Cryo-EM structure of enteric adenovirus HAdV-F41 highlights structural divergence among human adenoviruses

Marta Pérez-Illana[1], Marta Martínez[1], Gabriela N. Condezo[1], Mercedes Hernando-Pérez[1], Casandra Mangroo[2], Martha Brown[2,3], Roberto Marabini[4], Carmen San Martín[1,*]

[1]Department of Macromolecular Structures. Centro Nacional de Biotecnología (CNB-CSIC). Madrid (Spain)

[2]Department of Laboratory Medicine and Pathobiology, University of Toronto. Toronto, Ontario (Canada)

[3]Department of Molecular Genetics. University of Toronto. Toronto, Ontario (Canada)

[4]Escuela Politécnica Superior. Universidad Autónoma de Madrid. Madrid (Spain)

*Corresponding author: carmen@cnb.csic.es

Abstract

Enteric adenoviruses are one of the main causes of viral gastroenteritis in the world. To carry out a successful infection, the virions must withstand the harsh conditions found in the gut. This requirement suggests that capsid stability must be different from that of other adenoviruses. We have determined the structure of a human enteric adenovirus, HAdV-F41, at 4.0 Å resolution by single particle averaging cryo-electron microscopy, and compared it with that of other adenoviruses with respiratory (HAdV-C5) and ocular (HAdV-D26) tropisms. While the overall structures of hexon, penton base and internal minor coat proteins IIIa and VIII are conserved, we observe partially ordered elements reinforcing the vertex region, which suggests their role in enhancing the physicochemical capsid stability of HAdV-F41. Unexpectedly, we find an organization of the external minor coat protein IX different from all previously characterized human and non-human mastadenoviruses. Knowledge of the structure of enteric adenoviruses can provide a starting point for the design of vectors suitable for oral delivery or intestinal targeting.

Main

Adenoviruses (AdVs) are pathogens which can be engineered to become therapeutic tools (Appaiahgari and Vrati, 2015; Baker *et al.*, 2018). Over 100 types of human AdVs (HAdV) have been described so far (http://hadvwg.gmu.edu/; https://talk.ictvonline.org/), and grouped in seven species (HAdV-A to G). HAdVs cause acute infections in the respiratory and gastrointestinal tracts, as well as conjunctivitis. Most AdV infections are subclinical, but they can cause significant morbidity and even mortality in immunocompromised patients (Lion, 2014). There are currently no vaccines for use in the general population or approved antiviral therapies for AdV infections.

High resolution (3.2-3.5 Å) structures are available only for two HAdVs: HAdV-C5 and HAdV-D26 (Dai *et al.*, 2017; Liu *et al.*, 2010; Yu *et al.*, 2017). The AdV capsid stands out among the non-enveloped viruses because of its large size (~950 Å, 150 MDa), triangulation number (*pseudo* T = 25), and complex composition. The facets are formed by 240 trimeric hexons, while pentamers of penton base protein fill the vertices. Receptor-binding trimeric fibres of various lengths, depending on the virus type, project from the penton bases (Nicklin *et al.*, 2005). Minor coat proteins IIIa and VIII on the inner capsid surface, and IX on the outer one, contribute to modulate the quasi-equivalent icosahedral interactions. The membrane lytic protein VI competes with genome condensing protein VII for binding to an internal cavity in hexons (Dai *et al.*, 2017; Hernando-Perez *et al.*, 2020).

HAdVs in species F are unusual in that they have a narrow enteric tropism. HAdV-F40 and HAdV-F41 are one of the three major causes of viral gastroenteritis in young children, together with rota- and noroviruses (Chhabra *et al.*, 2013; Corcoran *et al.*, 2014). They account for more than half of the AdVs identified in stools of immunocompetent symptomatic children (Brandt *et al.*, 1985; Brown, 1990; Uhnoo *et al.*, 1984). Differences in the capsid structure of HAdV-F, relative to HAdV-C, may confer stability to gastric conditions en route to the intestine *in vivo*. Indeed, infectivity assays after incubation in acidic conditions have shown that HAdV-F41 is

stable, and even possibly activated, at low pH, in contrast to HAdV-C5 (Favier *et al.*, 2004). The fact that HAdV-F viruses code for two different fibres has been related with their enteric tropism, although the exact relation is unclear (Favier *et al.*, 2002; Kidd *et al.*, 1993).

