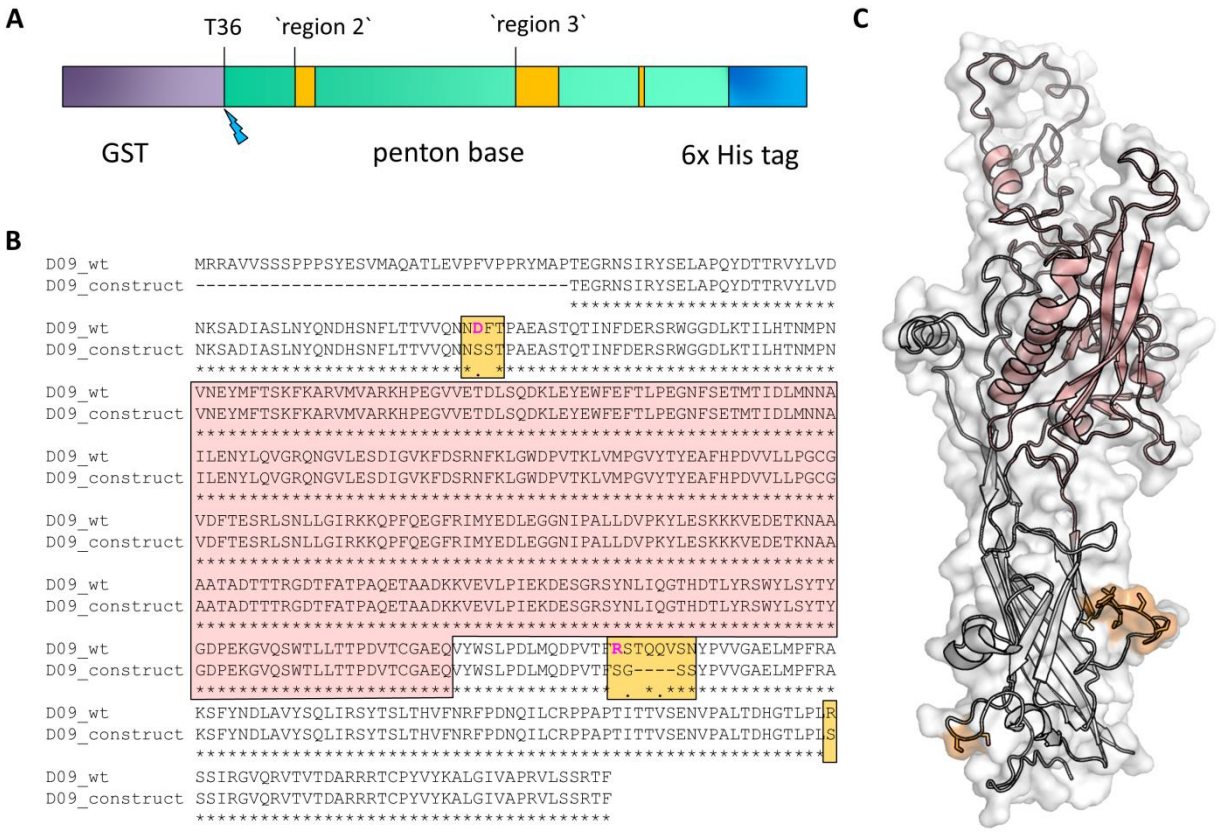


Figure 1 | Penton base construct for expression in E. coli. ***A*** *The construct contains a cleavable N-terminal GST tag and a non-cleavable C-terminal hexahistidine tag for immobilized metal affinity chromatography (IMAC). The location of a human rhinovirus 3C protease (HRV3C) binding site is indicated with a blue symbol; mutations that were introduced to prevent dodecahedron formation are highlighted in orange.* ***B*** *Sequence alignment of the HAdV-D09pb deletion variant with the wildtype sequence. The insertion domain used to solve the cryo-EM structure of monovalent interactions is highlighted in red; mutations that were introduced to prevent dodecahedron formation are highlighted in orange.* ***C*** *Homology model generated with SWISS-MODEL based on the HAdV-C02pb structure (PDB ID 1X9T). The outward-facing side of the monomeric pb is displayed as cartoon. Insertion domain and mutations are highlighted as in B.*

We altered several regions of the native HAdV-D09pb sequence (Uniprot ID Q9EA46_ADE09) in order to prevent the self-assembly into dodecahedral particles (**Figure 1**). In particular, two loops at the base of pb are known to form a homo-interface between neighboring pentamers and have been coined 'Region 2' and 'Region 3', respectively [4,10]. The key residue of 'Region 2' is an aspartic acid (D100 in HAdV-D03, D87 in HAdV-D09) that is conserved in all HAdV types and lies within an 'NDFT' motif in both species B and D HAdVs. This residue is involved in a salt bridge with the key residue of 'Region 3', an arginine (R425 in HAdV-B03, R400 in HAdV-D09) situated within an 'RSTQ' motif in both species. The mutation of either of these two residues partially disrupted dodecahedron formation of HAdV-B03. In order to thoroughly disrupt the interface potentially formed by these two motifs, we mutated residues DF 87-88 of 'Region 2' to serines, and replaced eight residues in 'Region 3' containing the RSTQ motif with a minimal SGGS motif. In addition to the 'Region 2' and 'Region 3' motifs, the N-terminus of HAdV-B03pb was found to critically stabilize dodecahedral particles by a strand swapping mechanism between neighboring pentamers similar to that seen in polyomavirus capsids. In HAdV-B03, the first β-strand of the central β-sandwich starts at residue 64 – however, a truncation of the first 47 amino acids is already enough to substantially decrease dodecahedron formation. To account for this, and for the fact that the disordered N-terminus is prone to proteolytic cleavage, we truncated the first 34 AA of our HAdV-D09 construct (**Figure 1**, equivalent to truncating the first 48 AA in HAdV-B03, which has a longer N-terminus).

Expression and purification in *E. coli* Rosetta

Protein expression performed in *E.coli* Rosetta (batch 17) gave reasonable amounts of soluble, overexpressed protein that could be further purified by GST affinity chromatography (**Figure 2**). In contrast, Nickel affinity chromatography using the full-length construct gave mostly poorly folded and aggregated protein, since the GST tags effectively shield the His tag in assembled oligomers (data not shown). The GST tag was removed using 20U HRV3C protease (Sigma Aldrich) per mg of protein in the presence of 1 mM DTT. An additional Nickel affinity chromatography step did not increase the purity significantly and was therefore omitted (data not shown).

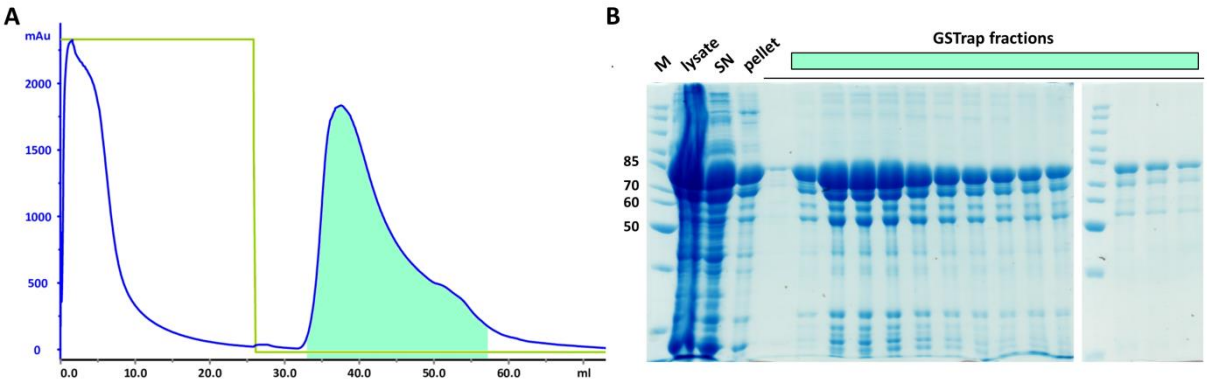


Figure 2 | Expression, lysis, and GST affinity purification of HAdV-D09pb. ***A*** *Chromatogram of a GST affinity purification run using 10.6 mg of cell culture pellet. The elution buffer was freshly prepared and applied through tube* ***A****. The fractions highlighted in cyan were pooled.* ***B*** *Corresponding SDS-PAGE of protein expression, cell lysis, and GST chromatography. Fractions 8-35 (cyan highlight as in A) were pooled. The samples were diluted 1:5 in order to not overload the gel.*

Expression in *E. coli* results in partial C-terminal truncation

The purified protein showed a characteristic set of co-purifying bands in SDS-PAGE. Mass spectrometry analysis demonstrated that in fact all of these bands contain partial sequences of the expression construct (**Figure 3**). Truncations were assessed using MALDI and functional assays (ability to bind GST columns, IMAC after unfolding, time course of HRV3C digest).

While a small subset of pb proteins is cleaved within the RGD loop region, the other truncations occurred at the C-terminus of the protein. C-terminal truncations appear already during protein expression and most probably originate from preliminary termination of translation. Cloning of the construct into a T7 vector, testing of several protease inhibitors, and extensive trials using different *E. coli* cell lines as well as different expression and induction conditions did not improve the expression quality (data not shown). While the main truncated species only lack a hidden β-strand that seems to not be critical for protein stability, the proteolytic digest of the RDG loop is of critical importance for protein quality. Protein unfolding in 6M guanidinium hydrochloride trials served to separate full-length from truncated protein through the C-terminal his tag, but refolding was unsuccessful and is challenging due to the large insertion domain (data not shown). Attempts to purify truncated pb proteins from fully expressed versions using careful IMAC steps were also unsuccessful (data not shown). The most probable reason for this is that the truncated pb monomers are still folded and incorporated into pentamers along with full-length pb versions that still contain the His tag. In this case, a fraction of only 20% truncated protein would result in the loss of 100% of the pb protein if distributed statistically.

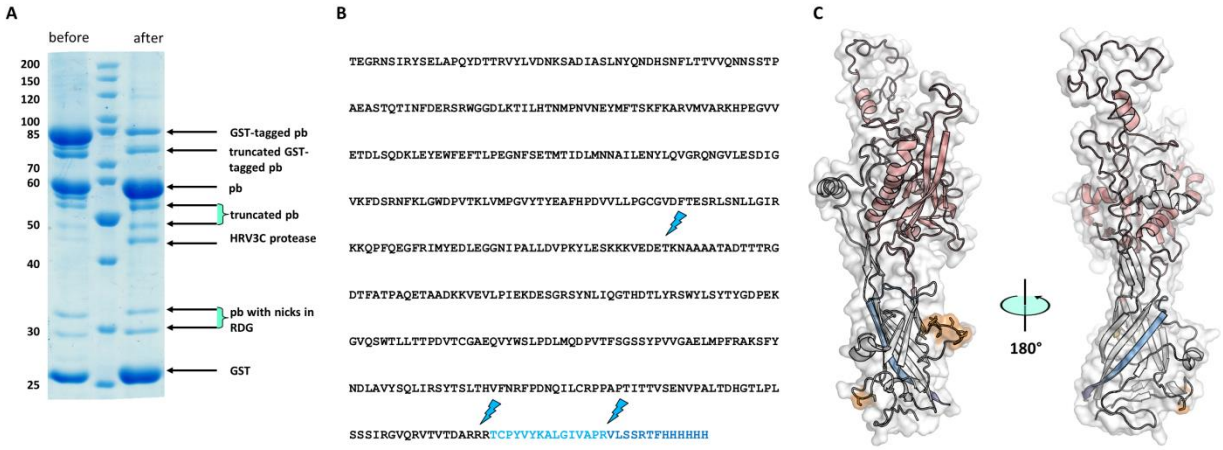


Figure 3 | Characteristic truncations occurring during expression of HAdV-D09pb. ***A*** *SDS-PAGE of HAdV-D09pb before and after HRV3C-cleavage. The bands were analyzed and assigned using MALDI. Longer incubation times result in near-complete proteolytic cleavage.* ***B*** *Approximate locations of the truncations as assessed by MALDI. The C-terminal truncations are likely to occur by incomplete translation, while the RGD-loop nicks (upper flash) are most likely the result of proteolytic hydrolysis during expression.* ***C*** *The untranslated C-terminal region is coding for a β-strand hidden inside the lower pentamer interface.*

Expression from *E. coli* results in poor protein folding

The quality and folding of purified HAdV-D09pb was assessed using a combination of anion exchange chromatography (IEX), size exclusion chromatography (SEC), and negative-stain transmission electron microscopy (TEM) (**Figure 4**). IEX resulted in the separation of several peaks that were subsequently analyzed by SEC and TEM. Some fractions contained the bacterial chaperone GroEL, which has been reported to co-purify with HAdV-B07pb upon expression in *E. coli* and is generally an indicator for the present of substantial amounts of misfolded protein (**Figure 4C,E**). GroEL could be separated from pb by anion exchange chromatography (IEX) using a shallow NaCl gradient on a MonoQ column. However, the remaining pb still

eluted in a broad peak upon SEC (**Figure 4B**). TEM analysis of the main peak fraction showed that the particles are indeed oligomeric but slightly bigger than expected. Furthermore, they do not exhibit visible pentameric features (**Figure 4D**).

