Structural Studies of N-Acetylmuramic Acid Kinases from the Bacterium *Tannerella forsythia* and of the Human Adenovirus Penton Base-Integrin Entry Receptor Complex

Dissertation

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

vorgelegt von
Aleksandra Stasiak
aus Warschau, Polen

Tübingen
2023

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

Tag der mündlichen Qualifikation:	27.10.2023
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter/-in:	Prof. Dr. Thilo Stehle
2. Berichterstatter/-in:	Prof. Dr. Dirk Schwarzer

“The time has come,” the Walrus said,
“To talk of many things:
Of shoes — and ships — and sealing-wax —
Of cabbages — and kings —”
Lewis Carroll, *Through the Looking Glass*

— Of penton bases, integrins
In ratios varying bound
Of sugars for bacterial growth
In enzyme crystals found

Table of Contents

Selected Abbreviations

AAV	Adeno-associated virus
AdV	Adenovirus
AMP-PCP	5′-adenylyl methylenediphosphonate
AMP-PNP	Adenylyl-imidodiphosphate
BSA	Buried surface area
C_5	Fivefold rotational symmetry
CD	Circular dichroism
CryoEM	Cryoelectron microscopy
CTF	Contrast transfer function
EM	Electron microscopy
FMDV	Foot-and-mouth disease virus
FSC	Fourier shell correlation
GlcNAc	N-acetylglucosamine
IMAC	Immobilised metal ion affinity chromatography
HAdV(-C5)	Human adenovirus (species C, type 5)
hpi	Hours post infection
HVR loop	Highly variable loop
kB	Kilobase pairs
MOI	Multiplicity of infection
MS	Mass spectrometry
MurK	Murein kinase
MurNAc	N-acetylmuramic acid
pb	Penton base
pb mut	Penton base dodecahedra non-forming mutant
pb wt	Penton base wild type
PDB	Protein Data Bank
PGN	Peptidoglycan
RGD motif	Arginine-glycine-aspartic acid motif
SAA	Surface-accessible area
SEC	Size exclusion chromatography
T	Triangulation number

Standard naming conventions are followed for amino acids and common biochemical molecules. Other abbreviations are defined in the text when they first appear.

Abstract

Structural biology is a branch of science dedicated to the determination and analysis of the shapes of biological molecules. This work describes two projects aiming to determine the structures of complexes formed by proteins from pathogens.

Project 1 – MurNAc kinases of *Tannerella forsythia*

T. forsythia is a Gram-negative bacterium associated with late-stage periodontitis, unable to synthesise the sugar N-acetylmuramic acid (MurNAc), required for the synthesis of the peptidoglycan (PGN) layer. The sugar kinase murein kinase (MurK) has been previously shown to 6-phosphorylate MurNAc after import into the cytoplasm. Here, the crystal structures of MurK and its paralogue K1058, in complex with their MurNAc substrate, are presented, leading to the identification of the basis of MurNAc specificity in sugar kinases. This is mediated by an absence of bulky residues introducing steric hindrance against the lactyl ether group. Results of kinetic assays of the kinases agreed with previously published MurK apparent kinetic parameters (K_M 180 µM, V_{max} 80 µmol min^{-1} mg^{-1}) and showed that K1058 had a higher binding affinity for MurNAc than MurK (apparent K_M 30 µM), but much lower efficiency (apparent V_{max} 0.34 µmol min^{-1} mg^{-1}). Hence, the natural substrate of K1058 remains to be determined. While the potential of MurK as a druggable target for periodontitis is uncertain, these first structures of MurNAc kinases in complex with their substrate provide new insights into their specificity, which could be applied to drug design or research on glycan processing in other organisms.

Project 2 – Human adenovirus and its integrin receptor

Human adenoviruses (HAdV) are double-stranded DNA viruses, responsible for a wide range of symptomatic diseases such as keratoconjunctivitis, respiratory, and gastrointestinal illnesses. Their broad tropism and large carrying capacity also make them attractive gene delivery vectors for vaccine or gene therapy applications.

To initiate infection, HAdV bind to an attachment receptor on the host cell surface, then enter the cell by engaging a transmembrane integrin protein. This work aimed to structurally characterise the complex between the HAdV penton base (pb) capsid protein from HAdV-C5, HAdV-D9, and HAdV-B35, and the $\alpha_v\beta_3$ integrin entry receptor. The purification of both complex components from insect cells was optimised successfully, but preparations of pb formed higher-order oligomers and had to be maintained at a low concentration. Complex formation was confirmed with negative

stain electron microscopy (EM) and cryoelectron microscopy (CryoEM). CryoEM 3D reconstructions of the HAdV-C5 penton base in complex with $\alpha_v\beta_3$ included a pentameric pb structure at resolution below 5 Å and pb at lower resolution (below 10 Å) with separate, featureless electron density, in some cases linked to the pb by an elongated bridge. The density could not be identified as the integrin with certainty due to the lack of detail. Initial attempts to purify the complex and remove the unbound pentamers that weaken the signal, failed. Future work will include purification of optimised pb constructs and structural characterisation of the complex.

Zusammenfassung

Strukturbiologie ist ein Teilgebiet der Wissenschaft, das sich mit der Aufklärung und Analyse der Struktur biologischer Moleküle beschäftigt. Diese Arbeit beschreibt zwei Projekte, deren Ziel die Aufklärung der Strukturen der Proteine aus Pathogenen, im Komplex mit anderen Molekülen, war.

Projekt 1 – Die Zucker-Kinasen der Tannerella forsythia

T. forsythia ist ein mit fortgeschrittener Parodontitis verbundenes gramnegatives Bakterium, das den Zucker N-Acetylmuraminsäure (MurNAc), der für die Synthese des Peptidoglykan notwendig ist, nicht herstellen kann. Es wurde zuvor gezeigt, dass die Zucker-Kinase Murein Kinase (MurK), MurNAc nach ihrem Import ins Zytoplasma an C6 phosphoryliert. In dieser Arbeit werden Kristallstrukturen von der mit MurNAc komplexierten MurK und ihr Paralog K1058 vorgestellt, was die Identifizierung des Prinzips der MurNAc-Spezifität in Zucker-Kinasen ermöglicht. Diese Spezifität beruht auf der Abwesenheit der größeren Aminosäureseitenketten, die eine sterische Hinderung gegen die Lactylether Gruppe verursachen würde. Kinetische Messungen der Kinasen haben mit früher veröffentlichten kinetischen Parameter übereingestimmt (K_M 180 µM, V_{max} 80 µmol min^{-1} mg^{-1}) und gezeigt, dass K1058 eine höhere Bindungsaffinität (K_M 30 µM) als MurK für MurNAc hat, aber auch eine viel niedrigere Effizienz (V_{max} 0.34 µmol min^{-1} mg^{-1}), deshalb muss das natürliche Substrat von K1058 noch bestimmt werden. Obwohl das Potential von MurK als Zielmolekül für Arzneimittel gegen Parodontitis unsicher ist, schaffen diese ersten Strukturen der MurNAc-Kinasen komplexiert mit ihrem Substrat neue Einblicke in ihre Spezifität, was bei der Medikamentenentwicklung oder Forschung über Glykanverarbeitung bei anderen Organismen Anwendungen haben könnte.

Projekt 2 – Human Adenovirus und sein Integrin Rezeptor

Humane Adenoviren (HAdV) sind eine Gruppe von Viren mit einer doppelsträngigen DNA, die für ein Vielzahl von Krankheiten wie Keratoconjunctivitis, Atemwegentzündungen oder Gastroenteritis, verantwortlich sind. Mit ihrem breiten Tropismus und großer Verpackungskapazität sind sie auch gut geeignet, um als Vektoren für Impfstoffe oder Gentherapie verwendet zu werden.

Um eine Infektion auszulösen, binden HAdV zuerst an einen Anheftungsrezeptor auf der Wirtzelloberfläche, dann dringen sie in die Zelle durch Bindung an das Transmembranprotein Integrin ein. Dieses Projekt hatte als Ziel, die strukturelle

Charakterisierung des Komplexes zwischen dem HAdV-Kapsidprotein Penton Base (pb) aus HAdV-C5, HAdV-D9, und HAdV-B35, und dem Infektionsrezeptor Integrin $\alpha_v\beta_3$. Die Reinigung der beiden Komplexbestandteile aus Insektenzellen wurde erfolgreich optimiert, pb bildete aber Oligomere höherer Ordnung und konnte deswegen nicht konzentriert werden. Die Komplexbildung wurde durch Negativkontrastierung-Elektronenmikroskopie und Kryoelektronenmikroskopie (CryoEM) bestätigt. CryoEM 3D-Modelle des HAdV-C5 Penton Base Protein komplexiert mit $\alpha_v\beta_3$ zeigten eine pb Pentamer Struktur mit einer Auflösung unter 5 Å und ein pb Monomer mit niedrigerer Auflösung (unter 10 Å) und eine weitere getrennte, merkmalslose Elektronendichte mit niedrigerer Auflösung, die in einigen Fällen mit dem pb Pentamer durch eine längliche Brücke verbunden ist. Die Dichte konnte wegen des Fehlens der Unterscheidungsmerkmale nicht als Integrin identifiziert werden. Erstversuche, den Komplex zu reinigen und dabei die ungebundene Pentamere, die das Signal abschwächen, abzutrennen, sind gescheitert. Zukünftige Arbeit wird die Reinigung der optimierten pb Konstrukten und die strukturelle Charakterisierung des Komplexes umfassen.

Authorship Statement – Part 1 (*T. forsythia* MurNAc kinases)

All experiments presented in the "Results – Part 1" section and not listed below have been performed by the author. Dr Georg Zocher operated the crystallisation robots, fished the diffracting MurK/MurNAc crystal, performed the initial processing steps of the structure solution of K1058/MurNAc and MurK/MurNAc to obtain initial density and solve the twinning issue, calculated the apparent kinetic values from the enzymatic assay data to produce **Figure 19**, and identified the mechanism of MurNAc specificity. Karolin Gogler optimised the purification of MurK and K1058, crystallised them, and, together with Phillipp Fink, solved and refined the MurK apo, K1058 apo, and MurK/AMP-PCP structures under Dr Zocher's supervision. Niklas J. Bayer fished the diffracting K1058/MurNAc crystals. Jasmin Freytag performed the CD spectroscopy measurements with the students of the Strukturaufklärung von Biomolekülen module. Dr Elena Ostertag and Katja Vonmetz also helped with the crystallisation procedures.

This work is included in the research article:
Stasiak AC, Gogler K, Borisova M, Fink P, Mayer C, Stehle T, Zocher G. N-acetylmuramic acid recognition by MurK kinase from the MurNAc auxotrophic oral pathogen *Tannerella forsythia*. J Biol Chem. 2023 Sep;299(9):105076.

Where the author has produced the figures, with exception of Figure 6 by Dr Georg Zocher, obtained the crystals of and built the MurK/MurNAc and K1058/ MurNAc structures, deposited the data in the Protein Data Bank (PDB), and performed the kinetic assays and some of the protein purifications, as well as co-wrote the article, co-analysed and co-interpreted the results.

Accepted publication	Yes
List of authors	Mayer C, Stehle T, Zocher G.
Position of candidate in list of authors	First
Scientific ideas by the candidate (%)	30%
Data generation by the candidate (%)	60%
Analysis and Interpretation by the candidate (%)	50%
Paper writing done by the candidate (%)	45%

Authorship Statement – Part 2 (Human adenovirus penton base and its integrin receptor)

All experiments presented in Results – Part 2 (Human adenovirus) have been performed by the author, except as listed here. The penton base mutant DNA constructs have been designed by Dr A. Manuel Liaci and the integrin construct has been designed, cloned, and used to generate baculovirus stocks by Dr Karolina Cuppelli. The penton base HAdV-C5 wild type construct was a kind gift of Professor Niklas Arnberg, University of Umeå. Negative stain electron microscopy was performed by Rebecca Stahl, ZMBP, Tübingen and Dr Katharina Hipp, MPI, Tübingen. CryoEM experiments were performed at the University of Helsinki CryoEM unit: samples were vitrified by Dr Pasi Laurinmäki, Dr Benham Lak, and Benita Löflund, and data were collected by Dr Pasi Laurinmäki and Dr Aušra Domanska, in some cases including on-the-fly motion and CTF corrections. **Figure 37a** has been provided by Dr Aušra Domanska. Assistance with the interpretation of data and their processing was provided by Professor Sarah Butcher, Dr Aušra Domanska, and Dr Justin Flatt, of the Structural Virology lab at the University of Helsinki.

1. Introduction

Part 1 – Sugar kinases of *Tannerella forsythia*

1.1. Bacteria as pathogens

Bacteria are unicellular organisms that constitute a domain of life separate from eukaryotes and archaea (1,2). They are characterised by lack of a nucleus and the presence of a cell wall. Like viruses, they are ubiquitous with regards to their habitats, but unlike them, they are able to synthesise protein, have a metabolism and replicate by their own cellular machinery (3). Bacteria are divided into Gram-positive and Gram-negative species, a distinction based on absorption of a positively charged dye that reflects the structure of their cell wall (reviewed in 4).

Despite what may seem to be relative simplicity compared to multicellular or eukaryotic organisms, bacteria can cause pathology in other living organisms, and this is a major reason for research interest in them. Diseases such as cholera and tuberculosis (TB) remain major causes of death in the developing world, and food-borne or opportunistic infections pose a danger, especially in hospital settings, in the developed world, too. The re-emergence of TB is also starting to be a concern in high-income countries, particularly among deprived populations, such as the homeless (reviewed in 5). Overall, TB is estimated to have caused 1.5 million deaths in 2020, with 10 million new infections and many of the treatment goals off track, largely due to the coronavirus disease 2019 (COVID-19) pandemic (6). It is difficult to estimate the exact bacterial and viral contributions to two major causes of death: respiratory and diarrhoeal diseases. However, at least 64% of lower respiratory infection deaths were modelled to be due to pneumococcal or *Haemophilus influenzae* infections (7), and the contribution of bacteria such as *Shigella*, *Campylobacter*, *Salmonella* and cholera to the diarrhoeal mortality was also considerable (8).

1.1.1. Antibiotic resistance and search for new targets

The over- and misuse of antibiotics over the past decades has predictably caused a steady increase in infections with bacteria resistant to multiple, or even all, known antibiotic classes, such as methicillin-resistant *Streptococcus aureus* (MRSA) (reviewed in 9), TB (6) or pan-drug resistant strains (reviewed in 10). Over a million deaths worldwide have been ascribed to antimicrobial resistance in 2019 (11). Though studies have shown that low-cost, low-tech solutions can be a way to prevent many issues (e.g., strict handwashing enforcement in hospital staff helped decrease the MRSA infections in British healthcare settings, 12), they cannot act as cures for already

infected patients. Strict usage control and relatively short lifetime of new antibiotics to prevent resistance reduces the commercial interest in their development.

In the search for new treatments, not only new compounds and targets, but also secondary targets that will prevent or revert resistance, can be explored. As an example, resistance to β-lactam antibiotics can be combated and overcome by synergistic combinations of the β-lactam with a compound inhibiting the enzymes involved in induction of resistance (reviewed in 13). The mechanism of action of both β-lactams and resistance-inducing pathways centres on the processes of bacterial cell wall synthesis and recycling, making them an important object of study. This is a significant challenge because the particulars differ between bacterial species, requiring specific rather than generalist solutions and the elucidation of details (reviewed in 14).

1.1.2. Bacterial cell wall-recycling cycle

The bacterial cell wall is a three-dimensional molecular structure forming a barrier between the bacterial cytoplasm and the outside and maintaining cell shape and osmotic pressure (reviewed in 15). It is composed of peptidoglycan (PGN), a peptide-sugar polymer composed of lipid II, a polyglycan stem of alternating N-acetylmuramic acid (MurNAc) and N-acetylglucosamine (GlcNAc) hexose units, cross-linked with pentapeptide stems and attached to undecaprenyl phosphate. PGN production and recycling are crucial to maintaining the homeostasis of cell wall structure, which protects the organism from the outside environment and has to be continuously renewed, or expanded on cell division (14). In Gram-negative bacteria, the PGN layer is sandwiched between two lipid bilayers, while in Gram-positive bacteria it is on the outside of a single lipid bilayer. These structures can be further modified by the addition of other molecules such as wall techoic acids or lipopolysaccharide, mediators of pathogenicity and immune avoidance for the pathogen (reviewed in 16). These modifications vary between different species and taxonomic categories, for instance, the exact composition of the pentapeptide stem differs between Gram-positive and negative bacteria, and glycans in the stem can undergo deacetylation or acetylation (reviewed in 17,18).

De novo synthesis of PGN begins with fructose-6-P in the cytoplasm, which is converted into GlcN-6P (glucosamine-6-P) and GlcN-1-P, with the first committed step of PGN synthesis being the generation of GlcNAc-1-P and UDP-GlcNAc by the bifunctional enzyme GlmU (GlcNAc-1-P uridyltransferase; see Figure 1). UDP-GlcNAc is then converted into MurNAc-containing compounds to which the peptide moieties

are attached. They in turn are attached to the lipid precursor to be flipped into the periplasmic space or the outside (reviewed in 14,19). Bacterial cell wall synthesis is a good target for antibiotics because this structure is absent from eukaryotic cells. MurNAc in particular is not present in eukaryotic cells, making its processing steps a selective target.

The integrity of the cell wall is crucial for bacterial survival and reproduction. The structure is in a constant state of flux, with up to 50% of muropeptides recycled in a single generation period in some species. Multiple enzymes participate in this process: lytic transglycosylases cleave the GlcNAc-MurNAc chain into disaccharide units of GlcNAc-anhMurNAc with the pentapeptide still attached, which are then transported back across the membrane. There, the enzyme NagZ splits the disaccharide stem in two, yielding GlcNAc and anhMurNAc peptide. Amidases then release the peptide chain, and the products are shuttled back into the PGN synthesis pathway, where phosphorylation and other enzymatic steps lead to the synthesis of UDP-MurNAc, a crucial precursor of PGN (reviewed in 14,20). An additional level of complexity is the synthesis and recycling of PGN by a network of enzymatic pathways, in a series of steps that can vary between species. For instance, certain MurNAc-processing enzymes bypassing the UDP-GlucNAc synthesis stage are present in *Pseudomonas putida* and absent in other Gram-negative bacteria, providing the organism with a way of maintaining cell wall synthesis even when some steps are blocked by antibiotics (21).

The cell wall synthesis and recycling processes are vital not only for structural reasons, but also because these pathways serve as signals and blocking them with antibiotics will lead to resistance induced by accumulation of unusual by-products. For instance, in many Gram-negative bacteria β-lactam resistance is mediated by the protein AmpC, the expression of which has been proposed to be induced by the presence of atypical anhydropentapeptides in the cytoplasm. These molecules are a result of the inactivation of one of the penicillin-binding proteins (PBP) by the antibiotic and signal issues with cell-wall recycling (14). The exact identity of metabolite pools and cycles and their relationships are not yet clear in many cases, which is why understanding them will be an important step towards combating antibiotic resistance and discovering synergistically acting compounds. For instance, resistance to Fosfomycin decreases in culture on blocking cell wall synthesis (22), suggesting that a combined application

of Fosfomycin with a cell wall synthesis inhibitor could be a therapeutic approach to overcome resistance.

1.2. *Tannerella forsythia* and its importance

Tannerella forsythia is an anaerobic Gram-negative bacterium of the *Bacteroidetes* phylum and *Tannerellaceae* family, originally isolated from patients suffering from different stages of periodontitis and other oral conditions such as gingivitis and endodontic infections (reviewed in 23,24). While its contribution to periodontitis is the main reason behind research interest in it, it has also been found in other disease states, including cardiac cadaver tissues (25), suggesting potential links between the bacterium and cardiac or neurological conditions (26). Its presence in calcified dental plaque (calculus) allows studies of its lineages and evolution in archaeological samples (27,28), from as long ago as pre-Columbian America (29).

Periodontitis is a disease of gums and adjoining tissue, where an inflammatory state resulting from imbalance of the oral microbiome leads to decay of underlying bones and, consequently, loss of teeth (reviewed in 30). The burden of severe periodontitis has been estimated to 11% of the global population, making it the sixth most prevalent condition affecting about 743 million people in 2010 (31). Additionally, some reports indicate connections of periodontitis with diseases such as diabetes, oral cancer (32) and Alzheimer's (26), though the causal relationships are still largely unclear.

Traditional treatment approaches include root scaling and planing, though other interventions, like antibiotics and immune modulators, have been applied with some success. The mechanisms of the disease are complex, with the autoinflammatory component driving the pathogenesis, but presence of particular microbial species being a strong indicator of disease. A reduction in their population does not necessarily correlate with recovery, and this interplay makes the design of targeted, non-mechanical interventions a challenge (33).

Dysbiosis, or the displacement of healthy bacterial flora of the oral cavity with periodontitis-promoting bacteria is thought to be a crucial step for the progress of the disease (34). *Tannerella forsythia* forms a part of the so-called "red complex" with *Tannerella denticola* and *Porphyromonas gingivalis* (35,36). Somewhat counter-intuitively, the periodontic "red complex" is not named after the colour of the infected tissues, but rather is the final step in the rainbow-ordered progression of microbial assemblies over the course of the disease (i.e., it is preceded by the orange and yellow complexes, considered "bridging pathogens", present at earlier stages). Studies have

shown that rather than a single red complex pathogen being the causative agent, all three (and, potentially, more) contribute to disease course (reviewed in 37). A number of their virulence factors, such as protein-degrading enzymes, have been proposed to contribute to late-stage periodontitis. Additionally, they interact co-operatively: *P. gingivalis* is only found in the presence of *T. denticola*, possibly relying on it for nutrients. Importantly for this work, *T. forsythia* seems to be similarly dependent on the complex members for a supply of N-acetylmuramic acid (MurNAc) (38).

1.3. *T. forsythia* kinases MurK and K1058

A peculiarity of *T. forsythia* is the absence of the enzymes MurA (UDP-N-acetylglucosamine 1-carboxyvinyltransferase), MurB (UDP-N-acetylenolpyruvoyl-glucosamine reductase), and some of the Glm family enzymes, which initiate the synthesis of the MurNAc component of the PGN from GlcN and, earlier, the transformation of fructose-6-P to GlcN (39,40). This results in strict auxotrophy for the sugar, which is sourced from the environment of the bacterium, and as such MurNAc-processing steps are a viable target for antimicrobials (41). The exception to this is the observation that some isolates from cats and dogs do not have this dependence on MurNAc (42). This fits with the presence of GlmM, MurA and MurB-homologous sequences in sequencing results from canine specimens (43).

An alternative pathway must exist to provide MurNAc for cell-wall synthesis, scavenging metabolites of other bacterial species from the environment. The conventional MurP (phosphotransferase) system known from *E. coli* and other species has not been identified in the genome. Instead, the sugar is taken up by the transporter MurT (44) and phosphorylated by the ATP-dependent kinase MurK (Murein kinase). MurNAc-6-P can then be processed to GlcNAc-6-P by the etherase MurQ, and to GlcN-6-P by the enzyme NagZ (44). It remains unclear if and how these products feed into the PGN synthesis pathway in the absence of traditional enzymes, given that MurK knockout mutants seem to increase cell wall synthesis (41). Homologues of the AmgK and MurU proteins that convert MurNAc into MurNAc-1-P and UDP-MurNAc have been identified in the *T. forsythia* genome (45), which would provide a way to incorporate MurNAc into the PGN pathways (see Figure 1). Additionally, sialic acid scavenging can replace external MurNAc supply in biofilm forms of *T. forsythia*, with homologues of enzymes able to produce GlcNAc and GlcNAc-1-P, though it is unknown how these are processed to produce MurNAc (46).

The kinases MurK and K1058, which share 44% sequence similarity and are encoded by genes separated only by a single PGN-recycling gene cluster in the genome (47), are the object of structural investigation in this work. Both phosphorylate MurNAc, and some affinity for GlcNAc in K1058 was also reported. The biochemical properties of MurK have been described by Hottmann and colleagues (41) and in a doctoral thesis (47), also containing the first study of K1058. This protein has been less extensively characterised and its participation in the PGN synthesis pathways is unclear.

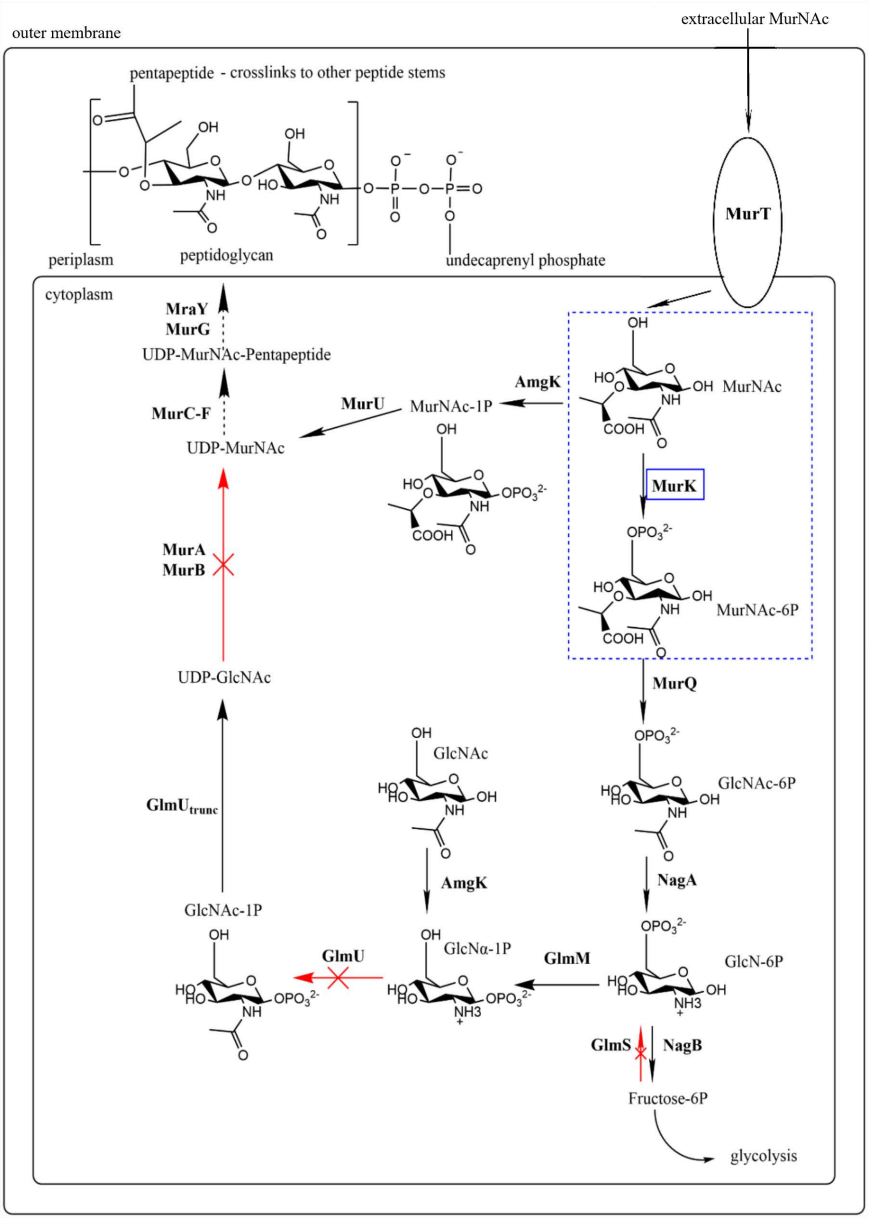


Figure 1. Selected reactions of peptidoglycan synthesis and recycling in *T. forsythia* cytoplasm. The reaction catalysed by MurK, the subject of the project, is highlighted by the **blue** box. **Red** arrows depict reactions catalysed by enzymes generally absent in *T. forsythia* (47).

1.4. Objectives and expected outcome - Part 1

The goal was to establish the atomic structures of the *T. forsythia* enzymes MurK and K1058 as ternary complexes, with their sugar substrates and ATP analogues. Based on this, the mechanism of specificity was to be identified. Previously established purification protocols were used to produce large quantities of high-purity protein for crystallisation and X-ray diffraction studies. Crystallisation and structure solution of the apo forms of both enzymes and of the complex of MurK with an ATP analogue had been performed earlier (48) by Karolin Gogler and Philip Fink. Since previous soaking experiments have been unsuccessful, suggesting a large conformational change causing crystal cracking, co-crystallisation approaches would be employed to determine high-resolution structures of the protein-sugar complexes. The active site would then be identified from the substrate binding location and comparison to structural homologues performed to establish the basis for difference in kinetic specificity of the enzymes.

Before this work, MurK has been shown to be involved in the MurNAc processing pathways, with knockouts accumulating intermediate products and having a growth advantage in MurNAc-abundant conditions, which suggested its involvement in catabolic pathways. It was also biochemically characterised to show that it had a strong preference for MurNAc, with K_M of 113 µM, over GlcNAc, with K_M of 116 mM, at short (45 minutes) incubation periods, as established by NADH-coupled assays (41). K1058 was only examined for substrate preferences and found to be able to phosphorylate both GlcNAc and MurNAc without a marked difference after a 16-hour incubation (47). This work aimed to replicate these results as well as establish the kinetic parametres for K1058 catalysis of MurNAc phoshorylation.

Part 2 – Human adenovirus penton base in complex with its integrin receptor

1.5. Human adenovirus

Viruses are infectious, parasitic molecular machines that require the structures of a cell of a host organism to replicate and are unable to synthesise proteins independently of it (49). They exhibit a striking diversity in size, type of nucleic acid, capsid structure, host species, replication strategies and symptoms elicited. While the prevailing view seems to be that in light of their crucial dependence on the host for proteins synthesis, they are not living organisms themselves, this is not a consensus (49–53). Regardless of this controversy, largely influenced by semantic definitions of life (54), which attempt to assign the products of evolution into categories it does not follow (55,56), viruses continue to be a focus of research for a number of reasons, as outlined below.

Human adenoviruses (HAdV) are non-enveloped icosahedral viruses of the genus *Mastadenovirus* in the family of *Adenoviridae,* with a DNA genome of about 36 to 38 kilobase pairs (kB) and a capsid of between 70 and 90 nanometres (nm) in diameter (49). Members of this family are known to infect many other animals such as fish and snakes, and within the genus *Mastadenovirus* itself the host diversity spans mammals from bats to primates and dolphins (reviewed in 57). Species belonging to the genus are distinguished based on genome organisation and phylogenetic distance (*Human Mastadenovirus A* to *G*), and these are further divided into types (e.g., the HAdV type C5 discussed in this work belongs to species C). Nearly 60 HAdV types have been isolated, and more than 100 genotypes have been identified from sequencing data. The name comes from the adenoid (gland) tissues from which the virus was originally isolated (58,59).

Originally, HAdV was divided into serotypes, differentiated by the absence of antibody cross-neutralisation. Currently, given the challenges of serotyping, the term “type” is used to denote isolates of significantly divergent sequences and phenotypes, though the exact criteria are still a subject of discussion (58,60,61). Tropism for cell types varies between adenovirus types and species (summarised in 62), which results in different disease phenotypes (see below) and is mediated partly by the diversity of attachment receptors of different types.

1.5.1. HAdV as a pathogen

In the process of viral replication, the machinery of the host cell is modified and disrupted, which often causes disturbance to its function, leading to pathology, i.e.,

disease symptoms, in the organism (49). Diseases of viral aetiology are a significant cause of morbidity and mortality globally (63–65), more so due to the recent COVID-19 pandemic caused by Severe Acute Respiratory Syndrome Coronavirus type 2 (SARS-CoV-2, 66). While this is the latest example of a virus-caused pandemic, it has been preceded by multiple others, such as Human Immunodeficiency Virus/Acquired Immunodeficiency Syndrome (HIV/AIDS), which came to public attention in the 1980s (reviewed in 67), or numerous influenza pandemics, notably that of 1918 (reviewed in 68). Of these, the former is estimated to have resulted in 36 million deaths (69), while the latter in about 50 million deaths (68).

Viral pathogenicity is not restricted to pandemic viruses. Prior to the emergence of SARS-CoV-2, the most lethal, and the one causing the loss of the most disability-adjusted life years (DALY), pathogen in the European Union had been estimated to be seasonal influenza (70). The developing world is disproportionally burdened by deaths caused by gastroenteric and respiratory virus infections, particularly among children under five (71). Non-lethal disease is also a significant issue, for instance the comparatively mild illnesses caused by common cold viruses were estimated to have caused economic losses of 40 billion dollars in the United States in 2000 due to healthcare costs and lost productivity (72). Additionally, the symptoms of disease caused even by viruses considered mild, such as enteroviruses, parechoviruses, or adenoviruses, may still be severe and life-threatening for the elderly, children or the immunocompromised (reviewed in 73,74), and the range of antivirals available is also limited.

A special concern for public health is the emergence of viruses, from a non-human host species into humans (Ebola virus, SARS-CoV-1 and 2), or the gain in importance due to environmental or societal changes (enterovirus D-68, Zika virus) (49). The COVID-19 pandemic exemplifies the threat posed by viral spillover from wild animals, in this case likely bats (75).

While those viral diseases that affect primarily humans tend to be the focus of attention, viruses infect all known taxonomical domains and kingdoms of life (49). Of particular economic and environmental concern are diseases of animals and plants: for example, the cost of the 2001 outbreak of the foot-and-mouth disease virus (FMDV) among cattle in Britain has been estimated to 3 billion euros (76). Recurring epidemics also threaten crop yields, for instance of sharka disease caused by the plum pox virus (reviewed in 77) or the papaya ringspot virus (reviewed in 78).

HAdV is of research interest due to its clinical importance. It can cause a range of symptoms and disease severities: for instance, types F40 and F41 are linked to gastrointestinal disease, types E4, C5, and B7 to respiratory infections, types D36 and D37 to epidemic keratoconjunctivitis (EKC), and others to urinary infections (summarised in 62). Although the symptoms are typically self-resolving and non-threatening, an infection can still pose a danger to those immunocompromised or immunosuppressed (reviewed in 79). Of practical concern are outbreaks in highly crowded conditions, such as the military, where deaths due to pneumonia-like disease occur sporadically in otherwise healthy people (80). As a result, a vaccine against HAdV types 4 and B-7 was developed for the exclusive use of the US military (81).

1.5.2. HAdV as a tool

In addition to having wide-ranging effects on health, viral infection of host cells can provide important insights into the mechanics of functioning cells, and it can disrupt them or be used to deliver vaccines or therapies.

Viruses have become a major tool of biotechnology through their application in gene delivery, whether in oncotherapy, vaccines or gene therapy. The principle of gene therapy: introduction of new, functional genes to an organism that is lacking them, can be realised by viruses, as their replication cycle involves introducing foreign genes to host cells (reviewed in 82). Removing pathogenic and replicative genes and substituting them with the desired sequences enables the production of a particle that can enter cells and deliver the contents to be employed for protein production.

Hundreds of clinical trials have utilised adenoviruses for vaccine, gene- or oncotherapy (83). Their advantages include a relatively large loading capacity of nucleic acids compared to smaller viruses such as adeno-associated virus (AAV), no evidence of integration into the genome, with the nucleic acid instead transcribed independently, large-scale production technologies, and a wide range of tropisms that could be targeted to specific tissues or organs (reviewed in 82,84).

However, though the clinical trials have been many, successes have been relatively few. The COVID-19 vaccines have recently brought viral vectors to prominence (85,86), with type HAdV-D26 in Johnson & Johnson (87), and types C5 and D26 in Sputnik V (88). Apart from these, a component of an Ebola vaccine regimen (89); a single bladder cancer therapy approved by the Food and Drug Administration in December 2022 (90); as well as two head and neck cancer therapies only approved in China (91,92); are the sole approved HAdV-based products. Key issues encountered

with HAdV gene delivery vectors, similarly to other viruses, include targeting, with the tendency to accumulate in the liver rather than target tissues (93), as well as prevalence of immunity to the vectors (reviewed in 94). Given that over 60% of surveyed adults have antibodies to HAdV-C5 (95), the vectors can be neutralised before reaching their targets, and their immunogenicity decreases the effectiveness of multiple usages, too (96). Immunogenicity was also the reason for the 1999 death of a patient in an adenoviral gene therapy clinical trial (97), which led to a decrease in interest and support for the use of this vector for some time (reviewed in 98). Similarly, recent reports of deaths due to cerebral thrombosis following the administration of adenovirus COVID-19 vaccines have raised concerns about safety (99).

Clinical trial dosing had been optimised, preventing death-inducing levels of pathogenicity (100), and the mechanism of blood clot formation has been proposed, enabling diagnosis and treatment (101,102). Use of less prevalent HAdV types, such as type D26, or B35 (103), chimeras with components from multiple types (104), or non-human AdVs (e.g., chimpanzee AdV in the COVID-19 vaccine, 105), have been ways to avoid patient immunity, while increased understanding of entry processes has improved targeting (106). The widespread and largely successful use of AdV-vectored vaccines has likely opened the door for their further applications. Further studies of HAdV can enhance the therapeutic applications, for instance through understanding of processes involved in the initiation of infection.

1.5.3. HAdV attachment and entry

In order to hijack the machinery of the host cells, the viral genetic material, possibly with other functional components, has to be delivered into the cytoplasm or the nucleus. There, the genome will be replicated, new virion proteins produced, to be followed by assembly and exit from the cell. Infection begins with engaging the outside of the host cell from the extracellular matrix or from the outside of the organism altogether. This attachment step is required for all those to follow (except in plant and fungal viruses, 49), involving a specific interaction between a virus capsid (glyco)protein and a receptor molecule present on the cell surface. Proteins are a common type of receptor molecule, being perhaps the predominant focus of research attention, but glycans and lipids, or compound molecules such as glycoproteins and glycolipids, also participate in virus binding (reviewed in 107).

After attachment to cell surface, entry into the cell is required for a productive replication cycle. This may occur through a cellular process for the transport of

macromolecules: for instance, the virus being packaged into a lipid vesicle and undergoing endocytosis via clathrin-dependent or independent mechanisms. In enveloped viruses, the outer lipid-based envelope layer can fuse with the cell membrane, enabling the nucleocapsid to be released into the cell (reviewed in 108).

These initial interactions with cellular receptors are mediated by viral glycoproteins protruding from the envelope in the case of enveloped viruses, or by capsid proteins in the case of non-enveloped viruses. In the latter case, these proteins may either form an integral part of the basic capsid shape (e.g., in polyomaviruses or picornaviruses, where the proteins forming the icosahedral capsid bind the receptors, 109,110) or protrude away from the centre of the capsid (e.g., fibre proteins of the adenovirus).

Attachment and entry do not necessarily utilise a single viral protein or single receptor: more than one viral or host protein may be required to complete these multi-step processes. For instance, a receptor and a co-receptor (or an attachment and an entry receptor) may be needed, as in the case of HIV or adenovirus, and these may be engaged by different viral proteins. Moreover, it may be possible for a single virus species or genus to enter cells via multiple receptor molecules: adenoviruses have a range of attachment receptors between their different types, one of the reasons for their broad tropism. Since the cell surface molecules serving as receptors have physiological functions and are conserved, many molecules are receptors for multiple viruses (reviewed in 107). Interplay between receptors and attachment proteins can be complex and dependent on the context, as exemplified by ACE2 (angiotensin-converting enzyme) and TMPRSS2 (transmembrane serine protease subtype 2) in SARS-CoV-2 (111).

Given that attachment and entry are steps required for productive viral infection, they are also potential key targets for control measures. Interfering with them is a mechanism of some antiviral drugs, such as the anti-HIV Enfuvirtide, which binds to a virus surface glycoprotein and prevents it from conformational change required for membrane fusion (112). Another anti-HIV compound, Maraviroc, binds to the cellular entry receptor, CCR5 (C-C chemokine receptor type 5), which prevents virus binding to the receptor, and blocks entry (113).

Understanding of attachment and entry mechanisms can also be used to improve the efficiency and targeting of viral gene delivery vectors. Designing vectors that use as receptors molecules prevalent on particular cell populations (e.g., cancer cells such as neuroblastomas overexpressing polysialic acids (114,115), cell types that are missing

or underexpressing the gene of interest in gene therapy (116) or immune cells for better vaccine efficacy (117) are all potential applications.

In the HAdV replication cycle (see Figure 2), initial attachment is mediated by the binding of the fibre knob protein protruding from the capsid (see below) to an attachment receptor. So far, a number of these, both protein and glycan, have been identified (reviewed in 118). Among these, some have also been structurally characterised in complex with fibres from different HAdV types (reviewed in 119): polysialic acid (114) and GD1a (120), among the sugars, and proteins coxsackie and adenovirus receptor (CAR, 121), cluster of differentiation protein 46 (CD46, 122), and desmoglein-2 (DSG-2, 123). This initial attachment step is followed by entry via a clathrin-mediated endocytosis pathway, enabled by the binding of the penton base capsid protein to proteins from the integrin family (see 1.5.3). Receptor-mediated signalling initiates the endocytosis process required for viral entry into the host cell (124).

Once HAdV enters the early endosome, a disruption to normal organelle function occurs, which causes the virus to be transported to the nuclear pore region via the microtubule network. The process of uncoating, initiated on binding the attachment receptor, ends when the genome is transported into the nucleus for transcription (reviewed in 125). Transcribed RNA is exported into the cytoplasm for translation in stages, after which the produced proteins enter the nucleus where they participate in transcription of further genes and regulate it, a cycle that repeats two more times (49). Structural proteins of the virus are finally imported into the nucleus to form immature capsids and are released from the host cell, generally as part of cell death processes.

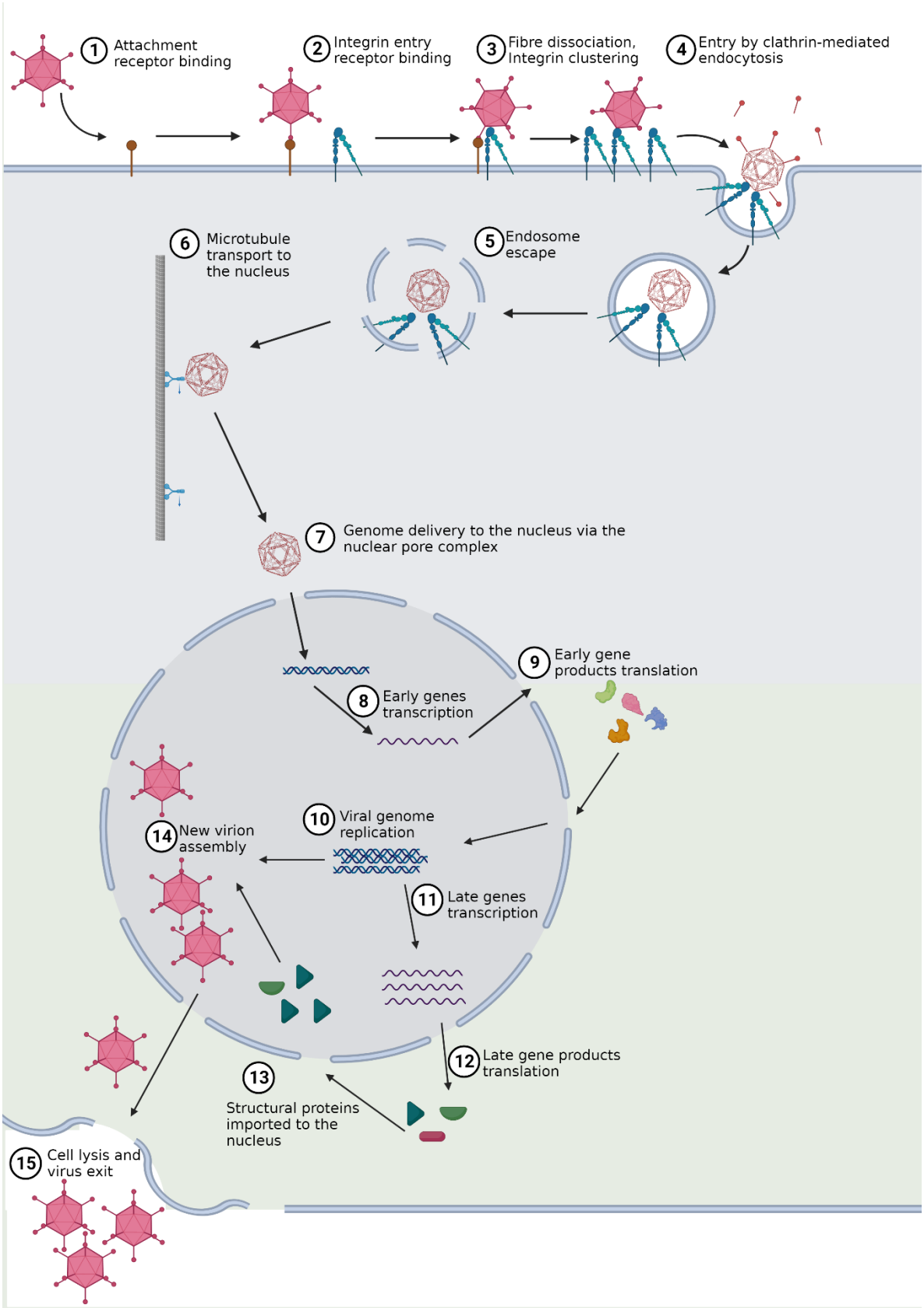


Figure 2. A schematic representation of the HAdV replication cycle. Engagement with one of the possible attachment receptors leads to the binding of the integrin entry receptor and endocytosis of the virion. Figure created with BioRender.com, redrawn and adapted from (126) and (127).

1.5.4. Structural studies of viruses with integrins

Integrins are a family of transmembrane proteins implicated in a wide range of processes, such as inside-out and outside-in signalling in transport, growth, immunity and cancer. They also serve as receptors for numerous viruses, such as FMDV, herpesviruses, and HAdV (reviewed in 128). Multiple integrins have been structurally characterised in closed and open conformations (129–131). They are composed of an α and a β subunit, in 24 different combinations of the known 10 α and 8 β subunits (reviewed in 132; see Figure 3). They are all characterised by large conformational shifts and flexibility, with a closed conformation associated with inactive signalling and a transition to extended-closed and extended-open states related to active signalling. On the outside of the cell, binding of ligands and metal ions, such as Mg^{2+}, Ca^{2+}, Mn^{2+}, to the globular head can cause changes of the elongated "legs" (133). The intracellular domains can also be separated by ligand binding. A number of integrin receptors have been identified for HAdVs: $\alpha_v\beta_3$ and $\alpha_v\beta_5$ (134), $\alpha_v\beta_1$ (135), $\alpha_M\beta_2$ (136). Arginine-glycine-aspartic-(RGD)-binding integrins are thought to interact with all HAdV types (137) based on the presence of the RGD motif in most HAdV penton bases as part of a flexible loop, except for types F40 and F41, which interact with α_6-containing integrins (138).

To date, there have been relatively few structural studies of viral interactions with their integrin receptors. Apart from the three adenovirus-integrin studies (see section 1.5.4), studies of FMDV (140), human parechovirus 1 (141), and coxsackievirus A9 (142) full capsids in complex with integrins have been published. All three concern the interactions of an entire virus capsid, and used cryoelectron microscopy (cryoEM). Of these, the latter two have been conducted before the "resolution revolution", yielding reconstructions at a resolution of 8-10 Å, which confirmed that the integrins ($\alpha_v\beta_3$ or $\alpha_v\beta_6$) bind to the capsid in the vicinity of an RGD motif. They also showed a range of conformations of the bound integrin, but given the resolution and issues with averaging and multiple occupancies, were unable to elucidate the details of molecular contacts.

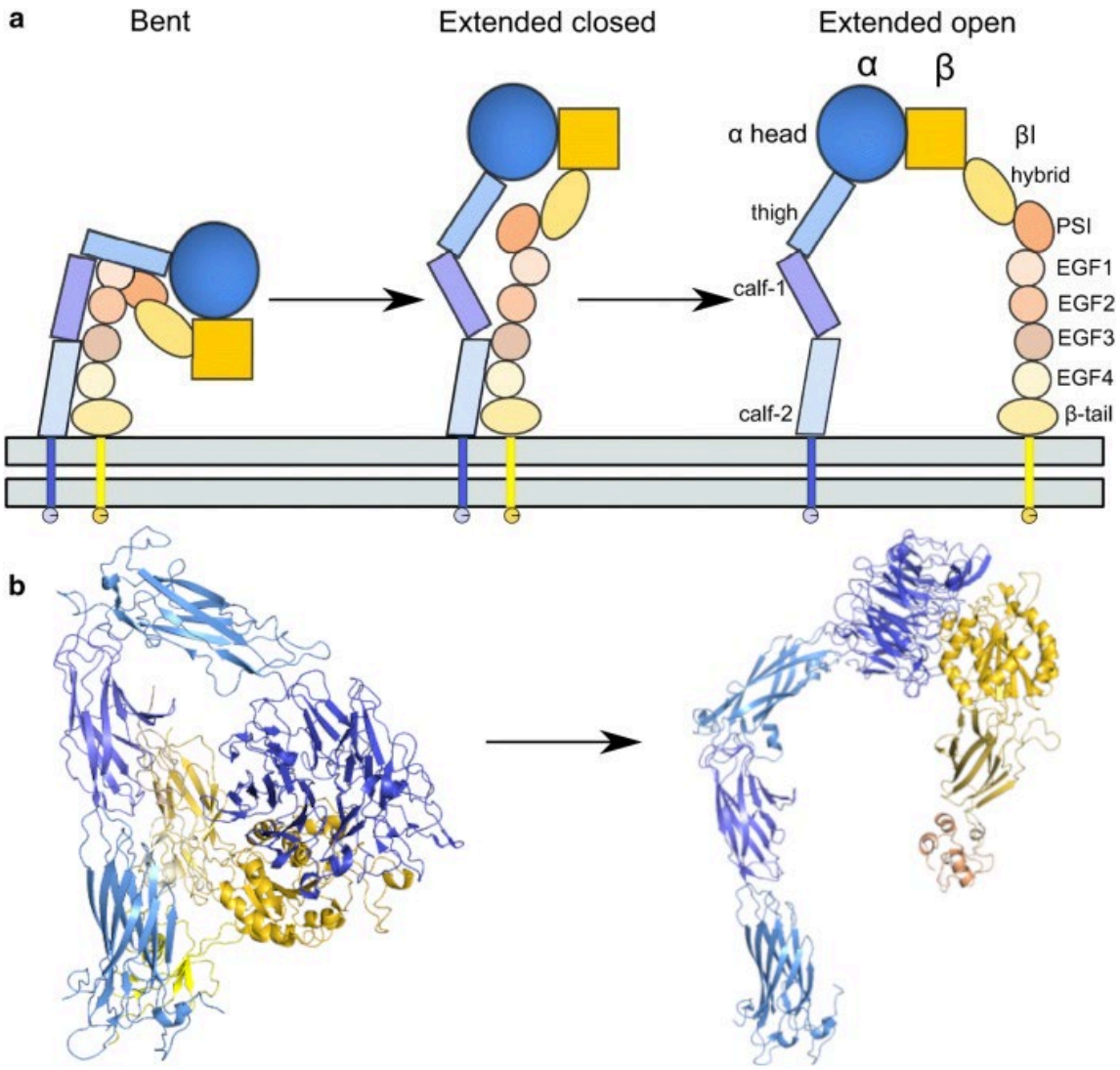


Figure 3. The integrin $\alpha_v\beta_3$ a. A diagram of the conformational changes in integrin $\alpha_v\beta_3$. Cell membrane in grey. Redrawn from (131) and previously published in (119), reproduced on a Creative Commons-BY license (https://creativecommons.org/licenses/by/4.0/). **b.** X-ray structures of two conformations of integrin $\alpha_v\beta_3$: bent (129) and extended open (130) stabilised by an antibody (not shown for clarity). Figures generated with PyMol (139). Domain colouring as in panel a. Not all domains were resolved due to flexibility.

The FMDV study resolves the FMDV-integrin $\alpha_v\beta_6$ binding by using localised reconstruction to refine the binding sites separately from the icosahedrally symmetrical capsid, comparing between capsids of two virus serotypes. These are resolved at a resolution between 3 and 4 Å, with the integrins resolved at between 8.5 and 12 Å. The number of integrins bound to the maximum 60 binding sites per capsid ranged between 5 to 55 integrins, with differences in distribution between the serotypes. The protein fragment mediating the binding was identified as the GH loop of the VP1 capsid protein, containing the RGD motif and further hydrophobic interactions, though binding via an N-linked sugar on the integrin has also been proposed. The integrin was primarily in an open conformation, demonstrating

typical flexibility between the different subclasses. Molecular details of the integrin's interaction could not be described in detail due to insufficient resolution.

1.5.5. HAdV capsid structure

Virus capsids are constructed of repetitive protein motifs, reducing the size of nucleic acid needed to encode a capsid large enough to contain the genetic information. As a result, many capsids exhibit highly symmetrical features, most commonly helical or icosahedral. An icosahedron (see Figure 4a) is a platonic solid composed of 20 triangular faces, which can be generated by applying 5-3-2 rotational symmetry to an icosahedral asymmetric unit, repeated sixty times (generally three times per face) in the solid. Viral architecture is also based on this property, relying on the principles of quasiequivalence, which allow an identical protein to form a range of similar, but not identical contacts, enabling participation in both pentameric assemblies at the vertices and hexameric subunits on the faces (143).

While the simplest icosahedral capsids consist of 60 copies of the same protein, others have many more proteins per asymmetric unit. The triangulation number *T* refers to both the number of capsomeres per triangular facet (not necessarily identical to an icosahedron face) and protein subunits within a viral asymmetric unit, which can be multiplied by 60 to yield the total number of major capsid proteins in the virus (143). The capsid of HAdV is icosahedral (pseudo *T*=25), composed of three repeating major capsid proteins and a number of minor capsid proteins known as cement proteins. The *T* number is referred to as pseudo 25, because the hexameric capsomers are not composed of six, but rather three, identical hexon proteins, which means that the number of proteins in the asymmetric unit is 13, and not the expected 25 (144).

The hexon is the most abundant protein in the HAdV capsid, 720 copies in total. At each of the 12 capsid vertices is a protein complex known as the penton (see Figure 4b). The penton consists of a pentameric penton base interacting with the hexon and trimeric fibre in the centre of the five penton base copies, with a knob protruding away from the centre of the capsid (see Figure 4c). The minor capsid proteins, among other functions, stabilise the interactions between the major capsid proteins, mostly from the inside, though protein IX forms a network on the outside (145,146).

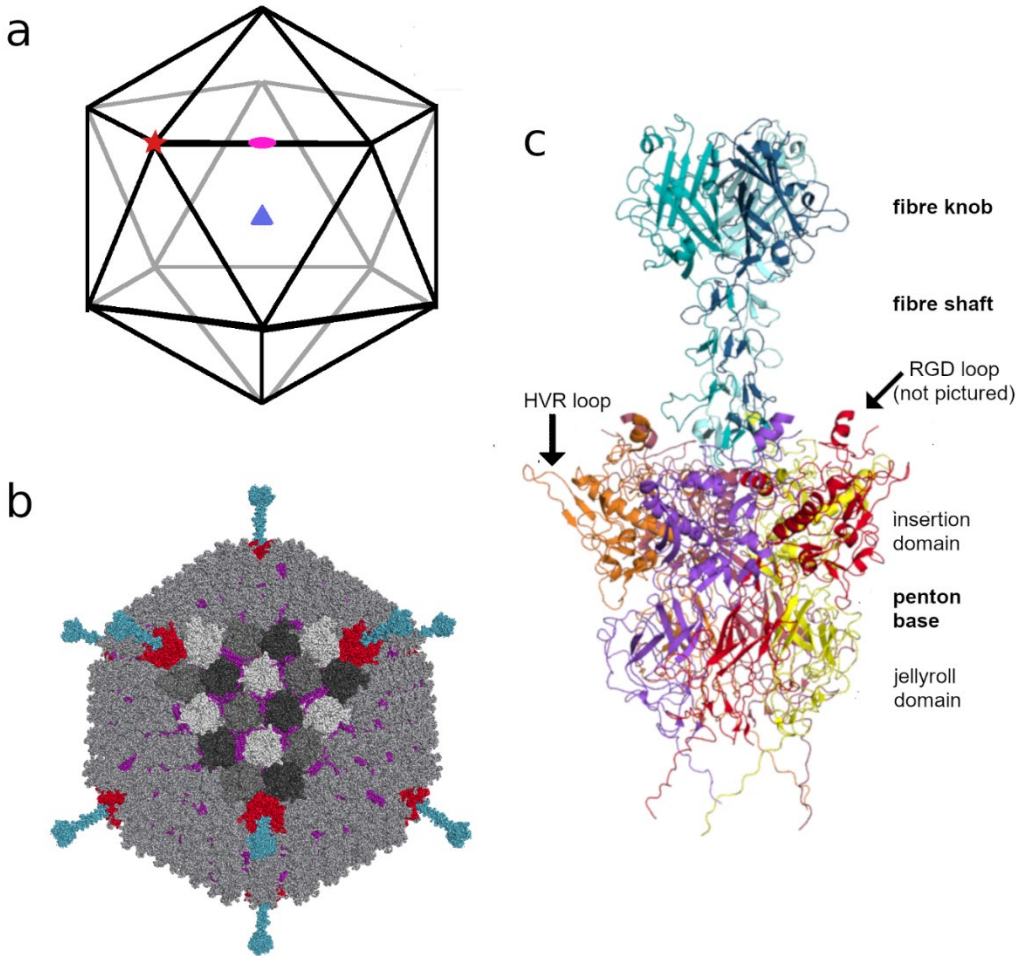


Figure 4. HAdV icosahedral capsid structure and the vertex penton. a. An icosahedron, with a fivefold symmetry axis at the vertex marked with a **red** star, a threefold symmetry axis in the centre of the face with a **blue** triangle, and a twofold symmetry axis at the edge with a **pink** ellipse. **b.** HAdV-C5 capsid (146), with the penton reconstructed from (147–149). Fibre shaft length based on HAdV-D26, structure from HAdV-C2. Hexons in **grey** (shades of grey on one of the icosahedral faces), penton bases in **red**, fibres in **blue**, and minor capsid proteins in **purple**. **c.** Human adenovirus penton, reconstruction composed of atomic structures HAdV-C5 penton base and HAdV-C2 fibre, fitted into HAdV-D26 penton reconstruction (147,149,150) Panels **(b,c)** generated with PyMol (139).

1.5.6. HAdV penton

The HAdV penton base is an elongated protein between 50 and 60 kiloDaltons (kDa) in size. It consists of a jellyroll and an insertion domain, the latter protruding away from the capsid centre (151). The insertion domain contains two loops: the highly variable (HVR) loop and the RGD loop, named for the arginine-glycine-aspartic acid motif it contains. Both of these loops differ strongly in sequence between different HAdV types, unlike the rest of the penton base, which typically exhibits 90% conservation (152). The RGD loop ranges from 30 to 100 amino acids in length in different types, exhibiting different degrees of flexibility. The RGD motif is an integrin-binding sequence,

mediating interaction with the entry receptor (153), with the exception of types HAdV-F40 and F41 (138,154). It is unclear what impact loop length has on tropism and receptor interactions in different HAdV types and comparing binding of pentons with different loop lengths to integrin receptors was one of the goals of this project.

In the virus capsid, the penton base is a pentamer located on the vertices. An attachment receptor-engaging trimeric fibre protein is bound in the middle of the pentamer, forming the penton (155,156,157, see Figure 4b). Diversity in these proteins is one of the mediators of differences between tropism of HAdV types. The five repeats of the penton base have significance for the entry receptor binding, as there are five RGD loops per vertex, providing up to five binding sites for the integrin (151). How many integrins actually bind is a research question this project has attempted to address.

A property of the penton base is the formation of dodecahedra during the HAdV replication cycle. Dodecahedra are twelve-faced platonic solids (Figure 5a), which in this case are composed of twelve penton base pentamers (60 copies of the penton base in total; Figure 5b), often also containing 12 fibre trimers (158). Moreover, when producing only the penton base in a bacterial or insect cell expression system, fibre-less dodecahedra can be formed, too (159). This often produces a mixture of pentons, dodecahedra, and smaller penton multimers (160). Obtaining a homogenous preparation for structure determination experiments has been one of the challenges of this project.

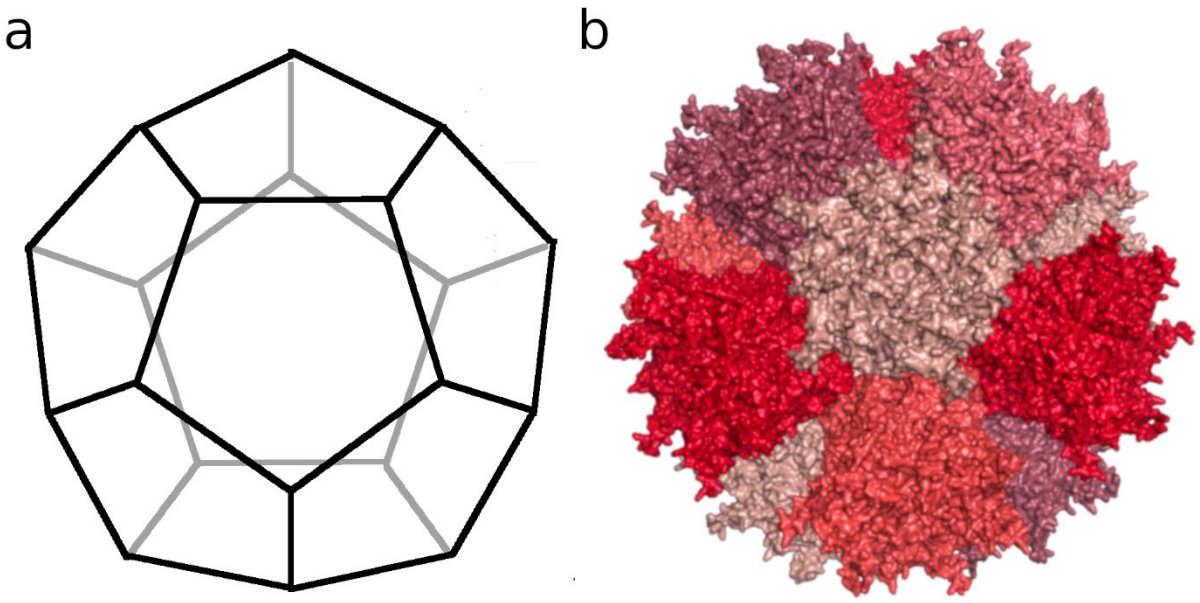


Figure 5. The HAdV Penton base dodecahedron. a. A dodecahedron. **b.** Structure of the HAdV-C2 penton base-only dodecahedron (151). Panel **b** generated with PyMol (139).

A number of structural mechanisms are thought to be involved in the formation of dodecahedra, and their contribution seems to vary between HAdV types. Firstly, some types do not form dodecahedra spontaneously (e.g., HAdV-C2, 151), while others may form them depending on production conditions, e.g., HAdV-B3; (160,161). Dodecahedra formation can also be induced by a change in conditions, for instance, HAdV-C2 penton base forms dodecahedra in its crystallisation solution (151). The crystal structure of the HAdV-B3 dodecahedron and subsequent mutational studies has led to the conclusion that dodecahedra formation is driven by a strand-swapping mechanism of the N-termini of neighbouring pentons, together with additional bonds formed by residues aspartic acid 100 and arginine 425. Change in amino acids 59 to 61 resulted in complete abolition of dodecahedra production, while other changes, including removing the first 47 amino acids, decreased it significantly, suggesting the N-terminus to be crucial (160). However, spontaneous proteolysis at the N-terminus until amino acid 37 has been proposed as the trigger for dodecahedra formation in penton bases of the same type by some of the same authors (161). While it is possible that the additional ten amino acids mediate between driving and inhibiting dodecahedra formation, this discrepancy within a single HAdV type highlights the problem of establishing and controlling the mechanism, a challenge that is magnified by the diversity between types.

From a physiological perspective, dodecahedra produced during viral replication are thought to deliver additional receptor-engaging proteins to neighbours of the infected cell. This may result in weakening poorly accessible tight junctions where some attachment receptors occur and remodelling the cell (162,163). In a biotechnological context, dodecahedra have been proposed as delivery vectors for small molecules (159,164).

1.6. Objectives and Expected Outcome – Part 2

The goal of the research project was to structurally characterise the HAdV type C5, D9 and B35 penton base interactions with the integrin $\alpha_v\beta_3$: the types were selected for their clinical (C5 is a common respiratory pathogen, 58) or technological relevance (application of both C5 and B35 as a gene delivery vector), and the diversity of RGD loop lengths: 99 amino acids in C5, 54 in D9, and 77 in B35 .

Given the high flexibility of the integrin leg domains and the predicted multiple stoichiometries of the complex (the pentameric penton base has five RGD loops, each potentially binding an integrin), cryoEM was chosen as the experimental method. X-ray crystallography requires a highly concentrated, homogeneous sample to obtain high-quality crystals. The variety of conformations that the integrin assumes, combined with the high chance of multiple stoichiometries of the complex, would limit the chances of producing well-diffracting crystals. CryoEM is a method better adapted to visualising multiple conformational states in the same sample and specimens of high molecular weight and has the additional advantage of requiring less material at a lower sample concentration.

Previous research included a negative-stain electron microscopy (EM) structure of the penton base in complex with the integrin. However, the applied method yielded lower-resolution (about 20 Å) structures, unable to resolve atomic details above general protein shapes (165). Moreover, a monomeric penton base protein was used, which made it impossible to establish the stoichiometry of the binding and allowed for a greater number of conformations. Therefore, the first goal of the project was to express and purify the penton base proteins in insect cells as pentamers, an approach that has failed in *E. coli*, but would allow to mimic the environment in the capsid more closely. The expression and purification of the integrin binding partner was to be re-established and optimised based on protocols from previous work in the lab.

Once purification of both components had been successful, complex formation under different conditions would be examined using size exclusion chromatography (SEC) to identify the stoichiometry of binding. This question has been previously investigated in cryoEM structures of the whole capsid in complex with integrins, but given the low resolution and application of five-fold symmetry, the data on this were averaged out (166,167). This produced a structure where one integrin was bound to one RGD motif, five integrins per pentameric penton. Such a model is probably sterically impossible due to the size and flexibility of the leg domains. Similar questions relate to four

integrins per penton, which previous surface plasmon resonance (SPR) studies of virus-integrin complexes suggested this was the answer (167). This leaves values between 1 and 3 as possibilities, likely with all possibilities present, but questions remain about preferences, especially considering that integrin signalling can be clustering-mediated.

Finally, the optimised complexes would be used in cryoEM studies and the resulting structures give insights into the mechanisms of binding between the penton base and its receptor on an atomic level, particularly into the influence of the differences in RGD loop length between HAdV serotypes.

1.7. Structural Biology Methods

Structural biology is a branch of science that focuses on the shapes of biological molecules and their relationship with function. This provides information on the mechanics underlying biochemical reactions: for instance, clarifying reaction mechanisms and structure-function relationships. Such an understanding can enable design of therapeutic compounds to interact with a particular process, explain the significance of mutations, or the course of evolution. As examples, the HIV drug saquinavir was designed based on the structure of the viral protease, as was the influenza neuraminidase inhibitor Zanamivir and Oseltamivir (168,169).

The major experimental techniques of structural biology are X-ray crystallography, cryogenic electron microscopy and nuclear magnetic resonance (NMR), which can be employed to establish atomic-resolution structures of biomolecules, particularly proteins and their binding partners (reviewed in 170). Together with a range of accessory techniques, such as small-angle X-ray scattering (SAXS), hydrogen-deuterium exchanging with mass spectrometry (MS), or electron paramagnetic resonance spectroscopy, they can be used to characterise molecules and complexes of a wide range of sizes, also in a dynamic context. Moreover, more accurate protein structure-prediction software was recently made available in the form of the neural network-based AlphaFold2 (171).

Every structural biology technique has its own characteristics when it comes to achieving optimal results and applications, enabling a complementary approach (170). The viral part of the thesis used cryoEM, while the bacterial part employed X-ray crystallography, choices explained in the description of the research questions. Figure 9 shows the generalised workflows of both techniques.

1.7.1. X-Ray Crystallography

X-ray crystallography is a well-established, over a century-old, technique for macromolecular structure solution (172) that uses diffraction patterns recorded when shooting X-rays at a crystal of the molecule of interest. The indices of the diffraction spots, together with externally provided phase information, are used to calculate the reverse Fourier transform of the sum of diffracted waves to generate an electron density map, which is then iteratively refined (170).

1.7.1.1. Protein crystallisation

Crystals are solids formed of highly regular arrangements of molecules in very similar orientations. The molecules are arranged in a periodic lattice, which can be defined as

a single unit cell undergoing translational symmetry operations. The unit cell is a parallelepiped, and the relationships between its edges and angles define the seven lattice systems. If the type of centring, where additional lattice points are located on the faces, not just vertices of the unit cell, is also considered, there are 14 Bravais lattices. Combining the 14 Bravais lattices with crystallographically and chemically (e.g., the chirality of proteins excludes mirror symmetry in most biologically relevant samples) allowed symmetry operations other than translation produces space groups, which define the symmetry of the unit cell in relation to its constituent asymmetric units. The unit cell may be divided into sets of planes, where a plane (hkl) intersects the x-axis of the unit cell at $\frac{x}{h}$, the y-axis at $\frac{y}{k}$, and z-axis at $\frac{z}{l}$. The values of h, k, and l, scaled to be given as the smallest possible integers, are known as Miller indices, and represent a set of equivalent planes, which are parallel to each other and defined by integer multiples of the intercept values (173).

Among the advantages of using crystals for X-ray diffraction experiments are the amplification of signal by the repeating arrangement of atoms and maintaining the protein in a hydrated, non-denatured state. Protein crystals differ from crystals of inorganic compounds, which are typically formed by strong charge-based ionic interactions, in their reliance on the weaker hydrogen bonding for cohesion. As a result, protein crystals are smaller and more delicate, while the necessity of maintaining the proteins in their native state introduces restrictions on the methods that could otherwise be used for crystal growth. To obtain optimal quality (highly ordered repetition of unit cells with minimal deviations from the symmetry), crystals should increase in size slowly to maintain order, but the end products should be possibly large, to enable handling and to amplify the signal.

Popular methods rely on controlled supersaturation, nucleation, and growth, to achieve the phase transition from liquid to crystalline solid (reviewed in 174), for instance by vapour diffusion of supersaturated protein solutions in closed systems. A drop of highly concentrated protein solution (the optimal concentration varies between proteins) is mixed in a 1:1 ratio with a precipitant solution and maintained in a closed environment with a much larger reservoir of the precipitant solution. The precipitant concentration in the protein-containing drop is equilibrated to the concentration in the reservoir by water evaporation, gradually increasing the concentration of both protein and precipitant. This leads to gradual supersaturation of the solution, enabling crystal nucleation and growth (see **Figure 6**). If the supersaturation is too high, the protein

precipitates without forming crystals. At more moderate levels, the supersaturated solution forms crystal nuclei that may initiate crystal growth: however, too many nuclei will prevent the growth of larger crystals. At even lower levels, in the so-called metastable zone, crystal growth occurs, but without nuclei formation. The goal of the vapour diffusion methods is to slowly reach supersaturation levels suitable for crystal nucleation, but not precipitation, and, as the nucleation removes protein from the solution, to then reach the lower supersaturated concentration enabling stable crystal growth, but not further nucleation. If this proves difficult and only numerous crystals too small for diffraction can be grown, the nucleation stage can be omitted by supplying nuclei seeds externally to the protein-precipitant drop at the start of the experiment, for instance by adding microcrystals from previous experiments to a lower initial protein concentration.

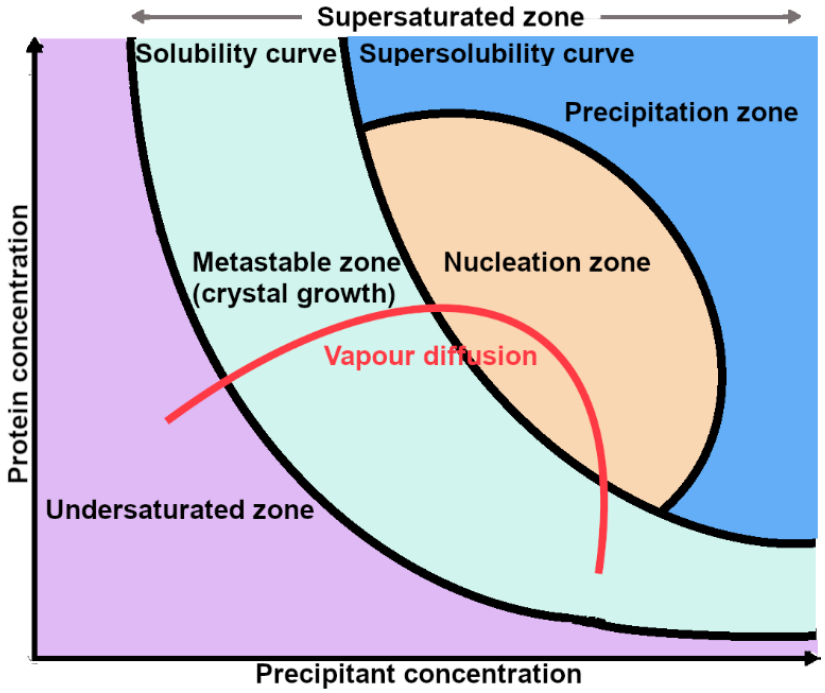


Figure 6. Schematic phase diagram of protein crystal formation. **Red**, concentration changes in the course of a successful vapour diffusion crystallisation. Redrawn from (174).

In practice, crystal production can be very challenging: even when crystals can be obtained, they may diffract poorly due to disorder in the packing of molecules, or well-diffracting crystals may be twinned, i.e., contain two intergrown crystal lattices related by symmetry operators (175). Both these issues have been encountered in the bacterial project.

1.7.1.2. From X-ray diffraction by protein crystals to structure solution

The diffraction experiment is performed at low temperatures, which enhances the results by reducing the impact of radiation damage, the movement of the atoms within the crystal, and the background from the solution in which the crystals are stored. Once the crystals have been produced, they can be harvested and shot with X-ray beams to record diffraction patterns necessary for structure determination. To prevent the formation of ice crystals by the mother liquor in which the protein crystals were grown and within the crystal solvent channels, they are harvested with a mounting loop into a cryoprotectant solution containing glycerol or another substance reducing ice nucleation and flash-frozen, typically in liquid nitrogen. This ensures the ice around the crystal is vitreous, not crystalline, reducing its unnecessary signal, and that ice crystals do not form, destroying protein crystals (reviewed in 176).