Here we use cryo-EM to obtain the structure of the HAdV-F41 virion, and compare it with the two other known HAdV structures. We interpret the structural differences in the context of the high physicochemical stability of the enteric AdV capsid.

Results

Physicochemical stability of HAdV-F41 virions

To evaluate the stability of HAdV-F41 virions, we assessed their infectivity after treatment with synthetic gastric, intestinal, or gastric followed by intestinal fluids. GFP fluorescence measured by flow cytometry was used to determine the number of infected cells in HEK293 cultures. Neither gastric nor intestinal conditions reduced infectivity. Interestingly, consecutive incubation in synthetic gastric and intestinal fluid resulted in a clear increase of GFP expression, indicating enhanced infectivity (**Fig. S1a**). That is, not only is HAdV-F41 resistant to acidic pH as previously shown (Favier *et al.*, 2004), but it is also resistant to salt and proteases, and even activated by the combined action of several of these factors.

We also examined the physical stability of HAdV-F41 particles in comparison to HAdV-C5 using extrinsic fluorescence of the DNA intercalating agent YOYO-1 to characterize capsid disruption as a function of temperature. The fluorescence emission of YOYO-1 increased with temperature for both specimens, as genomes became exposed to the solvent (**Fig. S1b)**. While the half-transition temperature ($T_{0.5}$) estimated for HAdV-C5 was 47ºC (Hernando-Perez *et al.*, 2020), for HAdV-F41 the maximum rate of DNA exposure happened at near 51ºC. This result indicates that HAdV-F41 particles are thermally more stable than HAdV-C5.

Overall HAdV-F41 structure

The cryo-EM map of HAdV-F41 at 4.0 Å resolution (**Fig. 1 and S2, Table S1**) shows the general characteristics common to all previous AdV structures: a ~950 Å diameter *pseudo* T=25 icosahedral particle with 12 trimeric hexon capsomers per facet, plus the pentameric pentons at the vertices. The N-terminus of protein IX lays on the outer capsid surface, in the valleys formed by hexon trimers (**Fig. 1b top**). Minor coat proteins IIIa and VIII, plus small fragments of VI and core protein VII, can be observed on the inner surface of the icosahedral shell,

oriented towards the virion core (**Fig. 1b bottom**). Due to their flexibility and symmetry mismatch with the penton base, fibres cannot be resolved by cryo-EM when imposing icosahedral symmetry as done here.

Except for proteins VI and VIII, the HAdV-F41 structural polypeptides are shorter than their HAdV-C5 counterparts (**Table S2**). Sequence identity varies between 52 and 78%. The predicted isoelectric points for hexon and penton base suggest a more basic capsid surface for HAdV-F41 (Favier *et al.*, 2004), except for the contribution of the external cementing protein IX, which is remarkably more acidic (**Table S2**). On the inner capsid surface, proteins IIIa, VI and VIII are more acidic in the enteric AdV.

Major coat protein: hexon

The architecture of the HAdV-F41 hexon is very similar to that of HAdV-C5 (Dai *et al.*, 2017) (**Table S3**, **Fig. 2a**). The double jelly roll in each monomer that forms the trimer pseudo-hexagonal base is highly conserved. As expected (Crawford-Miksza and Schnurr, 1996), the main differences between the structures are in the loops located in the hexon towers, the so-called hyper variable regions (HVRs) which constitute serotype-specific epitopes exposed on the virion surface (**Fig. 2a-b, and Table S4**). As in HAdV-C5, HVRs in HAdV-F41 are flexible (**Fig. S3**), but we were able to fully trace all of them except HVR4 (**Fig. 2a-b, Table S5**). The 32-residue acidic loop in HVR1, unique to HAdVs in species C (Ebner *et al.*, 2005), has not been traced in any of the available HAdV-C hexon structures. We were able to completely trace HVR1 in HAdV-F41, probably because its shorter length limits flexibility (**Fig. 2b, Table S4**). Lack of the long acidic region reflects on the absence of a large negatively charged patch in the hexon towers of HAdV-F41 (**Table S4**), and could have an influence in the interplay with host factors under the extreme pH conditions in the gastrointestinal tract. However, the acidic stretch is also absent in HAdV-D26, which has ocular tropism and where HVR1 is also wholly traced (Yu *et al.*, 2017). Residue Tyr784 also has RMSD > 2 Å between HAdV-C5 and HAdV-F41,