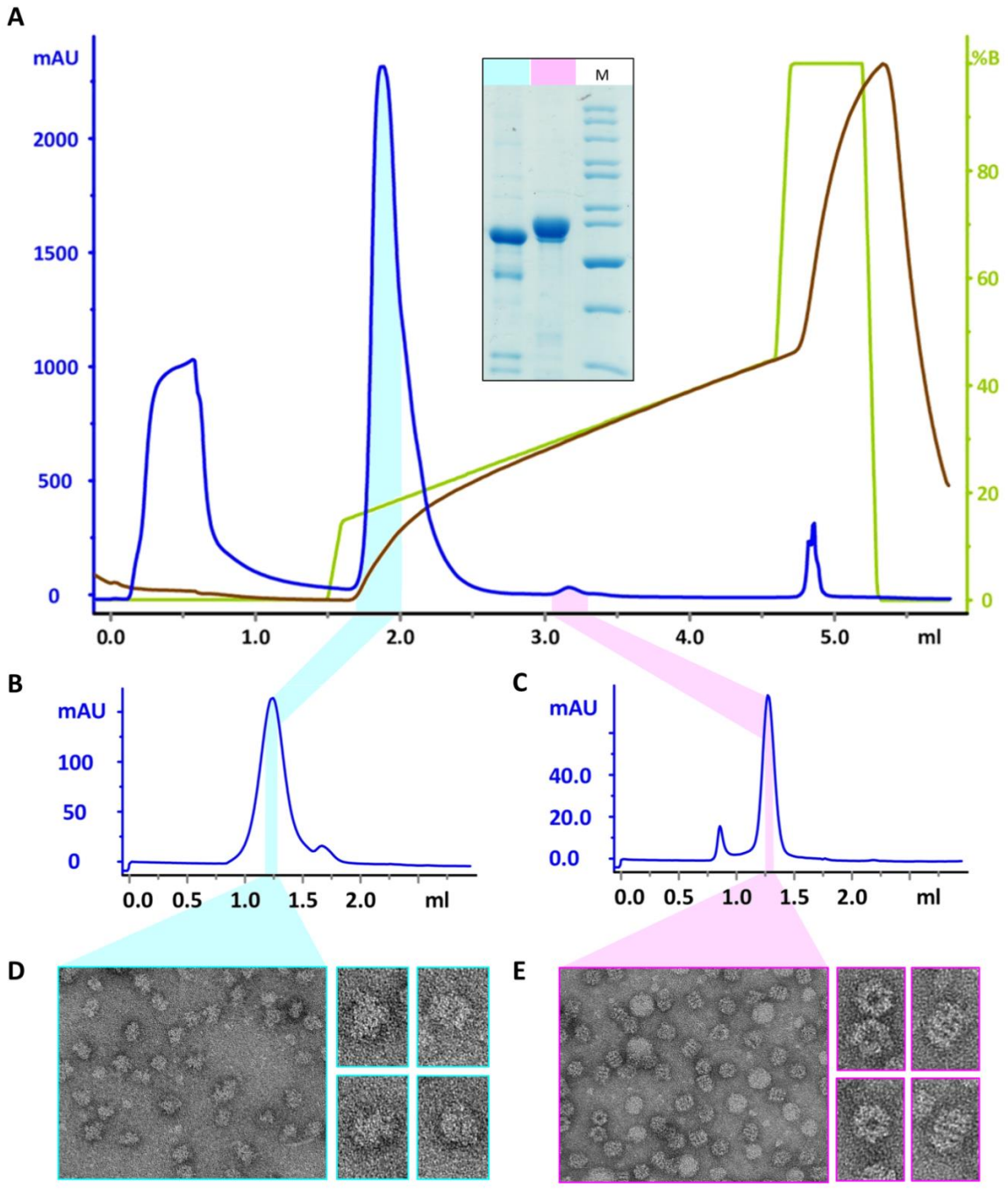


Figure 4 | Ion exchange chromatography, size exclusion chromatography, and negative stain TEM of HAdV-D09pb. ***A*** *Anion exchange chromatography served to separate pb (cyan) from co-purified GroEL (magenta). UV_{280} (in mAU) and theoretical buffer B percentage are depicted on the Y axes. The actual salt gradient is depicted by its conductivity (in mS/cm) as brown line. Inlay: SDS-PAGE analysis of the respective peak fractions from B and C corresponding to HAdV-D09pb and GroEL. The fractions indicated by color were pooled and subjected to size exclusion chromatography (B, C).* ***B*** *HadV-D09pb elutes in a broad peak at a volume corresponding to ~1100 kDa, indicating a high degree of inhomogeneity. The highlighted fraction was analyzed by negative-stain electron microscopy.* ***C*** *GroEL elutes in a sharp peak at the expected volume. The highlighted fraction was analyzed by negative-stain electron microscopy.* ***D*** *HAdV-D09pb particles form oligomers, but these are inhomogeneous and do not exhibit pentameric features. Selected particles are magnified.* ***E*** *GroEL particles exhibit characteristic heptameric features in negative-stain EM pictures.*

In summary, we have developed a strategy to purify HAdV-D36pb expressed in *E. coli*. However, the expression of HAdV pb in bacteria is a challenging task that involves high loss due to faulty protein expression and results in poorly folded oligomers. Previous studies reported the necessity to use GST tags for soluble protein production. In contrast, our data imply that the size of the tag could impair correct pentamer formation. Although the construct has a total length of 717 amino acids (81.65 kDa), the protein is soluble and produced in large amounts by *E. coli*. However, a fraction of about 20% (estimated from SDS-gels) appears to be C-terminally truncated. Since these protein versions are still folded and incorporated into the stable oligomeric quaternary arrangement, they cannot efficiently be separated from fully expressed protein. In addition, the folding of pb is complicated by the large insertion domain that forms the integrin interaction interface. It is also possible that the extensive mutations introduced in our construct hamper the folding process. Folding problems are often indicated by an upregulation and co-purification of chaperones such as the GroEL observed here. Both the C-terminal truncations and co-purification of GroEL were also observed for HAdV-B07, and are thus likely to be a general problem of pb production in *E. coli*. Therefore, we conclude that *E.coli* is not a suitable system for the purification of HAdV pb.

Construct design for the expression of HAdV-D09 and HAdV-A12 penton base in insect cells

Based the findings made for *E. coli* expression, we designed a new construct with more moderate sequence alterations for expression of HAdV-D09pb in *S. frugiperda* cells. This construct contains only a single amino acid substitution. We chose to substitute R400 with a glutamate, since this mutation was reported to efficiently reduce dodecamer formation while still resulting in acceptable protein yields in Sf9 cells. In addition, we chose to further truncate the N-terminus beyond the 'SELA' motif that is thought to mediate the first inter-pentamer contacts. Furthermore, since protein solubility has not been reported to be an issue in insect cell systems, we replaced the bulky GST tag with an N-terminal hexahistidine tag that can be cleaved off using TEV protease, and consequently omitted the C-terminal His tag.

In addition to HAdV-D09, we designed a construct for HAdV-A12pb expression. Working with HAdV-A12pb is advantageous for several reasons. For one, the protein does not form penton dodecahedra and thus does not necessitate mutations in 'Region 2' or 'Region 3'. Secondly, both its variable and RGD loops are among the shortest found in HAdVs. However, the interaction of HAdV-A12pb with integrins seems to be less efficient than for HAdV-D09, and although HAdV-A12 is generally assumed to interact with α_v integrins, specific interactions have so far only been reported for $\alpha_v\beta_5$. We did not include any mutations into the HAdV-A12pb construct, but decided to also truncate the N-terminus since it is reported to be prone to proteolytic digest by cellular proteases. In order to ensure efficient protease cleavage, a short artificial linker of 6 AA is also included. The codon usage of both constructs has been optimized for insect cells.

We are currently in the process of cloning these new constructs. Future experiments will build on the purification procedure established for *E. coli* protein production in order to produce pb pentamers that are of sufficient quality for single-particle cryo-EM studies. The stoichiometry of the pb/$\alpha_v\beta_3$ integrin complex will be assessed by SEC and small-angle X-Ray scattering (SAXS) experiments.

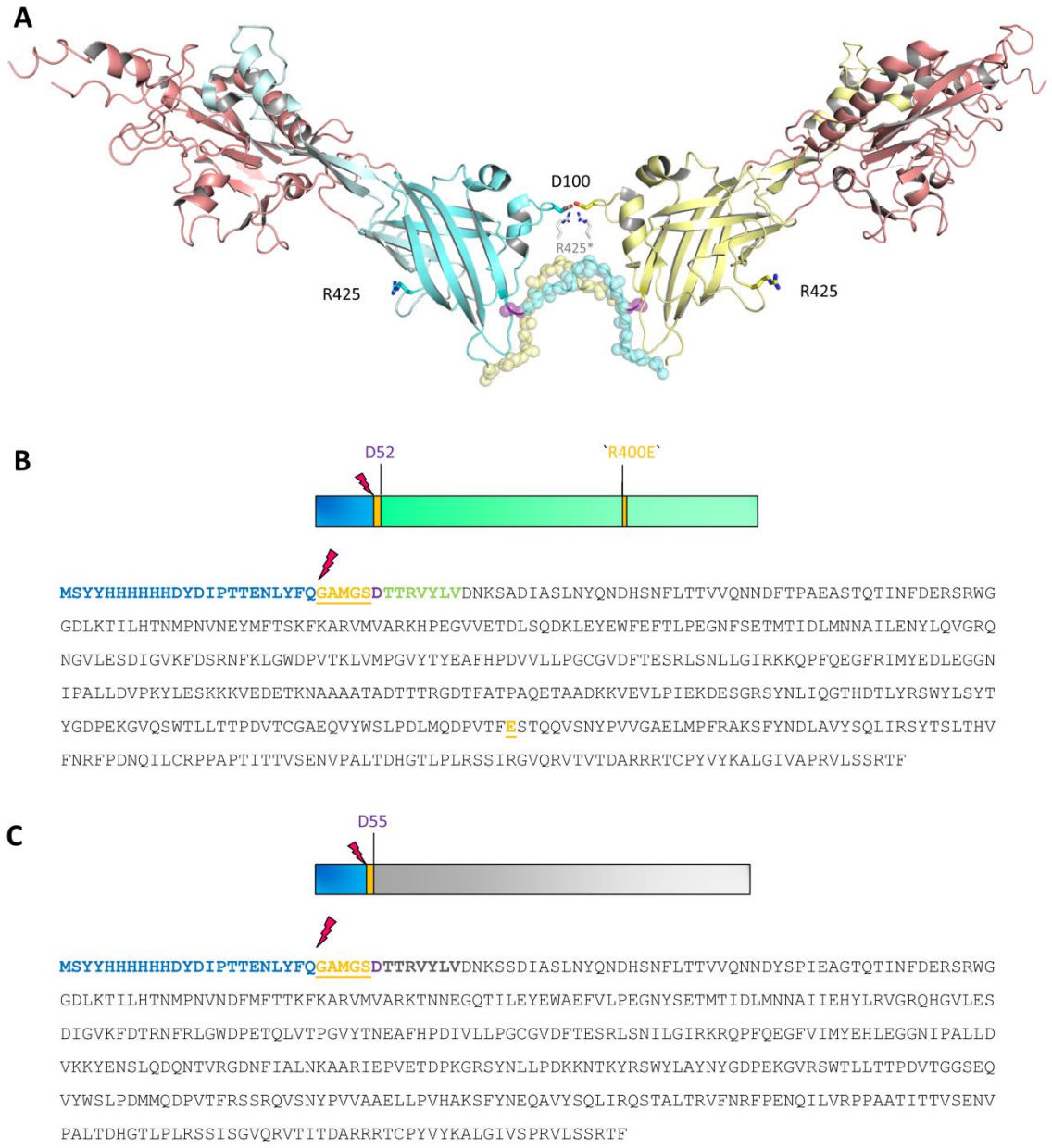


Figure 5 | New constructs for insect cell expression. ***A*** *Inter-pentamer interface observed in HAdV-B03 dodecahedra. Residues D100 and R425 are shown as sticks. R425 sidechains located on the respective neighboring monomers reach into the interface to form a double salt bridge (gray sticks). The unstructured N-terminal amino acids that are part of the strand swapping interface are highlighted as transparent surface. In the new constructs, the equivalent of R425 is mutated to disrupt the salt bridge, and the highlighted parts of the N-terminus are deleted. The location of the first amino acid to be included is highlighted in purple. The insertion domain is shown in red as above.* ***B*** *Schematic representation of the new construct for HAdV-D09pb. The construct starts at D52 and possesses an N-terminal His-tag (blue) and a TEV cleavage site (arrow). Artificial amino acids present at the N-terminus after cleavage are highlighted in orange. The only mutation is R400E (equivalent to the R425E mutant of HAdV-B03, also highlighted in orange).* ***C*** *Schematic representation of the new construct for HAdV-A12pb. The construct starts at D55 and possesses an N-terminal His-tag (blue) and a TEV cleavage site (arrow). Artificial amino acids present at the N-terminus after cleavage are highlighted in orange. No mutations are included. Both constructs are codon-optimized.*

Materials and Methods

Cloning

The gene construct for *E. coli* expression was ordered from Eurofins Genomics (Ebersberg, Germany). The construct is present in a pGEX6P1 vector that contains an IPTG-inducible *tac* promoter and ampicillin (amp) resistance.

Genes for expression in *S. frugiperda* were codon-optimized using the 'GENEius' web tool (Eurofins Genomics) and constructs were ordered in a pFastBac HT B vector.

Expression and cell lysis

For precultures, 20 ml LB containing 50 µg/ml ampicillin and 34 µg/ml chloramphenicol were inoculated with a glycerol stock of *E.coli* Rosetta containing the pGEX6P1-HAdV-D09pb construct and grown over night at 37°C, 160 rpm in a Falcon tube. The precultures were added to 1.5 L LB containing the same selection markers, grown to an OD_{600} of 0.6 at 37°C, and protein expression was induced with 0.1 mM IPTG. Proteins were expressed for approx. 17h at 20°C, 100 rpm. The total volume of the expression cultures varied between 3 and 6 L. Cells were harvested for 10 min at 7000 rpm, 4°C in a SLC-4000 rotor and resuspended in 3 mL lysis buffer (150 mM NaCl, 100 mM HEPES pH 8, 5 mM DTT) per gram bacterial pellet under the addition of 3 µL benzonase, 1 mM $MgCl_2$, and 1 mM PMSF or 1 EDTA-free protease inhibitor tablet (Roche). Lysis was performed with a sonicator (3 runs at 40% amplitude, 3s pulse, 4s pause, 90s total per run) followed by 45 min of centrifugation at 18500 rpm, 4°C in an SS-34 rotor to remove cell debris. The supernatant was filtered using a 0.22 µm cellulose membrane and a vacuum pump.

