

Guide to the Structure and Function of the Adenovirus Capsid

By Logan Thrasher Collins

For this guide, I will explain the fundamental biology of adenovirus capsid proteins with an emphasis on the context of gene therapy. While the guide is meant primarily for readers with an interest in applying adenovirus to gene therapy, it will not include much discussion of the techniques and technologies involved in engineering adenoviruses for such purposes. If you are interested in learning more about adenovirus engineering, you may enjoy my review paper “Synthetic Biology Approaches for Engineering Next-Generation Adenoviral Gene Therapies” [1]. Here, I will focus mostly on the capsid of human adenovirus serotype 5 (Ad5) since it is the most commonly used type of adenovirus employed in gene therapy research, but I will occasionally describe other types of adenoviruses when necessary. Many of the presented concepts remain the same or similar across other types of adenoviruses.

The adenovirus consists of an icosahedral protein capsid enclosing a double-stranded DNA (dsDNA) genome. It possesses 12 fiber proteins which protrude from the capsid and helps to facilitate cellular transduction. Adenoviruses are nonenveloped and approximately 90 nm in diameter (not including the fibers). The Ad5 genome is about 36 kb in size. Major capsid proteins of the adenovirus include the hexon, penton, and fiber. The minor capsid proteins are protein IIIa, protein VI, protein VIII, and protein IX. Inside the capsid, there are core proteins including protein V, protein VII, protein μ (also known as protein X), adenovirus proteinase (AVP), protein IVa2, and terminal protein (TP) [2]. There are also many proteins expressed during adenovirus infection which are not incorporated into mature capsids, including the E1A proteins (289R, 243R, 217R, 171R, and 55R), the E1B proteins (52k and 55k), the adenoviral DNA polymerase, and more [3].

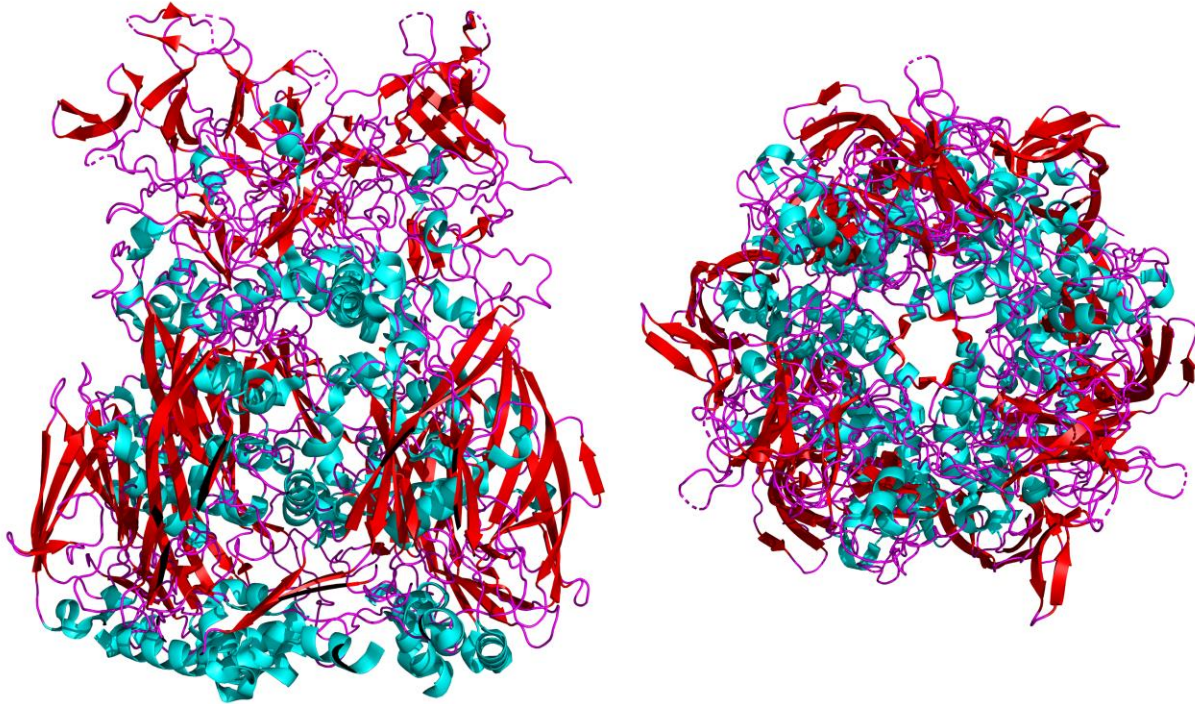
Ad5's genome contains a variety of transcriptional units which are expressed at different times during the viral life cycle [3]. The E1A, E1B, E2A, E2B, E3, and E4 transcriptional units are expressed early during cellular infection. Their proteins are involved in DNA replication, transcriptional regulation, and suppression of host immune responses. The L1, L2, L3, L4, and L5 transcriptional units are expressed later in the life cycle. Their products include most of the capsid proteins as well as other proteins involved in packaging and assembly. Each transcriptional unit can produce multiple mRNAs through the host's alternative splicing machinery.