and could be involved in the different interaction with protein IX (see IX section below). Besides the HVRs, other regions already described for HAdV-C5 and D26 also display variability among the twelve hexon monomers in the icosahedral asymmetric unit (AU) in HAdV-F41 (**Fig. S3**).

Vertex capsomer: penton base

The architecture of the HAdV-F41 penton base protein is very similar to that of HAdV-C5 (Dai *et al.*, 2017) (**Table S3, Fig. 2c)**. As in previously solved structures (Liu *et al.*, 2010; Yu *et al.*, 2017), the first 34 amino acids could not be traced (**Table S5, Fig. 2d**). This region is either flexible, or buried within the noisy, non-icosahedral core density. However, weak density suggests its location in HAdV-F41 (see below, section on additional internal densities). The protein is 63 amino acids shorter than in HAdV-C5 (**Table S2**). This length difference corresponds mainly to a hypervariable loop at the periphery of the pentamer, which in HAdVC-5 bears the integrin-binding RGD sequence motif (**Fig. 2c-d**). All HAdVs bear this RGD motif, except HAdV-F40 and F41, which instead have RGAD and IGDD (Albinsson and Kidd, 1999), and HAdV-D60, which has a deletion of the loop (Robinson *et al.*, 2013). The RGD loop is highly flexible, precluding its tracing in any of the available structures. In HAdV-F41 the IGDD-containing region is 57 residues shorter than the HAdV-C5 RGD loop, but could not be traced either (**Table S5, Fig. 2c-d**), indicating that, despite being much shorter, it is also flexible. Lack of the RGD sequence is not a determinant for enteric tropism, as this motif is present in other enteric AdVs such as HAdV-A31 (UniProtKB entry: D0Z5S7_ADE31) and HAdV-G52 (A0MK51_9ADEN). In spite of lacking the RGD motif, it has recently been shown that HAdV-F41 can bind to laminin-binding integrins (Rajan *et al.*, 2018). The variable loop (Glu148-Leu156), one of the least conserved regions in the penton base sequences and likewise exposed at the periphery of the pentamer, has also a different conformation in HAdV-F41 (**Fig.**

2c-d). The role of this loop is unknown, but it has been proposed as a site to be engineered for gene therapy (Zubieta *et al.*, 2005).

The N-terminal peptides of AdV fibres (**Fig. 2d**) bind to the groove formed at the interface between monomers in the penton base pentamer. Both penton base and fibre residues involved in this interaction are conserved (Zubieta *et al.*, 2005) and we do not observe structural differences in this region. However, we notice a possible difference in the interaction between the start of the fibre shaft and a ring of hydrophobic residues at the centre of the pentamer (Liu *et al.*, 2011; Zubieta *et al.*, 2005). In this region, fibre sequences are not conserved (**Fig. 2d**). In the penton base of HAdV-F41, the hydrophobic residues forming the ring are conserved, except for HAdV-C5 Phe489 which is instead a polar residue (Ser426) in HAdV-F41 (**Fig. 2d-e**). Next to Ser426, Thr427-Thr430 have a large RMSD from the equivalent residues in HAdV-C5 (**Fig. 2c**). This region forms part of a larger stretch (Tyr419- Leu442) which undergoes a conformational rearrangement upon fibre binding (Zubieta *et al.*, 2005). These observations suggest a different fibre-penton binding mode in HAdV-F41, with less hydrophobic and more hydrogen bridging interactions than in HAdV-C5. We also find some differences between HAdV-C5 and HAdV-F41 in the interactions between penton base monomers, and between penton base and the surrounding peripentonal hexons, which are described in **Fig. S4**. Finally, at the pentamer cavity oriented towards the viral core, Arg47 residues in HAdV-C5 form a positively charged ring absent in HAdV-F41, which has Gly45 instead (**Fig. 2f**). This charge variation suggests a different interaction between penton base and the viral genome, or its packaging machinery.