GST affinity chromatography

The filtered supernatant was loaded at least twice over a new 5 ml GSTrap column at a flow rate of 1 ml/min on a peristaltic pump, then applied to an ÄKTA prime instrument and washed with lysis buffer at the same flow rate until UV_{280} reached baseline level. The protein was eluted in a single step by using lysis buffer supplied with 10 mM reduced glutathione (GSH). Peak fractions were analyzed by SDS-PAGE, pooled accordingly, and stored at -80°C until further use. The protein concentration was assessed using a Nano Drop with a theoretical extinction coefficient of 110,130 and a molecular weight of 81.65 kDa. Typical protein yields were about 20 mg per L cell culture, elution resulted in protein concentrations of 1.5-4 mg/mL depending on how much lysate was applied.

HRV3C protease digest

Batches of 5-10 mg of the frozen GSH eluate were diluted to approx. 0.5 mg/ml in lysis buffer containing 1 mM DTT into a 15 mL Falcon tube. 20 units HRV3C protease (Sigma) per mg protein were added, and the setup was incubated over night at 4°C (60° skewing at 5 rpm). The cleavage efficiency was assessed *via* SDS-PAGE.

Anion exchange chromatography (IEX) on a MonoQ column

5-20 mg of the digested protein were concentrated to at least 2 mg/mL using a 100 kDa concentration filter (Millipore) at spin rates of 3000 rpm in 5 min intervals under frequent manual resuspension to prevent adhesion of the protein to the filter material. In the process, the buffer was changed to IEX buffer a (100 mM NaCl, Tries pH 7) to guarantee efficient immobilization, and excess GST was removed. An extinction coefficient of 67,270 and a molecular weight of 54.82 kDa were used for concentration assessment at the Nano Drop. The protein was filtered and applied to a MonoQ column (GE healthcare) installed on an ETTAN system in 3-4 batches à 500 µL. Prior to elution, the setup was washed with several column volumes of buffer A. A steep initial gradient to 15% buffer B (1M NaCl, Tries pH 7) was followed by a slow gradient (4 mL at 0.1 mL/min) to 45% buffer B in order to effectively separate HAdV-D09pb and GroEL. Intact protein eluted from fractions B2 and D2-D8 (using a self-made program) were used for further experiments.

Size Exclusion Chromatography

The IEX fractions B2 (HAdV-D09pb) and GroEL (D2-D9) were used separately. 25 µL of B2 were applied directly to a Superose 6 column on an ETTAN system using IEX buffer A as running buffer and standard flow rates. Fractions D2-D8 from all four batches were pooled and concentrated to 25 µL in a 100 kDa filter (Millipore). The peak fractions from the respective chromatograms were prepared and directly subjected to negative stain EM.

Negative-stain transmission electron microscopy (TEM)

Freshly prepared peak fractions from SEC runs were diluted 1:10 in SEC buffer, applied to poolroom- and carbon-coated 400 mesh grids and stained with 1% uranyl acetate (UA). The samples were analyzed with a JEOL 1400plus TEM-Microscope (JEOL, Japan) at 120 kV and micrographs were acquired with a Tiete Team F-416 CMOS camera using the Digital Micrograph software (Gaytan, US) and a pixel size of 1.97 Å on the objective scale.

Acknowledgments and Funding

The authors are thankful to Dr. David Veesler for helpful discsussions. This work has been funded by the SFB 685 of the German Research Foundation and the Swedish Research Council (2013-2753 and 2013-8616). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Wickham TJ, Mathias P, Cheresh DA, Nemerow GR (1993) Integrins alpha v beta 3 and alpha v beta 5 promote adenovirus internalization but not virus attachment. Cell 73: 309-319.
2. Wickham TJ, Filardo EJ, Cheresh DA, Nemerow GR (1994) Integrin alpha v beta 5 selectively promotes adenovirus mediated cell membrane permeabilization. J Cell Biol 127: 257-264.
3. Nemerow GR, Cheresh DA, Wickham TJ (1994) Adenovirus entry into host cells: a role for alpha(v) integrins. Trends Cell Biol 4: 52-55.
4. Zubieta C, Schoehn G, Chroboczek J, Cusack S (2005) The structure of the human adenovirus 2 penton. Molecular cell 17: 121-135.
5. Burckhardt CJ, Suomalainen M, Schoenenberger P, Boucke K, Hemmi S, et al. (2011) Drifting motions of the adenovirus receptor CAR and immobile integrins initiate virus uncoating and membrane lytic protein exposure. Cell Host Microbe 10: 105-117.
6. Lindert S, Silvestry M, Mullen TM, Nemerow GR, Stewart PL (2009) Cryo-electron microscopy structure of an adenovirus-integrin complex indicates conformational changes in both penton base and integrin. Journal of virology 83: 11491-11501.
7. Snijder J, Reddy VS, May ER, Roos WH, Nemerow GR, et al. (2013) Integrin and defensin modulate the mechanical properties of adenovirus. Journal of virology 87: 2756-2766.
8. Berk AJ (2013) Adenoviridae. In: Knipe DM, Howley, P. M., editor. Fields Virology. 6 ed. Philadelphia: LIPINCOTT WILLIAMS & WILKINS. pp. 1704-1731.
9. Fender P, Boussaid A, Mezin P, Chroboczek J (2005) Synthesis, cellular localization, and quantification of penton-dodecahedron in serotype 3 adenovirus-infected cells. Virology 340: 167-173.
10. Szolajska E, Burmeister WP, Zochowska M, Nerlo B, Andreev I, et al. (2012) The structural basis for the integrity of adenovirus Ad3 dodecahedron. PLoS One 7: e46075.
11. Norrby E (1969) The structural and functional diversity of Adenovirus capsid components. J Gen Virol 5: 221-236.
12. Lu ZZ, Wang H, Zhang Y, Cao H, Li Z, et al. (2013) Penton-dodecahedral particles trigger opening of intercellular junctions and facilitate viral spread during adenovirus serotype 3 infection of epithelial cells. PLoS Pathog 9: e1003718.
13. Wang H, Beyer I, Persson J, Song H, Li Z, et al. (2012) A new human DSG2-transgenic mouse model for studying the tropism and pathology of human adenoviruses. J Virol 86: 6286-6302.
14. Wang H, Li Z, Yumul R, Lara S, Hemminki A, et al. (2011) Multimerization of adenovirus serotype 3 fiber knob domains is required for efficient binding of virus to desmoglein 2 and subsequent opening of epithelial junctions. J Virol 85: 6390-6402.
15. Xiong JP, Stehle T, Zhang R, Joachimiak A, Frech M, et al. (2002) Crystal structure of the extracellular segment of integrin alpha Vbeta3 in complex with an Arg-Gly-Asp ligand. Science 296: 151-155.
16. Chiu CY, Mathias P, Nemerow GR, Stewart PL (1999) Structure of adenovirus complexed with its internalization receptor, alphavbeta5 integrin. J Virol 73: 6759-6768.
17. Veesler D, Cupelli K, Burger M, Graber P, Stehle T, et al. (2014) Single-particle EM reveals plasticity of interactions between the adenovirus penton base and integrin alphaVbeta3. Proc Natl Acad Sci U S A 111: 8815-8819.
18. Roelvink PW, Kovesdi I, Wickham TJ (1996) Comparative analysis of adenovirus fiber-cell interaction: adenovirus type 2 (Ad2) and Ad9 utilize the same cellular fiber receptor but use different binding strategies for attachment. J Virol 70: 7614-7621.
19. Arnberg N, Kidd AH, Edlund K, Olfat F, Wadell G (2000) Initial interactions of subgenus D adenoviruses with A549 cellular receptors: sialic acid versus alpha(v) integrins. J Virol 74: 7691-7693.
20. Bai M, Harfe B, Freimuth P (1993) Mutations that alter an Arg-Gly-Asp (RGD) sequence in the adenovirus type 2 penton base protein abolish its cell-rounding activity and delay virus reproduction in flat cells. J Virol 67: 5198-5205.
21. Bai M, Campisi L, Freimuth P (1994) Vitronectin receptor antibodies inhibit infection of HeLa and A549 cells by adenovirus type 12 but not by adenovirus type 2. J Virol 68: 5925-5932.
22. Bal HP, Chroboczek J, Schoehn G, Ruigrok RW, Dewhurst S (2000) Adenovirus type 7 penton purification of soluble pentamers from Escherichia coli and development of an integrin-dependent gene delivery system. Eur J Biochem 267: 6074-6081.

Author Contributions:

- Initial constructs for *E. coli* expression were designed and ordered by KC
- Constructs for expression in insect cells were designed and ordered by AML
- Protein purification was established by and carried out by AML
- Biophysical measurements were established by AML and carried out by AML and PB
- Negative stain electron microscopy was carried in the laboratory of YS
- MALDI analysis was carried out in the Laboratory of Dr. Hubert Kalbacher

UNPUBLISHED MANUSCRIPT

The Adenoviral E4ORF1 Protein – Structural and Functional Studies on a Viral Powerhouse Protein

Liaci AM[1], Hehl L[1], Lott A[1], Stierhof Y[2], Stehle T[1,3]

[1]Interfaculty Institute of Biochemistry, University of Tuebingen, 72076 Tuebingen, Germany; [2]Center for Plant Molecular Biology, University of Tuebingen, 72076 Tuebingen, Germany; [3]Department of Pediatrics, Vanderbilt University School of Medicine, Nashville 37232, TN, USA

Abstract

The E4ORF1 gene product (E4ORF1p) of human adenoviruses (HAdVs) represents an important component of the viral early gene machinery. Together with other early gene products, E4ORF1p stimulates the host cell to generate optimal conditions for virus replication. The protein is distantly related to the human dUTPase, but has lost its enzymatic activity. Instead, E4ORF1p interacts with various cellular partners through its PDZ-binding motif and a proposed 'Domain 2' to induce cell-transforming activity. Its functions include the constitutive activation of the phosphatidylinositol 3-kinase pathway and the stimulation of MYC-dependent gene expression. Both of these pathways boost the cellular glucose and nucleotide metabolism, with a result similar to the Warburg-effect in cancer in order to provide building blocks for the generation of virus progeny. Additionally, E4ORF1p of species A-D deregulates epithelial cell polarity by associating with several components of the tight junction complex. HAdV-D36 E4ORF1p was identified as the key obesity-causing factor of the virus due to its ability to induce the differentiation of pre-adipocytes into mature adipocytes. Due to its transforming and potential adipogenic effects, the protein represents a promising new drug target, and since it efficiently promotes glucose uptake, it is also considered a potential anti-diabetic agent.

E4ORF1p homologues from different HAdV species have been extensively studied from a functional point of view, and many assumptions have been drawn about the protein's structure and biophysical properties. As such, E4ORF1 is assumed to exist as functionally distinct trimers and monomers. Here, we present the first comprehensive biophysical characterization of E4ORF1p from HAdV species C. The protein forms stable oligomers, and we did not detect any signs of monomerization. Further, we report for the first time the crystallization of HAdV-C05 E4ORF1p and outline the strategy for structure solution. The resulting structure will lead to novel and important insights into how the protein is able to simultaneously carry out its diverse functions, and will give valuable cues on how to inhibit and engineer it for therapeutic use.

Introduction

Human adenoviruses (HAdVs) are non-integrating DNA viruses that develop a complex relationship with their host cell after infection. The host cell reacts to HAdV infections e.g. by inducing signaling through the surface proteins used for attachment and entry or by triggering immune responses and apoptotic pathways [1]. Using several splice variants of their early gene products, HAdVs are able to counteract these measures. For example, the viruses downregulate antigen presentation on MHC class I molecules or the actions of the tumor suppressor proteins retinoblastoma and p53. At the same time, the virus deregulates the host cell metabolism in favor of increased glycolysis, depolarization, nucleotide synthesis, and ultimately entry into the S phase and replication of virus progeny DNA. In species D HAdVs such as HAdV-D09 and D36, the gene product of E4ORF1 (E4ORF1p) is the main factor driving the host cell to maximum metabolic capacity, rather than the more prominent E1A proteins [2-4]. HAdV-D09 E4ORF1p is able to transform fibroblast cell lines [3], while HAdV-D36 E4ORF1p is sufficient to induce the differentiation of the mouse preadipocyte cell line 3T3-L1 and is thought to be the main player in HAdV-D36-induced obesity [4,5].