Major capsid proteins

Hexon

Adenovirus hexon represents the main structural component of the capsid. It is encoded as one of the products of the Ad5 L3 gene. Each capsid contains 240 trimers of the hexon protein (720 monomers) and each facet of the icosahedron consists of 12 trimers [4]. The lower part of each hexon monomer consists of two eight-stranded β barrels linked by a β -sheet. The eight-stranded β -barrels are known as jellyroll domains. In between the β -strands, long loops are present. These loops contain the seven hypervariable regions (HVRs) of the hexon, which differ in sequence composition between distinct adenovirus

types. The loops form the upper portion of each hexon. HVR1 of Ad5 includes a 32-residue acidic loop which might be involved in neutralizing host defensins. The valley between the loop towers of Ad5 has been shown to interact with coagulation factors as well as to bind to the CD46 cellular receptor as an alternative cell entry mechanism.

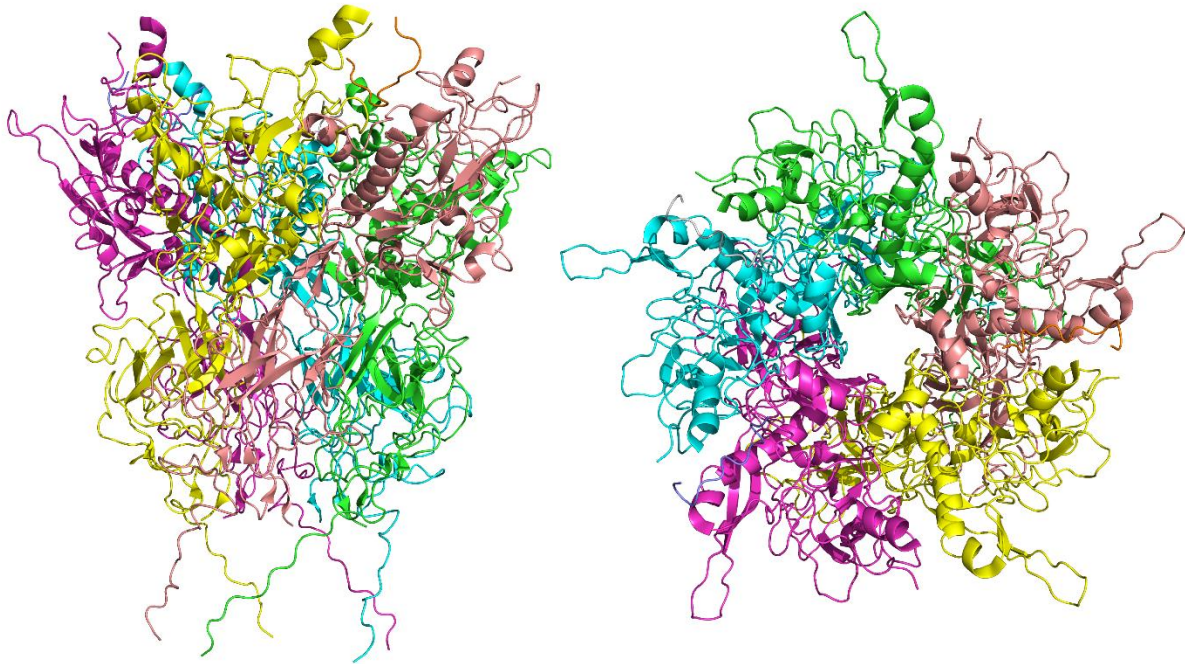


Here, the structure of the Ad5 hexon trimer is shown from a side view and from a top view (PDB 1P30). All β -sheets are red, α -helices are cyan, and loops are magenta. Jellyroll domains are visible at the base of the side view and the HVR loops can be seen in the upper half of the side view. In the top view, the hexagonal shape of the hexon is clearly visible. The N- and C- termini are both located near the bottom of the hexon (adjacent to the inside of the virion). Some disordered regions are shown as dashed lines.

Penton

The 12 pentons serve to fill pentagonal gaps within the icosahedral capsid (which arise due to the geometry of the hexons) [4]. Penton is encoded as one of the products of the Ad5 L2 gene. Each penton also acts as a base onto which a fiber protein is anchored. Adenovirus pentons are pentamers, with each monomeric subunit consisting of a single jellyroll domain for the lower part and both a hypervariable loop and a variable loop at the top. In Ad5 and many other human adenoviruses, the penton hypervariable loop includes an RGD amino acid sequence. RGD is both an α_v integrin binding motif and is a target for adenovirus neutralization by the enteric defensin HD5. Importantly, the penton's RGD motif is essential for cellular transduction into clathrin-coated pits [5]. RGD may also play some role in endosomal escape. The other penton variable loop (distinct from the hypervariable loop) is poorly understood from a functional standpoint. Both the hypervariable loop and the variable loop might serve as decent sites for sequence modification in the context of gene therapy vectors. The penton N-terminal domain consists of approximately 50 amino acid sequence which extends into the inside of the

adenovirus virion. This sequence is mostly disordered except for the part nearest to the jellyroll domain (residues 37-51 in Ad5), which interacts with two copies of protein IIIa.