Proteins IIIa and VIII

Minor coat proteins IIIa and VIII line the HAdV-F41 internal capsid surface. Five copies of protein IIIa are located beneath each vertex, bridging the penton and the peripentonal hexons (**Fig. 1b**), and presumably interacting with packaging proteins during assembly (Condezo *et al.*,

2015; Ma and Hearing, 2011). The extent of protein IIIa traced and its architecture are very similar to those of HAdV-C5 (**Tables S3, S5**), with the domain organization previously described (**Fig. 3a-b**) (Liu *et al.*, 2010). HAdV-F41 protein IIIa has an eight amino acid insertion at the N-terminus with predicted α-helical structure (residues 2-9, **Fig. 3c**), which could form extra intramolecular interactions in the peripentonal region and contribute to make the vertex region more stable. However, we did not observe interpretable density for this insertion, or for the appendage domain (APD) traced in HAdV-D26 (Yu *et al.*, 2017) (**Fig. 3c**). The largest difference with HAdV-C5 corresponds to a kink between the 4th and 5th turns in the helix connecting the GOS-glue to the VIII-binding domain **(Fig. 3a-b).** This kink may cause differences in the interactions between neighbouring molecules beneath the vertex (see next section).

Two copies of protein VIII are present in each AU of the AdV icosahedral shell. One of them interacts with protein IIIa and contributes to stabilize the vertex, while the second one helps keep the non-peripentonal hexons together in the central plate of the facet (**Fig. 1b**). HAdV-F41 protein VIII is highly conserved compared to HAdV-C5, both in sequence and structure (**Tables S2-S3 and Fig. 3d**). Only the central region of the protein, cleaved by the maturation protease (AVP) (Mangel and San Martín, 2014), presents sequence divergence (**Fig. 3e**, and see next section).

Proteins VI and VII, and additional internal elements

Analysis of density unoccupied by the molecules traced so far in the HAdV-F41 cryo-EM map (remnant density, RD) revealed several features of interest (**Fig. 4a**). At the rim of the central cavity of the hexon trimers, we observe discontinuous weak density (RD1) that, by similarity with HAdV-C5, corresponds to peptides pVIn and pVIIn$_2$, cleaved from precursor proteins pVI and pVII during maturation (Dai *et al.*, 2017; Hernando-Perez *et al.*, 2020). Since side chains are poorly defined in these regions due to lack of order and low occupancy, we have only

tentatively traced four copies of pVIn and two of $pVIIn_2$, out of the twelve possible equivalent sites in the AU (**Fig. 1b, Fig. 4b and Table S5**) (Hernando-Perez *et al.*, 2020). Protein VI is longer in HAdV-F41 than in HAdV-C5 (**Table S2**) and bears a particular epitope common to enteric AdVs (species A and F) in its central domain (residues 114-125) (Grydsuk *et al.*, 1996). However, we do not observe any interpretable density that could inform about either extra interactions between the longer VI chain and hexon, or the central region traced in HAdV-C5 (Dai *et al.*, 2017).