X-rays are a form of electromagnetic radiation at wavelengths between 0.1 and 100 Å. They may be generated when an accelerated external electron displaces an electron from a lower-energy atomic orbital, causing other electrons to drop from a higher-energy orbital and emit the extra energy as radiation. Such radiation can be generated via X-ray tubes or other laboratory equipment; however, synchrotron radiation is typically the most optimal for macromolecular crystallography experiments. Synchrotron radiation is emitted when electrons are accelerated along curved paths, such as in synchrotron storage rings. It produces a more coherent beam with a higher flux and a single wavelength, which decreases the time needed to record data, increases its quality, and enables the solution of structures with larger unit cells (reviewed in 177).

Shooting an X-ray beam at a crystal diffracts the rays, the divergence of which can be recorded as diffraction spots by a detector or on film. The diffraction is caused by the electron clouds of the molecules in the crystal, with the signal amplified by the periodic arrangement of the crystal. The relationship between the experimentally obtained diffraction spots and the electron density, ρ, of the molecule in the crystal can be described mathematically (summarised in 178). Electron density may be expressed as a Fourier sum of complex three-dimensional wave functions:

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} F_{hkl}e^{-2\pi i(hx+ky+lz)}$$

Where x, y, z are the coordinates of a point, V is the volume of the unit cell, h, k, and l the Miller indices of a set of equivalent planes in the crystal, and F_{hkl} the structure factor

function for the (*hkl)* planes. F_{hkl} is the sum of the scattering contributions of the *hkl* planes by all atoms in the unit cell and also a three-dimensional wave function, a Fourier sum:

$$F_{hkl} = \sum_{j=1}^{n} f_j e^{2\pi i(hx_j + ky_j + lz_j)}$$

where *j* is an atom in the unit cell, f_j is the scattering factor of *j* (dependent on the element of the atom), and x_j, y_j, and z_j are the coordinates of the atom in the unit cell as well as the indices of the diffraction spots with respect to an arbitrary origin. The structure factor can also be expressed as a vector on a complex plane, as it is the sum of the vectors of scattering by single atoms. Substituting the vector expression into the electron density function:

$$\rho(x, y, z) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} |F_{hkl}| e^{-2\pi i(hx + ky + lz - \alpha_{hkl})}$$

where α_{hkl} is the phase of F_{hkl}. The phases cannot be generated from the experimental data but establishing of the other parameters necessary to describe a wave function: frequency and amplitude is possible. Frequency is proportional to the intensity of the reflections, while the amplitude is determined by h, k, l, which, from Bragg's law, are not only Miller plane indices in real space, but also the coordinates of the diffraction spots in reciprocal space.

Bragg's law (Figure 7) determines that only at certain angles can the waves interfere constructively:

$$2d_{hkl} \sin\theta = n\lambda$$

where d_{hkl} is the distance between planes with indices the integer multiples of h, k, and l, *n* an integer number, and λ the wavelength.

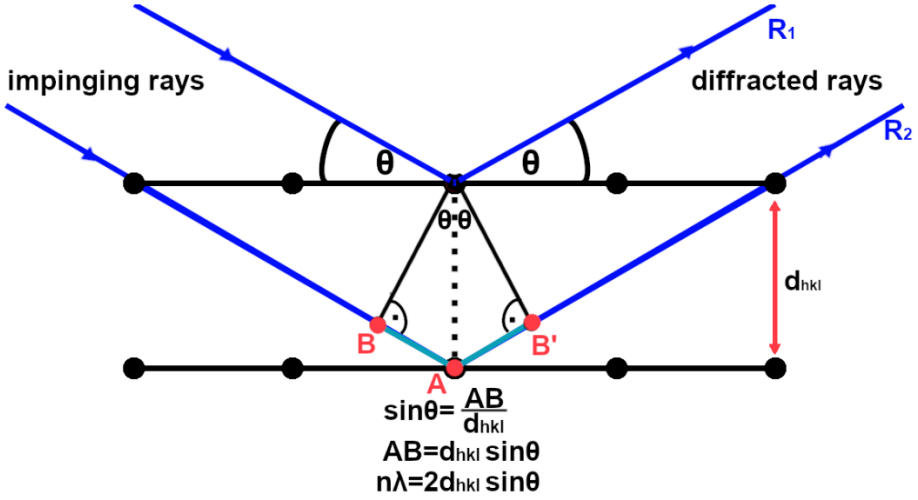


Figure 7. A diagram illustrating Bragg's law. Rays may interfere productively only at certain angles for a given wavelength λ, otherwise they are out of phase. d_{hkl}, distance between equivalent planes, θ, angle at which the ray impinges and is reflected from the plane, R_1, R_2: rays. Redrawn and adapted from (179).

The Ewald sphere (see **Figure 8**) describes the relationship between the allowable reflections according to Bragg's law and the reciprocal lattice of the crystal (reviewed in 180). The diffraction spots may be considered as occupying an infinite lattice system reciprocal to the real space lattice in the crystal. Consequently, for reflections of rays scattered under the Bragg condition, the distance between a diffraction spot and an origin located along the original beam on a sphere with the radius $\frac{1}{\lambda}$ will be $\frac{1}{d_{hkl}}$ and the spot will be defined by a vector normal to the set of (*hkl*) planes in the crystal. This system shifts as the crystal is rotated to produce more allowable reflections.

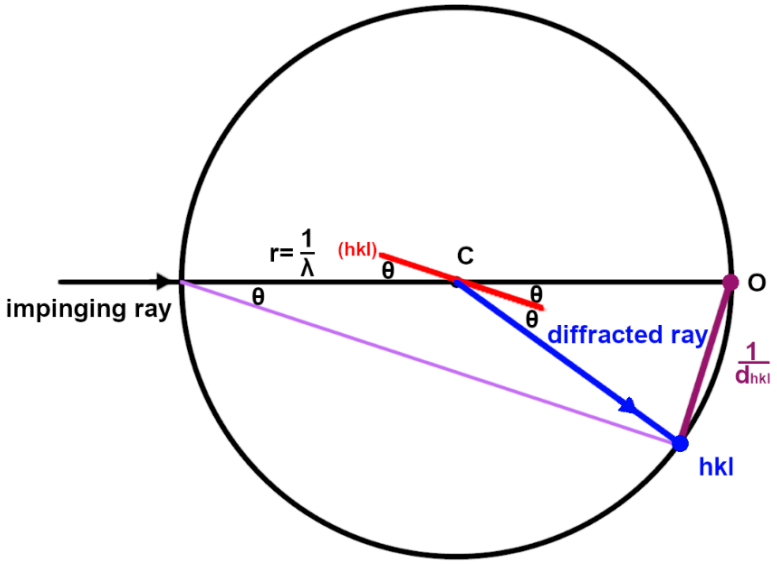


Figure 8. Ewald sphere construction. In a circle (or sphere in 3D) of radius $\frac{1}{\lambda}$, where λ is the wavelength, the reflection **hkl** is $\frac{1}{d_{hkl}}$ distant from the origin of the reciprocal lattice and the vector between O and **hkl** is perpendicular to the (**hkl**) set of planes in the crystal from which

the ray was diffracted. C, crystal in the centre of the circle. O, origin of the reciprocal lattice and the intercept of the undiffracted ray with the circle. θ, the Bragg angle at which the ray is diffracted by the plane. Redrawn and adapted from (180).

Due to the absence of information on phases in the diffraction pattern data, it is necessary to supply them externally to establish the electron density. This is known as the phase problem (summarised in 181) and may be solved experimentally by generating and comparing the reflections of heavy-atom derivatives of the proteins, or by substituting the phases calculated from the known structure of a related protein to obtain initial density. The latter approach is known as molecular replacement (MR) and has been used to solve both crystallographic structures in this work. In both cases, the Patterson function is applied to compare vectors representing distances between atoms.

Once a molecular model can be built into the initial electron density, it can be used to calculate new, more accurate phases. The iterative steps of model building and refinement in reciprocal space should produce a final model. The accuracy of the model may be assessed by a set of metrics quantifying the agreement between the model and experimental data, most prominently the R and R_{free} values, which compare the experimental and calculated reflection intensities of the model.

The disadvantages of X-ray crystallography are the necessity to obtain large amounts of highly pure protein for crystallisation trials, and the largely unpredictable process of crystal growth (170). The latter was a major challenge throughout the bacterial part of this project, despite the prior successful crystallisation of the proteins in their unbound forms and of MurK in complex with an ATP analogue (48). Moreover, large, flexible or non-homogeneous protein assemblies are not suitable for crystal formation, the reason that made this technique unsuitable for the viral project in this thesis.

Nonetheless, once the barriers of purification and crystallisation have been overcome, X-ray crystallography generally allows for high-resolution structures, providing a solution to the phase problem is found (181). This has become easier with the introduction of high-accuracy modelling with AlphaFold2, even in cases where there are no homologous structures solved experimentally. Additionally, a range of software helps automate the process of structure solution (e.g., CCP4, 182), while a set of widely accepted metrics provides means of quality assessment of the resulting structures (183). Finally, and crucially for the bacterial project in this thesis, X-ray crystallography is well-suited to the solution of structures of small (<50-100 kDa)

proteins, something that is more challenging with cryoEM, requiring the use of extensive computational resources and highest-performance equipment.

1.7.2. Cryo-EM

Cryogenic electron microscopy is a technique for obtaining high-resolution structures, generally of larger (>50 kDa) proteins and macromolecular complexes by imaging frozen samples with transmission electron microscopy (reviewed in 170,184). A focused beam of electrons is directed onto a specimen flash-frozen in liquid ethane, focusing it with a magnetic lens and recording the signal, onto samples. Similarly as in cryocrystallography, where crystals are flash-frozen for protection from radiation damage (see section 1.7.1), flash freezing produces vitreous, or amorphous, non-crystalline ice, maintaining the macromolecular structure pre-existing in solution also in the vacuum environment. Additionally, the coherence of the electron beam is of similar importance for the quality of results as that of the X-ray beam in crystallography. The samples, located in ice-filled holes on microscopy grids, are then imaged by EM, producing 2D micrographs of the molecule of interest in different orientations. Initial processing of the micrographs is performed to account for noise and convolution with the artefacts of the imaging process such as the point spread function (PSF) and the contrast transfer function (CTF) or movement of the sample caused by the electron beam (184).

Using powerful computational techniques, the resulting sets of 2D projections, particles, are excised from the micrographs and classified according to orientation. The members of resulting classes are averaged, which allows for visual inspection and identification of “low-quality” classes giving poor averages, which do not contain similar particles. A subset of particles from the 2D classes can be used to produce an initial low-resolution 3D model, or this can be achieved by low-pass filtering a homologous structure (184).

The high-quality particles are then used to calculate a higher-resolution 3D model by establishing the relative orientations of their transforms in Fourier space and then back-projecting the particles onto the initial model to obtain an improved, more detailed model in real space. This model can be used as a starting volume for the same process, in an iterative cycle until no more improvements in resolution are made. The 3D reconstruction can then be adjusted with a number of further techniques such as particle polishing (185) and map sharpening (186) to optimise the resolution.

Historically, the resolutions obtained by cryoEM were generally lower than those from X-ray crystallography, reaching only about 10 Ångstroms. This changed dramatically in the previous decade, when the introduction of direct electron detectors greatly enhanced the signal-to-noise ratio (reviewed in 187) and automation of data collection increased the sizes of data sets, leading to a phenomenon dubbed the “Resolution Revolution” (188). The 2017 Nobel Prize in Chemistry was awarded for development of the method and the number and resolution of structures are both rising (189).

Advantages of cryoEM include the ease of sample preparation compared to X-ray crystallography, as less material is needed and there is no need for crystallisation, the possibility of resolving large complexes, particularly transient or flexible ones, and the visualisation of samples in their natively solvated state (184).

However, though resolution barriers are constantly being broken, achieving up to 1.2 Å (190), the average resolution for cryoEM structures is still below the average value for X-ray structures (191,192), and structures for proteins smaller than 100 kDa are rare and challenging to solve, while those under 50 kDa rely on augmenting their molecular weight (M_w) with tags or scaffolds (191, reviewed in 193). Additionally, the data processing stage can be challenging or introduce bias: for instance, an inappropriate initial model may result in a reconstruction with features absent in the dataset (194). This issue is compounded by the lack of widely accepted metrics to evaluate the quality of the structures as used in crystallography (195). Given the relative youth of the field, these are still being developed (196), and new programmes and approaches are introduced regularly, supplying solutions to earlier problems, but also increasing the challenge of keeping up with the state of the art for the nonexpert users.

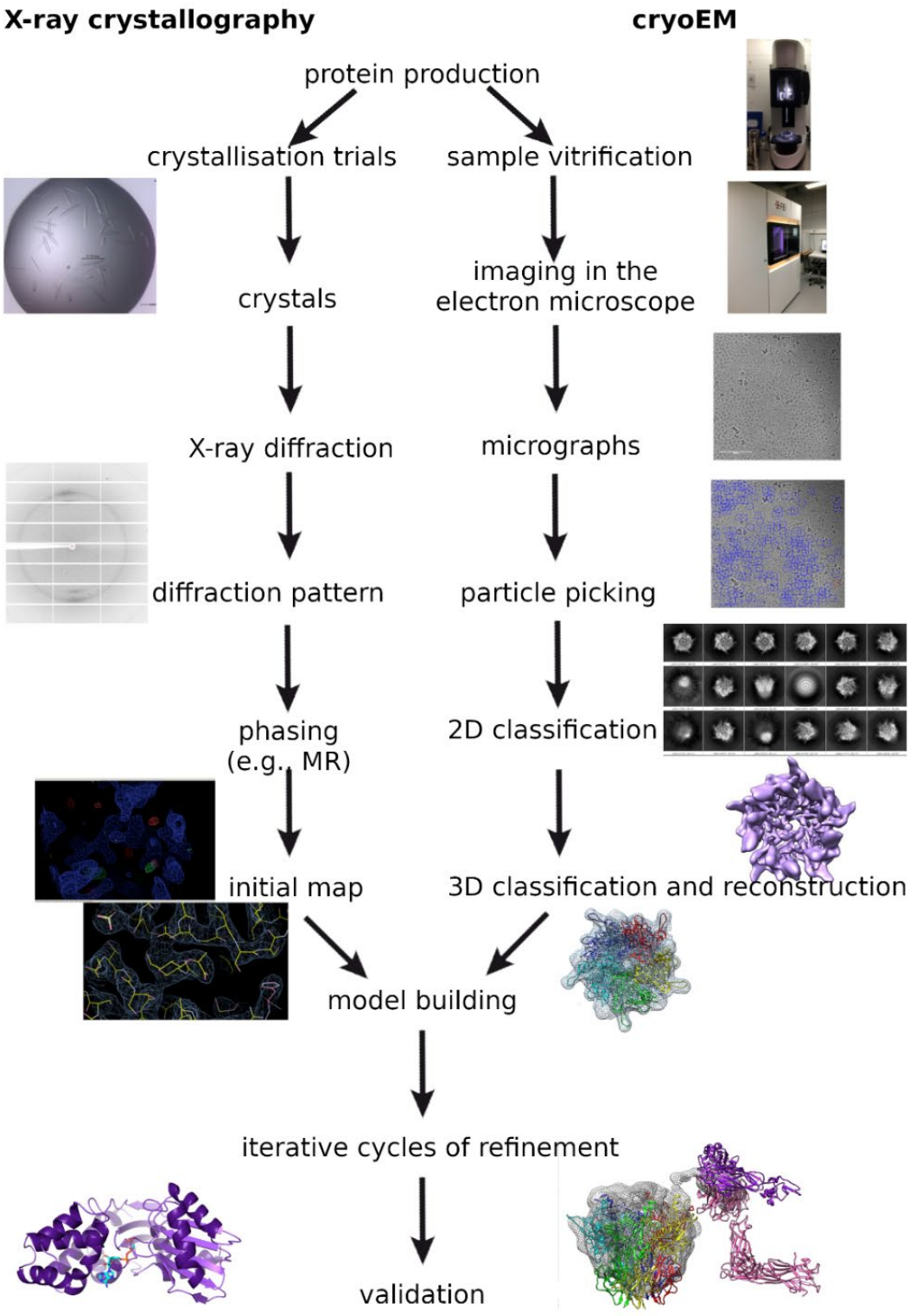


Figure 9. X-ray crystallography and cryoEM workflows. Both techniques rely on iterative cycles of refinement to arrive at the final model from experimental data. MR: molecular replacement.

2. Materials and methods

2.1. Materials

2.1.1. Chemicals

All chemicals were analytical grade and manufactured by Sigma-Aldrich (Taufkirchen, Germany), Roth (Karlsruhe, Germany), Roche (Basel, Switzerland), ThermoFisher Scientific (St. Louis, Missouri, USA), or Merck (Darmstadt, Germany). Crystal screens were manufactured by Molecular Dimensions (Wizard I/II, III/IV, JCSG, Morpheus; Sheffield, UK) or Hampton Research (Crystal I/II; Aliso Viejo, USA). Manufacturers of reagents key for the experiments are specified in the Methods section within the description of the protocols.

2.1.2. Constructs

Table 1. DNA constructs used in this work.

Protein encoded (Abbreviation)	NCBI Protein ID (Amino acid changes)	Plasmid	Tags	Tag Cleavage Site	Source
MurK	WP_046825531.1	pET28a	C-terminal-6xHis-tag	-	Mayer group, IMIT Tübingen
MurK I69T; **MurK** P134A; **MurK** I69T P134A	WP_046825531.1 I69T; P134A; I69T P134A	pET28a	C-terminal-6xHis-tag	-	BioCat, Heidelberg, Germany
K1058	WP_041590480.1	pET28a	C-terminal-6xHis-tag	-	Mayer group, IMIT Tübingen
K1058 T69I; **K1058** A133P; **K1058** T69I A133P	WP_041590480.1 T69I; A133P; T69I A133P	pET28a	C-terminal-6xHis-tag	-	BioCat, Heidelberg, Germany
HAdV-C5 penton base wild type (pb5wt)	AGT78086.1	pFastBac	N-terminal-6xHis-tag	TEV protease	Arnberg group, Umeå University
HAdV-C5 penton base dodecahedra non-forming mutant (pb5mut)	AAA42519.1 truncated beyond the SELA motif aa 65-571, R452E	pFastBac	N-terminal 6xHis-tag	TEV protease	BioCat, Heidelberg, Germany

HAdV-D9 penton base dodecahedra non-forming mutant (pb9mut)	CAI05964.1 truncated beyond the SELA motif aa 52-519, R400E	pFastBac	N-terminal 6xHis-tag	TEV protease	BioCat, Heidelberg, Germany
HAdV-B35 penton base dodecahedra non-forming mutant (pb35mut)	AAN17481.2 truncated beyond the SELA motif aa 65-561, R442E	pFastBac	N-terminal 6xHis-tag	TEV protease	BioCat, Heidelberg, Germany
Extracellular domains of **human integrin $\alpha_v\beta_3$**	α_v (aa 1-987) and β_3 (aa 1-718) subunits encoded in separate loci and dimerised with a Fos-Jun domain, see previous publication (165)	pFastBacDual	C-terminal: 6xHis-tag on α_v, Strep tag on β_3	HRV 3C protease	Cloned and used for production of baculovirus stocks by K. Cupelli (197)

2.1.3. Bacterial strains

In-house propagated chemically competent *E. coli*: MAX Efficiency DH10Bac (ThermoFisher Scientific) were used for Bacmid production of baculovirus expression, MAX Efficiency DH5α (ThermoFisher Scientific) for DNA isolation, BL21DE3 (Novagen, Darmstadt, Germany) for protein expression

2.1.4. Cell lines

Sf9 (*Spodoptera frugiperda*) and High Five (Hi5; *Trichoplusia ni*) insect cell lines were acquired from ThermoFischer Scientific.

2.1.5. Buffers

MurK/K1058 IMAC wash buffer
25 mM Tris-HCl, 500 mM NaCl, 25 mM imidazole, 10% (v/v) glycerol, pH 8.0

MurK/K1058 IMAC elution buffer
25 mM Tris-HCl, 500 mM NaCl, 500 mM imidazole, 10% (v/v) glycerol, pH 8.0

MurK SEC buffer
200 mM NaCl, 50 mM 2-(N-morpholino)ethanesulfonic acid (MES), pH 6

K1058 SEC buffer
750 mM NaCl, 25 mM Tris-HCl, pH 8.5

HAdV penton base WT lysis buffer (Zubieta protocol, 151)
25 mM Tris, 150 mM NaCl, pH 7.5

HAdV penton base WT lysis buffer 2 (Arnberg group protocol, modified, 138)
25 mM Tris, 500 mM NaCl, 10% glycerol, 0.2% Triton-X 100, 10 mM imidazole, 2 mM β-mercaptoethanol, pH 8

HAdV penton base WT HEPES lysis buffer (M. Liaci's protocol, 198)
150 mM NaCl, 100 mM HEPES, 5 mM dithiothreitol (DTT), 10 mM imidazole, benzoase, $MgCl_2$, 1 protease inhibitor tablet, pH 8

HAdV penton base WT IMAC wash buffer (Zubieta protocol, 151)
25 mM Tris, 150 mM NaCl, 20 mM imidazole, pH 7.5

HAdV penton base WT IMAC wash buffer 2 (Arnberg group protocol, modified, 138)
25 mM Tris, 500 mM NaCl, 20 mM imidazole, 2 mM β-mercaptoethanol, pH 8

HAdV penton base WT IMAC HEPES wash buffer (M. Liaci's protocol, 198)
150 mM NaCl, 100 mM HEPES, 5 mM DTT, 20 mM imidazole, pH 8

HAdV penton base WT IMAC elution buffer (Zubieta protocol, 151)
25 mM Tris, 150 mM NaCl, 250 mM imidazole, pH 7.5

HAdV penton base WT IMAC elution buffer 2 (Arnberg group protocol, modified, 138)
25 mM Tris, 150 mM NaCl, 500 mM imidazole, pH 7.5

HAdV penton base WT IMAC HEPES elution buffer (M. Liaci's protocol, 198)
150 mM NaCl, 100 mM HEPES, 5 mM DTT, 500 mM imidazole, pH 8

HAdV penton base WT IMAC HEPES SEC buffer (M. Liaci's protocol, 198)
150 mM NaCl, 100 mM HEPES, 5 mM DTT, pH 8

HAdV penton base mutant high salt lysis buffer
500 mM NaCl, 100 mM HEPES, 5 mM DTT, 10 mM Imidazole, benzoase, $MgCl_2$, 1 protease inhibitor tablet, pH 8

HAdV penton base mutant IMAC high salt wash buffer
500 mM NaCl, 100 mM HEPES, 20 mM imidazole, pH 8

HAdV penton base mutant IMAC high salt elution buffer
500 mM NaCl, 100 mM HEPES, 500 mM imidazole, pH 8

HAdV penton base mutant SEC high salt buffer
500 mM NaCl, 100 mM HEPES, pH 8

Integrin $\alpha_v\beta_3$ IMAC wash buffer
300 mM NaCl, 25 mM Tris-HCl, 5 mM $CaCl_2$, 10 mM imidazole, pH 8

Integrin $\alpha_v\beta_3$ IMAC elution buffer
300 mM NaCl, 25 mM Tris-HCl, 5 mM $CaCl_2$, 500 mM Imidazole, pH 8

Integrin $\alpha_v\beta_3$ SEC buffer
150 mM NaCl, 25 mM Tris-HCl, 1 mM $CaCl_2$, pH 8

SDS-PAGE (Sodium dodecyl sulfate polyacrylamide gel electrophoresis) resolving gel
10/12/15% v/v acrylamide solution, 25% v/v 1.5 M Tris-HCl pH 8.8, 0.1% v/v SDS, 0.05% v/v N-tetramethylethylenediamine (TEMED), 0.1% v/v ammonium persulphate (APS)

SDS-PAGE stacking gel
4% acrylamide, 25% v/v 1.5 M Tris-HCl pH 6.8, 0.1% v/v SDS, 0.1% v/v N-tetramethylethylenediamine (TEMED), and 0.1% v/v APS

Protein Sample SDS-PAGE buffer x4
22.6% v/v glycerol, 226 mM Tris pH 6.8, 1% SDS, 90 mM EDTA pH 8, 4.5% v/v β-mercaptoethanol, 22.6% w/v bromophenol blue

SDS-PAGE running buffer
Rothiphorese TBE Buffer, diluted x10 (Roth)

TAE agarose gel buffer
40 mM Tris, 20 mM acetic acid, 1 mM EDTA

Western blotting buffer
25 mM Tris, 156 mM glycerol, 20% v/v methanol

Tris-buffered saline (TBS) x10
200 mM Tris pH 7.5, 1.5 M NaCl

Cross-linking/complex vitrification buffer
210 mM NaCl, 37 mM HEPES, 0.8 mM $CaCl_2$

Bacterial culture

S.O.C. medium: 2% w/v tryptone, 0.5% w/v yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose

LB medium: 25 g/L LB Lennox or Miller

LB-Agar: 15 g/L agar, 25 g/L LB Lennox or Miller

2.2. Methods

2.2.1. Bacterial protein expression

BL21(DE3) competent cells (50 µL) were transformed with 100 ng MurK or K1058-encoding plasmid by heat shock (30 minutes on ice, 45 seconds at 42 °C, 2 minutes on ice), then incubated with 900 µL LB-medium (25 g/L) at 37°C with shaking at 700 rpm for 1 hour. The mixture was either streaked on a LB-agar/kanamycin plate, then a 120 mL LB medium with kanamycin (LB/kan; 50 µg/mL) starter culture was inoculated from a single colony after overnight growth, or the starter culture would be grown directly overnight from the transformed bacteria, shaking at 110 rpm. The starter culture would then inoculate 2 or 4 L LB/kan, grown at 37°C, 100 rpm, until OD_{600} of 0.8-1.1 was reached. Then, the temperature was reduced to 18°C, and IPTG added to a final concentration of 0.5 mM to induce expression. The cultures were harvested after 18-22 hours by centrifugation at 9220 x g at 4°C for 10 minutes, the supernatant discarded, and the bacterial pellets stored at -80°C or lysed directly for purification.

2.2.2. MurK/K1058 protein purification

All purification steps were performed at 4°C or on ice. Samples for SDS-PAGE electrophoresis were taken throughout. The pellet was resuspended in MurK/K1058 IMAC wash buffer and lysed with a sonicator (Branson 250 Digital Sonifier Cell Disruptor, Marshall Scientific, Hampton, US) by pulsing at 50% for 0.5 second with 2.5 second rests for 5 minutes pulse time total. The lysate was centrifuged at 34500 x g for 45 minutes. The supernatant was loaded onto a 5 mL HisTrap-FF nickel column (pre-equilibrated with water, 1 M imidazole, and MurK/K1058 IMAC wash buffer; GE Healthcare, Chicago, USA, or Cytiva, Marlborough, USA) at a rate of 0.8 mL/min, then loaded circularly for 1 hour at 1.5 mL/min. The column was then installed on an ÄKTA Prime system (GE Healthcare), washed with the wash buffer, then 5 and 10% of the MurK/K1058 IMAC elution buffer until a base line of absorbance at 280 nm was reached. The protein was eluted with 80% of the elution buffer and collected in 2 mL fractions. The highest-concentrated fractions with the best purity were pooled based on Abs_{280} peaks and SDS-PAGE analysis, then concentrated to a volume of 3-5 mL using 10 kDa concentrators pre-washed with water and SEC buffer, at 1470 x g in 10-minute steps. The concentrate was filtered with a 0.22 µm filter before loading onto a HiLoad SEC Superdex 200 16/60 preparative grade column (Cytiva) installed on an ÄKTA Basic system (GE Healthcare), prewashed with the appropriate degassed SEC buffer. The protein was eluted with MurK or K1058 SEC buffer and collected into 1- or

2-mL fractions, which were pooled based on purity. The pool was concentrated to 1-3 mL as described and stored at -80°C after flash-freezing.

2.2.3. **Polyacrylamide gels and SDS-PAGE electrophoresis**

10, 12 or 15% polyacrylamide gels were prepared inhouse in batches of 12. The resolving gel solution was poured in between glass plates and covered with isopropanol to polymerise. The stacking gel solution was poured on top of the solidified separating gel and sealed with 10- or 15-well combs. Once polymerised, the gels were stored between glass plates wrapped in moist tissues at 4°C. Gels were run with 15 µL of protein samples loaded into wells (boiled at 95 °C for 5 minutes, then centrifuged, with x4 concentrated SDS-PAGE protein sample buffer), except for pellet samples, which were resuspended in the concentrated sample buffer, and supernatant samples, both of which were loaded at 5 µL/well to avoid overloading. Gels were run at the voltage of 150 to 250 V, increasing the speed after the stacking gel had been cleared. The gel was stained with a fast commercially bought Coomassie-based stain, then washed in water 2 or 3 times and scanned.

2.2.4. **Crystallisation of MurK and K1058 in complex with sugar substrates**

Enzyme-substrate complexes were formed by adding MurNAc (final concentration 10 mM, dissolved in 1 M HEPES pH 7.5) or GlcNAc (final concentration 10 mM, dissolved in appropriate SEC buffer) and ATP analogues 5′-adenylyl methylenediphosphonate (AMP-PCP) or adenylyl-imidodiphosphate (AMP-PNP; final concentration 2 mM, dissolved in appropriate SEC buffer) to MurK or K1058 at a concentration of 6-12 mg/mL. Crystallisation plates were set up with the Freedom Evo (Tekan, Männedorf, Switzerland) or Gryphon (Art Robbins Instruments, Sunnyvale, USA) robots, and monitored at least monthly for crystals.

Initial screens were set up in a 96-well sitting drop format with commercially available screens (see section 2.1.1) or with a fine screen based on a previously identified crystallisation condition of MurK (0.1 M ammonium sulfate, bis-tris propane/citric acid 0.1 M, pH 5.5, 16% polythelene glycol 3350, 48). The protein solution was mixed in a 1:1 ratio with the crystallisation solution. Each screen was set up both at 4 and 20°C, with each protein in complex with one of the ATP analogues and with the MurNAc sugar substrate (MurK) or with MurNAc or GlcNAc substrates (K1058). Crystals were harvested by fishing and stored in cryoprotective solutions (fresh crystallisation solutions with 15% glycerol for K1058/MurNAc, based on (199), and 20% glycerol for MurK/MurNAc) in liquid nitrogen before transport to the synchrotron in a dry shipper.

20% glycerol for MurK/MurNAc) in liquid nitrogen before transport to the synchrotron in a dry shipper.

2.2.5. Seeding

Seeding was performed by harvesting microcrystals or crystals from conditions that had yielded non-diffracting crystals before and adding them to a polytetrafluoroethylene seeding bead (Hampton Research) in the appropriate crystallisation buffer, vortexing, and storing at -80°C for future use. The seeding beads were added to the crystallisation drop at a volume of 2.5 nL per 50 nL condition of a 10^{-2} or 10^{-3} diluted seed stock.

2.2.6. Fine screens

Fine screens were designed by varying the parameters of a hit condition: pH by 0.2 units within 1 unit of the hit, precipitant percentage within 10 percentage points in 2-point steps, and additive concentration, if any, was decreased or increased, two or three times. Fine screens were either prepared with the Tekan robot and then pipetted in the 500+500 nL well format or pipetted by hand for a hanging drop setup with 500 µL reservoir volume and 1+1 µL drop volume, optionally with 0.2 µL seeding solution.

2.2.7. X-ray crystallography data collection, processing, structure solution, and model building

Data were collected remotely at the X06Da beamline at the Swiss Light Source (SLS; Villigen, Switzerland). Data reduction was performed with the XDS/XSCALE package (200) and the phase problem was solved by molecular replacement with the previously solved structures: the apostructure of K1058 for the K1058/MurNAc structure, K1058/MurNAc for MurK/MurNAc. The N- and C-terminal domains were used as independent search models to account for the large conformational shift on sugar binding. Molecular replacement was performed with MOLREP implemented in CCP4 (182), with model building in Coot (201) followed by reciprocal space refinement in REFMAC5 (202). Ligands were only introduced after multiple cycles of refinement in which positive density was visible in a simulated annealing (F_o-F_c) difference omit map in the putative active site identified in the previously solved MurK/AMP-PCP structure. The MurK/MurNAc crystal was almost perfectly merohedrally twinned, and its structure solution involved simulated annealing with PHENIX.refine (203) with the model target geometry of the apostructure of MurK instead of REFMAC5 runs.

The structures were validated with the PDB server (204) and deposited in the PDB under accession codes 8ow9 and 8ow7 for MurK/MurNAc and K1058/MurNAc, respectively.

2.2.8. MurK/K1058 MurNAc phosphorylation coupled assay

The assay coupled formation of MurNAc-6P by MurK to reduction of NADH by lactate dehydrogenase, supplied with substrate by pyruvate kinase reducing ADP generated by the phosphorylation of MurNAc (see Figure 10). The change in absorbance at 340 nm (NADH absorbance range) could be measured with a spectrophotometer. 320 µL batches of a master mix (final concentrations: 50 mM HEPES pH 7.5, 10 mM $MgCl_2$, 5 mM ATP, 1 mM phosphoenolpyruvate, 0.2 mM NADH, 100 U/mL pyruvate kinase, 70 U/mL lactate dehydrogenase from rabbit muscle) were stored frozen at -80°C after reproducibility was confirmed by comparing reaction rates obtained when using non-frozen aliquots. After thawing, the master mix was added to 40 µL MurNAc solution in 1M HEPES for a final concentration of 0.0125, 0.025, 0.05, 0.075, 0.1, 0.2, 0.4, 0.6, 0.8, 1, 2 or 3 mM of the sugar and equilibrated to room temperature for 2 minutes. Finally, 40 µL of protein were added to a final amount of 178 ng (MurK) or 17.6 µg (K1058), concentrations 0.44 ng/µL or 44 ng/µL, respectively (original concentration 8.5-8.9 mg/mL, MurK was diluted 1:2000, K1058 1:20, both with 1 M HEPES pH 7). The decrease in Abs_{340} (data points measured every 10 seconds) were used to determine the reaction rates from the steepest part of the slope over the course of 100-200 seconds (according to 205,206). These were then converted from Abs/s to µmol of NADH/min/mg of enzyme used. Absorbance was converted to µmols by applying the Beer-Lambert law, with the extinction coefficient for NADH 6.22×10^3 M^{-1} cm^{-1}, and the length of the light path 1 cm. The obtained slopes were then plotted as single data points on a reaction rate against MurNAc concentration plot. The data were analysed with matplotlib in python and kinetic parameters established using Michaelis-Menten kinetics and substrate inhibition-adjusted Michaelis-Menten kinetics.

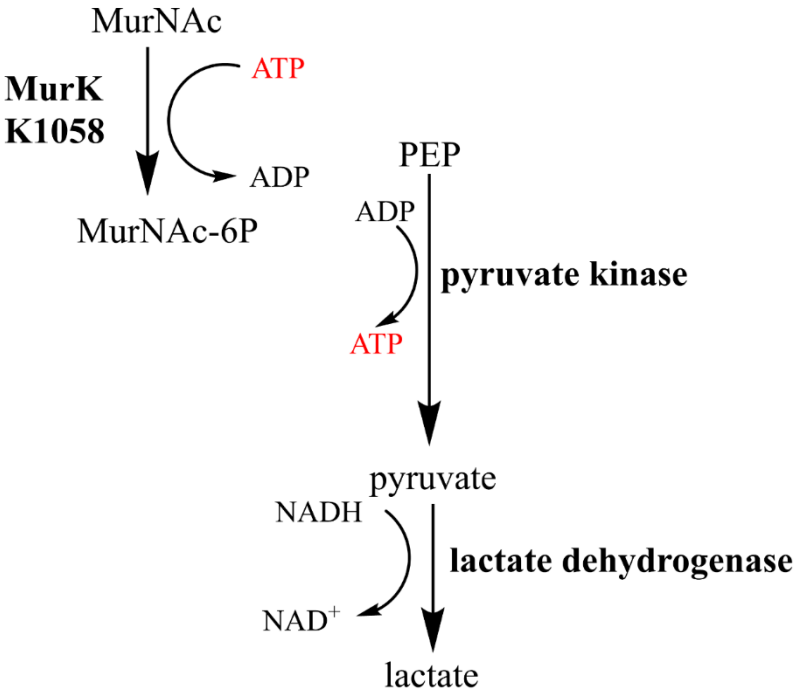


Figure 10. The coupled enzyme assay quantifying MurNAc phosphorylation by MurK and K1058. ATP is oxidised to ADP in the MurK/K1058 reaction. It is reduced back to ATP by pyruvate kinase. The pyruvate kinase-produced pyruvate is reduced by lactate dehydrogenase to lactate, oxidising NADH to NAD^+ in the process. The change in absorbance at 340 nm caused by the depletion of NADH is the measurable output.

2.2.9. Purifications of MurK/K1058 mutants

Constructs encoding MurK P134A, MurK I69T, MurK I69T P134A, K1058 T69I, K1058 A133P, K1058 T69I A133P (see Table 1) were purified according to the protocols described above for the corresponding wild type proteins. Stability was tested by flash-freezing and thawing the proteins three times in rapid succession, then comparing fresh and thrice-frozen samples by analytical SEC to confirm peak homogeneity. Folding was confirmed by circular dichroism.

2.2.10. Circular dichroism

Spectra for the protein solutions were recorded at wavelengths ranging from 250 to 195 nm on a Jasco J-720 (Jasco, Tokyo, Japan) machine, with 100 milidegrees sensitivity, in continuous scanning mode and 50 nm/min scanning speed, at 20 °C, in a 1 mm path length quartz cell under nitrogen gas to prevent the formation of radicals. Ten spectra were recorded per sample and averaged. The proteins were diluted 1:20 with water for a concentration of 0.1 – 0.3 mg/mL and the spectrum of the blank of buffer diluted 1:20 with water was subtracted from the protein spectrum.

2.2.11. Baculovirus stock generation

The penton base genes encoded in pFastBac (see Table 1) were transformed into DH10Bac cells as described above (see section 2.2.1), then plated on LB-Agar plates containing 50 µg/mL kanamycin, 10 µg/mL tetracycline, 7 µg/mL gentamycin, 100 µg/mL BluoGal and 40 µg/mL IPTG and incubated at 37°C for 48 hours. White colonies (indicating the gene-coding insert had been successfully introduced into the baculovirus genome-coding bacmid) were picked and grown in 10-25 mL LB or, preferentially, S.O.C. medium with the three antibiotics. The bacmids were purified using the PureLink HiPure Plasmid DNA Miniprep Kit (Invitrogen, Waltham, USA) according to the manufacturer's instructions. They were either stored at 4°C for up to a week or stored at -20°C for longer periods before transfection.

The presence of the insert was confirmed by polymerase chain reaction (PCR) with pUC/M13 forward and reverse primers on 100 ng of the isolated bacmid, with the master mix according to the composition recommended by the manufacturer's instructions for the *Taq* polymerase (New England Biolabs, Ipswich, USA). The PCR programme was as given in the Bac-to-Bac manual (207).

For transfections into insect cells, protocols provided by the manufacturer in the *Bac-to-Bac Baculovirus Expression System* manual (207) were followed. Briefly, Sf9 cells at initial density of 1.5-2 x 10^6 cells/mL were seeded in six-well plates at a final density of 8 x 10^5 cells/mL and left to attach for 15 minutes at room temperature. The medium was then exchanged to plating medium (Grace's Insect Cell Medium without antibiotics, ThermoFisher, with 1.5% foetal bovine serum, Invitrogen). Complexes of bacmid DNA (1-2 µg/well) with Cellfectin I or II (ThermoFisher; in one case, though out of date by about 10 years, the transfection was still successful), diluted in plating medium were incubated for 20 minutes and added dropwise to the wells. The plating medium was exchanged to regular serum-free antibiotic-supplemented medium after 5 hours. A mock transfection with medium and Cellfectin without DNA was included, and cells were monitored for signs of infection. The medium was harvested after 5 to 7 days and stored protected from light at 4°C as a P1 stock. The P1 stocks for the integrin $\alpha_v\beta_3$ had been generated previously (197) and were only amplified.

The stocks were amplified to P2 and P3 by infecting suspension cultures of Sf9 cells at an estimated multiplicity of infection (MOI) of 0.1, assuming the titre of the P1 stock to be 5 x 10^6. The supernatants were harvested after 5 to 7 days by centrifuging at

1470 x g for 5 minutes, filtered with a 0.22 µm filter, and stored protected from light at 4°C.

2.2.12. Insect cell culture

Sf9 and High Five cells were acquired from ThermoFisher Scientific after attempts at transferring previously prepared stocks to serum-free suspension culture from adherent serum-supplemented culture failed. The new stocks were grown in the appropriate medium (Sf-900 II SFM or Express Five SFM, for Sf9 and High Five, respectively; both ThermoFisher Scientific) directly in suspension culture with shaking (110 rpm), without the addition of serum, with penicillin and streptomycin, at 20 °C. After an initial uninterrupted passage of 7 days, the cells were split every two to three days by diluting an aliquot into fresh medium for an initial seeding density of 3-5x10^5 cells/mL. The High Five cells were supplemented with heparin at a decreasing concentration for the first few passages to prevent cell clumping.

Initial attempts at adherent culture involved seeding the cells in appropriate flasks, maintained in a non-humidified incubator without CO_2 at 20 °C, treatment with trypsin and scraping off for seeding at confluence. Transfer to 6- or 24-well plates for expression tests or transfections was also performed this way, with the cells attaching after 30 minutes.

Cell stocks were prepared at early passages (P-5, P-7) by freezing cultures in 1 mL aliquots in medium supplemented with 10% dimethylsulfoxide (DMSO) in an insulated cryobox stored first at -20, then -80° C, and finally transferred to liquid nitrogen tanks for long-term storage. The stocks were brought up as described for newly acquired cells.

2.2.13. qPCR for baculovirus stock titration

The baculoQUANT kit for viral DNA extraction and qPCR (quantitative PCR; Oxford Expression Technologies, Oxford, UK) was used to establish stock titres according to manufacturer's protocols. Briefly, the DNA was extracted using chemical lysis and heat shock, then added to a master mix (Brilliant II qPCR Low ROX Master Mix) with baculovirus-specific primers for the gene *gp64*, fluorescent dye (6FAM) and quencher (IowaBlack FQ). The reaction was performed in duplicate or triplicate for each virus stock. The amplification was performed according to manufacturer's protocol, with a single activation step of 10 minutes at 95 °C, followed by 40 cycles of 15-second denaturation at 95 °C and 60-second annealing and extension at 60 °C. C_t values for

the reactions were compared to the standard curve generated using the virus sample supplied, on the basis of which the titres in viral particles/mL were established.

2.2.14. Baculovirus insect cell protein expression

Sf9 or High Five cells at a density of 2 x 10^6 cells/mL were infected with the P3 stock at an MOI of 2.5 or 5. The optimal timepoint for harvesting was established during expression tests by western blotting (see section 2.2.15) of samples taken at 24-hour intervals and the cells were harvested after 96 hours by centrifugation at 1490 x g for 5 minutes. The pellets were stored frozen at -80°C or lysed immediately.

2.2.15. Expression tests and western blotting

50 µL samples were collected from Sf9 and High Five cells at 24-hour intervals between 0 and 120 hours post infection of a standardised number of cells with penton base-coding virus, at MOIs ranging from 1 to 10, including a non-infected negative control. The cells were pelleted by centrifugation and either lysed directly by diluting with 4xSDS-PAGE sample buffer or by incubating the lysates or supernatants with Ni^{2+} beads from SpinTrap columns (GE Healthcare), which were washed with appropriate buffers, then mixed with the 4xSDS-PAGE sample buffer and loaded onto the gel instead of eluting. SDS-PAGE gels were run as described above (Section 2.2.3), then soaked in WB buffer (15 minutes), transferred with semi-dry transfer onto a PVDF (polyvinylidene difluoride) membrane pre-soaked in methanol (10 minutes), water (5 minutes), and WB buffer (10 minutes), sandwiched in water-moistened blotting paper. The transfer ran for 1 hour at 20 V. The membrane was then blocked in 5% w/v milk powder solution in TBS with 1% Tween20 (TBS-T) for an hour at room temperature with rocking, incubated with rocking at room temperature for 1 hour or overnight at 4°C, with a primary anti-His-tag mouse antibody diluted to 1:5000 or 1:10000 in a milk powder solution. The membrane was washed in four 5-minute intervals with TBS-T, with agitation, then incubated with 1:5000 secondary anti-mouse goat antibody for 1 hour and washed again. ELC Amersham Western Blotting Detection reagents (GE Healthcare) were applied to the blot (1.5 mL) to visualise the result. Imaging was performed with the ChemiDoc (BioRad, Hercules, USA) was used for visualising western blots. by exposing the membrane for 5 minutes or an automatically determined period on the chemiluminescence setting.

2.2.16. Penton base WT and mutant purification protocols

All steps were performed at 4°C. The pelleted Sf9 cells were resuspended in the appropriate lysis buffer (see section 2.1.5), lysed with a microfluidizer (2 bar, 600-800 psi) or a sonicator (Branson 250 Digital Sonifier Cell Disruptor, Marshall Scientific; 10% pulses with 1 second pulse and 1 second rest for 2 minutes total pulse time), then centrifuged for 45 minutes at 34500 x g. The supernatant was loaded onto a HisTrapFF column, pre-washed with water and IMAC wash buffer, by loading continuously for 12 to 16 hours at a speed of 1 mL/min. The column was mounted onto an ÄKTA Prime system, where it was washed to baseline with wash buffer, then with 5% elution buffer, then eluted with a gradient or stepwise up to 100% concentration of the elution buffer. Fractions were collected and analysed by SDS-PAGE. The ones with the highest and purest protein content were pooled and concentrated (50 kDa concentrator with a regenerated cellulose membrane, prewashed with water and SEC buffer, 10-minute concentration steps at 1470 x g) while exchanging buffer. After concentrating had proven to favour multimer formation (see section 3.2.3), this step was omitted, limiting the pool volume to 10 mL.

The pooled protein was filtered with a 0.22 µm filter, loaded onto a HiLoad Superdex 200 (SD200) 16/60 or Superdex 200 26/60 preparative grade column (Pharmacia, later Cytiva) installed on a Pharmacia or an ÄKTA Basic system, prewashed with the appropriate degassed SEC buffer, and eluted. Fractions were collected, concentrated as described above or not, analysed for multimer formation and homogeneity, flash-frozen in liquid nitrogen and stored at -80°C. Stability and multimer formation was examined by analytical SEC, and flash-freezing, regular freezing at -80°C, and slow freezing in a cryobox were compared.

2.2.17. Integrin $\alpha_v\beta_3$ purification

A 500 mL culture of Sf9 cells at a density of $2x10^6$ cells/mL was infected and harvested after 120 hours as described for the penton base constructs (section 2.2.14), except the supernatant was collected after centrifugation rather than the pellet, since the protein was secreted. The protocol had been published previously (165), though modifications were introduced. After centrifugation, the supernatant was not concentrated or filtered, as previously with 8 L culture volumes, but loaded directly onto a pre-washed 1 mL HisTrapFF column, continuously at 0.8 mL/min overnight. The IMAC was then performed as described for the penton bases (section 2.2.16), with elution by continuous gradient of elution buffer from 5 to 100% concentration. Fractions

were analysed by SDS-PAGE and those containing both integrin subunits were pooled, then dialysed overnight into SEC buffer in a 100 kDa cut-off membrane. Cleavage of the dimerization domain with human rhinovirus (HRV) 3C protease was performed overnight, adding 10 µg HRV3C per 100 µg integrin). After the 3C cleavage, trypsin digest (with 0.3 µg enzyme per 100 µg integrin) was performed for 30 minutes to remove non-degraded α subunit, then quenched with x10 trypsin inhibitor. The products were filtered with a 0.22 µm filter and loaded onto an SD200 16/60 or 26/60 column prewashed with the appropriate degassed SEC buffer, then eluted. The peaks identified by absorbance and SDS-PAGE were pooled and concentrated with a 100 kDa cut-off concentrator to 0.3-0.5 mg/mL, flash-frozen and stored at -80°C.

2.2.18. Analytical SEC

Analytical SEC was performed on the ÄKTA Ettan system (GE Healthcare) with the SD200 (up to 600 kDa resolution range), Superose6 (up to 5000 kDa), or Superose6 Increase columns (also up to 5000 kDa; all columns GE Healthcare), depending on the expected size of the protein and its multimeric assemblies. The higher molecular range columns were generally used for the penton bases and their complexes with the integrins. 50 µL of the sample (centrifuged or filtered to avoid aggregates) were loaded onto the column prewashed with the appropriate degassed buffer and eluted.

2.2.19. Complex formation

The penton base was dialysed in dialysis membranes with either 3 or 50 kDa molecular weight cut-off (depending on availability) into $\alpha_v\beta_3$ SEC buffer or cross-linking buffer for 1 hour at 4°C, with stirring. The two complex components were incubated with rocking for 1 hour or overnight at a 1:1, 1:5, or 1:10 penton base:integrin molar ratio. An aqueous solution of 1 M $MnCl_2$ was added to a final concentration of 2 mM for integrin activation.

2.2.20. Negative stain electron microscopy

1-2 µL of the sample (diluted with buffer at 1:3, 1:10, or 1:20 sample:buffer concentration) were spotted onto a glow-discharged continuous carbon film-coated grid and allowed to evaporate for 2 minutes. Next, three drops of water were added, as were three drops of 1% uranium acetate, with the grid allowed to sit for 5 minutes for evaporation each time, then blotted. The grids were imaged in a 120 kV Tecnai G2 Spirit BioTWIN transmission electron microscope with a TVIPS F416 camera (FEI,

Hillsboro, USA) at a magnification x70000 or a JEM-1400plus (Jeol, Tokyo, Japan) transmission electron microscope at a magnification of x75000.

2.2.21. CryoEM sample preparation

The protein complexes were incubated at 4 °C in the presence of 2 mM manganese ions at 1:1, 1:5, or 1:10 penton:integrin molar ratios for 1.5 hour or overnight, with rocking. For the largest dataset of HAdV-C5 penton base:integrin 1:5 complex, the complex was erroneously diluted 1:1 with distilled water rather than integrin buffer. Three µL of a sample were spotted (Leica) or 0.1 µL were sprayed onto a glow-discharged Quantifoil 1.2/1.3 holey carbon grid (Quantifoil Instruments, Jena, Germany) at room temperature with 90% humidity, blotted with filter paper for 1.5 seconds, then plunged in liquid ethane using an automated vitrification setup: a Leica vitrification robot (Wetzlar, Germany) or the VitroJet Cryosol (Weert, Netherlands). The grids were stored in liquid nitrogen until data collection.

2.2.22. CryoEM data collection and processing

Data collection was performed at the University of Helsinki CryoEM Unit, Finland. Movies were collected on a Talos Arctica microscope (ThermoFisher Scientific, Waltham, USA) with a Falcon III direct electron detector camera (ThermoFisher) at 200 kV, 30 or 40 frames per movie, electron dose 1.03 electrons per $Å^2$ per frame, defocus between -1.2 and 3 µm, varied in 0.3 µm increments, pixel size 0.97 Å, magnification 150000x. They were motion-corrected either with on-the-fly processing or MotionCor2 (208), implemented in Scipion 2 Diocletian (209), in which all further processing steps were performed as well. CTF correction was performed with CTFfind4 (210), and micrographs where the comparison between estimated and calculated CTF suggested excessive drift were discarded. Particles were picked using Xmipp 3 (211) trained with a sample of manually picked particles and curated micrographs, extracted using Xmipp 3 with a box size of 200 pixels. The particles were assigned to 2D classes using Relion 2 (212) and classes with clear averages were reclassified until stable. Particles from stable classes were used for a 3D classification with 10 classes in Relion 2, using the structure of penton base from HAdV type 5 (PDB ID: 3izo) low-pass filtered to 60 Å as the initial model. The 3D classification was repeated with particles constituting the largest classes re-classified into 10 further classes until the largest class was stable. The Relion autorefine protocol was then applied to the largest 3D class, with or without imposition of C5 symmetry. Published atomic models of the appropriate proteins were fitted into the densities in UCSF Chimera (213) and cross-correlation values for a

model of similar resolution compared. Resolution was estimated with the gold standard threshold of FSC 0.143 (214), as implemented in Relion 2. See Table 7 for other data collection and processing parameters.

2.2.23. Cross-linking of the penton:integrin complex

The integrin was buffer-exchanged to remove Tris, which would inhibit the cross-linking reaction. 100 µL of 0.3 mg/mL integrin were thawed from –80 °C and dialysed, in dialysis cassettes with a molecular weight cutoff of either 3 or 10 kDa (depending on availability), in 100 mL of the Cross-linking SEC buffer for an hour at 4 °C, with stirring. The 1:5 penton base:integrin complex was then formed overnight as described above (section 2.2.19 Complex formation).

The cross-linker DSS-H12/D12 (disuccinimidylsuberate, Creative Molecules Inc., San Francisco, USA) was dissolved in dimethylformamide (DMF) for a stock of 50 mM DSS. 1 µL of the stock was added to the complex to a final concentration of 0.5 mM and incubated for 1 hour at 800 rpm at room temperature. The reaction was then quenched with 2 M ammonium bicarbonate for 15 minutes at 800 rpm. The sample was then loaded onto a TSKgel G6000PWXL SEC column (Tosoh Bioscience, Tokyo, Japan; separation range 40 kDa - 6 MDa) and peak fractions analysed by SDS-PAGE, visualised by silver staining according to published protocol (215). Samples were selected for vitrification or negative stain (see sections 2.2.20 Negative stain electron microscopy or 2.2.21 cryoEM sample preparation).

Alternatively, the cross-linker BS3 (bissulfosuccinimidyl suberate; Molecular Bioscience, Boulder, USA), dissolved in water, was added to the complex at concentrations of 0.5, 1, 2.5 or 5 mM. The samples were incubated at 4 °C for an hour, then quenched with 1 M Tris and run on the Superose6 analytical SEC column. Samples were analysed by SDS-PAGE, silver stained with a SilverQuest kit (Invitrogen) according to manufacturer’s protocols, and selected for negative stain.

3. Results

3.1. Part 1 – *T. forsythia* sugar kinases MurK and K1058

3.1.1. MurK and K1058 purification

The course of purification was very similar for K1058 and MurK, without any major differences. The final preparative SEC step showed a major peak (Figure 11a), which corresponded to a 30 kDa band on the SDS-PAGE gel (Figure 11b), showing sufficient purity and yield for crystallisation trials.

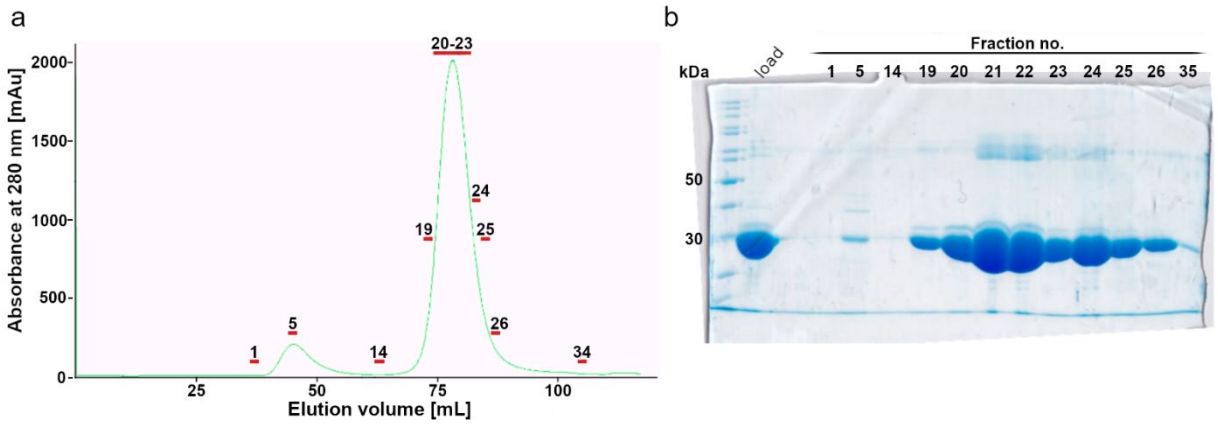


Figure 11. An example purification of K1058, closely resembling purification of MurK, produced a highly pure protein. a. Absorbance at 280 nm measured during the preparative SEC step of the purification of K1058. Fractions analysed in **(b)** marked in **red**. **b.** SDS-PAGE gel with samples from SEC of K1058. Expected M_w of a K1058 monomer is 33 kDa.

Analytical SEC on concentrated fractions from the major peak of the preparatory SEC (Figure 12) showed a single peak running at a volume corresponding to the molecular weight of 65 kDa, which is consistent with a dimer, (the protein monomer is 32.4 kDa for MurK and 33.0 kDa for K1058). The typical yields were 10 mg/mL for K1058 and 8.5-12.5 mg/mL for MurK.

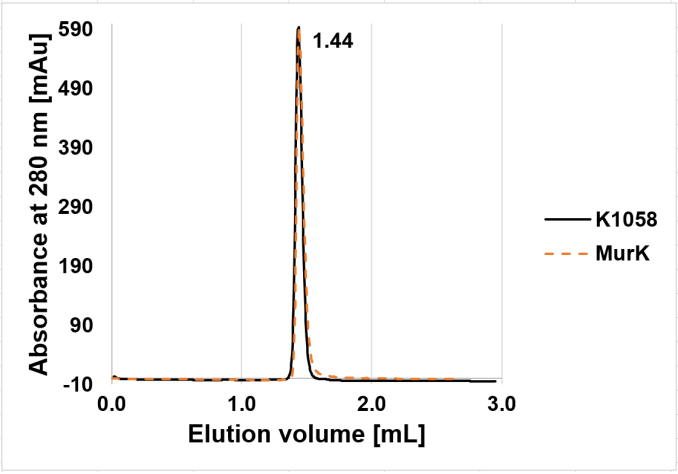


Figure 12. Analytical SEC of purified K1058 (black line) and MurK (orange dashes). The peaks at elution volume of 1.44 mL correspond to M_w of 65 kDa, which is consistent with a dimer for both proteins.

3.1.2. K1058/MurNAc and MurK/MurNAc crystallisation

A number of initial hits were identified from commercial screens (see: 2.1.1 Chemicals) and fine screens set up on their basis. These were: K1058/MurNAc/AMP-PCP in Morpheus condition 2-27/G3 (imidazole/MES pH 6.5, 0.1 M carboxylic acid mix, 12% v/v glycerol, 6% w/v polythelene glycol 4000; PEG4000), K1058/MurNAc/AMP-PCP in Morpheus condition 2-31/G7 (0.1 M carboxylic acids, 0.1 M sodium HEPES:MOPS pH 7.5, 12% v/v ethylene glycol 6% w/v PEG8000), a fine screen based on the original MurK crystallisation condition (0.1 M BIS-TRIS propane buffer pH 5.5 or 6, 14-18% v/w PEG3350, 0.08-0.18 M ammonium sulfate, 48). The crystals were either too small, did not diffract or could not be replicated for optimisation. Large (240 µm by 100 µm by 10 µm) crystals were identified in the MurK/MurNAc/AMP-PNP fine screen, in 0.1 M BIS-TRIS propane buffer pH 5.5, 16% v/w PEG3350, 0.16 M ammonium sulfate (Figure 13a), but did not diffract above 8 Å. Similar issues with diffraction and ice in loops during data collection were encountered with MurK/MurNAc/AMP-PCP in PEGIon condition II-30/G6, (0.2 M ammonium tartrate dibasic pH 7.0, 20% w/v polyethylene glycol 3350; Figure 13b).

Ultimately, the only crystals of K1058/MurNAc that diffracted to high resolution took about nine months to grow (monitored monthly) in the initial screen Crystal I/II H2, in condition H2 with 1.5 M ammonium sulfate, 0.1 M Tris pH 8.5, 12% v/v glycerol at 4°C and AMP-PCP. They were long, thin, and flat, growing to a length of about 150-200 µm by 20 µm by 10 µm (Figure 13c). Smaller crystals, diffracting to 4.5 Å, were present in the same condition at 20°C after five months. The single diffracting crystal for MurK/MurNAc, with AMP-PCP (see Figure 13d) took about six months to grow and was harvested from condition I-38/D2 of the initial Wizard III/IV screen (15% PEG10000, 100 mM trisodium citrate pH 5.5, 2% v/v 1,4 dioxane), with the dimensions 65 µm by 25 µm by 5 µm.

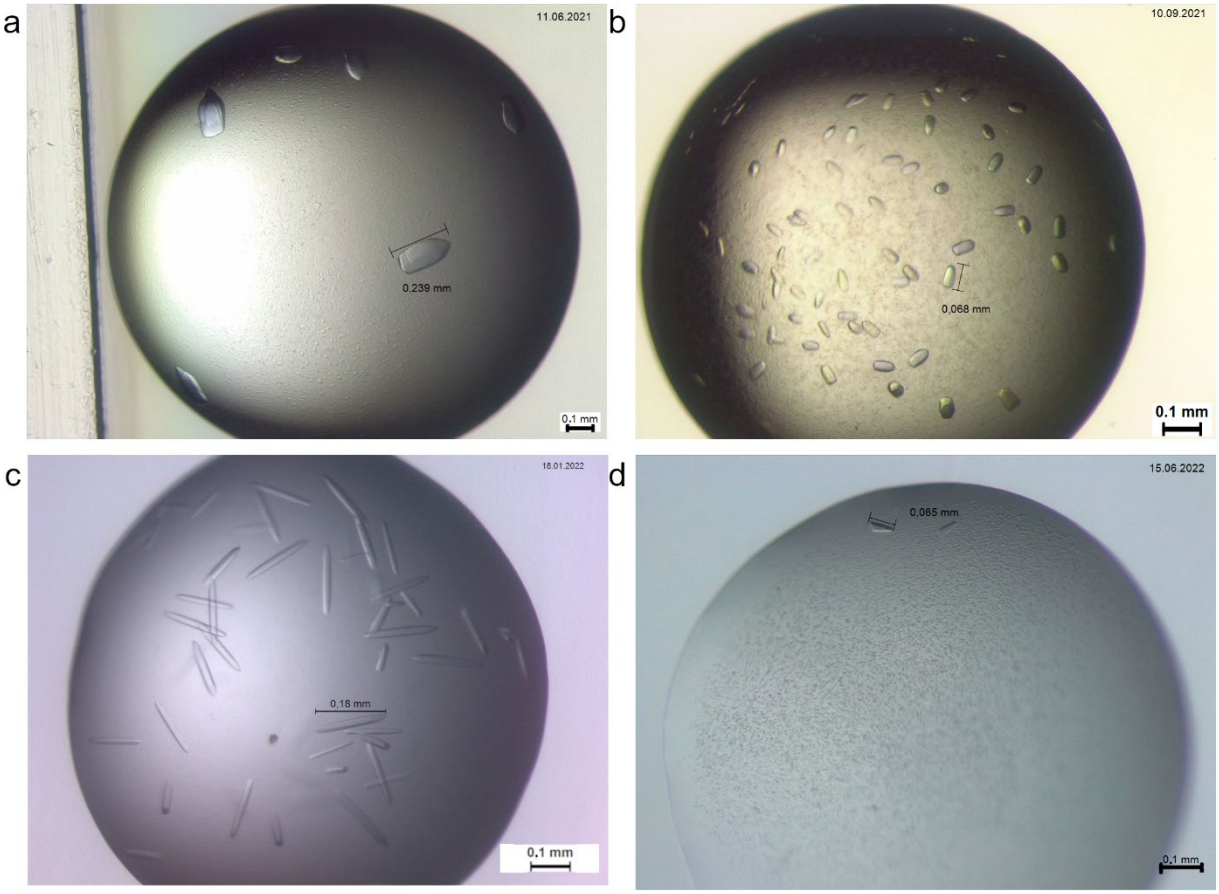


Figure 13. Representative crystals of the K1058/MurNAc and MurK/MurNAc complexes. **a.** MurK/MurNAc/AMP-PNP crystals growing within hours in 0.1 M BIS-TRIS propane buffer pH 5.5, 16% v/w PEG 3350, 0.16 M ammonium sulfate at 20 °C. Protein concentration 9 mg/mL. No diffraction above 8 Å was observed. **b.** MurK/MurNAc/AMP-PCP crystals growing after two days in 0.2 M ammonium tartrate dibasic pH 7.0, 20% w/v polyethylene glycol 3350 at 20 °C. Protein concentration 8 mg/mL. No diffraction was observed. **c.** Diffracting crystals (3-3.1 Å) of K1058/MurNAc with AMP-PCP, grown over nine months in 1.5 M ammonium sulfate, 0.1 M Tris pH 8.5, 12% v/v glycerol, AMP-PCP at 4°C. Protein concentration 9 mg/mL **d.** Diffracting crystal (2.7 Å) of K1068/MurNAc with AMP-PCP, grown over six months in 15% PEG10000, 100 mM trisodium citrate pH 5.5, 2% v/v 1,4 dioxane at 4 °C. Protein concentration 8 mg/mL. Scale bar in the bottom right corner in all panels is 0.1 mm.

3.1.3. **MurK and K1058 structure solution and refinement**

The MurK/MurNAc crystal diffracted to a resolution of 2.7 Å, but exhibited almost perfect merohedral twinning (twin fraction 0.46), which likely contributed to the electron density being of lower quality in two out of the six chains in the asymmetric unit (see Table 2). Nevertheless, the chains could be traced, except for the final amino acids of the C-terminal, and unambiguous density for the ligand was present in the four remaining chains. The dataset for K1058/MurNAc was merged from two crystals grown in the same condition, with a final resolution of 3.06 Å. Density for all six chains of the asymmetric unit was continuous until the final amino acids of the C-terminal and sugar density was visible in all chains. In both cases, no density for the ATP analogue was

present where it had been present in the MurK/AMP-PCP structure (48), despite its presence in the initial protein solution when the crystallisation trials were set up.

Table 2. Data collection and refinement statistics for the MurK/MurNAc and K1058/MurNAc structures. * denotes twinning in the MurK/MurNAc crystal. The K1058/MurNAc structure was obtained by combining two datasets. Values for the highest resolution shell given in parenthesis.

	MurK/MurNAc*	**K1058/MurNAc**
Data collection		
Space group	I4* (apparent space group: $I4_1$ 2 2)	P 3_1 2 1
Cell dimensions		
a, *b*, *c* (Å)	113.43, 113.43, 294.76	143.75, 143.75, 210.95
α, β, γ (°)	90, 90, 90	90, 90, 120
Resolution (Å)	49.13 - 2.7 (2.87-2.70)	29.85 - 3.06 (3.14-3.06)
No. unique reflections	49814 (8018)	48103 (3521)
Redundancy	14.1 (14.5)	23.6 (20.5)
R_{meas} *(%)*	40.8 (338)	45.6 (331)
$I / \sigma(I)$	8.1 (1.2)	8.1 (1.0)
Completeness (%)	98.5 (98.6)	99.9 (100)
$CC_{1/2}$	99.4 (46.6)	99.3 (34.8)
Wilson B-factor (Å^2)	65	62
Refinement		
R_{work} / R_{free}	0.248 / 0.277	0.235 / 0.259
B-factors (Å^2):		
Chain A/B/C/D/E/F	72/48/72/51/71/99	74/68/73/113/103/92
Ligand	66/40/-/48/61/-	62/57/58/105/78/76
R.m.s. deviations:		
Bond lengths (Å)	0.009	0.001
Bond angles (°)	1.85	0.366
Ramachandran:		
Favoured [%]	90.8	94.1
Outliers [%]	1.9	0.4

Both enzymes, as already shown in Karolin Gogler's work (48), consisted of two α/β domains, a smaller N-terminal (amino acids 1-103) and a larger C-terminal one (amino acids 104-284) and formed dimers (see Figure 14), which were confirmed to be biologically relevant by analytical SEC and PISA (Proteins, Interactions, Structures and Assemblies, 216) analysis. The elongated active site is located between the two domains, likely bringing the phosphate group of the phosphate donor and the hydroxyl group of the C6 carbon in proximity.

Since no density for AMP-PCP was present in MurNAc-containing structures, this could only be simulated by superimposing the AMP-PCP-liganded structure onto the MurNAc-liganded one. It was found that the distance between the reactive groups was suitable for the reaction to occur (3.4 Å). The substrates were bound in separate parts of the active site and did not overlap.

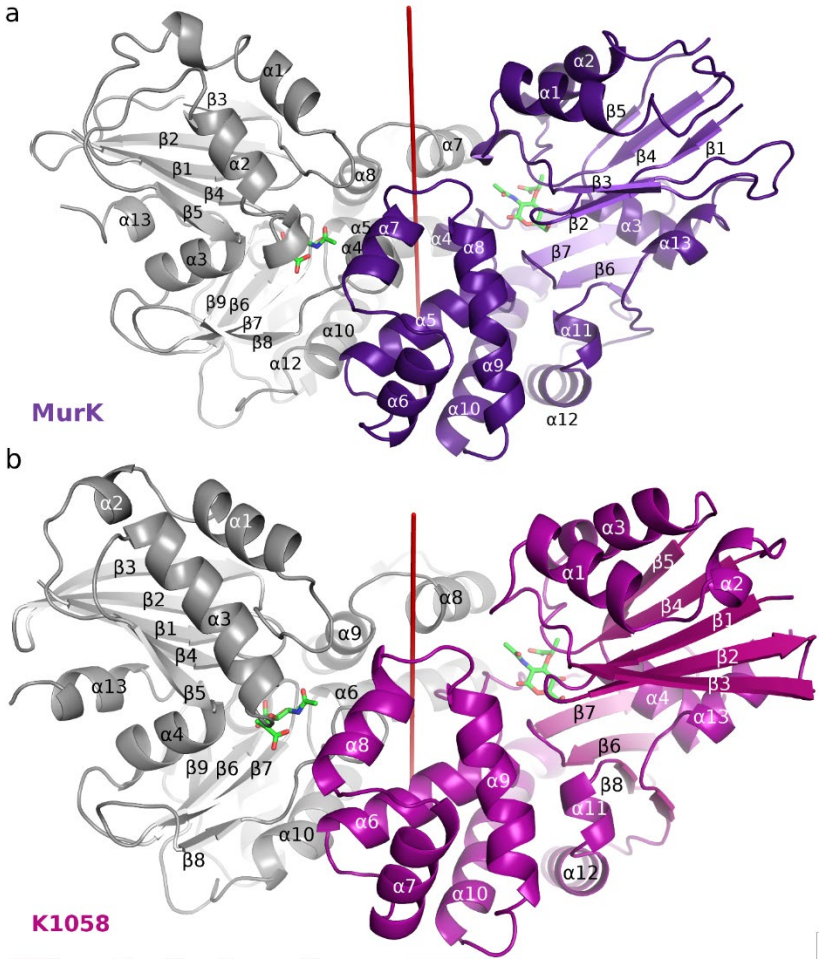


Figure 14. Structures of MurK and K1058 in complex with MurNAc. Overview of the MurK/MurNAc **(a)** and K1058/MurNAc **(b)** structures as dimers, shown as cartoon, with MurNAc as **green** sticks. One copy of the protein chain is shown in **purple** (MurK) or **magenta** (K1058), the other in grey. Red line represents the twofold symmetry axis of the dimer. Figures prepared in PyMol (139) and adapted from (217).

The MurNAc active site was similar in the two proteins (see Figure 15), with highly conserved aspartate residues at positions 95 and 141 (or 94 and 140, respectively, for K1058). Aspartate 95 formed two hydrogen bonds, with C4 and C6 hydroxyl groups of the sugar. Other hydrogen bonds were mediated by asparagine 119 (with C4 hydroxyl group) and the backbone amide of glycine 68 (with the oxygen of the ether bond). Since the resolution of either structure did not allow for the identification of water

molecules in the electron density, two water-mediated hydrogen bonds were postulated for the MurK active site based on waters in the higher-resolution MurK/AMP-PCP structure: serine 133 and an amide backbone interaction of leucine 135, both interacting with the carboxyl functional group on the sugar.

These four direct interactions were also present in the K1058 (see Figure 15b), where an additional hydrogen bond was formed by asparagine 32. An asparagine was also present at the corresponding position in MurK, however, the distance between it and the sugar was too large (3.6 Å) for an interaction. No water-mediated hydrogen bonds could be inferred based on the apo structure with waters for K1058.