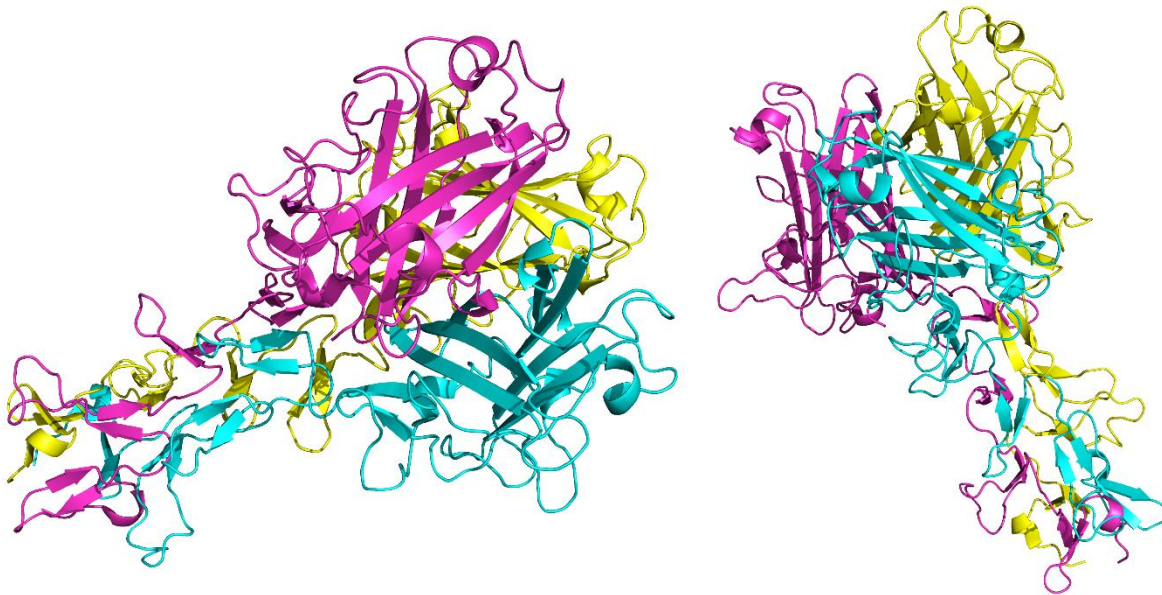


Here, the structure of the Ad5 penton is shown from side and top views (PDB 3IZO). Coloration is by subunit. In the side view, the intravirion N-terminal domains are visible at the bottom, the jellyroll domains can be seen as the groups of β -sheets in the middle, and the loops are present at the upper region. The top view clearly illustrates the pentagonal symmetry of the penton. It should be noted that, in this structure, some of the loops are missing due to the difficulty of reconstructing them at high resolution. Of special relevance here is that the loop with the RGD sequence should be located at the top of the penton (in the gap between the uppermost α -helix and a nearby loop which both terminate prematurely).

Fiber

Ad5's 12 trimeric fibers are anchored onto the tops of the pentons [4]. They are encoded as a product of the L5 gene. These fibers initiate cellular transduction through binding of the knob domain to cellular receptors. The primary receptor for Ad5 is the coxsackievirus and adenovirus receptor (CAR). That said, it should be noted that Ad5's fiber knob can also bind to alternative receptors such as vascular cell adhesion molecule 1 and heparan sulfate proteoglycans. For Ad5, the fiber is about 37 nm in length, but other adenoviruses can have shorter or longer fibers [6]. Fibers consist of an N-terminal tail domain, a shaft domain, and a C-terminal knob (also called head) domain [4]. The three N-terminal tails anchor into some of the clefts between penton monomers, likely via a hydrophobic ring region. The shaft consists of a structure known as a trimeric β -spiral. Shaft flexibility plays a role in cellular transduction by facilitating interaction of the penton with its integrin receptor after binding of the knob to CAR. Many adenovirus fibers are known to have hinges at the third β -repeat from the N-terminal tail domain [7]. These hinges arise from an insertion of a few extra amino acids within the third β -repeat which disrupts its structure

and allows for it to flex. The C-terminal knob domain consists of an antiparallel β -sandwich and is responsible for trimerization of the fiber [4]. Its C-termini are oriented back towards the capsid of the adenovirus.