L-shaped densities at the icosahedral 3-fold (I3) axis and the local 3-fold (L3) axis between hexons 2, 3 and 4 (**Fig. 4a,** RD2) were also observed in HAdV-D26, but their assignment is uncertain (Yu *et al.*, 2017). In HAdV-F41, we observe two additional RDs that were not reported in previous studies (Dai *et al.*, 2017; Yu *et al.*, 2017). RD3 and RD4 are located near the two independent copies of protein VIII and have similar shapes, but RD3 (located beneath the penton region) has stronger density. Both RD3 and RD4 are near the gap in protein VIII left by AVP cleavage during maturation (**Fig. 3d-e**, and **Fig. 4a, c**). In RD3, we could model two poly-Ala peptides. One of them could correspond to a 21 residue α-helix, while the other one is an extended 23 residue peptide (**Fig. 4c, top**). These lengths correlate well with those of the pVIII peptides cleaved by AVP (**Fig. 3e**). In HAdV-F41, one of the cleaved peptides is 6 residues longer than in HAdV-C5 and HAdV-D26, has low sequence conservation (**Fig. 3e**), and secondary structure predictions suggest that, unlike in HAdV-C5, it may have some propensity to form an α-helix, although with low confidence (**Fig. S5**). In RD4, the density is weaker and only a 10 residue extended peptide could be modelled (**Fig. 4c, bottom**). We hypothesize that at least part of RD3 and RD4 correspond to these excised peptides of protein VIII, which in HAdV-F41 would be ordered and reinforcing the network of contacts on the inner capsid surface (particularly beneath the vertex), therefore contributing to a higher capsid stability. RD3 and RD4 also connect to density corresponding to pVIn, and could account for its untraced first four residues (**Table S5** and **Fig. 4c, black dotted lines**). Fragmented density in RD3, which is not present in RD4, seems to connect to the first traced residue in penton base,

Glu35, and runs parallel to the IIIa connecting helix. We propose that this density corresponds to the untraced N-terminus of penton base (**Fig. 4c, pink dotted line**), that would interact with the kinked connecting helix in IIIa, a contact which would not happen in HAdV-C5, where the helix is not kinked (**Fig. 3a)** and the sequence of the penton base N-terminal sequence differs (**Fig. 2d).**

Determining the identity of weak, discontinuous density patches in AdV is problematic, as there are many virion components that do not follow icosahedral symmetry (packaging proteins, protein VI, half of protein IIIa, cleaved peptides, core proteins), making the combinatorial problem of sequence assignment unsolvable (Condezo *et al.*, 2015; Hernando-Perez *et al.*, 2020; San Martín, 2012). Nevertheless, our interpretation suggests that the central peptides of pVIII, the N-terminus of penton base, and the kinked IIIa helix would collaborate to strengthen the HAdV-F41 vertex region in a manner different to that previously observed in other HAdVs.

External cementing network: protein IX

Previous studies (Liu *et al.*, 2010; Yu *et al.*, 2017) have shown that each of the twelve monomers of protein IX in a facet of the AdV capsid presents an extended conformation and is composed of three domains. The N-terminal domains of three IX molecules associate in a triskelion-shaped feature. One of the four triskelions in each facet occupies the valley between hexons at the I3 axis, while the other three lay at the L3 axes formed by hexons 2, 3 and 4 in the AU (**Fig. 1b and Fig. 5a**). The central domain of protein IX is highly flexible, and has also been termed “rope” domain. In HAdV-C5 and D26, the rope domains of three conformationally unique IX molecules, each one originating in a different triskelion, crawl around the hexons on the surface of the icosahedral facet until they reach the edge. There, the C-terminal domains of these three molecules join a fourth one coming from the neighbouring facet to form a coiled coil with three parallel and one anti-parallel α-helices. There are two of these four-helix bundles per facet edge **(Fig. 5a).** In non-human mastadenoviruses, the rope domain is shorter (**Fig. 5b**), and

the C-terminal domains of IX form coiled coils with only three parallel α-helices located directly on top of their N-terminal triskelions (four such bundles per facet), as exemplified in the bovine adenovirus BAdV-3 structure (**Fig. 5a**) (Cheng *et al.*, 2014; Hackenbrack *et al.*, 2016; Reddy, 2017; Schoehn *et al.*, 2008). Of all HAdV-F41 structural proteins, protein IX is the least similar to its HAdV-C5 counterpart (**Table S2**). We find that this protein also displays a conformation different from all previously reported mastadenovirus structures.

We have traced residues Val10 to Gly59 (**Fig. 5c**), which form the triskelion in a manner very similar to that of HAdV-C5**.** As previously reported (Liu *et al.*, 2010), the triskelion is underpinned by a core of hydrophobic residues, Tyr20-Leu21 in HAdV-F41 (**Fig. 5b, d**). The first residues traced fold over the triskelion centre. Notably, this arrangement implies that Phe12 and Phe17 add two tiers of hydrophobic interactions to the triskelion core (**Fig. 5d**). In HAdV-C5, one of the two phenylalanines is absent (Val11, **Fig. 5b**), and the region containing the other (Phe6) could not be traced, implying lack of icosahedral order (**Fig. 5d)**. Both phenylalanine residues (Phe8 and Phe13, **Fig. 5b**) are modelled in HAdV-D26 protein IX, but in a different conformation, oriented outwards from the triskelion core (**Fig. 5d**).