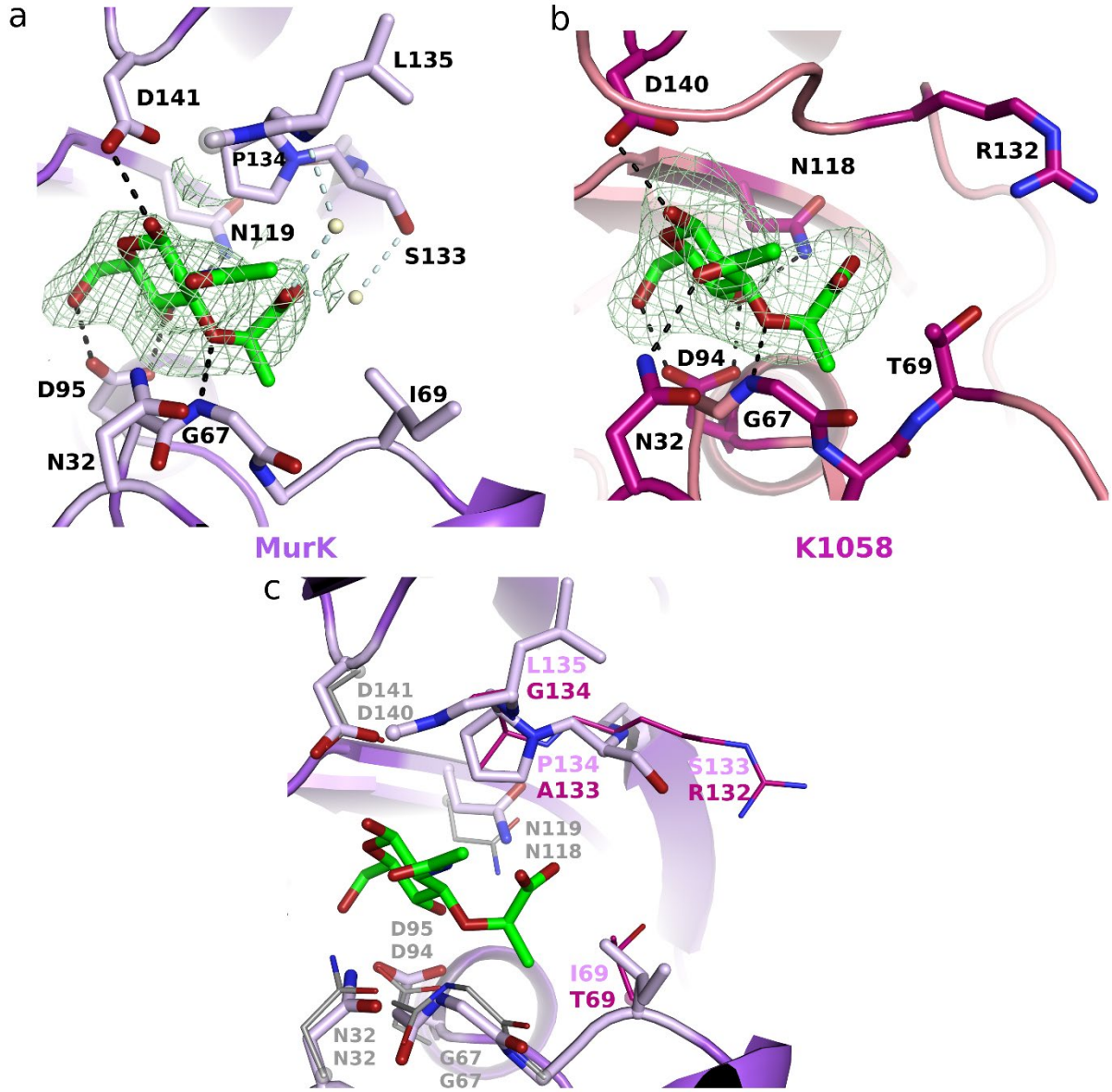


Figure 15. The MurNAc binding sites in MurK and K1058. Close-ups of the MurK **(a)** and K1058 **(b)** binding sites and MurNAc. Residues forming contacts are shown as sticks, predicted hydrogen bonds are shown as dashes. Hydrogen bonds and water molecules proposed based on the MurK/AMP-PCP structure are shown in light blue and ivory, respectively. Density for the ligands from the F_o-F_c difference maps, at 3σ level, is shown as mesh. **c.** Superposition of the active sites from **(a)** and **(b)**. Conserved residues forming hydrogen bonds between the two kinases are shown in grey. K1058 residues shown as lines. Residues that do not form hydrogen bonds and do not differ between the kinases shown as cartoon only. Helices α3, α4, α10, sheets β6, β7, β8, and loops β3α1, α10β9 (panel **a**), helices α1, α4, sheets β6, β7, β8, and loops β7β8, α10β9, α12α13 (panel **b**), helices α3, α4, α 13, sheets β6, β7, β9, loop β3α1, α3b6, b7b8, α10β9, α12α13 (panel **c**) have been partially or completely removed for clarity. Figure adapted from (217).

The consequences of the differences in the MurNAc binding sites between MurK and K1058 were largely unclear (Figure 15c). An isoleucine residue was located at position 69 in MurK instead of a threonine 69 in K1058, but both of these were too far for interactions with the ligand. A proline at position 134 in MurK is bulkier than the alanine at the corresponding position in K1058, but an arginine residue at position 132 of K1058 is bulkier compared to a serine 133 residue in MurK. In K1058, arginine 132 could be mediating additional contacts, however, it was pointing away from the sugar and the active site in this structure.

3.1.4. Movement of N-terminal and buried surface area changes of MurK and K1058 on MurNAc binding

On comparison of the apo and MurNAc-bound structures as well as modelling by the DynDom server (218), a large movement of the smaller N-terminal domain could be observed in both proteins. This was a rotation of over 12°, together with a movement of about 4-8 Å, depending on the residue, which brought the two domains closer together into a more closed conformation (see Figure 16). The movement was not observed on binding of the AMP-PCP to MurK, which remained largely the same as the apoenzyme (Figure 16c).

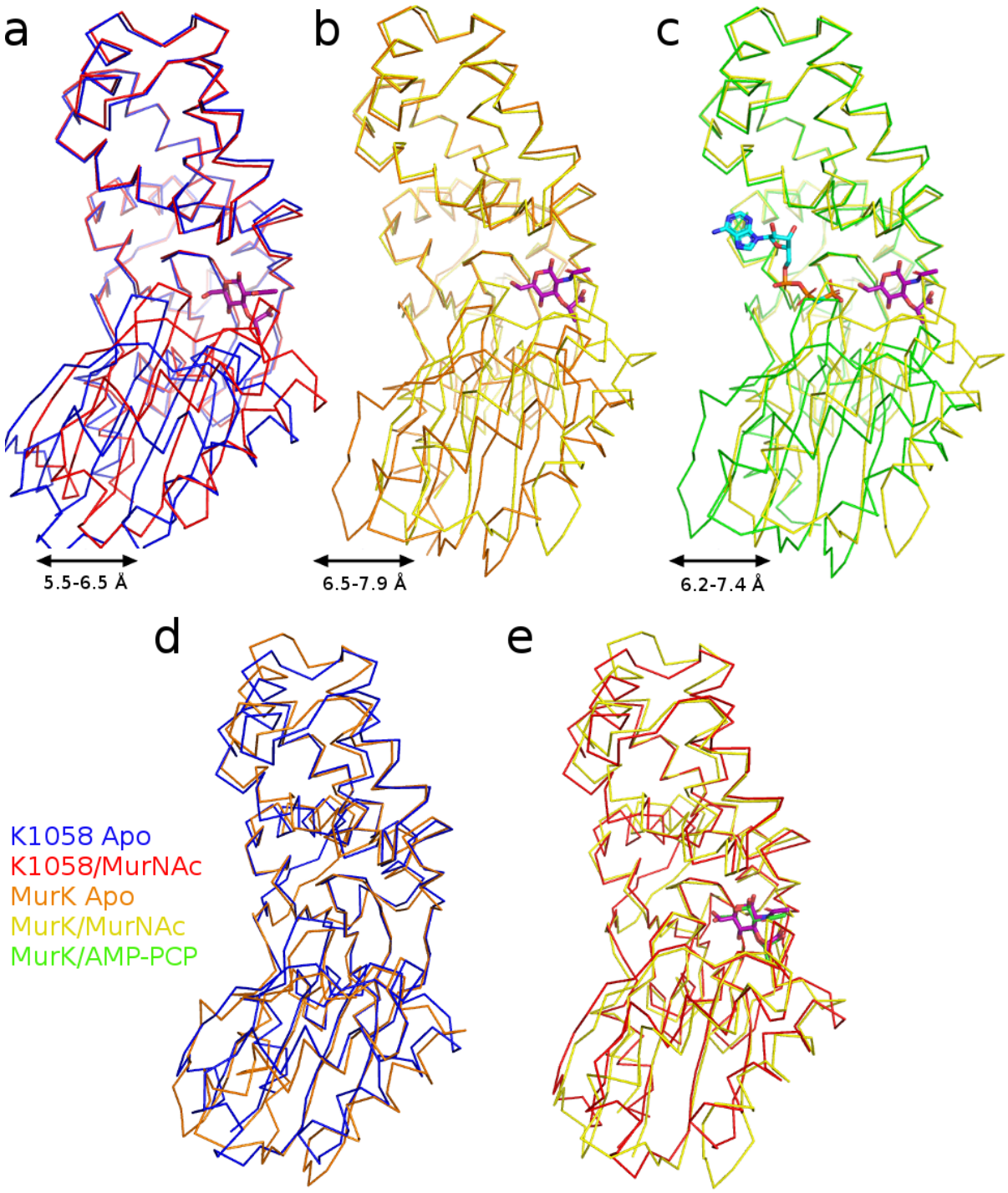


Figure 16. N-terminal domain movement on binding of MurNAc, but not AMP-PCP, in MurK and K1058. a. K1058 apo (blue) and K1058/MurNAc (red), superimposed on the C-terminal domain (residues 90-272) **b.** MurK apo (orange) and MurK/MurNAc (yellow), superimposed on the C-terminal domain (residues 91-273). **c.** MurK/AMP-PCP (green) and MurK/MurNAc (yellow), superimposed on the C-terminal domain (residues 91-273). **d.** MurK apostructure (orange) and K1058 apostructure (blue), superimposed on the C-terminal domains. **e.** MurK/MurNAc (yellow) and K1058/MurNAc (red), superimposed on the C-terminal domains. Figure adapted from (217).

PISA analysis (219) showed a change in the dimer interface formation on binding of the sugar, too, with the estimated buried surface area increasing from 1200 Å^2 to 1600 Å^2 (see **Table 3**). The N-terminal domain movement on MurNAc binding brings the N-terminal domains in the dimer closer together, resulting in stronger interactions between the interfacing residues (present especially in helices α4, α5, and α7 in MurK and corresponding helices α5, α6, and α8, as well as loop β8α5 in K1058) and in new interactions, particularly in loop α1β3 in MurK.

Table 3. Surface area of dimer interfaces of MurK and K1058 (all values in [Å^2]). **SAA**: Solvent-accessible area; **BSA**: Buried surface area. K1058 apo has a single chain in the asymmetric unit, which is predicted to form a symmetrical interface, so a single value is given. The highest-confidence interfaces were chosen for comparison.

	Apo		**With MurNAc**	
MurK	**SAA**	**BSA**	**SAA**	**BSA**
Chain 1	13390	**1163**	13290	**1590**
Chain 2	13600	**1155**	13520	**1610**
K1058				
Chain 1	13240	**1210**	12840	**1620**
Chain 2	-	-	13150	**1630**

3.1.5. Structural homologues of MurK and K1058

A Dali search (Distance matrix alignment, 220) identified structural homologues of the kinases. Some, including the two best matches (Z-score up to and above 35, root mean square deviation: 2.0: see Table 4), were targets of structural genomics projects and as such had not been functionally characterised. They did, however, originate in related (genus *Bacteroides*) bacteria or another member of the red complex (*P. gingivalis*), with the latter predicted to be a glucosamine kinase. Other high-scoring matches included sugar kinases from organisms such as humans, the thermophilic archaeon *Sulfolobus tokodaii*, or other pathogenic bacteria like *Haemophilus influenzae* or *Pasteurella brucellosis*. Non-kinase proteins included the MreB cell shape-determining protein and cell division protein FtsA.

Table 4. Best Dali (220) matches for MurK/MurNAc and K1058/MurNAc structures

Name	Organism	PDB Code	Z-score	r.m.s.d. [Å]	Similarity to
Conserved hypothetical protein Q8A1P1	*Bacteroides thetaiotaomicron*	1zxo	35.9	2.0	MurK/MurNAc
Putative N-acetylglucosamine kinase PG1100	*Porphyromonas gingivalis*	1zbs	34.9	2.1	MurK/MurNAc
Broad-specificity hexokinase	*Sulfolobus tokodaii*	2e2o	23.4	2.8	MurK/MurNAc
N-acetylglucosamine kinase (NAGK)	Human	2ch5	22.6	2.8	MurK/MurNAc
Probable N-acetylglucosamine kinase	*Chromobacterium violaceum*	1zc6	22.3	2.7	MurK/MurNAc
Cell shape-determining protein MreB	*Spiroplasma eriocheiris*	7e1g	17.2	2.7	MurK/MurNAc
N-acetylmannosamine kinase	*Haemophilus influenzae*	6jdb	15.8	3.3	MurK/MurNAc
Glucokinase	*Streptomyces griseus*	3vgl	15.6	3.7	MurK/MurNAc
Rod shape-determining protein MreB	*Caulobacter crescentus*	4czf	15.1	3.1	MurK/MurNAc
Conserved hypothetical protein Q8A1P1	*Bacteroides thetaiotaomicron*	1zxo	36.5	1.9	K1058/MurNAc
Putative N-acetylglucosamine kinase PG1100	*Porphyromonas gingivalis*	1zbs	36.4	2.0	K1058/MurNAc
Broad-specificity hexokinase	*Sulfolobus tokodaii*	2e2q	25.5	2.5	K1058/MurNAc
N-acetylglucosamine kinase (NAGK)	Human	2ch5	23.7	2.7	K1058/MurNAc
Probable N-acetylglucosamine kinase	*Chromobacterium violaceum*	1zc6	22.6	2.9	K1058/MurNAc
Cell shape-determining protein MreB	*Spiroplasma eriocheiris*	7e1g	17.7	2.6	K1058/MurNAc
Rod shape-determining protein MreB	*Caulobacter crescentus*	4cze	17.0	2.9	K1058/MurNAc
N-acetylmannosamine kinase	*Haemophilus influenzae*	6jdb	16.9	3.3	K1058/MurNAc
Cell division protein FtsA	*Thermotoga maritima*	1e4g	16.7	2.9	K1058/MurNAc

The percentage sequence similarity (Table 5) between the two solved structures was over 44% between MurK and K1058 and ranged from 16.5 to 24% for comparisons with the two closest functionally characterised matches: the *S. tokodaii* glucokinase (PDB ID: 2e2e) and human NagK kinase (PDB ID: 2ch5).

Table 5. Percentage sequence similarity of MurK and K1058 to the closest functionally characterised homologues

Protein	**MurK**	**K1058**	**Human NagK (PDB: 2ch5)**	***S. tokodaii* (PDB: 2e2o)**
MurK	100.00	44.2	17.5	22.2
K1058	44.2	100.0	16.7	17.8
Human NagK (PDB: 2ch5)	17.5	16.7	100.0	23.8
S. tokodaii (PDB: 2e2o)	22.2	17.8	23.8	100.0

Superposing the *S. tokodaii* glucokinase (PDB ID: 2e2o) and human NagK kinase (PDB ID: 2ch5) onto K1058 showed similarity in the overall fold (Figure 17a and c). This included the active site (Figure 17b,d), where the two aspartates (94 and the catalytic 140 in K1058) were conserved. Asparagine 32, which formed a hydrogen bond with the MurNAc in K1058, but did not at the homologous position in MurK, was also conserved. Asparagine 118 was conserved in the human, but not the archaeal enzyme. The archaeal kinase had an additional hydrogen bond formed by aspartate 71, and a glycine homologous to glycine 67 forming a hydrogen bond via the backbone nitrogen in MurK (Figure 18a).

This additional aspartate would likely clash with the N-acetyl group of the MurNAc, precluding its processing by the *S. tokodaii* glucokinase. No such clashes are present in the human NagK structure, and, indeed, the enzyme has recently been shown to process MurNAc dipeptide in addition to GlcNAc (221).

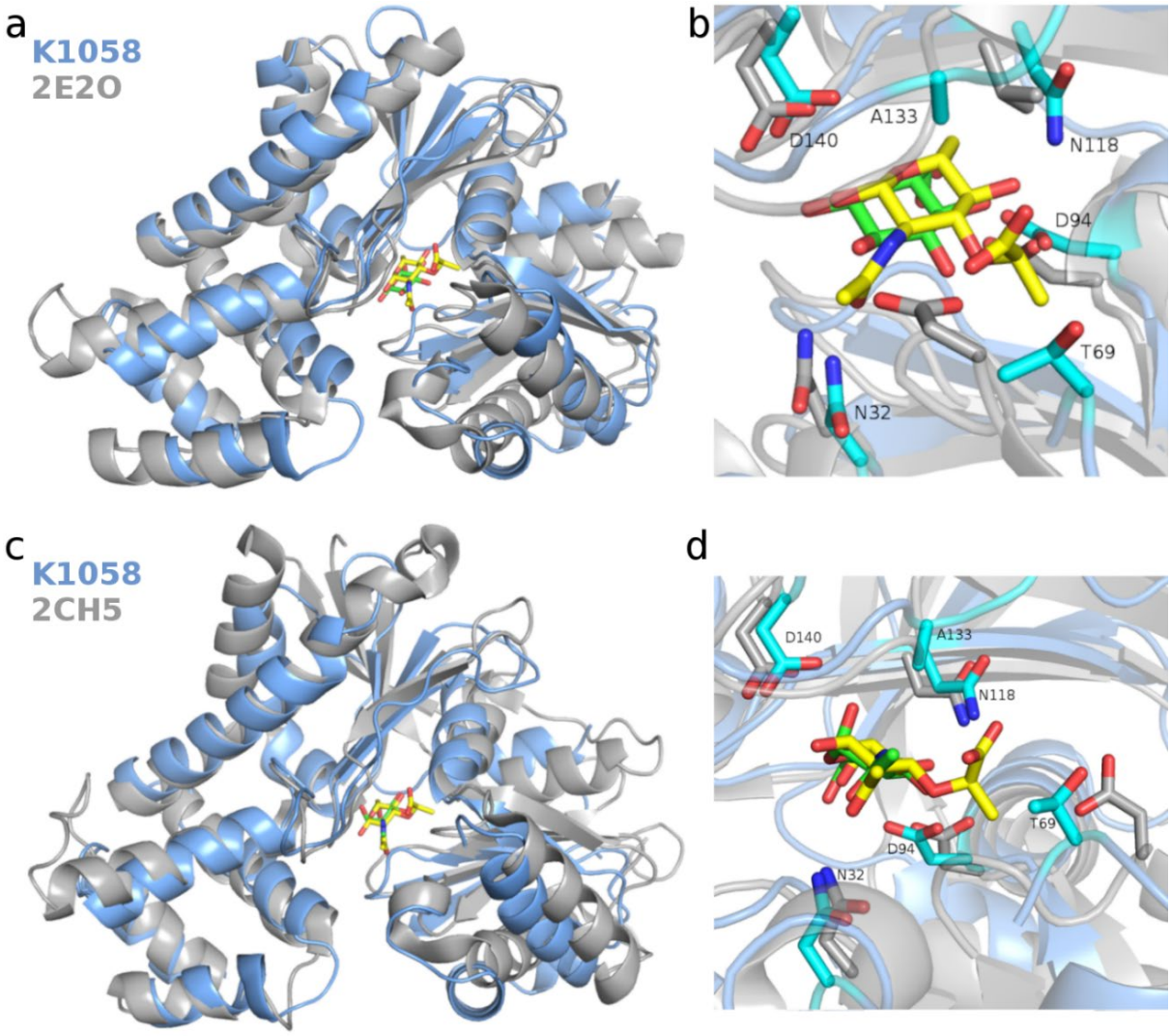


Figure 17. K1058 and its two closest functionally characterised structural homologues with substrates. a. K1058 (**blue**; MurNAc in **yellow**) and the broad-specificity hexokinase (**grey**; glucose in **green**) from S. tokodaii (PDB ID: 2e2o), superimposed on the C-terminal. **b.** Close-up of the active site from the superposition in **(a)** Labels refer to the K1058 residues. **c.** K1058 (**blue**; MurNAc in **yellow**) and the human NagK kinase (**grey**; GlcNAc in **green**; PDB ID: 2ch5), superimposed on the C-terminal. **d.** Close-up of the active site from the superposition in **(c)** Labels refer to the K1058 residues. Figure adapted from (217).

Comparison of binding sites to other structurally similar bacterial sugar kinases has suggested that specificity for MurNAc is mediated not by the presence of specific residues in the active site, but rather by their absence: glucokinases have an aspartate or glutamate residues in the active site, which would clash with the carboxylic acid group, as in the *S. griseus* glucokinase (Figure 18b) or, more rarely, a similarly clashing histidine in this position, as in *H. influenzae* mannosamine kinase (Figure 18c). The lack of bulky side chains at corresponding positions (serine 133, proline 134, leucine 135) lining the cavity allows for accommodation of MurNAc in MurK.

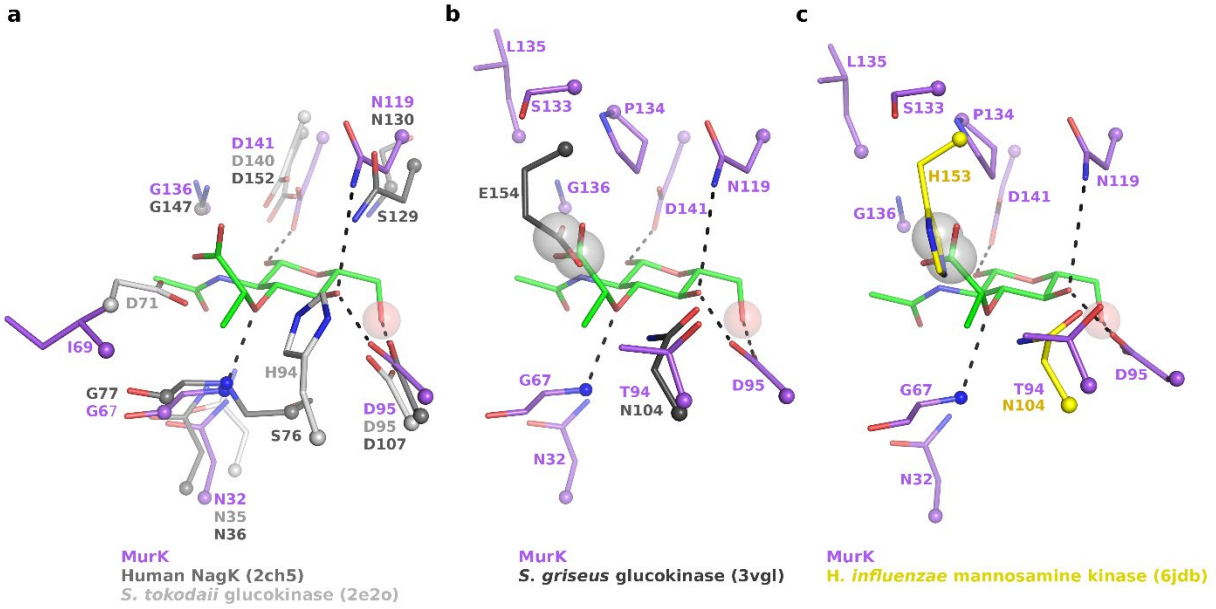


Figure 18. Superimposition of the active sites of MurK and other sugar kinases. a. the closest functionally characterised homologues of MurK (**purple**, MurNAc in **green**) and K1058, the human GlcNAc NagK kinase (**dark grey**) and *S. tokodaii* glucokinase (white). **b.** Active site residues mediating MurNAc binding or MurNAc specificity in MurK (**purple**) and those potentially clashing with this sugar in *S. griseus* glucokinase (**black**, PDB ID: 3vgl). **c.** Active site residues mediating MurNAc binding or MurNAc specificity in MurK (**purple**) and those potentially clashing with this sugar in *H. influenzae* mannosamine kinase (**yellow**, PDB ID: 6jdb). All superimpositions on the active site sugars only, only MurNAc from the MurK/MurNAc structure shown for clarity. Figure adapted from (217).

3.1.6. Coupled MurNAc phosphorylation assays

Each replicate reaction measurement showed a decrease in absorbance at 340 nm corresponding to the depletion of NADH, the slope initially increasing in steepness with time, and a plateau was reached at the higher MurNAc concentrations (Figure 19a). Given the reversible capacities of the LDH as an enzyme, this was assumed to correspond to the depletion of MurNAc, and not NADH, which would have been replenished. The highest slope values were not recorded at the very beginning of the experiment, as would be the case in direct assays measuring initial rates, but later. This period is known as lag time and is related to the enzyme reactions being coupled (205,206): at the beginning of the experiment, the intermediate substrates have not been produced in a sufficient quantity for the subsequent reactions to become saturated. While the lag time can be calculated by independently establishing the rates of the coupled reactions (206), in this case it was estimated to be about 150 seconds

by comparing the derivatives of the experimental curves to identify the time points for the fastest rate, which is also a published approach (205).

MurNAc assays showed a two-fold difference in MurNAc maximum phosphorylation velocity (V_{max}) between MurK and K1058 (80 and 0.34 µmol min^{-1} mg^{-1}, respectively). The K_M values were in the same order of magnitude (180 and 28 µM for MurK and K1058, respectively), with lower values corresponding to stronger binding in K1058, regardless of whether the model accounted for substrate inhibition (see Table 6). However, the k_{cat} and V_{max} values were much lower than for MurK, corresponding to the slow reaction speed observed during the measurements, which required using a hundred times more enzyme than in the case of MurK to obtain curves and reach substrate depletion in a comparable time scale.

The kinetic parameters were compared for models including or excluding substrate inhibition (see Figure 19b and c). For MurK, the curve fitted better in a model including substrate inhibition, while, in the case of K1058, the K_i error was larger than the K_i value, which, together with the poor fit of the line, suggested the classical Michaelis-Menten model was more applicable in this case. The influence of substrate inhibition in MurK was only evident at the highest (1-3 mM) MurNAc concentrations, as confirmed by the high K_i value (5.6 ± 2.3 µM).

Table 6. Kinetic parameters for MurK- and K1058-catalysed phosphorylation of MurNAc. * denotes the substrate-inhibited model.

Protein	K_m [µM]	V_{max} [µmol min^{-1} mg^{-1}]	k_{cat} [s^{-1}]	K_i [mM]	k_{cat}/K_m [s^{-1} M^{-1}]
MurK	110 ± 20	62 ± 2.7	32 ± 1.4	-	300000 ± 56000
MurK*	180 ± 40	80.0 ± 8.0	42 ± 4.2	5.6 ± 2.3	230000 ± 56000
K1058	28 ± 10	0.34 ± 0.02	0.18 ± 0.01	-	6400 ± 2900
K1058*	55 ± 22	0.38 ± 0.04	0.20 ± 0.02	20 ± 23	3600 ± 1500

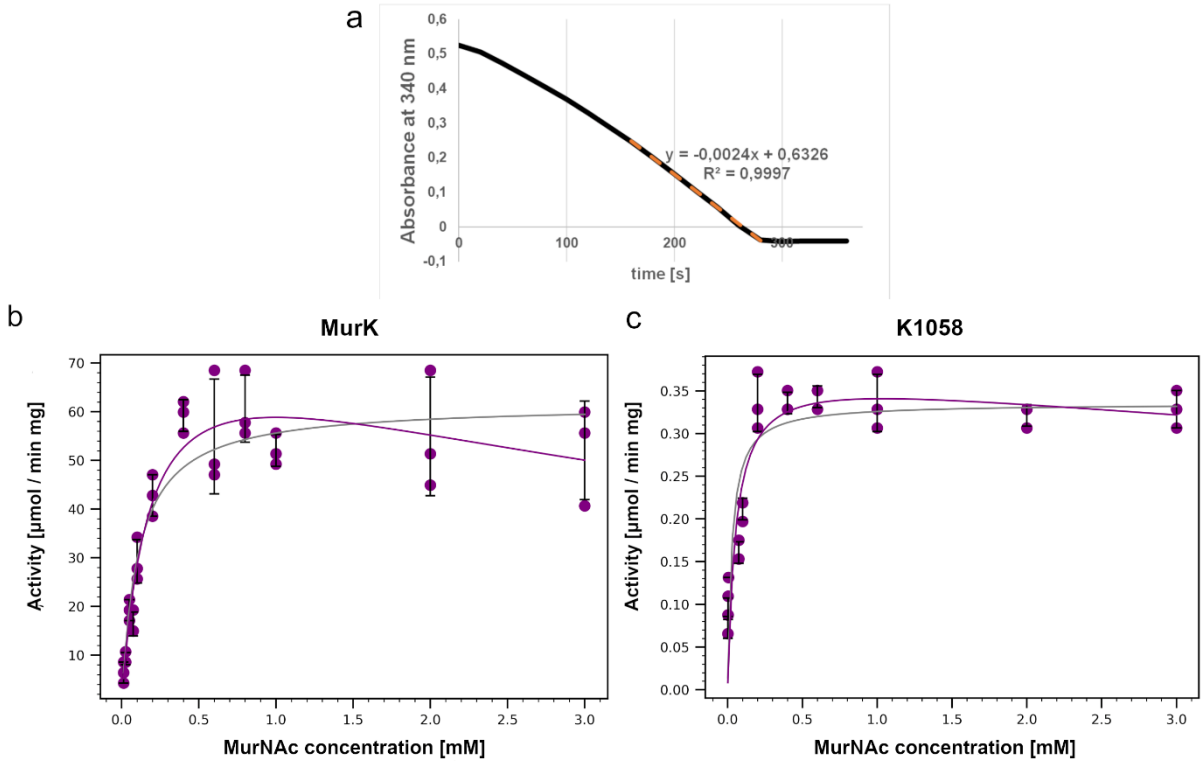


Figure 19. Results of coupled assays for MurNAc phosphorylation catalysed by MurK and K1058. a. An example raw data measurement for a MurK reaction with 1 mM MurNAc concentration. Absorbance at 340 nm, in **black**, and the trend line fitted to the fastest linear section, in **orange**. **b.** The Michaelis-Menten saturation curves for MurK-catalysed phosphorylation of MurNAc. **Grey**, a classical Michaelis-Menten saturation curve. **Purple**, a Michaelis-Menten saturation curve accounting for substrate inhibition. **c.** The Michaelis-Menten saturation curves for K1058-catalysed phosphorylation of MurNAc, as in **(b)**. **Grey**, a classical Michaelis-Menten saturation curve. **Purple**, a Michaelis-Menten saturation curve accounting for substrate inhibition. Each data point represents the slope of the trendline from a measurement as in **a.**, converted to µmol/min/mg from Abs/second. Error bars represent one standard deviation value. Figure provided by Dr Georg Zocher.

3.1.7. Purification of MurK and K1058 mutants

The six binding site mutants (see 2.2.9) were purified successfully according to protocols for the wild type proteins, with some variations in yield but similar purity. CD measurements indicated folded α-helical and β-sheet structures, with local ellipticity minima at 210 and 220 nm (see Figure 20a). Stability was examined by analytical SEC after three freeze-thaw cycles and showed no influence of the freeze-thawing cycles on the homogeneity or height of the peak (Figure 20b) in all six cases. Preliminary mass spectrometry experiments were performed in the Mayer lab to compare the enzyme specificities after mutations. However, they were abandoned, given that K1058 did not show the expected preference for GlcNAc over MurNAc in the MS or preliminary kinetic assays. In fact, its ability to process GlcNAc seemed limited, and the only data confirming this was a 16-hour endpoint radioactivity assay, which did not provide information on kinetics. Overall, this suggested the specificity-switch assumptions (from MurNAc-exclusive MurK to MurNAc and GlcNAc-processing K1058), on which the active site residue-swapping mutant design had been based, were false and this line of investigation was abandoned.

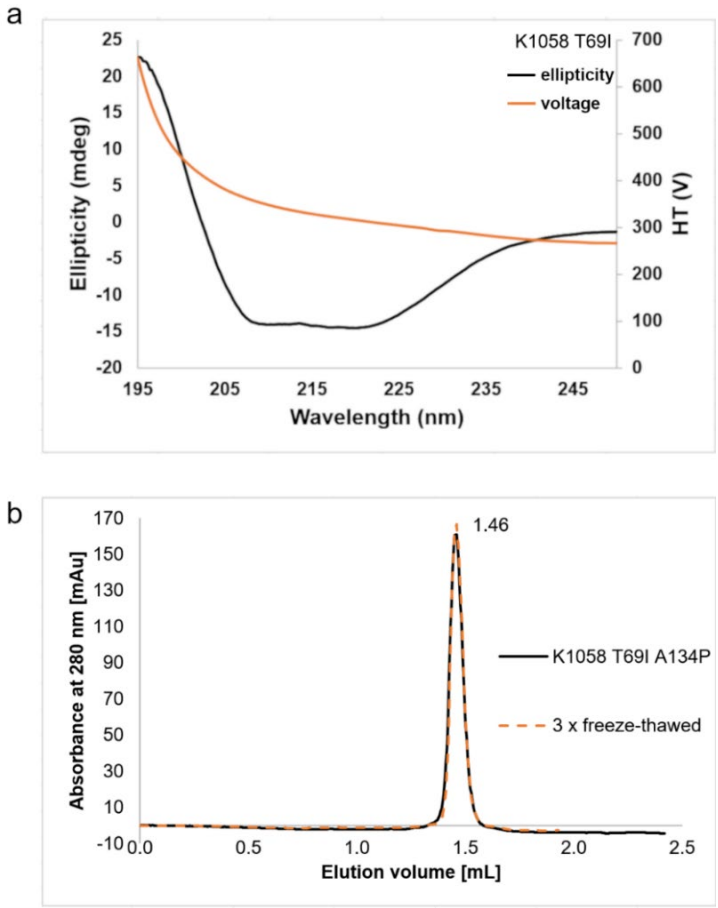


Figure 20. Example characterisation of purified chimeric active site mutants. **a.** CD spectrum of K1058 T69I mutant, showing ellipticity peaks at 210 and 220 nm wavelength, corresponding to folded protein with 79% α-helical and 21% β-sheet content according to analysis with BeStSel (222). **b.** Superose200Inc Analytical SEC of K1058 T69I A134P after purification and concentration (**black line**) and after three cycles of flash-freezing in liquid nitrogen and thawing (orange dashes), showing no change in the elution profile. Elution volume 1.46 mL corresponds to a M_w of 60 kDa on this column, about the expected weight for a dimer (65 kDa).

3.2. Part 2 – Studies of HAdV penton base in complex with the integrin entry receptor

3.2.1. Expression, purification, and characterisation of wild type HAdV-C5 penton base

Western blotting to verify and optimise expression of the penton bases showed that the proteins were expressing, with signal both in Sf9 and High Five cells (Figure 21). The signal for the His-tagged HAdV-C5 penton base was stronger in Sf9 lysates than in High Five lysates (Figure 21b), so the Sf9 cells were used for all further expressions of the HAdV penton bases. The optimal MOI and timepoint for harvest were identified as 2.5 virus particles/cell and 96 hours post infection, respectively, as they corresponded to the most prominent bands. MOI of 2.5 showed signal that was similar in strength to that from MOI of 5, so the former was selected as a more efficient use of the baculovirus stocks.

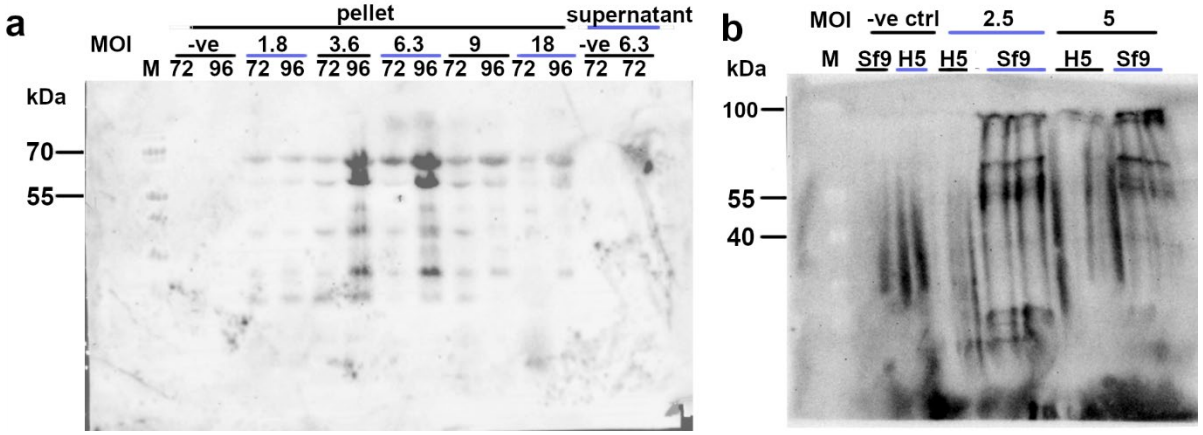


Figure 21. Expression of HAdV-C5 penton base WT in High Five and Sf9 cell lines under different conditions. a,b. Western blots of HAdV-C5 penton base WT in lysates and supernatant from High Five (H5) cells infected with different MOIs (**a**), and lysates of High Five and Sf9 cells infected with different MOIs (**b**), with an anti-His-tag primary antibody, confirming expression and showing stronger signal in Sf9 cells compared to High Five cells. The expected M_w for the uncleaved HADV-C5 penton base wild type is 65.8 kDa. 72 and 96 refer to hpi; M – marker.

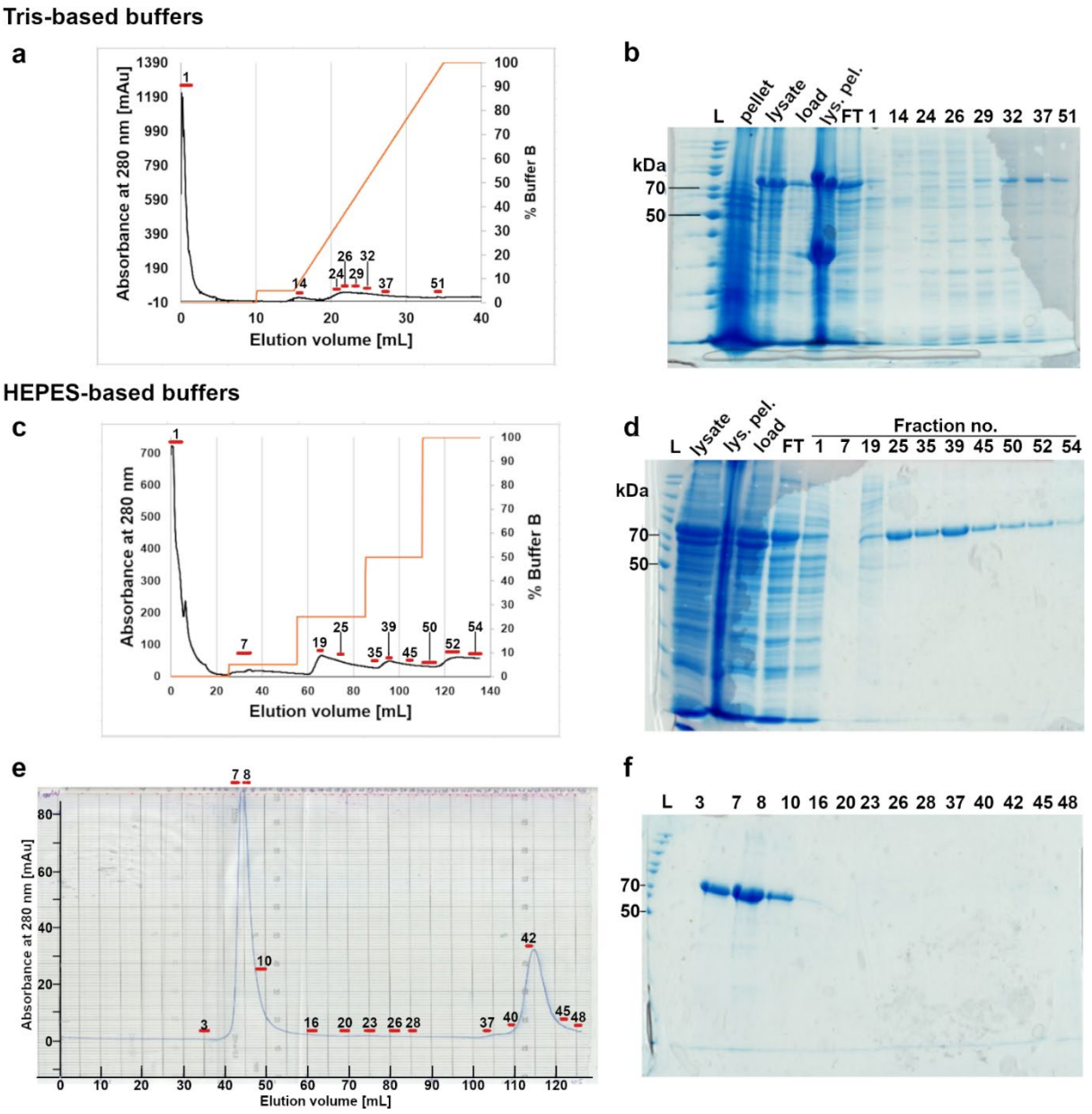


Figure 22. Purification of HAdV-C5 penton base WT in different buffers. a. IMAC step of the purification of HAdV-C5 penton base WT in Tris buffers, showing a small impurities peak and a small peak at 50% elution buffer. **b.** A large amount of the penton base remains in the pellet fraction after lysis and centrifugation, and little of it elutes, with a lot of impurities, as shown on the SDS-PAGE of samples from the lysis and IMAC steps of HAdV-C5 penton base WT purification in Tris buffers. **c.** IMAC step of the purification of HAdV-C5 penton base WT in HEPES buffers. **d.** SDS-PAGE of samples from the lysis and IMAC steps of HAdV-C5 penton base WT purification in HEPES buffers, showing a clear band in the fractions corresponding to the peak. The protein elutes in the void volume of the column, which indicates the M_w of aggregates or higher-order oligomers. **e.** Preparative SEC step of the purification of HAdV-C5 penton base WT in HEPES buffers, showing a single protein peak in the void volume and a smaller one corresponding to the DTT in the buffer. **f.** SDS-PAGE of fractions from the preparative SEC in **(e)**, showing protein of the right M_w is present only in the void volume peak. Expected M_w for the HAdV-C5 penton base WT is 66 kDa. Chromatography fractions analysed by SDS-PAGE marked in **red**.

Extensive attempts were made to find the appropriate purification conditions for the HAdV penton base (Figure 22). Initial attempts with the wild type constructs (focusing on penton base from HAdV type 5) used a number of buffers in search for one that would solubilise the protein. These were based on buffers (section 2.1.5) previously used in penton base purifications in literature and other communications: the HAdV-C2 penton base crystal structure (Tris-based; Zubieta buffer); the type C5, F40, andF 41 HAdV penton bases wild type (used in the Arnberg group purification, 138, personal communication), adjusted by substituting the phosphate buffer, unsuitable for structural studies, with Tris); and the HEPES-based buffer used by M. Liaci (198) for bacterial purifications of the penton base.

The HEPES-based buffer resulted in the best solubility and yields of the wild type protein (see Figure 22c and d). However, the resulting protein eluted completely in the void volume of the preparative SEC (Figure 22e) and analytical SEC (Figure 23a), despite running at the expected molecular weight on the SDS-PAGE gels (Figure 22f), suggesting aggregate or multimer formation.

Negative stain EM showed large, somewhat regular assemblies, which didn't fit perfectly to dodecahedra in shape, with a preference for assemblies of between five and seven pentamers (see Figure 23b), though the size fitted well (28 nm diameter for HAdV-C2 penton base, the monomer of which has the M_w 58 kDa, 151). Given the prevalence and size of these multimers in the sample, it was decided to proceed with purifications of mutants designed to prevent the formation of dodecahedra based on the crystal structure of the penton base dodecahedra of HAdV-C2 (198) and HAdV-B3 (160).

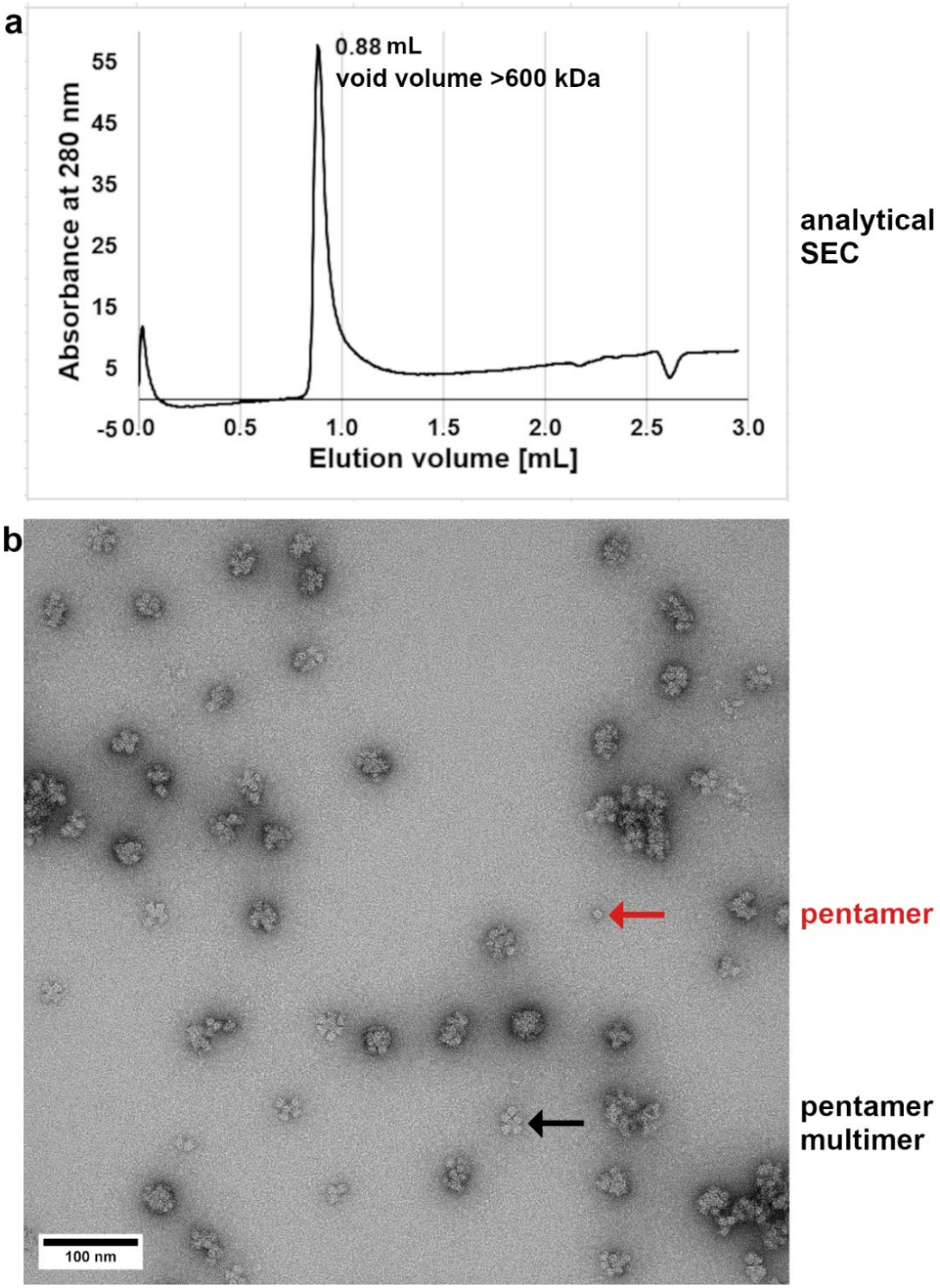


Figure 23. Characterisation of purified HAdV-C5 penton base WT. a. Analytical SEC (SD200Increase) of purified and concentrated HAdV-C5 penton base WT, showing a single peak in the void volume, above the 600 kDa maximum resolution range of the column. **b.** Negative stain images of purified and concentrated HAdV-C5 penton base WT, showing they are somewhat regular assemblies of pentameric penton bases, of a similar size to previously observed dodecahedra (expected size: 28 nm), though not as ordered and with a preference for assemblies of five-seven pentamers. An example assembly of six pentamers (**black** arrow) and a single pentamer (**red** arrow) have been marked on the micrograph.

3.2.2. Expression, purification, and characterisation of the non-dodecahedral mutant of the HAdV-C5 penton base

The mutant form solubilised and purified well in the IMAC step in the HEPES-based buffer (Figure 24a,b). However, in the preparative SEC, two peaks were visible beyond void volume (Figure 24c). This was the case even when only the later half of the lower M_w peak was collected and pooled: the two-peak pattern would be present in the analytical SEC, too (Figure 24d). This suggested the existence of an equilibrium in the purified sample. The higher M_w-peak ran at 1.03 mL, roughly corresponding to 930 kDa (the highest weight in the available calibration standards is 600 kDa, hence values above this are estimates), which could match a pentamer triplet, a smaller assembly than the full dodecahedron, but still a higher oligomer than a single pentamer. The second, lower-M_w peak, could likely be the pentamer, with the elution volume of 1.19 mL corresponding to the M_w of 390 kDa, as compared to the expected M_w of 300 kDa (expected M_w of the mutant penton base monomer was 59.2 kDa).

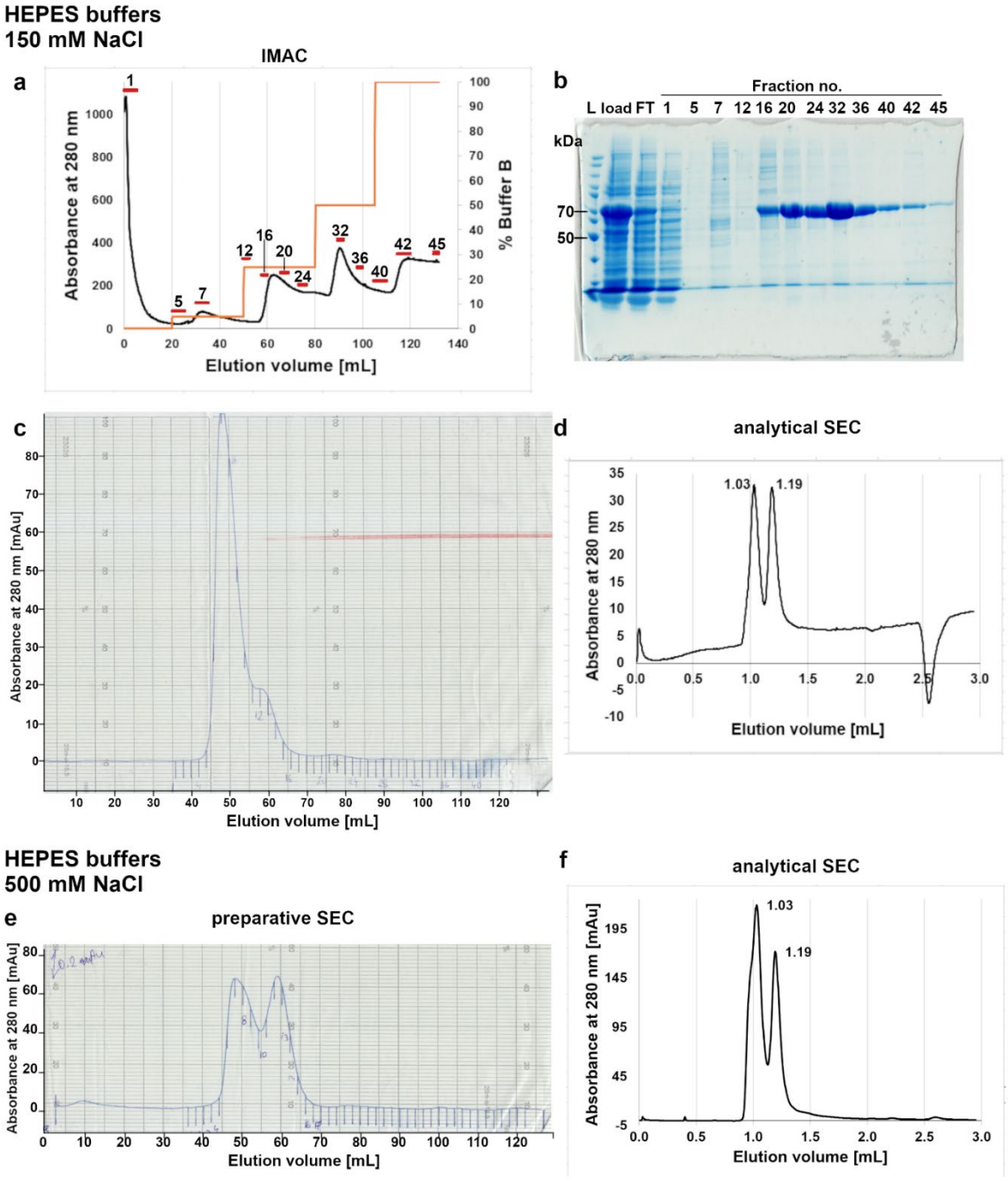


Figure 24. Purification of the dodecahedra non-forming mutant of HAdV-C5 penton base and multimer formation in different salt concentrations. a. IMAC step of the purification of HAdV-C5 penton base mutant in HEPES buffers, with peaks at about 25% and 50% elution buffer. **b.** SDS-PAGE of samples from the lysis and IMAC steps of HAdV-C5 penton base mutant purification in HEPES buffers. **c.** Preparative SEC step of the of HAdV-C5 penton base mutant in low salt HEPES buffer shows two peaks. **d.** Analytical SEC (SD200 Increase) of purified and concentrated HAdV-C5 penton base mutant in low salt (150 mM NaCl) HEPES buffer. One peak (1.19 mL, 390 kDa) roughly corresponds to the M_w of the pentamer, the other to a triple pentamer (1.03 mL, 930 kDa). **e.** Preparative SEC of HAdV-C5 penton base mutant in high salt (500 mM) HEPES buffer **f.** Analytical SEC (SD200 Increase) of purified and concentrated second half of the lower-M_w peak HAdV-C5 penton base mutant in high salt The two peaks are as in **(d).** The expected M_w of HAdV-C5 mutant penton base is 59 kDa.

3.2.3. Control of multimer formation in the non-dodecahedral mutant of the HAdV-C5 penton base

Attempts were made to elucidate the mechanisms governing multimer formation and equilibrium, as SEC experiments showed a varying relationship between the peaks, with the higher-M_w peak larger or smaller than the desired lower-M_w peak. Modifying the HEPES-based buffers to contain high (500 mM) rather than low (150 mM) NaCl resulted in a more favourable ratio, making it the buffer of choice (Figure 24e). However, concentrating the lower peak after preparative SEC again produced two peaks (Figure 24f). Concentrating the sample in even higher (1M) NaCl buffers failed to improve the ratio.

Moreover, it was observed that omitting concentrating the sample after the IMAC step resulted in the most favourable ratio of the peaks seen (Figure 25a,b). This, however, meant that the volume of material for preparative SEC was reduced to the maximum loading volume of 10-12 mL for a SD200 26/60 column. This, together with the inability to concentrate the fractions after the SEC step, led to the final concentration of the protein not exceeding 0.1-0.5 mg/mL (compared to 10-15 mg/mL in concentrated samples). Since not all material from penton base-containing IMAC fractions could be pooled for SEC, the IMAC peaks were compared to identify ones with the best higher:lower M_w peak ratio (best singlet pentamer content).

Elution with a gradient of 0-100% IMAC elution buffer (10-500 mM imidazole) resulted in three separate peaks, which could be better separated with a step elution (steps of 25%, 50%, and 100% elution buffer). The contents of the two largest (25 and 50%) peaks (Figure 25c) were compared using analytical SEC (the 100% peak was the smallest in terms of absorbance and the bands showed the least intensity in the SDS-PAGE gels, hence was left out). The 25% peak consisted of a wider range of species, including a broad 1.27 mL peak (252 kDa; Figure 25d), which could correspond to a range of complete and incomplete pentamers. However, concentrating this peak produced higher multiple-pentamer M_w species (data not shown), precluding the use of this material in further purification steps. The 50% peak was the one with the best presumed triplet to singlet pentameric pentamer ratio (Figure 25e) and it was pooled accordingly in future purifications. These differences were impossible to identify from denaturing SDS-PAGE gels, where all fractions contained a strong band at the same M_w of about 70 kDa, roughly corresponding to a monomeric penton base.

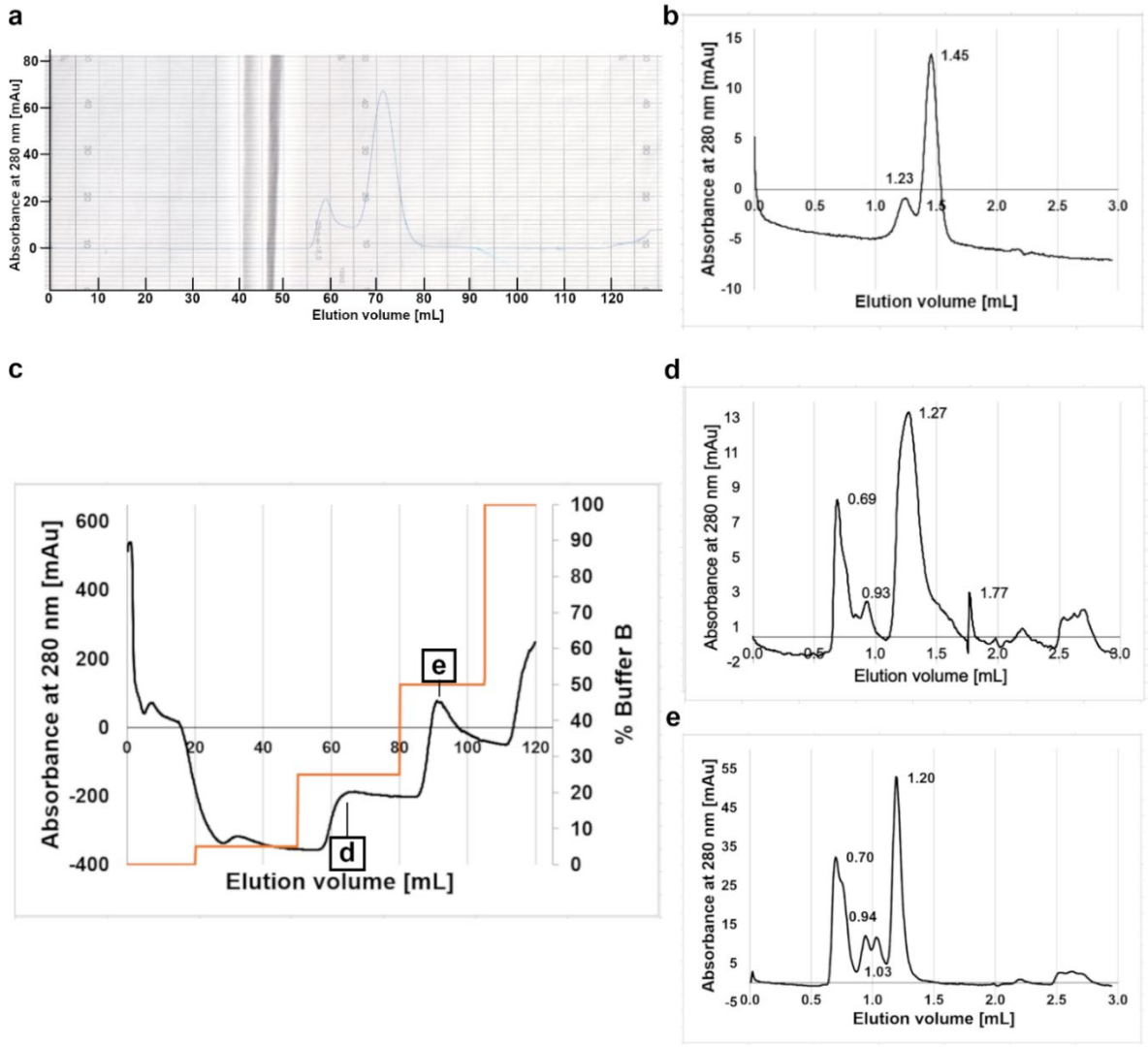


Figure 25. Purification and optimisation of the pentamer singlet to pentamer triplet ration in the HAdV-C5 penton base dodecahedra non-forming mutant a. Preparative SEC of HAdV-C5 penton base mutant in high salt (500 mM NaCl) HEPES buffer without concentrating the sample. **b.** Analytical SEC (Superose6) of HAdV-C5 penton base dodecahedra non-forming mutant in high salt HEPES buffer without concentrating the sample. 1.23 mL estimated to be 2120 kDa, 1.46 mL 418 kDa; likely corresponding to the SD200Increase 1.03- and 1.19-mL peaks, respectively **c.** An IMAC purification of the HAdV-C5 penton base dodecahedra non-forming mutant. Fractions analysed in **(d)** and **(e)** are indicated. **d.** Analytical SEC (SD200 Increase) of the 25% elution buffer peak from **(c)**, with the largest, broad peak (1.27 mL; 252 kDa) eluting at a lower molecular weight than expected for the pentamer, though extending to higher weights, too. **e.** Analytical SEC (SD200 Increase) for the 50% elution buffer peak from **(c)**, with the largest peak (1.20 mL, 370 kDa) corresponding to the pentamer and smaller peaks corresponding to larger multimers (1.03 mL, 931 kDa). The void volume of the column is 0.88 mL, so the earlier peaks (0.70 mL) are impurities from previous runs.

To check if it would be possible to concentrate the protein between the IMAC and SEC steps in this way, an additional affinity step of ion exchange was applied. IEX had been used previously in penton base purifications (151), however, in this case, it did not improve the pentamer content, and it was abandoned (data not shown).

To verify if storage or freezing conditions could affect the formation of higher-order oligomers, analytical SEC was conducted on differently treated samples (Figure 26). Prolonged (7 days) storage at 4°C did not lead to degradation, but rather formation of additional higher-order oligomers (Figure 26a), and freezing itself did not affect the peak multimer ratio (Figure 26b). Three methods of freezing tested also showed no difference (data not shown). Hence, all samples were flash-frozen and stored at -80°C before use.

Importantly, no method to break up the higher-order oligomers once formed has been identified, nor any conditions in which concentrating would not lead to formation of oligomers. Increasing of salt concentration to 1 M NaCl, reducing pH to 5.5 in succinic acid-based buffer identified from thermal shift assays for either concentration or the preparative SEC step, and a variety of reducing agents, were tested, among others. Ultimately, the protocol with 500 mM salt and no concentrating was accepted for HAdV-C5 penton base mutant.

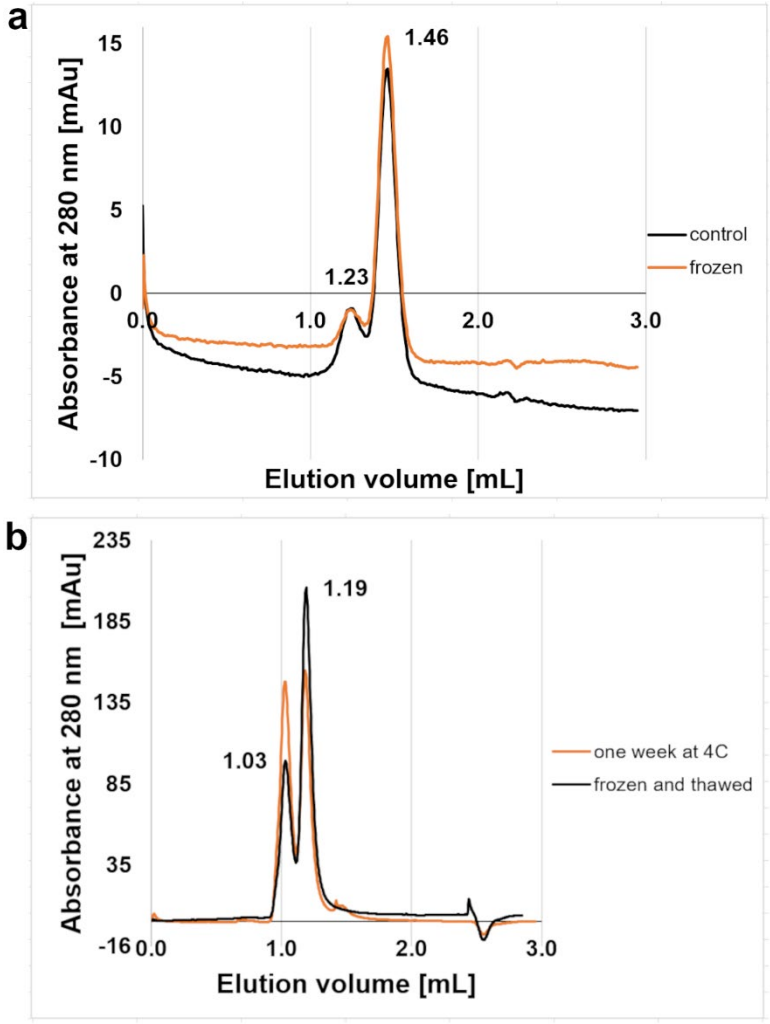


Figure 26. Influence of freezing and long-term storage of HADV-C5 penton base dodecahedra non-forming mutant on its oligomeric state. a. Analytical SEC (Superose6) on purified HAdV-C5 penton base mutant frozen in liquid nitrogen (1.23 mL: 2120 kDa, 1.46 mL: 418 kDa; likely corresponding to the SD200Increase peaks, as the estimates for Superose6 are unreliable above 600 kDa). **b.** Analytical SEC (SD200Increase) of the purified and concentrated HAdV-C5 penton base mutant stored at 4°C for a week, compared to a sample frozen immediately after purification, showing the increase of pentamer triplets with time (1.03 mL: 931 kDa, 1.19 mL: 390 kDa).

3.2.4. Expression and purification of the integrin $\alpha_v\beta_3$ according to an optimised protocol

Purification of $\alpha_v\beta_3$ was re-started and optimised from protocols previously developed in the group (197,198). Expression was confirmed with western blotting, where the expected two bands (one for α-, 120 kDa, the other one for β-subunit, at 100 kDa) were visible (Figure 27a). Optimisation for MOI and harvest time gave the same results as previously (197): MOI of 5, 120 hours post infection, and Sf9 as the cell line.

Since the insect cell culture protocols have been optimised after previous attempts, with lower-passage number cells and serum-free culture, the integrin was successfully purified from a smaller volume (300-500 mL) of culture supernatant than previously (8 L). Consequently, the step of concentrating the supernatant to 1.5 L by filtering it with a tangential flow device could be removed from the protocol. The reduction in culture volume proved successful, with a visible Abs_{280} peak in the initial IMAC step (Figure 27b) and appropriate bands on the gel (Figure 27c,d), and significantly facilitated the purification process for the integrin.

The next step according to established protocols was Strep-tag affinity chromatography purification. However, this step diluted the protein further without enhancing its purity (data not shown), so it was omitted, further simplifying the purification. HRV 3C protease cleavage to remove the Fos-Jun dimerization domains and His- and Strep-tags from the α_v and β_3 subunits, respectively (for construct design details, see Table 1 and 165), was successful, as was later trypsin cleavage to remove the heterogeneity caused by spontaneous N-terminal degradation of the α_v subunit. This required optimisation to prevent digestion of all purified material, but once the short (30 minutes) time and low enzyme concentration (about 0.3 µg trypsin per 100 µg integrin) were identified, it was also successful. The integrin was recovered either by a reverse IMAC step or dialysed into the appropriate SEC buffer and purified by preparative SEC. The $\alpha_v\beta_3$ peak was concentrated to 0.3-0.5 mg/mL (integrins do not tolerate higher concentrations well). The yields were around 1-1.2 mg/L of culture, a notable improvement from the previous results of 0.03-0.08 mg/L.

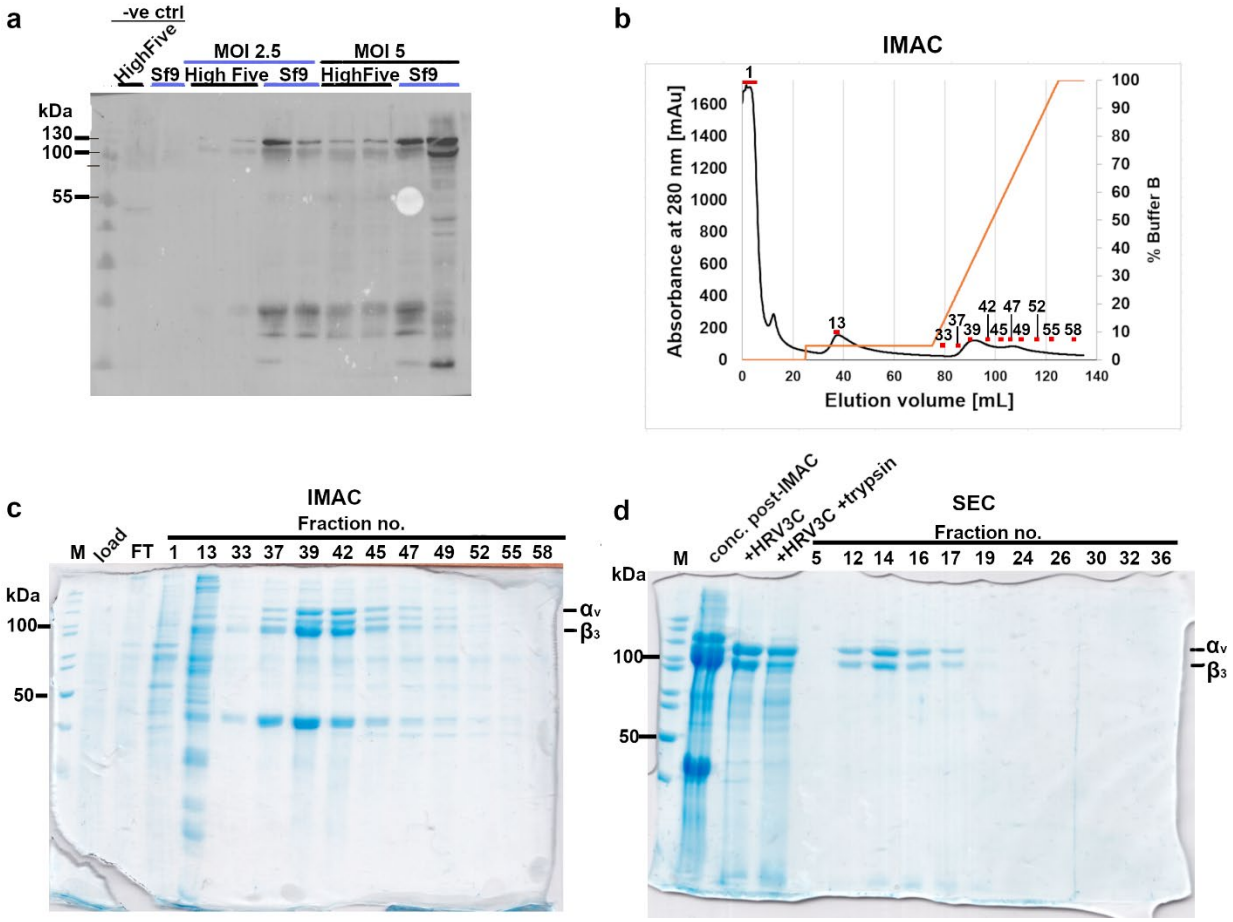


Figure 27. **Expression and purification of the integrin $\alpha_v\beta_3$ from insect cells. a.** Western blot with a primary anti-His-tag antibody of cell culture supernatants, incubated with and eluted from Ni^{2+} beads. The two bands running at higher molecular weights above 100 kDa correspond to the α_v subunit (expected M_w 130 kDa) and its spontaneous cleavage product (the β_3 subunit is not His-tagged). **b.** IMAC purification of $\alpha_v\beta_3$. Fractions analysed on gel in panel **(c)** are indicated in **red** and labelled. **c.** SDS-PAGE of the IMAC purification, with the two expected bands at 130 and 100 kDa for subunits α_v and β_3, respectively; degrading material from the α_v subunit and a 40 kDa impurity previously identified as cathepsin L are also present. **d.** SDS-PAGE of HRV3C and trypsin digest, after HRV3C removal of the dimerization domains with tags, and trypsin digest. In this case, the trypsin digest has not changed the composition visibly. The remaining lanes show the fractions for the SEC step, with two bands for the two subunits in the peak fractions.

3.2.5. Formation and characterisation of the HAdV-C5 non-dodecahedral mutant:integrin $\alpha_v\beta_3$ complex

The two complex components were run on the analytical SEC columns after incubation together in a 1:1 or 1:5 ratio or separately (Figure 28). The unavoidably low concentration of both components caused the Abs_{280} signal to be weak in the measurements, typically between 1 and 10 mAu per peak. However, the reproducibility of the elution volumes of these peaks could reflect the presence of new species of a particular molecular weight on complex formation.

Many of the penton base samples suffered from heterogeneity: the pentamer triplet peak was prominent before the purification protocol had been optimised. Nevertheless, a peak shift towards higher M_w on adding the complex components together could be seen, indicating the formation of a new, larger species. The peak was broad, spanning 0.3-0.4 mL from 0.9 mL (void volume) to 1.2-1.3 mL, corresponding to estimated molecular weight from 5 megaDaltons (MDa) at the void volume to below 670 kDa (reference thyroglobulin with a M_w of 670 kDa eluted at 1.31 mL). The estimated M_w, particularly for higher molecular weights, are approximate, as the calibration of analytical SEC columns did not involve species above 670 kDa. The expected M_w for the complex would be between 500 kDa and 1.3 MDa (the M_w of the pentamer singlet is 300 kDa, to which between one and five 200 kDa-integrins can be bound).

The broad peak suggests a range of stoichiometries, which would agree with the fact that more than one integrin can bind to the five available RGD loops. Typically, a peak corresponding to the M_w of the singlet pentamer would still be present in the complex sample, suggesting not all of the pentamer participated in the binding. This was the case also when oversaturating the pentamer with a 1:8 pentamer:integrin molar ratio (data not shown). However, the signal extending to the exclusion limit of the column would suggest aggregate, as the expected maximum molecular weight for the complex is well beyond that.

Overall, the evidence for complex formation from analytical SEC was rather weak and difficult to interpret, given the weak signal and variability of the peak shapes given by the components and the complex.

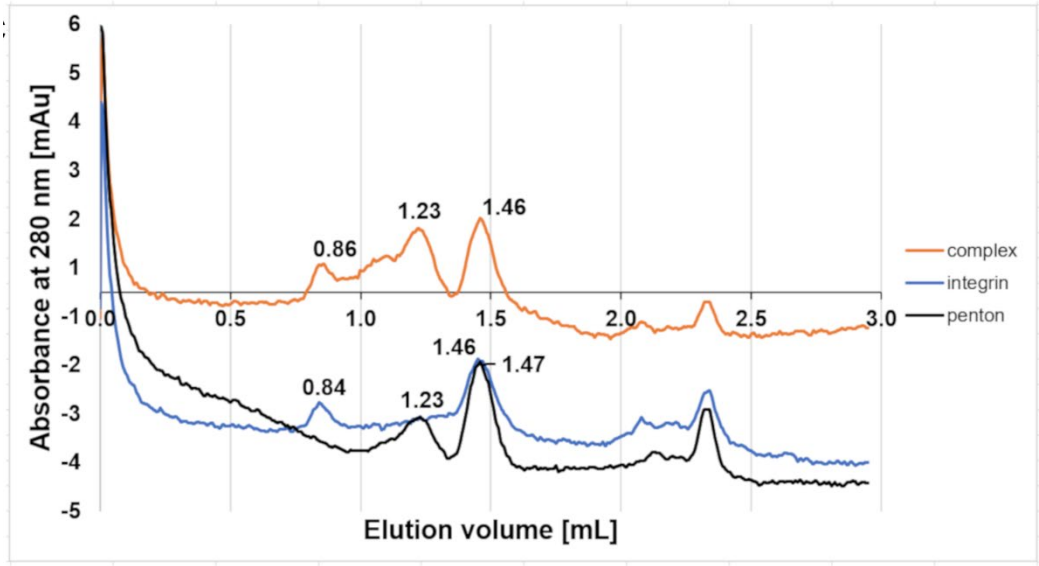


Figure 28. Example analytical SEC of a complex formation attempt between HAdV-C5 penton base dodecahedra non-forming mutant and integrin $\alpha_v\beta_3$. Analytical SEC (Superose6) comparison of the complex components separately and incubated together at a 1:5 ratio shows a peak shift towards higher M_w. Contamination of the penton base sample with higher-order oligomers is an issue. Estimated M_w: 1.23 mL – above 670 kDa, 1.46 mL – 418 kDa. Expected M_w range for the stoichiometries of the complex: 500-1300 kDa.