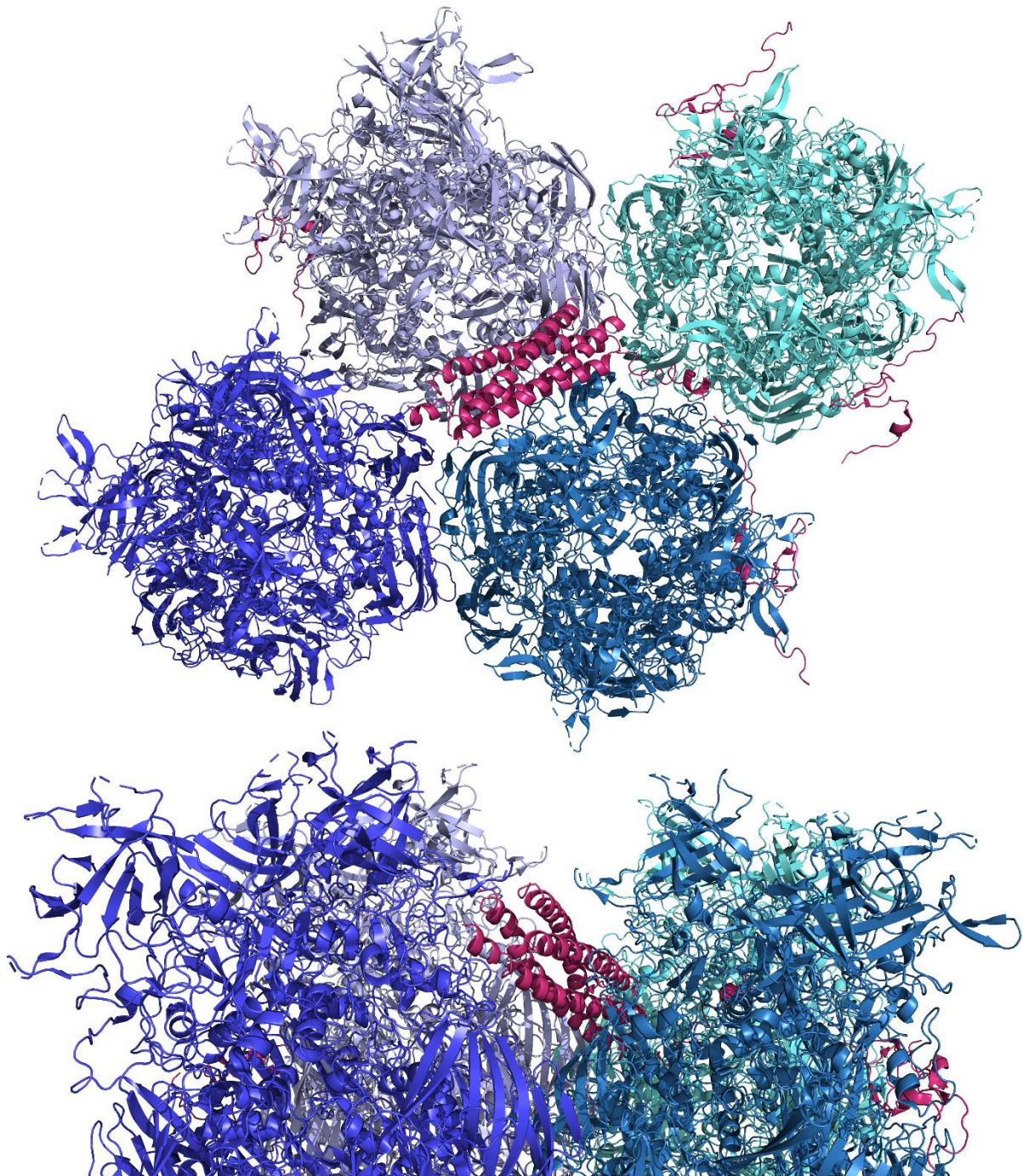


Here, part of the structure of an Ad2 fiber is shown from two perspective views (PDB 1QIU). Though there are structures of the Ad5 fiber components available, only the above Ad2 fiber structure has been assembled into a complex with and made publicly available. The Ad2 fiber is highly similar to the Ad5 fiber. Both Ad5 and Ad2 fibers have 22 β -repeats. Only a few β -repeats are included in the above structures, but that should be enough to grant an intuitive understanding of the general fiber organization.