Protein IX in HAdV-F41 has a five-residue insertion at the N-terminus when compared to HAdV-C5 (four-residue compared to HAdV-D26). The N-terminal domain is even shorter in BAdV-3 (**Fig. 5b**). We hypothesize that, in HAdV-F41, the region containing the two phenylalanine residues is more ordered than in HAdV-C5, and in a different conformation from HAdV-D26, due to the presence of these extra residues. Weak density capping the triskelion at the I3 axis could account for the presence of the longer N-terminus (**Fig. 6a-b**). Both the dense network of hydrophobic residues at the triskelion core and the presence of extra residues at the N-terminus of IX would reinforce the intermolecular interactions within the triskelions, likely contributing to stabilize the protein IX trimer. It has previously been shown that the triskelion is sufficient to provide capsid thermostability in HAdV-C5 (Vellinga *et al.*, 2005). More stable triskelions may enhance the physicochemical stability of HAdV-F41.

The flexibility of the rope domain makes it difficult to observe: only one of the four copies per AU has been traced for HAdV-D26 or HAdV-C5 (only partially for the latter) (Dai *et al.*, 2017; Yu *et al.*, 2017). We do not observe density with enough definition to trace any of the rope domains in HAdV-F41. Surprisingly, in our map there is no density corresponding to the C-terminal 4-helix bundle at the icosahedron edges, although secondary structure prediction indicates a helical propensity. Instead, we observe blurry density protruding in a radial orientation between the towers of hexons 2, 3 and 4, on top of the triskelions at the L3 axis (**Fig. 6b** and **S6**). The presence of this protrusion would suggest that in HAdV-F1 the protein IX C-terminal domains could be arranged in a similar way to the shorter protein IX in BAdV-3, forming a three-helix bundle directly adjacent to each triskelion (**Fig. 5a**). However, in our HAdV-F41 map there is no equivalent weak density near the triskelion at the 3-fold icosahedral axis (**Fig. 6b** and **S6**), indicating that the organization of IX is not the same as in BAdV-3. Focused classification did not yield any subset of particles with a protrusion at the I3 axis (**Fig. S7**). Therefore, the arrangement of the C-terminal domain of HAdV-F41 protein IX does not seem to follow either the typical human or the non-human AdV architectures previously observed.

In both human and non-human AdVs with solved structures, heptad repeats are involved in the formation of protein IX C-terminal helix bundles (Cheng *et al.*, 2014; Liu *et al.*, 2010; Yu *et al.*, 2017). While HAdV-C5 and D26 have five complete heptad repeats and BAdV-3 has four, prediction algorithms find only 3 complete repeats in the C-terminal sequence of HAdV-F41 protein IX (**Fig. 6c**). The shorter series of heptad repeats hinders the formation of a stable, ordered helix bundle in HAdV-F41 (**Fig. S8**). An intriguing question is why the helix bundle, even if disordered, would form at the L3 axis, and not at the facet edge as in the other HAdVs. Sequence alignment indicates that the rope domain in HAdV-F41 is 3 residues shorter than in HAdV-C5 and D26 (**Fig. 5b**). This difference in length, together with the shorter heptad repeat region, may be responsible for hindering interaction of the three parallel α-helices originating from one facet with the fourth one, coming in anti-parallel orientation from the adjacent facet.

This consideration would explain why the four-helix bundle is not formed at the facet edges, but the question why there is no density at the centre of the facet that could correspond to a disordered three-helix bundle remains.