Negative stain electron micrographs of the complex samples showed pentameric species, some of which were surrounded by elongated, two-subunit proteins (Figure 29). This suggests that the complex was formed successfully. The number of integrins that could be identified around a pentamer varied between one and two (with three in some unclear cases, which could be more than one pentamer), though the stoichiometry of the complex as set up was 1:5. The sample was not homogenous, with some assemblies consisting of more than one pentamer, and the integrin assuming multiple conformations, as expected. The penton base showed a preference for a vertical orientation, and the horizontal side views were fewer.

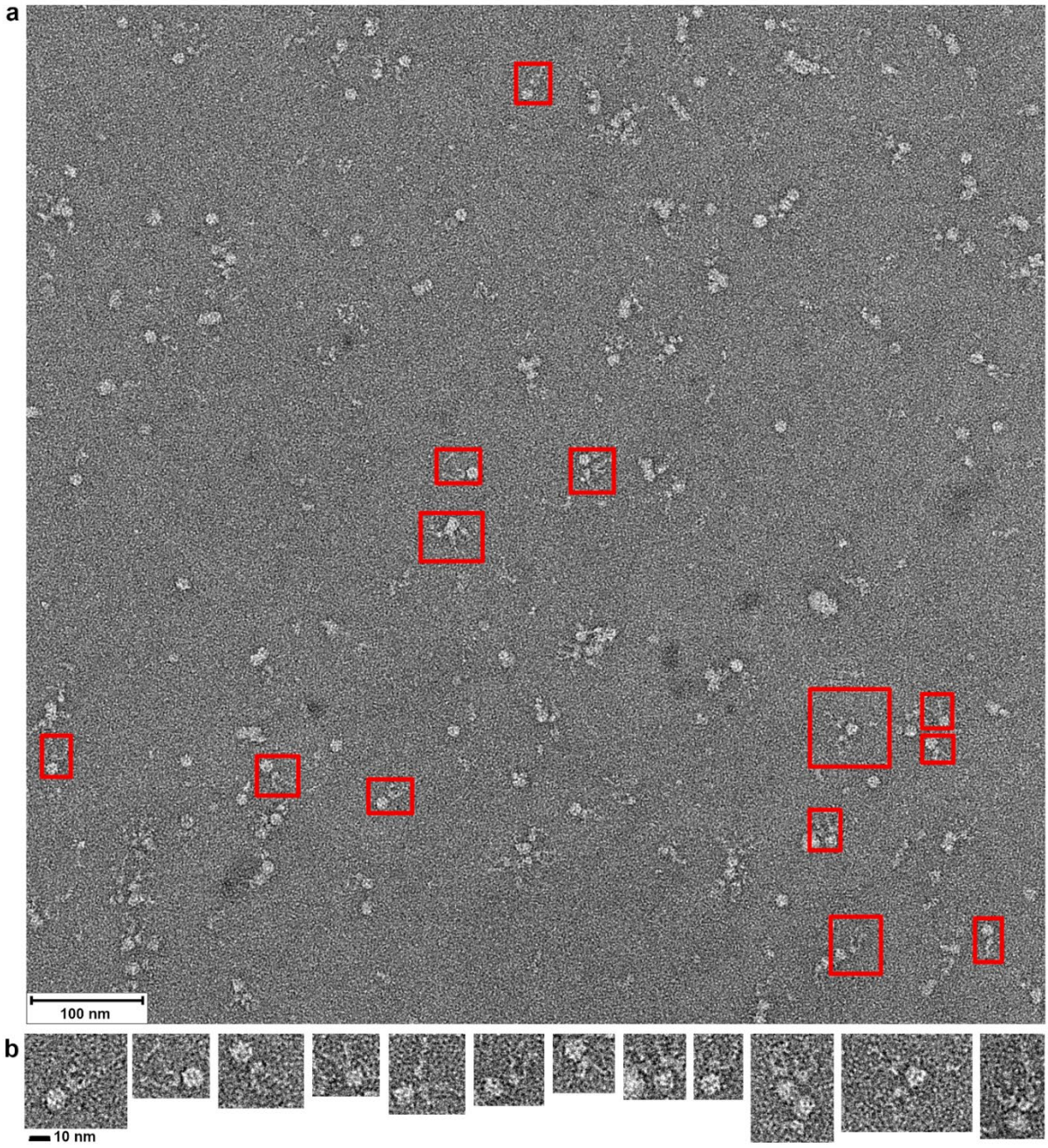


Figure 29. Negative stain EM of the complex of HAdV-C5 penton base dodecahedra non-forming mutant with integrin $\alpha_v\beta_3$. a. Negative stain EM images of the complex components incubated together overnight (1 to 5 pentameric penton base to integrin molar ratio) show pentameric shapes with elongated leg-like structures attached, corresponding to integrins. **b.** Enlarged segments of the micrographs, showing examples of the pentamer with integrins in more detail.

3.2.6. CryoEM data collection and processing

Initial data collections, on samples of 1:1 (210 movies) and 1:5 (141 movies) penton:integrin complexes showed the material was too concentrated on the grid. This positioned the pentamer into a top orientation, which resulted in 2D classes of only or almost only the top view (data not shown). Therefore, the best data set of 826 movies

was collected of the penton:integrin complex at a 1:5 stoichiometry, diluted 1:2 (see Table 7 for statistics)

Table 7. CryoEM data collection and processing statistics

	HAdV-C5 penton base	**HAdV-C5 with extra density**
Magnification	150000	150000
Voltage (kV)	200	200
Electron exposure ($e^-/Å^2$)	40	40
Defocus range (μm)	-1.2 to -3	-1.2 to -3
Pixel size (Å)	0.97	0.97
Symmetry imposed	C5	C1
Particles after 2D classification	147300	147300
Final particles	52070	6126
Map resolution	4.73 Å	9.70 Å
FSC threshold	0.143	0.143
Initial model (PDB code)	3izo, low-pass filtered to 60 Å	3izo, low-pass filtered to 60 Å

The data were processed as described above (see Section 2.2.22). Briefly, movies were aligned to produce micrographs, particle images were picked, sorted into 2D classes, and 3D reconstructions were obtained from particles in high-quality 2D classes (Figure 30).

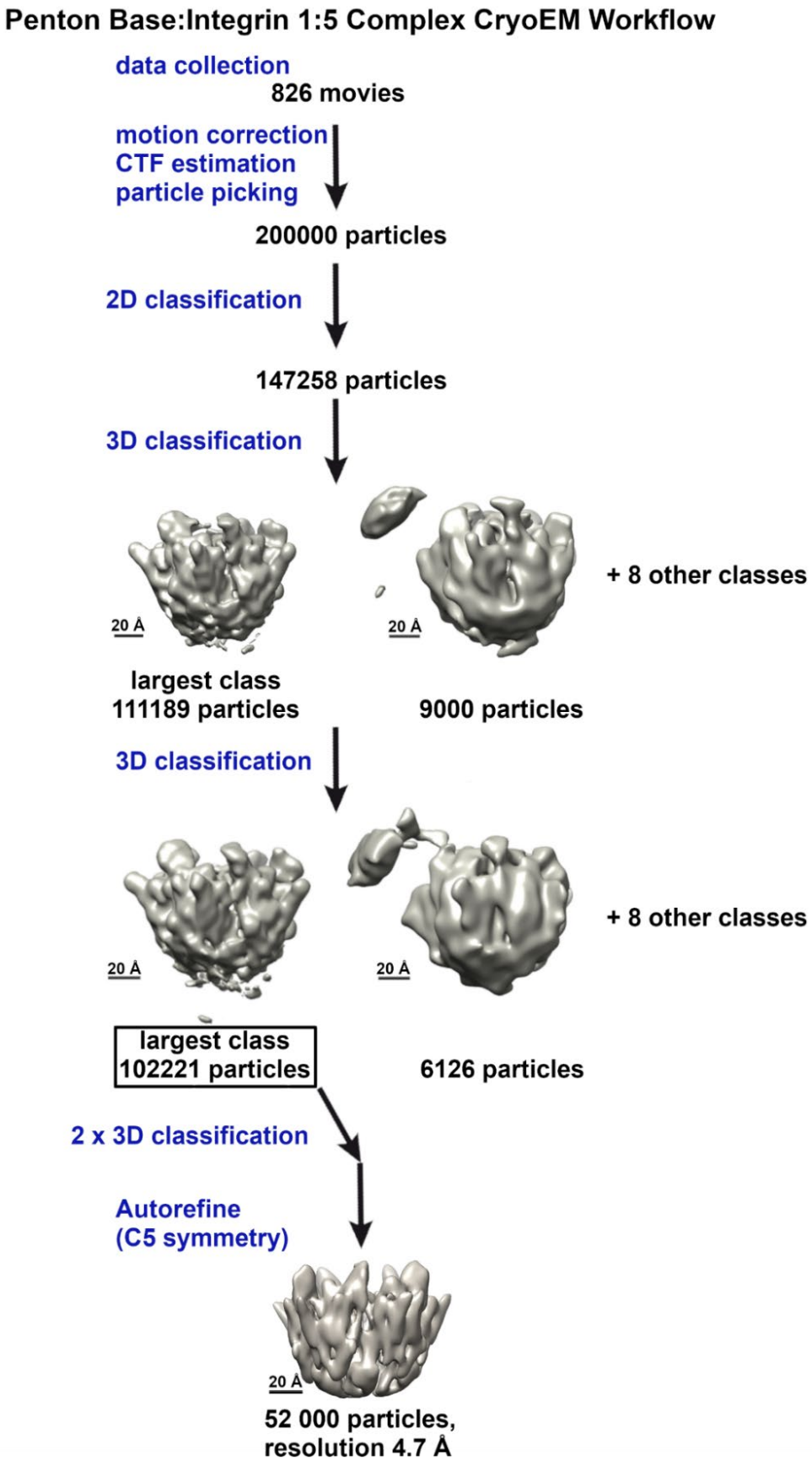


Figure 30. Overview of the workflow and results of the CryoEM data processing. All reconstructions shown at an isocontour level between 2σ and 3σ.

CryoEM micrographs of the complex (Figure 31a) visualised the penton base in pentameric form, in some cases with the integrin visible next to it, however, the inherent lack of contrast made it difficult to identify this protein. As a result, the final approach to particle picking was to pick all visible penton bases with a box size large enough (200 pixels; Figure 31b) to account for any potential integrin heads, then separate the bound and unbound penton bases during the *in silico* classification stages.

The averages of the 2D classes showed sharp, clearly pentameric features (see Figure 31c), with a preference for top views shown by the prevalence of such classes. Nevertheless, some side views were also present, as were bottom views. Cycles of discarding the particles that formed classes with featureless averages (likely picked erroneously) and subsequent re-classification led to a final set of about 147000 particles, which all belonged to classes with clear averages.

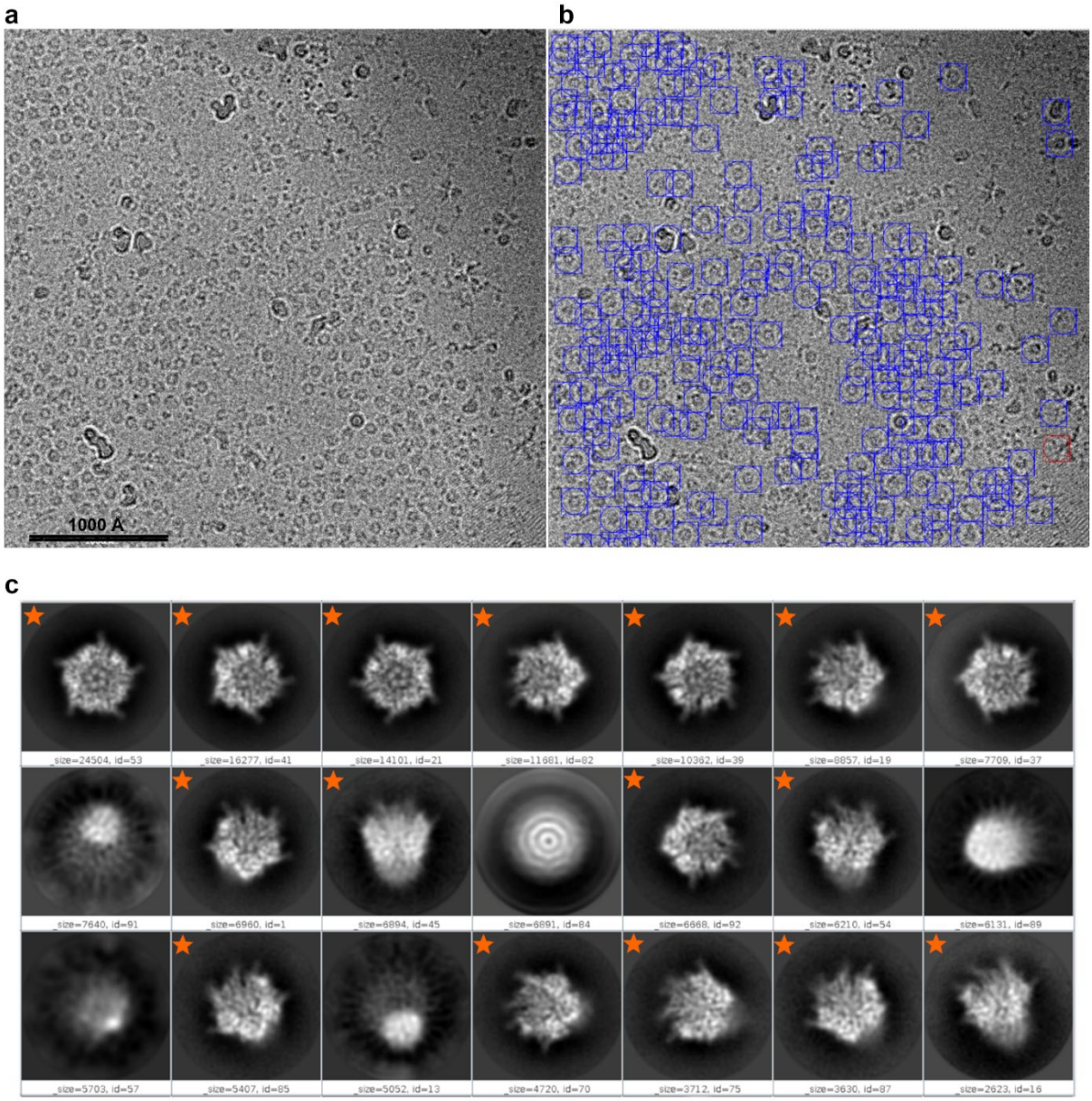


Figure 31. CryoEM data collection and reconstruction of the HAdV-C5 penton base-integrin $\alpha_v\beta_3$ complex structure, 1:5 penton base to integrin ratio. a. Micrograph of the sample, showing the presence of the penton base. The integrin is less clear. **b.** The same micrographs as in **(a)**, with particles picked by centring on the pentamers with a 200 px box. **c.** Sample of the largest 2D class averages calculated from the picked particles. No symmetry has been applied. There is a preference for top views of the pentamer, but classes with side and bottom views are also present. The integrin is not visible. Class averages for classes that were used in further processing marked with an **orange star**.

3D classification without applying symmetry produced a set of classes with pentameric features (Figure 30). The resolution-adjusted models from the X-ray structure of the HAdV-C5 penton base could be fitted into them with a cross-correlation of 0.75, and, in some cases, extra density could be seen, corresponding to unresolved flexible RGD loops, as for the largest class containing 111000 particles (Figure 32a). Further 3D classification rounds were performed until the largest 3D class was stable at 53000

particles reconstruction, with C_5 symmetry imposed at the final autorefinement stage, producing a penton base reconstruction (Figure 32b), with resolution estimated to be 4.7 Å (Figure 32c). The density for five HRV loops was clearly visible, however, the capsid-proximal lower part of the pentamer was largely absent, possibly due to flexibility.

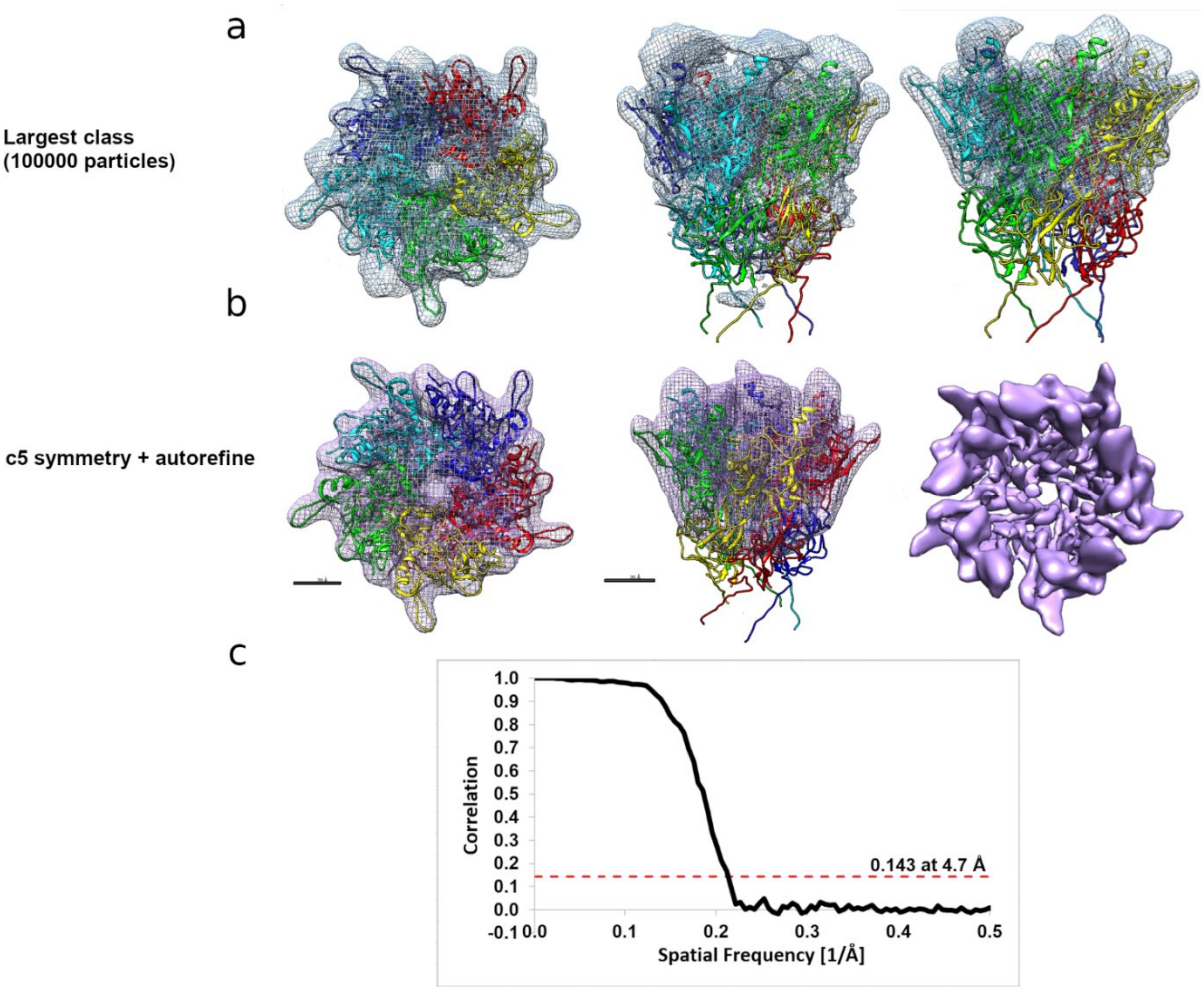


Figure 32. **CryoEM 3D reconstructions of the HAdV-C5 pentamer. a.** The largest (110000 particles) 3D class of the first classification, with an X-ray structure of the HAdV-C5 penton base (PDB: 3izo) fitted into the density. **b.** Reconstruction of the penton base, with the features enhanced by imposing a five-fold symmetry, with a resolution of 4.7 Å. Scale bar corresponds to 20 Å. **c.** FSC plot for the reconstruction from **(b)**, showing the estimated resolution according to the gold standard method (195).

A smaller 3D class (9000 particles) obtained in the first 3D classification showed additional density above the insertion domain that was not connected to that of the pentamer (Figure 33). However, the density was comparatively featureless, and could not be easily identified as the integrin or a neighbouring penton base.

Re-classification of the particles from the largest 3D class (Figure 32a) into 10 new 3D classes yielded another set of pentameric classes, one of which (a small one, with 6000 particles and estimated resolution of 9.7 Å) had additional density that was connected to the penton base by an elongated bridge (Figure 33). Again, the density lacked features that would identify it, making the description of the binding mode impossible. Attempts to use further protocols such as autorefine resulted in the disappearance of the bridge and loss of distinguishing features. Using the bridged class as the reference model did not provide any larger or more detailed classes with this feature. It is notable that the bridge was always linked to where the RGD loop of the penton base would be (it is not resolved in crystal structures due to the flexibility).

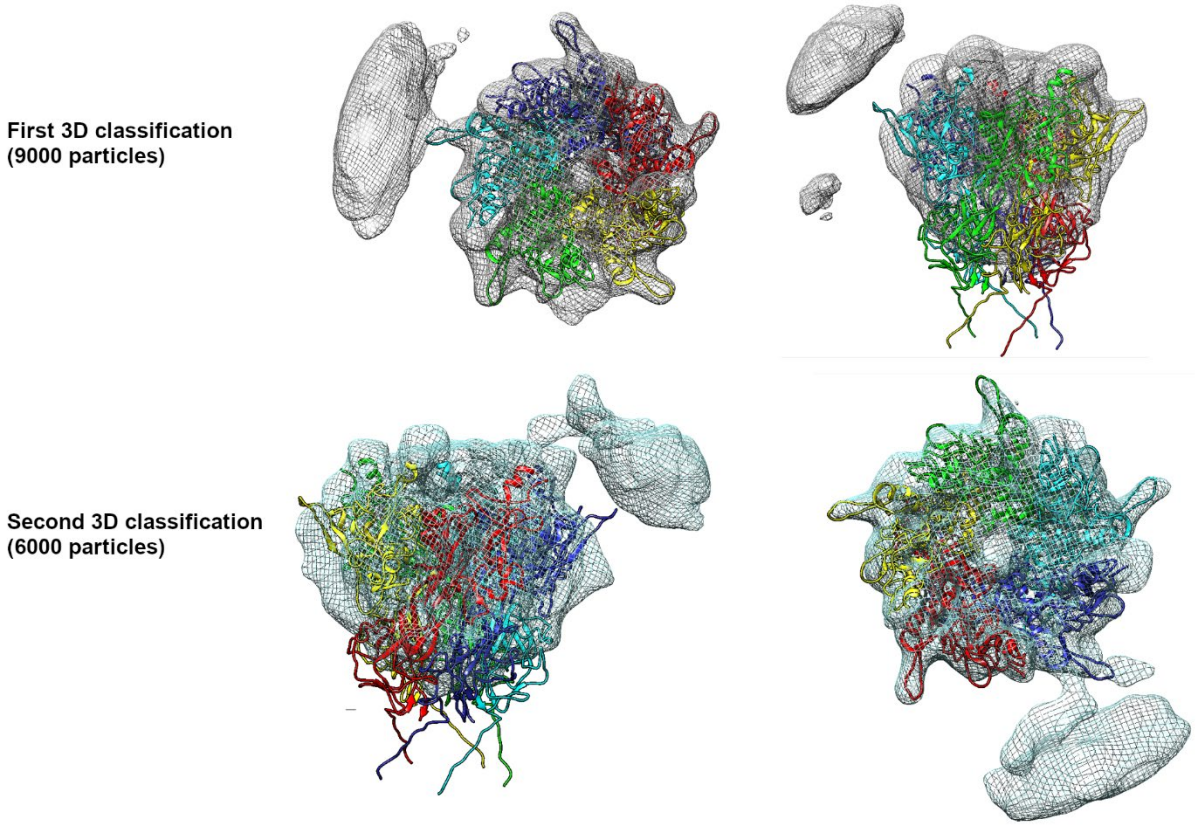


Figure 33. Smaller classes from the cryoEM 3D classification, showing the penton base and another feature adjacent to the penton base. In the case of reclassification of particles from the first classification, the extra density is linked to the density of the penton base.

An integrin head could partly be fitted into this additional density (Figure 34). The cross-correlation between the resolution-adjusted density generated from an integrin head X-ray structure and the extra density was 0.55. The legs are likely too flexible to be resolved, confirmed by attempts to classify particles picked with a 300-pixel box, instead of 200-pixel, to extend the reconstruction, which failed. The lack of detail made further interpretation of the non-pentamer additional density impossible, though the

bridge and correlation score would suggest it is the integrin, with the bridge being the penton base RGD loop. This linking feature is also present at 2σ, but not visible at the 3σ confidence threshold, further weakening its significance as evidence.

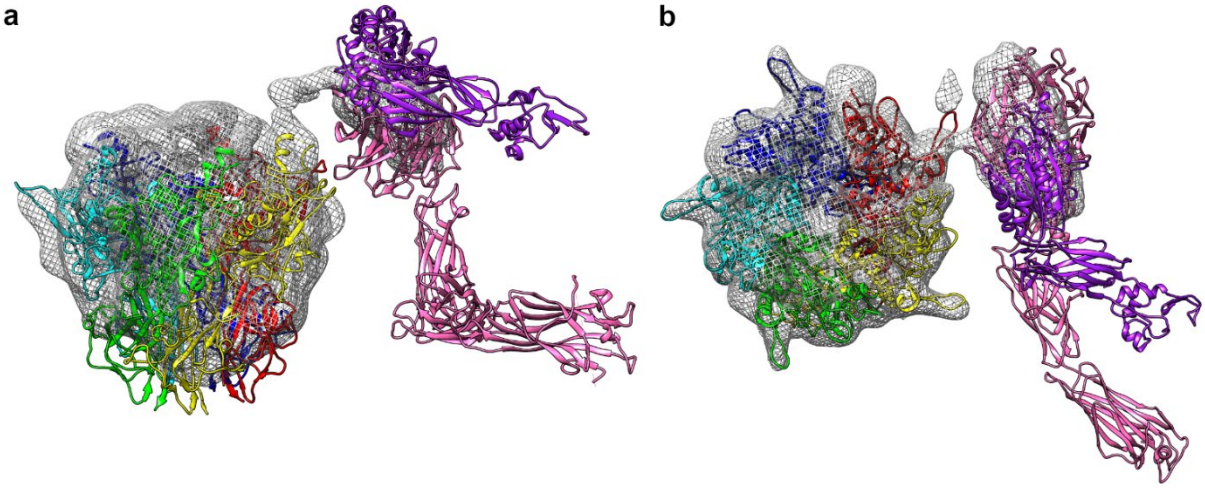


Figure 34. 3D class with extra density continuous to the penton base, fitted with the penton base and integrin structures (PDB ID: 3lm9), side **(a)** and top **(b)** view. The head of the integrin matches the density to a certain extent, but the lack of higher-resolution structural features could accommodate other proteins, including possibly another penton base.

3.2.7. Expression, purification of non-dodecahedral mutant of the HAdV-D9 penton base, and attempts at structural investigations of its complex with the integrin $\alpha_v\beta_3$

Penton base from HAdV-D9 has an RGD loop shorter by nearly 50 amino acids compared to HAdV-C5. As such, it is likely to be less flexible and the RGD loop-integrin interaction easier to resolve with structural techniques. Initial western blotting showed the protein to be expressing in Sf9 cells and purifying in the first IMAC step (Figure 35a). The highest signal was in the 48-h sample, which led to the initial shortening of the expression time from 96 hours. However, difficulties with solubility and higher-order oligomer formation, which could not be resolved with: lower salt (150 mM NaCl) buffer, phosphate buffers, 25 mM Tris-HCl pH 7.5 buffering component, harvesting freshly without freezing, SEC analysis of the protein after IMAC (major peak >600 kDa) to identify the pentamer singlets, overnight dialysis in SEC buffer to remove imidazole, incubating at room temperature for 1 hour or 1 week, and incubating with 3 mM β-mercaptoethanol as reducing agent (data not shown), led to a return to the original 96 hour expression time. Under these conditions and after generating new baculovirus stock, the high-salt HEPES buffer produced good solubility and high yields, the same as for penton base from HAdV-C5 (Figure 35b,c).

The multimeric assemblies remained an issue, as seen in preparative SEC, where the protein eluted in two peaks (Figure 35d). Analytical SEC of the lower-M_w peak (sample C3, Figure 35d,e) showed a range of species in peaks that run together (from 1.10 to 1.5 mL, corresponding to a large range between over 450 kDa to nearly 5 MDa at the 0.9 mL void volume, with a major peak at 1.28 mL above 670 kDa volume) and an additional, much smaller peak at 1.61 mL (between 440 and 158 kDa, possibly corresponding to the penton base; issues with calibration of the column made it impossible to estimate M_w, which could only be compared to the elution volume of the standard proteins).

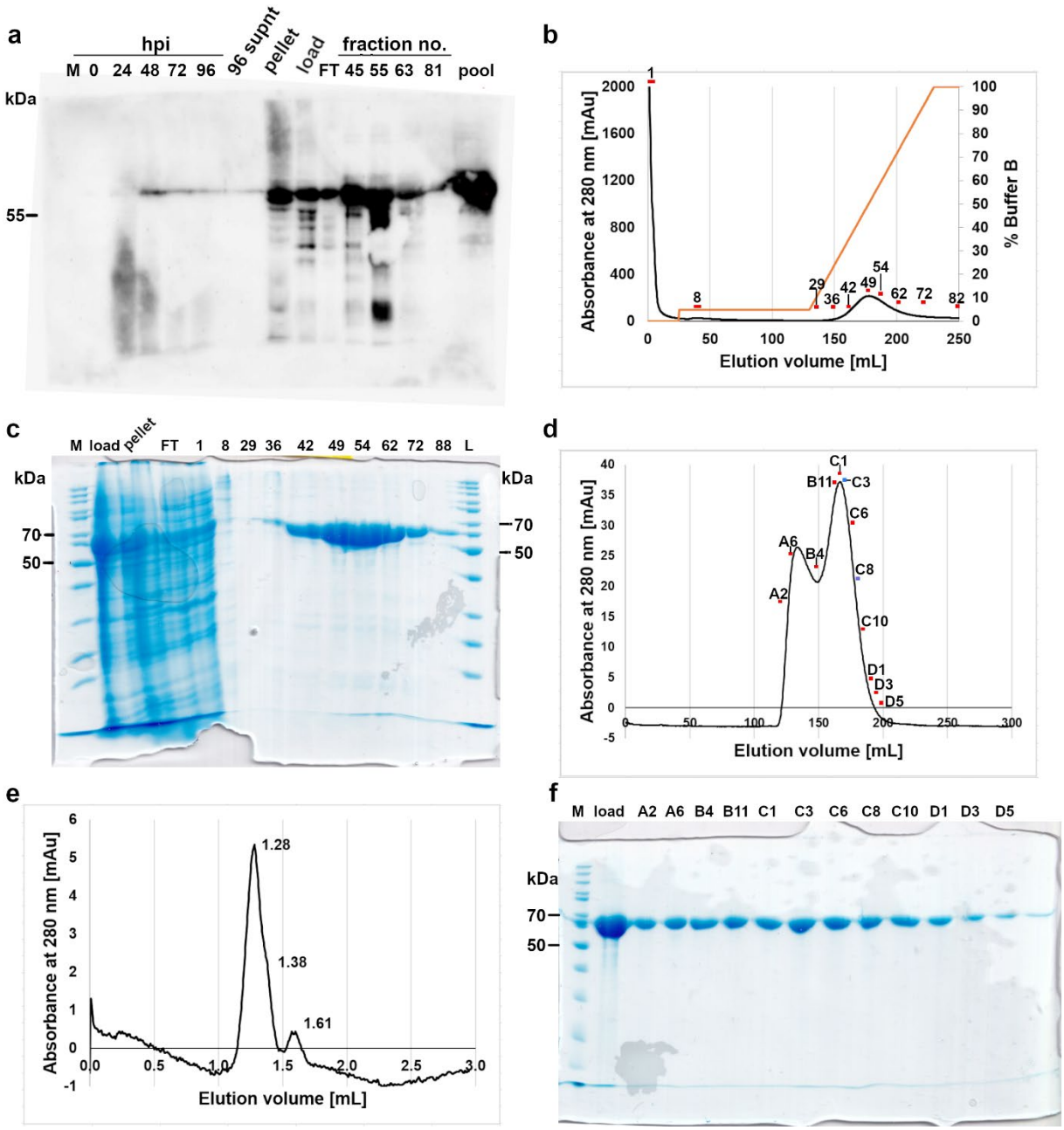


Figure 35. Expression, purification, complex formation, and cryoEM data for the HAdV-D9 non-dodecahedral mutant penton base. a. Western blotting with an anti-His-tag primary antibody of cell lysates from an expression in Sf9 cells and an IMAC purification of HAdV-D9 penton base. **b.** The IMAC step of the purification of HAdV-D9 penton base. Samples analysed in **(c)** marked in **red**. **c.** SDS-PAGE gel with samples collected during the IMAC. **d.** The preparative SEC step of the purification of HAdV-D9 penton base, showing two conjoined peaks. Samples analysed in (f) marked in **red**, fractions for (e) and negative stain (Figure 36) marked in **blue**. **e.** Analytical SEC (Superose6 Increase) of the lower-M_w peak from **(d)** showing a broad peak, or possibly two, suggesting a range of species. 1.28 mL elution volume corresponding to a M_w above 670 kDa, 1.61 mL between 440 and 158 kDa could correspond to a 265 kDa pentamer. **f.** SDS-PAGE gel with samples taken during the SEC (d), showing the protein runs as the same species for both peaks on the gel. Predicted M_w of HAdV-D9 penton base mutant is 55.7 kDa.

According to the SEC experiments, the homogeneity of purified HAdV-D9 penton base mutant was insufficient for structural studies. However, negative-stain EM of the lower-molecular weight arm of the lower-M_w peak (fraction C8, Figure 35d and Figure 36a) showed a nearly-homogenous, pentameric sample, at a concentration that enabled visualisation by EM. This suggests variation in the multimeric content of the preparative SEC peaks, likely at the level of distinct fractions.

Formation of the complex of the integrin with the penton base from HAdV-D9 (1:5 penton base:integrin ratio) was examined using analytical SEC. The Abs_{280} values were still low (1-10 mAu range) due to the limited concentrations of the components, but the peaks were sharper compared to the experiments with HAdV-C5 penton base (see section 3.2.5).

The penton base sample showed a single, broad peak, ranging from 1.05 to 1.46 mL (Figure 36b), suggesting a range of molecular weights and differing considerably from the profile in previous analytical SEC runs (Figure 35e). The peak absorbance was highest at 1.23 mL elution volume, compared to the 670 kDa thyroglobulin reference at 1.38 mL. In the complex sample, there was a single peak, at the void volume. This was a clear peak shift compared to the components, suggesting there is a change in the molecular weight, which would suggest a single species is present. However, the peak occurs at the void volume (0.91 mL), indicating a much higher molecular mass than expected, possibly aggregate, given that the higher limit of the column is 5 MDa. The maximum M_w for the complex is 1.3 MDa (the penton base pentamer is 265 kDa, that of the integrin 200 kDa after cleavage, and there are maximum five integrin binding sites on the pentamer), which is well below the exclusion volume. A 1.56 mL peak corresponding to the integrin decreases on incubation with the penton base, but does not disappear completely, suggesting not all of it is bound or involved in the aggregation.

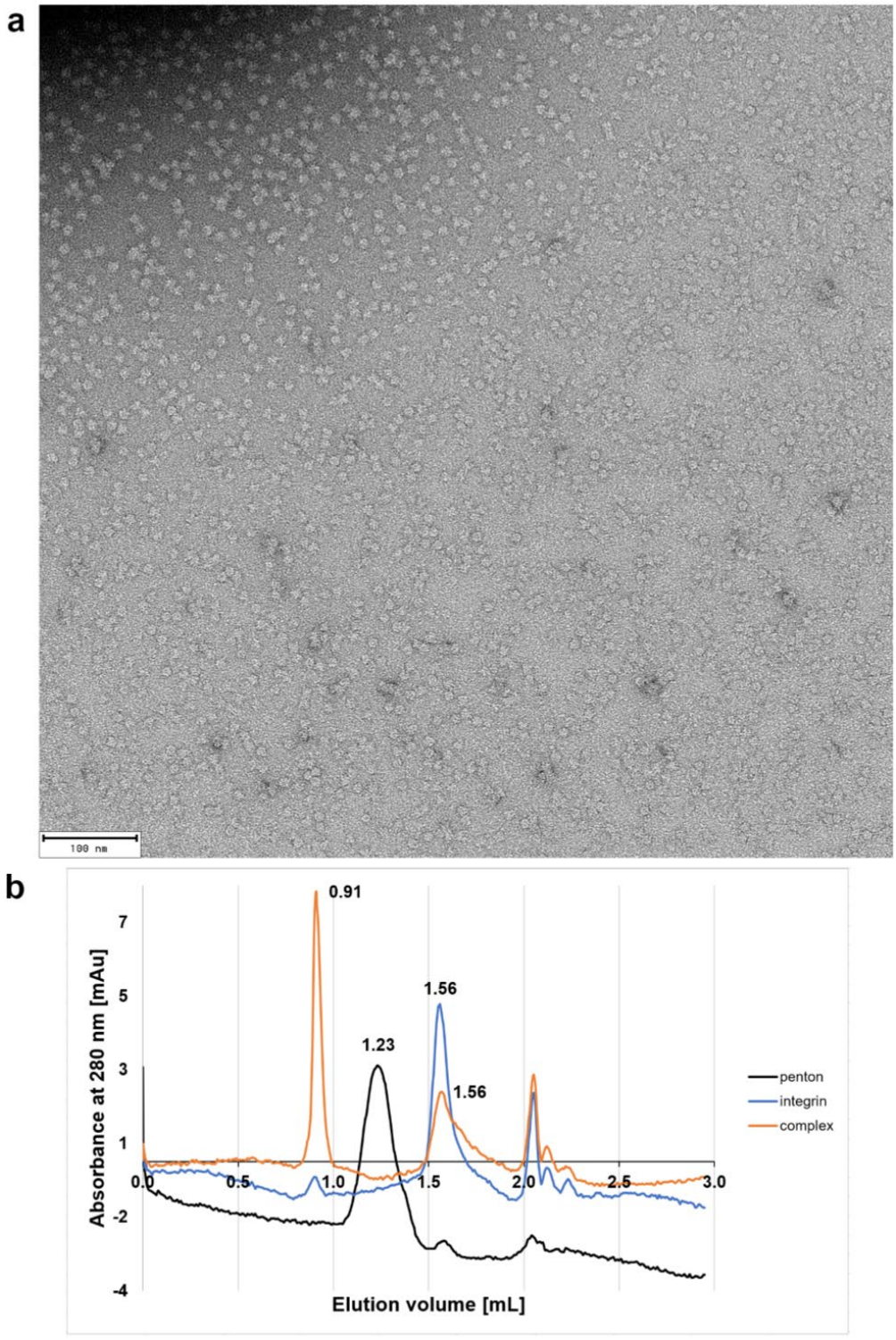


Figure 36. Negative stain EM of purified HAdV-D9 penton base dodecahedra non-forming mutant and its complex formation with integrin $\alpha_v\beta_3$. a. Negative stain electron micrographs of fraction C9 from **Figure 35d,** showing a largely homogenous, pentameric protein. **b.** Analytical SEC (Superose6 Increase) of the 1:5 penton:integrin complex and its components, showing a clear peak shift towards higher M_w compared to the complex components. The resulting peak elutes in the void volume (0.91 mL), which points to a very high (5 MDa or higher), potentially aggregated, M_w (maximum M_w of 1.3 MDa expected for the complex). The penton base sample peak is at 1.23 mL, above the 1.38 mL corresponding to 670 kDa and the calibration limit of the column.

Negative stain images of the complex samples showed exclusively the penton base (data not shown), likely due to the integrin being lost in that complex preparation attempt. CryoEM studies on the HAdV-D9 penton base:integrin complex, both at 1:5 and 1:10 stoichiometry, suffered from material thawing cycles during transport and the absence of material in the vitrified samples (Figure 37a). Some pentamers were present in the micrographs: for two sets, only around 200 particles could be picked, in each case from 30 micrographs. In another sample, which had thawed in transport, more material was present, enabling the automated picking of 17500 particles. Most of these were likely picked erroneously, and 2D classification averages revealed blurry, circular or faintly pentameric, features, seen almost exclusively as top views (Figure 37b). Further rounds of 2D classification to remove the contaminating particles did not improve class averages. The quality of the data did not warrant further processing. Similar issues were encountered with the new HAdV-C5 penton base:integrin complex samples processed and imaged at the same time (data not shown).

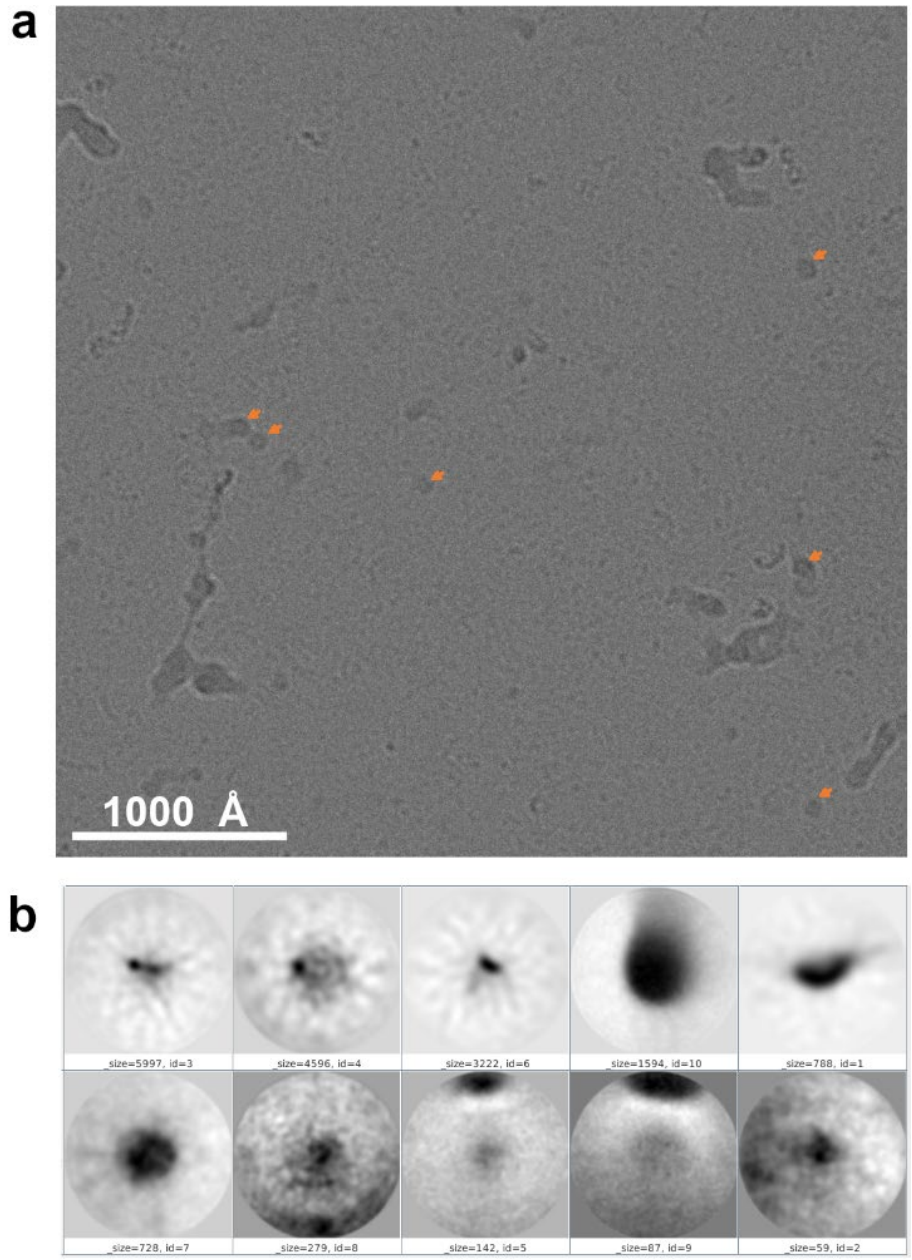


Figure 37. **cryoEM data collection on the HAdV-D9 penton base:integrin complex samples. a.** Micrograph of the 1:5 HAdV-D9 penton base:integrin complex, showing little pentameric material. Possible pentamers have been marked by **orange** arrows. **b.** 2D class averages of the particles picked, showing blurry, very faintly pentameric features, compared to the sharp classes of the HAdV-C5 penton base mutant dataset (see **Figure 31c**). Panel **(a)** provided by Dr Aušra Domanska, University of Helsinki CryoEM Facility.

3.2.8. Expression and purification attempts of the non-dodecahedral mutant of HAdV-B35 penton base

Penton base from HAdV-B35 was also expressed and purified according to the same protocol (HEPES-based buffers) as the penton bases from HAdV-D9 and HAdV-C5. The elution profile for IMAC (Figure 38a) looked similar to that of the other homologues, with three peaks, out of which the 50% elution buffer one was the highest. However,

in the fractions containing the most material, there were multiple bands on the gel in addition to the prominent expected 58.5 kDa M_w penton base monomers, which were the most prominent (Figure 38b). These additional bands were running at about 40 kDa, suggesting degradation.

The elution profile for the preparative SEC had a different higher M_w to lower M_w peak ratio to the other two penton bases, with the higher M_w peak having greater Abs_{280} than the lower M_w peak (Figure 38c). This could indicate less singlet pentamers and more of the higher order oligomers. Analytical SEC on the lower M_w peak (Figure 38d) showed virtually the same profile as the one for the lower M_w peak in the HAdV-D9 penton base: a broad peak from 1.13 to 1.5 mL, corresponding to a large range between over 450 kDa to below 5 MDa at the 0.9 mL void volume, with the highest at 1.28 mL above 670 kDa volume, and second, smaller peak at 1.57 mL, with 440 kDa ferritin running at 1.54 mL (compare section 3.2.7). Negative stain analysis showed a range of multimers larger than pentamers and generally heterogenous (Figure 38e).

While attempts had been made to collect cryoEM and initially process datasets of a HAdV-B35 penton base:integrin complex, too, they suffered from similar issues of too little material, thawing during transport, and only around 300 particles could be picked from the 30 micrographs taken. As in some of the datasets with penton base HAdV-D9 described above, this made them unsuitable for 2D classification 3D reconstructions (data not shown).

3.2.9. Attempts of purifying and imaging a cross-linked complex

Cross-linking the complex with DSS was attempted to enable a separation of the different stoichiometries present and removal of aggregates by SEC. The cross-linking step was generally successful, with silver stain gels showing additional higher-M_w bands in the cross-linked samples. The original bands were still present and there was no accumulation of most of the material in the pockets, which would suggest excessive cross-linking (Figure 39a,b). However, SEC separation of the cross-linked complex typically resulted in low signal and samples the contents of which could not be visualised with silver stain on the SDS-PAGE.

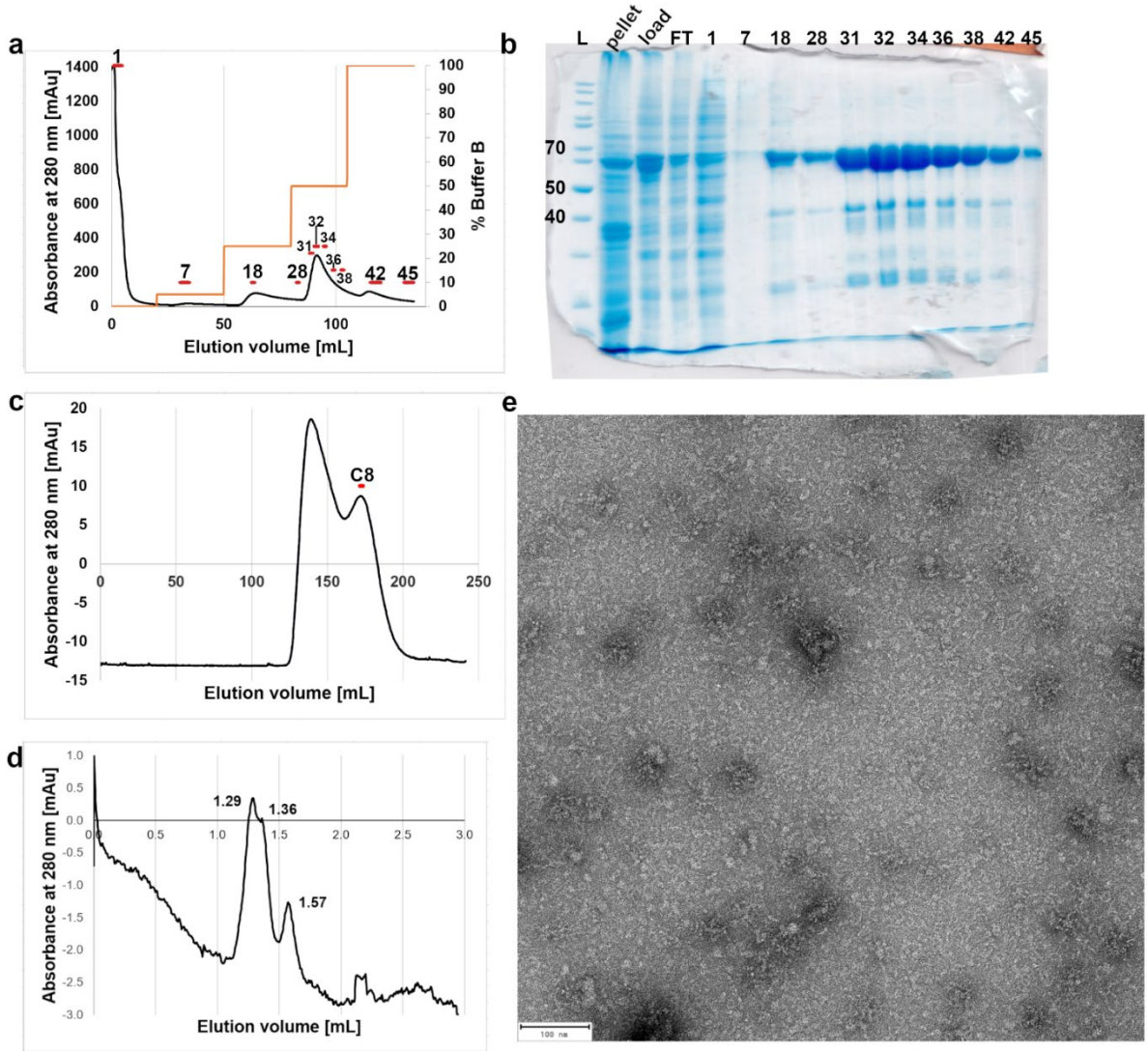


Figure 38. Purification and negative stain electron microscopy of HAdV-B35 penton base non-dodecahedral mutant. a. IMAC step of the purification of HAdV-B35 penton base, showing the highest peak eluting at 50% elution buffer. Fractions analysed in (b) marked in **red**. **b.** SDS-PAGE gel with samples taken during the IMAC, showing the protein at the expected M_w is eluting, although there are lower M_w species, particularly at 40 kDa, which were absent in the purifications for other HAdV types. **c.** Preparative SEC step of the purification of HAdV-B35 penton base, showing two peaks, with the higher-M_w peak the more prominent one. Fraction analysed in **(d)** and **(e)** marked in **red**. **d.** Analytical SEC (Superose6 Increase), showing a broad peak that could be two conjoined peaks, similar to the HAdV-D9 penton base analytical SEC elution profile (**Figure 35e**). **e.** Negative stain micrograph of the sample from **d.**, showing a range of species larger than pentamers and lacking pentameric features. Scale bar in the left hand-corner is 100 nm. Expected M_w for HAdV-B35 penton base is 58.5 kDa.

In this regard, the experiment shown in Figure 39b,c, showing both clear peaks on the chromatogram and bands identifiable as the complex components on the gel, was the exception. As the column used for this experiment had not been calibrated, it wasn't possible to assign M_w values to the species. The presence of both integrin subunits and penton base in the conjoined peaks 1 and 2, shown by SDS-PAGE, would suggest the smaller peaks corresponded to different complex stoichiometries, and the tallest

peak to free penton base or integrin. All peaks eluted well after the estimated 5 mL void volume of the column, corresponding to 6 MDa.

CryoEM on the SEC fractions of the cross-linked complex did not yield data of sufficient quality for 3D reconstructions (Figure 39d), with little material present and a preference for top views and a tendency of the penton bases to aggregate. Other attempts, using BS3 as the cross-linker, failed, producing no interpretable signal on the SEC or gel, and nothing but large aggregates visible in negative stain micrographs (data not shown).

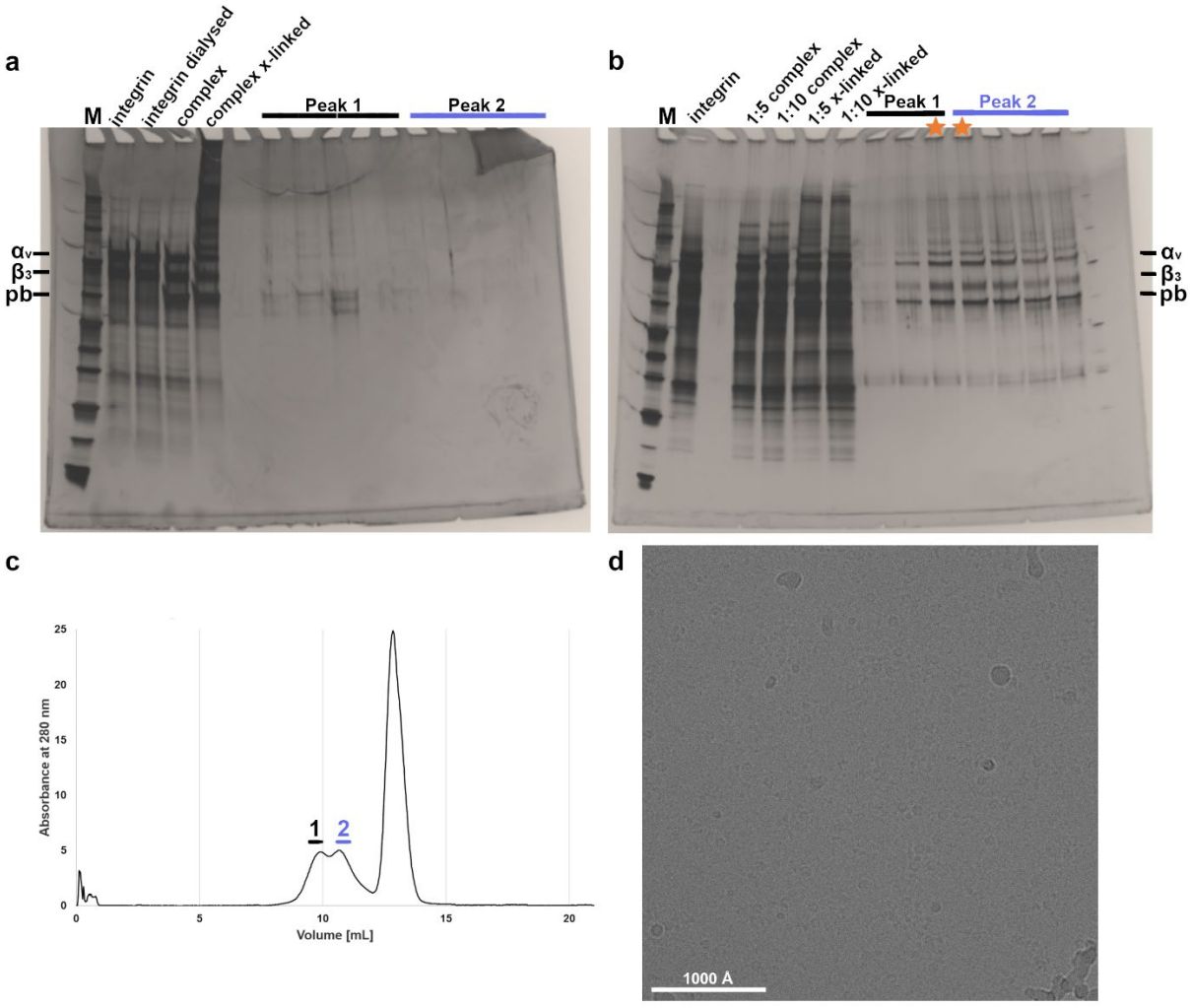


Figure 39. Cross-linking of the HAdV-C5 penton base mutant-integrin complex. a,b. silver stain SDS-PAGE gels of the complex components and DSS-H12/D12 cross-linked samples. pb, penton base. Lanes with samples vitrified for cryoEM marked by **orange** stars. **c.** SEC analysis (TSKgel G6000PWXL column) of the cross-linked sample of 1:5 HAdV-C5 penton base mutant:integrin complex from **(b)**. The void volume of the column was at about 5 mL and corresponded to 6 MDa. **d.** cryoEM micrograph of the sample from peak 2, showing the top views of pentamers, likely aggregated, which did not produce enough views for a reconstruction.

4. Discussion

4.1. Part 1 – *T. forsythia* kinases MurK and K1058

Previously optimised protocols for the production of large amounts of highly pure MurK and K1058 (Figure 11, Figure 12) enabled their crystallisation in complex with the sugar MurNAc. The production of well-diffracting crystals was challenging (Figure 13), with both complexes requiring at least six months for crystal growth. The resolution of the structures was somewhat limited (2.7 Å and 3.1 Å for MurK and K1058, respectively), precluding the modelling of waters, and the MurK crystals additionally suffered from merohedral twinning (see Table 2). Nevertheless, both provided information about the protein fold and binding of MurNAc. The overall structures were very similar to the previously solved apostructures, with the glucokinase fold typical for the ASKHA superfamily (Figure 14) within the ribonuclease H-like group (223). The protein forms a dimer, which was predicted to be biologically relevant by PISA and confirmed by analytical SEC (Figure 12). Comparison of apo, ATP analogue-, and MurNAc-bound structures showed that on MurNAc binding the N-termini of the proteins shift closer to the C-terminus, forming a more closed conformation (Figure 16), and the buried surface area of the dimer increased by 400 $Å^2$ (Table 3). The residues involved in hydrogen bond formation were the same in the two enzymes, with the exception of asparagine 32, which could only form hydrogen bonds with the MurNAc in K1058, and the binding sites of the two enzymes only differed by four residues (Figure 15). A Dali search for structural homologues revealed close matches among uncharacterised predicted proteins (one of which was a predicted GlcNAc kinase) from related bacteria, but also archaeal and human kinases (Table 4). Comparison of the active sites (Figure 8,9) from these and other, structurally similar, bacterial kinases showed that MurNAc binding is enabled by the absence of bulky residues in the active site, which would clash with the lactyl ether group of the MurNAc. Such residues (for instance glutamic acid 154 in *S. griseus* glucokinase or histidine 153 in *H. influenzae* mannosamine kinase) are present in kinases that are known to have other substrates and are unlikely to process MurNAc. Moreover, they are absent in the human NagK protein, which was recently shown to 6-phosphorylate both MurNAc and muramyl dipeptide (224). Coupled kinetic assays showed MurK had similar apparent kinetic parameters as established previously: a K_M of 180 µM (200 µM in reference 41) and substrate inhibition at the highest concentrations (1-3 mM MurNAc; Table 6, Figure 19). The previously uncharacterised K1058 had a better affinity (K_M 28 µM), but much worse

reaction speed (V_{max} 0.34 µmol min^{-1} mg^{-1}, compared to 80 µmol min^{-1} mg^{-1} in MurK), which correlated with the higher enzyme concentrations required to record signal in the assays. Active site mutants could be purified, were stable and folded (Figure 20), but preliminary data on K1058 specificity challenged the relevance of their design, so they were not included in the assays.

4.1.1. First crystal structures of kinases in complex with MurNAc suggest a specificity mechanism

New crystallisation conditions had to be identified for both MurNAc-kinase complexes rather than optimising conditions in which the apo and MurK/AMP-PCP structures crystallised, making the process more laborious (Figure 13). This was likely due to the large conformational change (Figure 16) in the protein on MurNAc binding, which prevented soaking of the crystals of the apo form with the sugar. Crystal damage or cracking on soaking with the sugar substrate has been reported also for human NagK (225) and *S. tokodaii* glucokinase (226). The N-terminal shift towards the C-terminal on substrate binding, forming a more closed conformation, has been observed for multiple glycokinases (225–230). This promotes the exclusion of water from the active site and is also consistent with the model of induced fit (231), according to which conformational change in an enzyme induced by substrate binding increases affinity for the substrate, allowing for selectivity and promoting the completion of the reaction. The rotation of the N-terminus towards the C-terminus on sugar binding is typically greater in other glycokinases than the 12° observed in MurK and K1058. While the human N-acetylmannosamine kinase also exhibits a 12° rotation (228), the N-terminal domains in the human NagK (on GlcNAc binding) (225), *S. tokodaii* broad-specificity glucokinase (on glucose binding) (226), *Pasteurella multocida* NagK (229), and *Plesiomonas shigelloides* NagK (230), rotate by 20-26° compared to their apo forms. In some cases, further rotations have been reported on phosphate binding to the enzyme-sugar complex, for instance 16° in *P. shigelloides* NagK-GlcNAc to AMP-PNP (230), 12° *Streptomyces griseus* glucokinase to AMP-PNP (232), and possibly on ADP binding to *S. tokodaii*-glucose (226). Other enzymes, such as the human N-acetylmannosamine kinase, do not exhibit conformational changes on phosphate binding by the enzyme-sugar complex (228). Although only a small shift is observed in MurK on binding AMP-PCP in relation to the apoenzyme structure, the lack of a ternary complex structure or data on the preferred binding order makes it impossible to exclude that a conformational change occurs on phosphate binding to the MurK-MurNAc

complex. For *P. shigelloides* NagK, binding to AMP-PNP alone elicits a small conformational change, and a 16° shift is only present on its binding to the NagK-GlcNAc complex, although the order of binding has been shown to be random (230). Therefore, it is possible that further conformational changes would also take place in the ternary complex of MurK or K1058.

While the inability to obtain the structure of a ternary complex reduced the amount of information on the catalytic mechanisms of the enzymes, superposition of MurK/AMP-PCP and MurK/MurNAc (See Figure 4 in Research Article, Appendix, 217), on the C-terminus due to the conformational shift of the N-terminus, provided a credible model. The two substrates bound in separate locations of the binding site and the distance between the reactive groups (3.4 Å) would enable the reaction to take place. In the proposed catalytic mechanism, aspartic acid 95 polarises the oxygen atom of the hydroxyl group at MurNAc C6, leading to the oxygen's nucleophilic attack on the γ-phosphate of the ATP. This type of mechanism has been proposed for multiple sugar kinases, and confirmed by mutational studies proving that the homologous aspartic acid is essential (230,233). The similarity of K1058 and K1058/MurNAc structures to those of MurK would suggest the ternary complex and reaction mechanism would likely follow the same principles.

No amino acids obviously mediating MurNAc specificity by engaging the N-acetyl or lactyl ether groups were present among the predicted hydrogen bonds in the active site of MurK (Figure 15a). Binding to MurNAc was similar in the two kinases (Figure 15), with the binding site conserved except for four residues, which did not have an immediately obvious influence. A hydrophobic, phenylalanine-lined cavity accommodated the methyl of the N-acetyl group, the oxygen of which formed an additional bond with asparagine 32 in K1058 only. This asparagine is conserved between the *T. forsythia* kinases, the human NagK (225), and *S. tokodaii* broad-specificity glucokinase, which can phosphorylate GlcNAc (226). The presence of this interaction in K1058, but not MurK, could at least partly explain the improved binding affinity of K1058 for MurNAc in the activity assays (Table 6), and suggest K1058's ability to process GlcNAc could be better than MurK, as suggested previously (47).

Comparison to structural homologues enhanced the interpretation of the structure. Top structural matches (Table 4) represented sugar kinases from the three domains of life: bacteria, archaea (*S. tokodaii* glucokinase), and eukaryotes (human NagK kinase). Given the importance of sugar metabolism to basic life processes, this deep

conservation is to be expected. Comparison of active sites revealed that sugar kinases known to process other substrates, such as glucose, typically had bulky residues in the active site, which would clash with the MurNAc lactyl ether group (Figure 18). This is unlike the serine 133 - proline 134 - leucine 135 residues and arginine 132 – alanine 133- glycine 134 residues in MurK and K1058, respectively, which lined the cavity, with the larger side chains pointing away from the active site, leaving space for the lactyl ether group. This observation likely enables the prediction of MurNAc-binding ability from sugar kinase structures, suggesting the top two Dali matches of bacterial hypothetical proteins and the human NagK. While there are no functional data for the hypothetical proteins, NagK has recently been shown to have a physiologically relevant ability of phosphorylating muramyl dipeptides (224).

Overall, the structures and modelling were able to provide a cohesive description of the binding sites and reaction mechanism, as well as suggest an explanation for the specificity.

4.1.2. Activity assays pose questions about K1058 function and specificity

The assays were not completely robust, being adapted from an earlier protocol, without optimisation by ensuring the saturation of some of the reagents (different ATP, NADH concentrations had not been tested, assuming they would be replenished by the reversible activity of LDH). Protein freezing and thawing was kept to a minimum, as, together with prolonged (months) storage at -80 °C, it was observed to result in issues with reproducibility. While the activity seemed highest in freshly purified protein that had not been frozen, it was impossible to make all 30 measurements of up to 45 minutes immediately after purification. Therefore, all protein samples underwent only a single freeze-thaw cycle, and all measurements were made within days of purification and freezing.

The results within this setup were reproducible for both proteins, including between different batches of purified protein. Additionally, MurK exhibited apparent kinetic parameters similar to previously reported (41); Table 6), and substrate inhibition at the highest MurNAc concentrations (1-3 mM) was also observed. Substrate inhibition at those concentrations is unlikely to be physiologically relevant, though it could perhaps be involved in regulating MurK activity in a feedback loop mechanism limiting its activity when the substrate is highly abundant (234).

K1058 has a greater binding affinity, but a much lower efficiency than MurK. This raises questions about its role and natural substrates *in vivo*. A few potential explanations are discussed below.

Firstly, K1058 could simply be an enzyme that has a low efficiency, and its function is duplicated by the more efficient MurK. While this would not confer an obvious selective advantage (and bacteria seem to be under selective pressure to lose genes, 235), paralogous enzymes can occur in an organism. In *Bacillus subtilis*, a pair of homologue (over 70% sequence similarity) glutamate dehydrogenases with very similar kinetic properties are regulated by two different promoters. They are expressed complementarily depending on the conditions and unable to substitute for each other when expressed under the other's promoter, as one of the proteins is more sensitive to glutamate concentrations and pH (236). Here, the low efficiency of K1058 (perhaps coupled with the substrate inhibition of MurK at high MurNAc concentrations) could be a mechanism to limit the 6-phosphorylation of MurNAc under conditions of high MurNAc abundance. A switch from a high to a low-efficiency enzyme, by transcriptional control or another mechanism, would ensure the limiting of production and prevent formation of excessive product, with the two kinases complementing each other.

Secondly, K1058 could have a different natural substrate and only incidentally have the ability to phosphorylate MurNAc. Earlier studies showed that it was unable to process a number of sugar substrates: glucose, 1,6-anhydroMurNAc, glucosamine, and galactose, only phosphorylating MurNAc and GlcNAc (47), similarly to MurK (41). These results from extended 16-hour assays led to the conclusion that MurK was specific for MurNAc only, and K1058 for both GlcNAc and MurNAc. Later mass spectrometry measurements showed both proteins phosphorylated GlcNAc, and preliminary kinetic assays suggested the efficiency was very low, lower than for MurNAc, in both cases (Marina Borisova, personal communication). Although these preliminary assays suffered from multiple issues, including reproducibility, their indications regarding limited GlcNAc-processing ability of both kinases (together with the inability to obtain a GlcNAc-bound structure of K1058 despite extensive trials) decreased the priority of studying this reaction. A direct comparison of the MurNAc- and GlcNAc-phosphorylating abilities of K1058 with full kinetic assays for both substrates and both enzymes would have improved our understanding of this property but was not undertaken due to time constraints.

K1058's localisation immediately upstream of the *Tanf_08345-08370* gene cluster responsible for peptidoglycan scavenging (237) (which in turn is directly upstream of the MurKTQ operon), would suggest it is involved in PGN recycling processes. It was postulated as a GlcNAc 6-phosphorylating enzyme in a model of PGN recycling (237), which would be coherent with there being no characterised *T. forsythia* enzymes with this function, and its location in the genome, but conclusive evidence is missing so far. The potentially poor kinetic parameters of GlcNAc phosphorylation by K1058 are not necessarily an argument against this being its natural substrate. Research into enzyme evolution has shown that enzymes do not evolve towards optimal kinetic parameters (238,239), but their evolution is constrained by multiple other factors (240), and greater efficiency does not necessarily equate to a selective advantage (241). The drive for evolutionary optimisation is particularly limited if an enzyme functions within a pathway that contains other, less efficient enzymes (242). In fact, on examination of kinetic parameters deposited for thousands of enzymes, it was found that the majority of the enzymes were only "moderately efficient", particularly when involved in secondary metabolism (238). For a GlcNAc kinase involved in the PGN recycling network, where GlcNAc-6P can already be generated from MurNAc-6P by MurQ (44), and other enzymes use GlcNAc as a substrate as well, the evolutionary pressure for optimisation may be limited.

Alternatively, as the lactyl ether group is not constrained in the active site, substrates with further functional groups added to the MurNAc moiety could be considered. K1058's structural similarity to the muramyl dipeptide-phosphorylating human NagK, as well as space in its active site for the additional functional groups, raised the possibility this could be the substrate. A preliminary mass spectrometry experiment with muramyl dipeptide was performed, but no phosphorylated product was detected (Marina Borisova, personal communication).

4.1.3. The immediate drug potential of the *T. forsythia* kinases is uncertain

The validity of MurK as a drug target is uncertain. Since *T. forsythia* MurK knockout mutants had a growth advantage under MurNAc-depleted conditions and showed little difference under MurNAc-supplemented conditions (41), MurK inhibitors would not be effective against *T. forsythia*. The MurK and MurQ proteins are responsible for shunting MurNAc to a catabolic pathway that ultimately provides additional substrates for glycolysis without direct involvement in PGN synthesis. While the *T. forsythia* MurNAc dependency seems a promising target for the development of drugs, the system is in

fact rather robust. The bacterium can use different sources of the sugar, importing both MurNAc and MurNAc-containing disaccharide products of PGN degradation, via the inner-membrane transporters MurT (44) and AmpG (243), respectively. Therefore, blocking MurNAc import would require inhibiting at least two proteins. The most promising target seems to be the postulated *Tf_AmgK/Tf_MurU* bypass pathway, which, given the absence of many of the classical *de novo* PGN synthesis genes, is likely the main route for PGN production in *T. forsythia* (45). In *Pseudomonas putida*, AmgK phosphorylates MurNAc and GlcNAc, which then serve as substrates for MurU, which generates the committed PGN precursors, UDP-MurNAc or UDP-GlcNAc (21). However, not only does the viability of these genes as targets need to be confirmed experimentally, but *T. forsythia* was also reported to have the ability to use sialic acids instead of MurNAc when grown in biofilms (46), a form strongly associated with periodontitis. Finally, the absence of this MurNAc dependence in some *T. forsythia* strains (42), together with the MurA/MurB sequences being present in some canine isolates (244), suggests that acquisition of MurA/MurB-encoding genes could easily occur via horizontal gene transfer, leading to resistance.

Moreover, periodontitis is a complex disease. Originally, the presence of particular bacterial species, particularly the red complex, has been associated with disease progression (245). However, more recent models postulate that the interplay between species is more important for the development of dysbiosis. The presence of defined groups of bacterial species is not strictly required for the development of the disease (37), with early pathology strongly mediated by inflammation and immune responses (246). Treatment of periodontitis relies on mechanical interventions, and antibiotic use is only recommended as a potential adjunctive treatment (247), as there are indications that it can be beneficial in certain cases (248). Targeting *T. forsythia* could possibly be another adjunctive intervention but, given the difficulty of drug delivery to biofilms or deep dental pockets, it seems unlikely to be practical or sufficiently impactful against the intricate factors that drive the condition. Successful selective antibiotic treatments for periodontitis, such as amixicile, have targets shared by anaerobic bacteria, reducing the population of multiple species associated with pathogenicity (249). However, should reports of the presence of *T. forsythia* correlating with conditions such as Alzheimer's disease or cardiac disease (reviewed in 34) be linked to pathogenesis, interest in specific therapeutic agents could increase.

Overall, structure-informed design of MurK inhibitors is unlikely to be of immediate use for treatment against *T. forsythia* and periodontitis. Nevertheless, other pathogens, for which MurNAc kinases or related enzymes are essential, could be identified and investigated, and inhibitors could also be of use in research on glycan processing pathways. For instance, the function of the *P. gingivalis* and *B. thetaiotaomicron* enzymes identified as MurK and K1058 structural homologues (Table 4), could be examined. Since these organisms are not auxotrophic for MurNAc, these genes are unlikely to be essential, but research on them could provide information on MurNAc processing. MurNAc metabolism is an attractive drug target, as it is absent in humans, which could limit treatment side effects.