Minor capsid proteins

Protein IX

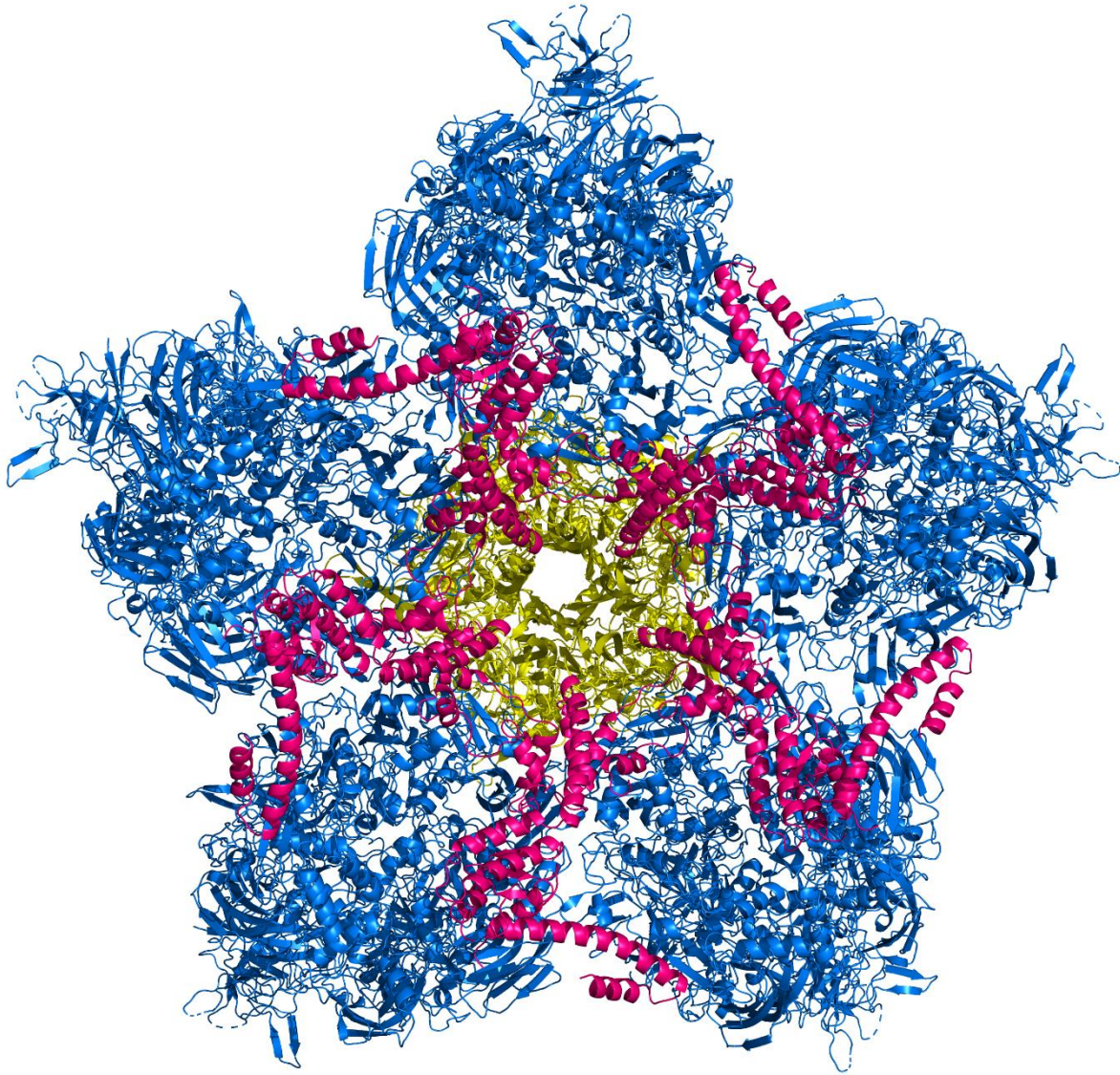
Ad5 protein IX (pIX) is a 140 amino acid protein found nestled between hexons which confers greater thermostability to the capsid relative to mutants lacking pIX [4]. There are 240 copies of pIX in the capsid. It has an N-terminal domain, a rope domain, and a C-terminal domain. The N-terminal domains of three pIX monomers interlace to form a triskelion structure in the valleys between some of the hexons. The rope domain (also called linker domain) is often disordered and connects the N- and C-terminal domains. The C-terminal domain is an α -helix which forms a coiled-coil structure with the helices of other copies of pIX monomer. This coiled coil consists of four α -helices (three parallel and one antiparallel), each from a different pIX monomer. Four triskelions and three α -helix bundles are present in each icosahedral facet of the capsid. It should be noted that some of the triskelions take on slightly different structural features depending on which hexons they are associated with within a given facet [8]. Though all of the C-termini of pIX are exposed on the capsid surface, they can still be described as resting within crevices between hexons. Because of this, spacer peptides are usually necessary when engineering Ad5 pIX-fusions such that the added protein is elevated out of the crevices [9].