The 14 kDa E4ORF1p, which is present in all HAdV types with the exceptions of F40 and F41, is distantly related to the human enzyme dUTPase [6] and shows a remarkable versatility in both structure and function. Its manifold effects include increased glucose uptake, depolarization, and the induction of PI3K signaling that leads to increased nucleotide metabolism (summarized in [7]). In order to accomplish these tasks, E4ORF1p interacts with several different proteins through its two major interaction motifs. The protein is reported to occur as both a monomer and a trimer, and both oligomerization states have different functions in the cytosol and the nucleus [8]. For one, E4ORF1p possesses a C-terminal class I PDZ-binding motif (PBM) which enables it to bind PDZ-containing scaffolding proteins such as MUPP1, ZO-2, and MAGI-1 as a monomer, most likely through a canonical PDZ-PBM interaction called β-sheet augmentation [9-13]. These proteins assemble large complexes at tight junctions and within the cytoplasm and are important for the establishment of cell polarity and the secretion of surface receptors [14]. Similarly, cellular transformation induced by protein E6 of human papillomaviruses also depends on the interaction of its C-terminal PBM with scaffolding proteins such as MUPP1, MAGI-1, and potentially Dlg1 [12,15]. Interestingly, E4ORF1p interacts with the PDZ I and III domains of MAGI-1, both of which also interact with vesicular CAR^{ex8} and control the secretion of this apical attachment factor variant in epithelial cells [12,16]. As a trimer, E4ORF1p of HAdV species A-D interacts with the epithelial protein Discs large 1 (Dlg1) through its PBM [17]. E4ORF1 and Dlg1 constitutively activate phosphatidylinositol 3-kinase (PI3K) in the cytosol by forming a ternary complex that migrates to the plasma membrane and induces PIP_3/Akt signaling. The formation of this complex involves interactions between PI3K and several non-PBM residues of E4ORF1p that have been proposed to co-locate in a so-called 'Domain 2' [18]. Downstream effects of this signaling pathway include an increase of cellular metabolism, enhanced viral replication, and the induction of oncogenicity and cell survival [19]. In addition, Akt signaling induces the secretion of the glucose importer GLUT4 that is stored in secretory vesicles, hence inducing cellular glucose uptake and mirroring the Warburg effect in cancer [20]. At the same time, E4ORF1p (and E4ORF6p) of HAdV-C05 engages and activates the broadly active transcription factor MYC in the nucleus that influences about 30% of the cellular transcriptome [21]. Downstream effects of MYC activation include the upregulation of metabolic genes and a strongly enhanced glycolytic flux. Together, these effects reduce the cellular polarization and turn the infected cell into an active powerhouse, creating optimal conditions for virus replication. Because of its ability to increase glucose uptake by adipose and skeletal muscle cells, the protein is currently under investigation as a potential anti-diabetes agent, and due to its proposed role in the induction of obesity it is a potentially druggable target itself [22].

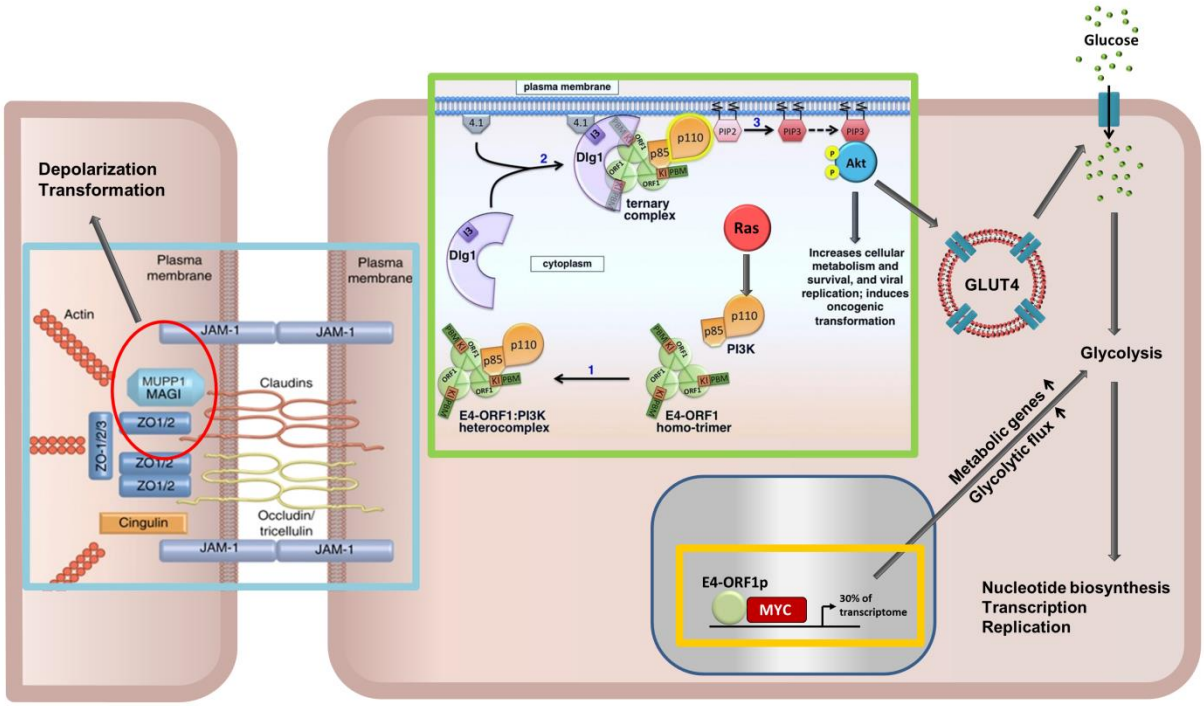


Figure 1 | Reported functions of E4ORF1p. ***Blue Box:*** *monomeric E4ORF1p (here denoted as E4-ORF1 or E4-ORF1p) is reported to interact with tight junction proteins MUPP1, ZO-2, and MAGI-1 (red circle) and to cause depolarization of the host cell.* ***Green Box:*** *Trimeric E4ORF1p (green) is reported to interact with the PI3K heterodimer (orange) and Dlg-1 (violet), and recruits the complex to the plasma membrane, thereby constitutively activating PI3K downstream signaling through the PIP3/Akt pathway. In adipose tissue and skeletal muscle, one of the functions of Akt signaling is the fusion of GLUT4-containing vesicles with the plasma membrane, leading to increased glucose uptake. In the liver, influence on GLUT2 sequestration decreases glucose output (omitted for clarity).* ***Yellow box:*** *E4ORF1p interacts both indirectly and directly with MYC, thereby boosting the expression of metabolic MYC downstream genes and shunting glucose metabolites into the nucleotide synthesis pathway. The blue box was taken from [14] (© 2007, Elsevier B. V.), the green box from [17] (© 2014, Thai et al.), and the yellow box is based on [21] (© 2014, Elsevier B. V.), all with permission.*

From a structural point of view, this plethora of functions requires some unique features. As such, a metastable trimer interface is necessary to fulfill all the reported functions at the same time. Additionally, the protein needs to be able to enter the nucleus, and to form a recognizable 'Domain 2' despite its small size. Information about the structure and biophysical properties of E4ORFp from different HAdV species would contribute significantly to our understanding of how this protein accomplishes all these tasks, and would give valuable information for the development of drugs addressing E4ORF1p and the engineering of the protein as a potential diabetes drug. Therefore, we set out to structurally characterize E4ORF1 proteins from different species using X-ray crystallography, with a special focus on species C HAdVs. This study presents an efficient purification protocol from *E. coli* and an extensive biophysical characterization of the protein. In addition, we present an analysis of a model based on human dUTPase and conditions suitable for the crystallization of HAdV-C05 E4ORF1p.

Results and Discussion

Sequence Analysis and Molecular Modelling

E4ORF1p is highly conserved within the individual HAdV species (sequence identity: HAdV-A >85%; HAdV-B >94%; HAdV-C 100%; HAdV-D >83% and >89% if leaving out HAdV-D49), while its protein sequence differs considerably between species (lowest score 38.4% between HAdV-G52 and HAdV-D49; highest score 70.2% between HAdV-E04 and both HAdV-B07 and B34) (Figure 2A). This variability suggests that there might be different roles of the protein depending on the species.

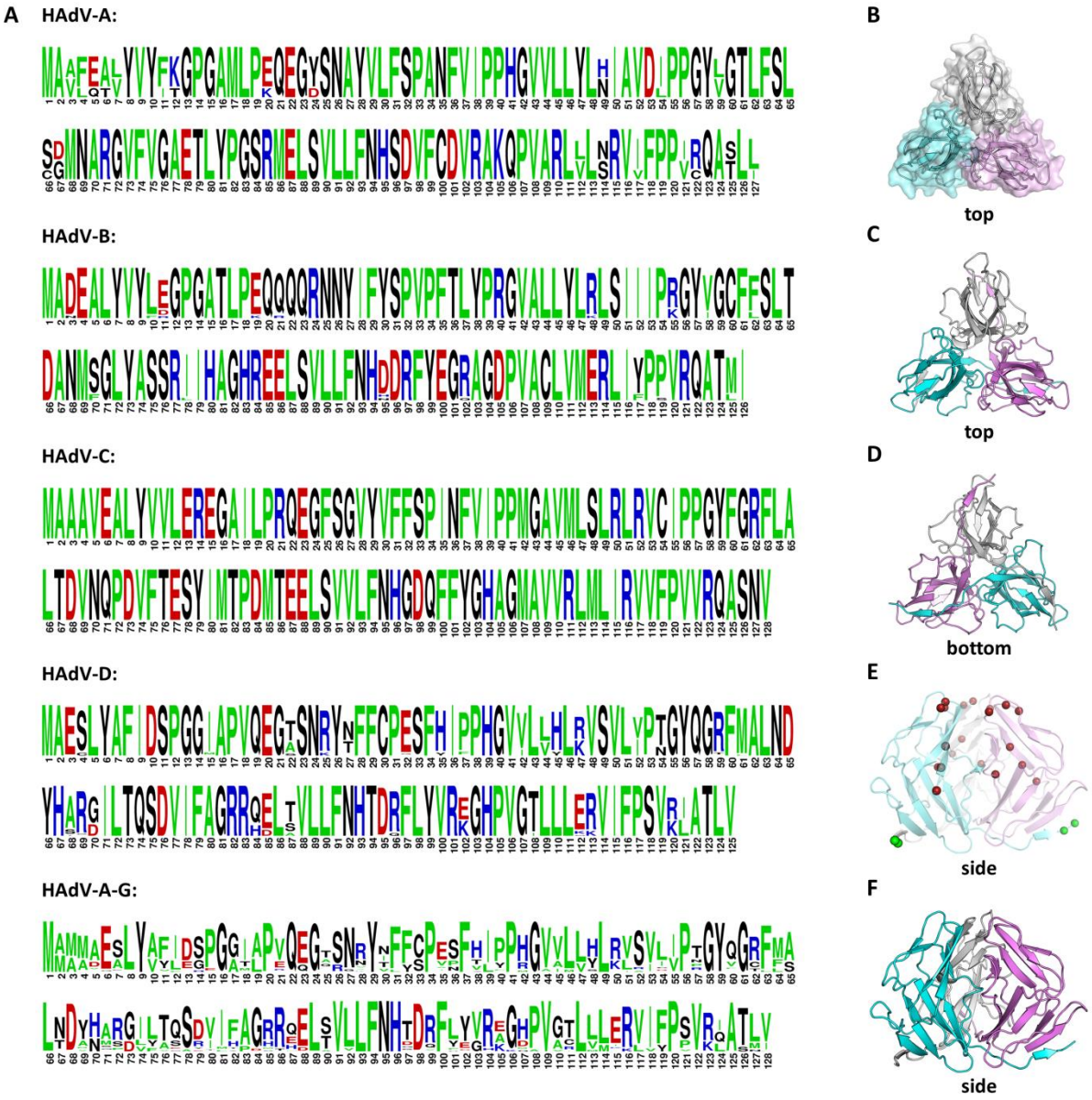


Figure 2 | Sequence analysis and modelling of HAdV E4ORF1p. ***A*** *Sequence variation of E4ORF1p among HAdVs. The figure was generated with WebLogo [23]. Blue: positively charged; Red: negatively charged; Green: apolar; black: polar. HAdV types D19p, D28, D37, and D64 are omitted from the alignment.* ***B-F*** *Different views of the model generated for HAdV-C05 E4ORF1p based on* C. variabilis *dUTPase (PDB-ID 3SO2). The chains are colored individually.* ***B, C*** *Top view with and without surface representation.* ***D*** *Bottom view featuring the proposed 'strand swapping' of C-termini observed for dUTPase.* ***E*** *Location of the residues comprising the proposed 'Domain 2' (red) and PBM (green) in the model. Most 'Domain 2' residues are buried inside the protein. The locations of the respective C_α atoms are depicted as spheres.* ***F*** *Side view. The model was generated with SWISS-MODEL.*

Notably, HAdV types D19p, D37, and D64 (a recombinant of types D19, D22, and D37; formerly listed as type D19) are reported to possess a single base deletion in the E4ORF1 gene (one adenine is missing from a stretch of eight) that causes a frame shift. For this reason, only the C-terminal half of the protein is assumed to be expressed and was termed '7.8kDa protein' [24]. It is currently unclear whether this is due to three independently occurring sequencing errors in the genome, or whether it represents a physiologically relevant truncation/deletion of E4ORF1p from these types. When assuming a sequencing error, all three types cluster closely together in a full sequence alignment, and all of them possess a corneal tropism. However, other HAdVs with corneal tropism (HAdVs D08, D53, D54, and D56) as well as the sequentially highly similar HAdV-D17 E4ORF1 all do not seem to possess this base deletion. In addition, the deposited HAdV-D28 genome sequence possesses a single base substitution at another position that introduces a stop codon, and it is unclear whether this represents a sequencing error, as well.