4.1.4. Potential further experiments

A line of inquiry that presents itself is establishing the function of K1058: performing kinetic assays to characterise its interactions with GlcNAc and generating knockout mutants to examine their viability under normal and MurNAc-depleted conditions. Advantages or disadvantages in growth under different conditions, together with mass spectrometry data on accumulated byproducts, could indicate whether the enzyme is involved in MurNAc or PGN processing, or in different pathways altogether. MurK activity with GlcNAc should also be characterised kinetically and compared to the low activity reported previously (41). These results would yield a picture of the interactions of these two enzymes with GlcNAc and clarify whether they may be of significance *in vivo*.

Kinetic activity assays of the two closest structural homologues of MurK and K1058, the biochemically uncharacterised putative bacterial kinase proteins from *P. gingivalis* and *B. thetaiotaomicron* (Table 4), could also be performed. They could confirm whether MurNAc acceptance can be predicted from the structure of the enzyme active site, which needs to accommodate the lactyl ether functional group. This in turn could lead to mining for other potential MurNAc kinases in structure prediction databases, such as those generated by AlphaFold2 (171), and identification of further proteins of interest.

Finally, inhibitor design could be attempted, for instance by substituting functional groups that would form stronger interactions with the protein, for the N-acetyl and lactyl ether moieties of the sugar backbone.

4.2. Part 2 – Towards a structural understanding of HadV penton base-integrin complex

This project aimed to establish the complex structure of the HAdV penton base capsid protein with its receptor, the integrin $\alpha_v\beta_3$. The mutant HAdV type 5 and type 9 penton base and integrin $\alpha_v\beta_3$ were successfully expressed and purified from insect cells, although the penton bases suffered from sample heterogeneity, forming multimeric assemblies. SEC data suggested, but did not conclusively prove, formation of the complex, which was confirmed with negative stain and cryoEM micrographs. CryoEM micrographs contained both unbound penton base and the penton base:integrin complex at different integrin stoichiometries, which increased the challenge of data processing, particularly at the stage of particle picking. CryoEM reconstructions yielded clearly identifiable density for the pentameric penton base and, in some of the minor 3D classes, extra density lacking distinctive feature. This density was in some cases linked to the penton by an elongated bridge, but it could not be identified as the integrin with complete certainty. Attempts to enrich the complex at given stoichiometries by SEC or SEC with cross-linking were thwarted by the low concentration of the complex components.

4.2.1. Protein purification: issues of multimerization and low concentration

After initial optimisation of the expression and purification of HAdV-C5 penton base wild type produced heterogenous multimers (Figures 21-23), HAdV-C5 and HAdV-D9 penton base dodecahedra non-forming mutants could be successfully purified when solubilised in HEPES-based buffers and high salt (Figures 24 and 35). These conditions differed from the previously published ones for HAdV-C2 penton base that used Tris buffers (151,250), in which HAdV-C5 wild type was mostly insoluble (Figure 22ab). The HAdV-C2 construct was 50 residues shorter than the HAdV-5 wild type at the N-terminus, but there was only a difference of six amino acids in the following 481 residues. This would suggest that these differences, likely the absence of the N-terminus, together with the expression conditions (absence of a tag in HAdV-C2 penton base, expression in High Five cells), were enough to result in different solubility for HAdV-C5 penton base wild type.

Another study reported the purification of exactly the construct of wild type HAdV-C5 used in this work, which has been kindly shared by the authors (138, Annasara Lenman, personal communication); but without describing the buffer composition. Although it had been impossible to establish the experimental protocol with certainty,

they were probably phosphate-based buffers unsuitable for structural biology. The yields were low (Mikael Lindberg, University of Umeå Protein Purification platform, personal communication) and the protein assembled into heterogeneous multimers ranging from pentamers to dodecahedra (138). The reconstructed buffers had Tris substituted for phosphate buffers, but also did not improve solubility, and, given the lack of certainty as to the components, were not optimised further.

Ultimately, HEPES-based buffers were required to properly solubilise and purify the large amounts of HAdV-C5 penton base, both wild type (Figure 22) and dodecahedra non-forming mutant (Figure 24) needed for structural studies. The issue of multimer formation could be mitigated by increasing salt concentration (Figure 24) and omitting any concentration steps (Figure 25). The ionic composition of a solution influences interactions between proteins, particularly virus capsid proteins. For instance, the poxviral scaffolding protein D13 self-assembles into cage- or tube-like structures under low-salt conditions, and forms trimers in high salt (251), owing to specific charge interactions and overall charge distribution patterns. Similar observations have been made regarding HAdV-B3 dodecahedron stability, which dissociated into pentameric penton singlets on dialysis into buffer with 370 mM or 1 M NaCl for 24 hours at 37 °C (252). Since, among others, a charged interaction between amino acids conserved to arginine 452 (mutated in the penton base mutant constructs, see below) and aspartate at position 100 in HAdV-C2 penton have been shown to mediate dodecahedra formation (151,160), the increase in salt concentration likely interferes with interactions between pentamers. Unlike for capsid proteins of other viruses (253–255), mediation of the interactions by calcium ions is unlikely, as EDTA has been reported to have no influence on dodecahedra stability (252).

Additionally, concentrating the protein would bring the pentamers into closer proximity, increasing the chance of interactions occurring. No solution was found to break up triplets or higher-order oligomers of pentamers once formed, which necessitated a low concentration of the protein in the final preparations. HAdV-B3 dodecahedra have been reported to dissociate into free singlet pentamers under certain conditions (252). However, these are likely too harsh (incubation at pH 10.4, or at 37 °C for 24 hours) for samples to maintain the homogeneity needed for structural studies, while incubation for extended periods of time at 4 °C promoted the formation of triplets, not singlet pentamers (see Figure 26). Since the integrin also favours low concentrations, and cryoEM does not require protein concentrations as high as crystallography, this

was not an issue that prohibited complex production and data collection. However, the verification of complex formation was made more difficult by low absorbance signals in analytical SEC (Figure 28) or SDS-PAGE (Figure 39), and the purification of the complex by SEC was impossible (Figure 39). Additionally, monitoring of sample contents would have been easier with higher concentrations of both proteins, helping to avoid problems with sample preparations and empty grids in some of the cryoEM data collections (Figure 37).

The lack of easily accessible, reliable methods of multimeric state verification was a major hurdle: for most of the project it was assumed that the preparative and analytical SEC provided sufficient information regarding the molecular weight of the purified penton bases. However, as observed in the purifications of HAdV-D9 and HAdV-B35 (Figure 35, Figure 38), a sample could present as multiple peaks on the analytical SEC, varying between fractions, then be nearly homogeneous under negative stain electron microscopy (Figure 36), but this was not reliably reproducible (Figure 37). Size estimation of the species was also problematic, especially given the lack of column calibration above 600 kDa, leaving the values from 600 kDa to 5 MDa without clear assignment. In hindsight, a combination of a method establishing absolute molecular weight, such as SEC-multi-angle light scattering (SEC-MALS), with negative stain electron microscopy, and possibly native PAGE, would likely have been more effective. Alternatively, ultracentrifugation could have been used to separate the different molecular weights of the multimers, which could even remove the need to design non-dodecahedral mutants.

Another issue was that the mutants designed to prevent the formation of dodecahedra still showed a tendency towards forming multimers, tentatively identified as pentamer triplets based on analytical SEC and negative stain results (Figures 24-26, 28, 29, 38). While the formation of full dodecahedra does not seem to have occurred, the multimers still make structural studies challenging, compromising sample homogeneity, complex formation or characterisation due to the necessity of maintaining low protein concentrations, as discussed above.

The mutants had a truncated N-terminus beyond the SELA motif (about 50-60 amino acids) and the arginine involved in dodecahedra formation (arginine 452 in HAdV-C5 penton base, see Table 1) was changed to a glutamic acid. This design was based on the crystal structures of HAdV-C2 (151) and HAdV-B3 (160) penton base dodecahedra, which identified the residues homologous to aspartic acid 100 and

arginine 452 in HAdV-C2 as mediating a charge-based interaction, as well as a strand-swapping mechanism between pentamers, present only in HAdV-B3, initiated by N-terminal residues $_{59}$ELS$_{61}$ (ELA in the penton bases used in this work).

The HAdV-C2 penton base formed dodecahedra only under particular crystallisation conditions (151), unlike the HAdV-B3 penton base, where dodecahedra occur during infection, and form spontaneously under some purification conditions (160). The absence of the N-terminus strand-swapping mechanism, initiated by the serine-glutamic acid-leucine-serine (or, as also proposed, the homologous serine-glutamic acid-leucine-alanine residues in other HAdV types) motif and involving additional residues upstream, in the HAdV-C2 structure, has been suggested as the reason for this difference (160). Since changing the glutamic acid-leucine-serine motif to aspartic acid-valine-alanine abolished the formation of dodecahedra completely, but the structure showed it involved amino acids 49 to 61, and N-terminal truncation of the initial 47 amino acids also reduced dodecahedra formation (160), in the mutant constructs the N-terminus was truncated beyond the homologous serine-glutamic acid-leucine-alanine motif to abolish this interaction. Given the tendency to spontaneous proteolysis at the N-terminus (151,161,252), this appeared to be a practical solution to both improve protein stability and prevent multimer formation. However, these two approaches have not been shown to function synergistically, and it was the change to aspartic acid-valine-alanine by itself, with the N-terminus undergoing only spontaneous proteolysis (mostly of the first 9 amino acids), led to the production of singlet pentamers. Removing the N-terminus could be counterproductive, as an earlier study of HAdV-B3 penton dodecahedron structure found that a truncation of the first 37 amino acids was required for dodecahedron formation (161). The postulated reason for this was the inability of the disordered N-terminus to fit inside the limited space of the dodecahedron and potentially disrupting contacts between pentamer multimers. The later findings by some of the same authors that, conversely, the proteolysis of the first 47 (though not 40 or 42) amino acids limited dodecahedra formation, were explained as those ten amino acids mediating strand-swapping (160). Still, the complete removal of amino acids that could have obstructed contacts between the pentamer singlets was not necessarily beneficial and could have additional consequences. However, it seems unlikely to promote the formation of dodecahedron intermediates such as pentamer triplets: triplets do not enclose a limited volume where the N-termini need to fit.

Further, the choice of altered residues could also be examined. The arginine 452 residue was shown to be involved in a charge-based interaction with aspartic acid 100 in the HAdV-C2 structure (151), and its mutation to a glutamate reduced dodecahedra formation (160). However, the interactions of other residues were also involved in stabilising this region. In addition to the interaction with arginine 452, the backbone oxygen of aspartic acid 100 mediated contacts with threonine 454 and serine 455 (151). One of the differences between some dodecahedra non-forming and forming types is the amino acid change from serine 455 to arginine 455 (one of the six differences between HAdV-C2 and HAdV-C5 outside of the N-terminus) (160). An auxiliary charge-based interaction between arginine 455 and aspartic acid 100 has been postulated in the HAdV-B3 penton base dodecahedron structure, although the resolution was not high enough to verify this (160). It seems possible that, in the absence of arginine 452, arginine 455 could substitute for it in the interaction with aspartic acid 100 to a certain extent. The negative stain micrographs of HAdV-B3 penton base with amino acid changes influencing dodecahedron formation showed that the sample of aspartic acid 100 to arginine (D100R) variant contained more singlet pentamers, whereas arginine 452 to aspartic acid (R452E) contained full dodecahedra and smaller multimeric assemblies (160). It has been proposed that the repelling interaction between the three arginines serves to additionally destabilise the dodecahedra of D100R (160).

Therefore, it seems that designing constructs with a D100R amino acid change instead of R452E would be a promising approach. The serine-glutamic acid-leucine-alanine motif could also be mutated, changing glutamic acid-leucine to aspartic acid-valine, and the N-terminus truncated by 47 amino acids to avoid heterogeneity introduced by spontaneous N-terminal degradation. While the most robust way to confirm the mechanism of multimer formation by the penton base and the effect of the introduced amino acid changes in current and future constructs would be to crystallise the penton base pentamer triplets, this would likely not be time- or cost-effective.

Alternatively, and more simply, considering the variation between HAdV types, penton bases known not to form dodecahedra or only forming them under particular conditions, such as HAdV-C2, which requires 1.5 M ammonium sulfate and 10% dioxane (151), could have been chosen. However, it has been suggested that given the absence of amino acids characteristic for penton bases of dodecahedra non-forming serotypes, all penton bases are able to form dodecahedra under appropriate

conditions (250). Since it is likely that concentration steps required in protein purification could drive these processes, designing mutants of serotypes of interest seems the better course of action. Additionally, integrin specificity is not known for all HAdV types, which could introduce the requirement to establish a purification of an integrin different than $\alpha_v\beta_3$ to obtain a complex.

The optimisation of the expression and purification of the integrin $\alpha_v\beta_3$ was successfully achieved, resulting in great practical improvement. The previous workflow (165), while feasible, involved handling litres of culture, increasing the difficulty and cost. The improvements were simple and not directly related to the purification protocol itself, but to the expression in insect cell culture. Once this was optimised to serum-free conditions, reducing the presence of contaminants in the supernatant into which the integrin was expressed, and to shaking culture instead of adherent culture, the supernatant volume required for a purification decreased more than ten times. This, together with the reduced contamination by other proteins, allowed for exclusion of multiple steps from the protocol, simplifying it noticeably.

4.2.2. The HAdV penton base-integrin complex can be formed, but is difficult to characterise

Verifying complex formation with analytical SEC indicated that the molecular weight composition changed on incubating the components together (Figure 28, 36). However, the lack of reproducibility and of calibration for higher values made it difficult to determine with certainty what the values were, and whether they corresponded to particular stoichiometries. The presence of higher-order multimers confounded this, contributing to the presence of a broad range of higher-M_w specimens, which could not be identified as complexes, penton multimers, or aggregates.

The formation of the complex was successfully confirmed by negative stain and cryoEM, both of which showed the (generally singlet) penton bases clustering with elongated, two-legged structures (Figures 29, 31), corresponding to previously published images of the integrins (165). Due to the lack of contrast in cryoEM micrographs, it was easier to distinguish the complexes in negative stain images, where they showed 1:1 and 1:2 pentameric penton base:integrin stoichiometry, with some putative 1:3 species. The latter were difficult to distinguish from possible 2:2 species, as the singlet penton base was not completely homogenous.

Multiple stoichiometries of the complex are to be expected based on prior results and the presence of five integrin binding sites on a singlet pentamer of the penton base

(165–167). The data here are not of high enough quality to draw quantitative conclusions other than that there seemed to be preferentially between one and two integrins bound per penton base out of the maximum number of five (Figures 29, 31). This would be expected if the potential steric clashes between integrins in open conformation are considered, indicating it could be difficult to fit above three integrins around the penton. The ratio on setting up the complex was typically 1:5 penton base pentamer singlet to integrin, and the analytical SEC profile did not change on an increase to 1:8 (data not shown), indicating the complex had been saturated with integrin. Conversely, HAdV entry by endocytosis has been proposed to be mediated by integrin clustering-induced signalling (167), which would be expected to involve binding to more than two integrins. Moreover, the long RGD loop of some HAdV types, including HAdV-C5, could extend away from the penton, increasing the space available and the number of integrins that could be accommodated around it.

While care was taken that as many of the particles picked for cryoEM data processing as possible were in complex with integrins, this was only partly successful. Only a minority of 3D classes showed density additional to the penton base, and only some of these classes showed a connection between it and the pentamer (Figures 30, 33, 34). This density was largely featureless compared to the neighbouring penton base (9.7 Å estimated overall resolution). While it was possible to fit the crystal structure of the head of the integrin $\alpha_v\beta_3$ into this density (Figure 34), with an acceptable cross correlation coefficient, this was also possible with other proteins, for instance another penton base, and so is not fully convincing as an identification. However, a pentamer would be exhibiting additional symmetrical features, which are not seen in this case, and no other proteins were present in the sample.

Nevertheless, no molecular details of the binding can be resolved at the obtained resolutions. The presence of the extra density and connecting bridge with the RGD loop is dependent on noise levels: however, they are all present in the range between 2 and 3σ isocontour level, which makes them more credible. Considering the dense clustering of the sample on the micrographs, it could be a possibility that this extra density is a product of signal from both the integrin and extra pentamers. However, the bridge ending at the penton base density where the flexible integrin-engaging RGD loop begins would suggest the connection is to the receptor.

The multiple stoichiometries present in the sample were the greatest challenge for structural data processing and reconstruction. Particularly, the presence of unbound

pentamers in the largest dataset was a pronounced issue: they dominated the 3D reconstructions, likely averaging out many particles with extra density, with new bridged extra density classes only obtained in the second round of the reconstructions (Figures 30, 32). While further *in silico* purification approaches should be tried, especially given the rapid development of new methods in cryoEM, obtaining a sample without unbound pentamers would be the easiest solution. This has been attempted by oversaturating the pentamer, with a 1:10 pentamer:integrin ratio (unsuccessful data collections, see Figure 37), or purification of a single stoichiometry by SEC, for instance with prior cross-linking (see Figure 39). Both of these approaches failed initially; however, they could succeed with more optimisation or be adapted to other methods. For instance, as the cross-linking of the complex was successful, cross-linking mass spectrometry (256) could be applied to determine if there are other interacting residues of the penton base and the integrin beyond the RGD loop.

4.2.3. Potential future experiments and optimisation strategies

To obtain improved data on both the stoichiometry and molecular details of the binding, a number of approaches could still be attempted. Even with the current protein preparations, a more extensive data collection of negative stain EM on the complex sample could enable 2D classification of the species and more reliable information on preferred complex stoichiometries. Native mass spectrometry (257) or SEC-MALS could also be used to determine the stoichiometries present in the sample.

However, improvements in the concentration and homogeneity of the HAdV penton base component could be introduced, with more reliable techniques to establish the M_w and multimeric status, optimised constructs, or purifying penton bases with a reduced predisposition to dodecahedra formation (HAdV-C2), as detailed above. With a highly concentrated, homogeneous penton base sample, quantitative methods could be used to establish binding affinity, such as surface plasmon resonance or microscale thermophoresis, and the quality of analytical SEC measurements could be improved. Additionally, SEC could be used to purify the complex at a higher concentration by incubating with the integrin before the preparative SEC step, and purifying particular stoichiometries of a cross-linked complex would also be easier. CryoEM data could be collected on an electron microscope with higher voltage than the Arctica (200 kV), such as a Krios (300 kV), for a higher resolution of the reconstruction.

Secondly, the penton base component of the complex could be replaced with a more physiologically relevant species: the intact capsid of HAdV. CryoEM is a technique

well-suited to large, symmetrical macromolecular assemblies such as viruses, and a structure of the whole capsid with the integrin entry receptors would provide more information on the context of the binding. Structures of whole HAdV in complex with their integrin receptors have been determined before improvements in cryoEM resolution (166,167), and the current techniques could provide more detailed, higher-resolution results. However, an issue in the case of the HAdV capsid could be the need for events that precede integrin binding. The initial interaction of the fibre knob with an attachment receptor, associated with the initiation of the dissociation of the fibre from the penton base (258), is presumed to be finalised by the “untwisting” of the penton base within the pentamer on integrin binding, loosening the interaction with the shaft (125,166). The continued presence of the fibre, due to the lack of this initial interaction, which is stronger than the fibre-penton base one and so thought to play a role in their dissociation (259), could restrict the penton base movements. This could perhaps be solved by pre-incubating the capsid with an attachment receptor. However, this would introduce additional molecular species to the preparation, or the need for another purification step afterwards.

The presence of 12 pentons, one at each HAdV capsid vertex, with 60 binding sites for the integrin in total, introducing 60 rather than five possibilities of occupancy, could also be regarded as an issue. However, this can be addressed during data processing: localised reconstruction (260) protocols allow for extracting from a structure a fragment that does not follow the general symmetry, and processing it separately. Thus, icosahedral symmetry could be applied to the whole capsid, then the vertices could be refined without imposing a higher symmetry. Still, the original problem of up to five different occupancies would likely remain and need to be addressed at the levels of 2D and 3D classifications, increasing the challenge.

4.2.4. Outlook

The project has not met its initial aims, namely the determination of the stoichiometries and high-resolution structure of the HAdV penton base-integrin $\alpha_v\beta_3$ complex. The production of one of its components has proven to be intertwined with a different question, that of the mechanism of multimer formation by HAdV penton base. This was investigated with a practical application in mind, looking for approaches that would result in homogenous penton base preparations, rather than to provide a scientifically robust explanation for the observations on its multimeric states under different conditions. Still, the major conclusions provide protocols to reliably produce the two

complex components in insect cells, albeit at low concentrations, and suggest an interaction occurs between them in solution. These results indicate a range of possible further steps to improve sample quality and enable stoichiometry determination and structure solution, leading to novel, publishable findings.

Acknowledgments

If it takes a village to raise a child, it probably takes something like a small German university town to finish a doctoral thesis. Here, I would like to thank some of its literal and metaphorical inhabitants for helping me arrive at this milestone.

My supervisor, Professor Thilo Stehle, has had faith in me from the beginning, giving me this opportunity and guiding me through it with scientific rigour, patience, good humour, making sure I made the most of it.

Professor Sarah Butcher has introduced me to CryoEM and supported me scientifically and otherwise, during my stay in Helsinki and away. Her wisdom, kindness, and time have been invaluable.

I would also like to thank Professor Dirk Schwarzer for kindly agreeing to review this thesis.

The kinases project would not have been successful without Dr Georg Zocher's enormous knowledge, hard work, and the willingness to share his expertise with a beginner like me. I would also like to thank everyone who has worked on the project before me, particularly Karolin Gogler and Phillipp Fink, and Professor Christoph Mayer and his group.

I would also like to thank all the members, past and present, of the Stehle lab for a remarkable working environment, which has shown me kindness and support over the years. I have learned so much about biochemistry from so many people here. I am especially grateful to Dr Michael Strebl for teaching me protein purification from scratch with patience and good humour.

The University of Helsinki Structural Virology lab welcomed me for an unexpectedly COVID-extended secondment on CryoEM and gladly shared their expertise. In particular, I'd like to thank Dr Aušra Domanska and Dr Justin Flatt for discussion on data processing; Dr Maria Anastasina for prolonged access to the computing cluster; and Dr Pasi Laurinmäki, Dr Benham Lak, and Benita Löflund for sample preparation and data collection.

I would like to thank the EU Horizon 2020 programme for three years of funding within the project ViBrANT; the Swiss Light Source for synchrotron time, and the Finnish Centre for Scientific Computing for computational resources.

I am grateful to Professor Niklas Arnberg and Dr Annasara Lenman, who have kindly provided the wild type constructs of the HAdV penton base and information about their purification, and to Dr Katharina Hipp and Ms Rebecca Stahl for negative stain EM of the penton bases.

I'd like to thank my high school teachers, Monika Koblak, Dr Kuba Mikoszewski, and Ms Ewa Wink-Zapalska, for enabling me to organise the remote school visits and their encouragement and faith in me.

Among the many people that I met during my time here, with whom I have shared good times and who have supported me with conversations, kindness and compassion, I'd like to particularly thank: Jasmin Freytag, for help in a myriad ways when I needed it most; Katja Vonmetz for near-infinite patience, proofreading for German (including the abstract of this thesis) and English, and excellent plant-sitting; Dr Marie Sorin for all the travel and conversations; Lea Hansen-Palmus for her companionship, help with

moving and countless trips; Dr Natascha Bartlick for her sense of humour and welcoming me into the lab; Nina Pfahler for her care and inspiring energy, Katie Weimer for her friendship and wisdom; my fellow ViBrANT ESRs: Sarah, Ina, Diana, Arno, and Theis; Dr Mumina Javed for all the time spent together and discussions.
The non-biologists in my life, for years of friendship, intellectual companionship, and support I have received from them: Eleanor (and the Victorian ferns!), Sarah C., Marissa, Magda; and, of course, the people whom I have met at the schools in Raszyńska street and who have stood by me for well more than a decade: Ola “Chudzo”, Wanda, Karolina, Pola, Piotrek, and, naturally, Michał, ready to discuss not just troubleshooting computing clusters at 1 am, but also to analyse and solve any problem real or theoretical. I am so much less alone thanks to all of you.
My family, especially my uncle Darek for his continued interest and support.

And, finally and most obviously, to my mum, who is the best mother I can imagine and whose unconditional love, support, patience, sense of humour, and wisdom make everything else worth it. Nie zaprzeczaj, bo teraz będę mieć papiery, żeby wiedzieć lepiej, nawet jeśli twoich nigdy nie pobiją.

I apologise to anyone I might have forgotten. This has been a long, hard endeavour, which has taught me so much, and I am very grateful to everyone who has helped me throughout.

Reader, I crystallised it.

- Aleksandra/Ola/Aleks

References

1. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci USA. 1990 Jun;87(12):4576–9.
2. Williams TA, Cox CJ, Foster PG, Szöllősi GJ, Embley TM. Phylogenomics provides robust support for a two-domains tree of life. Nature Ecology & Evolution. 2019 Dec 9;4(1):138–47.
3. Baron S, editor. Introduction to Bacteriology. In: Medical Microbiology. Galveston, TX: University of Texas Medical Branch at Galveston; 1996.
4. Silhavy TJ, Kahne D, Walker S. The Bacterial Cell Envelope. Cold Spring Harbor Perspectives in Biology. 2010 May 1;2(5):a000414–a000414.
5. Carvalho AC, Migliori GB, Cirillo DM. Tuberculosis in Europe: a problem of drug resistance or much more? Expert Review of Respiratory Medicine. 2010 Apr 9;4(2):189–200.
6. World Health Organisation. Global tuberculosis report 2021 [Internet]. Geneva: World Health Organisation; 2021. Available from: http://apps.who.int/bookorders.
7. Troeger C, Forouzanfar M, Rao PC, Khalil I, Brown A, Swartz S, et al. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory tract infections in 195 countries: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet Infectious Diseases. 2017 Nov 1;17(11):1133–61.
8. Troeger C, Forouzanfar M, Rao PC, Khalil I, Brown A, Reiner RC, et al. Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet Infectious Diseases. 2017 Sep 1;17(9):909–48.
9. Lee AS, De Lencastre H, Garau J, Kluytmans J, Malhotra-Kumar S, Peschel A, et al. Methicillin-resistant *Staphylococcus aureus*. Nat Rev Dis Primers. 2018 May 31;4(1):18033.
10. Karakonstantis S, Kritsotakis EI, Gikas A. Pandrug-resistant gram-negative bacteria: A systematic review of current epidemiology, prognosis and treatment options. Journal of Antimicrobial Chemotherapy. 2020 Feb 1;75(2):271–82.
11. Murray CJ, Ikuta KS, Sharara F, Swetschinski L, Robles Aguilar G, Gray A, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. The Lancet. 2022 Feb 12;399(10325):629–55.
12. Duerden B, Fry C, Johnson AP, Wilcox MH. The control of methicillin-resistant *Staphylococcus aureus* blood stream infections in England. Open Forum Infectious Diseases. 2015 Apr 1;2(2).
13. Bush K, Bradford PA. β-lactams and β-lactamase inhibitors: An overview. Cold Spring Harbor Perspectives in Medicine. 2016 Aug 1;6(8).
14. Dik DA, Fisher JF, Mobashery S. Cell-Wall Recycling of the Gram-Negative Bacteria and the Nexus to Antibiotic Resistance. Chemical Reviews. 2018 Jun 27;118(12):5952–84.
15. Vollmer W, Bertsche U. Murein (peptidoglycan) structure, architecture and biosynthesis in *Escherichia coli*. Biochimica et Biophysica Acta - Biomembranes. 2008 Sep;1778(9):1714–34.
16. Fisher JF, Mobashery S. Constructing and deconstructing the bacterial cell wall. Protein Science. 2020 Mar 1;29(3):629–46.
17. Vollmer W. Structural variation in the glycan strands of bacterial peptidoglycan. FEMS Microbiology Reviews. 2008 Mar;32(2):287–306.

18. Vollmer W, Blanot D, De Pedro MA. Peptidoglycan structure and architecture. FEMS Microbiology Reviews. 2008 Mar;32(2):149–67.
19. Barreteau H, Kovač A, Boniface A, Sova M, Gobec S, Blanot D. Cytoplasmic steps of peptidoglycan biosynthesis. FEMS Microbiology Reviews. 2008 Mar;32(2):168–207.
20. Dhar S, Kumari H, Balasubramanian D, Mathee K. Cell-wall recycling and synthesis in *Escherichia coli* and *Pseudomonas aeruginosa* – Their role in the development of resistance. Journal of Medical Microbiology. 2018 Jan 1;67(1):1–21.
21. Gisin J, Schneider A, Nägele B, Borisova M, Mayer C. A cell wall recycling shortcut that bypasses peptidoglycan de novo biosynthesis. Nature Chemical Biology. 2013;9(8):491–3.
22. Borisova M, Gisin J, Mayer C. Blocking Peptidoglycan Recycling in *Pseudomonas aeruginosa* Attenuates Intrinsic Resistance to Fosfomycin. Microbial Drug Resistance. 2014 Jun;20(3):231–7.
23. Tanner ACR, Izard J. *Tannerella forsythia*, a periodontal pathogen entering the genomic era. Periodontology 2000. 2006;42(1):88–113.
24. Tanner ACR, Listgarten MA, Ebersole JL, Strzempko MN. Bacteroides forsythus sp. nov., a Slow-Growing, Fusiform *Bacteroides sp*. from the Human Oral Cavity. International Journal of Systematic Bacteriology. 1986 Apr 1;36(2):213–21.
25. Wood N, Johnson RB. Recovery of periodontopathogenic bacteria from embalmed human cadavers. Clinical Anatomy. 2005;18(1):64–7.
26. Piekut T, Hurla M, Banaszek N, Szejn P, Dorszewska J, Kozubski W, et al. Infectious agents and Alzheimer's disease. Journal of Integrative Neuroscience. 2022 Mar 1;21(2).
27. Philips A, Stolarek I, Handschuh L, Nowis K, Juras A, Trzciński D, et al. Analysis of oral microbiome from fossil human remains revealed the significant differences in virulence factors of modern and ancient Tannerella forsythia. BMC Genomics. 2020 Jun 15;21(1).
28. Fagernäs Z, Salazar-García DC, Haber Uriarte M, Avilés Fernández A, Henry AG, Lomba Maurandi J, et al. Understanding the microbial biogeography of ancient human dentitions to guide study design and interpretation. FEMS Microbes. 2022 Mar 17;3.
29. Bravo-Lopez M, Villa-Islas V, Rocha Arriaga C, Villasenõr-Altamirano AB, Guzmán-Solís A, Sandoval-Velasco M, et al. Paleogenomic insights into the red complex bacteria *Tannerella forsythia* in Pre-Hispanic and Colonial individuals from Mexico: T. forsythia aDNA in Mexico. Philosophical Transactions of the Royal Society B: Biological Sciences. 2020;375(1812).
30. Darveau RP. Periodontitis: A polymicrobial disruption of host homeostasis. Nature Reviews Microbiology. 2010 Jun;8(7):481–90.
31. Kassebaum NJ, Bernabé E, Dahiya M, Bhandari B, Murray CJL, Marcenes W. Global burden of severe periodontitis in 1990-2010: A systematic review and meta-regression. Journal of Dental Research. 2014 Nov 11;93(11):1045–53.
32. Xiao L, Zhang Q, Peng Y, Wang D, Liu Y. The effect of periodontal bacteria infection on incidence and prognosis of cancer: A systematic review and meta-analysis. Medicine. 2020 Apr 1;99(15):e19698.
33. Abdelbary MMH, Schittenhelm F, Yekta-Michael SS, Reichert S, Schulz S, Kasaj A, et al. Impact of Three Nonsurgical, Full-Mouth Periodontal Treatments on Total Bacterial Load and Selected Pathobionts. Antibiotics. 2022 May 19;11(5):686.
34. Sedghi LM, Bacino M, Kapila YL. Periodontal Disease: The Good, The Bad, and The Unknown. Frontiers in Cellular and Infection Microbiology. 2021 Dec 7;11.

35. *Porphyromonas gingivalis*, *Treponema denticola*, and *Tannerella forsythia* the 'red complex', a prototype polybacterial pathogenic consortium in periodontitis.
36. LKent Jr R. RL: Microbial complexes in subgingival plaque. Vol. 25, J Clin Periodontol. 1998 p. 134–44.
37. Hajishengallis G, Lamont RJ. Beyond the red complex and into more complexity: The polymicrobial synergy and dysbiosis (PSD) model of periodontal disease etiology. Molecular Oral Microbiology. 2012 Dec;27(6):409–19.
38. Wyss C. Dependence of proliferation of *Bacteroides forsythus* on exogenous N-acetylmuramic acid. Infection and Immunity. 1989;57(6):1757–9.
39. Friedrich V, Pabinger S, Chen T, Messner P, Dewhirst FE, Schäffer C. Draft genome sequence of *Tannerella forsythia* type strain ATCC 43037. Genome Announcements. 2015;3(3).
40. Ruscitto A, Sharma A. Peptidoglycan synthesis in *Tannerella forsythia*: Scavenging is the modus operandi. Molecular Oral Microbiology. 2018 Apr 1;33(2):125–32.
41. Hottmann I, Mayer VMT, Tomek MB, Friedrich V, Calvert MB, Titz A, et al. N-Acetylmuramic Acid (MurNAc) Auxotrophy of the Oral Pathogen *Tannerella forsythia*: Characterization of a MurNAc Kinase and Analysis of Its Role in Cell Wall Metabolism. Frontiers in Microbiology. 2018 Jan 26;9(JAN):1–12.
42. Hudspeth MK, Hunt Gerardo S, Maiden MFJ, Citron DM, Goldstein EJC. Characterization of *Bacteroides forsythus* strains from cat and dog bite wounds in humans and comparison with monkey and human oral strains. Journal of Clinical Microbiology. 1999;37(6):2003–6.
43. CoilID DA, Jospin G, Darling AE, Wallis C, Davis IJ, Harris S, et al. Genomes from bacteria associated with the canine oral cavity: A test case for automated genome-based taxonomic assignment. PLoS ONE. 2019 Jun 1;14(6).
44. Ruscitto A, Hottmann I, Stafford GP, Schäffer C, Mayer C, Sharma A. Identification of a novel N-acetylmuramic acid transporter in *Tannerella forsythia*. Journal of Bacteriology. 2016;198(22):3119–25.
45. Hottmann I, Borisova M, Schäffer C, Mayer C. Peptidoglycan Salvage Enables the Periodontal Pathogen *Tannerella forsythia* to Survive within the Oral Microbial Community. Microbial Physiology. 2021;31(2):123–34.
46. Roy S, Douglas CWI, Stafford GP. A novel sialic acid utilization and uptake system in the periodontal pathogen *Tannerella forsythia*. Journal of Bacteriology. 2010 May;192(9):2285–93.
47. Hottmann I. Cell wall metabolism of the N-acetylmuramic acid auxotrophic oral pathogen Tannerella forsythia. University of Tübingen; 2018.
48. Gogler K. Structural characterization of bacterial cell wall recycling kinases. University of Tübingen; 2018.
49. Skalka AM, Flint J, Rall GF, Racaniello VR. Principles of Virology, Volume I: Molecular Biology [Internet]. Washington, DC: American Society of Microbiology; 2015. Available from: http://www.asmscience.org/content/book/10.1128/9781555819521
50. Moreira D, López-García P. Ten reasons to exclude viruses from the tree of life. Nature Reviews Microbiology. 2009 Apr 9;7(4):306–11.
51. Harris HMB, Hill C. A Place for Viruses on the Tree of Life. Frontiers in Microbiology. 2021;11(January):1–16.
52. Forterre P. Defining Life: The Virus Viewpoint. Origins of Life and Evolution of Biospheres. 2010;40(2):151–60.

53. Dupré J, Guttinger S. Viruses as living processes. Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences. 2016;59:109–16.
54. Koonin EV. Defining life: An exercise in semantics or a route to biological insights? Journal of Biomolecular Structure and Dynamics. 2012;29(4):603–5.
55. Koonin EV, Starokadomskyy P. Are viruses alive? The replicator paradigm sheds decisive light on an old but misguided question. Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences. 2016;59:125–34.
56. Szostak JW. Attempts to define life do not help to understand the origin of life. Journal of Biomolecular Structure and Dynamics. 2012;29(4):599–600.
57. Harrach B, Tarján ZL, Benkő M. Adenoviruses across the animal kingdom: a walk in the zoo. FEBS Letters. 2019;593(24):3660–73.
58. Benko M, Aoki K, Arnberg N, Davison AJ, Echavarria M, Hess M, et al. ICTV Virus Taxonomy Profile: *Adenoviridae* 2022. Journal of General Virology. 2022;103(3):1–2.
59. Rowe WP, Huebner RJ, Gilmore LK, Parrott RH, Ward TG. Isolation of a Cytopathogenic Agent from Human Adenoids Undergoing Spontaneous Degeneration in Tissue Culture. Experimental Biology and Medicine. 1953 Dec 1;84(3):570–3.
60. Aoki K, Benkö M, Davison AJ, Echavarria M, Erdman DD, Harrach B, et al. Toward an Integrated Human Adenovirus Designation System That Utilizes Molecular and Serological Data and Serves both Clinical and Fundamental Virology. Journal of Virology. 2011 Jun;85(11):5703–4.
61. HAdV Working Group. Criteria for a New HAdV Type [Internet]. [cited 2023 Feb 14]. Available from: http://hadvwg.gmu.edu/index.php/criteria-for-a-new-hadv-type/
62. Sallard E, Zhang W, Aydin M, Schröer K, Ehrhardt A. The Adenovirus Vector Platform: Novel Insights into Rational Vector Design and Lessons Learned from the COVID-19 Vaccine. Viruses. 2023;15(1):204.
63. Armitage C. The high burden of infectious disease. Nature. 2021 Oct 28;598(7882):S9–S9.
64. Robine JM. By the numbers: counting the costs of infectious illness. Nature. 2021 Oct 28;598(7882):S18–9.
65. Naghavi M, Abajobir AA, Abbafati C, Abbas KM, Abd-Allah F, Abera SF, et al. Global, regional, and national age-sex specifc mortality for 264 causes of death, 1980-2016: A systematic analysis for the Global Burden of Disease Study 2016. The Lancet. 2017;390(10100):1151–210.
66. Fronteira I, Sidat M, Magalhães JP, de Barros FPC, Delgado AP, Correia T, et al. The SARS-CoV-2 pandemic: A syndemic perspective. One Health. 2021;12.
67. De Cock KM, Jaffe HW, Curran JW. The evolving epidemiology of HIV/AIDS. Aids. 2012;26(10):1205–13.
68. Hsieh YC, Wu TZ, Liu DP, Shao PL, Chang LY, Lu CY, et al. Influenza pandemics: Past, present and future. Journal of the Formosan Medical Association. 2006;105(1):1–6.
69. UNAIDS. Fact Sheet - World AIDS Day 2021 [Internet]. 2021 [cited 2022 Jun 23]. Available from: aidsinfo.unaids.org.
70. Cassini A, Colzani E, Pini A, Mangen MJJ, Plass D, McDonald SA, et al. Impact of infectious diseases on population health using incidence-based disability-adjusted life years (DALYs): Results from the burden of communicable diseases in Europe study, European Union and European economic countries, 2009 to 2013. Eurosurveillance. 2018;23(16):1–20.

71. Wang H, Abajobir AA, Abate KH, Abbafati C, Abbas KM, Abd-Allah F, et al. Global, regional, and national under-5 mortality, adult mortality, age-specific mortality, and life expectancy, 1970-2016: A systematic analysis for the Global Burden of Disease Study 2016. The Lancet. 2017;390(10100):1084–150.
72. Fendrick AM, Monto AS, Nightengale B, Sarnes M. The economic burden of non-influenza-related viral respiratory tract infection in the United States. Archives of Internal Medicine. 2003;163(4):487–94.
73. Englund J, Feuchtinger T, Ljungman P. Viral Infections in Immunocompromised Patients. Biology of Blood and Marrow Transplantation. 2011;17(1 SUPPL):S2–5.
74. Tao L, Fill MMA, Banerjee R, Humphries RM. Notes from the Field: Cluster of Parechovirus Central Nervous System Infections in Young Infants — Tennessee, 2022. MMWR Morbidity and Mortality Weekly Report. 2022;71(30):977–8.
75. Temmam S, Vongphayloth K, Baquero E, Munier S, Bonomi M, Regnault B, et al. Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. Nature. 2022;604(7905):330–6.
76. Thompson DK, Muriel P, Russell D, Osborne P, Bromley A, Rowland M, et al. Economic costs of the foot and mouth disease outbreak in the United Kingdom in 2001. Revue Scientifique et Technique de l'OIE. 2002 Dec 1;21(3):675–87.
77. Jones RAC. Global plant virus disease pandemics and epidemics. Plants. 2021;10(2):1–41.
78. Tripathi S, Suzuki JY, Ferreira SA, Gonsalves D. Papaya ringspot virus-P: Characteristics, pathogenicity, sequence variability and control. Molecular Plant Pathology. 2008;9(3):269–80.
79. Lion T. Adenovirus infections in immunocompetent and immunocompromised patients. Clinical Microbiology Reviews. 2014;27(3):441–62.
80. Kajon AE, Moseley JM, Metzgar D, Huong H, Wadleigh A, Ryan MAK, et al. Molecular Epidemiology of Adenovirus Type 4 Infections in US Military Recruits in the Postvaccination Era (1997–2003). J INFECT DIS. 2007 Jul;196(1):67–75.
81. Hoke CH, Snyder CE. History of the restoration of adenovirus type 4 and type 7 vaccine, live oral (Adenovirus Vaccine) in the context of the Department of Defense acquisition system. Vaccine. 2013 Mar;31(12):1623–32.
82. Lee CS, Bishop ES, Zhang R, Yu X, Farina EM, Yan S, et al. Adenovirus-mediated gene delivery: Potential applications for gene and cell-based therapies in the new era of personalized medicine. Genes and Diseases. 2017;4(2):43–63.
83. The Journal of Gene Medicine [Internet]. 2021 [cited 2022 Dec 9]. Gene Therapy Clinical Trials Worldwide. Available from: https://a873679.fmphost.com/fmi/webd/GTCT
84. Bulcha JT, Wang Y, Ma H, Tai PWL, Gao G. Viral vector platforms within the gene therapy landscape. Sig Transduct Target Ther. 2021 Feb 8;6(1):53.
85. Daussy CF, Pied N, Wodrich H. Understanding post entry sorting of adenovirus capsids; a chance to change vaccine vector properties. Viruses. 2021;13(7).
86. Milligan ID, Gibani MM, Sewell R, Clutterbuck EA, Campbell D, Plested E, et al. Safety and immunogenicity of novel adenovirus type 26-and modified vaccinia Ankara-vectored Ebola vaccines: A randomized clinical trial. JAMA - Journal of the American Medical Association. 2016;315(15):1610–23.
87. Sadoff J, Gray G, Vandebosch A, Cárdenas V, Shukarev G, Grinsztejn B, et al. Safety and Efficacy of Single-Dose Ad26.COV2.S Vaccine against Covid-19. New England Journal of Medicine. 2021 Jun 10;384(23):2187–201.
88. Logunov DY, Dolzhikova IV, Shcheblyakov DV, Tukhvatulin AI, Zubkova OV, Dzharullaeva AS, et al. Safety and efficacy of an rAd26 and rAd5 vector-based

heterologous prime-boost COVID-19 vaccine: an interim analysis of a randomised controlled phase 3 trial in Russia. The Lancet. 2021;397(10275):671–81.
89. European Medicines Agency. Zabdeno [Internet]. 2020 [cited 2023 Feb 14]. Available from: https://www.ema.europa.eu/en/medicines/human/EPAR/zabdeno
90. Food and Drug Administration. FDA approves first adenoviral vector-based gene therapy for high-risk Bacillus Calmette-Guérin unresponsive non-muscle invasive bladder cancer [Internet]. 2022 [cited 2023 Feb 14]. Available from: https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-first-adenoviral-vector-based-gene-therapy-high-risk-bacillus-calmette-guerin
91. Zhang WW, Li L, Li D, Liu J, Li X, Li W, et al. The First Approved Gene Therapy Product for Cancer Ad-p53 (Gendicine): 12 Years in the Clinic. Human Gene Therapy. 2018;29(2):160–79.
92. Liang M. Oncorine, the World First Oncolytic Virus Medicine and its Update in China. Current Cancer Drug Targets. 2018 Jan 15;18(2):171–6.
93. Di Paolo NC, van Rooijen N, Shayakhmetov DM. Redundant and synergistic mechanisms control the sequestration of blood-born adenovirus in the liver. Molecular Therapy. 2009;17(4):675–84.
94. Fausther-Bovendo H, Kobinger GP. Pre-existing immunity against Ad vectors: Humoral, cellular, and innate response, what's important? Human Vaccines and Immunotherapeutics. 2014 Oct 1;10(10):2875–84.
95. Mast TC, Kierstead L, Gupta SB, Nikas AA, Kallas EG, Novitsky V, et al. International epidemiology of human pre-existing adenovirus (Ad) type-5, type-6, type-26 and type-36 neutralizing antibodies: Correlates of high Ad5 titers and implications for potential HIV vaccine trials. Vaccine. 2010 Jan 22;28(4):950–7.
96. Alonso-Padilla J, Papp T, Kaján GL, Benko M, Havenga M, Lemckert A, et al. Development of novel adenoviral vectors to overcome challenges observed with HAdV-5-based constructs. Molecular Therapy. 2016;24(1):6–16.
97. Raper SE, Chirmule N, Lee FS, Wivel NA, Bagg A, Gao GP, et al. Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer. Molecular Genetics and Metabolism. 2003;80(1–2):148–58.
98. Zhao Z, Anselmo AC, Mitragotri S. Viral vector-based gene therapies in the clinic. Bioengineering and Translational Medicine. 2022;7(1):1–20.
99. See I, Lale A, Marquez P, Streiff MB, Wheeler AP, Tepper NK, et al. Case Series of Thrombosis With Thrombocytopenia Syndrome After COVID-19 Vaccination—United States, December 2020 to August 2021. Annals of Internal Medicine. 2022 Apr 1;175(4):513–22.
100. Harvey BG, Maroni J, O'Donoghue KA, Chu KW, Muscat JC, Pippo AL, et al. Safety of Local Delivery of Low- and Intermediate-Dose Adenovirus Gene Transfer Vectors to Individuals with a Spectrum of Morbid Conditions. Human Gene Therapy. 2002 Jan;13(1):15–63.
101. Venier LM, Clerici B, Bissola AL, Modi D, Jevtic SD, Radford M, et al. Unique features of vaccine-induced immune thrombotic thrombocytopenia; a new anti-platelet factor 4 antibody-mediated disorder. International Journal of Hematology. 2022;
102. Cines DB, Bussel JB. SARS-CoV-2 Vaccine–Induced Immune Thrombotic Thrombocytopenia. New England Journal of Medicine. 2021 Jun 10;384(23):2254–6.
103. Gao W, Robbins PD, Gambotto A. Human adenovirus type 35: Nucleotide sequence and vector development. Gene Therapy. 2003;10(23):1941–9.

104. Someya K, Xin KQ, Ami Y, Izumi Y, Mizuguchi H, Ohta S, et al. Chimeric adenovirus type 5/35 vector encoding SIV gag and HIV env genes affords protective immunity against the simian/human immunodeficiency virus in monkeys. Virology. 2007;367(2):390–7.
105. Voysey M, Clemens SAC, Madhi SA, Weckx LY, Folegatti PM, Aley PK, et al. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. The Lancet. 2021 Jan 9;397(10269):99–111.
106. Schmid M, Ernst P, Honegger A, Suomalainen M, Zimmermann M, Braun L, et al. Adenoviral vector with shield and adapter increases tumor specificity and escapes liver and immune control. Nature Communications. 2018;9(1):1–16.
107. Maginnis MS. Virus–Receptor Interactions: The Key to Cellular Invasion. Journal of Molecular Biology. 2018;430(17):2590–611.
108. Marsh MP, Campos SK, Baker ML, Chen CY, Chiu W, Barry MA. Cryoelectron Microscopy of Protein IX-Modified Adenoviruses Suggests a New Position for the C Terminus of Protein IX. Journal of Virology. 2006;80(23):11881–6.
109. Neu U, Maginnis MS, Palma AS, Ströh LJ, Nelson CDS, Feizi T, et al. Structure-function analysis of the human JC polyomavirus establishes the LSTc pentasaccharide as a functional receptor motif. Cell Host and Microbe. 2010;8(4):309–19.
110. Belnap DM, McDermott BM, Filman DJ, Cheng N, Trus BL, Zuccola HJ, et al. Three-dimensional structure of poliovirus receptor bound to poliovirus. Proceedings of the National Academy of Sciences of the United States of America. 2000;97(1):73–8.
111. Ou T, Mou H, Zhang L, Ojha A, Choe H, Farzan M. Hydroxychloroquine-mediated inhibition of SARS-CoV-2 entry is attenuated by TMPRSS2. PLoS Pathogens. 2021;17(1 January):1–15.
112. Jamjian MC, McNicholl IR. Enfuvirtide: First fusion inhibitor for treatment of HIV infection. American Journal of Health-System Pharmacy. 2004;61(12):1242–7.
113. Dorr P, Westby M, Dobbs S, Griffin P, Irvine B, Macartney M, et al. Maraviroc (UK-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor CCR5 with broad-spectrum anti-human immunodeficiency virus type 1 activity. Antimicrobial Agents and Chemotherapy. 2005;49(11):4721–32.
114. Lenman A, Liaci AM, Liu Y, Frängsmyr L, Frank M, Blaum BS, et al. Polysialic acid is a cellular receptor for human adenovirus 52. Proceedings of the National Academy of Sciences. 2018;115(18):E4264–73.
115. Martin NT, Wrede C, Niemann J, Brooks J, Schwarzer D, Kühnel F, et al. Targeting polysialic acid-abundant cancers using oncolytic adenoviruses with fibers fused to active bacteriophage borne endosialidase. Biomaterials. 2018;158:86–94.
116. Castello R, Borzone R, D'Aria S, Annunziata P, Piccolo P, Brunetti-Pierri N. Helper-dependent adenoviral vectors for liver-directed gene therapy of primary hyperoxaluria type 1. Gene Therapy. 2016;23(2):129–34.
117. Sharma PK, Dmitriev IP, Kashentseva EA, Raes G, Li L, Kim SW, et al. Development of an adenovirus vector vaccine platform for targeting dendritic cells. Cancer Gene Therapy. 2018 Feb 15;25(1–2):27–38.
118. Arnberg N. Adenovirus receptors: implications for targeting of viral vectors. Trends in Pharmacological Sciences. 2012 Aug;33(8):442–8.
119. Stasiak AC, Stehle T. Human adenovirus binding to host cell receptors: a structural view. Medical Microbiology and Immunology. 2020 Jun 29;209(3):325–33.

120. Nilsson EC, Storm RJ, Bauer J, Johansson SMC, Lookene A, Ångström J, et al. The GD1a glycan is a cellular receptor for adenoviruses causing epidemic keratoconjunctivitis. Nature Medicine. 2011;17(1):105–9.
121. Bewley MC. Structural Analysis of the Mechanism of Adenovirus Binding to Its Human Cellular Receptor, CAR. Science. 1999 Nov 19;286(5444):1579–83.
122. Cupelli K, Müller S, Persson BD, Jost M, Arnberg N, Stehle T. Structure of Adenovirus Type 21 Knob in Complex with CD46 Reveals Key Differences in Receptor Contacts among Species B Adenoviruses. Journal of Virology. 2010 Apr 1;84(7):3189–200.
123. Vassal-Stermann E, Effantin G, Zubieta C, Burmeister W, Iseni F, Wang H, et al. CryoEM structure of adenovirus type 3 fibre with desmoglein 2 shows an unusual mode of receptor engagement. Nature Communications. 2019 Dec 12;10(1):1181.
124. Nemerow GR, Stewart PL. Role of α v Integrins in Adenovirus Cell Entry and Gene Delivery. Microbiology and Molecular Biology Reviews. 1999 Sep;63(3):725–34.
125. Nemerow GR, Stewart PL. Insights into adenovirus uncoating from interactions with integrins and mediators of host immunity. Viruses. 2016;8(12):1–15.
126. Greber UF, Flatt JW. Adenovirus Entry: From Infection to Immunity. Annu Rev Virol. 2019 Sep 29;6(1):177–97.
127. Giménez-Roig J, Núñez-Manchón E, Alemany R, Villanueva E, Fillat C. Codon Usage and Adenovirus Fitness: Implications for Vaccine Development. Front Microbiol. 2021 Feb 10;12:633946.
128. Hussein HAM, Walker LR, Abdel-Raouf UM, Desouky SA, Montasser AKM, Akula SM. Beyond RGD: virus interactions with integrins. Archives of Virology. 2015;160(11):2669–81.
129. Xiong J ping, Stehle T, Diefenbach B, Zhang R, Scott DL, Joachimiak A, et al. Crystal Structure of the Extracellular Segment of Integrin αVβ3. Science. 2001;294(5541):339–45.
130. Borst AJ, James ZM, Zagotta WN, Ginsberg M, Rey FA, DiMaio F, et al. The Therapeutic Antibody LM609 Selectively Inhibits Ligand Binding to Human αVβ3 Integrin via Steric Hindrance. Structure. 2017;25(11):1732-1739.e5.
131. Cormier A, Campbell MG, Ito S, Wu S, Lou J, Marks J, et al. Cryo-EM structure of the αvβ8 integrin reveals a mechanism for stabilizing integrin extension. Nat Struct Mol Biol. 2018 Aug;25(8):698–704.
132. Hynes RO. Integrins. Cell. 2002 Sep;110(6):673–87.
133. Takagi J, Petre BM, Walz T, Springer TA. Global conformational earrangements in integrin extracellular domains in outside-in and inside-out signaling. Cell. 2002;110(5):599–611.
134. Mathias P, Wickham T, Moore M, Nemerow G. Multiple adenovirus serotypes use alpha v integrins for infection. Journal of virology. 1994;68(10):6811–4.
135. Li E, Brown SL, Stupack DG, Puente XS, Cheresh DA, Nemerow GR. Integrin αvβ1 Is an Adenovirus Coreceptor. Journal of Virology. 2001 Jun;75(11):5405–9.
136. Huang S, Kamata T, Takada Y, Ruggeri ZM, Nemerow GR. Adenovirus interaction with distinct integrins mediates separate events in cell entry and gene delivery to hematopoietic cells. Journal of Virology. 1996;70(7):4502–8.
137. Mathias P, Galleno M, Nemerow GR. Interactions of Soluble Recombinant Integrin αvβ5 with Human Adenoviruses. Journal of Virology. 1998 Nov 1;72(11):8669–75.
138. Rajan A, Persson BD, Frängsmyr L, Olofsson A, Sandblad L, Heino J, et al. Enteric Species F Human Adenoviruses use Laminin-Binding Integrins as Co-Receptors for Infection of Ht-29 Cells. Scientific Reports. 2018;8(1):1–14.

139. The {PyMOL} Molecular Graphics System, Version~2.5. New York: Schrödinger, LLC; 2021.
140. Kotecha A, Wang Q, Dong X, Ilca SL, Ondiviela M, Zihe R, et al. Rules of engagement between αvβ6 integrin and foot-and-mouth disease virus. Nature Communications. 2017;8(May).
141. Seitsonen J, Susi P, Heikkilä O, Sinkovits RS, Laurinmäki P, Hyypiä T, et al. Interaction of alphaVbeta3 and alphaVbeta6 integrins with human parechovirus 1. Journal of virology. 2010;84(17):8509–19.
142. Shakeel S, Seitsonen JJT, Kajander T, Laurinmaki P, Hyypia T, Susi P, et al. Structural and Functional Analysis of Coxsackievirus A9 Integrin αvβ6 Binding and Uncoating. Journal of Virology. 2013;87(7):3943–51.
143. Prasad BVV, Schmid MF. Principles of Virus Structural Organization. In: Rossmann MG, Rao VB, editors. Viral Molecular Machines [Internet]. Boston, MA: Springer US; 2012 [cited 2023 Sep 11]. p. 17–47. (Advances in Experimental Medicine and Biology; vol. 726). Available from: http://link.springer.com/10.1007/978-1-4614-0980-9_3
144. San Martín C. Latest Insights on Adenovirus Structure and Assembly. Viruses. 2012 May 21;4(5):847–77.
145. Yu X, Veesler D, Campbell MG, Barry ME, Asturias FJ, Barry MA, et al. Cryo-EM structure of human adenovirus D26 reveals the conservation of structural organization among human adenoviruses. Science Advances. 2017;3(5):1–12.
146. Kundhavai Natchiar S, Venkataraman S, Mullen TM, Nemerow GR, Reddy VS. Revised Crystal Structure of Human Adenovirus Reveals the Limits on Protein IX Quasi-Equivalence and on Analyzing Large Macromolecular Complexes. Journal of Molecular Biology. 2018;430(21):4132–41.
147. Abrishami V, Ilca SL, Gomez-Blanco J, Rissanen I, De La Rosa-Trevín JM, Reddy VS, et al. Localized reconstruction in Scipion expedites the analysis of symmetry mismatches in cryo-EM data. Progress in Biophysics and Molecular Biology. 2021 Mar;160:43–52.
148. Henry LJ, Xia D, Wilke ME, Deisenhofer J, Gerard RD. Characterization of the knob domain of the adenovirus type 5 fiber protein expressed in *Escherichia coli*. Journal of virology. 1994 Aug;68(8):5239–46.
149. van Raaij MJ, Mitraki A, Lavigne G, Cusack S. A triple β-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. Nature. 1999 Oct;401(6756):935–8.
150. Liu H, Wu L, Zhou ZH. Model of the trimeric fiber and its interactions with the pentameric penton base of human adenovirus by cryo-electron microscopy. Journal of Molecular Biology. 2011;406(5):764–74.
151. Zubieta C, Schoehn G, Chroboczek J, Cusack S. The structure of the human adenovirus 2 penton. Molecular Cell. 2005;17(1):121–35.
152. Madisch I, Hofmayer S, Moritz C, Grintzalis A, Hainmueller J, Pring-Akerblom P, et al. Phylogenetic Analysis and Structural Predictions of Human Adenovirus Penton Proteins as a Basis for Tissue-Specific Adenovirus Vector Design. Journal of Virology. 2007;81(15):8270–81.
153. Bai MEI, Harfe B, Freimuth P. 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. 1993;67(9):5198–205.
154. Albinsson B, Kidd AH. Adenovirus type 41 lacks an RGD α(v)-integrin binding motif on the penton base and undergoes delayed uptake in A549 cells. Virus Research. 1999;64(2):125–36.

155. Stewart PL, Burnett RM, Cyrklaff M, Fuller SD. Image reconstruction reveals the complex molecular organization of adenovirus. Cell. 1991;67(1):145–54.
156. van Oostrum J, Burnett RM. Molecular composition of the adenovirus type 2 virion. Journal of Virology. 1985;56(2):439–48.
157. Laver WG, Wrigley NG, Pereira HG. Removal of pentons from particles of adenovirus type 2. Virology. 1969 Nov;39(3):599–605.
158. Norrby E, Skaaret P. Comparison between soluble components of adenovirus types 3 and 16 and of the intermediate strain 3-16 (the San Carlos Agent). Virology. 1968;36(2):201–11.
159. Fender P, Ruigrok RWH, Gout E, Buffet S, Chroboczek J. Adenovirus dodecahedron, a new vector for human gene transfer. Nature Biotechnology. 1997 Jan;15(1):52–6.
160. Szolajska E, Burmeister WP, Zochowska M, Nerlo B, Andreev I, Schoehn G, et al. The Structural Basis for the Integrity of Adenovirus Ad3 Dodecahedron. Verma C, editor. PLoS ONE. 2012 Sep 25;7(9):e46075.
161. Fuschiotti P, Schoehn G, Fender P, Fabry CMS, Hewat EA, Chroboczek J, et al. Structure of the dodecahedral penton particle from human adenovirus type 3. Journal of Molecular Biology. 2006;356(2):510–20.
162. Trotman LC, Achermann DP, Keller S, Straub M, Greber UF. Non-classical export of a adenovirus structural protein. Traffic. 2003;4(6):390–402.
163. Fender P, Hall K, Schoehn G, Blair GE. Impact of Human Adenovirus Type 3 Dodecahedron on Host Cells and Its Potential Role in Viral Infection. Journal of Virology. 2012;86(9):5380–5.
164. Zochowska M, Paca A, Schoehn G, Andrieu JP, Chroboczek J, Dublet B, et al. Adenovirus dodecahedron, as a drug delivery vector. Current Research in Pharmaceutical Technology. 2011;4(5):141–60.
165. Veesler D, Cupelli K, Burger M, Graber P, Stehle T, Johnson JE. Single-particle EM reveals plasticity of interactions between the adenovirus penton base and integrin αVβ3. Proceedings of the National Academy of Sciences. 2014 Jun 17;111(24):8815–9.
166. Lindert S, Silvestry M, Mullen TM, Nemerow GR, Stewart PL. Cryo-Electron Microscopy Structure of an Adenovirus-Integrin Complex Indicates Conformational Changes in both Penton Base and Integrin. Journal of Virology. 2009;83(22):11491–501.
167. Chiu CY, Mathias P, Nemerow GR, Stewart PL. Structure of adenovirus complexed with its internalization receptor, αvβ5 integrin. J Virol. 1999;73(8):6759–68.
168. Wlodawer A, Vondrasek J. INHIBITORS OF HIV-1 PROTEASE: A Major Success of Structure-Assisted Drug Design. Annual Review of Biophysics and Biomolecular Structure. 1998 Jun;27(1):249–84.
169. Itzstein M, Kok GB, Pegg MS, Dyason J, Jin B, Van Phan T, et al. Rational design of potent sialidase-based inhibitors of influenza virus replication. 1993.
170. Egli M. Diffraction techniques in structural biology. Current Protocols in Nucleic Acid Chemistry. 2016;2016:7.13.1-7.13.41.
171. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021 Aug 26;596(7873):583–9.
172. Brooks-Bartlett JC, Garman EF. The Nobel Science: One Hundred Years of Crystallography. Interdisciplinary Science Reviews. 2015 Jul 3;40(3):244–64.
173. Rupp B. Biomolecular crystallography: principles, practice, and application to structural biology. New York: Garland Science; 2010.

174. Chayen NE. Turning protein crystallisation from an art into a science. Current Opinion in Structural Biology. 2004 Oct;14(5):577–83.
175. Nespolo M. Tips and traps on crystal twinning: How to fully describe your twin. Crystal Research and Technology. 2015;50(5):362–71.
176. Pflugrath JW. Practical macromolecular cryocrystallography. Acta Crystallogr F Struct Biol Commun. 2015 Jun 1;71(6):622–42.
177. Ealick SE, Walter RL. Synchrotron beamlines for macromolecular crystallography. Current Opinion in Structural Biology. 1993 Oct;3(5):725–36.
178. Wlodawer A, Minor W, Dauter Z, Jaskolski M. Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. FEBS J. 2013 Nov;280(22):5705–36.
179. Stan C, Beavers C, Kunz M, Tamura N. X-Ray Diffraction under Extreme Conditions at the Advanced Light Source. QuBS. 2018 Jan 23;2(1):4.
180. Dauter Z. Data-collection strategies. Acta Crystallogr D Biol Crystallogr. 1999 Oct 1;55(10):1703–17.
181. Cowtan K. Phase Problem in X-ray Crystallography, and Its Solution. In: Encyclopedia of Life Sciences [Internet]. Wiley; 2003. Available from: https://onlinelibrary.wiley.com/doi/10.1038/npg.els.0002722
182. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP4 suite and current developments. Acta Crystallographica Section D Biological Crystallography. 2011 Apr 1;67(4):235–42.
183. Brown EN, Ramaswamy S. Quality of protein crystal structures. Acta Crystallographica Section D Biological Crystallography. 2007 Sep 1;63(9):941–50.
184. Cheng Y, Grigorieff N, Penczek PA, Walz T. A primer to single-particle cryo-electron microscopy. Cell. 2015;161(3):439–49.
185. Zivanov J, Nakane T, Scheres SHW. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. IUCrJ. 2019;6:5–17.
186. Fernández JJ, Luque D, Castón JR, Carrascosa JL. Sharpening high resolution information in single particle electron cryomicroscopy. Journal of Structural Biology. 2008;164(1):170–5.
187. Nogales E, Scheres SHW. Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. Molecular Cell. 2015;58(4):677–89.
188. Kühlbrandt W. The resolution revolution. Science. 2014;343(6178):1443–4.
189. Nobel Prize Outreach AB. NobelPrize.org. [cited 2023 Feb 21]. The Nobel Prize in Chemistry 2017. Available from: https://www.nobelprize.org/prizes/chemistry/2017/summary/
190. Nakane T, Kotecha A, Sente A, McMullan G, Masiulis S, Brown PMGE, et al. Single-particle cryo-EM at atomic resolution. Nature. 2020 Nov 5;587(7832):152–6.
191. Electron Microscopy Data Bank. Statistics [Internet]. [cited 2023 Feb 22]. Available from: https://www.ebi.ac.uk/emdb/statistics/emdb_resolution_year
192. RCSB PDB. PDB Statistics: PDB Data Distribution by Resolution [Internet]. [cited 2023 Feb 22]. Available from: https://www.rcsb.org/stats/distribution-resolution
193. Lander GC, Glaeser RM. Conquer by cryo-EM without physically dividing. Biochemical Society Transactions. 2021 Nov 1;49(5):2287–98.
194. Henderson R. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. Proceedings of the National Academy of Sciences of the United States of America. 2013 Nov 5;110(45):18037–41.
195. Henderson R, Sali A, Baker ML, Carragher B, Devkota B, Downing KH, et al. Outcome of the first electron microscopy validation task force meeting. In: Structure. 2012. p. 205–14.

196. Lawson CL, Kryshtafovych A, Adams PD, Afonine PV, Baker ML, Barad BA, et al. Cryo-EM model validation recommendations based on outcomes of the 2019 EMDataResource challenge. Nature Methods. 2021 Feb 1;18(2):156–64.
197. Cupelli AK. Structural Investigations of Adenovirus-Host Cell Receptor Interactions and the Cathepsin B Protease from *Trypanosoma brucei*. University of Tübingen; 2011.
198. Liaci AM. Structural and Functional Studies on the Early Steps of Polyomavirus and Adenovirus Life Cycles. University of Tübingen; 2017.
199. McFerrin MB, Snell EH. The development and application of a method to quantify the quality of cryoprotectant solutions using standard area-detector X-ray images. Journal of Applied Crystallography. 2002;35(5):538–45.
200. Kabsch W. XDS. Acta Crystallographica Section D Biological Crystallography. 2010 Feb 1;66(2):125–32.
201. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta crystallographica Section D, Biological crystallography. 2010 Apr;66(Pt 4):486–501.
202. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. REFMAC 5 for the refinement of macromolecular crystal structures. Acta Crystallographica Section D Biological Crystallography. 2011 Apr 1;67(4):355–67.
203. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta crystallographica Section D, Biological crystallography. 2010 Feb;66(Pt 2):213–21.
204. Gore S, Sanz García E, Hendrickx PMS, Gutmanas A, Westbrook JD, Yang H, et al. Validation of Structures in the Protein Data Bank. Structure. 2017;25(12):1916–27.
205. Uhr ML. Coupled enzyme systems: Exploring coupled assays with students. Biochemical Education. 1990;18(1):48–50.
206. Storer AC, Cornish Bowden A. The kinetics of coupled enzyme reactions. Applications to the assay of glucokinase, with glucose 6 phosphate dehydrogenase as coupling enzyme. Biochemical Journal. 1974;141(1):205–9.
207. Invitrogen. Bac-to-Bac® Baculovirus Expression System. 2013;1–78.
208. Zheng SQ, Palovcak E, Armache JP, Verba KA, Cheng Y, Agard DA. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nature Methods. 2017 Apr 27;14(4):331–2.
209. de la Rosa-Trevín JM, Quintana A, del Cano L, Zaldívar A, Foche I, Gutiérrez J, et al. Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. Journal of Structural Biology. 2016;195(1):93–9.
210. Rohou A, Grigorieff N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. Journal of Structural Biology. 2015;192(2):216–21.
211. De la Rosa-Trevín JM, Otón J, Marabini R, Zaldívar A, Vargas J, Carazo JM, et al. Xmipp 3.0: An improved software suite for image processing in electron microscopy. Journal of Structural Biology. 2013;184(2):321–8.
212. Kimanius D, Forsberg BO, Scheres SHW, Lindahl E. Accelerated cryo-EM structure determination with parallelisation using GPUS in RELION-2. eLife. 2016;5:1–21.
213. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. Journal of computational chemistry. 2004 Oct;25(13):1605–12.

214. Scheres SHW, Chen S. Prevention of overfitting in cryo-EM structure determination. Nature Methods. 2012 Sep 29;9(9):853–4.
215. O'Connell KL, Stults JT. Identification of mouse liver proteins on two-dimensional electrophoresis gels by matrix-assisted laser desorption/ionization mass spectrometry ofin situ enzymatic digests. Electrophoresis. 1997;18(3–4):349–59.
216. Krissinel E, Henrick K. Inference of Macromolecular Assemblies from Crystalline State. Journal of Molecular Biology. 2007 Sep;372(3):774–97.
217. Stasiak AC, Gogler K, Borisova M, Fink P, Mayer C, Stehle T, et al. N-acetylmuramic acid recognition by MurK kinase from the MurNAc auxotrophic oral pathogen *Tannerella forsythia*. Journal of Biological Chemistry. 2023 Sep;299(9):105076.
218. Veevers R, Hayward S. Methodological improvements for the analysis of domain movements in large biomolecular complexes. Biophysics and Physicobiology. 2019;16:328–36.
219. Duarte JM, Bliven S, Lafita A, Capitani G, Burley SK. Automated evaluation of quaternary structures from protein crystal structures. Acta Crystallographica Section A Foundations and Advances. 2017;73(a1):a117–a117.
220. Holm L. Dali server: structural unification of protein families. Nucleic Acids Research. 2022 Jul 5;50(W1):W210–5.
221. Stafford CA, Gassauer A, de Oliveira Mann CC, Tanzer MC, Fessler E, Wefers B, et al. Phosphorylation of muramyl peptides by NAGK is required for NOD2 activation. Nature. 2022 Aug 24;
222. Micsonai A, Moussong É, Wien F, Boros E, Vadászi H, Murvai N, et al. BeStSel: webserver for secondary structure and fold prediction for protein CD spectroscopy. Nucleic Acids Research. 2022 Jul 5;50(W1):W90–8.
223. Cheek S, Zhang H, Grishin NV. Sequence and Structure Classification of Kinases. Journal of Molecular Biology. 2002 Jul;320(4):855–81.
224. Stafford CA, Gassauer AM, De Oliveira Mann CC, Tanzer MC, Fessler E, Wefers B, et al. Phosphorylation of muramyl peptides by NAGK is required for NOD2 activation. Nature. 2022 Sep 15;609(7927):590–6.
225. Weihofen WA, Berger M, Chen H, Saenger W, Hinderlich S. Structures of Human N-Acetylglucosamine Kinase in Two Complexes with N-Acetylglucosamine and with ADP/Glucose: Insights into Substrate Specificity and Regulation. Journal of Molecular Biology. 2006;364(3):388–99.
226. Nishimasu H, Fushinobu S, Shoun H, Wakagi T. Crystal structures of an ATP-dependent hexokinase with broad substrate specificity from the hyperthermophilic archaeon *Sulfolobus tokodaii*. Journal of Biological Chemistry. 2007 Mar 30;282(13):9923–31.
227. Sigrell JA, Cameron AD, Mowbray SL. Induced fit on sugar binding activates ribokinase. Journal of Molecular Biology. 1999 Jul;290(5):1009–18.
228. Martinez J, Nguyen LD, Hinderlich S, Zimmer R, Tauberger E, Reutter W, et al. Crystal structures of N-acetylmannosamine kinase provide insights into enzyme activity and inhibition. Journal of Biological Chemistry. 2012;287(17):13656–65.
229. Gangi Setty T, Sarkar A, Coombes D, Dobson RCJ, Subramanian R. Structure and Function of N-Acetylmannosamine Kinases from Pathogenic Bacteria. ACS Omega. 2020 Dec 8;5(48):30923–36.
230. Roy S, Vivoli Vega M, Ames JR, Britten N, Kent A, Evans K, et al. The ROK kinase N-acetylglucosamine kinase uses a sequential random enzyme mechanism with successive conformational changes upon each substrate binding. Journal of Biological Chemistry. 2023 Apr;299(4):103033.