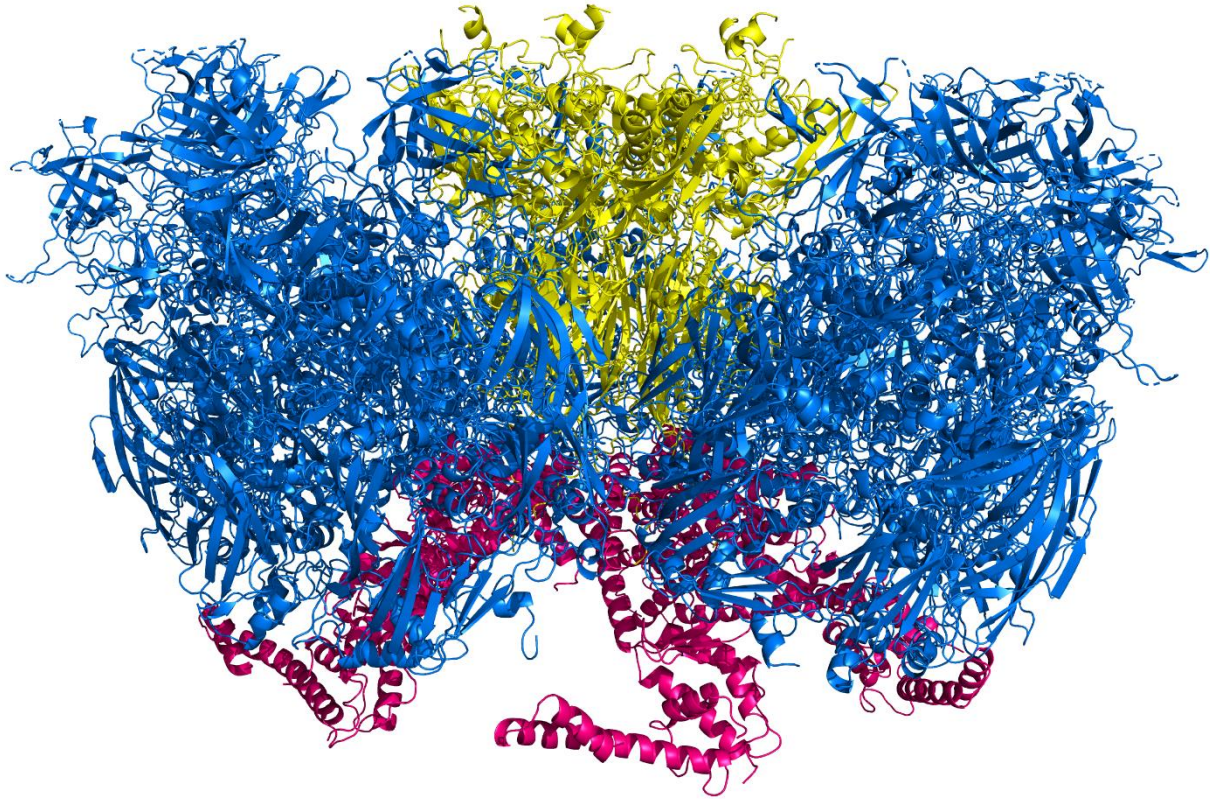
Here, four copies of Ad5 pIX are shown interlacing among four hexons (top and side views) (PDB 6B1T). The C-terminal domain α -helical bundle of pIX is clearly visible. The N-terminal domain triskelion structures are not visible in these views. Hexons are portrayed in cool colors and the pIX copies are shown in magenta. Some disordered regions are shown as dashed lines.

Protein IIIa

The Ad5 protein IIIa (pIIIa) plays a structural role in stabilizing the capsid from the inside [4]. Five copies of pIIIa are found under each vertex of the Ad5 capsid. It is 585 amino acids in length, but only residues 7 to 300 have been structurally traced at high resolution. Its N-terminal domain connects the penton and the five adjacent hexons. (These are known as the peripentonal hexons. The peripentonal hexons plus the penton are collectively named the group-of-six or GOS) Its C-terminal domain binds protein VIII (another structural protein which will be discussed later). The traced part of the pIIIa structure consists of two globular domains connected by a long α -helix.



Above, traced parts of five pIIIa proteins are shown on the underside of a part of the Ad5 capsid (perspective is from the interior) (PDB 6B1T). Hexons are colored blue, the penton is colored yellow, and pIIIa is colored bright pink. The same structure is shown below from a side perspective.