Currently, no structural data of any of the HAdV E4ORF1p variants are available, and there are no close homologs from other organisms. In order to form initial hypotheses on protein structure and hydrophobicity, we searched for lower homology models for HAdV-C05 E4ORF1p. The nuclear splice variant of human dUTPase shows a sequence identity between 17 and 22%. A number of bacterial and dUTPase homologues identified with SWISS-MODEL [25] display a slightly higher sequence identity. The best hit, trimeric dUTPase from *Chlorella variabilis*, had a sequence identity of 29.8%. Its structure (PDB-ID 3SO2) was used as a template for homology modelling with SWISS-MODEL. With a QMEAN score of -3.15 and a global mean quality estimate (GMQE) of 0.65, the expected model quality is below average, and its overall arrangement is very similar to that found in the template protein. However, the result is highly similar to that obtained by other homology modelling programs (I-TASSER [26], RaptorX [27], HHPred [28]) and may provide useful information about the overall arrangement of the E4ORF1 protein. The predicted quaternary state of the protein is a trimer. Interestingly, despite being cytosolic proteins, all E4ORF1p variants show a very high grand average of hydropathy (GRAVY) index, which is a measure for protein hydrophobicity [29]. Monomeric, globular proteins usually have a much lower GRAVY index, which might indicate the presence of a hydrophobic homo-protein interface that cannot be easily dissolved in an aqueous environment. HAdV-C05 E4ORF1p consists of an unusually high fraction hydrophobic residues (7.8% A, 4.7% I, 9.4% L, 5.5% M, 8.6% F, 3.9% Y, 14.8% V, total 54,7%). The core structure of the homology model consists of four β-sheets that together form two β-sandwiches (**Figure 2 C-E**). Both sandwiches are exclusively held together by hydrophobic contacts. The trimer interface is large (~ 1000Å^2) and accounts for a ΔG value of about -18 kcal/mol per homo-interface according to the PISA server [30], which is higher than values found for the trimeric HAdV fiber knob domains. The predicted interface is relatively hydrophobic (P-value 0.36). Despite the fact that many hydrophobic residues are buried in the protein core and the trimer interface, there is a large hydrophobic patch on the water-exposed protein surface (**Figure 3**). According to the model, the residues comprising 'Domain 2' are mostly buried within the protein, and only few of them would be accessible to PI3K in a trimer. The model predicts a 'strand swapping mechanism (**Figure 2D-E**) that is regularly observed in dUTPase homologues and would serve to further stabilize the trimer.

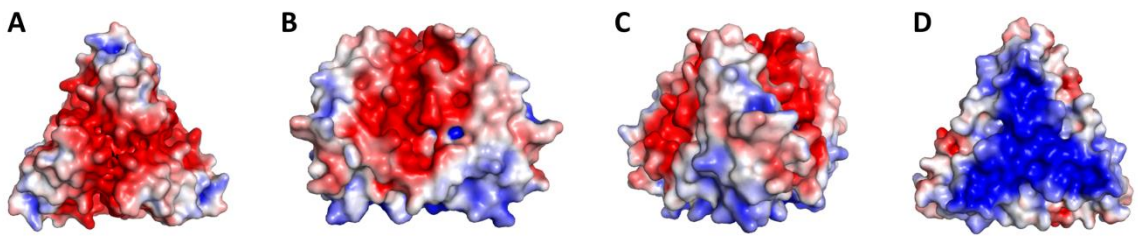


Figure 3 | Electrostatic surface potential (ESP) calculated for the homology model. *The ESP was calculated* at ±3 kT/e *using the Pymol APBS plugin. Positive areas are blue, negative areas red.* ***A*** *top view* ***B, C*** *side view* ***D*** *bottom view.*

E4ORF1p expression constructs

Owing to the small size, potential trimerization, and expected hydrophobicity of the protein, we decided to test three N-terminal tag systems: a small purification tag system that does not interfere with oligomerization, as well as two solubility tags of different sizes that might enhance expression and solubilisation (Figure 4). We chose to start with His-tagged constructs of all candidate types, since these would give an adequate idea about the solubility of the protein itself, and, in a second step, to test if the expression and solubilisation of the best construct can be further enhanced using small ubiquitin-related modifier (SUMO) or maltose-binding protein (MBP) tags. Cloning of His-tagged proteins from HAdV types A12, B03, C05, D36, E04, and G52 was performed from genomic DNA samples (received as a gift from the laboratory of Prof. Niklas Arnberg). We used a restriction-ligation protocol in a pETM11 vector modified to possess a TEV cleavage site. All sequences are listed in [31].

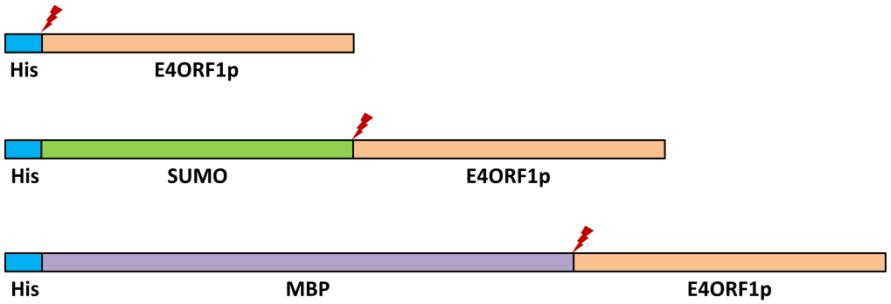


Figure 4 | Constructs for E4ORF1p expression. *The TEV cleavage site is indicated with red arrows. The relative sizes of the bars approximately indicate the relative length of the construct domains. The His-tagged constructs were cloned for HAdV types B03, E04, C05, A12, D36, and G52. The SUMO-and MBP-tagged constructs were cloned only for HAdV-C05.*

We tested the solubility and overexpression of all constructs using a range of standard buffers and *E. coli* cell lines while varying a set of standard parameters (expression time and temperature, induction time point, and concentration of IPTG, see [31]). All tested constructs showed low expression levels, and visible overexpression was only observed for types E04, C05, D36, and G52. Most of the protein was found in the cell pellet, and the largest fractions of soluble protein were achieved for HAdV-E04 and –C05 (Figure 5). Some proteins, such as HAdV-G52 E4ORF1p, formed inclusion bodies. HAdV-C05 E4ORF1p emerged as the best candidate, and the most efficient buffer contained 20 mm imidazole, 50 mM LiCl, and 100 mM ethanolamine (pH 8.5). The optimal expression time was 5h, and the expression strain did not have a significant influence. Using these conditions, a sufficient fraction of the protein could be solubilized. An initial purification protocol was established with a yield of about 0.5 mg from 6L of bacterial culture (data not shown).

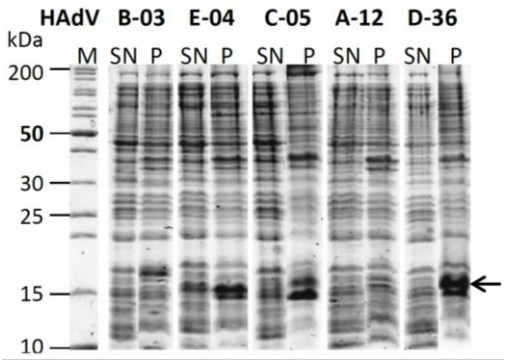


Figure 5 | Solubility of His-tagged E4ORF1p variants. *Cells were lysed by sonication and separated into supernatant (SN) and pellet (P) fractions. The black arrow indicates the expected size of the protein. The highest solubility was observed for HAdV-E04, -C05, and –D36. Figure adapted from [31].*

PDZ domains are among the most ubiquitous protein domains and found throughout all kingdoms of life. A possible explanation for the low yield and insolubility is therefore that the PBM of E4ORF1 targets it to insoluble complexes, thereby contributing to protein toxicity and low solubility. However, an HAdV-C05 E4ORF1 construct with a deleted PBM did not show an altered behavior upon expression (data not shown).

Solubility Tags Improve Protein Expression but not Folding Quality

In an attempt to improve the solubility and expression levels, we designed a second generation of HAdV-C05 E4ORF1p constructs featuring N-terminal solubility tags. Both constructs showed much higher levels of overexpression, but the resulting proteins were almost completely insoluble in most buffers. However, both constructs could be readily solubilized with low concentrations (0,2% w/v) of N-lauroylsarcosine (NLS) if a high dilution of about 50 ml/mg cell pellet was applied (Figure 6A). We tested the ability of several detergents to solubilize both constructs, but apart from the strongly denaturing agent sodium dodecyl sulfate (SDS), only NLS was able to promote solubilisation, while other anionic detergents such as DOC generally failed to do so. Wherever possible, the experiments with NLS were conducted at room temperature in order to stay below the critical micelle concentration (CMC). The construct featuring an N-terminal MBP tag proved difficult to work with, since the tag could not readily be cleaved off and immobilization using nickel or amylose affinity chromatography was not readily reproducible (data not shown). SUMO-tagged E4ORF1p, in contrast, was readily immobilized on a nickel affinity column and eluted at high concentrations of imidazole. To efficiently remove the detergent, the protein was dialyzed at a v/v ratio of at least 1:1000 against a buffer without detergent or imidazole, immobilized again, and washed with 500-1000 mL of the same buffer at flow rates of up to 1 mL/min before elution in a single step. Lastly, the eluate was dialyzed a second time. After this procedure, there was no NLS left in the sample, as was assessed by liquid chromatography-mass spectrometry (LC-MS). The protein could be at least partially digested with Ulp1 protease, which is indicative for a folded SUMO tag. In size exclusion chromatography (SEC) runs, however, both tagged and partially untagged protein as a higher-order species of about 500-600 kDa in size in a broad peak (Figure 6B). Similar results were obtained in dynamic light scattering (DLS) experiments. This species was also quantitatively present at low protein concentrations (data not shown).

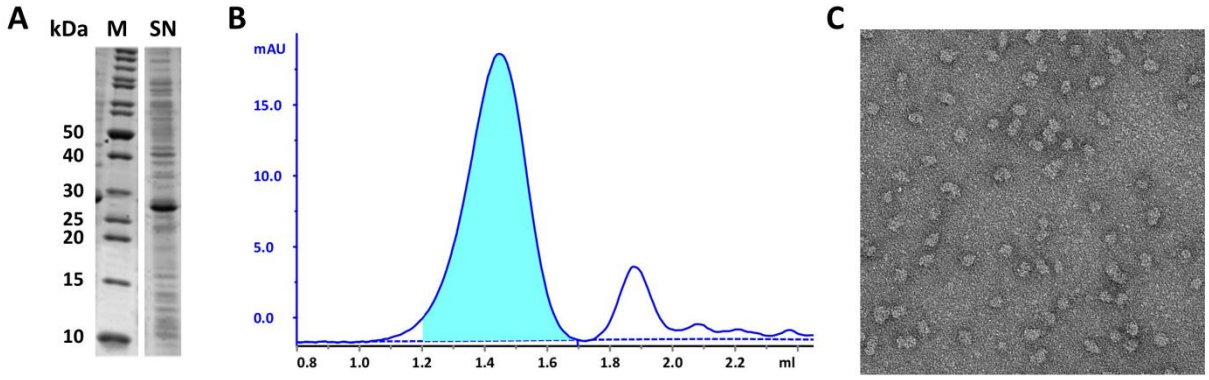


Figure 6 | Key steps of HAdV-C05 SUMO-E4ORF1p purification. ***A*** *About 50% of the protein (M_w 28.1 kDa) could be readily solubilized using 0.2% NLS and a buffer/cell ratio of 50 mL per mg cell pellet. M= marker, SN = supernatant.* ***B*** *Size exclusion chromatography of the tagged protein on a Superose 6 column after Nickel affinity chromatography and thorough dialysis. The protein elutes at a theoretical volume of 500-600 kDa in a broad peak (highlighted in cyan). The smaller peak at 1.88 mL corresponds to overexpressed SUMO. mAU = milli absorbance unit at λ=280 nm.* ***C*** *Negative stain electron microscopy of the peak fraction from B showed that all particles are of similar size, but exhibit little structural homogeneity.*

In theory, at least some of the functions of E4ORF1 might benefit from the formation of such large superstructures. Therefore, we performed negative-stain electron microscopy to test whether these large assemblies (comprised of 15-20 monomers) represent a physiological oligomerization state (Figure 6C). Indeed,

nearly all particles exhibited a similar size, but there was no clear structural homogeneity. Therefore, we concluded that SUMO-tagged HAdV-C05 E4ORF1 is most likely poorly folded and forms aggregates that are limited in size by an outer layer of folded SUMO tags, which keeps the assemblies in solution.

Purification of His-tagged HAdV-C05 E4ORF1p

Since both solubility tags seemed to enhance expression levels at the cost of protein folding, we resorted to scaling up and optimizing the purification of the His-tagged construct. We were able to develop a working purification protocol with improved yields compared to initial trials (Figure 7). The protocol includes the scaling up of bacterial cultures to 18L in order to minimize losses by unspecific binding of the protein to surfaces and thorough cell lysis using a cell homogenizer with the established lysis buffer system. These steps are followed by an initial nickel affinity chromatography step with a reduced imidazole content to account for poor immobilization, proteolytic cleavage of the protein using TEV protease, and a second nickel affinity chromatography step to remove the His-tagged TEV protease and uncleaved protein. A final size exclusion chromatography step served to remove aggregated protein and leftover impurities. This improved protocol resulted in a yield of about 8 mg pure protein from 18L bacterial culture.