231. Koshland DE. Enzyme flexibility and enzyme action. J Cell Comp Physiol. 1959 Dec;54(S1):245–58.
232. Miyazono K ichi, Tabei N, Morita S, Ohnishi Y, Horinouchi S, Tanokura M. Substrate recognition mechanism and substrate-dependent conformational changes of an ROK family glucokinase from *Streptomyces griseus*. Journal of bacteriology. 2012 Feb;194(3):607–16.
233. Aleshin AE, Kirby C, Liu X, Bourenkov GP, Bartunik HD, Fromm HJ, et al. Crystal structures of mutant monomeric hexokinase I reveal multiple ADP binding sites and conformational changes revelant to allosteric regulation. Journal of Molecular Biology. 2000;296(4):1001–15.
234. Reed MC, Lieb A, Nijhout HF. The biological significance of substrate inhibition: A mechanism with diverse functions. Bioessays. 2010 Apr 22;32(5):422–9.
235. Koskiniemi S, Sun S, Berg OG, Andersson DI. Selection-Driven Gene Loss in Bacteria. Casadesús J, editor. PLoS Genet. 2012 Jun 28;8(6):e1002787.
236. Noda-Garcia L, Romero Romero ML, Longo LM, Kolodkin-Gal I, Tawfik DS. *Bacilli* glutamate dehydrogenases diverged via coevolution of transcription and enzyme regulation. EMBO Reports. 2017 Jul;18(7):1139–49.
237. Ruscitto A, Honma K, Veeramachineni VM, Nishikawa K, Stafford GP, Sharma A. Regulation and molecular basis of environmental muropeptide uptake and utilization in fastidious oral Anaerobe *Tannerella forsythia*. Frontiers in Microbiology. 2017 Apr 12;8(APR).
238. Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, et al. The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. Biochemistry. 2011 May 31;50(21):4402–10.
239. Davidi D, Longo LM, Jabłońska J, Milo R, Tawfik DS. A Bird's-Eye View of Enzyme Evolution: Chemical, Physicochemical, and Physiological Considerations. Chem Rev. 2018 Sep 26;118(18):8786–97.
240. Labourel F, Rajon E. Resource Uptake and the Evolution of Moderately Efficient Enzymes. Ozkan B, editor. Molecular Biology and Evolution. 2021 Aug 23;38(9):3938–52.
241. Newton MS, Arcus VL, Gerth ML, Patrick WM. Enzyme evolution: innovation is easy, optimization is complicated. Current Opinion in Structural Biology. 2018 Feb;48:110–6.
242. Heckmann D, Zielinski DC, Palsson BO. Modeling genome-wide enzyme evolution predicts strong epistasis underlying catalytic turnover rates. Nat Commun. 2018 Dec 10;9(1):5270.
243. Mayer VMT, Tomek MB, Figl R, Borisova M, Hottmann I, Blaukopf M, et al. Utilization of different MurNAc sources by the oral pathogen *Tannerella forsythia* and role of the inner membrane transporter AmpG. BMC Microbiology. 2020;20(1):1–15.
244. Coil DA, Jospin G, Darling AE, Wallis C, Davis IJ, Harris S, et al. Genomes from bacteria associated with the canine oral cavity: A test case for automated genome-based taxonomic assignment. Moreno-Hagelsieb G, editor. PLoS ONE. 2019 Jun 10;14(6):e0214354.
245. Socransky SS, Haffajee AD, Cugini MA, Smith C, Kent RL. Microbial complexes in subgingival plaque. J Clin Periodontol. 1998 Feb;25(2):134–44.
246. Bartold PM, Van Dyke TE. An appraisal of the role of specific bacteria in the initial pathogenesis of periodontitis. J Clin Periodontol. 2019 Jan;46(1):6–11.
247. Sanz M, Herrera D, Kebschull M, Chapple I, Jepsen S, Berglundh T, et al. Treatment of stage I–III periodontitis—The EFP S3 level clinical practice guideline. J Clinic Periodontology. 2020 Jul;47(S22):4–60.

248. Abdulkareem A, Abdulbaqi H, Gul S, Milward M, Chasib N, Alhashimi R. Classic vs. Novel Antibacterial Approaches for Eradicating Dental Biofilm as Adjunct to Periodontal Debridement: An Evidence-Based Overview. Antibiotics. 2021 Dec 22;11(1):9.
249. Hutcherson JA, Sinclair KM, Belvin BR, Gui Q, Hoffman PS, Lewis JP. Amixicile, a novel strategy for targeting oral anaerobic pathogens. Sci Rep. 2017 Sep 5;7(1):10474.
250. Zubieta C, Blanchoin L, Cusack S. Structural and biochemical characterization of a human adenovirus 2/12 penton base chimera. FEBS Journal. 2006;273(18):4336–45.
251. Hyun J, Matsunami H, Kim TG, Wolf M. Assembly mechanism of the pleomorphic immature poxvirus scaffold. Nat Commun. 2022 Mar 31;13(1):1704.
252. Zochowska M, Paca A, Schoehn G, Andrieu JP, Chroboczek J, Dublet B, et al. Adenovirus Dodecahedron, as a Drug Delivery Vector. Jagetia GC, editor. PLoS ONE. 2009 May 15;4(5):e5569.
253. Speir JA, Munshi S, Wang G, Baker TS, Johnson JE. Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron microscopy. Structure. 1995 Jan;3(1):63–78.
254. Overman SA, Kristensen DM, Bondre P, Hewitt B, Thomas GJ. Effects of Virion and Salt Concentrations on the Raman Signatures of Filamentous Phages fd, Pf1, Pf3, and PH75. Biochemistry. 2004 Oct 1;43(41):13129–36.
255. Chen PL, Wang M, Ou WC, Lii CK, Chen LS, Chang D. Disulfide bonds stabilize JC virus capsid-like structure by protecting calcium ions from chelation. FEBS Letters. 2001 Jul 6;500(3):109–13.
256. Piersimoni L, Kastritis PL, Arlt C, Sinz A. Cross-Linking Mass Spectrometry for Investigating Protein Conformations and Protein–Protein Interactions─A Method for All Seasons. Chem Rev. 2022 Apr 27;122(8):7500–31.
257. Leney AC, Heck AJR. Native Mass Spectrometry: What is in the Name? Journal of the American Society for Mass Spectrometry. 2017;28(1):5–13.
258. Nakano MY, Boucke K, Suomalainen M, Stidwill RP, Greber UF. The First Step of Adenovirus Type 2 Disassembly Occurs at the Cell Surface, Independently of Endocytosis and Escape to the Cytosol. J Virol. 2000 Aug;74(15):7085–95.
259. Burckhardt CJ, Suomalainen M, Schoenenberger P, Boucke K, Hemmi S, Greber UF. Drifting motions of the adenovirus receptor CAR and immobile integrins initiate virus uncoating and membrane lytic protein exposure. Cell Host and Microbe. 2011;10(2):105–17.
260. Ilca SL, Kotecha A, Sun X, Poranen MM, Stuart DI, Huiskonen JT. Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes. Nature Communications. 2015;6:4–11.

Appendix: Publications

Research Article

Stasiak AC, Gogler K, Borisova M, Fink P, Mayer C, Stehle T, Zocher G. N-acetylmuramic acid recognition by MurK kinase from the MurNAc auxotrophic oral pathogen *Tannerella forsythia*. J Biol Chem. 2023 Sep;299(9):105076.

Reviews

Stasiak AC, Stehle T. Human adenovirus binding to host cell receptors: a structural view. Medical Microbiology and Immunology. 2020 Jun 29;209(3):325–33.

*Sorin MN, *Kuhn J, ***Stasiak AC**, Stehle T. Structural insight into non-enveloped virus binding to glycosaminoglycan receptors: A review. Viruses. 2021;13(5):1–11.

* - these authors contributed equally

JBC RESEARCH ARTICLE

N-acetylmuramic acid recognition by MurK kinase from the MurNAc auxotrophic oral pathogen *Tannerella forsythia*

Received for publication, May 8, 2023, and in revised form, July 14, 2023 Published, Papers in Press, July 20, 2023,
https://doi.org/10.1016/j.jbc.2023.105076

Aleksandra Cecylia Stasiak[1], Karolin Gogler[1], Marina Borisova[2], Phillipp Fink[1], Christoph Mayer[2], Thilo Stehle[1], and Georg Zocher[1,*]

From the [1]Interfaculty Institute of Biochemistry, and [2]Interfaculty Institute of Microbiology and Infection Medicine, Organismic Interactions/Glycobiology, University of Tuebingen, Tuebingen, Germany

Reviewed by members of the JBC Editorial Board. Edited by Chris Whitfield

The bacterial cell wall consists of a three-dimensional peptidoglycan layer, composed of peptides linked to the sugars *N*-acetylmuramic acid (MurNAc) and GlcNAc. Unlike other bacteria, the pathogenic *Tannerella forsythia*, a member of the red complex group of bacteria associated with the late stages of periodontitis, lacks biosynthetic pathways for MurNAc production and therefore obtains MurNAc from the environment. Sugar kinases play a crucial role in the MurNAc recycling process, activating the sugar molecules by phosphorylation. In this study, we present the first crystal structures of a MurNAc kinase, called murein sugar kinase (MurK), in its unbound state as well as in complexes with the ATP analog β-γ-methylene adenosine triphosphate (AMP-PCP) and with MurNAc. We also determined the crystal structures of K1058, a paralogous MurNAc kinase of *T. forsythia*, in its unbound state and in complex with MurNAc. We identified the active site and residues crucial for MurNAc specificity as the less bulky side chains of S133, P134, and L135, which enlarge the binding cavity for the lactyl ether group, unlike the glutamate or histidine residues present in structural homologs. In establishing the apparent kinetic parameters for both enzymes, we showed a comparable affinity for MurNAc (K_m 180 μM and 30 μM for MurK and K1058, respectively), with MurK being over two hundred times faster than K1058 (V_{max} 80 and 0.34 μmol min^{-1} mg^{-1}, respectively). These data might support a structure-guided approach to development of inhibitory MurNAc analogs for pathogen MurK enzymes.

Tannerella forsythia is a Gram-negative oral pathogen that is part of the red complex of bacterial species associated with late stages of periodontitis, an inflammatory disease of gums and bones supporting teeth (1). Like other bacteria, *T. forsythia* possesses a three-dimensional peptidoglycan (PGN) network in the cell wall, consisting of N-acetylmuramic acid (MurNAc) and GlcNAc sugars and peptides that are connected by glycosidic, peptide, and amide bonds (2). Unlike other bacteria, however, the organism depends on MurNAc as a growth factor (3) (except for some canine and feline isolates (4)), since it is unable to synthesize MurNAc due to the absence of biosynthetic genes (5). To maintain its PGN layer, *T. forsythia* relies on scavenging MurNAc from the environment (6, 7), which involves the action of periplasmic *N*-acetylmuramidases (8), the sugar transport across the cell membrane (7, 9), and the synthesis of UDP-MurNAc *via* an uridylyltransferase (10). The *T. forsythia* ATP-dependent murein sugar kinase (MurK, NCBI Reference Sequence: WP_046825531.1) has previously been identified as a part of the Tf_*murTKQ* operon (Fig. 1) coding for enzymes responsible for import and cytoplasmic processing of MurNAc (11). MurK was shown to have a strong kinetic preference for MurNAc, yielding MurNAc 6-phosphate (MurNAc-6P; see Fig. 1), though with the ability to phosphorylate GlcNAc, albeit with very low affinity and reaction rate (11). As it is genetically linked to the MurNAc 6-phosphate etherase (MurQ), which converts MurNAc-6P to GlcNAc 6-phosphate, its apparent role is mainly to utilize MurNAc as a GlcNAc source rather than to yield UDP-MurNAc to sustain cell wall synthesis. Consistent with this view, a *T. forsythia murK* deletion mutant was shown to block MurNAc catabolism and allow the direction of MurNAc solely to PGN biosynthesis, resulting in a growth advantage in MurNAc-depleted medium (11). In addition to MurK, a second putative MurNAc kinase–encoding gene (NCBI Reference Sequence: WP_041590480.1) was identified in the genome of *T. forsythia*. This kinase, provisionally referred to as K1058 in this work, is encoded by a gene (*Tanf_RS08300*) located directly upstream of the PGN-recycling gene cluster located directly upstream of the Tf_*murTKQ* operon (12) and shares 44% sequence identity with MurK (Fig. S1, Table S1).

Sugar kinases catalyze the activation of sugar molecules by phosphorylation, commonly though not exclusively using ATP as the terminal phosphate group donor (13), and have been divided into five structural classes: hexokinases, repressor-ORF kinases (ROKs), ribokinases, GHMP (galactokinase, homoserine kinase, mevalonate kinase, and phosphomevalonate kinase) kinases, and phosphatidylinositol kinases (14). All these families employ the same reaction mechanism, where a conserved aspartic acid residue deprotonates the hydroxyl group, which then forms a transient pentavalent state with the ATP molecule, before the ADP moiety leaves. The presence of divalent metal ions such as magnesium or zinc is also a common requirement (14). In addition to phosphorylating

* For correspondence: Georg Zocher, georg.zocher@uni-tuebingen.de.

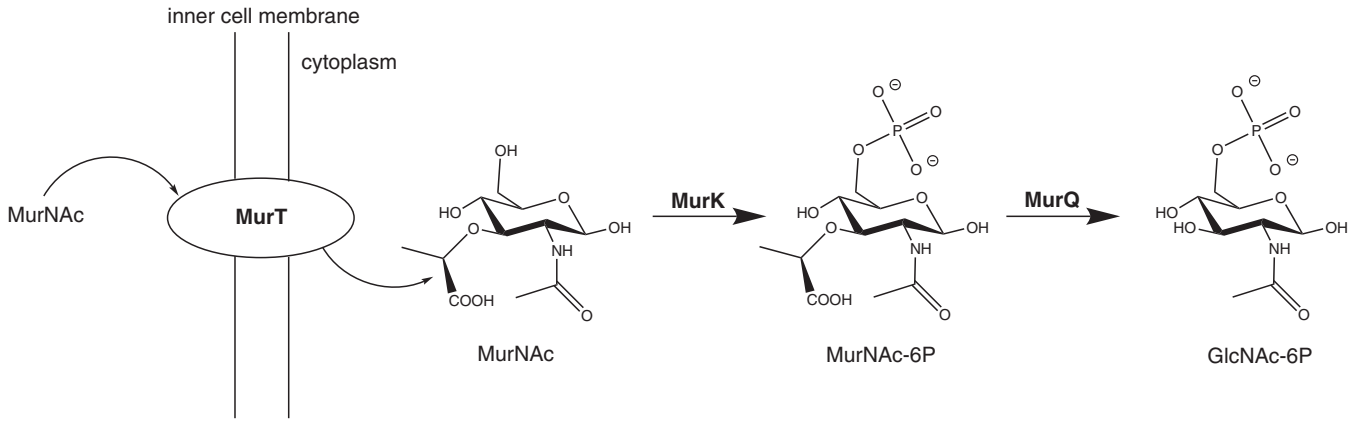


Figure 1. The MurTQK pathway in *Tannerella forsythia*: extracellular MurNAc is transported into the cytoplasm by MurT. MurNAc phosphorylation to MurNAc-6P is catalyzed by kinases MurK and K1058 (the relation of the latter to the MurTQK pathway is unclear), using ATP as a phosphate group donor. MurNAc-6P is converted to GlcNAc-6-P by MurQ. MurK, murein sugar kinase; MurNAc, N-acetylmuramic acid; MurNAc-6P, N-acetylmuramic acid-6 phosphate.

glucose, glycokinases have many other sugar substrates, such as fructose, *N*-acetylmannosamine, GlcNAc, allose (13, 15), or even more complex glycan molecules such as muramyl dipeptides. Recently, the human GlcNAc kinase *N*-acetylglucosamine kinase (NagK) was shown to act as a MurNAc kinase that specifically phosphorylates muramyl dipeptide and thereby activates this immunogenic ligand for bacterial recognition *via* the nucleotide-binding oligomerization domain-containing protein 2 (NOD2) receptor (16). Besides their role in primary metabolism, sugar kinases also catalyze secondary metabolism reactions, generating products necessary for other pathways (15). In bacteria, such reactions enable the organism to produce molecules for cell wall synthesis or reshaping, crucial in the formation of the barrier between the environment and the organism (5, 17–19).

MurNAc-specific kinases have not yet been structurally characterized, although previous work has functionally characterized the MurNAc- and GlcNAc-specific MurK from *Clostridium acetobutylicum* (20). Here, we present high-resolution crystal structures of MurK as an apo enzyme, in complex with an ATP analog, and in complex with its MurNAc substrate, based on which we model the ternary complex to describe the mode of enzymatic catalysis. To compare MurK and K1058, we expanded our investigation to structures of K1058 in its apo and MurNAc-bound forms. Structural and sequence data showed that MurK and K1058 both belong to the group of ribonuclease H-like kinases and are members of the ASKHA (acetate and sugar kinase/hsc70/actin) superfamily (21). The conserved catalytic aspartate residue is present in both enzymes. Based on the data, we identified a nonconserved patch potentially important for MurNAc recognition in MurK. To gain insight into their roles in the cell wall metabolism in *T. forsythia*, we established the apparent kinetics of both enzymes for their ability to turn over MurNAc.

Results

Structure determination of MurK and K1058

MurK and K1058 are two sugar kinases of the cell wall recycling system of *T. forsythia*, sharing a sequence identity of 44% (Fig. S1, Table S1) and separated by a single PGN-scavenging operon within the genome. To investigate the relationship between this sequence similarity, we determined the structure of both enzymes. We produced and purified MurK and K1058 by recombinant expression as His_6-tagged proteins in *Escherichia coli*. Both kinases eluted as highly pure, dimeric enzymes from a size-exclusion chromatography (SEC) column with an apparent molecular mass of approximately 60 kDa (molecular mass for MurK and K1058 is 32.4 and 32.3 kDa, respectively). Both enzymes were successfully crystallized. We obtained triclinic crystals for MurK (termed MurKapo), which contain four protomers in the asymmetric unit (ASU) that form two biological dimers. K1058 crystallized in a tetragonal space group (termed K1058apo) with a single protomer in the ASU. The biological corresponding dimer interface as is observed for MurKapo results from the application of the crystallographic two-fold axis. To obtain insight into substrate binding and catalysis of the kinases, we determined substrate-bound structures of MurK and K1058. Initial attempts of soaking MurNAc into MurKapo or K1058apo crystals with or without an ATP analog failed, probably due to an induced fit accompanying MurNAc binding. Instead, we pursued cocrystallization with MurNAc and/or β-γ-methylene adenosine triphosphate (AMP-PCP), a less hydrolyzable analog of the phosphate donor ATP. We successfully cocrystallised MurK with AMP-PCP (termed MurK/AMP-PCP) and established its structure at 2.12 Å resolution. Crystals of MurK/AMP-PCP were monoclinic and contain a biological dimer in the ASU. The production of MurNAc-bound crystals of MurK and K1058 was very challenging but eventually yielded crystals of MurK and K1058 in complex MurNAc (termed MurK/MurNAc and K1058/MurNAc, respectively). Both crystal forms diffract to approximately 3 Å resolution and contain six protomers in the ASU corresponding to three biological dimers. Crystals of MurK/MurNAc suffered from perfect merohedral twinning.

Although, both substrates (AMP-PCP and MurNAc) were used to produce MurK/MurNAc and K1058/MurNAc crystals, a ternary complex could not be established, probably due to a very long crystallization time (6–9 months, in contrast to

3 days for the MurK/AMP-PCP crystals), during which the ATP analog seems not to have remained stable in the crystallization conditions. X-ray data showed continuous electron density for both kinases and allowed tracing of the complete protein chains in all structures (Figs. 2 and S2) except for the final C-terminal residues (for K1058 and MurK at least residues 1–277). The electron density for AMP-PCP and MurNAc in the corresponding structures unequivocally identified both ligands and allowed for placing of the compounds with full occupancy into the active sites of all chains of the enzymes (Fig. 3), beside MurK/MurNAc. Here, two chains showed weak electron density probably due to merohedral twinning, which allowed the sugar to be placed only in four out of six chains of the ASU.

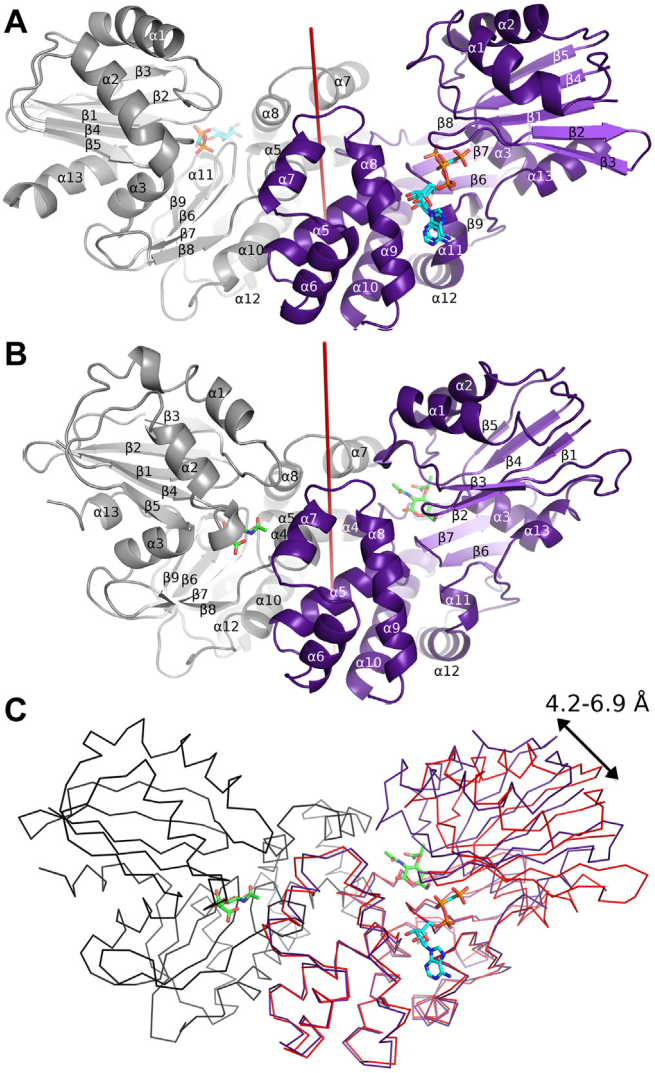


Figure 2. The structures of *Tannerella forsythia* MurK in complex with its substrates. *A*, cartoon representation of the structure of MurK as a dimer in complex with the ATP analog AMP-PCP (*blue* and *orange*). Two conformations are shown for AMP-PCP in one of the chains, differing mainly in the position of the terminal phosphoryl group. One of the chains is shown in *gray*, the other in *purple* (chain B). *B*, cartoon representation of the structure of MurK as a dimer in complex with the MurNAc substrate (*green*). Color scheme as in (*A*). *C*, movement of MurK on binding the sugar substrate: ribbon representation of the ATP analog-bound MurK (*red*) and MurNAc-bound MurK (*purple*). The smaller N-terminal domain shifts into a closed conformation by 4.5 to 6.9 Å. The other chain of the MurK/MurNAc dimer is showed as *black ribbon* for reference. MurK, murein sugar kinase; MurNAc, N-acetylmuramic acid.

In summary, we established the unliganded structures of MurK (MurKapo) and K1058 (K1058apo), a MurK structure with the phosphate analog AMP-PCP (MurK/AMP-PCP), and structures of both MurK and K1058 with MurNAc (MurK/MurNAc and K1058/MurNAc, respectively). Data collection and model quality criteria of all structures are summarized in Table 1.

Overall architecture of MurK kinases

According to the classification by Cheek *et al.* (21), the kinase MurK belongs to the group of ribonuclease H-like kinases and is a member of the ASKHA (acetate and sugar kinase/hsc70/actin) superfamily. MurK is a member of the glucokinase family of hexokinases that phosphorylate their conjugate sugar substrate at the opposite site of the anomeric C-atom (14), which is the hydroxyl group at atom C6 of MurNAc (Fig. 1). Structurally, MurK possesses the glucokinase fold (Figs. 2, *A* and *B* and S3), sharing the same topology characterized by a small N-terminal α/β domain (residues 1–104) and a large C-terminal α/β domain (residues 104–284). The core motif of both domains resembles the core motif of the ribonuclease H fold, a three-layered α/β/α-motif with ββββαβαβα topology, where the second β-strand runs antiparallel to the central β-sheet. In contrast to the N-terminal core motif, seven α-helices (α4 to α10) are inserted into the larger C-terminal domain of the enzyme after the third β-strand β8. These insertions contain the majority of amino acids that constitute the two-fold symmetrical dimeric interface of the MurK.

In agreement with other glucokinases (22), a detailed analysis of all three crystal lattice systems observed for MurK using Evolutionary Protein Protein Interface Classifier (EPPIC) and Protein Data Bank Proteins, Interfaces, Structures and Assemblies (PDB PISA; (23, 24)) revealed a dimeric assembly of the enzyme (Fig. 2, *A* and *B*), which was confirmed by SEC experiments. A detailed C_α-RMSD analysis of all four chains in MurKapo showed that both protomers of a dimer possess a virtually identical conformation that is slightly different to the second dimer of the ASU (Fig. S4). The dimerization interface of the apo enzyme buries a surface of around 1160 Å^2 and involves just two residues (F34 and F35) of the N-terminal domain, whereas all other contacts to the adjacent protomer were made by several locations on the α-helical insertion of the C-terminal domain. These include residues in the helices α4-α9 and the loop region in between β8 and α4, as well as residues in the loop region α10 to β9. The interactions of the dimeric assembly are detailed in Tables S2–S4. The dimer buried surface area increased to 1600 to 1700 Å^2 for MurNAc-bound structures, an effect that is also reported for the human *N*-acetylmannosamine kinase by Martinez *et al.* (25).

Comparison to other sugar kinases

As expected, a search for structurally related proteins (26) revealed similarities to glucokinases (some not biochemically characterized), ROK hexokinases, but also to actin-related

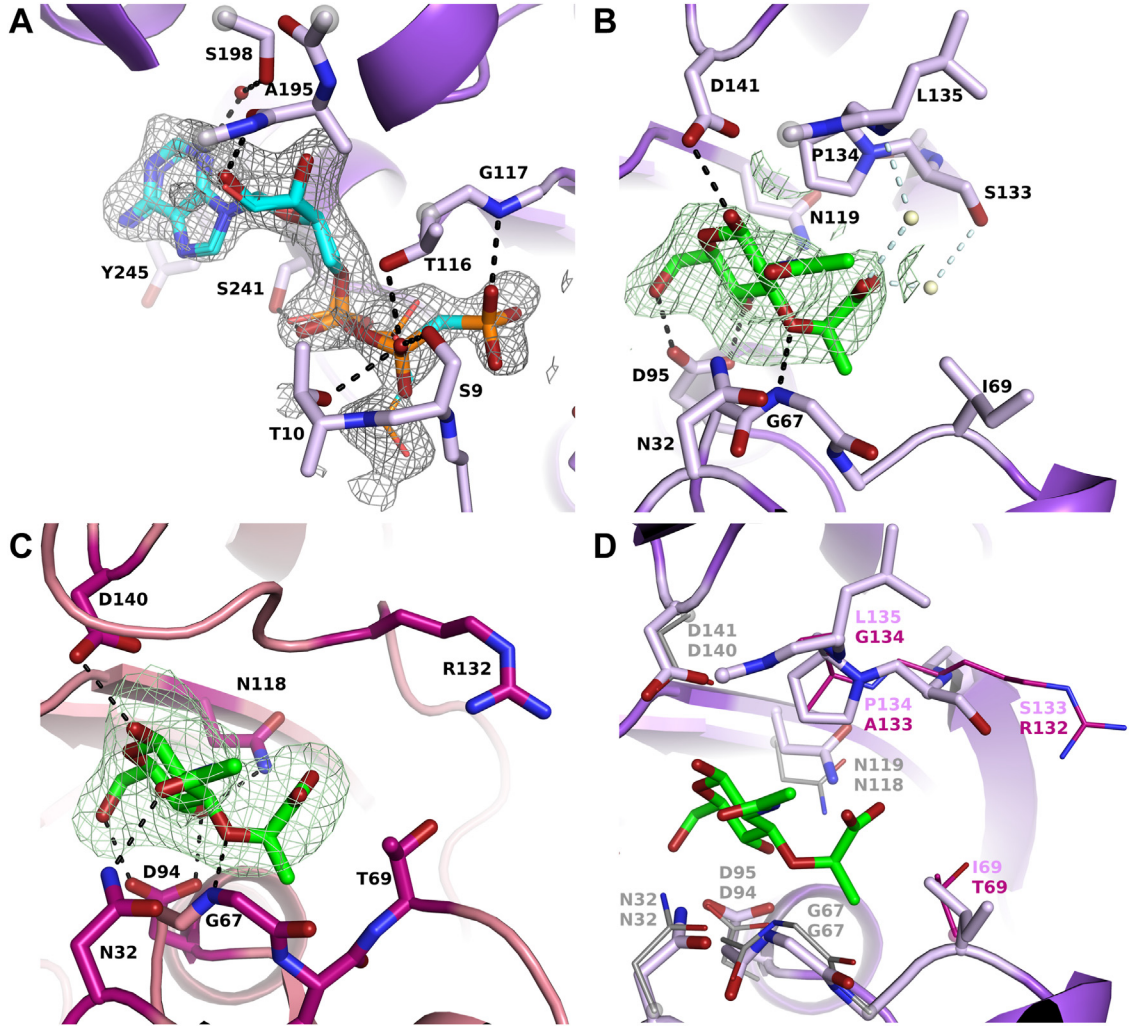


Figure 3. Substrate-binding sites of MurK and K1058. Binding sites of MurK with AMP-PCP bound (*A*), MurK with MurNAc bound (*B*), and K1058 with MurNAc bound (*C*). Residues presumed to be involved in binding and substrates (MurNAc, *green*; AMP-PCP, *blue* and *orange*) are shown as *sticks*. Alternative conformation of AMP-PCP, pointing away from the MurNAc-binding site, is shown as lines. *Black dashes* represent proposed hydrogen bonds. *Red spheres* represent water molecules. In (*B*), *pale blue dashes* and *ivory spheres* represent water molecules and water-mediated hydrogen bonds postulated only based on their presence in MurNAc/AMP-PCP structure. Density from the simulated annealing Fo-Fc omit map shown as mesh at 3σ (both MurNAc structures) or 2.5σ (MurK/AMP-PCP structure). Best-quality density is chosen from multiple chains, and density in other chains is shown in Fig. S6. *D*, superposition of the MurK and K1058 MurNAc-binding sites. MurK residues are shown as *purple sticks*, and K1058 residues are shown as *lines*. Only amino acids forming hydrogen bonds or differing between the binding sites of the two proteins are shown. Residues which are conserved between the two sites are shown in *gray* for K1058, and different residues are shown in *magenta*. Helices α3, α8, α9, α12, α13, beta sheet β9 and β10, loops β1β2, α7α8 (*A*), helices α3, α4, α10, sheets β6, β7, β8, and loops β3α1, α10β9 (*B*), helices α1, α4, sheets β6, β7, β8, and loops β7β8, α10β9, α12α13 (*C*), helices α3, α4, α13, sheets β6, β7, β9, loop β3α1, α3b6, β7β8, α10β9, α12α13 (*D*) have been partly or completely removed for clarity. MurK, murein sugar kinase; MurNAc, N-acetylmuramic acid.

proteins with C_α-RMSDs ranging from 1.2 Å to around 4.2 Å. We focused our comparison on sugar kinases. Here, the closest structural homologs to MurK are two proteins, both targets of the Northeast Structural Genomics project (PDB ID 1ZXO and 1ZBS), both possessing a very low C_α-RMSD of 1.1 Å in relation to MurKapo (Table S5). These proteins are not biochemically characterized, and their natural substrates are therefore unknown; however, 1ZBS has been described as a putative GlcNAc kinase. Based on structural similarity, it is possible that MurNAc would be a substrate candidate. All other related sugar kinases are structurally less well conserved, with C_α-RMSD values above 2.8 Å, which is the C_α-RMSD for human GlcNAc kinase (22), the closest hit for which the substrate specificity is known and for which complex structures with ADP and GlcNAc were also established. In comparison to the human GlcNAc kinase, a member of the glucokinase family, the C terminus of MurK is substantially shortened. For detailed comparison of the active site, we focused on hits with established substrate specificities and structurally characterized sugar binding.

Induced fit of MurK kinase upon substrate binding

Soaking experiments with MurKapo crystals using MurNAc failed, probably due to a rearrangement of the N-terminal domain upon substrate uptake ultimately leading to severe crystal damage. The induced fit of MurK is triggered by binding of MurNAc into the cleft-shaped active site and results in a combined sheered rotation move of the N-terminal domain as a rigid body closing the active site over the

Table 1
Data collection and refinement statistics for MurK and K1058 structures

Data collection	MurKapo	MurK/AMP-PCP	MurK/MurNAc	K1058apo	K1058/MurNAc
Space group	P1	P 2_1	I4*	I 4_1 2 2	P 3_1 2 1
Cell dimensions					
a, *b*, *c* (Å)	58.76, 59.51, 96.71	61.70, 55.65, 86.00	113.43, 113.43, 294.76	131.94, 131.94, 89.22	143.75, 143.75, 210.95
α,β,γ (°)	74.30, 75.27, 65.95	90, 93.60, 90	90, 90, 90	90, 90, 90	90, 90, 120
Resolution (Å)	49.50-2.05 (2.05-2.17)	48.6-2.12 (2.12-2.18)	49.13-2.7 (2.87-2.7)	49.22-2.00 (2.00-2.05)	29.85-3.06 (3.14-3.06)
No. of unique reflections	68,635 (11,045)	33,324 (2435)	49,814 (8018)	26,844 (1946)	48,103 (3521)
Redundancy	2.4 (2.4)	6.8 (7.0)	14.1 (14.5)	26.6 (26.2)	23.6/20.5
R_{meas} (%)	9.7 (65.3)	17.9 (141.6)	40.8 (338)	18.8 (267.2)	45.6/(331.3)
$I/\sigma(I)$	9.1 (1.9)	10.4 (1.5)	8.1 (1.2)	19.6 (1.5)	8.1 (1.0)
Completeness (%)	96.1 (95.7)	100 (100)	98.5 (98.6)	100 (100)	99.9/100
$CC_{1/2}$	99.7 (75.1)	99.6 (57.4)	99.4 (46.6)	99.9 (53.4)	99.3 (34.8)
Wilson B-factor (Å^2)	38	36	65	40	62
Refinement					
R_{work}/R_{free}	0.207/0.254	0.185/0.233	0.248/0.277	0.177/0.206	0.235/0.259
B-factors (Å^2)					
Chain A/B/C/...	34/35/39/42	36/33	72/48/72/51/71/99	42	74/68/73/113/103/92
Ligand	-	58/34	66/40/-/48/61/-	-	62/57/58/105/78/76
RMS deviations					
Bond lengths (Å)	0.003	0.011	0.009	0.007	0.001
Bond angles (°)	1.06	1.79	1.85	1.46	0.366
Ramachandran					
Favored [%]	94.6	94.2	90.8	96.4	94.1
Outliers [%]	0.6	1.1	1.9	0.7	0.4

Asterisk (*) denotes a twinned structure.

phosphate acceptor sugar (Figs. 2*C* and S5). The induced fit of MurK can be visualized by superposing the C-terminal domain of the unliganded MurKapo ("open" conformation) to its corresponding C-terminal domain of MurK/MurNAc ("closed" conformation). The best fit utilizes residue range 106 to 266 and results in a C_α-RMSD value of 0.67 Å for 152 aligned residues, showing virtually identical backbone conformation of all the aligned residues of that domain (see Fig. 2*C*). The transition from open to closed conformation is based on helix α3, which slides over the C-terminal β-sheet towards the active site, resulting in a rotation by approximately 12° (calculated by DynDom (27), Fig. S5) and a maximum translation of 6.9 Å measured for the C_α distance of residue S88 when comparing open and closed states. Beside loops α1β4 and β1β2, of which loop β1β2 pushes towards the sugar-binding site, a C_α-RMS difference plot of the N-terminal domain from open to closed state would classify the adjustment as a rigid body move as indicated by the total C_α-RMSD value of 1.1 Å for all residues of the N-terminal domain.

Only minor rearrangements for ATP binding were observed, with an C_α-RMSD value of 0.78 Å, aligning all residues of the N-terminal domain of MurKapo and MurK in its AMP-PCP–bound state. This is in agreement with the observation that ATP is predominantly held in place by residues of the C-terminal domain. With respect to the dimeric state of the enzyme, the induced fit is also a concentric move of both N-terminal domains towards each other. In the open state, the adjacent protomer does not contribute to the neighboring active site, but the induced fit accompanying the MurNAc uptake enhances the reshaping of the active site by rigidifying the dimeric interface with additional contacts from the N-terminal protomer. This involves contacts made by residues from loop β3α1 and helix α1 (residues 34–39) and residues from loop α1β4 (residues 70,72 und 73) to enlarge the buried surface area of the dimer by almost 500 Å^2 (Table S2). Similar kinase dynamics on sugar binding are widely reported, for example, the human NagK (22), *Sulfolobus tokodaii* broad-specificity sugar kinase (28), *N*-acetylmannosamine kinases from pathogenic bacteria *Pasteurella multocida* and *Haemophilus influenzae Pm*-NanK, *Hi*-NanK (29), as well as *N*-acetylmannosamine kinases from humans (25). The observation that a strong induced fit takes place only during MurNAc binding is in agreement with the suggested random sequential binding mechanism (21) of sugar kinases. The closed conformation is essential for catalytic turnover, as only in this enzymatic state, the reaction center is shielded and allows for tight interactions with MurNAc.

The active site and activity of MurK kinase

The active site is located in a channel between the N- and C-terminal domains of the V-shaped enzyme (Fig. 2*A*). The channel compresses upon MurNAc binding to the enzyme (Fig. 2*B*), likely to enhance the interactions with the sugar molecule. The binding site can be subdivided into two binding regions, the ATP-binding site and the MurNAc-binding site (Figs. 3, *A* and *B* and 4). The ATP-binding site is formed by residues of the loop β6β7 (I113, G115, T116, G117), helix α5 (G145), helices and loops between helices α8 and α11 (A195, G196, S198, P199 and A202, S241, V242, Y244, Y245) on the C-terminal domain. The interactions of the N-terminal domain are limited to residues of the loop β1β2 (G8, S9, T10, and K11), which interact with all phosphoryl groups of ATP (Fig. 3*A*).

The AMP-PCP is clearly bound in both protomers of the ASU of the monoclinic crystals (Fig. S6). Both chains possess an identical conformation (Fig. S7). AMP-PCP is bound at the N-terminal domain *via* a conserved DxGxT motif with interactions of the side chain of S9 upstream of the conserved T10 residue, both forming hydrogen bonds to the β-phosphate group. This common motif is also found in the kinases of *Bacillus subtilis* (30) and *Naegleria fowleri* (31) and,

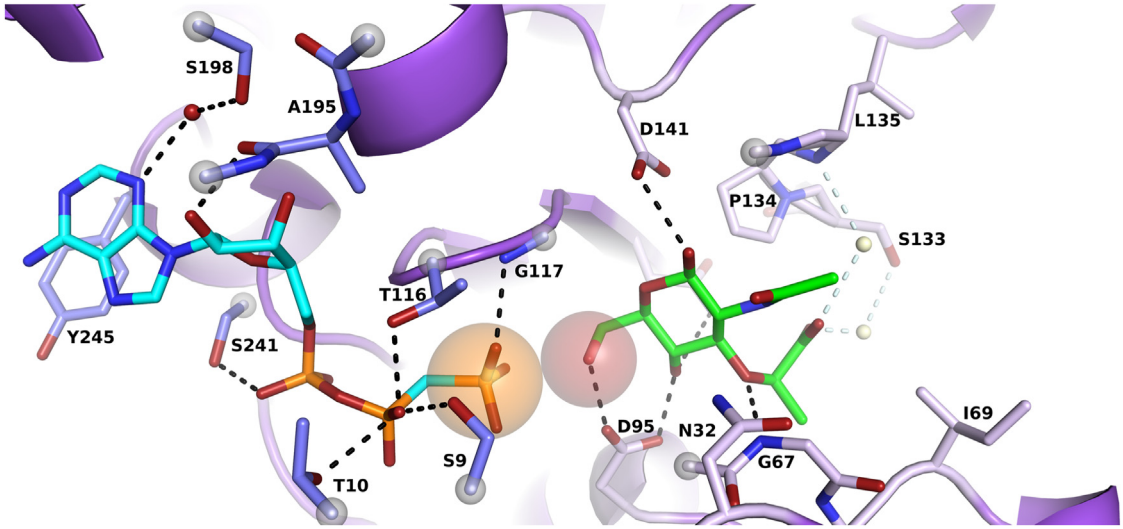


Figure 4. Structurally informed model of the active site of MurK with both substrates bound. MurK/AMP-PCP is superimposed on the C-terminal (residues 110–273) of MurK/MurNAc (shown as *cartoon*). Residues involved in binding are shown as *sticks*: *pale purple* for the MurK/MurNAc structure, *blue* for those from the MurK/AMP-PCP structure. *Dashes* represent proposed hydrogen bonds, and *spheres* represent waters (color scheme as in Fig. 3). *Orange* and *red spheres* represent reaction centers for phosphorus and oxygen, respectively. MurK, murein sugar kinase; MurNAc, N-acetylmuramic acid.

functionally conserved with a serine residue, in human NagK. All other interactions with AMP-PCP are located on the C-terminal domain. T116, also conserved in the homologs, is located on the opposite site of the phosphate entity and forms a hydrogen bond with the β-phosphate of AMP-PCP and the γ-phosphate to the backbone amino group of G117. The adenosine is bound mainly *via* van der Waals interactions contributed by nonpolar residues but forms hydrogen bonds *via* the carbonyl atom of A195 towards the hydroxy group at C2 atom of ribose and a water-mediated hydrogen bond *via* S198 to the adenosine ring system. In addition, the aromatic ring system is aligned coplanar to tightly pack against side chain Y244. Finally, S241, part of a conserved IGSV sequence (residues 238–241), stabilizes the α-phosphate by hydrogen bonding. Magnesium ions have been shown to be important for the function of MurK (11) and other sugar kinases (25, 28). We were not able to unequivocally identify a magnesium ion in any of the MurK structures. Typically, magnesium is found at the terminal phosphate groups to enhance the electrophilic character of the γ-phosphate but also to stabilize the leaving group ADP.

The MurNAc-binding site is located on the opposite site of the ATP molecule and shaped by residues of the loop β1β2 (S9), loop β3α1 (N32, F34), loop β4α2 (G67, I69), and loop β5α3 (D95) on the N-terminal domain. On the C-terminal domain, interactions involve strand β7 (N119), loop α4α5 (D141), and loop β8α4 (S133, P134, L135) (Fig. 3*B*). Structure-based modeling (Fig. 4) suggested that the C6 hydroxy group of MurNAc points towards the terminal phosphate of ATP in the center of the cavity, which is in perfect agreement with the biochemical reaction catalyzed by the enzyme. The anomeric carbon of the glucosamine ring of MurNAc adopts a β-configuration and is held in place by a strictly conserved catalytic residue D95 that establishes hydrogen bonding to the hydroxy groups of C4 and C6. This polarizes the reactive oxygen atom and facilitates the nucleophilic attack on the γ-phosphate group of ATP. In the N-terminal domain, residue S9 and residues of loop β3α1 (N32, F34, F35), of loop β4α2 (A66, G67, C68, I69), and of loop β5α3 and helix α3 (T94, M96) form the MurNAc cavity that is completed by interactions on the C-terminal domain to residues 117 to 119 and the region including residues 133 to 136. These interactions are predominantly van der Waals contacts. Only four residues specifically recognize the functional groups of MurNAc. G67 forms a hydrogen bond to the ether bond of the lactyl ether entity *via* its backbone amide, whereas D95, N119, and D141 utilize their side chains to recognize the hydroxy groups at atom C4 and C6, C4, and C1, respectively (see Fig. 3*B*). The importance of the strictly conserved catalytic D95 is highlighted by its clamp-like interaction with hydroxy groups both at atom C4 and atom C6, a feature that is also reported in other sugar kinases ((22, 32, 33) and PDB ID 8DTC). Both functional groups (*N*-acetyl and lactyl ether) of the MurNAc ring system are not specifically recognized by hydrogen bonding. The *N*-acetyl group dips into a rather hydrophobic cavity that is formed by the backbone of G67 and a set of three phenylalanine side chains (F34, F35, F137). Residue N32, a potential hydrogen bond donor to interact with the carbonyl function of the *N*-acetyl group as reported for the human NagK, is too far away (3.6 Å) to contribute to its recognition but lines the bottom of this cavity. In comparison to NagK, the amide functional group is also not saturated by hydrogen bonding due to amino acid substitution of G145 in NagK with P134 in MurK. The lactyl ether points towards the entrance of the channel. No direct protein interactions are observed to account for the negative charge of the carboxyl group besides shape complementarity of the protein, predominately created by P134 on top of which the lactyl ether sits. When superposing the MurK–AMP-PCP complex to MurK/MurNAc, potential water-mediated interactions of two

structurally conserved water molecules with the carboxyl group of MurNAc seem to be likely based on their presence in the high-resolution structure of MurK/AMP-PCP. One of the modeled water molecules bridges the carboxy entity *via* the side chain of S133, whereas the other water molecule would be bonded by an amide backbone interaction to L135.

We compared the binding mode of available hexokinases and ROK enzyme structures by superposing the conserved catalytic aspartate residues from different proteins (Fig. 5). These included the human NagK (22), the hexokinase from *S. tokodaii* (28), human hexokinase I (32), the inorganic polyphosphate/ATP-glucomannokinase from *Arthrobacter* sp. *strain KM* (33), the fructokinase from *B. subtilis* (30), the ROK glucokinase from *Streptomyces griseus* (34), glucokinase from *N. fowleri* (31), the glucokinase from *Acanthamoeba castellanii* (PDB ID: 8DTC), the *N*-acetylmannosamine kinase from *H. influenzae* (29), and the glucokinase from *Trypanosoma cruzi* in complex with a GlcNAc-linked inhibitor (PDB ID: 7S2P). It is noteworthy that all glucose molecules or glucose derivatives adopt a virtually identical chair conformation with β-configuration at the anomeric C-atom when bound to their corresponding enzymes, with the exception of GlcNAc binding to NagK (22), which adopts an α-configuration.

When comparing MurNAc kinases to glucose or GlcNAc kinases, analysis indicates the reason for MurNAc specificity to be related to two positions (Fig. 5). Most obviously, residues S133, P134, and L135 in MurK allow for MurNAc binding by providing sufficient space for the lactyl ether substituent. Except for the human NagK (22) and the hexokinase from *S. tokodaii* (28), the corresponding position in the homolog enzymes listed above carries a bulky residue that is either charged (mostly a glutamate residue) or polar (H153, (29)) that would sterically clash with lactyl ether entity, abolishing the binding of MurNAc (see Fig. 5). A second position crucial for MurNAc recognition arises from residue T94. Here, the homolog sugar kinases carry a polar side chain (asparagine, except for hexokinase from *S. tokodaii*, where a histidine residue is found; Fig. 5) that is involved in binding the hydroxy group at C4 of the sugar molecule. This interaction is not observed for MurK and would result in steric stress towards the lactyl ether substituent, abolishing MurNAc binding. In hexokinase from *S. tokodaii*, the sterical clash to the lactyl ether moiety is moreover reinforced by substitution of I69D which would additionally interfere with MurNAc binding *via* charge repulsion.

The reaction pathway can be inferred from superposing both substrate structures, MurK/MurNAc and MurK/AMP-PCP (see Fig. 4). The terminal phosphate group of ATP is in close proximity to the hydroxy group of MurNAc, with a distance of 3.4 Å to the phosphorus atom that is attacked during the reaction. The clamp-like binding by D95 polarizes the hydroxyl group to facilitate a nucleophilic engagement of the oxygen atom at C6 of MurNAc. Such a polarization would facilitate a nucleophilic attack on the terminal phosphate group of ATP. During catalysis, D95 probably acts as a catalytic base to abstract the proton of the nascently generated MurNAc-6P. The acid-base catalysis of this type of sugar phosphorylation has been previously confirmed by site-directed mutagenesis, for example, human kinase I almost completely loses catalytic activity (99% activity loss) when the

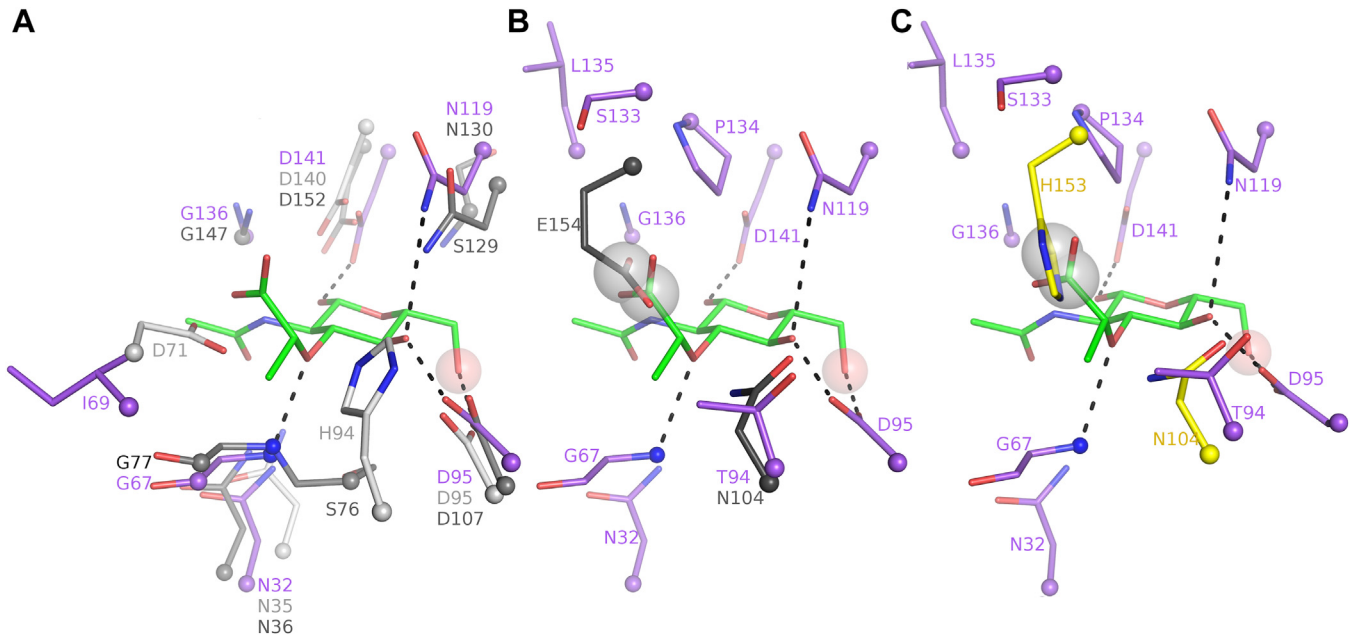


Figure 5. Sugar-binding residues in the active site of MurK and its structural homologues. *A*, active site residues mediating MurNAc binding in MurK (*purple*), broad-specificity kinase from *Sulfolobus tokodaii* (*white*; PDB ID: 2E2O), and human NagK (*gray*; PDB ID: 2CH6). Structures are superimposed on the sugar (*green*). *B*, active site residues mediating MurNAc binding or MurNAc specificity in MurK (*purple*) and those causing clashes in *Streptomyces griseus* glucokinase (*black*, PDB ID: 3VGL). *C*, active site residues mediating MurNAc binding or MurNAc specificity in MurK (*purple*) and those causing clashes in *Haemophilus influenzae* mannosamine kinase (*yellow*, PDB ID: 6JDB). Structures in all panels were superimposed only on the sugar molecules present in the active sites to demonstrate the similarities and differences in sugar binding. MurK, murein sugar kinase; MurNAc, N-acetylmuramic acid; NagK, N-acetylglucosamine kinase.

catalytic aspartate residue is replaced by an alanine or the human mannosamine kinase loses catalytic activity on mutating the aspartate to alanine or asparagine (25, 35, 36).

Based on the structural data, we reinvestigated the enzymatic activities that were reported for MurK acting as MurNAc kinase with a K_M value of 113 μM and a catalytic turnover of k_{cat}/K_M of 69,910 $M^{-1}s^{-1}$ with little acceptance of the enzyme for GlcNAc (11). Our activity measurements are in agreement with previous measurements (Table 2) and show an affinity towards MurNAc with similar apparent K_M values (110 ± 20 μM for strict Michaelis–Menten kinetics and 180 ± 40 μM taking substrate inhibition into account). The substrate-inhibition model fit the data better (Fig. 6), a result that was also reported in the studies by Hottman *et al.* (11). The structural basis for substrate inhibition remains to be determined.

Crystal structure and activity of MurK paralog K1058

Based on genomic sequence analysis, a second putative MurNAc kinase, termed K1058, was identified in *T. forsythia*, sharing a sequence similarity of 44% to MurK. In an attempt to establish the structural differences between the two MurKs, the K1058 ortholog was cloned and the protein heterologously expressed and purified in a similar procedure as described for MurK. We determined the structure of K1058 to establish the structural differences (Figs. 3, *C* and *D* and S8). We were able to obtain structures of the apo enzyme (K1058apo) and K1058 in complex with MurNAc (K1058/MurNAc) (Fig. S2). A superposition showed a close relationship between K1058 and MurK, with an observed C_α-RMSD value of 2.1 Å over the total protein length. An in-depth active site comparison of both MurNAc-liganded structures showed a very similar binding of MurNAc as in MurK, with the sugar molecule slightly tilted around the bond between O1 and C5 (Fig. 3, *C* and *D*). When comparing all residues of the sugar-binding site, differences can be found in two locations. Firstly, residue I69 of MurK is replaced by a less bulky and more polar threonine (T69) in K1058. Secondly, in the region involved in lactyl ether recognition, residues S133 and P134 are substituted by an arginine and an alanine, respectively. All other residues in the MurNAc-binding site are identical and occupy conformations very close to MurK (Fig. 3*D*). We speculate that the substitutions should not severely structurally influence the affinity of MurNAc binding to K1058. The residues I69 and P134 in MurK give an excellent shape complementarity, which corresponds to residues T69 and R132 in K1058, respectively. Hence, structural data would not suggest a decrease in MurNAc affinity. Moreover, R132 in K1058 seems to be an optimal candidate to bind MurNAc *via* its carboxy group. Although we do not observe such an interaction, probably because the side chain conformation would have to be in an "bent back" setup, a partial charge compensation is possible, which might compensate for the loss in affinity by the loss of water-mediated hydrogen bonding to S133 in MurK. We confirmed our observations by activity measurements with K1058 (Table 2). Indeed, the affinity for K1058 is higher than MurK, with an apparent K_M value of 28 ± 10 μM, but the catalytic turnover is severely reduced and suggests that MurNAc might not be the natural substrate of K1058. To exclude enzyme aggregation as the reason for this observation, we performed CD spectroscopy and analytical SEC experiments (Fig. S9) on K1058. These revealed folded protein and did not show significant aggregation even on incubation of the enzyme for 4 days at 4 °C and after freezing and thawing. A reason for this unexpected observation is still to be determined, but we speculate that the subtle changes in MurNAc binding are mainly caused by the exchanges at the positions S133 and P134 in MurK to R132 and A133 in K1058. The backbone geometry is rigidified by P134, resulting in a tighter packing of this region towards MurNAc and therefore favors MurNAc binding to MurK in comparison to K1058. In summary, the enzymatic assays showed that K1058 has low efficiency for MurNAc compared to MurK, although its affinity is higher. One possible explanation for the unusual kinetics of K1058 would be that the enzyme needs to be (allosterically) activated for full efficiency. Currently, we would assume that the natural substrate of K1058 is not MurNAc, rather it may act as an inhibitor. A speculative explanation for the observation of high affinity but low turnover of MurNAc could be a controlling mechanism to downregulate K1058 processing of the natural substrate in the presence of MurNAc.

Discussion

Structural analyses of MurK revealed that the enzyme belongs to the group of ribonuclease H-like kinases, with a fold that is very similar to that of other sugar kinases, particularly those classified as ROK kinases (despite MurK not belonging to that family based on sequence). Ribonuclease H-like kinases show a remarkable similarity across domains of life, with the top matches for structural similarity being an uncharacterized bacterial protein (PDB IDs: 1ZXO) and a putative NagK (PDB code: 1ZBS), followed by a broad-specificity glucokinase from the hyperthermophilic archeon *S. tokodaii* (28) and the human NagK in complex with GlcNAc (22). The broad structural conservation of sugar phosphorylation is unsurprising, given its role in basic metabolic functions such as glycolysis. Our

Table 2
Apparent kinetic parameters for MurK and K1058 with MurNAc as a substrate

Protein	K_M [μM]	v_{max} [μmol min^{-1} mg^{-1}]	k_{cat} [s^{-1}]	K_i [mM]	k_{cat}/K_M [M^{-1} s^{-1}]
MurK	110 ± 20	62 ± 2.7	32 ± 1.4	-	300,000 ± 56,000
MurK*	180 ± 40	80 ± 8.0	42 ± 4.2	5.6 ± 2.3	230,000 ± 56,000
K1058	28 ± 10	0.34 ± 0.02	0.18 ± 0.01	-	6400 ± 2900
K1058*	55 ± 22	0.38 ± 0.04	0.20 ± 0.02	20 ± 23	3600 ± 1500

Both the Michaelis–Menten and substrate-inhibited models (marked with *) were fitted into the experimental data.

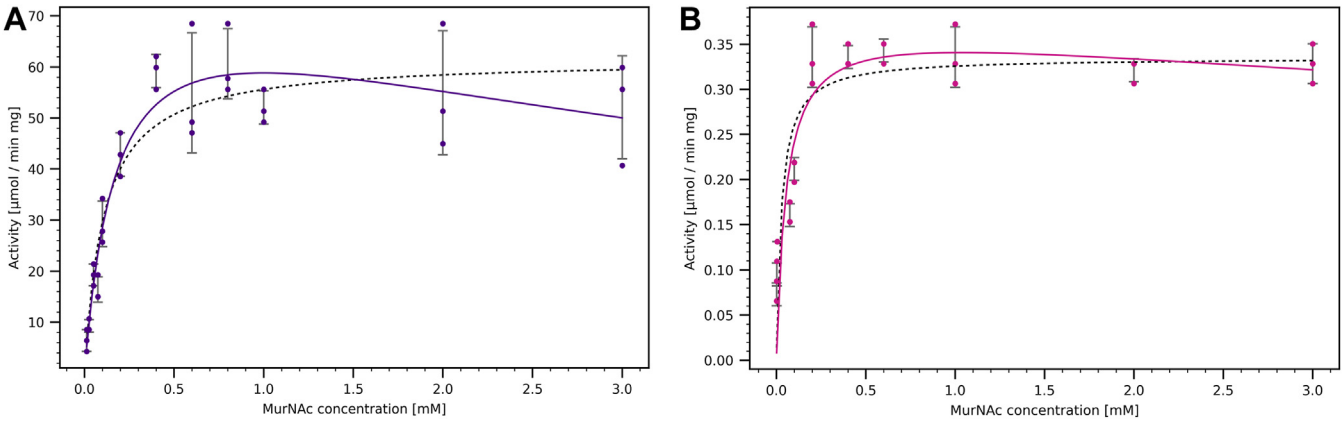


Figure 6. Kinetics of MurK and K1058 for the phosphorylation of MurNAc. The activity depending on MurNAc concentration for MurK (*A*) and K1058 (*B*) is shown. The *black dashed line* is fitted to the Michaelis–Menten model, and the *purple* (MurK) or *magenta* (K1058) line is fitted to the substrate-inhibited model. All measurements were performed in triplicate, and each point represents the fastest reaction rate observed in the UV spectra. Where only two data points are visible per concentration, the values of two points were overlapping. MurK, murein sugar kinase; MurNAc, N-acetylmuramic acid.

work presents the first structure of a MurNAc kinase, which broadens the general understanding of specificity for sugars within the kinase family. The MurK structure enabled us to identify residue patterns in the protein that are crucial to accommodate MurNAc. Based on these patterns, we are now able to determine whether a hexokinase of interest is capable of binding MurNAc or MurNAc derivatives. Although we cannot identify MurNAc binding from sequence analysis, MurNAc binding can be inferred from the vastly increasing amounts of structural data based on *in silico* folding approaches by AlphaFold2 (37). Our analysis of currently available experimental data of sugar kinases in complex with their substrates suggests that the majority of deposited hexokinases are not able to use MurNAc as a substrate due to steric clashes with bulky polar or bulky charged residues at positions corresponding to region 133 to 136 or at positions corresponding to T94 of MurK. On the other hand, the structure-based MurNAc recognition pattern enables us to hypothesize that both proteins from structural genomic consortium projects (1ZXO and 1ZBS) are kinases at least able to phosphorylate MurNAc, if MurNAc is not their natural substrate. Another example concurring with the MurNAc specificity regions is NagK. Comparing the structural data of NagK with MurK indicates no conflict for MurNAc binding and would suggest that NagK is able to accommodate MurNAc as a potential phosphate acceptor, although NagK initially was characterized as a GlcNAc kinase. Indeed, recent studies by Stafford *et al.* (16) identified NagK as being essential for the immunostimulatory activity of muramyl dipeptides by their phosphorylation at C6 of the MurNAc entity. The reaction is comparable to MurK and in excellent agreement with the specificity-controlling positions obtained from structural data of MurK. This work therefore broadens our view of how hexokinases structurally achieve their specificities for their conjugate substrates.

The catalytic turnover of MurK was previously investigated *via* TLC assays (11). The experiment showed that the kinase is capable of phosphorylating both MurNAc and GlcNAc in an extended assay with an incubation time of 16 h, and MurK showed a clear preference for MurNAc. When comparing K1058 and MurK with GlcNAc kinases, for example, human NagK (22), a structural reason for GlcNAc specificity cannot be concluded from the available data on both enzymes. Noteworthy, the only obvious difference in binding the *N*-acetyl entity of GlcNAc seems to be related to a tryptophan residue (W38 in human NagK), which is functionally replaced by a phenylalanine residue in both MurK and K1058. The preference for MurNAc over GlcNAc therefore seems to be the binding of the lactyl ether moiety.

The data presented here enable the structure-based development of MurNAc analogs as specific inhibitors of MurNAc kinases. These could be helpful to differentiate their cellular functions and to evaluate their potential as drug targets in organisms relying on such enzymes, particularly in combination with other metabolic targets. A similar approach using GlcN-derivatized inhibitors such as benzoyl glucosamine or hydroxyphenyloxopropyl glucosamine aims to block the *Leishmania* glucose kinases from *Leishmania braziliensis* for the treatment of leishmaniasis (38) or, for example, carboxybenzyl glucosamine to specifically inhibit the glucokinase and hexokinase from *T. cruzi*, a pathogen causing Chagas disease (39). The approach is promising, resulting in growth inhibition IC_{50} values in the low micromolar range (39).

In conclusion, we determined the first structure of a MurNAc kinase, MurK, from *T. forsythia*, a pathogen of the red complex associated with the late stages of periodontitis. We showed by kinetic analysis that both MurK and K1058 are capable of phosphorylating MurNAc, although with different enzymatic affinities and catalytic turnover rates. The natural substrate of K1058 remains to be determined. A structure comparison with other glucokinase and hexokinases from different sources showed relaxed steric restraints at the lactyl ether entity, which serves as a basis for the specificity of MurNAc preference. Our findings broaden our understanding of how kinases evolved substrate specificity and might lead to a structure-guided development of new MurNAc analogs to

specifically block MurNAc-dependent kinases of pathogenic bacteria for antibiotic treatment.

Experimental procedures

Protein expression and purification

The cloning of MurK has been described previously (11), and the K1058 construct was generated according to the same protocol. Briefly, an 855-bp DNA fragment containing the *T. forsythia* K1058 gene (*Tanf_RS08300)* was amplified by PCR from the bacterial genome using the forward primer 5′GCGCCATGGCGAAATTAATAGCAGAAAGCGGATCA-ACG3′ and reverse primer 5′GCGCTCGAGGCTTT CCGGATCATCCGGAGGGAAGGC3′. The PCR product and target vector pET28a (Novagen) were digested with NcoI and XhoI (New England Biolabs) restriction enzymes and ligated into pET28a with T4 DNA ligase (Thermo Fisher Scientific). To obtain large amounts of high purity proteins for crystallization trials, MurK and K1058 were expressed in *E. coli* BL21(DE3) cells. First, chemically competent cells were transformed by heat shock with plasmid DNA (approx. 100 ng) coding for the desired protein genes. Next, an overnight culture of LB medium (120 ml) containing 50 μg/ml kanamycin (all chemicals Sigma-Aldrich, unless otherwise stated) was inoculated with either a freshly transformed cell suspension or with cells from a glycerol stock. The overnight culture was then used for large scale protein production in LB-Kanamycin medium (2 L or 4 L). Here, cells were grown at 37 °C to an A_{600} of 1.1, then expression was induced by addition of IPTG (final concentration of 0.5 mM), while the temperature was reduced to 18 °C. The cells were harvested by centrifugation after 20 h of incubation and either stored at −80 °C or directly used for purification of the target protein. All purification steps were conducted at 4 °C. The cells were resuspended in immobilized metal ion affinity chromatography (IMAC) wash buffer (25 mM Tris–HCl, 500 mM NaCl, 25 mM imidazole, 10% (v/v) glycerol, pH 8.0). Cell lysis was performed by sonication. The lysate was clarified by centrifugation at 34,500*g* for 45 min, and the resulting supernatant was loaded onto a His-trap column (5 ml; GE Healthcare) pre-equilibrated with IMAC wash buffer. The column was then installed onto an ÄKTA prime system and washed with IMAC wash buffer, and IMAC wash buffer was supplemented with 5% and 10% of the elution buffer (as IMAC wash buffer, with 500 mM imidazole) until UV-monitoring showed that all impurities and unspecific binders were removed. The target proteins were eluted with 80% (v/v) elution buffer and fractionated based on inspection of UV absorbance at 280 nm. The resulting fractions were analyzed by SDS-PAGE and pooled by purity. The solution was concentrated (molecular weight cut-off of 10 kDa) by centrifugation (Amicon, Merck). Typically, 4 to 6 ml of the target protein solution was obtained at concentration between 2 to 6 mg/ml. The protein was loaded onto an SD200 16/60 SEC column equilibrated with SEC buffer MurK (200 mM NaCl, 50 mM MES, pH 6) or SEC buffer K1058 (750 mM NaCl, 25 mM Tris-HCl, pH 8.5) and eluted as 2 ml fractions. These were analyzed by SDS-PAGE, and pure protein fractions were pooled and concentrated. Typical yields were 12 mg for MurK and 5 to 10 mg for K1058 per 1 L culture. Analytical SEC was performed on an SD200 Increase 3.2/300 column pre-equilibrated with the appropriate SEC buffer by loading 50 μl of the sample and eluting it with the same buffer.

Crystallization experiments

Initial screens were performed in a 96-well plate format as sitting drop experiments at two different temperatures (4 °C and 20 °C) using either a Tecan Freedom Evo 150 (Tecan) or as Gryphon pipetting robot (Art Robins Instruments). The protein solution was mixed in a 1:1 ratio with the corresponding reservoir solution and placed over the reservoir. Initial crystallization hits from commercial screens (Molecular Dimensions; Hampton Research) were optimized for crystal quality. For K1058, the best crystals were obtained in a solution of ammonium sulfate (0.02–0.08 M; Roth), MES monohydrate (0.1 M, pH 6.3–6.9), and PEG 5000 (16%–21% (w/v)), whereas MurK crystallized in bis-tris propane/citric acid mixture (0.1 M, pH 4.5), ammonium sulfate (0.3 M), and PEG3350 (16% (w/v)). Crystals grew within a few days to weeks in both cases.

As our attempts of soaking the apo crystals were unsuccessful due to crystal packing issues, we performed cocrystallization experiments using nonhydrolyzable substrate analogs AMP-PNP (adenylyl-imidodiphosphate; Roche) or AMP-PCP and MurNAc or GlcNAc and preincubated for 30 min with the kinases to obtain substrate-bound kinase crystals. The substrate solutions were dissolved in the appropriate SEC buffer at concentrations of 100 mM (sugar substrates) or 20 mM (phosphate analogs). MurNAc, which would precipitate in the MurK SEC buffer due to the pH, was dissolved in 1 M Hepes, pH 7.5, instead. The substrate solutions were added in a 1:10 ratio to the purified kinases, resulting in final concentrations of 2 to 10 mM for the crystallization trials. We obtained MurK–AMP-PCP complex crystals in 0.1 bis-tris propane/citric acid mixture at pH 5.5, 0.12 to 0.14 M ammonium sulfate, and 14 to 18% w/v PEG3350 within weeks, while the K1058–MurNAc complex could be crystallized after 6 months in 1.5 M ammonium sulfate (Roth), 0.1 M Tris pH 8.5, 12% (v/v) glycerol, and the MurK–MurNAc complex after 9 months in 15% PEG 10000, 100 mM trisodium citrate pH 5.5, 2% v/v 1,4-dioxane. MurK/MurNAc crystals resulted from cocrystallization and grew within 6 months to maximum dimensions of 60 × 25 × 5 μm^3 in an apparent space group $I4_122$ as a result of almost perfect merohedral twinning.

X-ray diffraction experiments and structure solution

The crystals were cryoprotected by harvesting into a fresh solution of the crystallization condition with the addition of a cryoprotectant solution (20%–30% (v/v) glycerol, except for K1058/MurNAc, where 15% (v/v) glycerol was used based on (40)), mounted, and flash-frozen in liquid nitrogen for data collection at the synchrotron. Data was collected at beamline I03 at Diamond Light Source (Oxford) or at beamline X06 Da at the Swiss Light Source . Data reduction was performed

using the XDS/XSCALE package (41). The phase problem for the first structure (MurKapo) was solved by molecular replacement using PDB entry with accession code 1ZBS after adjustments to the sequence using CHAINSAW (42) and to the structure coordinates using PYMOL (43). The MurK/AMP-PCP and K1058 structures were solved using MurKapo structure for molecular replacement, while for K1058/MurNAc, K1058apo was the model and K1058/MurNAc was used for MurK/MurNAc (see below). Initial models were refined using a simulated annealing approach as implemented in PHENIX (44) to reduce model-induced bias. The models were then completed using several cycles of real space model corrections in COOT (45), followed by reciprocal space refinement as implemented in REFMAC5 (46). Solvent molecules were placed in several cycles by a combination of the COOT:Find_waters algorithm followed by refinement with REFMAC5. Unequivocal positive density was observed in a simulated annealing (F_o-F_c)-difference omit map for the three ligand-bound structures before the appropriate ligands were placed into it.

Due to merohedral twinning, the MurK/MurNAc structure was the most challenging case. The crystal structure of MurK/MurNAc was solved in space group I4 using K1058/MurNAc as a template with the sugar removed to avoid introducing bias and the C- and N-terminal domains placed individually in the crystal lattice. Further refinement involved simulated annealing with PHENIX.refine using the model target geometry of MurKapo to restrain the refinement. The simulated annealing omit map density unequivocally showed the presence of MurNAc in four out of six protomers in a virtually identical position as observed for K1058/MurNAc. The density of the remaining two protomers is generally of lower quality probably due to crystal disorder caused by twinning. The final steps of refinement for K1058, MurK, and MurK/AMP-PCP included Translation-Libration-Screw-rotation parameterization as analyzed by TLSMD (47). During the final steps of refinement, substrate molecules were placed based on a simulated annealing (F_o-F_c)-difference omit map and subsequently refined with REFMAC5.

All models were validated with tools implemented in COOT and the PDB validation server (48). Figures were generated with PYMOL, and coordinates and structure factor amplitudes were deposited in the PDB under ID codes 8QQW, 8QQX, 8OW9, 8QQK, 8OW7 for MurK, MurK in complex with AMP-PCP, MurK in complex with MurNAc, K1058, and K1058 in complex with MurNAc, respectively.

CD spectroscopy

To verify the folding of K1058, we used CD spectroscopy on a Jasco J-720 spectropolarimeter. We investigated freshly thawed K1058 enzyme that was used for enzymatic assays. To determine the folding stability, we incubated the sample for 4 days at 4 °C and repeated the measurement. The measurements were performed at a protein concentration of 0.3 mg/ml, in a glass cuvette with a light pathway of 1 mm. CD spectra were recorded by ten repeated measurements in a wavelength range of 195 nm to 250 nm. These measurements were averaged and subtracted from the buffer measurements. The resulting CD spectra were analyzed by BETSEL (49).

Enzymatic activity assays

The apparent kinetic parameters of MurK and K1058 for the phosphorylation of MurNAc with ATP were determined by a coupled enzyme assay (20). In this assay, the formation of ADP was directly, stoichiometrically coupled to NADH oxidation by pyruvate kinase and lactate dehydrogenase. In a 400 µl QS cuvette with a 10 mm light path, 50 mM Hepes (pH 7.5, Roth), 10 mM $MgCl_2$ (Merck), 5 mM ATP, 1 mM phosphoenolpyruvate, 0.2 mM NADH, 100 U/ml of pyruvate kinase, and 70 U/ml of lactate dehydrogenase (from rabbit muscle, 0.1 U of pyruvate kinase/µl and 0.07 U of lactate dehydrogenase/µl) were incubated with MurNAc at final concentrations ranging from 0.0125 or 0.025 to 3 mM for MurNAc. For every individual measurement, we used freshly thawed enzyme (MurK or K1058) from the same batch of protein preparation to maintain reproducibility. The hydrolysis of ATP was initiated by the addition of 178 ng of MurK and 17.6 µg K1058. The final volume of the reaction was 400 µl: 320 µl master mix, 40 µl protein (8.9 mg/ml diluted to 1:2000 with 1 M Hepes pH 7, 178 ng per reaction), and 40 µl MurNAc at 10 times the final concentration diluted in 50 mM Hepes pH 7. The reactions were then incubated at 20 °C for 45 min or until no change in absorbance was observed. The change of NADH absorbance was monitored at 340 nm in a spectrophotometer (WTW Photometer PhotoLab 7600 UV-Vis, Xylem Analytics). Apparent kinetic parameters were evaluated using matplotlib as implemented in python by fitting the experimental data to the Michaelis–Menten equation and Michaelis–Menten kinetics including substrate inhibition. We used a molar extinction coefficient of NADH at 340 nm of 6.22×10^3 M^{-1} cm^{-1} for calculation of the V_{max} and k_{cat} values.

Data availability

The coordinates and structure factors for the X-ray structures have been deposited in the Protein Data Bank under accession codes 8QQW, 8QQX, 8OW9, 8QQK, 8OW7 for MurK, MurK in complex with AMP-PCP, MurK in complex with MurNAc, K1058, and K1058 in complex with MurNAc, respectively.

Supporting information—This article contains supporting information (26, 50, 51).

Acknowledgments—We thank the two anonymous reviewers for their constructive comments, which have significantly improved the paper. We would like to acknowledge the Swiss Light Source (Villigen, CH) and the Diamond Light Source (Didcot, UK) for beam time and technical support. We thank Isabel Hottmann for support in the initial phase of the project. A. C. S. would like to thank Niklas J. Bayer for help fishing the crystals. Furthermore, infrastructural support from the Cluster of Excellence EXC 2124 Controlling

Microbes to Fight Infections, project ID 390838134, is kindly acknowledged.

Author contributions—A. C. S., K. G., M. B., P. F., and G. Z. investigation; A. C. S., C. M., T. S., and G. Z. formal analysis; A. C. S., C. M., T. S., and G. Z. writing–original draft.