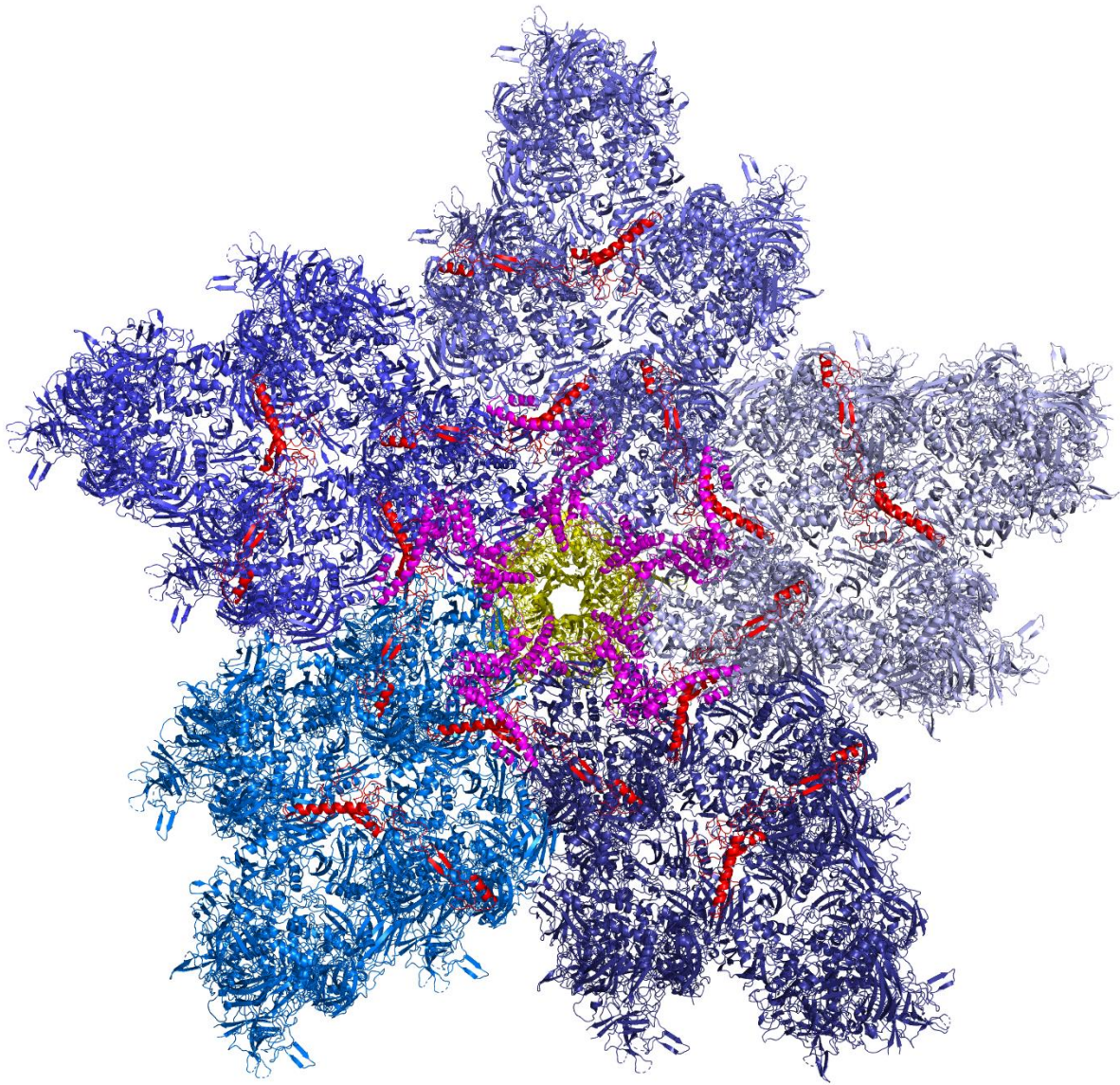
Protein VI

Ad5 protein VI (pVI) starts out as 250 amino acids long but is cleaved by AVP at two sites, yielding multiple peptides [4]. The first site is after residue 33 and the second is after residue 239. The middle part contains a predicted amphipathic α -helix (residues 34-54) which inserts into host endosomal membrane. This alters the membrane's curvature and helps facilitate lysis of the endosome, allowing the adenovirus to escape into the cytosol. The middle part also contains a domain (residues 109-143) which sometimes binds to the inner surface of the capsid in the cavities between certain hexons. The N-terminal peptide pVI_N also binds to cavities between hexons. It has been suggested that this affinity hides the first pVI cleavage site in these cavities, preventing release of the membrane lytic peptide. During intracellular trafficking, environmental changes may allow adenovirus protein VII (a core protein) to outcompete pVI for the binding sites between hexons, causing release of the membrane lytic peptide. Finally, the C-terminal peptide pVI_C is a cofactor which helps activate AVP. The pVI_C peptide binds covalently to AVP and slides along the adenoviral genome, using the DNA as a track to reach all of the substrates in the core and the inner capsid surface. There are approximately 360 copies of protein VI in the Ad5 virion. Unfortunately, high-resolution structural data on pVI are scarce due to its variable position in the adenovirus virion.

Protein VIII

Ad5's protein VIII (pVIII) also contributes to structurally stabilizing the adenoviral capsid from the interior [4]. It starts as a 227-residue protein which is cleaved by AVP at three sites, yielding two large peptides and two small peptides. The two large peptides stay

together and bind between hexons. Some pVIII copies wedge between pIIIa and the peripentonal hexons, helping to connect the peripentonal hexons to the next set of surrounding hexons. Some pVIII copies are located underneath the nine hexons on the middle face of each icosahedral facet (known as the group-of-nine or GON). An interesting aspect of pVIII-hexon interactions is that pVIII can engage in β -sheet augmentation, where a β -strand from pVIII is incorporated into one of the jellyroll domains of a nearby hexon. Not much is known about the two smaller peptides from pVIII except that these peptides do not appear to bind the capsid in a symmetric fashion.



Here, the traced parts of pVIII (red) are shown interwoven into a piece of the Ad5 capsid from an interior perspective (PDB 6B1T). Hexons are shown in shades of blue, the penton is shown in yellow, and pIIIa is displayed in bright pink.