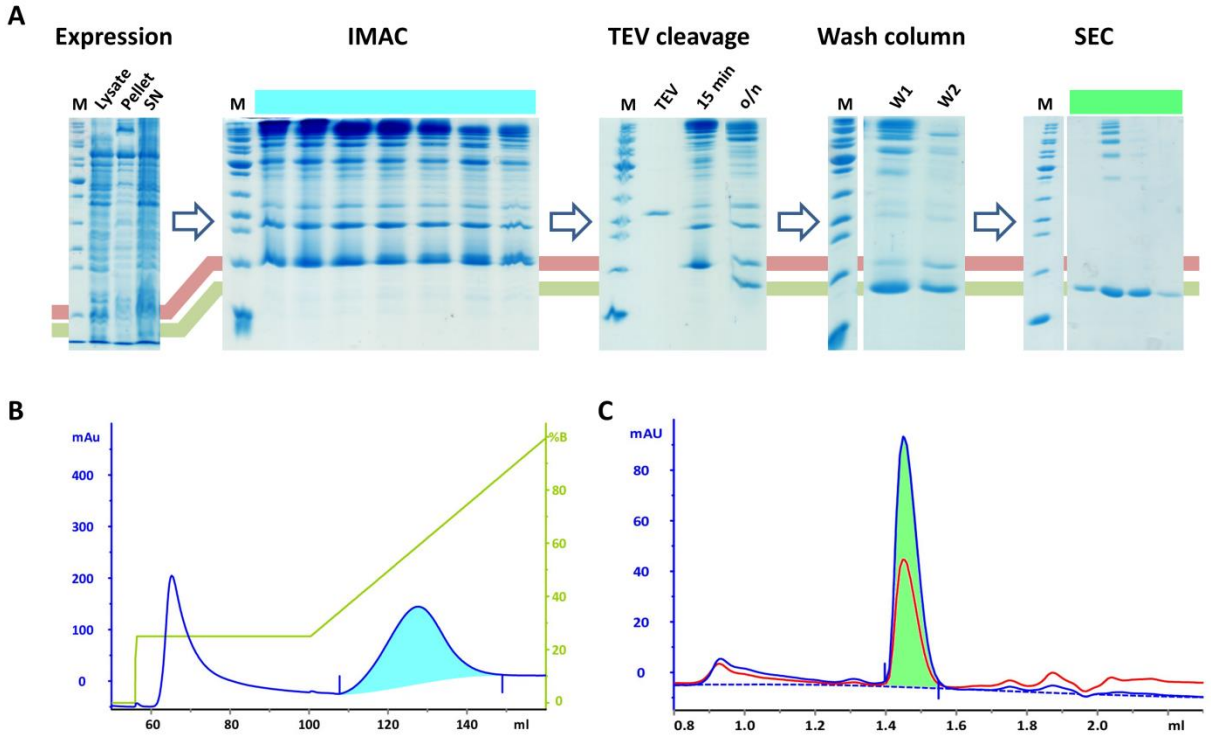


Figure 7 | Key steps of HAdV-C05 His-E4ORF1p purification. ***A*** *Flowchart of the purification progress. The size corresponding to the tagged protein is highlighted by a red band, the size corresponding to (monomeric) untagged protein by a green band. IMAC and SEC fractions were pooled according to the coloring in B and C. Protein expression was followed by thorough cell lysis using a cell homogenizer, Nickel affinity chromatography (IMAC), TEV cleavage during dialysis overnight (o/n), a second IMAC step of which the wash (W) was collected, and a quantitative size exclusion chromatography (SEC) step in which uncleaved proteins and remaining impurities were separated. The pure protein shows a distinct laddering pattern resulting from gel artifacts, as discussed in Figure 8.* ***B*** *Chromatogram of the first IMAC step. The immobilized protein was washed with 25% elution buffer (65 mM imidazole) and eluted in a gradient up to 100% elution buffer (200 mM imidazole). The concentration of elution buffer is shown as a green line (right axis). The peak highlighted in blue was pooled for further use.* ***C*** *Chromatogram of a representative SEC run on a Superdex 200 column using the ETTAN system. The void volume is 0.9 mL, the protein elutes at a volume corresponding to 87 kDa. The peak highlighted in green was used for protein crystallization. The blue line in B,C represents the UV absorption at λ=280 nm (left axis), the red line at λ=254 nm (same scale).*

Biochemical and Biophysical Characterization

HAdV-C05 E4ORF1 possesses several characteristic features. Due to the absence of tryptophan residues, the relative UV absorbance at 254 nm (theoretical extinction coefficient ε= 3412) is about half the value obtained at 280 nm (ε=7450), while the difference is usually much more pronounced for tryptophane-containing proteins (Figure 8A). This distinctive correlation can be monitored and used to judge the purity of the protein. The protein quantitatively eluted at a volume corresponding to about 80kDa from Superdex 200 columns, as compared to 59 kDa when using a Superdex 75 column and 113.5 kDa on a Superose 6 column using the same sample. This size range implies that E4ORF1p forms stable oligomers (n=3-9), but definite proof of the oligomerization stoichiometry cannot be provided at this point.

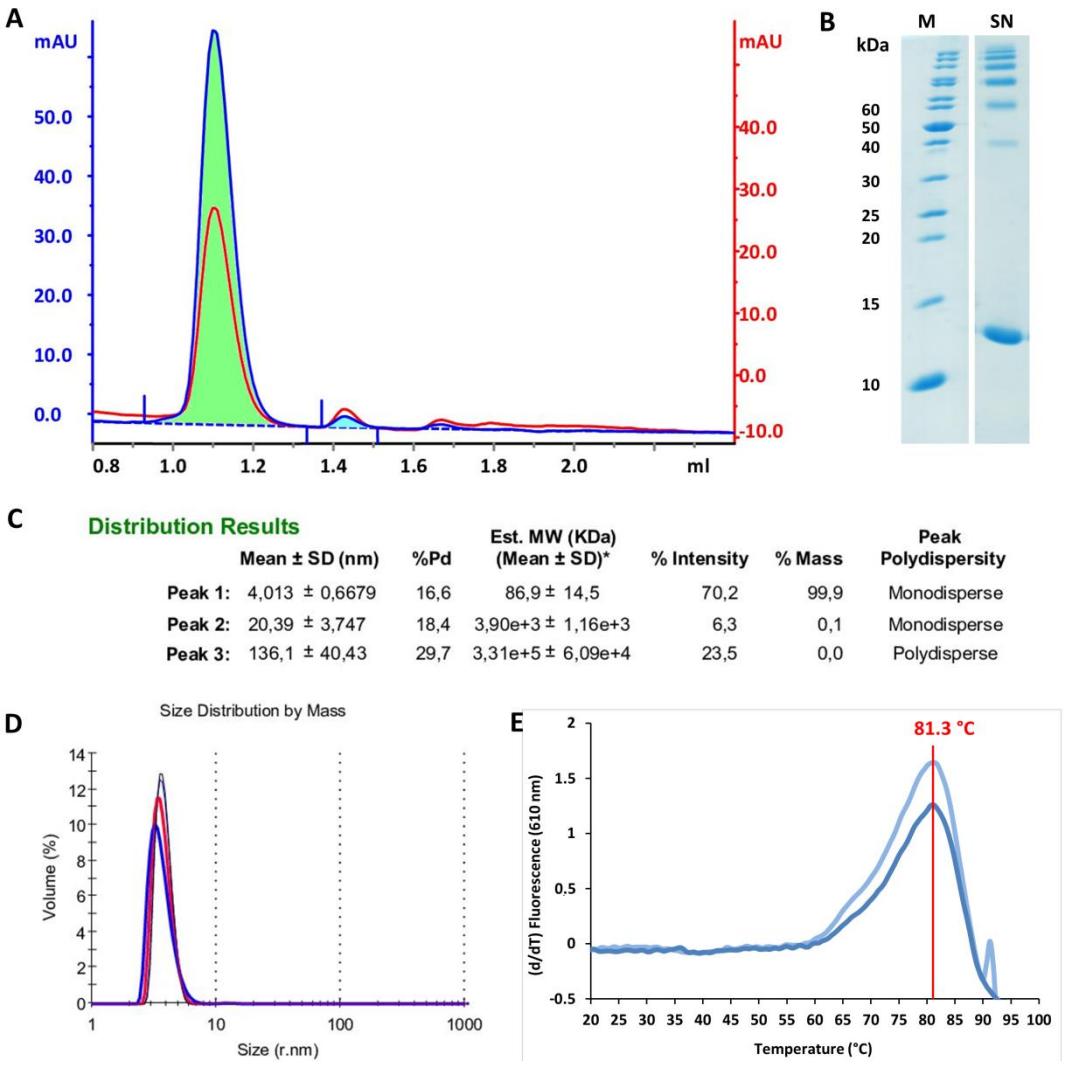


Figure 8 | Biochemical and biophysical properties of HAdV-C05 E4ORF1p. ***A*** *Size exclusion chromatography runs using a Superdex 75 column (here on an ETTAN system). E4ORF1p elutes at a size roughly corresponding to 59 kDa (green peak) according to the standard laboratory calibration curve. Due to the lack of tryptophanes, the extinction at λ=254 nm (red line) is about half the extinction at λ=280 nm (blue line, same scale). A small peak (about 1.8% for the volume of the big peak) elutes at a size corresponding to 14 kDa. However, the spectral wavelength correlation is different, and it is unclear if this peak is formed by monomeric E4ORF1p.* ***B*** *Pure HAdV-C05 E4ORF1p displays distinct laddering bands occurring at multiples of the protein weight (14.4 kDa) on SDS gels. However, this laddering does not correlate to the SEC elution volume.* ***C*** *,****D*** *Dynamic light scattering indicated that the protein is 99.9% monodisperse. Five independent measurements were averaged (colored curves in D) the results are presented in a standard representation. SD = standard deviation; %Pd = percentage of polydispersity.* ***E*** *Differential scanning fluorimetry displays a broad melting profile with a single* T_m *of 81.3°C. The first derivative of the fluorescence signal measured at 610 nm is plotted against the temperature. The measurement was carried out as a duplicate at a protein concentration of 1 mg/mL.*

Upon re-chromatographing the SEC peak fraction, a small peak corresponding to 14 kDa occurs, which might indicate the formation of a monomeric fraction (the M_w of the protein is 14.4 kDa) (**Figure 8A**). However, this peak contains only about 1.8% of the volume of the larger oligomer peak, and its spectral absorption ratio at 280 and 254 nm is quite unusual for a protein. We will repeat this experiment with higher protein concentrations that allow for an SDS-PAGE analysis of the small peak. In SDS-PAGE runs, even the pure protein showed discrete bands occurring at multiples of the molecular weight of a monomer (**Figure 8B**). All of these bands were shown to contain E4ORF1p by mass spectrometry (data not shown). This 'laddering' pattern occurs in all constructs regardless of the tag, and the bands change their size after proteolytic digest. It is unclear how this pattern emerges, but since there are at least nine distinctive bands, the multimerization does not correlate with the results obtained by SEC. Since the laddering is also poorly reproducible, it is likely to be a PAGE artifact.

We additionally subjected the pure protein to dynamic light scattering (DLS) shortly after SEC. The DLS data indicate that the protein is monodisperse in solution (**Figure 8C,D**), with a molecular weight in line with the SEC data. We are planning to assess the secondary structure content of HAdV-C05 E4ORF1p by circular dichroism (CD) spectroscopy. Unfortunately, in initial trials we have been unable to find a buffer condition that allows the recording of a full spectrum while keeping the protein folded (data not shown). Differential scanning fluorimetry (DSF) showed a broad melting profile with a single melting point at 81.3°C, which is very high for an intracellular protein (**Figure 8E**). The protein starts to melt at around 60°C and shows a long melting range.

Crystallization of HAdV-C05 E4ORF1p

We screened for suitable crystallization conditions using a commercially available sparse matrix screen (JCSG+ suite, Qiagen) at a protein concentration of 8-9 mg/mL. The protein readily crystallized in 19 out of 96 conditions (**Table 1**), and all crystals exhibited a highly similar morphology (**Figure 9A**). The first crystals appeared already after 1h (condition A9), while others only grew after 2 weeks (F5). The precipitating components are of heterogeneous nature: eight out of nineteen conditions contain small organic alcohols such as ethanol, glycerol, isopropanol, ethylene glycol, or MPD, while nine conditions contain PEG with a molecular weight of at least 3,350 g/mol. The conditions span a pH range between 5.5 and 9.0, however ten out of nineteen conditions have a pH of 8.5 or higher (including those without additional buffer), which is in agreement with the established buffer system for purification. In contrast, there is no detectable preference for any particular salt. All crystals are rod-shaped with dimensions of about 50 x 5 x 5 µm. Macroseeding of crystals from condition B9 increased the crystal size to about 120 x 12 x 12 µM. The identity of the protein forming the crystals was verified by SDS-PAGE (data not shown).