A remnant map showing weak density on the capsid surface provides a possible explanation for this unusual arrangement. At low threshold, we observe elongated density connecting one arm of the triskelions at the L3 axes with the triskelion at the centre of each facet (**Fig. 6a-b**). Extra density at the other two arms connects only the end of the modelled region to the centre of the same triskelion (**Fig. 6b**). We interpret that the elongated density corresponds to the rope domain of the central triskelion, which as in HAdV-C5 runs on the capsid surface towards the facet periphery. At the local-3fold triskelions however, lack of partner molecules with which to form a stable helix bundle at the icosahedron edges results in the rope domain turning back on itself. In this situation, the C-terminal regions of the three IX copies at the L3 triskelion would be joined by one copy coming from the facet centre. The flexibility and length of the rope domains, together with the extra N-terminal disordered residues capping all triskelions, and defective heptad repeats, would preclude formation of a stable, ordered helix bundle, producing weak density only at the peripheral triskelions in each facet (**Fig. 6b**).

According to our interpretation, in HAdV-F41 the rope domain would make a counter clockwise turn at the exit of the central triskelion, instead of clockwise as in HAdV-C5 and D26 (**Fig. 6d**). It has been proposed that protein IX in HAdV has two flexible bends that facilitate its coupling to the hexon contours (**Fig. 5b**) (Reddy, 2017). The different path followed by IX in our HAdV-F41 model may be determined by differences in the hexon residues interacting with IX near the first bend. In particular, Tyr784 in hexon 3, located near the end of the triskelion (residues 56-59), is one of the few residues presenting high RMSD when comparing the HAdV-F41 and HAdV-C5 hexon structures (**Fig. 2a** and **Fig. 6d**). HAdV-C5 also has a tyrosine in this position, but the upstream sequence is very different in both viruses, with a conspicuous threonine triplet in HAdV-F41 (residues 781-783); and protein IX has a one amino acid deletion

325 at this bend (**Fig. 5b**). All these changes may result in different interactions causing protein IX

326 to bend in a different direction in both viruses.

Discussion

The narrow enteric tropism of HAdV species F is not well understood. Infectivity analyses have shown that, unlike HAdV-C5, these viruses are not inactivated by acidic environments (Favier *et al.*, 2004), and a main player conferring this resistance to low pH seems to be the short fibre, one of the differentiating structural features of HAdV-F (Rodriguez *et al.*, 2013). Here we show that HAdV-F41 capsids are more thermostable than HAdV-C5, requiring higher temperatures to open up and expose its genome to the solvent; and that infectivity is stable, even activated, in simulated gastric and intestinal conditions. We have solved the structure of the HAdV-F41 capsid at near atomic resolution, and analysed it looking for possible determinants of its enhanced physicochemical stability.

The exact molecular basis of virus capsid stability is difficult to unravel, and is still a subject of intense investigation even for so-called simple viruses (Mateu, 2013). We describe some intermolecular contacts that could have a different nature in HAdV-F41 and HAdV-C5, but it is difficult to conclude if these differences will have an effect in capsid stability, due to the large complexity and sheer number of interactions present in AdV virions. However, we do find partially ordered features that suggest collaboration between the N-terminus of penton base, protein IIIa, pVIn and the cleaved peptides of VIII in stabilizing the vertex region. On the outer capsid surface, the protein IX triskelion has a stronger hydrophobic core than in HAdV-C5, while the rest of the protein is disordered, notably lacking the four-helix bundle at its C-terminus. It has previously been observed that the four-helix bundle in HAdV-C5 is easily disturbed, becoming disordered when protein IX was modified by a C-terminal fusion (Marsh *et al.*, 2006) or antibody labelling (Fabry *et al.*, 2009). It is possible that the lack of well-defined density for the C-terminal region of IX in our HAdV-F41 map is simply caused by partial disruption accidentally occurring during sample preparation. However, in studies where the HAdV-C5 helix bundle was not clearly visible, no weak density was observed in other capsid locations. Differences in sequence and weak densities in the map support the presence of an organization for protein IX in HAdV-F41 different from all described for other AdVs. Our

354 observations also imply that the formation of a well ordered helix bundle is not required to
355 enhance capsid stability, reinforcing the idea that the triskelion is the crucial part for the
356 cementing action of the protein (Vellinga *et al.*, 2005). The more exposed location of the protein
357 IX C-terminus in HAdV-F41 makes it an interesting locale for exogenous peptide insertion for
358 retargeting, epitope display or other biotechnological purposes (Matteson *et al.*, 2018).