Core proteins which interact directly with the capsid

Protein V

Adenovirus protein V (pV) is a positively charged protein which can form heterodimers with the pVII core protein [4]. That said, pV exists in a dimer-monomer equilibrium, so the binding to pVII is often transient. There also are direct associations between pV and the pVI capsid protein. These associations between pVII, pV, and pVI likely act to bridge the adenovirus core with the adenovirus capsid. In addition, pV-pVII heterodimers might interact with core protein μ . Each virion contains about 150 copies of pV. Most of the copies of pV are released during the beginning of uncoating. Interestingly, pV is not essential for adenovirus capsid assembly.

Protein VII

Protein VII (pVII) is a positively charged protein which plays a central role in condensing the adenovirus genome to fit into the capsid [4]. It has many arginine residues which contribute to its positive charge. AVP cleaves pVII at residues 13 and 24. The resulting middle peptide (including amino acids 13 through 24) might compete with pVI for hexon binding sites during adenovirus assembly. As mentioned earlier, environmental changes during intracellular trafficking may allow pVII to outcompete pVI for their hexon binding sites, causing release of the membrane lytic peptide from pVI cleavage. Though pVII acts as a functional analogue of the histone, it does not share much structural similarity with histones and does not replace histones when introduced into the cellular nucleus [2]. During infection, the viral genomic DNA as complexed with pVII is imported through nuclear pores. While pVII is important for condensing the adenoviral genome, it is not strictly required for assembly and packaging. In addition, pVII functions in signaling for the suppression of host innate immune responses. It binds to high mobility group B (HMGB) protein 1, a factor which is normally released from cells exposed to inflammation and which acts as a danger signal for the immune system. The adenoviral pVII prevents release of HMGB protein 1 and thereby dampens innate immune responses. Finally, pVII helps to regulate the progression of various steps during adenovirus genome replication.

Conclusion

This guide has centered on explaining the structures and functions of the Ad5 capsid proteins as well as the core proteins which are involved in key structural interactions with the capsid proteins. But this is only the beginning of learning about adenovirus biology. As mentioned in the introductory section, there are other core proteins including protein μ , the adenovirus proteinase, protein IVa2, and terminal protein which primarily interact with the adenovirus genome. Furthermore, the complex life cycle of the adenovirus requires numerous replication and packaging proteins (as well as interesting interactions with host cells) not covered here. Despite the specific focus of this guide, I hope that it is helpful to the reader for gaining a better idea of how the adenovirus capsid works. Perhaps this text will even provide a valuable bedrock of understanding for interested readers who are working on Ad5 capsid engineering projects.

References

- [1] L. T. Collins and D. T. Curiel, "Synthetic Biology Approaches for Engineering Next-Generation Adenoviral Gene Therapies," *ACS Nano*, Aug. 2021, doi: 10.1021/acsnano.1c04556.

- [2] S. Kulanayake and S. K. Tikoo, "Adenovirus Core Proteins: Structure and Function," *Viruses*, vol. 13, no. 3. 2021, doi: 10.3390/v13030388.
- [3] Y. S. Ahi and S. K. Mittal, "Components of Adenovirus Genome Packaging," *Frontiers in Microbiology*, vol. 7. p. 1503, 2016, [Online]. Available: <https://www.frontiersin.org/article/10.3389/fmicb.2016.01503>.
- [4] J. Gallardo, M. Pérez-Illana, N. Martín-González, and C. San Martín, "Adenovirus Structure: What Is New?," *International Journal of Molecular Sciences*, vol. 22, no. 10. 2021, doi: 10.3390/ijms22105240.
- [5] W. C. Russell, "Adenoviruses: update on structure and function," *J. Gen. Virol.*, vol. 90, no. 1, pp. 1–20, 2009, doi: <https://doi.org/10.1099/vir.0.003087-0>.
- [6] E. Vigne *et al.*, "Genetic manipulations of adenovirus type 5 fiber resulting in liver tropism attenuation," *Gene Ther.*, vol. 10, no. 2, pp. 153–162, 2003, doi: 10.1038/sj.gt.3301845.
- [7] S. A. Nicklin, E. Wu, G. R. Nemerow, and A. H. Baker, "The influence of adenovirus fiber structure and function on vector development for gene therapy," *Mol. Ther.*, vol. 12, no. 3, pp. 384–393, Sep. 2005, doi: 10.1016/j.ymthe.2005.05.008.
- [8] V. S. Reddy and G. R. Nemerow, "Structures and organization of adenovirus cement proteins provide insights into the role of capsid maturation in virus entry and infection," *Proc. Natl. Acad. Sci.*, vol. 111, no. 32, pp. 11715 LP – 11720, Aug. 2014, doi: 10.1073/pnas.1408462111.
- [9] J. Vellinga *et al.*, "Spacers Increase the Accessibility of Peptide Ligands Linked to the Carboxyl Terminus of Adenovirus Minor Capsid Protein IX," *J. Virol.*, vol. 78, no. 7, pp. 3470 LP – 3479, Apr. 2004, doi: 10.1128/JVI.78.7.3470-3479.2004.