Table 1 | Conditions of the JCSG+ Suite that resulted in HAdV-C05 E4ORF1p crystals

Well	Condition No.	Salt	Buffer	Precipitant	Final pH	Appeared after
E1	5	0.2 M $Mg(HCO_2)_2$	-	20% (w/v) PEG 3350	-	2d
B2	10	0.2 M $KHCO_2$	-	20% (w/v) PEG 3350		5d
F2	14	0.2 M NaSCN	-	20% (w/v) PEG 3350	-	1d
A3	17	-	0.1 M $Na(CH_3)_2AsO_2$ pH 6.5	40% (w/v) MPD; 5% (w/v) PEG 8000	-	1d
D3	20	0.2 M $MgCl_2$	0.1 M Tris pH 7.0	10% (w/v) PEG 8000	-	1d
H4	32	-	0.1 M Tris pH 8.5	20% (v/v) ethanol	-	1d
A5	33	-	0.1 M H_2NaPO_4 / HK_2PO_4 pH 6.2	25% (v/v) 1,2-propanediol; 10% (v/v) glycerol	-	1d
F5	38	0.2 M $MgCl_2$	0.1 M HEPES pH 7.5	30% (w/v) PEG 400	-	14d
B6	42	0.2 M $MgCl_2$	0.1 M Tris pH 8.5	20% (w/v) PEG 8000	-	1d
G6	47	0.14 M $CaCl_2$	0.07 M NaAc pH 4.6	14% (v/v) isopropanol; 30% (v/v) glycerol	-	1d
C7	51	0.2 M NaCl	0.1 M HEPES pH 7.5	10% (v/v) isopropanol	-	1d
B8	58	-	0.1 M Bicine pH 9.0	10% (w/v) PEG 6000	9.0	5d
G8	63	-	0.1 M Tris pH 8.5	20% (v/v) MPD	8.0	1d
A9	65	0.2 M $MgCl_2$	0.1 M Tris pH 8.5	50% (v/v) ethylene glycol	-	1h
B9	66	-	0.1 M Bicine pH 8.5	10% (v/v) MPD	9.0	1d
D10	76	0.2 M TMAO	0.1 M Tris pH 8.5	20% (w/v) PEG MME 2000	-	5d
F10	78	0.24 M Na malonate pH 7.0	-	20% (w/v) PEG 3350	-	6d
A12	89	0.2M NH_4Ac	0.1 M Bis-Tris pH 5.5	45% (v/v) MPD	-	1d
G12	95	0.2 M $MgCl_2$	0.1 M Bis-Tris pH 5.5	25% (w/v) PEG 3350		2d

The crystals diffracted with a low signal-to-noise ratio to about 3.6 Å at the X06DA beamline (Swiss Light Source, Switzerland) at an exposure rate of 0.8 - 1.5s per 0.1° (**Figure 9B**), which is not unexpected for crystals of this size. The unit cell is orthorhombic (unit cell dimensions: a = 66.6 Å, b = 83.6 Å, c = 132.7 Å; α = β = γ = 90°), and the most probable space group of the crystals is $C222_1$. Matthews Probability calculation, which estimates the most probable copy number based on protein molecular weight, space group, and unit cell volume, suggests that there are three copies of the protein in the asymmetric unit (Matthews coefficient 2.13, solvent content 42%, estimated probability 60%). Unfortunately, the current data quality does not allow the assessment of a self-rotation function (which would give experimental evidence about non-crystallographic protein symmetry) or molecular replacement methods. The best diffraction patterns were observed for larger crystals (e.g. the macroseeds), indicating that increasing the crystal size will also increase diffraction quality. Experiments to reproduce and improve crystal size, nucleation, and diffraction potency are currently ongoing. In initial fine screens, the protein preferably crystallized at 4°C in robotic 600 nL sitting drop setups at much lower concentrations (1.4 mg/mL), as opposed to hanging drop screens that resulted in crystal showers.

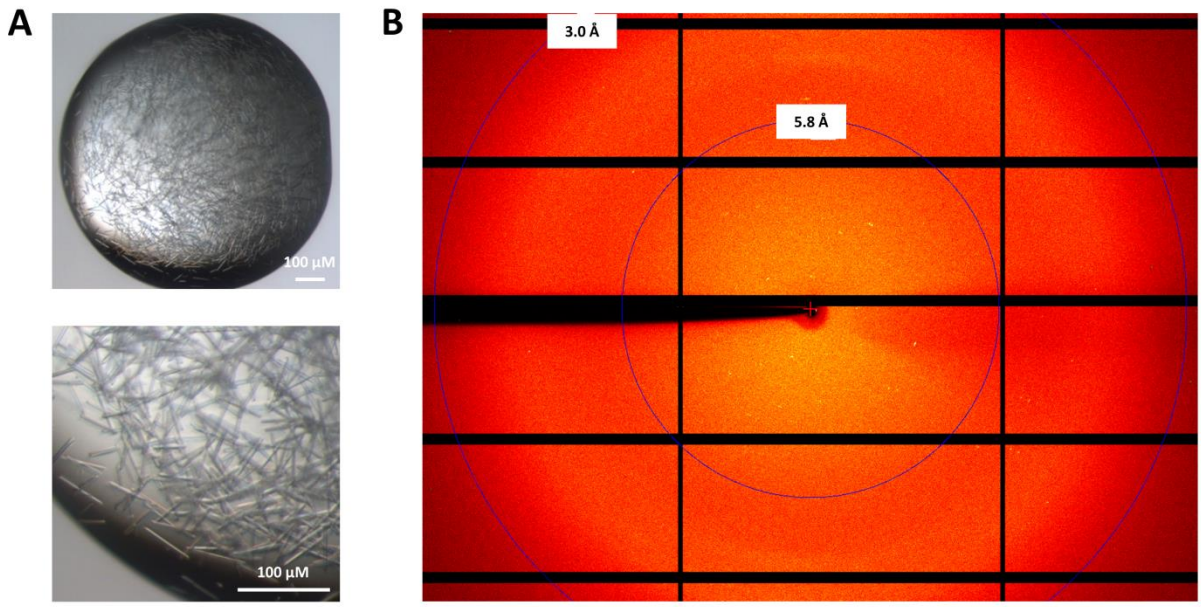


Figure 9 |Features of the preferred crystal form of HAdV-C05 E4ORF1p. ***A*** *This example shows condition C7 from the JCSG+ screen (Qiagen). Crystals are rod-shaped and about 50 x 5 x 5 µm in size. The scale bar indicated the different zoom levels.* ***B*** *Diffraction pattern of an E4ORF1p macroseeding crystal from refined condition B9 collected on a Pilatus 2M detector. The crystal was illuminated at full beam transmission (beamline X06DA) for 0.8s while rotating by 0.1°.*

Summary and Outlook

In summary, we have generated a working model for the E4ORF1p structure that is in line with currently available information, and we present a protocol for the purification of HAdV-C05 E4ORF1p as well as suitable conditions for its crystallization.

According to the data available in the literature to date, E4ORF1p is generally assumed to originate from the trimeric human dUTPase and to switch between a monomeric and trimeric form by an unknown mechanism [8]. To date, our available data for HAdV-C05 E4ORF1p do not show signs of significant monomer formation without the application of detergents. The publication that postulates this *modus operandi* used stringent detergent concentrations for the monomerization of the HAdV-D09 E4ORF1p (RIPA buffer containing a cocktail of the detergents SDS, sodium deoxycholate (DOC), and Triton X-100 or NP-40) [8]. However, the application of detergents seems to lead to the solubilisation of poorly folded protein, at least for the HAdV-C05 variant. It is plausible that, upon unfolding, the protein retains the functions conferred by its PBM while losing the ones that involve tertiary structure elements, such as the interaction with PI3K. The authors of the publication mentioned above that the presence of DOC induces protein trimerization. However, it seems equally possible that DOC simply prevents the protein from unfolding. We will monitor the folding, stability, and oligomerization of HAdV-C05 E4ORF1p in different conditions (pH, ionic strength, presence or absence of detergents or interaction partners) using a combination of CD spectroscopy, DSF, and size exclusion chromatography. As a complementary technique to X-Ray crystallography, we will subject the protein to small-angle X-Ray scattering (SAXS). This in-solution technique can give information about the rough shape and size of a macromolecule without the need to form crystal contacts. At the moment, we cannot exclude the possibility that E4ORF1p of different species behave differently in this respect, and the analysis of species D homologues will shed further light on the observed phenomena.

According to the current structural model, all the residues that constitute the proposed ‘Domain 2’ are either prolines or residues that are buried in the protein core or the trimer interface. These residues have been identified exclusively by a loss of function following mutagenesis [18]. The residues are spread throughout the protein and do not come together to form a recognizable domain or protein interface. In this light, it seems possible that mutating the residues affected protein integrity, and that a loss of PI3K interaction is indeed a

secondary effect. However, the expected accuracy of our model is relatively low, and only the crystal structure of E4ORF1p will provide robust data to evaluate this possibility.

We are planning to employ various crystallization techniques in order to improve crystal size and diffractive qualities, e.g. the screening of temperatures and protein concentration as well as seeding techniques. Once a dataset with an acceptable resolution (below 3.5 Å) is obtained, we will try to use molecular replacement (MR) in order to obtain initial model phases. However, while fast and easy to use, MR requires models with an r.m.s.d. below 2.5 Å, and the fidelity of atom coordinates of search models generally decreases below a sequence identity of about 30% [32]. Since all available structures possess a sequence identity lower than that, it is unclear if MR will produce phases of sufficient quality. Our modeled structure mainly consists of loops alternating with β-strands. The assessment of the β-strand content of HAdV-C05 E4ORf1p with CD spectroscopy might therefore serve as an indicator for model quality, and the SAXS envelope will provide information about the correct oligomerization state. If MR is not a feasible phasing option, we will exploit the high percentage of methionines (7 out of 130 residues, plus one cysteine, accounting for a calculated Bijvoet ratio of 1.9%) in HAdV-C05 E4ORF1p in order to employ experimental phasing using sulfur single wavelength anomalous diffraction (sulfur-SAD). Generally, a content of more than 5% sulfur-containing residues is considered optimal for this approach [33]. Alternatively, we will replace the methionines with selenomethionine (SeMet) in order to use multiple wavelength anomalous diffraction (MAD) for phasing. While generally decreasing the protein yield, SeMet phasing usually generates stronger anomalous signals than sulfur-SAD.

Our data provide useful information on the stoichiometry and stability of HAdV-C05 E4ORF1p. Its high thermal stability might be beneficial for its use as a potential diabetes drug. Future experiments will address the functional and structural variability of E4ORF1 proteins among different HAdV species.

Materials and Methods

Cloning

All sequences were obtained from the NCBI Nucleotide Database using the entry numbers listed by the HAdV working group (date: 11/2016) [34]. E4ORF1p variants of HAdV types B03, E04, C05, A12, D36, and G52 were amplified from genomic DNA (received as a gift from the laboratory of Prof. Niklas Arnberg). All primer sequences are listed in [31,35]. In the process, restriction sites for the endonucleases NcoI and XhoI were introduced at the 5' and 3' ends, respectively. These restriction sites were used for cloning into a pETM11 expression vector. This vector features a T7 promoter, kanamycin resistance, T7 sequences for forward and reverse sequencing (T7, T7 term), and an N-terminal His-tag that can be cleaved off by tobacco etch virus (TEV) protease. All initial constructs contain the full-length E4ORF1 and the artificial N-terminal sequence 'MKHHHHHHPMSDYDIPTTENLYFQGA'. After TEV cleavage, only the two C-terminal residues ('GA') of this stretch remain. The results were verified by Sanger sequencing. Except for HAdV-C05, all sequences corresponded to the listed genomic information. In the case of HAdV-C05, two conserved silent mutations were observed in all sequenced clones and likely represent a genetic variation of the template. All sequencing results can be found in [31]. All constructs were cloned into *E. coli* strains XL10 gold, LeMo, Arctic Express, and BL21 DE3.

MBP- and SUMO-tagged constructs were obtained by amplification of the inserts by PCR from the pETM11 vector construct. The 3' restriction site was replaced with a NotI site in the process. The amplified genes were inserted into pETM41 and pMH-SUMO vectors. Both vectors contain T7 promoters as well as a kanamycin resistance and T7 sequencing options. pETM41 features a TEV-cleavable N-terminal MBP tag, and pMH-SUMO an Ulp1-cleavable N-terminal SUMO tag. Both proteins leave only a short 'GA' stretch at the N-terminus after proteolytic digest.

The pBM (sequence 'ASNV') was deleted from His-tagged HAdV-C05 E4ORF1 in a site-directed mutagenesis reaction that replaced A125 with a stop codon using a standard lab procedure [36].

Calculation of extinction coefficients at 254 nm and GRAVY indices

Spectral data for tryptophane, tyrosine, and phenylalanine were obtained from the webpage of the Oregon Medical Laser Center (OMLC). Coefficients were then calculated *via* the same algorithm used by the Expasy ProtParam Server [37]. Extinction coefficients at 280 nm and GRAVY indices were obtained directly from the ProtParam web tool [29].

Purification of His-tagged HAdV-C05 E4ORF1p

The standard lab procedures used for expression tests are reported elsewhere [31,36]. *E.coli* BL21 DE3 proved to be the most suitable expression strain and was used for large-scale expression. The purification of His-tagged E4ORF1 is reported in [38], and will be outlined briefly: For expression, 6x 1.5 L LB medium containing 35 µg/mL kanamycin in 5L flasks were inoculated with an overnight culture grown at 37°C in LB (35 µg/mL kanamycin) at a dilution of 1:50. Cells were grown to an OD_{600} of 1.2-1.4 at 37°C, and protein expression was induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG). After 4-5 h of expression, cells were harvested and stored at -20°C over night. The pellets were resuspended in lysis buffer containing 10 mM imidazole, 1mM PMSF, 50 mM LiCl, and 100 mM triethanolamine (pH 8.5) at a ratio of 10 mL buffer per g of cells. Cell lysis by homogenizing the cells at least five times on an Avestin EmulsiFlex C3 instrument at a pressure of 700-800 bar. Cell debris was pelleted, and the supernatant was filtered, separated into four equal parts, and loaded three times onto four 1 mL Nickel affinity columns at 2 mL/min (GE healthcare) using a peristaltic pump. The columns were then mounted onto an ÄKTA prime and washed with lysis buffer until UV_{280} reached baseline levels. A first step with 25 % elution buffer (200 mM imidazole, 50 mM LiCl, and 100 mM triethanolamine (pH 8.5)) was applied in order to elute unspecifically bound protein from the column. E4ORF1p was eluted in a 60 mL gradient running to 100 % elution buffer. Elution factions were analyzed by SDS-PAGE and the concentration was estimated

accordingly. Fractions containing E4ORF1p were pooled, mixed with TEV protease (1/10 of the estimated protein mass), and dialyzed against at least 40x the pooled volume in buffer containing 50 mM LiCl, 100 mM triethanolamine (pH 8.5) using a 6 kDa dialysis tube (SpectraPor). The dialyzed protein was pumped 3x over a 1 mL Nickel affinity column in the absence of imidazole at 1 mL/min in order to get rid of un- or partially digested protein as well as the protease itself. The column was washed with five column volumes of lysis buffer, and flow through and wash were pooled and concentrated to a volume of 2.5-3 mL (corresponding to a concentration of 3-4 mg/mL). The sample was quantitatively applied to a Superdex 200 column on an ÄKTA basic system using dialysis buffer as liquid phase. The protein content of the main peak was checked by SDS-PAGE, and peak fractions were pooled and stored at -20 °C or -80 °C until further use.