Funding and additional information—A. S. was cofunded by the Marie Skłodowska-Curie Innovative Training Network Viral and Bacterial Adhesion Training Network (ViBrANT) under the Horizon2020 Programme, grant agreement no. 765042. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) - Project-IDs 314202130 (DACh programme) and 174858087 (GRK 1708, TP-B2) to C. M. Support for this work was provided by the DFG (TRR261), Project-ID 398967434.

Conflicts of interest—The authors declare that they no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: AMP-PCP, β-γ-methylene adenosine triphosphate; AMP-PNP, adenylyl-imidodiphosphate; ASKHA, acetate and sugar kinase/hsc70/actin; ASU, asymmetric unit; EPPIC, Evolutionary Protein Protein Interface Classifier; GHMP, galactokinase, homoserine kinase, mevalonate kinase and phosphomevalonate kinase; IMAC, immobilized metal ion affinity chromatography; MurK, murein kinase; MurNAc, N-acetylmuramic acid; MurNAc-6P, *N*-acetylmuramic acid 6 phosphate; MurQ, MurNAc 6-phosphate etherase; NagK, *N*-acetylglucosamine kinase; NOD2, nucleotide-binding oligomerization domain-containing protein 2; PDB, protein data bank; PGN, peptidoglycan; PISA, Proteins, Interfaces, Structures and Assemblies; ROK, repressor-ORF kinase; SEC, size-exclusion chromatography.

References

1. Holt, S. C., and Ebersole, J. L. (2005) Porphyromonas gingivalis, Treponema denticola, and Tannerella forsythia: the "red complex", a prototype polybacterial pathogenic consortium in periodontitis. *Periodontol. 2000* **38**, 72–122
2. Mayer, V. M. T., Hottmann, I., Figl, R., Altmann, F., Mayer, C., and Schäffer, C. (2019) Peptidoglycan-type analysis of the N-acetylmuramic acid auxotrophic oral pathogen Tannerella forsythia and reclassification of the peptidoglycan-type of Porphyromonas gingivalis. *BMC Microbiol.* **19**, 1–9
3. Wyss, C. (1989) Dependence of proliferation of Bacteroides forsythus on exogenous N-acetylmuramic acid. *Infect. Immun.* **57**, 1757–1759
4. Hudspeth, M. K., Hunt Gerardo, S., Maiden, M. F. J., Citron, D. M., and Goldstein, E. J. C. (1999) Characterization of Bacteroides forsythus strains from cat and dog bite wounds in humans and comparison with monkey and human oral strains. *J. Clin. Microbiol.* **37**, 2003–2006
5. Hottmann, I., Borisova, M., Schäffer, C., and Mayer, C. (2021) Peptidoglycan salvage enables the periodontal pathogen tannerella forsythia to survive within the oral microbial community. *Microb. Physiol.* **31**, 123–134
6. Ruscitto, A., Honma, K., Veeramachineni, V. M., Nishikawa, K., Stafford, G. P., and Sharma, A. (2017) Regulation and molecular basis of environmental muropeptide uptake and utilization in fastidious oral anaerobe tannerella forsythia. *Front. Microbiol.* **8**, 648
7. Mayer, V. M. T., Tomek, M. B., Figl, R., Borisova, M., Hottmann, I., Blaukopf, M., *et al.* (2020) Utilization of different MurNAc sources by the oral pathogen Tannerella forsythia and role of the inner membrane transporter AmpG. *BMC Microbiol.* **20**, 1–15
8. Borisova, M., Balbuchta, K., Lovering, A., Titz, A., and Mayer, C. (2022) NamZ1 and NamZ2 from the oral pathogen Tannerella forsythia are peptidoglycan processing Exo-b-N-acetylmuramidases with distinct substrate specificities. *J. Bacteriol.* **204**, 597–618
9. Ruscitto, A., Hottmann, I., Stafford, G. P., Schäffer, C., Mayer, C., and Sharma, A. (2016) Identification of a novel N-acetylmuramic acid transporter in Tannerella forsythia. *J. Bacteriol.* **198**, 3119–3125
10. Renner-Schneck, M., Hinderberger, I., Gisin, J., Exner, T., Mayer, C., and Stehle, T. (2015) Crystal structure of the N-acetylmuramic acid α-1-phosphate (MurNAc-α1-P) uridylyltransferase MurU, a minimal sugar nucleotidyltransferase and potential drug target enzyme in Gram-negative pathogens. *J. Biol. Chem.* **290**, 10804–10813
11. Hottmann, I., Mayer, V. M. T., Tomek, M. B., Friedrich, V., Calvert, M. B., Titz, A., *et al.* (2018) N-acetylmuramic acid (MurNAc) auxotrophy of the oral pathogen Tannerella forsythia: characterization of a MurNAc kinase and analysis of its role in cell wall metabolism. *Front. Microbiol.* **9**, 1–12
12. Ruscitto, A., and Sharma, A. (2018) Peptidoglycan synthesis in Tannerella forsythia: scavenging is the modus operandi. *Mol. Oral Microbiol.* **33**, 125–132
13. Conway, L. P., and Voglmeir, J. (2016) Functional analysis of anomeric sugar kinases. *Carbohydr. Res.* **432**, 23–30
14. Roy, S., Vega, M. V., and Harmer, N. J. (2019) Carbohydrate kinases: a conserved mechanism across differing folds. *Catalysts* **9**, 29
15. Rodionova, I. A., Yang, C., Li, X., Kurnasov, O. V., Best, A. A., Osterman, A. L., *et al.* (2012) Diversity and versatility of the thermotoga maritima sugar Kinome. *J. Bacteriol.* **194**, 5552–5563
16. Stafford, C. A., Gassauer, A., de Oliveira Mann, C. C., Tanzer, M. C., Fessler, E., Wefers, B., *et al.* (2022) Phosphorylation of muramyl peptides by NAGK is required for NOD2 activation. *Nature* **609**, 590–596
17. Vollmer, W., and Bertsche, U. (2008) Murein (peptidoglycan) structure, architecture and biosynthesis in Escherichia coli. *Biochim. Biophys. Acta* **1778**, 1714–1734
18. Dik, D. A., Fisher, J. F., and Mobashery, S. (2018) Cell-wall recycling of the gram-negative bacteria and the Nexus to antibiotic resistance. *Chem. Rev.* **118**, 5952–5984
19. Johnson, J. W., Fisher, J. F., and Mobashery, S. (2013) Bacterial cell-wall recycling. *Ann. N. Y. Acad. Sci.* **1277**, 54–75
20. Reith, J., Berking, A., and Mayer, C. (2011) Characterization of an N-acetylmuramic acid/N-acetylglucosamine kinase of clostridium acetobutylicum. *J. Bacteriol.* **193**, 5386–5392
21. Cheek, S., Zhang, H., and Grishin, N. V. (2002) Sequence and structure classification of kinases. *J. Mol. Biol.* **320**, 855–881
22. Weihofen, W. A., Berger, M., Chen, H., Saenger, W., and Hinderlich, S. (2006) Structures of human N-acetylglucosamine kinase in two complexes with N-acetylglucosamine and with ADP/glucose: insights into substrate specificity and regulation. *J. Mol. Biol.* **364**, 388–399
23. Duarte, J. M., Bliven, S., Lafita, A., Capitani, G., and Burley, S. K. (2017) Automated evaluation of quaternary structures from protein crystal structures. *Acta Cryst.* **73**, a117
24. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797
25. Martinez, J., Nguyen, L. D., Hinderlich, S., Zimmer, R., Tauberger, E., Reutter, W., *et al.* (2012) Crystal structures of N-acetylmannosamine kinase provide insights into enzyme activity and inhibition. *J. Biol. Chem.* **287**, 13656–13665
26. Holm, L. (2022) Dali server: structural unification of protein families. *Nucleic Acids Res.* **50**, W210–W215
27. Veevers, R., and Hayward, S. (2019) Methodological improvements for the analysis of domain movements in large biomolecular complexes. *Biophys. Physicobiol.* **16**, 328–336
28. Nishimasu, H., Fushinobu, S., Shoun, H., and Wakagi, T. (2007) Crystal structures of an ATP-dependent hexokinase with broad substrate specificity from the hyperthermophilic archaeon Sulfolobus tokodaii. *J. Biol. Chem.* **282**, 9923–9931
29. Gangi Setty, T., Sarkar, A., Coombes, D., Dobson, R. C. J., and Subramanian, R. (2020) Structure and function of N-acetylmannosamine kinases from pathogenic bacteria. *ACS Omega* **5**, 30923–30936
30. Nocek, B., Stein, A. J., Jedrzejczak, R., Cuff, M. E., Li, H., Volkart, L., *et al.* (2011) Structural studies of ROK fructokinase YdhR from Bacillus

subtilis: insights into substrate binding and fructose specificity. *J. Mol. Biol.* **406**, 325–342
31. Milanes, J. E., Suryadi, J., Abendroth, J., Van Voorhis, W. C., Barrett, K. F., Dranow, D. M., *et al.* (2019) Enzymatic and structural characterization of the naegleria fowleri glucokinase. *Antimicrob. Agents Chemother.* **63**, 1–16
32. Rosano, C., Sabini, E., Rizzi, M., Deriu, D., Murshudov, G., Bianchi, M., *et al.* (1999) Binding of non-catalytic ATP to human hexokinase I highlights the structural components for enzyme-membrane association control. *Structure* **7**, 1427–1437
33. Mukai, T., Kawai, S., Mori, S., Mikami, B., and Murata, K. (2004) Crystal structure of bacterial inorganic polyphosphate/ATP- glucomannokinase: insights into kinase evolution. *J. Biol. Chem.* **279**, 50591–50600
34. Miyazono, K., Tabei, N., Morita, S., Ohnishi, Y., Horinouchi, S., and Tanokura, M. (2012) Substrate recognition mechanism and substrate-dependent conformational changes of an ROK family glucokinase from streptomyces griseus. *J. Bacteriol.* **194**, 607–616
35. Arora, K. K., Filburn, C. R., and Pedersen, P. L. (1991) Glucose phosphorylation. Site-directed mutations which impair the catalytic function of hexokinase. *J. Biol. Chem.* **266**, 5359–5362
36. Aleshin, A. E., Kirby, C., Liu, X., Bourenkov, G. P., Bartunik, H. D., Fromm, H. J., *et al.* (2000) Crystal structures of mutant monomeric hexokinase I reveal multiple ADP binding sites and conformational changes revelant to allosteric regulation. *J. Mol. Biol.* **296**, 1001–1015
37. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589
38. Buechner, G. S., Millington, M. E., Perry, K., and D'Antonio, E. L. (2019) The crystal structure of glucokinase from Leishmania braziliensis. *Mol. Biochem. Parasitol.* **227**, 47–52
39. D'Antonio, E. L., Deinema, M. S., Kearns, S. P., Frey, T. A., Tanghe, S., Perry, K., *et al.* (2015) Structure-based approach to the identification of a novel group of selective glucosamine analogue inhibitors of Trypanosoma cruzi glucokinase. *Mol. Biochem. Parasitol.* **204**, 64–76
40. McFerrin, M. B., and Snell, E. H. (2002) The development and application of a method to quantify the quality of cryoprotectant solutions using standard area-detector X-ray images. *J. Appl. Crystallogr.* **35**, 538–545
41. Kabsch, W. (2010) Xds. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132
42. Stein, N. (2008) Chainsaw : a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Crystallogr.* **41**, 641–643
43. *The PyMOL Molecular Graphics System, Version~2.5.* (2021). Schrödinger LCC, New York, NY
44. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213–221
45. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501
46. Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., *et al.* (2011) REFMAC 5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367
47. Painter, J., and Merritt, E. A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D. Biol. Crystallogr.* **62**, 439–450
48. Gore, S., Sanz García, E., Hendrickx, P. M. S., Gutmanas, A., Westbrook, J. D., Yang, H., *et al.* (2017) Validation of structures in the protein data bank. *Structure* **25**, 1916–1927
49. Micsonai, A., Moussong, É., Wien, F., Boros, E., Vadászi, H., Murvai, N., *et al.* (2022) BeStSel: webserver for secondary structure and fold prediction for protein CD spectroscopy. *Nucleic Acids Res.* **50**, W90–W98
50. Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., *et al.* (2022) Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* **50**, W276–W279
51. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, 320–324

Supporting information

N-acetylmuramic acid recognition by MurK kinase from the MurNAc auxotrophic oral pathogen *Tannerella forsythia*

Authors:

Aleksandra Cecylia Stasiak[1], Karolin Gogler[1], Marina Borisova[2], Phillipp Fink[1], Christoph Mayer[2], Thilo Stehle[1], and Georg Zocher[1*]

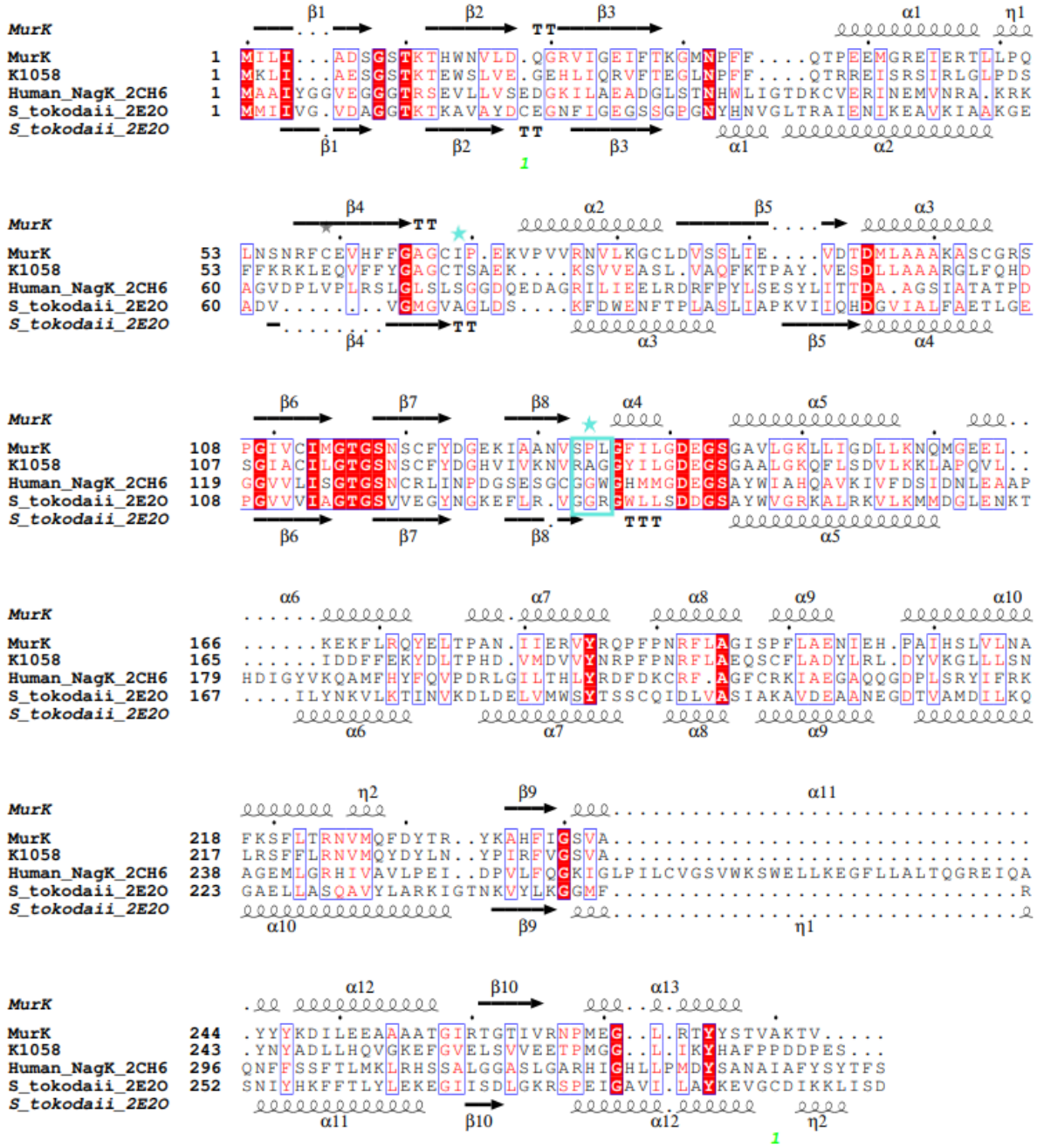
[1] Interfaculty Institute of Biochemistry, University of Tuebingen, D-72076 Tuebingen, Germany

[2] Interfaculty Institute of Microbiology and Infection Medicine, Organismic Interactions / Glycobiology, University of Tuebingen, D-72076 Tuebingen, Germany

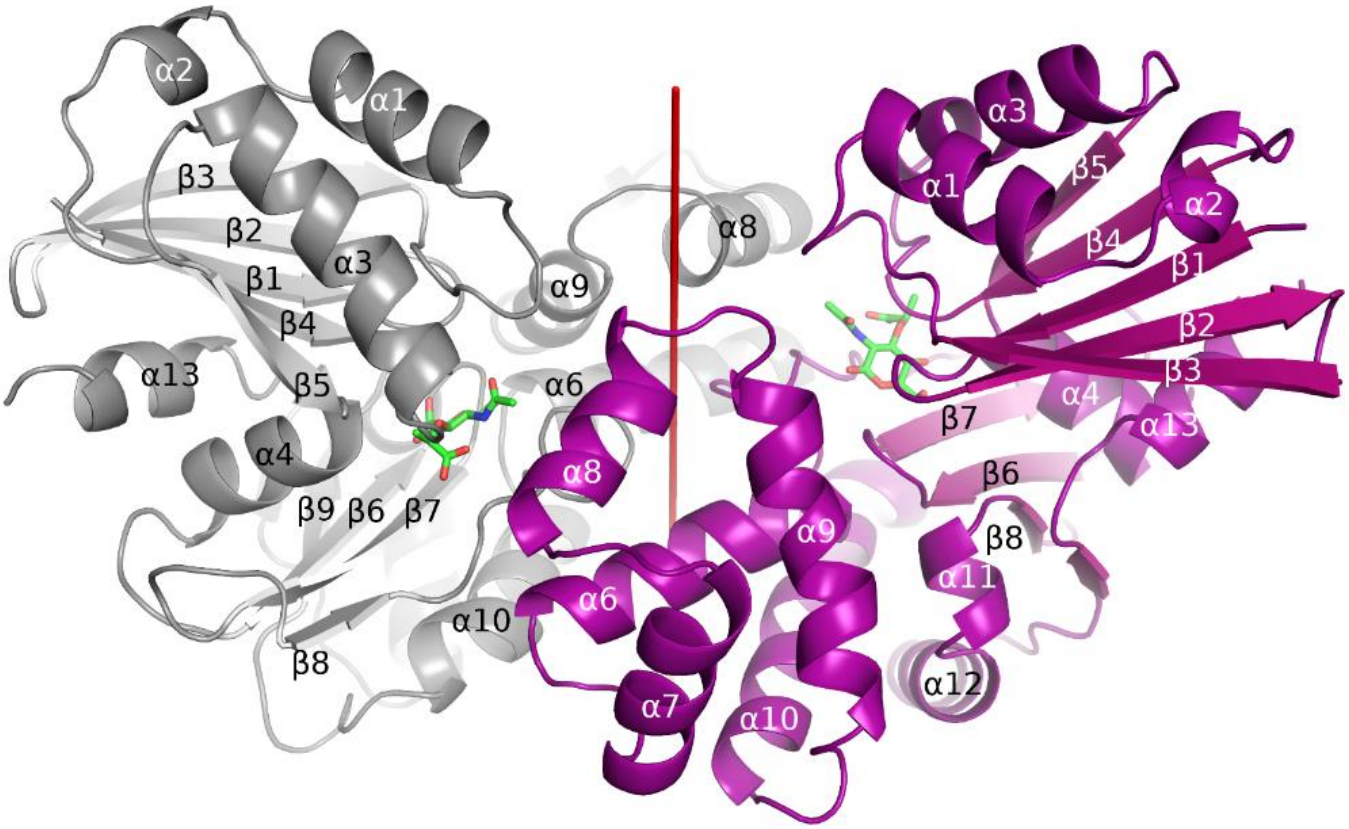
* Corresponding author: Dr. Georg Zocher (Interfaculty Institute of Biochemistry, University of Tuebingen, 72076 Tuebingen, Germany, e-mail: georg.zocher@uni-tuebingen.de, phone: +49 7071 2974762)

Running title: Structural analysis of MurNAc kinases

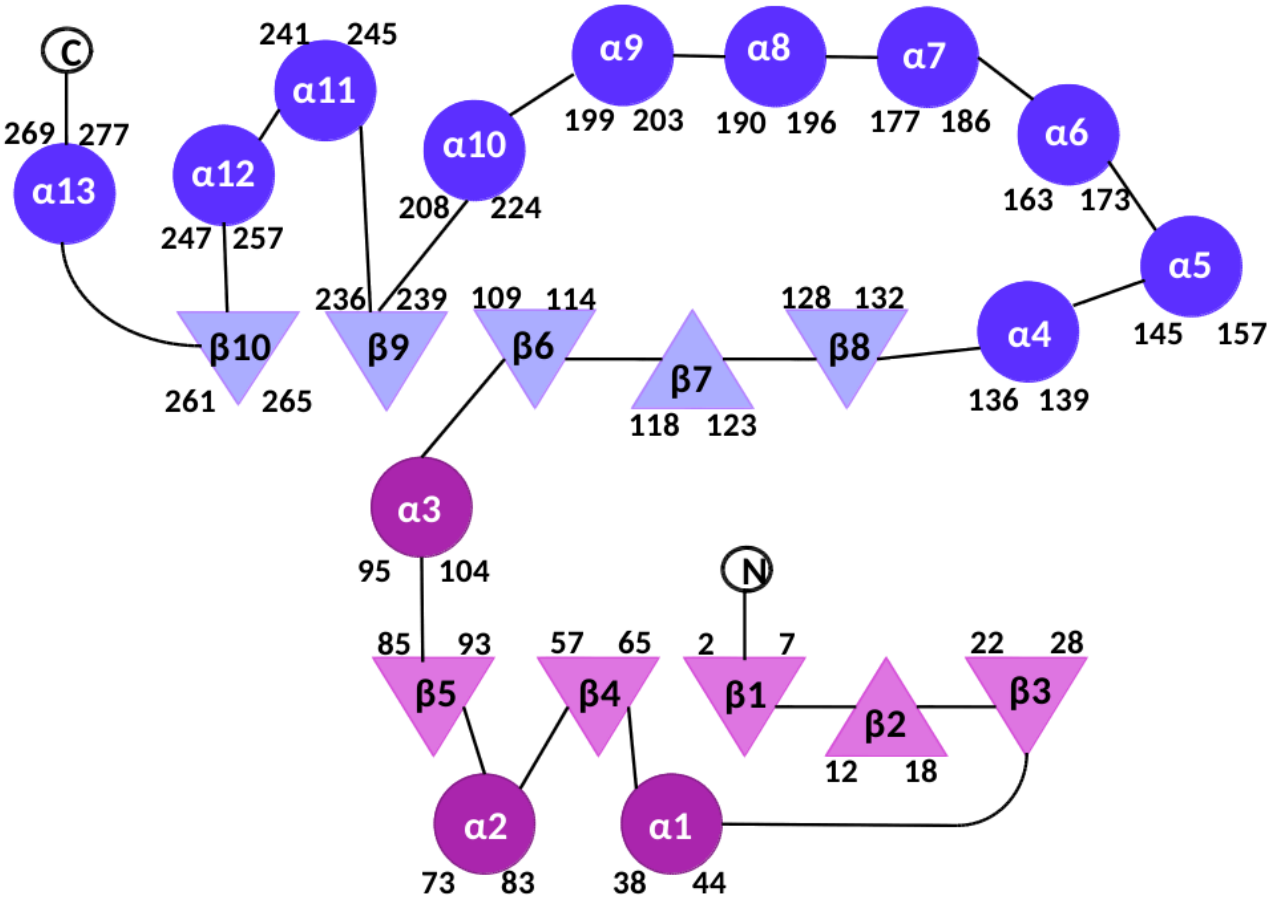
Keywords: ATPase, bacterial metabolism, cell wall, cell wall recycling, MurNAc kinase, periodontal disease, protein structure



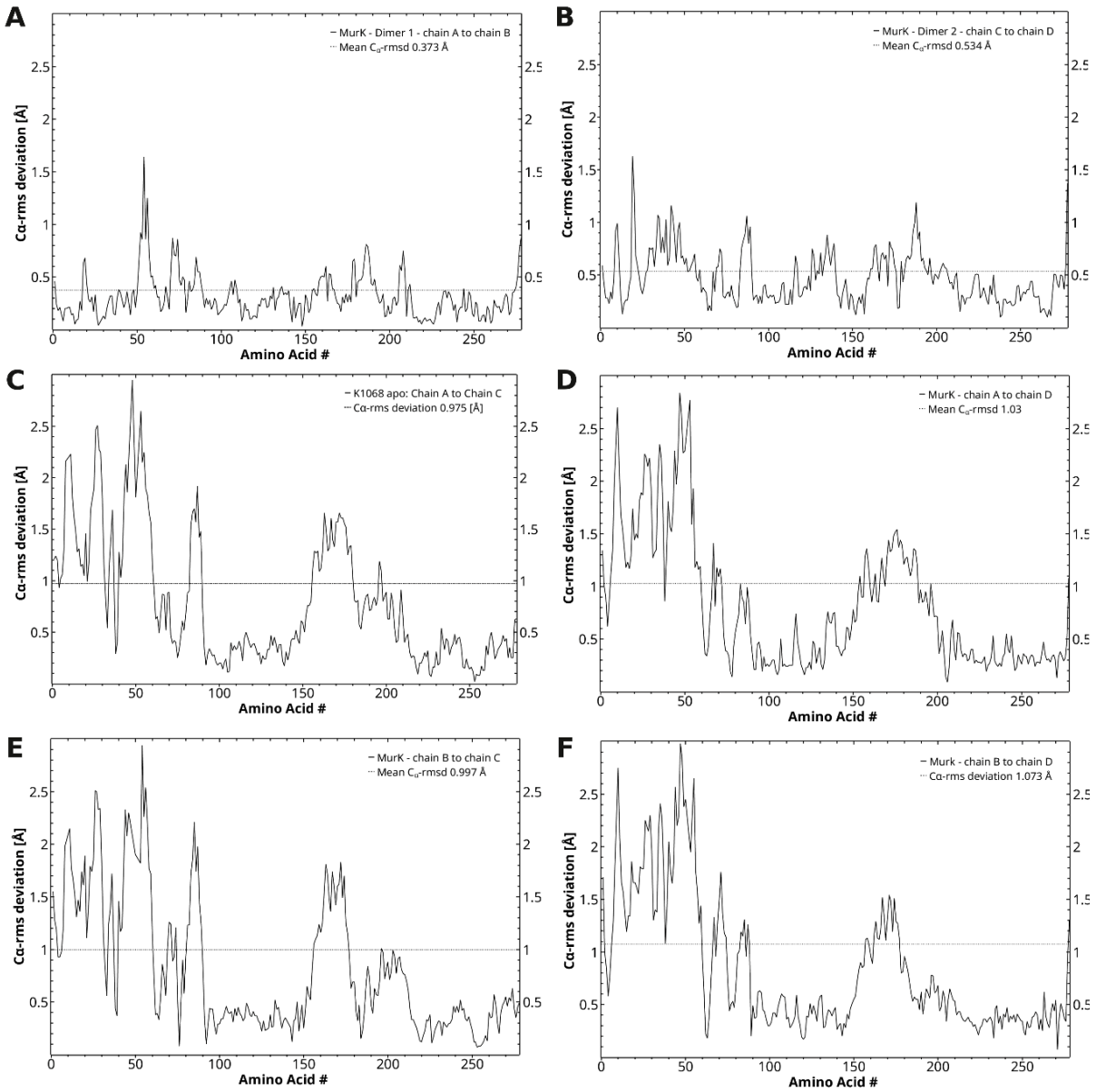
Supplementary Figure S1. Sequence alignment of MurK and K1058 with two closest structural homologues with known functions. The human N-acetylglucosamine kinase NagK (PDB ID: 2CH6) and the archaeal *S. tokodaii* broad-specificity sugar kinase (PDB ID: 2E2O). Cyan border and star indicate residues important for MurNAc recognition. Alignment generated with MUSCLE (50), figure with ESPRIPT (51).



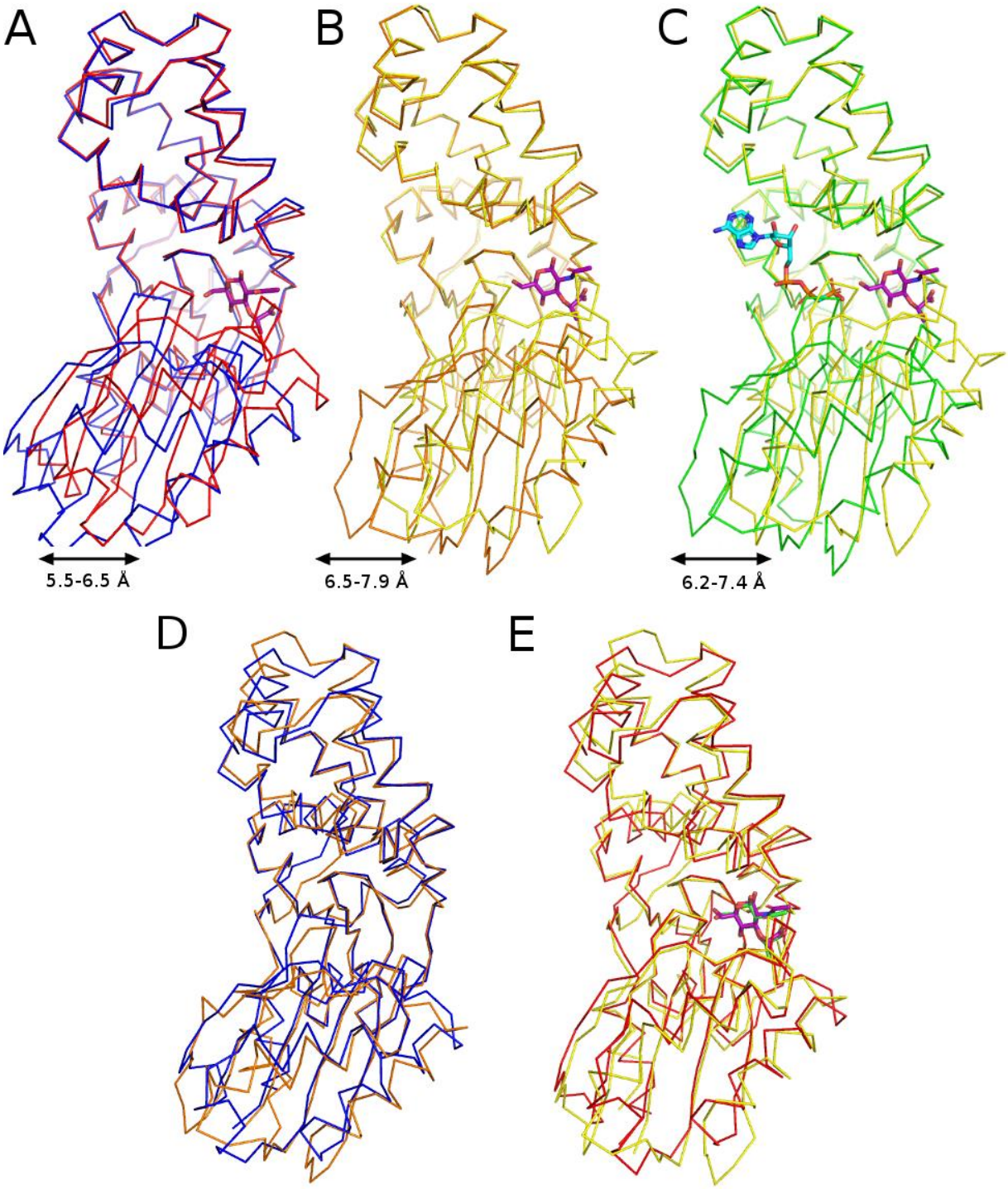
Supplementary Figure S2. The structure of K1058. Cartoon representation of the biological K1058 dimer (one chain in magenta, one in grey) in complex with the MurNAc substrate (green). The red line shows the two-fold symmetry axis.



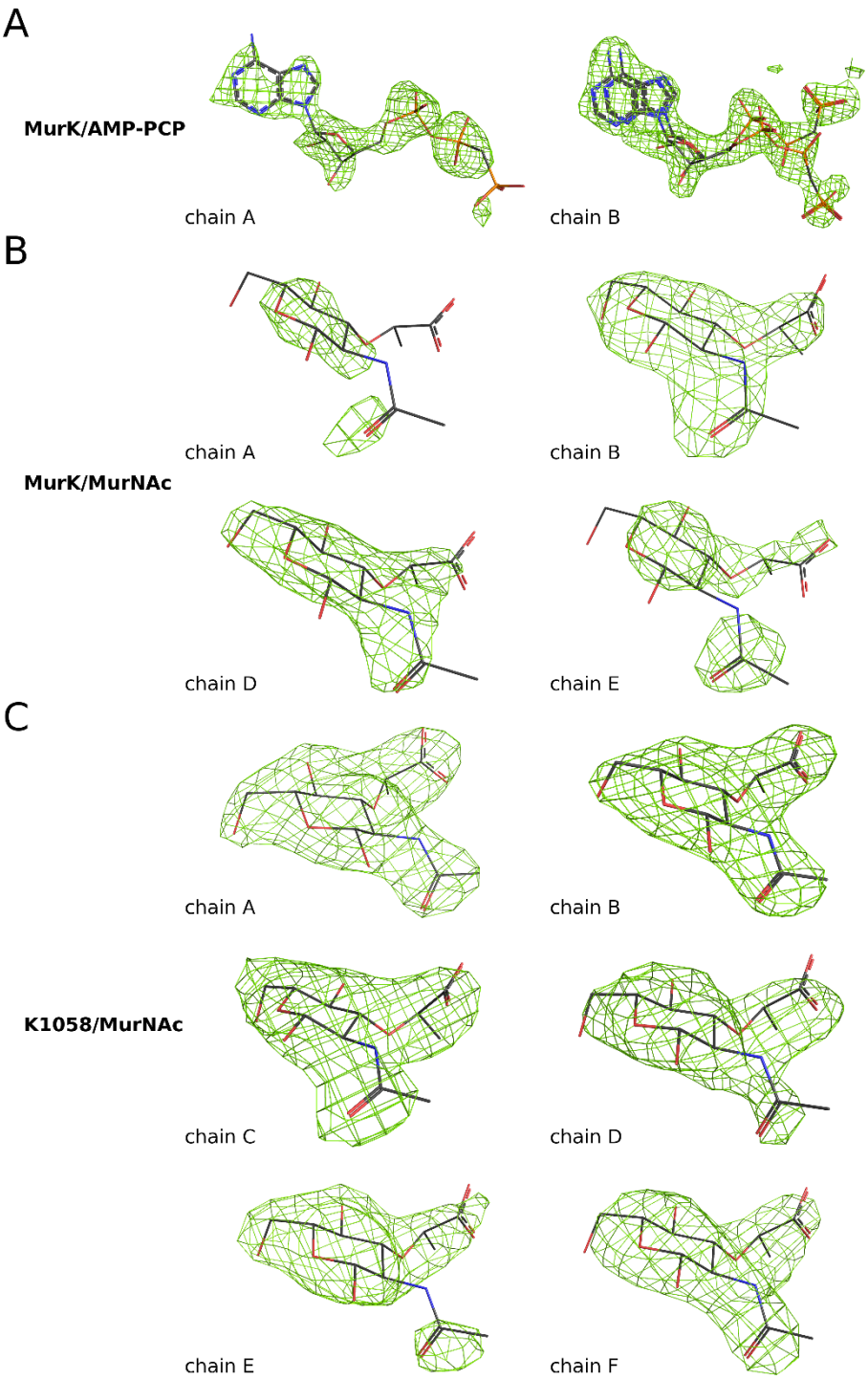
Supplementary Figure S3. Topology plot of MurK. N-terminal domain residues in magenta, C-terminal domain residues in violet.



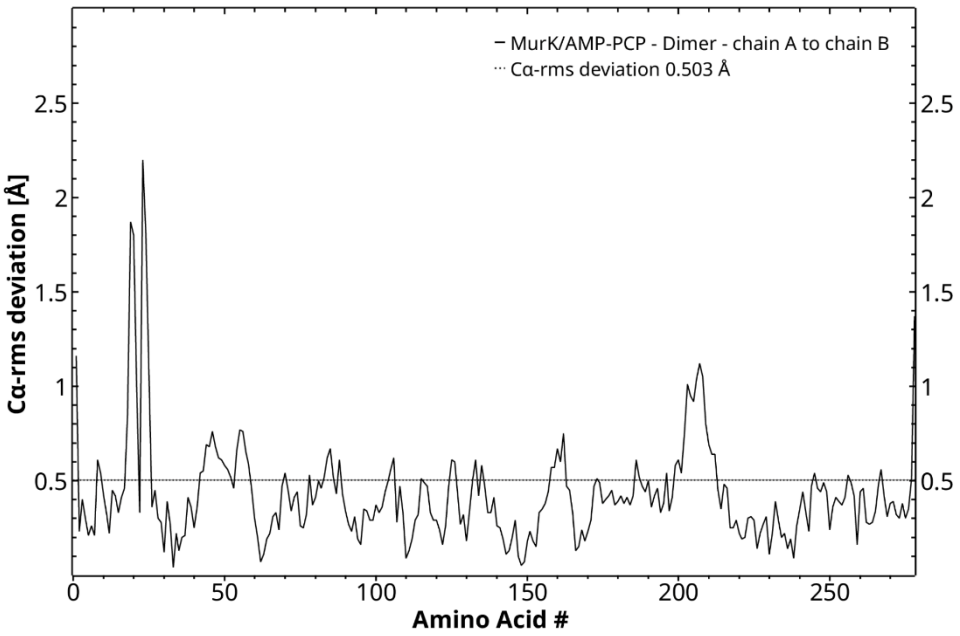
Supplementary Figure S4. Structural differences of the individual chains in the crystal structure of MurK. The triclinic crystal contains four protomers in the unit cell. Each chain was superposed on all others (A to F) and its corresponding amino acid C_α-RMSD was plotted. Both protomers in a biological dimer (chains A, B and chains C, D) possess a virtually identical conformation, as indicated by a very low mean C_α-RMSD. Comparing the chains of a dimer to the adjacent dimer showed a different conformation by increased C_α-RMSD values, especially in the flexible N-terminal domain and residues 160 to 180 of the C-terminal domain.



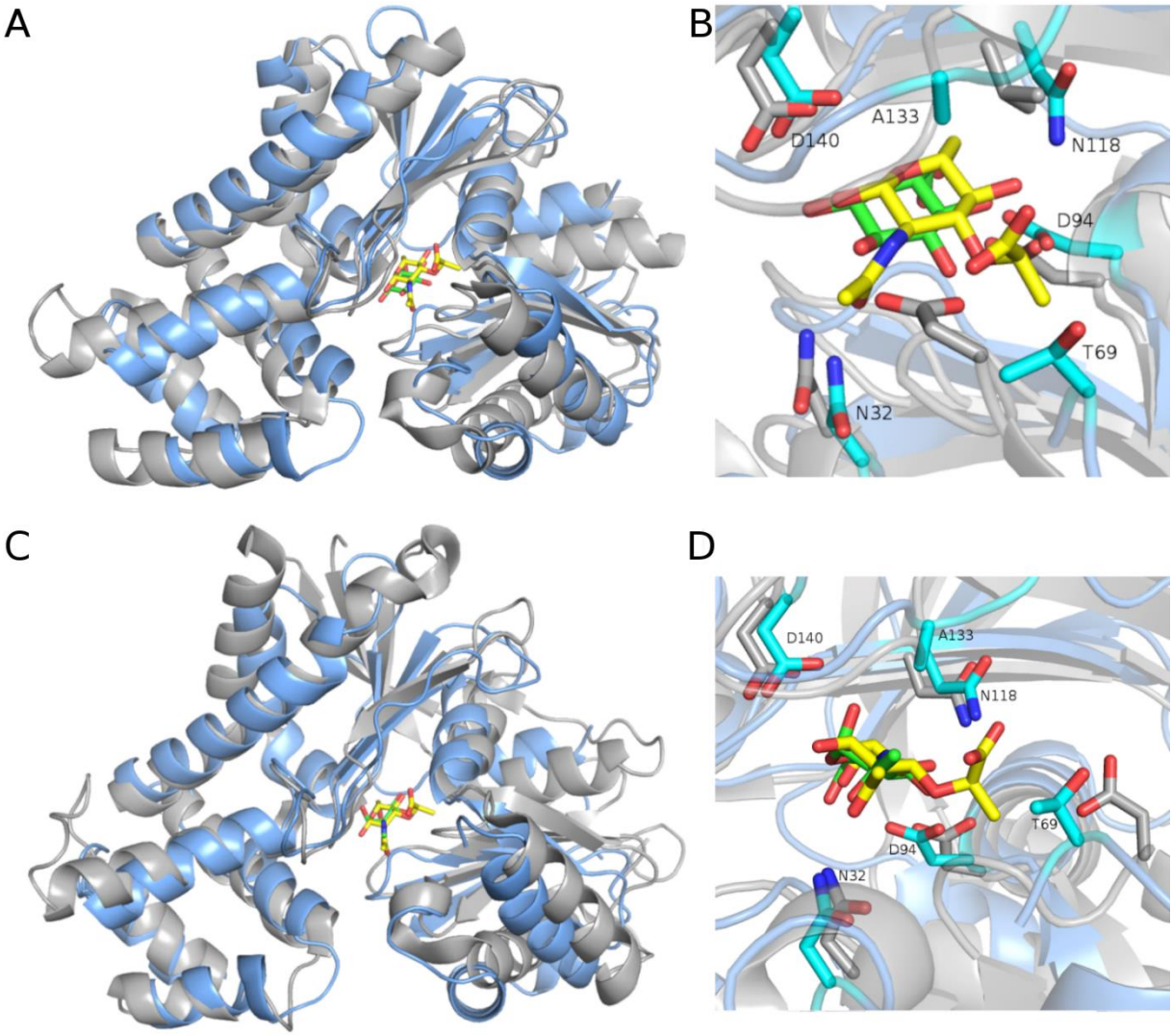
Supplementary Figure S5. N-terminal domain movement on MurNAc binding in MurK and K1058. (A) K1058 apo (blue) and K1058/MurNAc (red), superimposed to residues 91-272. (B) MurK apo (orange) and MurK/MurNAc (yellow), superimposed to residues 91-273. (C) MurK/AMP-PCP (green) and MurK/MurNAc (yellow), superimposed to residues 91-273. (D) MurK apostructure (orange) and K1058 apostructure (blue), superimposed on the C-terminal domains. (E) MurK/MurNAc (yellow) and K1058/MurNAc (red), superimposed on C-terminal domains.



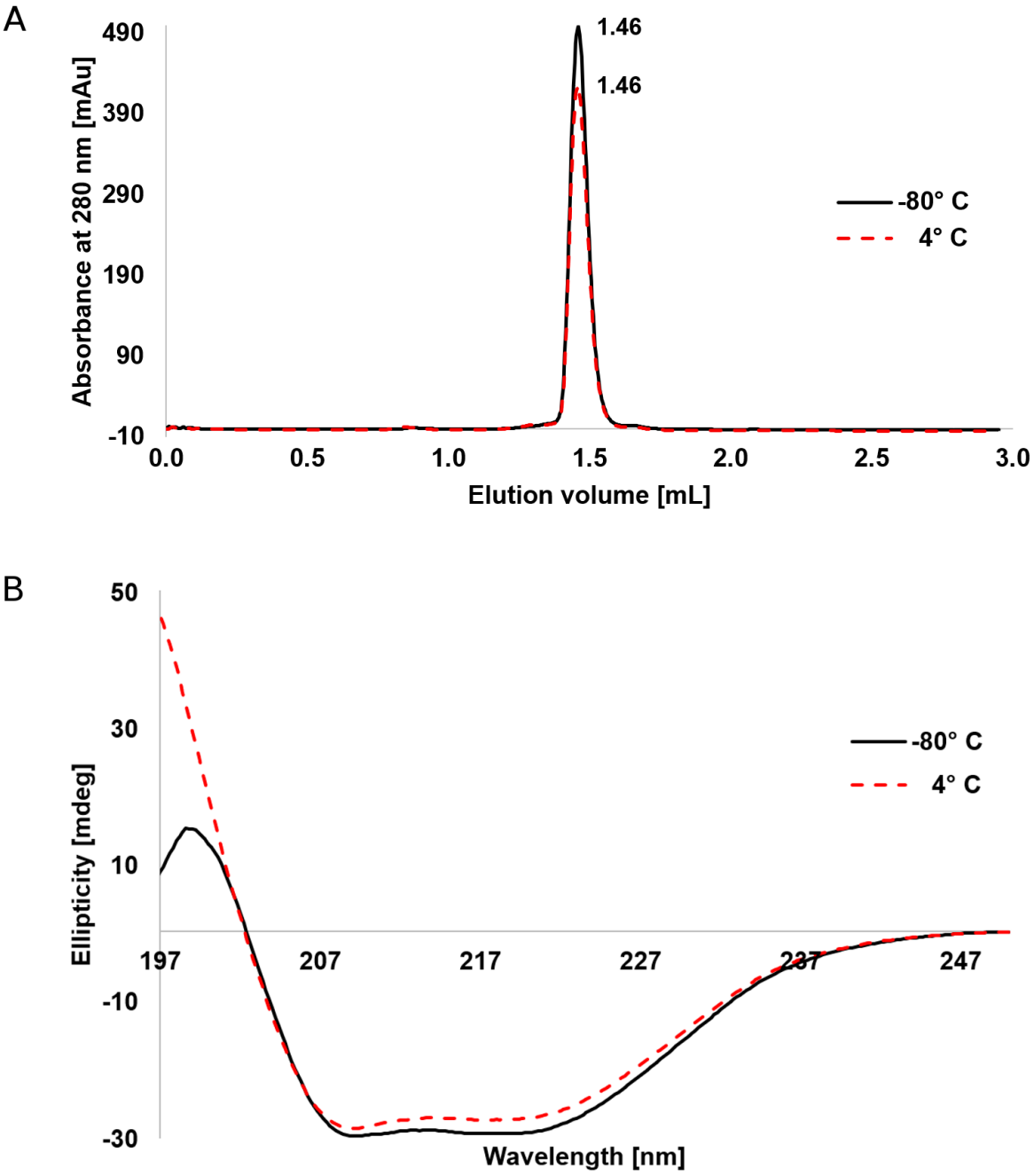
Supplementary Figure S6. **Substrate densities in the complex structures of MurK and K1058**. Unbiased simulated annealing difference electron density omit maps (green) at a contour level of 2.5σ or 3.0σ for AMP-PCP **(A)** or MurNAc in MurK/MurNAc and K1058/MurNAc **(B and C)**, respectively. Except for MurK/MurNAc, the electron density unequivocally shows the presence of the corresponding ligand in all chains of the asymmetric units. For the twinned crystal MurK/MurNAc the electron density is weak for two chains in general (chain C and F) and does not allow placing MurNAc in these chains. The density is acceptable for chain A and E, where MurNAc was placed. In chains B and D, the electron density unequivocally shows the presence of the substrate MurNAc and its conformation in the active site which is similar as observed for K1058/MurNAc.



Supplementary Figure S7. **Structural differences of the individual chains in the crystal of MurK/AMPPCP**. The monoclinic crystal contains a biological dimer in the asymmetric unit. Chain A was superposed to chain B and the corresponding C_α-RMSD was plotted. Both protomers possess a virtually identical conformation with a mean C_α-RMSD of 0.5 Å.



Supplementary Figure S8. Structural comparison of K1058. (A) cartoon representation of K1058 (blue) superimposed with the broad-specificity glucokinase from *S. tokoidaii* (2E2O; grey) on the C-terminal. (B) Close-up of the active site from (A), with active site residues of K1058 labelled. (C) cartoon representation of K1058 (blue) superimposed with the human NagK kinase (2CH6; grey). (D) Close-up of the active site from (C), with active site residues of K1058 labelled.



Supplementary Figure S9. Stability experiments and CD spectroscopy of K1058. (A) Kinase stability was investigated by analytical SEC using a SD200 increase 3.2/300 column. Frozen-thawed enzyme is compared to a sample that was stored at 4°C for four days. Although the intensities decreased within 4 days by 16%, no evident formation of aggregates could be observed. The elution volume of 1.46 mL corresponds to an apparent molecular mass of 60 kDa. (B) Both protein samples were investigated by CD spectroscopy to verify their folding. Although no severe stability issues were observed all enzymatic assay data were obtained from freshly thawed enzyme. Analysis of the CD spectra of both samples revealed an α-helical content of 36.7% and an antiparallel β-strand content of 18.7% for the freshly thawed K1058. After 4 days at 4 °C only subtle changes were detected yielding to an α-helical content of 35.6% and an antiparallel β-strand content of 12.9%.

Supplementary Table S1. Percentage sequence similarity between MurK and K058 and their closest characterised structural homologues in Supplementary Figure 1.

Protein	MurK	K1058	Human_NagK_2CH6	S_tokodaii_2E2O
MurK	100.00	44.24	17.50	22.22
K1058	44.24	100.00	16.73	17.82
Human_NagK_2CH6	17.50	16.73	100.00	23.81
S_tokodaii_2E2O	22.22	17.82	23.81	100.00

Supplementary Table S2. List of residues involved in the formation of the dimeric interface of MurK identified by PISA. The solvent accessible area (SAA), the buried surface area (BSA) and the fraction of the total surface area of the individual amino acid (AA) are given. Only residues with a percentage buried surface area over 80% are listed.

Structure	Chain:Residue	SAA [Å^2]	BSA [Å^2]	Fraction of BSA per AA
MurK APO	A:ARG 21	138	121	88%
MurK APO	A:PHE 27	110	104	95%
MurK APO	A:THR 48	47	40	84%
MurK APO	A:GLN 172	75	63	85%
MurK APO	A:TYR 173	59	57	97%
MurK APO	A:PHE 193	24	19	82%
MurK APO	A:ILE 197	8	8	100%
MurK APO	B:ARG 21	135	112	83%
MurK APO	B:PHE 27	110	102	93%
MurK APO	B:THR 48	34	29	86%
MurK APO	B:TYR 173	53	52	99%
MurK APO	B:ILE 197	8	8	100%
MurK APO	B:SER 277	38	30	81%

Supplementary Table S3. List of residues involved in the formation of the dimeric interface of MurK/AMPPCP identified by PISA. The SAA, the BSA and the fraction of the total surface area of the individual amino acid (AA) are given. Only residues with a percentage buried surface area over 80% are listed.

Structure	Chain:Residue	SAA [Å^2]	BSA [Å^2]	Fraction of BSA per AA
MurK AMPPCP	A:LEU 135	111	97	88%
MurK AMPPCP	A:ILE 138	124	114	91%
MurK AMPPCP	A:LEU 139	133	123	92%
MurK AMPPCP	A:ILE 153	14	13	96%
MurK AMPPCP	A:GLY 154	34	32	96%
MurK AMPPCP	A:LYS 158	155	153	99%
MurK AMPPCP	A:VAL 184	27	25	91%
MurK AMPPCP	A:LEU 194	3	3	100%
MurK AMPPCP	B:LEU 135	112	96	86%
MurK AMPPCP	B:ILE 138	126	115	91%
MurK AMPPCP	B:LEU 139	130	119	91%
MurK AMPPCP	B:ILE 153	14	13	95%
MurK AMPPCP	B:GLY 154	36	34	96%
MurK AMPPCP	B:LYS 158	155	154	99%
MurK AMPPCP	B:ILE 180	2	2	93%
MurK AMPPCP	B:VAL 184	27	25	93%
MurK AMPPCP	B:LEU 194	3	3	100%

Supplementary Table S4. List of residues involved in the formation of the dimeric interface of MurK/MurNAc identified by PISA. The SAA, the BSA and the fraction of the total surface area of the individual amino acid (AA) are given. Only residues with a percentage buried surface area over 80% are listed.

Structure	Chain:Residue	SAA [Å^2]	BSA [Å^2]	Fraction of BSA per AA
MurK MurNAc	A:PRO 33	14	14	100%
MurK MurNAc	A:PHE 34	109	95	86%
MurK MurNAc	A:LEU 135	113	98	87%
MurK MurNAc	A:ILE 138	130	114	88%
MurK MurNAc	A:LEU 139	131	117	90%
MurK MurNAc	A:ILE 153	14	14	95%
MurK MurNAc	A:GLY 154	32	31	96%
MurK MurNAc	A:LYS 158	142	142	100%
MurK MurNAc	A:ILE 180	7	7	98%
MurK MurNAc	A:VAL 184	30	26	85%
MurK MurNAc	A:LEU 194	7	7	100%
MurK MurNAc	A:ASN 225	4	4	100%
MurK MurNAc	B:PRO 33	20	18	91%
MurK MurNAc	B:PHE 34	107	87	82%
MurK MurNAc	B:LEU 135	115	104	91%
MurK MurNAc	B:ILE 138	133	122	92%
MurK MurNAc	B:LEU 139	127	111	87%
MurK MurNAc	B:ILE 153	14	14	94%
MurK MurNAc	B:GLY 154	32	31	97%
MurK MurNAc	B:LYS 158	147	147	100%
MurK MurNAc	B:ILE 180	10	9	90%
MurK MurNAc	B:VAL 184	33	30	90%
MurK MurNAc	B:TYR 185	171	138	81%
MurK MurNAc	B:LEU 194	6	6	100%
MurK MurNAc	B:ASN 225	6	6	100%

Supplementary Table S5. Highest-scoring DALI (26) matches for MurK and K1058 apo structures, sorted by Z-score.

Name	Organism	PDB Code	Z-score	RMSD [Å]	Similarity to
Conserved hypothetical protein Q8A1P1	*Bacteroides thetaiotaomicron*	1zxo	40.5	1.1	MurK
Putative N-acetylglucosamine kinase (PG1100)	*Porphyromonas gingivalis*	1zbs	40.3	1.1	MurK
N-acetylglucosamine kinase (NAGK)	Human	2ch5	23	2.9	MurK
Broad-specificity hexokinase	*Sulfolobus tokodaii*	2e2o	23	3	MurK
1.1.1.1. Putative N-acetylglucosamine kinase	*Chromobacterium violaceum*	1zc6	22.8	3.1	MurK
Cell shape-determining protein MREB	*Spiroplasma eriocheiris*	7e1g	17.9	2.7	MurK
N-acetylmannosamine kinase	*Haemophilus influenzae*	6jdb	17	3.5	MurK
Glucokinase	*Streptomyces griseus*	3vgl	16.6	3.2	MurK
α-skeletal actin	Rabbit	2yjf	15.6	3.3	MurK
Conserved hypothetical protein Q8A1P1	*Bacteroides thetaiotaomicron*	1zxo	39.1	1.2	K1058
Putative N-acetylglucosamine kinase (PG1100	*Porphyromonas gingivalis*	1zbs	38.1	1.5	K1058
Broad-specificity hexokinase	*Sulfolobus tokodaii*	2e2o	22.9	2.9	K1058
1.1.1.2. Putative N-acetylglucosamine kinase	*Chromobacterium violaceum*	1zc6	22.9	2.8	K1058
N-acetylglucosamine kinase (NAGK)	Human	2ch5	22.4	3	K1058
N-acetylmannosamine kinase	*Haemophilus influenzae*	6jda	17.1	2.9	K1058
Cell shape-determining protein MREB	*Spiroplasma eriocheiris*	7e1g	17.1	2.7	K1058
Glucokinase	*Streptomyces griseus*	3vgl	16.8	3.5	K1058

Medical Microbiology and Immunology (2020) 209:325–333
https://doi.org/10.1007/s00430-019-00645-2

REVIEW

Human adenovirus binding to host cell receptors: a structural view

Aleksandra Cecylia Stasiak[1] · **Thilo Stehle**[1,2]

Received: 16 August 2019 / Accepted: 14 November 2019 / Published online: 29 November 2019

Abstract
Human Adenoviruses (HAdVs) are a family of clinically and therapeutically relevant viruses. A precise understanding of their host cell attachment and entry mechanisms can be applied in inhibitor design and the construction of targeted gene delivery vectors. In this article, structural data on adenovirus attachment and entry are reviewed. HAdVs engage two types of receptors: first, an attachment receptor that is bound by the fibre knob protein protruding from the icosahedral capsid, and next, an integrin entry receptor bound by the pentameric penton base at the capsid vertices. Adenoviruses use remarkably diverse attachment receptors, five of which have been studied structurally in the context of HAdV binding: Coxsackie and Adenovirus Receptor, CD46, the glycans GD1a and polysialic acid, and desmoglein-2. Together with the integrin entry receptors, they display both symmetrical and asymmetrical modes of binding to the virus as demonstrated by the structural analyses reviewed here. The diversity of HAdV receptors contributes to the broad tropism of these viruses, and structural studies are thus an important source of information on HAdV-host cell interactions. The imbalance in structural data between the more and less extensively studied receptors remains to be addressed by future research.

Keywords Human adenovirus · Structural biology · X-ray crystallography · Cryoelectron microscopy · Review

Introduction

Human adenoviruses as pathogens and therapeutics

Human Adenoviruses (HAdVs) are a family of non-enveloped double-stranded deoxyribonucleic acid (dsDNA) viruses with genomes of about 35 kilobases (kb) [1]. They are causative agents of a wide range of illnesses, such as conjunctivitis, gastroenteritis and respiratory infections [2]. Over 100 types of HAdVs, classified into seven groups (A–G), have been reported to the HAdV Working Group (http://hadvwg.gmu.edu/). Viruses among these groups vary in pathology and molecular characteristics, for instance receptor specificity and host cell tropism [3, 4].

Due to their large genome capacity and the ability to infect different cell types, HAdVs have been extensively studied as gene delivery vectors, and HAdV family members have been used in hundreds of clinical trials in oncology, gene therapy, and vaccinology [5, 6]. A detailed understanding of HAdV host cell attachment and entry mechanisms facilitates the design of vectors with a particular tropism, as well as the rational design of antiviral compounds. Structural biology techniques, in particular, can provide information on the molecular details of virus binding to host cell, and can identify strategies for the disruption of such interactions. So far, structural analyses of HAdV particles and their components have revealed symmetrical and asymmetrical binding modes to HAdV receptors, and these will be reviewed below.

HAdV attachment and entry

The HAdV capsid possesses $T=25$ icosahedral symmetry, and consists of three major proteins: the hexon, the penton base and the fibre (see Fig. 1) [2, 7]. Both the fibre and

Edited by Volkhard A. J. Kempf.

This article is published as part of the Special Issue on ViBrANT-ITN, funded by the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant Agreement No. 765042.

✉ Thilo Stehle
thilo.stehle@uni-tuebingen.de

[1] Interfaculty Institute of Biochemistry, University of Tübingen, Hoppe-Seyler-Str. 4, 72076 Tübingen, Germany

[2] Department of Pediatrics, Vanderbilt University School of Medicine, Nashville, TN 37232, USA

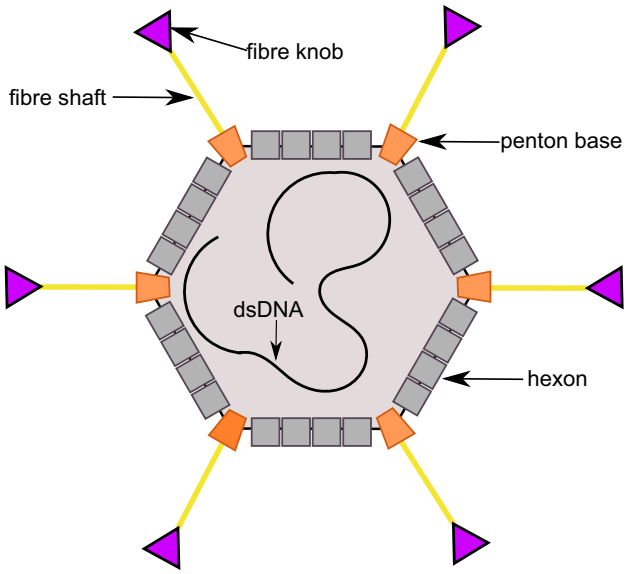


Fig. 1 A schematic drawing of Human Adenovirus, showing the major capsid proteins that interact with receptors. The fibre knob is in purple, the fibre shaft in yellow, the penton base in orange and the hexon in grey. The minor capsid proteins and non-structural proteins have been omitted for simplicity

penton base, forming the penton complex at the vertices of the icosahedron, can engage host cell receptors. The fibre can be divided into a C-terminal, globular structure, the "knob", which protrudes away from the capsid and mediates the initial interaction with attachment receptors, and an N-terminal, elongated "shaft" that anchors the fibre into the viral capsid. Binding of the knob to attachment receptors is followed by the penton base binding to the integrin entry receptor.

The fibre is a homotrimeric structure of between 60–80 kDa per monomer and its globular knob can engage different receptors [8–12]. The monomeric C-terminal fibre knob contains an eight-stranded antiparallel β-sandwich fold, comprised of two β-sheets, with multiple loops. The trimeric propeller-like knob is formed by intertwining β-sheets and has a deep depression at the centre, which thins into narrowing channels, exposing around 65% residues to the solvent [8, 13]. The knob is mounted on the shaft, which consists of repeats of a hydrophobic 15-residue sequence assembled into a trimeric β-spiral [14]. The number of repeats, and hence the length of the shaft, varies between HAdV types. Residues at the very N-terminal region of the fibre knob form a tail that interacts non-covalently with the penton base [15].

Multiple HAdV attachment receptors have been identified: the cell surface proteins Coxsackie and Adenovirus Receptor (CAR) [16, 17], CD46 [10], and desmoglein-2 (DSG-2) [9], as well as the glycans GD1a [11] and polysialic acid [18]. The interactions of all five receptors with HAdV have been established using structural biology techniques and are the focus of this review. Other adenovirus receptors, such as heparan sulphate glycosaminoglycans [4], or factors IX [19] and X [20], have also been described, but we currently lack detailed structural information about their modes of binding to the virus.

Once HAdV has attached to the cell surface, the fibre starts disassociating from the capsid, exposing the penton base [21]. The penton base forms the vertex pentamer (see Fig. 1), which binds to the integrin entry receptor and exploits integrin-mediated signalling to enter the cell by endocytosis. HAdVs have been shown to use multiple types of integrins as their receptors, again highlighting their broad tropism [22].

The penton monomer consists of two domains: a jelly-roll domain proximal to the virion centre, and a distal insertion domain [23]. The latter contains the variable, highly mobile RGD loop, so named because it contains an arginine-glycine-aspartic acid tripeptide sequence motif. The integrin-binding RGD motif mediates binding between the integrin and the penton base, with the exception of group F HAdV-40 and -41, where the interaction is presumed to take place in another manner due to the lack of this motif [24]. The length of the RGD loop varies significantly between strains, ranging from 36 amino acids for HAdV-12 to 99 amino acids for HAdV-5.

HAdV Interactions with attachment receptors

Coxsackie and adenovirus receptor (CAR)

CAR and CD46 have been the most extensively studied HAdV receptors, and their interactions with the fibre knob are particularly well understood. The first structure of a fibre knob bound to CAR was published in 1999, showing that the N-terminal domain of CAR engages the fibre knob with an unusually discontinuous interface that features large, solvent-filled areas [8]. Structural information about fibre knob binding to CD46 became available in 2007, revealing a binding interface that spans the N-terminal two CD46 domains and that likely alters the conformation of unliganded CD46 [25]. The interactions of the fibre knobs with CAR and CD46 have been reviewed, for example, in [26]. Given the lack of extensive developments since then, this review will only give a brief summary.

CAR is a transmembrane protein mediating cell–cell adhesion [27]. It is a member of the Junctional Adhesion Molecule (JAM) family, and present in tight junctions and on epithelial cells at their lateral surfaces. CAR can form homodimers [28] or heterodimers with numerous extracellular and intracellular proteins, for instance fibronectin and

JAM-L [29]. These interactions have been linked to processes such as T cell activation and cell adhesion, among others. CAR is a high-affinity receptor for HAdV groups A and C–F, as well as Coxsackievirus group B viruses [27].

CAR comprises two immunoglobulin-like extracellular domains, CAR-D1 and CAR-D2, both of which have been characterised structurally [28, 30]. The D1 domain forms direct contacts with the fibre knob. The crystal structure of D1 in complex with the knob from HAdV-12 shows one D1 domain bound per knob monomer, with binding occurring at the interface of two monomers rather than at the central cavity (see Fig. 2) [8]. Interactions are mediated by four flexible loops of the knob, as was shown by both structural and mutational studies [31–33]. The AB loop of the fibre knob is responsible for over a half of the protein–protein contacts and is, therefore, an important determinant of CAR receptor specificity [8].

In a recent publication, the structures of the fibre knobs of HAdV-26 and 48 (group D) were solved, and computer modelling, surface plasmon resonance (SPR) and inhibition assays were used to examine binding to potential receptors, primarily CAR, CD46 and DSG-2 [34]. The structure of the fibre knob bound to CAR and CD46 was simulated in silico using homology models. A limited flexibility of loops involved in receptor binding in CAR and CD46 was postulated based on the X-ray data and interaction calculations. Binding of a lower affinity to CAR and no significant binding to CD46 and DSG-2 were proposed based on loop conformation and this was supported by competition inhibition assays.

CD46 (human membrane cofactor protein)

Human CD46, or membrane cofactor protein (MCP), is a ubiquitously expressed transmembrane glycoprotein involved in multiple processes, such as complement response inhibition, fertilisation, and adaptive immunity regulation through its extracellular domains [35, 36]. CD46 is a member of the RCA (regulators of complement activation) family, and as such its role in binding complement proteins C3b and C4b is of particular importance since this interaction prevents the activation of complement response against autologous cells [36]. Complement system proteins are also pathogen receptors. CD46 can serve as a receptor for pathogenic bacteria and viruses, including *Neisseria*, measles, human herpesvirus-6, and some group B HAdVs [36]. RCA family members CD55 and CD21 serve as receptors for echoviruses [37] and Epstein-Barr virus [38], respectively.

The extracellular part of CD46 consists of four short consensus repeats (SCR, 1–4) and a serine, threonine and proline (STP)-rich region [39]. In an unbound form, SCR4 (proximal to cell membrane) is bent in relation to the almost-linearly arranged remaining subunits, forming a "hockey stick" shape. SCR1 and 2 have been identified as the domains interacting with the HAdV knob. When only these two subunits are expressed, a greater degree of flexibility is

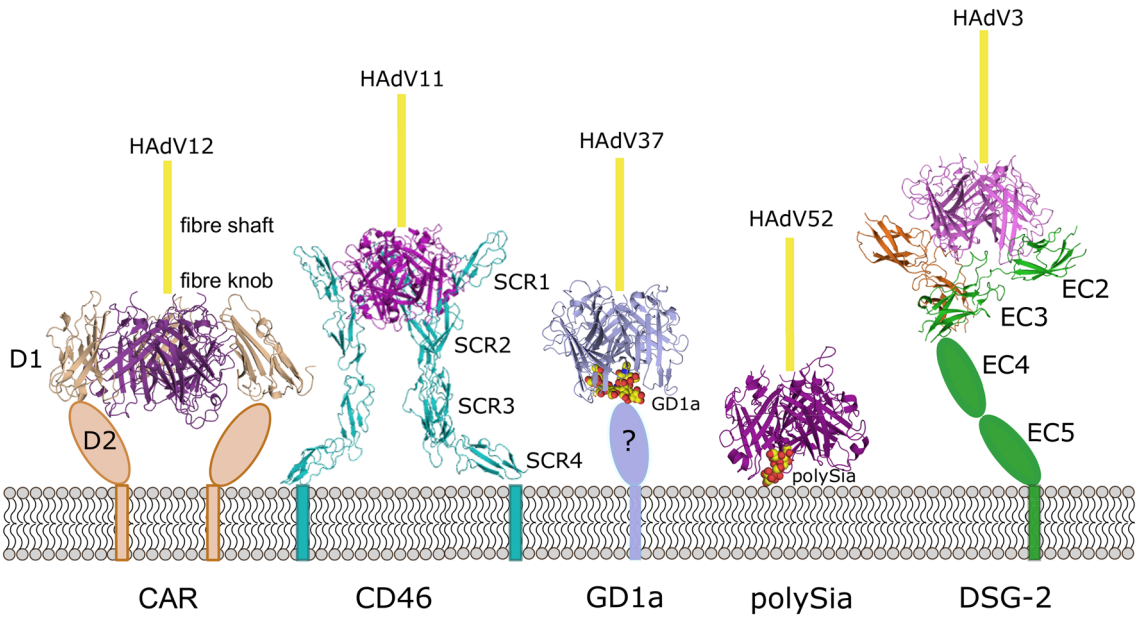


Fig. 2 Human adenovirus knob binding to attachment receptors. The trimeric knob is shown in shades of purple, the trimeric fibre shaft is represented by a yellow line (length not to scale). Domains not featured in structures are represented with cartoon shapes. CAR-D1 is bound to HAdV-12, PDB ID: 1KAC [8], CD46 to HAdV-11, PDB ID: 3O8E [39], GD1a to HAdV-37, PDB ID: 3N0I, [11], polySia to HAdV-52, PDB ID: 6G47 and two copies of DSG-2 to HAdV-3, PDB ID: 6QNU

observed between the two domains. However, when interacting with the HAdV-11 knob, SCR1 becomes almost colinear with SCR2 [25]. The interface of CD46 and this knob is remarkably long and flat, with the binding mediated not only by loops but also main-chain interactions and π–π stacking of crucial arginine residues, including a conserved Arg280 in the knob. HAdV-11 binds to CD46 with high affinity and comparisons between the structure of its fibre knob and that of the lower affinity group B HAdVs (HAdV 7, 14, 21) have provided insight into mechanisms and determinants of the interaction with CD46. These include absence of an Arg residue next to Arg280 involved in the stacking [40] and a change in position of surface loops which disrupt the receptor-virus interface [41]. In all fibre knob-CD46 complexes, the binding is symmetrical and involves one CD46 per one knob monomer. All CD46 molecules make identical contacts with the three fiber knob monomers (see Fig. 2).

Glycans: GD1a and polysialic acid

Glycans, or carbohydrates, constitute an essential class of molecules that is known to serve a wide range of physiological functions [42]. Glycans can be conjugated to other, non-saccharide moieties, such as proteins or lipids, and their diversity in structure, as well as chemical modifications, enables them to mediate functions in signalling, adhesion, and developmental processes. Their prominent display on many cell surfaces makes them a target for numerous pathogens [42].

The first structural data showing interactions of HAdVs with glycans identified three sialic-acid binding sites at the top of the fibre knobs of both HAdV-37 and HAdV-19p (prototypical), around the central cavity [43]. Sialic acid is a monosaccharide derived from neuraminic acid, which serves a range of functions, for instance in immune regulation, and is commonly found on cell surfaces [44]. Sialic acid-containing molecules are employed as receptors by numerous viruses, including influenza viruses, rotaviruses, polyomaviruses, paramyxoviruses, and HAdVs, primarily from group D [45]. In HAdVs, the sialic acid residues bound were moieties of a sialyl-lactose ligand, and both α(2,3)- and α(2,6)-linked ones could bind the knob. Unlike for CD46 and CAR, the binding sites for sialic acid were located on top of the fibre knob, one per knob monomer [43]. However, identification of the full glycan ligand of HAdV-37 as the GD1a glycan showed that two binding sites could be occupied by a single receptor [11] (see Fig. 2).

GD1a is a ganglioside (a glycosphingolipid) prominent in the mammalian nervous system [46]. The GD1a ganglioside is not the HAdV receptor itself, but rather the GD1a glycan is thought to be attached to a glycoprotein with an up to now unidentified protein conjugate. GD1a is Y-shaped and contains two terminal sialic acids, which engage the fibre knob at the top in a very similar manner for both. The sialic acid carboxylate group forms hydrogen bonds with residue Lys345, which was experimentally shown to be crucial for sialic-acid binding [11]. Residues from two fibre knob monomers are involved in binding one sialic acid moiety: one monomer provides hydrogen bonds and the other contributes van der Waals interactions and water-mediated hydrogen bonds, as well as water-mediated hydrogen bonds to galactose moieties anterior to the sialic acids. These glycan moieties bond to sialic acid pointed upwards, away from the binding site, thus leaving space for longer chains terminating in a sialic acid [11]. Knobs that bind sialic-acid are characterised by a highly positive electrostatic potential, which is thought to complement the negative charges of the receptor. No conformational change is seen on binding [43].

The divalent GD1a receptor engages two out of three possible binding sites in the knob. This asymmetry of binding differentiates the GD1a receptor from CAR and CD46, although it is not a unique feature among the HAdV receptors (see below).

Another HAdV type which uses a glycan as its receptor is HAdV-52, a member of group G, which has two different types of fibres: a short fibre and a long fibre, that are present in a 1:1 ratio on the virion [18]. The knob of the long fibre binds to CAR, while the knob of the short fibre binds polysialic acid, and this latter compound is the major attachment receptor for this virus type. The first crystal structure of short fibre knob in complex with a glycan bound 2-*O*-methyl-sialic acid and showed that the binding site is distinct from the one seen in HAdV-37. It is located where the EG and GH loops from two monomers come into contact, and there are again three identical binding sites per knob. The interactions between the knob and the receptor are hydrogen bond-mediated both by the backbone and side chains of the short fibre knob. The conserved RGN motif on the GH loop is of importance for this, although it seems exclusive to HAdV-52 [47]. The receptor was identified as α(2,8)-linked polysialic acid, and the structure obtained by X-ray crystallography showed the importance of transient electrostatic interactions with long glycan chains for the binding [18].

Polysialic acid (polySia) has been linked to many developmental functions, particularly in the human nervous system [48]. While only the non-reducing end was shown to interact with the knob in a stable manner, less stable and less directed electrostatic interactions were shown to also be of consequence to its binding [18]. A positively-charged border around the binding site and the binding site itself mediates contact with polyanionic sialic acid residues further away from the non-reducing end. Due to the transient nature of these interactions, they were not clearly visible in the structure, but simulations suggested that beyond a fifth sialic acid residue these interactions no longer increase. The

ability to bind sialic acid chains of different lengths may also assist in infecting different cell types and hosts. An arginine residue was also shown to be crucial to polySia binding by maintaining the appropriate charge, and three polySia binding was symmetrical.

In addition to the two binding sites described above, it is also worth noting that a third sialic acid binding site has been identified in canine adenovirus 2 [49], again highlighting the diversity and adaptability of the adenovirus fibre knob to variable receptors.

Desmoglein-2

DSG-2 is a transmembrane glycoprotein that belongs to the cadherin family and is involved in maintaining cell–cell adhesion in structures termed desmosomes, which are particularly important in tissues undergoing significant mechanical stress (e.g. heart muscle) [50]. In these junctions, adhesion is mediated by DSG-2 forming heterodimers with a related protein, desmocollin [51]. DSG-2 overexpression has been observed in a number of cancers, and the protein has been identified as a receptor for some group B HAdVs, that have been classified as group B-2 and include HAdV 3, 7, 11 and 14 [9].

The structure of the extracellular fragment of DSG-2 shows that the protein comprises five extracellular cadherin domains (EC1-5) in addition to a transmembrane segment and an intracellular domain [51]. These five domains (with EC1 distal to the cell membrane, and EC5 the most proximal) are linked linearly by Ca2+ ion-binding regions, and they form a curved shape due to a 100° bend between EC3 and EC4. One side of the protein is extensively glycosylated. In the crystal structure, DSG-2 forms homodimers (unlike the desmocollin-containing heterodimers favoured in vivo), linked by strand-swapping mechanisms [51].

A recently published cryo-electron microscopy (cryoEM) structure of the fibre knob-binding domains (EC2 and EC3) of DSG-2 in complex with the HAdV-3 fibre knob reveals that this interaction is also an example of a non-symmetrical receptor-fibre knob interaction [52]. In fact, two distinct stoichiometries of binding are observed: a 1:1, and a 2:1 receptor to fibre knob ratio (see Fig. 2). Importantly, no 3:1 binding is seen, indicating that the trimeric knob cannot engage three DSG-2 molecules at the same time, perhaps because of steric hindrance.

DSG-2 binds at the centre-top of the fibre knob, and in the 1:1 complex it interacts with two out of three monomers. The interactions are primarily mediated by knob loops, with one DSG-2 monomer contacting one knob monomer. The EC2 domain has a stabilising function, while EC3 participates in most of the interactions. Like for CD46, a positional shift is observed in the DSG-2 fragment on fibre knob binding: EC2 rotates by about 10°, once the first, essential interaction with EC3 has been established.

In the 2:1 complex the binding of the second DSG-2 molecule is mediated by very similar contacts, although the limited resolution (3.5 and 3.8 Å) makes it challenging to identify them and their nature clearly. An Asp261 fibre knob residue which is presumed to stabilise loop conformation was shown to be essential to receptor binding. However, it is doubtful that such 2:1 binding mode would be possible in vivo, among the desmocollin-DSG-2 heterodimers in cell–cell junctions [52].

The structural data on DSG-2 binding to HAdV show that the knob is likely too small, and the binding sites for DSG-2 are too close to each other, to allow for simultaneous binding of three DSG-2 molecules to the same knob. The observed asymmetrical, stoichiometrically non-uniform binding of DSG-2 to the fibre knob is of particular interest to studies of integrin-penton base binding (see below), which is also prone to forming asymmetric complexes of different stoichiometries. This decreases the chances of successfully obtaining crystals for X-ray studies, and increases the challenges of cryo-EM data processing as different forms of the complex need to be processed separately.

HAdV interactions with entry receptors—proteins of the integrin family

Integrins are a family of heterodimeric transmembrane proteins involved in signalling between the cell and its environment. They influence processes such as growth, development, immunity, and have been implicated in cancer [53]. Integrins consist of two extracellular subunits, the α and β chains, which each have single transmembrane domains and short cytoplasmic tails [54]. Eighteen different α and eight different β subunits have been identified in vertebrates, combining into 24 proteins. These subunits are composed of a number of smaller domains, which enable a remarkable degree of conformational flexibility of the protein: from a bent to an extended open conformation (see Fig. 3a). This shift is associated with activation of signalling [53].

Due to their ubiquity, integrins are exploited by a number of viruses as receptors, including the foot-and-mouth disease virus, members of the herpes virus family, reoviruses and the human papillomavirus-16 [55]. Adenoviruses also employ integrins as entry receptors. They engage integrins via the penton protein, and a penton-integrin interaction following attachment seems to be uniform among HAdVs. Interactions with integrins stimulate integrin clustering signalling for endocytosis, and virus entry into the host cell via the endocytic pathway is thought to be promoted through such receptor clustering. The integrin most extensively studied in the context of HAdV entry is $\alpha_v\beta_3$ [22], although other

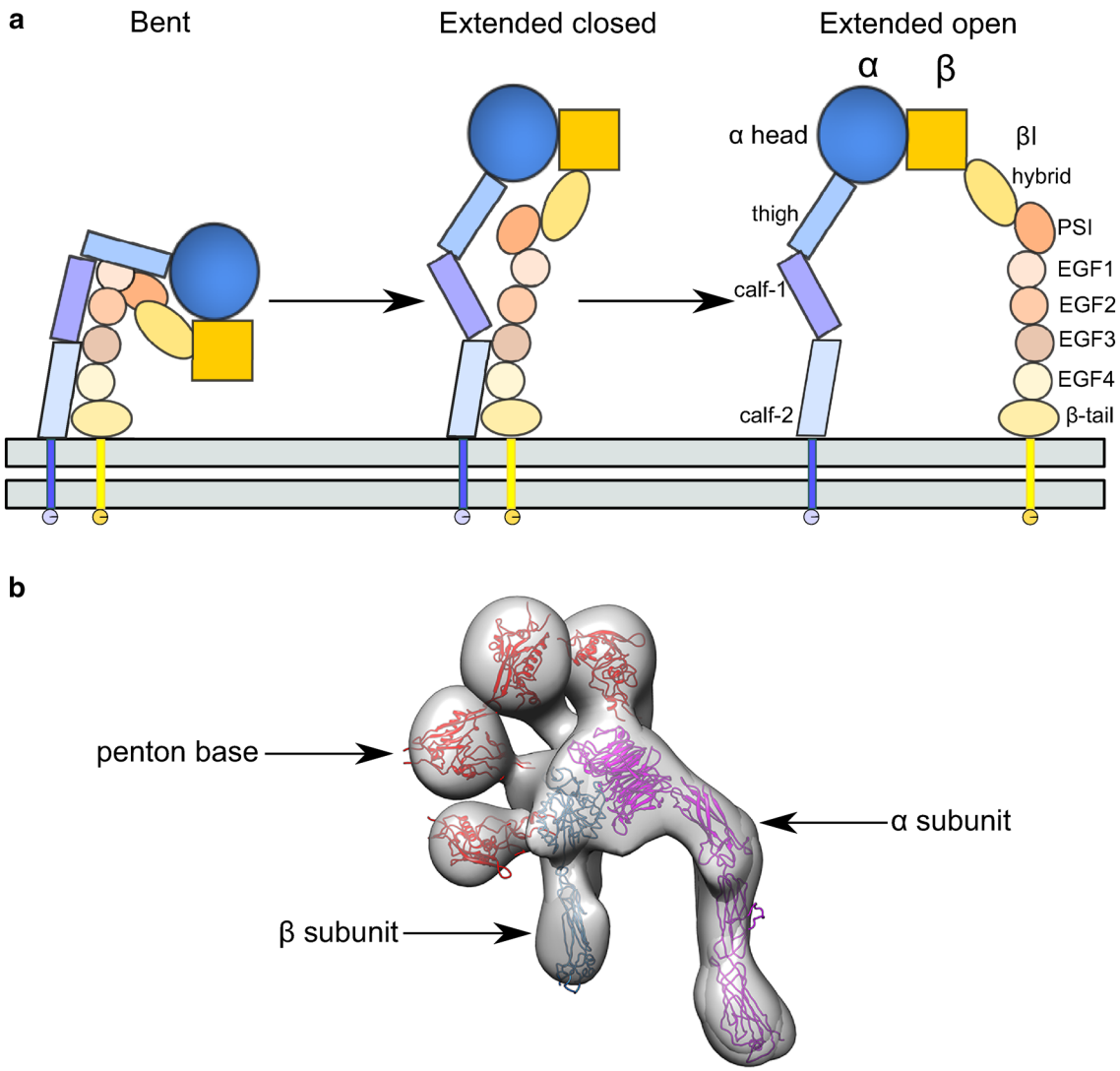


Fig. 3 **a** A cartoon representation of integrin conformational changes. **b** The different binding modes for HAdV-9 monomeric penton base to integrin $\alpha_v\beta_3$. The penton base is in red, the integrin α subunit is in purple, and the β subunit in blue. EMDB IDs: 5955–5973 [58]

integrins, such as $\alpha_v\beta_5$, $\alpha_v\beta_1$ and $\alpha_3\beta_1$ [22, 56] can also function in the same manner.

The first crystal structure of the integrin $\alpha_v\beta_3$ ectodomain showed a bent conformation, with a globular "head" where the N-terminals of the two subunits meet, then separating into two more loosely associated "legs" [57]. The α_v head consists of a seven-bladed β-propeller fold, while a metal ion-dependent adhesion site (MIDAS) of the β subunit mediates the binding between two legs at the "knee", contributing to the conformational change between bent and open conformation via shifts in the position of loops. The RGD motif-binding site is located in the head of the β subunit.

Integrins bind to the RGD motif on the flexible HAdV penton base loop (with the exception of HAdV-40 and 41, which lack the RGD motif and interact with integrins via an unknown mechanism [24]). While the pentameric penton base consists of five pentons and hence contains five RGD-containing loops, it is highly improbable that five integrins could bind the penton simultaneously due to steric clashes [58]. However, given the importance of integrin clustering for HAdV entry signalling, binding of multiple integrins is expected, probably with variable stoichiometries in solution. Given this asymmetry and heterogeneity, it is unlikely that high-quality crystals of a single type of HadV-integrin complex can be obtained, and electron microscopy is, therefore, the method of choice, as one can distinguish different complexes and process them separately. To date, two studies of whole virus complexes with different integrin receptors [59, 60], and one of monomeric penton base in complex with the integrin [58] have been published.

The imposition of icosahedral symmetry and lack of high-resolution structures of the integrin and penton base to dock

into the density at that time meant that it was initially reasoned that a full occupancy of five integrins would be present around the penton. This had been based on structures of HAdV-12 (group A) and HAdV-2 (group C) in complex with integrin $\alpha_v\beta_5$ at a resolution of about 20 Å, based on a ring of density above the pentamer and SPR measurements showing binding of 4.2 integrins per HAdV-2 pentamer [59]. However, after high-resolution integrin and penton base structures had been published it was recognised that at most four integrins, seen in an extended conformation and in different orientations, could fit around the HAdV-12 penton without steric clashes. The more flexible and extendable RGD loop of HAdV-2 penton was speculated to potentially accommodate five integrins [60].

It was also suggested that the fourth integrin would require a conformational shift in the penton base against the inter-domain "twist", and that this change may be responsible for the penton's disassociation from the capsid during HAdV entry and release of the fibre knob [60]. The different integrin orientations and varying occupancies contribute to asymmetry, making the complex structure more challenging to solve with the imposition of symmetry.

Single-particle reconstruction of negative-stain EM monomeric HAdV-9 (group D) penton base insertion domain only integrin $\alpha_v\beta_3$ showed that a variety of integrin conformations could bind the penton, from bent to extended [58]. Moreover, the monomeric penton was shown to bind the integrin in a number of different locations at the integrin head, showing the interactions to be more varied than expected (see Fig. 3b). It is unclear as yet if this variability is present in the more sterically constrained conditions of virus binding to host cell, and if the integrin conformational flexibility is restrained by the presence of transmembrane and intracellular domains.

Outlook

HAdVs are remarkable in the diversity of attachment receptors they employ, which is a contributing factor to their broad cellular tropism. Studies of their receptor-virus interactions using structural biology techniques have significantly advanced our understanding of HAdV attachment and entry. This knowledge can be used both in drug design, to combat the virus as a pathogen, and in rational gene vector development, to enhance the virus as a therapeutic agent. Prospects for the field include addressing the imbalance in data on particular receptors: while some of the attachment receptors have been extensively studied in their viral context (e.g. CAR, CD46), there is much less information available on others. Moreover, the integrin entry receptors, representing the second step of HAdV infection, remain to be examined more closely, as we lack a detailed view of the contacts that are being formed in this complex, and we also lack an understanding of how the geometry of binding dictates the stoichiometry of the interaction.

Acknowledgements ACS is funded by Viral and Bacterial Adhesin Network Training Innovative Training Network, part of the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant Agreement No. 765042. TS acknowledges support from the Baden-Württemberg Foundation (Glycobiology/Glycomics program).

References

1. Ghebremedhin B (2014) Human adenovirus: viral pathogen with increasing importance. Eur J Microbiol Immunol 4:26–33. https://doi.org/10.1556/EuJMI.4.2014.1.2
2. Flint SJ, Enquist LW, Racaniello VR, Skalka AM (2004) Adenoviruses. Principles of virology, 2nd edn. ASM Press, Washington, DC, pp 804–805, 846
3. Kremer EJ, Nemerow GR (2015) Adenovirus tales: from the cell surface to the nuclear pore complex. PLoS Pathog 11:1–8. https://doi.org/10.1371/journal.ppat.1004821
4. Arnberg N (2012) Adenovirus receptors: implications for targeting of viral vectors. Trends Pharmacol Sci 33:442–448. https://doi.org/10.1016/j.tips.2012.04.005
5. Baker A, Aguirre-Hernández C, Halldén G, Parker A (2018) Designer oncolytic adenovirus: coming of Age. Cancers (Basel) 10:201. https://doi.org/10.3390/cancers10060201
6. Dong W, Van Ginkel JWH, Au KY et al (2014) ORCA-010, a novel potency-enhanced oncolytic adenovirus, exerts strong antitumor activity in preclinical models. Hum Gene Ther 25:897–904. https://doi.org/10.1089/hum.2013.229
7. Nemerow GR, Stewart PL, Reddy VS (2012) Structure of human adenovirus. Curr Opin Virol 2:115–121. https://doi.org/10.1016/j.coviro.2011.12.008
8. Bewley MC (1999) Structural analysis of the mechanism of adenovirus binding to its human cellular receptor, CAR. Science 286:1579–1583. https://doi.org/10.1126/science.286.5444.1579
9. Wang H, Li ZY, Liu Y et al (2011) Desmoglein 2 is a receptor for adenovirus serotypes 3, 7, 11 and 14. Nat Med 17:96–104. https://doi.org/10.1038/nm.2270
10. Marttila M, Persson D, Gustafsson D et al (2005) CD46 is a cellular receptor for all species b adenoviruses except types 3 and 7. J Virol 79:14429–14436. https://doi.org/10.1128/jvi.79.22.14429-14436.2005
11. Nilsson EC, Storm RJ, Bauer J et al (2011) The GD1a glycan is a cellular receptor for adenoviruses causing epidemic keratoconjunctivitis. Nat Med 17:105–109. https://doi.org/10.1038/nm.2267
12. Henry LJ, Xia D, Wilke ME et al (1994) Characterization of the knob domain of the adenovirus type 5 fiber protein expressed in Escherichia coli. J Virol 68:5239–5246
13. 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. https://doi.org/10.1016/s0969-2126(94)00126-x
14. van Raaij MJ, Mitraki A, Lavigne G, Cusack S (1999) A triple β-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. Nature 401:935–938. https://doi.org/10.1038/44880
15. Cao C, Dong X, Wu X et al (2012) Conserved fiber-penton base interaction revealed by nearly atomic resolution cryo-electron microscopy of the structure of adenovirus provides insight into receptor interaction. J Virol 86:12322–12329. https://doi.org/10.1128/jvi.01608-12
16. Bergelson JM, Droguett G, Kurt-jones EA et al (1997) Isolation of a common receptor for coxsackie B viruses and adenoviruses 2 and 5. Science 275:1320–1324. https://doi.org/10.1126/science.275.5304.1320
17. 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 USA 94:3352–3356. https://doi.org/10.1073/pnas.94.7.3352
18. Lenman A, Liaci AM, Liu Y et al (2018) Polysialic acid is a cellular receptor for human adenovirus 52. Proc Natl Acad Sci 115:E4264–E4273. https://doi.org/10.1073/pnas.1716900115
19. Lenman A, Muller S, Nygren MI et al (2011) Coagulation Factor IX mediates serotype-specific binding of species a adenoviruses to host cells. J Virol 85:13420–13431. https://doi.org/10.1128/jvi.06088-11
20. Jonsson MI, Lenman AE, Frangsmyr L et al (2009) Coagulation Factors IX and X enhance binding and infection of adenovirus types 5 and 31 in human epithelial cells. J Virol 83:3816–3825. https://doi.org/10.1128/jvi.02562-08
21. Nemerow GR, Stewart PL (2016) Insights into adenovirus uncoating from interactions with integrins and mediators of host immunity. Viruses 8:1–15. https://doi.org/10.3390/v8120337
22. Wickham TJ, Mathias P, Cheresh DA, Nemerow GR (1993) Integrins αvβ3and αvβ5promote adenovirus internalization but not virus attachment. Cell 73:309–319. https://doi.org/10.1016/0092-8674(93)90231-E
23. Zubieta C, Schoehn G, Chroboczek J, Cusack S (2005) The structure of the human adenovirus 2 penton. Mol Cell 17:121–135. https://doi.org/10.1016/j.molcel.2004.11.041
24. Rajan A, Persson BD, Frängsmyr L et al (2018) Enteric species f human adenoviruses use laminin-binding integrins as Co-receptors for infection of Ht-29 Cells. Sci Rep 8:1–14. https://doi.org/10.1038/s41598-018-28255-7
25. Persson BD, Reiter DM, Marttila M et al (2007) Adenovirus type 11 binding alters the conformation of its receptor CD46. Nat Struct Mol Biol 14:164–166. https://doi.org/10.1038/nsmb1190
26. Cupelli K, Stehle T (2011) Viral attachment strategies: the many faces of adenoviruses. Curr Opin Virol 1:84–91. https://doi.org/10.1016/j.coviro.2011.05.024
27. Ortiz-Zapater E, Santis G, Parsons M (2017) CAR: a key regulator of adhesion and inflammation. Int J Biochem Cell Biol 89:1–5. https://doi.org/10.1016/j.biocel.2017.05.025
28. van Raaij MJ, Chouin E, van der Zandt H et al (2000) Dimeric structure of the coxsackievirus and adenovirus receptor D1 domain at 1.7 Å resolution. Structure 8:1147–1155. https://doi.org/10.1016/S0969-2126(00)00528-1
29. Matthäus C, Langhorst H, Schütz L et al (2017) Cell-cell communication mediated by the CAR subgroup of immunoglobulin cell adhesion molecules in health and disease. Mol Cell Neurosci 81:32–40. https://doi.org/10.1016/j.mcn.2016.11.009
30. Jiang S, Jacobs A, Laue TM, Caffrey M (2004) Solution structure of the coxsackievirus and adenovirus receptor domain 2. Biochemistry 43:1847–1853. https://doi.org/10.1021/bi035490x
31. Kirby I, Davison E, Beavil AJ et al (1999) Mutations in the DG loop of adenovirus type 5 fiber knob protein abolish high-affinity binding to its cellular receptor CAR. J Virol 73:9508–9514
32. Kirby I, Davison E, Beavil AJ et al (2000) Identification of contact residues and definition of the CAR-binding site of adenovirus Type 5 fiber protein. J Virol 74:2804–2813. https://doi.org/10.1128/jvi.74.6.2804-2813.2000
33. Jakubczak JL, Rollence ML, Stewart DA et al (2001) Adenovirus Type 5 viral particles pseudotyped with mutagenized fiber proteins show diminished infectivity of coxsackie B-adenovirus receptor-bearing cells. J Virol 75:2972–2981. https://doi.org/10.1128/jvi.75.6.2972-2981.2001
34. Baker AT, Greenshields-Watson A, Coughlan L et al (2019) Diversity within the adenovirus fiber knob hypervariable loops influences primary receptor interactions. Nat Commun 10:741. https://doi.org/10.1038/s41467-019-08599-y
35. Riley-Vargas RC, Gill DB, Kemper C et al (2004) CD46: expanding beyond complement regulation. Trends Immunol 25:496–503. https://doi.org/10.1016/j.it.2004.07.004
36. Cardone J, Le Friec G, Kemper C (2011) CD46 in innate and adaptive immunity: an update. Clin Exp Immunol 164:301–311. https://doi.org/10.1111/j.1365-2249.2011.04400.x
37. Bergelson JM, Chan M, Solomon KR et al (1994) Decay-accelerating factor (CD55), a glycosylphosphatidylinositol-anchored complement regulatory protein, is a receptor for several echoviruses. Proc Natl Acad Sci 91:6245–6248. https://doi.org/10.1073/pnas.91.13.6245
38. Fingeroth JD, Weis JJ, Tedder TF et al (1984) Epstein-Barr virus receptor of human B lymphocytes is the C3d receptor CR2. Proc Natl Acad Sci USA 81:4510–4514
39. Persson BD, Schmitz NB, Santiago C et al (2010) Structure of the extracellular portion of CD46 provides insights into its interactions with complement proteins and pathogens. PLoS Pathog 6:e1001122. https://doi.org/10.1371/journal.ppat.1001122
40. Persson BD, Muller S, Reiter DM 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. https://doi.org/10.1128/JVI.01967-08
41. Cupelli K, Muller S, Persson BD et al (2010) Structure of adenovirus type 21 knob in complex with CD46 reveals key differences in receptor contacts among species B adenoviruses. J Virol 84:3189–3200. https://doi.org/10.1128/jvi.01964-09
42. Varki A (2017) Biological roles of glycans. Glycobiology 27:3–49. https://doi.org/10.1093/glycob/cww086
43. Burmeister WP, Guilligay D, Cusack S et al (2004) Crystal structure of species D adenovirus fiber knobs and their sialic acid binding sites. J Virol 78:7727–7736. https://doi.org/10.1128/jvi.78.14.7727-7736.2004
44. Varki A (2007) Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. Nature 446:1023–1029. https://doi.org/10.1038/nature05816
45. Stencel-Baerenwald JE, Reiss K, Reiter DM et al (2014) The sweet spot: defining virus-sialic acid interactions. Nat Rev Microbiol 12:739–749. https://doi.org/10.1038/nrmicro3346
46. Schnaar RL (2016) Gangliosides of the vertebrate nervous system. J Mol Biol 428:3325–3336. https://doi.org/10.1016/j.jmb.2016.05.020
47. Lenman A, Liaci AM, Liu Y et al (2015) Human adenovirus 52 uses sialic acid-containing glycoproteins and the coxsackie and adenovirus receptor for binding to target cells. PLoS Pathog 11:1–23. https://doi.org/10.1371/journal.ppat.1004657
48. Galuska C, Lütteke T, Galuska S (2017) Is polysialylated NCAM not only a regulator during brain development but also during the formation of other organs? Biology (Basel) 6:27. https://doi.org/10.3390/biology6020027

49. Rademacher C, Bru T, McBride R et al (2012) A Siglec-like sialic-acid-binding motif revealed in an adenovirus capsid protein. Glycobiology 22:1086–1091. https://doi.org/10.1093/glycob/cws073
50. Dusek RL, Godsel LM, Green KJ (2007) Discriminating roles of desmosomal cadherins: beyond desmosomal adhesion. J Dermatol Sci 45:7–21. https://doi.org/10.1016/j.jdermsci.2006.10.006
51. Harrison OJ, Brasch J, Lasso G et al (2016) Structural basis of adhesive binding by desmocollins and desmogleins. Proc Natl Acad Sci 113:7160–7165. https://doi.org/10.1073/pnas.1606272113
52. Vassal-Stermann E, Effantin G, Zubieta C et al (2019) CryoEM structure of adenovirus type 3 fibre with desmoglein 2 shows an unusual mode of receptor engagement. Nat Commun 10:1181. https://doi.org/10.1038/s41467-019-09220-y
53. Campbell ID, Humphries MJ (2011) Integrin structure, activation, and interactions. Cold Spring Harb Perspect Biol 3:a004994. https://doi.org/10.1101/cshperspect.a004994
54. Luo B-H, Carman CV, Springer TA (2007) Structural basis of integrin regulation and signaling. Annu Rev Immunol 25:619–647. https://doi.org/10.1146/annurev.immunol.25.022106.141618
55. Hussein HAM, Walker LR, Abdel-Raouf UM et al (2015) Beyond RGD: virus interactions with integrins. Arch Virol 160:2669–2681. https://doi.org/10.1007/s00705-015-2579-8
56. Storm RJ, Persson BD, Skalman LN et al (2017) Human Adenovirus Type 37 Uses αVβ1 and α3β1 integrins for infection of human corneal cells. J Virol 91:2019–2020. https://doi.org/10.1128/JVI.02019-16
57. Xiong J, Stehle T, Diefenbach B et al (2001) Crystal structure of the extracellular segment of integrin αVβ3. Science 294:339–345. https://doi.org/10.1126/science.1064535.Crystal
58. Veesler D, Cupelli K, Burger M et al (2014) Single-particle EM reveals plasticity of interactions between the adenovirus penton base and integrin αVβ3. Proc Natl Acad Sci 111:8815–8819. https://doi.org/10.1073/pnas.1404575111
59. Chiu CY, Mathias P, Nemerow GR, Stewart PL (1999) Structure of adenovirus complexed with its internalization receptor, αvβ5 integrin. J Virol 73:6759–6768
60. Lindert S, Silvestry M, Mullen TM et al (2009) Cryo-electron microscopy structure of an adenovirus-integrin complex indicates conformational changes in both penton base and integrin. J Virol 83:11491–11501. https://doi.org/10.1128/jvi.01214-09

viruses

MDPI

Review

Structural Insight into Non-Enveloped Virus Binding to Glycosaminoglycan Receptors: A Review

Marie N. Sorin [1,2,†], Jasmin Kuhn [1,†], Aleksandra C. Stasiak [1,†] and Thilo Stehle [1,3,*]

[1] Interfaculty Institute of Biochemistry, University of Tuebingen, 72076 Tuebingen, Germany; marie.sorin@univ-nantes.fr (M.N.S.); jasmin.kuhn@uni-tuebingen.de (J.K.); aleksandra.stasiak@uni-tuebingen.de (A.C.S.)
[2] Faculté de Médecine, Université de Nantes, Inserm, Centre de Recherche en Transplantation et Immunologie, UMR 1064, ITUN, F-44000 Nantes, France
[3] Department of Pediatrics, Vanderbilt University School of Medicine, Nashville, TN 37232, USA
* Correspondence: thilo.stehle@uni-tuebingen.de
† These authors contributed equally.

Abstract: Viruses are infectious agents that hijack the host cell machinery in order to replicate and generate progeny. Viral infection is initiated by attachment to host cell receptors, and typical viral receptors are cell-surface-borne molecules such as proteins or glycan structures. Sialylated glycans (glycans bearing sialic acids) and glycosaminoglycans (GAGs) represent major classes of carbohydrate receptors and have been implicated in facilitating viral entry for many viruses. As interactions between viruses and sialic acids have been extensively reviewed in the past, this review provides an overview of the current state of structural knowledge about interactions between non-enveloped human viruses and GAGs. We focus here on adeno-associated viruses, human papilloma viruses (HPVs), and polyomaviruses, as at least some structural information about the interactions of these viruses with GAGs is available. We also discuss the multivalent potential for GAG binding, highlighting the importance of charged interactions and positively charged amino acids at the binding sites, and point out challenges that remain in the field.

Keywords: viruses; glycans; glycosaminoglycans; glycovirology; non-enveloped viruses; structural biology

Citation: Sorin, M.N.; Kuhn, J.; Stasiak, A.C.; Stehle, T. Structural Insight into Non-Enveloped Virus Binding to Glycosaminoglycan Receptors: A Review. *Viruses* **2021**, *13*, 800. https://doi.org/10.3390/v13050800

Academic Editor: Jacques Le Pendu

Received: 14 April 2021
Accepted: 26 April 2021
Published: 29 April 2021

1. Introduction

Viruses are infectious entities that require a living organism, a host, to support their replication. They can target any kind of organism, ranging from complex hosts such as animals or plants to microorganisms such as bacteria or archaea. Until now, more than 6500 types of viruses have been discovered, and more than 260 of these can infect humans [1]. Viruses are encoded by DNA or RNA genomes, which are packaged into and protected by a protein capsid. Enveloped viruses additionally contain a lipid bilayer shell that is derived from the host cell membrane, while non-enveloped viruses lack such a membrane envelope and are encased in a capsid protein shell.

In order to infect a host organism and replicate, specific virus proteins interact with receptor molecules on the host cell surface. Among these receptors, carbohydrate-based structures, or glycans, are used by many viruses to infect cells [2]. Such glycans can be grouped into three major groups: structures that terminate in sialic acids (sialylated glycans), neutral structures that are based on histo-blood group antigens (HBGAs), and structures known as glycosaminoglycans (GAGs). These GAGs are long, linear, negatively charged polysaccharides with molecular weights ranging from 10 to 100 kilodaltons (kDa), and they can be divided into two main types: sulfated GAGs, including heparan sulfate (HS), heparin, chondroitin sulfate (CH), and dermatan sulfate (DS); and non-sulfated GAGs, including hyaluronic acid [3].

HS and heparin are both linear polysaccharides composed of repeating units of uronic acid linked to D-glucosamine. HS is ubiquitously found on the cell surface or on mammalian extracellular matrix proteins, where it is involved in multiple biological processes, while heparin is secreted by mast cells and is medically used as an anticoagulant. HS and heparin share the same backbone, which is comprised of uronic acid and D-glucosamine, where the uronic acid moiety can either be α-L-iduronic acid (IdoA) or β-D-glucuronic acid (GlcA). Both components of the polysaccharide can be modified by sulfatation, deacetylation, and epimerization during biosynthesis [4]. There are three main differences between heparin and HS: (i) The uronic acid in heparin is predominantly IdoA, while it is GlcA in HS. (ii) D-glucosamine is mainly N-sulfated in heparin, while it is N-acetylated in HS. (iii) Heparin is composed of at least 70–80% of Ido(2S)-(1→ 4)-GlcNS(6S) disaccharides, while in HS 40–60% of the disaccharide units are GlcA-(1→ 4)-D-glucosamine, either N-acetylated or N-sulfated. Taken together, these differences indicate that heparin carries more sulfatation, and therefore has a higher overall negative charge compared to HS [5]. Chondroitin sulfate is also a linear polysaccharide composed of repeated units of uronic acid and N-acetylgalactosamine (GalNAc). Similar to HS, CS is found on the cell surface and in the extracellular matrix, and is involved in diverse physiological processes [6]. Dermatan sulfate, also known as chondroitin sulfate B, is the main GAG expressed in skin tissue. DS is structurally close to CH because it is composed of GalNAc, but also to HS and heparin because of the presence of IdoA in the disaccharide repeating unit [7].

While several non-enveloped viruses have been found to use GAGs as their receptors (e.g., adenoviruses [8] and rhinoviruses [9]), these results have not necessarily been verified in vivo, and it remains unclear whether the reported receptor specificity is the result of cell culture adaptation or is also present in nature [10]. This review aims to provide an update on the structural work characterizing the interaction of non-enveloped viruses with sulfated glycosaminoglycans, especially heparin. Among the structural studies that will be discussed, only structures with a resolution better than 4 Å will be considered in detail, and primarily only ones where electron density maps for the ligands are publicly available and can be inspected to assess the quality of the ligand fitting.

2. Adeno-Associated Virus

Adeno-associated viruses (AAVs) are single-stranded DNA viruses, members of the family *Parvoviridae*, with a small genome of about 4.5 kilobases. They require the host cell to be infected by another virus, typically an adenovirus, for productive replication, and they are not associated with human pathology. Their significance in research is connected to their potential as gene delivery vectors, associated with, among others, their ability to transduce non-dividing cells and lack of pathogenicity. The AAV capsid is about 26 nm in diameter, composed of 60 subunits, and has T = 1 icosahedral symmetry (see Figure 1a) [11]. The three capsid proteins, VP1, VP2, and VP3, all share most of the C-terminal sequence. Compared with the other two proteins, about ten times more copies of VP3 are found in the capsid. All three proteins consist of a jelly roll β-barrel fold, with loops between the barrel strands containing most of the sequences that differ among AAV strains and that are mainly responsible for receptor binding [11]. The loops form the triple-capsomer peaks close to the threefold symmetry axis (Figure 1b)—a location where multiple residues associated with heparin binding cluster [12–14].

While heparan sulfate was long considered to be the primary receptor for AAVs, recent research suggests that this role is in fact performed by the adeno-associated virus receptor [15], with GAGs serving as lower-affinity co-receptors of significance for attachment. Structural studies of GAG binding to AAV generally have suffered from difficulties in obtaining well-diffracting crystals of the virus–receptor complex, either by co-crystallization or soaking [16]. These difficulties have excluded X-ray crystallography as a method for the structural analysis of AAV–GAG complexes, and led to efforts of obtaining cryoelectron microscopy (cryoEM) structures instead.

Among the first structures of AAV–GAG complexes are two examples that were solved before the cryoEM "resolution revolution" [17], at lower resolutions of 18 Å [18] and 8 Å [19]. Both reports examined the structures of AAV type 2 in complex with heparin, and came to opposite conclusions, with the lower-resolution analysis postulating significant conformational shifts on GAG binding that were not observed in the higher-resolution structure. Additional studies of complexes of AAV-DJ (a mutant strain selected for resistance to most common human neutralizing antibodies, with a high capsid sequence similarity to AAV-2) at 5 Å [16] and 2.8 Å resolution [20] provided data that argue against a large conformational shift. In these cases, heparin analogues were used to mitigate the issues caused by multiple-site, asymmetric binding of the physiological ligand and to provide a clearer view of binding.

While the use of smaller, shorter GAG analogues such as fondaparinux (five sugar moieties) or sucrose octasulfate (SOS) is a promising approach to the problem of averaging out the non-icosahedral ligand density, the resulting structures do not provide high-resolution information on GAG binding to AAV. In the lower-resolution structure at 5 Å, the level of molecular details is low (as expected at this resolution), and while electron density for the ligand is visible, this density is largely globular and does not allow the assignment of specific contacts. Unfortunately, this situation is not improved for the 2.8 Å structure, probably due to low occupancy of the ligand. The viral capsid density allows for a clear delineation of the main chain and most side chains of the amino acids, but the ligand electron density is not detailed enough to reliably place the ligand or to deduce much information about its conformation (see Figure 1c).

The most detailed structural information on GAG binding to AAV is provided not through structural analyses but mutational studies. The residues identified by these studies as necessary (or of importance) for heparin binding can then be mapped onto the capsid or capsomer structures, where they do localize in the proximity of ligand densities from structural studies. Thus, positively charged residues at the threefold peak, primarily arginines (R587 and R590 from one capsomer and R486 and R489 from another—see Figure 1e; residue numbering according to AAV-DJ sequence), have been shown to play an important role in the cell binding of AAV-2, since a loss in infectivity was seen after their mutation [12]. These residues, along with others, form a positively-charged patch coming downwards from the peak (Figure 1d). GAG binding is often mediated by positive charges on the protein, with positively charged residues forming ionic interactions with the sulfur groups, and this is where the GAG density is found in the structures. Structural and mutational studies of AAV-6, a strain which binds both HS and sialic acid, have also highlighted the threefold peak as the location of an HS binding site [13,14]. Given that GAG molecules can achieve significant lengths, it is possible that one chain would engage multiple sites on a single capsid, winding in the clefts between the peaks. This would serve to increase the binding affinity and enhance the strength of the interaction, increasing the chance of entry.

Structural studies of AAV binding to GAGs form an interesting example of the challenges encountered in this field, and of characterizing interactions despite the unavailability of a high-resolution structure of the complex. Combining the high-resolution apostructures of AAV with extensive mutational data, an overview of the binding site locations and their residues has been obtained. A recent cryo-EM structure of AAV-DJ at a resolution below 1.6 Å [21], and a continuing interest in AAV as a gene therapy vector [22], shows this is an active field, both in structural biology and applied medicine; therefore, elucidating the details of AAV–GAG interactions at atomic (or near atomic) resolution would be an important development.

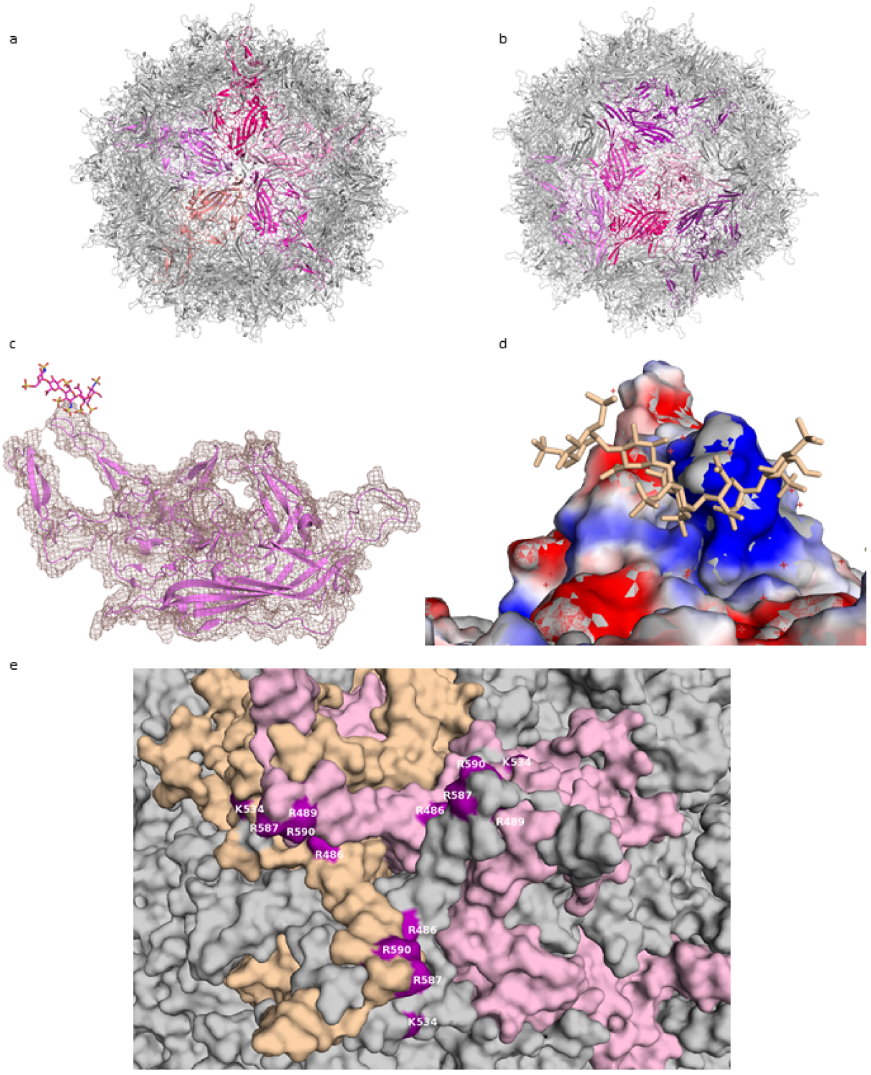


Figure 1. AAV-DJ capsid and its binding by GAG (PDB: 5UF6). (**a**) The viral capsid, with one constituting pentamer highlighted in shades of purple. (**b**) The viral capsid, with one constituting hexamer, and a threefold symmetry axis peak, highlighted in shades of purple. (**c**) A close-up of the density (salmon) for the structure at σ 1.5, showing that the capsid protein (purple) matches it closely and can be well resolved, while the ligand (raspberry) density is not visible at this level. (**d**) The charge distribution in the proximity of the threefold peak. Negatively charged residues are shown in red, positively charged in blue, ligand in wheat. (**e**) Close-up of the residues identified in mutational studies as contributing to binding HS, highlighted in purple. Capsomers in wheat, pink, and grey. Figures were created using PyMOL [23].

3. Human Papillomavirus

Human papillomaviruses (HPVs) are small, non-enveloped viruses containing a double-stranded DNA genome. Over 170 HPV types have been described, of which approximately 40 can be transmitted by sexual contacts, making HPV the most common sexually transmitted infection (STI) [24]. It is estimated that more than 80% of the world's population has at one point been in contact with HPV, although 90% of infections spontaneously resolve within the first two years and are often asymptomatic. However, some strains can cause warts or precancerous lesions (progressing to cancer), mainly of the genital or oropharyngeal tract. Strains HPV6 and HPV11 are the most common cause of genital warts, while 70% of cervical cancers are associated with HPV16 or HPV18 [25,26].

The HPV capsid is comprised of 360 copies of the major capsid protein L1 arranged in 72 capsomers of L1 pentamers, forming an icosahedral T = 7d geometry (Figure 2a,b). The minor capsid protein L2 is incorporated into the viral capsid in a not fully understood manner. The cell entry of HPV is mediated by binding of the viral capsid to cell surface proteoglycans as the primary receptors [27–29].

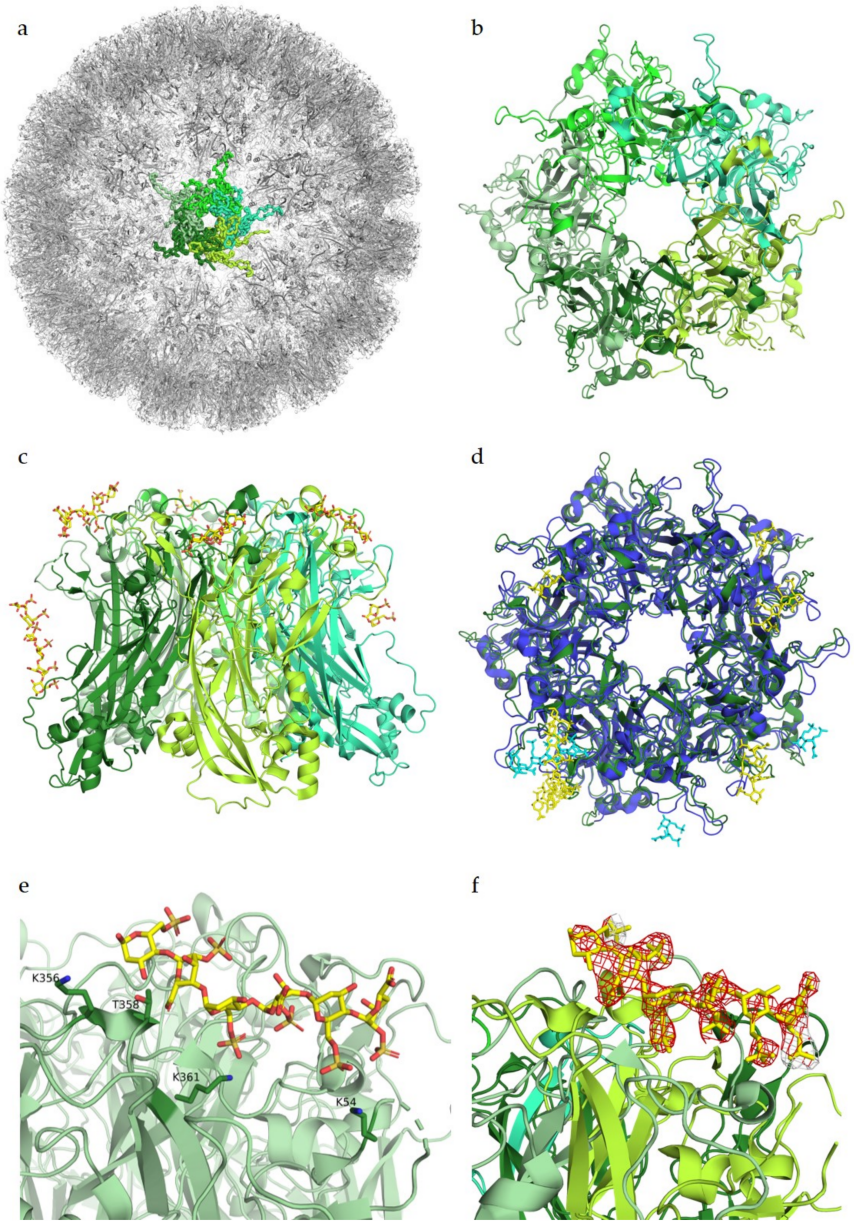


Figure 2. Glycan engagement of HPV16. (**a**) Capsid structure of HPV16 L1 pentamers (PDB ID 5KEP). L1 subunits of one pentamer are displayed in different shades of green. The highlighted pentamer is represented as a ribbon, the remaining capsid is depicted as cartoons. (**b**) Top view of a HPV16 L1 pentamer (PDB ID 5W1O). Subunits are displayed in cartoon representation and colored in different shades of green. (**c**) Side view of the pentamer. The heparin ligand is colored by atom type and represented with sticks in yellow. (**d**) Superposition of HPV16 L1 pentamer in green with its heparin ligand in yellow with HPV18 L1 pentamer (PDB ID 5W1X) in blue and its ligand in cyan. (**e**) Close-up view of one ligand binding site of HPV16. Ligand-binding residues K356, T358, K361, and K54 are represented as sticks. (**f**) Close-up view as in (**e**) of one ligand binding site with the negative F_o–F_c difference map of the ligand contoured at 3σ, colored in red. Figures were created using PyMOL [23].

Heparin binding assays identified largely positively charged peptide sequences at the C-terminus of L1, spanning amino acids G472 to L505, as the main interaction partners for GAGs [30]. Additional interaction studies determined similar residues of L1 as well as residues of L2 to bind to heparin, with binding constants in the high millimolar range [31]. However, both studies were only performed with small peptide fragments of both L1 and L2. Studies by Dasgupta et al. (2011) for the first time revealed structural details of the interactions between HPV and receptor candidates. Using co-crystallization experiments with heparin, these studies showed multiple heparan sulfate binding sites in the intact pentamers of both HPV16 and HPV18, demonstrating that these two strains display somewhat different oligosaccharide binding patterns (Figure 2d). The study identified key regions for ligand binding on surface loops (HI, FG, BC, EF, and a4) of the capsid protein L1, which are conserved between HPV16 and HPV18. These regions were determined to be located on the capsid-distal sides as well as on the capsid-proximal sides of the pentamer (Figure 2c). As more data are available for HPV16, we will mainly focus on HPV16 from here on. The crucial residues interacting with the negatively charged heparin in the HPV16 binding sites were identified as positively charged lysines that engage in charge–charge interactions with the sulfated groups of the receptors. These lysines form a mostly positively charged binding pocket in comparison to the highly negatively charged pore. For the two binding sites on the capsid-distal side of the pentamer, interaction partners were determined to be K54, K356, and K361 (Figure 2e), as well as K278 and N285 for a second binding site. Additionally, polar residues T358 and T266 were found to contact the glycan through polar interactions. To our knowledge, the characterized residues had not been identified before in binding assays as involved for glycan binding. Mutational experiments by Dasgupta et al. have confirmed K278 and K361 as essential residues for HPV16 pseudoviral infection, while mutations of the other structurally identified binding partners did not significantly influence infectivity [32].

The structure of HPV16 in complex with heparin has a resolution of 2.8 Å, meaning that residues and ligands should be clearly visible in the electron density, as previously described for adeno-associated viruses. In the difference map (F_o–F_c map, from PDB ID 5W1O), no negative difference density can be observed for the L1 pentamers; however, for the complex structure, a large amount of negative electron density around the heparin ligand is visible (Figure 2f). This negative density indicates that the modelled ligand structure does not fully agree with the experimentally observed data, perhaps because of high flexibility of the glycan or low occupancy, as both would show density in the initial map but would display negative density after refinement. Nevertheless, it is also possible that the glycan was placed incorrectly in the structure. Structural studies by Guan et al. (2017) identified a different heparan sulfate binding site, which overlaps with two of the previously identified binding sites [33]. However, no complex structure or difference maps are available for the newer study, and therefore cannot be considered for this review.

As already discussed for AAV complex structures, complex structures of HPV with heparin or HS published to date should be treated with caution, as the available electron density maps do not allow for unambiguous assignment of contacts between protein and ligand. It also becomes clear that flexibility and occupancy represent major challenges in structural elucidations of virus–GAG interactions.

4. Polyomavirus

The *Polyomaviridae* form a family of non-enveloped icosahedral dsDNA viruses. To date, fourteen human polyomaviruses have been identified. Infection mostly occurs during childhood, and the viruses typically persist in host cells asymptomatically. In the context of immunosuppression, polyomavirus family members can reactivate and lead to various diseases [34]. The BK (BKPyV) and JC (JCPyV) polyomaviruses were the first polyomaviruses to be discovered, and they induce polyomavirus-associated nephropathy and progressive multifocal leukoencephalopathy in immunocompromised patients, respectively [35,36]. The more recently discovered Merkel cell polyomavirus (MCPyV)

was found in Merkel cell carcinomas, and is the first confirmed human oncovirus within the *Polyomaviridae* family [37]. Similar to papillomaviruses, polyomaviruses have a capsid composed of 72 pentameric capsomers (referred to as VP1 pentamers) with an icosahedral T = 7d symmetry. However, polyomaviruses are smaller, with a diameter of around 50 nanometers compared to 60 nanometers for papillomaviruses. Moreover, the structures of the L1 and VP1 differ significantly. VP1 capsomers have a barrel-shaped morphology while papillomavirus L1 capsomers are mushroom-like protrusions with star-shaped pentamers [38].

When initiating infection, many polyomaviruses use glycans terminating in the sialic acid N-acetyl neuraminic acid (Neu5Ac) as their receptors. Interactions are mediated through surface loops at the outer margin of the major capsid protein VP1, and structural information is available for many complexes of polyomavirus VP1 pentamers with their cognate receptors [39–45]. These analyses have revealed that polyomaviruses differ in their specificities for sialylated glycan structures. For example, BKPyV binds to cells via sialylated structures found in b-series gangliosides, which carry two sialic acids on the inner galactose (Neu5Acα2-8Neu5Acα2-3Gal) (Figure 3b) [41]. On the other hand, JCPyV specifically engages glycans carrying terminal α2,6-linked Neu5Ac in the context of the monosialylated lactoseries tetrasaccharide c (LSTc) glycan [45]. MCPyV binds to yet another type of sialylated glycan carrying a terminal Neu5Acα2-3Gal motif [40].

For some polyomaviruses, additional interactions with non-sialylated glycans have been reported to be required for cell entry. For example, MCPyV requires both GAGs and sialylated glycans for entry, and these interactions occur sequentially, with GAGs serving as the primary receptor [46]. BKPyV and JCPyV have also been proposed to bind GAGs [47]. This recent study showed only a small decrease in the cell attachment of wild-type JCPyV and BKPyV virus-like particles (VLPs) on cells producing non-sialylated or hyposialylated glycans [47]. This result indicates that these two viruses do not require sialylated glycans for initial binding to cells. Further work showed that simultaneous removal of GAGs and sialylated glycans drastically reduced the binding of those VLPs to cells. Taken together, the results of this study suggest that wild-type BK and JC viruses can interact with either GAGs or sialylated glycans for initial cell attachment [47].

In the case of BKPyV, a structural study has investigated interactions of this virus with GAGs, solving the structure of BKPyV virus-like particles via cryo-electron microscopy (cryoEM) at 3.8 Å resolution. As expected, the BKPyV capsid was found to consist of 72 pentamers of the major capsid protein VP1 arranged in T = 7d icosahedral symmetry (Figure 3a). Six distinct VP1 conformations, all sharing the β-sandwich jelly roll fold, were found within the asymmetric unit of the capsid. The C-terminal arms are well resolved, and the structure of the assembled VLP shows how they are extended into adjacent capsomers and maintain a stable capsid structure, similar to what has been observed for other polyomavirus family members [39,48]. A structure of BKPyV VLPs in complex with heparin was also determined at a resolution of 3.6 Å. No structural differences at the protein level were seen between this structure and the unliganded BKPyV structure, establishing that heparin binding does not induce a structural change of the capsid. However, electron density corresponding to heparin was not visible at high resolution, and difference maps calculated to identify potential binding sites had to be computed at lower resolution, yielding a rather imprecise representation of the BKPyV–heparin complex at only 8 Å resolution. Difference density was observed between capsomers as well as at the top of each capsomer (Figure 3c). Density in between pentamers cannot be associated with GAG binding with certainty. However, this site is positively charged, which would fit the observations made for other viruses where GAG binding sites are indeed positively charged. The density observed in the pore of the pentamers could be attributed to the effects of symmetry averaging, but the authors do not completely discount the idea of GAG binding to the pore [49].

A detailed study conducted in 2011 investigated MCPyV cellular tropism and found that this virus sequentially engages GAGs and sialylated glycans. Through multiple

binding and transduction assays, it was shown that the initial attachment of MCPyV to cells requires GAGs such as HS and CS. Particularly, N-sulfated and/or 6-O-sulfated forms of HS are needed for infectious entry of the virus [46].

Structural information on GAG interactions with MCPyV is currently limited to a study by Bayer et al., which analyzed the binding of different GAGs to MCPyV particles and isolated VP1 capsomers with saturation transfer difference nuclear magnetic resonance spectroscopy (STD NMR). Interestingly, it was found that isolated, unassembled VP1 pentamers did not bind any of the tested GAGs, while the assembled virus capsid gave a strong binding signal. Isolated VP1 pentamers are thus not able to establish interactions with GAGs, in contrast to the interactions with sialylated glycans, which were analyzed using VP1 pentamers. Additional STD NMR experiments showed that MCPyV VLPs were able to bind HS-derived disaccharide, heparin-derived octosaccharide, DS-derived hexasaccharide, as well as the HS-derived pentasaccharide Arixtra(fondaparinux). Based on these observations, it is clear that MCPyV can bind a broad range of GAGs, and that the presence of sulfatation seems to be essential for interaction. Most importantly, GAG binding by MCPyV requires more than one VP1 pentamer, suggesting that the GAG binding sites most likely lie in the recessions that separate individual VP1 pentamers. However, attempts to visualize this interaction by solving the structure of MCPyV VLPs with GAGs through crystal soaking or cryoEM studies were not conclusive [50].

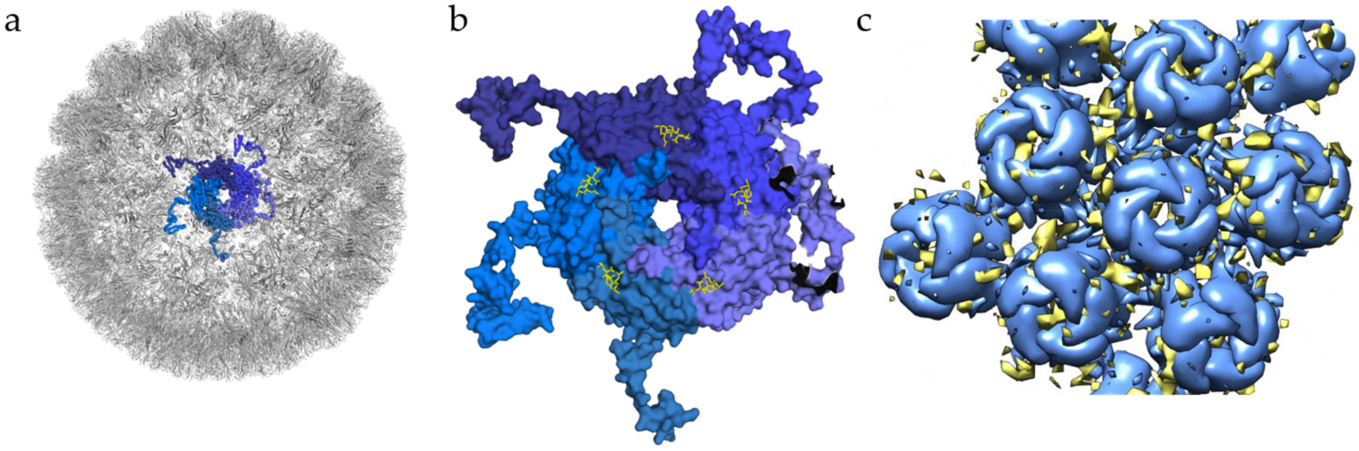


Figure 3. Glycan engagement of BKPyV. (**a**) Capsid structure of BKPyV (PDB ID 6ESB). The capsid is represented in the gray cartoon, except for the highlighted pentamer, where VP1 monomers are displayed as ribbons in different shades of blue. (**b**) Top view of BKPyV VP1 pentamers in interaction with GT1b. The pentamer is displayed as a surface colored in different shades of blue for each subunit. The double sialic acid part of GT1b is represented as yellow sticks. (**c**) Surface representation of BKPyV VP1 pentamers, in blue, associated with the difference map for heparin, in yellow (generated through subtraction of the unliganded VLP map (EMD-3946) from the heparin-VLP map (EMD-3945)). Figures were created using PyMOL [23] and Chimera [51].

5. Conclusions

In this review, we surveyed the available structural data on the binding of GAGs to non-enveloped viruses. It is clear that GAGs are exceedingly important glycans that play key roles in the entry processes of many viruses, including several that are relevant to human health. However, reliable and detailed structural information about contacts between GAGs and virus capsid proteins is essentially non-existent. More than anything, our review has therefore highlighted the challenges that hamper structural studies of GAG receptor binding to viruses. Crystallographic methods are in many cases unsuitable, primarily due to the difficulty of obtaining crystals—particularly with asymmetric, heterogeneous molecules such as HS. This moves the focus to cryo-EM. While it is developing rapidly, this method does not yet necessarily match crystallography's resolution and suffers from

a paucity of reliable metrics for assessing structure quality, particularly because of the variety of workflows applied for data processing, which can significantly influence the result. Perhaps more importantly, in the case of virus–GAG interactions, the icosahedral symmetry imposed during processing can result in the averaging out or weakening of the signal of the non-icosahedrally symmetrical ligand. Issues with ligand heterogeneity, and ligand binding to multiple sites on the same capsid also increase this challenge, further lowering the level of detail that can be resolved.

With the continuously increasing quality of cryoEM data and the development of new approaches to data processing, particularly algorithms that allow for non-icosahedral symmetry within a primarily icosahedral structure [52], these challenges can be addressed in the future. Using this approach, one could, for example, visualize longer and heterogeneous ligands that do not adhere to the rules of icosahedral symmetry when binding to virus capsids. Any detailed structural information would drastically increase our understanding of structural mechanisms of virus–GAG binding—an aspect of glycovirology that remains to be more fully explored.

Author Contributions: Conceptualization, M.N.S., J.K., and A.C.S.; writing—original draft preparation, M.N.S., J.K., A.C.S., and T.S.; visualization, M.N.S., J.K., and A.C.S.; supervision, T.S. All authors have read and agreed to the published version of the manuscript.

Funding: M.N.S. was funded by Ministère Français de l'Enseignement Supérieur, de la Recherche et de l´Innovation through l´Université de Nantes as well as the Deutscher Akademischer Austauschdienst and Deutsch-Französische Hochschule; J.K. was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG), grant number FOR2327; and A.C.S. was funded by Viral and Bacterial Adhesin Network Training Innovative Training Network, part of the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant Agreement No. 765042.

Acknowledgments: Molecular graphics and analyses were performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311. We apologize to our many colleagues whose work could not be discussed or cited here.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lefkowitz, E.J.; Dempsey, D.M.; Hendrickson, R.C.; Orton, R.J.; Siddell, S.G.; Smith, D.B. Virus taxonomy: The database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res.* **2018**, *46*, D708–D717. [CrossRef] [PubMed]
2. Suenaga, T.; Arase, H. Viral Interactions with Glycans. *Glycosci. Biol. Med.* **2014**, *785*, 785–794.
3. Gandhi, N.S.; Mancera, R.L. The Structure of Glycosaminoglycans and their Interactions with Proteins. *Chem. Biol. Drug Des.* **2008**, *72*, 455–482. [CrossRef] [PubMed]
4. Rabenstein, D.L. Heparin and heparan sulfate: Structure and function. *Nat. Prod. Rep.* **2002**, *19*, 312–331. [CrossRef]
5. Gallagher, J.T.; Walker, A. Molecular distinctions between heparan sulphate and heparin. Analysis of sulphation patterns indicates that heparan sulphate and heparin are separate families of N-sulphated polysaccharides. *Biochem. J.* **1985**, *230*, 665–674. [CrossRef]
6. Mikami, T.; Kitagawa, H. Biosynthesis and function of chondroitin sulfate. *Biochim. Biophys. Acta (BBA) Gen. Subj.* **2013**, *1830*, 4719–4733. [CrossRef]
7. Trowbridge, J.M.; Gallo, R.L. Dermatan sulfate: New functions from an old glycosaminoglycan. *Glycobiology* **2002**, *12*, 117R–125R. [CrossRef]
8. Dechecchi, M.C.; Melotti, P.; Bonizzato, A.; Santacatterina, M.; Chilosi, M.; Cabrini, G. Heparan Sulfate Glycosaminoglycans Are Receptors Sufficient To Mediate the Initial Binding of Adenovirus Types 2 and 5. *J. Virol.* **2001**, *75*, 8772–8780. [CrossRef]
9. Khan, A.G.; Pichler, J.; Rosemann, A.; Blaas, D. Human Rhinovirus Type 54 Infection via Heparan Sulfate Is Less Efficient and Strictly Dependent on Low Endosomal pH. *J. Virol.* **2007**, *81*, 4625–4632. [CrossRef]
10. Cagno, V.; Tseligka, E.D.; Jones, S.T.; Tapparel, C. Heparan Sulfate Proteoglycans and Viral Attachment: True Receptors or Adaptation Bias? *Viruses* **2019**, *11*, 596. [CrossRef]
11. Xie, Q.; Bu, W.; Bhatia, S.; Hare, J.; Somasundaram, T.; Azzi, A.; Chapman, M.S. The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 10405–10410. [CrossRef]
12. Kern, A.; Schmidt, K.; Leder, C.; Müller, O.J.; Wobus, C.E.; Bettinger, K.; Von Der Lieth, C.W.; King, J.A.; Kleinschmidt, J.A. Identification of a Heparin-Binding Motif on Adeno-Associated Virus Type 2 Capsids. *J. Virol.* **2003**, *77*, 11072–11081. [CrossRef]

13. Ng, R.; Govindasamy, L.; Gurda, B.L.; McKenna, R.; Kozyreva, O.G.; Samulski, R.J.; Parent, K.N.; Baker, T.S.; Agbandje-McKenna, M. Structural Characterization of the Dual Glycan Binding Adeno-Associated Virus Serotype 6. *J. Virol.* **2010**, *84*, 12945–12957. [CrossRef]
14. Xie, Q.; Lerch, T.F.; Meyer, N.L.; Chapman, M.S. Structure–function analysis of receptor-binding in adeno-associated virus serotype 6 (AAV-6). *Virology* **2011**, *420*, 10–19. [CrossRef]
15. Pillay, S.; Meyer, N.L.; Puschnik, A.S.; Davulcu, O.; Diep, J.; Ishikawa, Y.; Jae, L.T.; Wosen, J.E.; Nagamine, C.M.; Chapman, M.S.; et al. An essential receptor for adeno-associated virus infection. *Nat. Cell Biol.* **2016**, *530*, 108–112. [CrossRef]
16. Xie, Q.; Spilman, M.; Meyer, N.L.; Lerch, T.F.; Stagg, S.M.; Chapman, M.S. Electron microscopy analysis of a disaccharide analog complex reveals receptor interactions of adeno-associated virus. *J. Struct. Biol.* **2013**, *184*, 129–135. [CrossRef]
17. Hurtley, S.M. Continuing the resolution revolution. *Science* **2018**, *360*, 280–282. [CrossRef]
18. Levy, H.C.; Bowman, V.D.; Govindasamy, L.; McKenna, R.; Nash, K.; Warrington, K.; Chen, W.; Muzyczka, N.; Yan, X.; Baker, T.S.; et al. Heparin binding induces conformational changes in Adeno-associated virus serotype 2. *J. Struct. Biol.* **2009**, *165*, 146–156. [CrossRef]
19. O'Donnell, J.; Taylor, K.A.; Chapman, M.S. Adeno-associated virus-2 and its primary cellular receptor—Cryo-EM structure of a heparin complex. *Virology* **2009**, *385*, 434–443. [CrossRef]
20. Xie, Q.; Spear, J.M.; Noble, A.J.; Sousa, D.R.; Meyer, N.L.; Davulcu, O.; Zhang, F.; Linhardt, R.J.; Stagg, S.M.; Chapman, M.S. The 2.8 Å Electron Microscopy Structure of Adeno-Associated Virus-DJ Bound by a Heparinoid Pentasaccharide. *Mol. Ther. Methods Clin. Dev.* **2017**, *5*, 1–12. [CrossRef]
21. Xie, Q.; Yoshioka, C.K.; Chapman, M.S. Adeno-Associated Virus (AAV-DJ)—Cryo-EM Structure at 1.56 Å Resolution. *Viruses* **2020**, *12*, 1194. [CrossRef] [PubMed]
22. Wang, D.; Tai, P.W.L.; Gao, G. Adeno-associated virus vector as a platform for gene therapy delivery. *Nat. Rev. Drug Discov.* **2019**, *18*, 358–378. [CrossRef] [PubMed]
23. Schrödinger, L.L.C. The {PyMOL} Molecular Graphics System, Version~1.8. 2015. Available online: https://pymol.org/2/ (accessed on 14 April 2021).
24. Bzhalava, D.; Guan, P.; Franceschi, S.; Dillner, J.; Clifford, G. A systematic review of the prevalence of mucosal and cutaneous human papillomavirus types. *Virology* **2013**, *445*, 224–231. [CrossRef]
25. Lei, J.; Ploner, A.; Elfström, K.M.; Wang, J.; Roth, A.; Fang, F.; Sundström, K.; Dillner, J.; Sparén, P. HPV Vaccination and the Risk of Invasive Cervical Cancer. *N. Engl. J. Med.* **2020**, *383*, 1340–1348. [CrossRef]
26. Crosbie, E.J.; Einstein, M.H.; Franceschi, S.; Kitchener, H.C. Human papillomavirus and cervical cancer. *Lancet* **2013**, *382*, 889–899. [CrossRef]
27. Shafti-Keramat, S.; Handisurya, A.; Kriehuber, E.; Meneguzzi, G.; Slupetzky, K.; Kirnbauer, R. Different Heparan Sulfate Proteoglycans Serve asCellular Receptors for HumanPapillomaviruses. *J. Virol.* **2003**, *77*, 13125–13135. [CrossRef]
28. Buck, C.B.; Thompson, C.D.; Roberts, J.N.; Müller, M.; Lowy, D.R.; Schiller, J.T. Carrageenan Is a Potent Inhibitor of Papillomavirus Infection. *PLOS Pathog.* **2006**, *2*, e69. [CrossRef] [PubMed]
29. Rommel, O.; Dillner, J.; Fligge, C.; Bergsdorf, C.; Wang, X.; Selinka, H.-C.; Sapp, M. Heparan sulfate proteoglycans interact exclusively with conformationally intact HPV L1 assemblies: Basis for a virus-like particle ELISA. *J. Med. Virol.* **2005**, *75*, 114–121. [CrossRef]
30. Bousarghin, L.; Touzé, A.; Combita-Rojas, A.-L.; Coursaget, P. Positively charged sequences of human papillomavirus type 16 capsid proteins are sufficient to mediate gene transfer into target cells via the heparan sulfate receptor. *J. Gen. Virol.* **2003**, *84*, 157–164. [CrossRef]
31. Sun, J.; Yu, J.S.; Jin, S.; Zha, X.; Wu, Y.; Yu, Z. Interaction of synthetic HPV-16 capsid peptides with heparin: Thermodynamic parameters and binding mechanism. *J. Phys. Chem. B* **2010**, *114*, 9854–9861. [CrossRef]
32. Dasgupta, J.; Bienkowska-Haba, M.; Ortega, M.E.; Patel, H.D.; Bodevin, S.; Spillmann, D.; Bishop, B.; Sapp, M.; Chen, X.S. Structural Basis of Oligosaccharide Receptor Recognition by Human Papillomavirus. *J. Biol. Chem.* **2011**, *286*, 2617–2624. [CrossRef]
33. Guan, J.; Bywaters, S.M.; Brendle, S.A.; Ashley, R.E.; Makhov, A.M.; Conway, J.F.; Christensen, N.D.; Hafenstein, S. Cryoelectron Microscopy Maps of Human Papillomavirus 16 Reveal L2 Densities and Heparin Binding Site. *Structure* **2017**, *25*, 253–263. [CrossRef]
34. Maginnis, M.S. Human Polyomaviruses (Papillomaviridae). In *Encyclopedia of Virology*, 4th ed.; Bamford, D.H., Zuckerman, M., Eds.; Academic Press: Oxford, UK, 2021; pp. 518–527.
35. Hirsch, H.H. Polyomavirus BK nephropathy: A (re-)emerging complication in renal transplantation. *Am. J. Transplant.* **2002**, *2*, 25–30. [CrossRef]
36. Padgett, B.L.; Walker, D.L.; ZuRhein, G.M.; Eckroade, R.J.; Dessel, B.H. Cultivation of papova-like virus from human brain with progressive multifocal leucoencephalopathy. *Lancet* **1971**, *1*, 1257–1260. [CrossRef]
37. Feng, H.; Shuda, M.; Chang, Y.; Moore, P.S. Clonal Integration of a Polyomavirus in Human Merkel Cell Carcinoma. *Science* **2008**, *319*, 1096–1100. [CrossRef]
38. Belnap, D.M.; Olson, N.H.; Cladel, N.M.; Newcomb, W.W.; Brown, J.C.; Kreider, J.W.; Christensen, N.D.; Baker, T.S. Conserved Features in Papillomavirus and Polyomavirus Capsids. *J. Mol. Biol.* **1996**, *259*, 249–263. [CrossRef]

39. Stehle, T.; Yan, Y.; Benjamin, T.L.; Harrison, S.C. Structure of murine polyomavirus complexed with an oligosaccharide receptor fragment. *Nat. Cell Biol.* **1994**, *369*, 160–163. [CrossRef]
40. Neu, U.; Hengel, H.; Blaum, B.S.; Schowalter, R.M.; Macejak, D.; Gilbert, M.; Wakarchuk, W.W.; Imamura, A.; Ando, H.; Kiso, M.; et al. Structures of Merkel Cell Polyomavirus VP1 Complexes Define a Sialic Acid Binding Site Required for Infection. *PLOS Pathog.* **2012**, *8*, e1002738. [CrossRef]
41. Neu, U.; Allen, S.-A.A.; Blaum, B.S.; Liu, Y.; Frank, M.; Palma, A.S.; Ströh, L.J.; Feizi, T.; Peters, T.; Atwood, W.J.; et al. A Structure-Guided Mutation in the Major Capsid Protein Retargets BK Polyomavirus. *PLoS Pathog.* **2013**, *9*, e1003688. [CrossRef]
42. Khan, Z.M.; Liu, Y.; Neu, U.; Gilbert, M.; Ehlers, B.; Feizi, T.; Stehle, T. Crystallographic and glycan microarray analysis of human polyomavirus 9 VP1 identifies N-glycolyl neuraminic acid as a receptor candidate. *J. Virol.* **2014**, *88*, 6100–6111. [CrossRef]
43. Ströh, L.J.; Gee, G.V.; Blaum, B.S.; Dugan, A.S.; Feltkamp, M.C.W.; Atwood, W.J.; Stehle, T. Trichodysplasia spinulosa-Associated Polyomavirus Uses a Displaced Binding Site on VP1 to Engage Sialylated Glycolipids. *PLoS Pathog.* **2015**, *11*, e1005112. [CrossRef] [PubMed]
44. Ströh, L.J.; Rustmeier, N.H.; Blaum, B.S.; Botsch, J.; Rößler, P.; Wedekink, F.; Lipkin, W.I.; Mishra, N.; Stehle, T. Structural Basis and Evolution of Glycan Receptor Specificities within the Polyomavirus Family. *mBio* **2020**, *11*. [CrossRef] [PubMed]
45. Neu, U.; Maginnis, M.S.; Palma, A.S.; Ströh, L.J.; Nelson, C.D.; Feizi, T.; Atwood, W.J.; Stehle, T. Structure-Function Analysis of the Human JC Polyomavirus Establishes the LSTc Pentasaccharide as a Functional Receptor Motif. *Cell Host Microbe* **2010**, *8*, 309–319. [CrossRef] [PubMed]
46. Schowalter, R.M.; Pastrana, D.V.; Buck, C.B. Glycosaminoglycans and sialylated glycans sequentially facilitate Merkel cell polyomavirus infectious entry. *PLoS Pathog.* **2011**, *7*, e1002161. [CrossRef] [PubMed]
47. Geoghegan, E.M.; Pastrana, D.V.; Schowalter, R.M.; Ray, U.; Gao, W.; Ho, M.; Pauly, G.T.; Sigano, D.M.; Kaynor, C.; Cahir-McFarland, E.; et al. Infectious Entry and Neutralization of Pathogenic JC Polyomaviruses. *Cell Rep.* **2017**, *21*, 1169–1179. [CrossRef]
48. Liddington, R.C.; Yan, Y.; Moulai, J.; Sahli, R.; Benjamin, T.L.; Harrison, S.C. Structure of simian virus 40 at 3.8-A resolution. *Nature* **1991**, *354*, 278–284. [CrossRef] [PubMed]
49. Hurdiss, D.L.; Frank, M.; Snowden, J.S.; Macdonald, A.; Ranson, N.A. The Structure of an Infectious Human Polyomavirus and Its Interactions with Cellular Receptors. *Structure* **2018**, *26*, 839–847. [CrossRef]
50. Bayer, N.J.; Januliene, D.; Zocher, G.; Stehle, T.; Moeller, A.; Blaum, B.S. Structure of Merkel Cell Polyomavirus Capsid and Interaction with Its Glycosaminoglycan Attachment Receptor. *J. Virol.* **2020**, *94*, 1–17. [CrossRef]
51. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. [CrossRef]
52. Ilca, S.L.; Kotecha, A.; Sun, X.; Poranen, M.M.; Stuart, D.I.; Huiskonen, J.T. Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes. *Nat. Commun.* **2015**, *6*, 4–11. [CrossRef] [PubMed]