The purification of SUMO-tagged HAdV-C05 E4ORF1p is reported in detail elsewhere [38].

Protein crystallization and data collection

For initial crystallization, the protein was concentrated to 8-9 mg/mL, while lower concentrations (1-1.5 mg/mL) were used for later fine screen experiments. Crystal plates were set up using a TECAN freedom evo 150 liquid handling system and the JCSG+ Suite (Qiagen, Hilden). 300 nL of protein solution were mixed with 300 nL well solution, and a well size of 100 µL was used. For macroseeding, crystals from condition B9 were transferred into pre-equilibrated hanging drops of the same condition (drop size 2 µL, well size 500 µL) with a protein concentration of 6-8 mg/mL and allowed to grow for several days before repeating the procedure. Initial crystal plates were incubated at 20°C, while later fine screens were set up at 4°C. Crystals were harvested in a nylon loop (Hampton Research, USA) of appropriate size. Conditions A5 and F5 were flash frozen immediately in liquid nitrogen, while conditions B6, C7, F10, and H4 were transferred into a drop containing a standard cryoprotectant (100 mM LiCl, 100 mM Tris pH 8.5, 50% (v/v) glycerol) prior to freezing. For condition B9, a cryoprotectant containing 100 mM Bicine pH 9, 10% (v/v) MPD, and 40% (v/v) glycerol was used. The diffraction properties of the crystals were tested at the X06DA beamline (Swiss light source, Paul-Scherrer-Institut, Villigen, Switzerland). Crystals were illuminated for up to 1.5 s per 0.1° at full beam transmission.

Dynamic Light Scattering

The purified protein was concentrated to 2.2 mg/mL, centrifuged for 30 min at maximum speed in an Eppendorf centrifuge, and 25 µL of the sample were equilibrated at 25°C in a 3 mm cuvette before measuring its dynamic light scattering behavior in a Malvern Zetasizer Nano-ZS instrument. The experiment was repeated five times, and each experiment consisted of 15 runs.

LC-MS

The washed and dialyzed SUMO-E4ORF1p sample was diluted 1:50 in dialysis buffer and subjected to high pressure liquid chromatography using a water-methanol gradient on a C18 porous silica reversed phase column (Agilent 1100 HPLC instrument). Chromatography peaks were then subjected to negative-charge ESY-MS at a quadrupole MS system (3200 QTRAP, SCIEX, Darmstadt). The detection gate was calibrated with a known concentration of NLS.

Differential Scanning Fluorimetry

The pure protein was diluted to a concentration of 0.5 mg/mL and mixed with a final concentration of 5x SYPRO Orange (Invitrogen). 20 µL of protein solution were aliquoted in 96-well plate. The plate was sealed, centrifuged carefully, and equilibrated at rt for 5 min. The measurement was performed on a Roche LightCycler 480 II instrument using a self-made program. The sample was heated from 20 °C to 95°C in a continuous gradient with steps of 0.02 °C/s. The dye was excited at 498 nm, and the emission wavelength of 610 nm

served as a readout according to [39]. The readout frequency was 24 acquisitions per °C. The first derivative of the melting curve was plotted against the temperature, and the maximum of the resulting curve was detected manually. The melting curve of the HAdV-D36 fiber knob served as a positive control, a sample without protein as a negative control. The experiment was carried out as a duplicate.

Negative stain electron microscopy of SUMO-tagged E4ORF1p

The freshly prepared peak fraction from analytical SEC (**Figure 6B**) was diluted 1:4 in dialysis buffer. The samples were applied to piolofom- and carbon-coated 400 mesh grids and stained with 1% Uranyl Acetate (UA). The samples were analyzed with a JEOL 1400plus TEM-Microscope (JEOL, Japan) at 120 kV and micrographs were acquired with a Tietz TemCam F-416 CMOS camera using the Digital Micrograph software (Gatan, US) and a pixel size of 1.97 Å on the objective scale.

Acknowledgments & Funding

The authors are thankful to Benjamin Dannenmann and prof. Dr. Klaus Schultze-Osthoff for the use of the LightCycler 480 II instrument for DSF and practical advice on how to set up the program. Furthermore, the authors thank Judith Bauer and Prof. Dr. Harald Groß for the use of their LC-NMS system and practical advice on how to use it. We are also thankful to the Swiss Light Source (SLS) for beam time and beamline service. This work has been funded by the SFB 685 of the German Research Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Berk AJ (2013) Adenoviridae. In: Knipe DM, Howley, P. M., editor. Fields Virology. 6 ed. Philadelphia: LIPINCOTT WILLIAMS & WILKINS. pp. 1704-1731.
2. Javier R, Raska K, Jr., Shenk T (1992) Requirement for the adenovirus type 9 E4 region in production of mammary tumors. Science 257: 1267-1271.
3. Javier RT (1994) Adenovirus type 9 E4 open reading frame 1 encodes a transforming protein required for the production of mammary tumors in rats. J Virol 68: 3917-3924.
4. Rogers PM, Fusinski KA, Rathod MA, Loiler SA, Pasarica M, et al. (2008) Human adenovirus Ad-36 induces adipogenesis via its E4 orf-1 gene. Int J Obes (Lond) 32: 397-406.
5. Rathod M, Vangipuram SD, Krishnan B, Heydari AR, Holland TC, et al. (2007) Viral mRNA expression but not DNA replication is required for lipogenic effect of human adenovirus Ad-36 in preadipocytes. Int J Obes (Lond) 31: 78-86.
6. Weiss RS, Lee SS, Prasad BV, Javier RT (1997) Human adenovirus early region 4 open reading frame 1 genes encode growth-transforming proteins that may be distantly related to dUTP pyrophosphatase enzymes. J Virol 71: 1857-1870.
7. Miyake-Stoner SJ, O'Shea CC (2014) Metabolism goes viral. Cell Metab 19: 549-550.
8. Chung SH, Weiss RS, Frese KK, Prasad BV, Javier RT (2008) Functionally distinct monomers and trimers produced by a viral oncoprotein. Oncogene 27: 1412-1420.
9. Latorre IJ, Roh MH, Frese KK, Weiss RS, Margolis B, et al. (2005) Viral oncoprotein-induced mislocalization of select PDZ proteins disrupts tight junctions and causes polarity defects in epithelial cells. J Cell Sci 118: 4283-4293.
10. Weiss RS, Javier RT (1997) A carboxy-terminal region required by the adenovirus type 9 E4 ORF1 oncoprotein for transformation mediates direct binding to cellular polypeptides. J Virol 71: 7873-7880.
11. Frese KK, Lee SS, Thomas DL, Latorre IJ, Weiss RS, et al. (2003) Selective PDZ protein-dependent stimulation of phosphatidylinositol 3-kinase by the adenovirus E4-ORF1 oncoprotein. Oncogene 22: 710-721.
12. Glaunsinger BA, Lee SS, Thomas M, Banks L, Javier R (2000) Interactions of the PDZ-protein MAGI-1 with adenovirus E4-ORF1 and high-risk papillomavirus E6 oncoproteins. Oncogene 19: 5270-5280.

13. Lee SS, Glaunsinger B, Mantovani F, Banks L, Javier RT (2000) Multi-PDZ domain protein MUPP1 is a cellular target for both adenovirus E4-ORF1 and high-risk papillomavirus type 18 E6 oncoproteins. J Virol 74: 9680-9693.
14. Niessen CM (2007) Tight junctions/adherens junctions: basic structure and function. J Invest Dermatol 127: 2525-2532.
15. Massimi P, Gammoh N, Thomas M, Banks L (2004) HPV E6 specifically targets different cellular pools of its PDZ domain-containing tumour suppressor substrates for proteasome-mediated degradation. Oncogene 23: 8033-8039.
16. Excoffon KJ, Gansemer ND, Mobily ME, Karp PH, Parekh KR, et al. (2010) Isoform-specific regulation and localization of the coxsackie and adenovirus receptor in human airway epithelia. PLoS One 5: e9909.
17. Kong K, Kumar M, Taruishi M, Javier RT (2014) The human adenovirus E4-ORF1 protein subverts discs large 1 to mediate membrane recruitment and dysregulation of phosphatidylinositol 3-kinase. PLoS Pathog 10: e1004102.
18. Chung SH, Frese KK, Weiss RS, Prasad BV, Javier RT (2007) A new crucial protein interaction element that targets the adenovirus E4-ORF1 oncoprotein to membrane vesicles. J Virol 81: 4787-4797.
19. Kong K, Kumar M, Taruishi M, Javier RT (2015) Adenovirus E4-ORF1 Dysregulates Epidermal Growth Factor and Insulin/Insulin-Like Growth Factor Receptors To Mediate Constitutive Myc Expression. J Virol 89: 10774-10785.
20. Na HN, Hegde V, Dubuisson O, Dhurandhar NV (2016) E4orf1 Enhances Glucose Uptake Independent of Proximal Insulin Signaling. PLoS One 11: e0161275.
21. Thai M, Graham NA, Braas D, Nehil M, Komisopoulou E, et al. (2014) Adenovirus E4ORF1-induced MYC activation promotes host cell anabolic glucose metabolism and virus replication. Cell Metab 19: 694-701.
22. Hegde V, Na HN, Dubuisson O, Burke SJ, Collier JJ, et al. (2016) An adenovirus-derived protein: A novel candidate for anti-diabetic drug development. Biochimie 121: 140-150.
23. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14: 1188-1190.
24. Robinson CM, Shariati F, Gillaspy AF, Dyer DW, Chodosh J (2008) Genomic and bioinformatics analysis of human adenovirus type 37: new insights into corneal tropism. BMC Genomics 9: 213.
25. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, et al. (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 42: W252-258.
26. Yang J, Yan R, Roy A, Xu D, Poisson J, et al. (2015) The I-TASSER Suite: protein structure and function prediction. Nat Methods 12: 7-8.
27. Kallberg M, Wang H, Wang S, Peng J, Wang Z, et al. (2012) Template-based protein structure modeling using the RaptorX web server. Nat Protoc 7: 1511-1522.
28. Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33: W244-248.
29. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, et al. (1999) Protein identification and analysis tools in the ExPASy server. Methods Mol Biol 112: 531-552.
30. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol 372: 774-797.
31. Lott A (2015) Cloning, Expression and Purification of E4ORF1 gene constructs from human Adenoviruses [Bachelor Thesis]. Tübingen: University of Tübingen.
32. Rupp B (2010) Biomolecular crystallography : principles, practice, and application to structural biology. New York: Garland Science. xxi, 809 p. p.
33. Doutch J, Hough MA, Hasnain SS, Strange RW (2012) Challenges of sulfur SAD phasing as a routine method in macromolecular crystallography. J Synchrotron Radiat 19: 19-29.
34. Human Adenovirus Working Group (2016) Serotyping Tool.
35. Lott A (2015) Working towards the structural characterization of the E4orf1 gene product of human Adenovirus HAdV 52 - cloning, expression, first purification -. Tübingen: University of Tübingen.
36. Hehl L (2016) Working towards the Characterization of human Adenovirus protein E4-ORF1. Tübingen: University of Tübingen.
37. Gill SC, von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem 182: 319-326.
38. Hehl L (2016) Characterization of the Human Adenovirus protein E4-ORF1 [Bachelor Thesis]. Tübingen: University of Tübingen.
39. Niesen FH, Berglund H, Vedadi M (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat Protoc 2: 2212-2221.

Author contributions

- All constructs and experiments were planned by AML.
- The cloning and purification of all E4ORF1p constructs was performed by AL, LH, and AML.
- Crystallization experiments and the biophysical characterization of HAdV-C05 E4ORF1p were performed by LH and AML.
- *In silico* analysis of E4ORF1p sequences and homology modelling were performed by AML.
- Negative-stain TEM experiments were planned by AML and performed by LH and YS.

Copyright Statement

Some figures in this manuscript were taken from original presentations. All original publications are referenced in the respective figure subtext and are listed here. If the journals are not licensed under the freely distributable Creative Commons Attribution (CC BY) license, or do not provide free access to their content for dissertations, licenses were obtained from the s100.copyright.com website of RightsLink™.

- **Figure 1, blue box:** **Niessen, 2007, J. Invest. Dermatol.**

Supplier	Elsevier Limited The Boulevard, Langford Lane Kidlington, Oxford, OX5 1GB, IK
Customer Name	Antonio Manuel Liaci
License Number	**3994910058295**
License date	Nov 23, 2016
Licensed Content Publisher	Elsevier
Licensed Content Publication	Journal of Investigative Dermatology
Licensed Content Title	Tight Junctions/Adherens Junctions: Basic Structure and Function
Licensed Content Author	Carien M. Niessen
Licensed Content Date	November 2007
Original figure numbers	Figure 2

- **Figure 1, yellow box:** **Thai et al., 2014, Cell Metab.**

Supplier	Elsevier Limited The Boulevard, Langford Lane Kidlington, Oxford, OX5 1GB, IK
Customer Name	Antonio Manuel Liaci
License Number	**3994910390568**
License date	Nov 23, 2016
Licensed Content Publisher	Elsevier
Licensed Content Publication	Cell Metabolism
Licensed Content Title	Adenovirus E4ORF1-Induced MYC Activation Promotes Host Cell Anabolic Glucose Metabolism and Virus Replication
Licensed Content Author	Minh Thai,Nicholas A. Graham,Daniel Braas,Michael Nehil,Evangelia Komisopoulou,Siavash K. Kurdistani,Frank McCormick,Thomas G. Graeber,Heather R. Christofk
Licensed Content Date	1 April 2016
Original figure numbers	Figure 4I

- **Figure 1, green box:** **Kong *et al.*, 2014, PLoS Pathogens**

Published by the Public Library of Science. PLoS articles are licensed under the Creative Commons Attribution (CC BY) license.