359

Methods

Virus production

E1-deleted HAdV-F41-EGFP virus (kindly provided by Douglas Brough, GenVec/Precigen) was propagated in 2V6.11 cells (ATCC JHU67) with E4 orf6 protein induced with 1μg/ml ponasterone A 24 hours prior to infection. Cells were infected at an input multiplicity of infection (MOI) <0.1 fluorescent units (FU) per cell, and cultures were incubated until cytopathic effect (CPE) was complete or almost complete but before cells showed signs of disintegration. Flasks were hit sharply to dislodge infected cells from the surface and cells were harvested by centrifugation at 1600 x *g* for 10 minutes, washed with PBS, then re-suspended in a small volume of serum-free medium. Any cells remaining attached to the flask were re-fed with a 1:1 mixture of fresh medium and clarified medium from the infected flasks. If uninfected cells had become confluent, they were sub-cultured to allow for virus spread. The process was continued until CPE was complete, typically about 1-2 weeks post infection (p.i.). Progeny virus was released by five cycles of freezing-thawing and the cell lysate was clarified by centrifugation at 1600 x *g* for 10 minutes. Viruses were used in experiments at passage level 3.

Virus was purified by two cycles of caesium chloride (CsCl) density gradient centrifugation. Clarified cell lysate (~4.5ml) was underlaid with 4 ml CsCl (1.2 g/ml in 50mM Tris-HCl, pH 8.1) and 2 ml CsCl (1.4g/ml in 50mM Tris-HCl, pH 8.1). The gradient was centrifuged in a SW41 rotor at 120 000 x g at 4°C for 1 hour. Virions were collected from the interface of each step gradient and diluted in 50mM Tris-HCl to a density <1.2g/ml, as determined by measuring refractive index, and layered on top of a preformed gradient (1.2-1.4 g/ml) that was prepared using the Gradient Master (BioComp). The gradient was centrifuged in a SW41 rotor at 120,000 x *g* at 4°C for 2 hours. The viral band of mature virus particles was recovered and dialyzed at 4°C against 3 changes of storage buffer (50mM Tris-HCl, pH 8.1, 150mM NaCl, 10mM $MgCl_2$ and 10% glycerol) with a buffer change every 45 minutes. Optical density was measured with a NanoDrop® ND-1000 Spectrophotometer (v3.3) and virion concentration was calculated using

386 the equation: $1A_{260} = 1.1 \times 10^{12}$ particles/ml (Maizel *et al.*, 1968). Infectivity was determined by
387 endpoint dilution assays with HEK 293 cells, in duplicate, in 60-well Terasaki plates as
388 previously described (Brown, 1985). Assays were scored for the presence of green cells
389 (expressing EGFP), and titres, calculated by the Reed and Muench method (Reed and Muench,
390 1938), were expressed as fluorescent units (FU)/ml. Purified HAdV-F41 particles from two
391 different preparations were pooled and concentrated by centrifugation in a CsCl gradient and
392 stored in 20mM HEPES pH 7.8, 150 mM NaCl, 10% glycerol at -80ºC.

393 HAdV-C5 used as a control for extrinsic fluorescence experiments was the E1-E3 deleted, GFP
394 expressing Ad5/attP vector which is structurally wildtype (Alba *et al.*, 2007), propagated at
395 37ºC in HEK293 cells, harvested at 36 h post-infection and purified as described (Hernando-
396 Perez *et al.*, 2020).

397

398 *Infectivity assays under simulated gastrointestinal conditions*

399 HEK293 cells were seeded in 6-well plates at a density of $5x10^5$ cells/well one day prior to
400 infection. HAdV-41-EGFP preparations (~2×10^{11} particles/ml, 5×10^6 FU/ml) were dialyzed
401 against serum-free medium and frozen for a short period. Virus (50 µl) was then mixed 1:1 with
402 2x synthetic gastric fluid (SG) according to the US Pharmacopoeia at 0.01N HCl (4.0 g NaCl
403 and 6.4 g pepsin (Sigma) dissolved in 14 ml of 1N HCl, made up to 1 liter with ddH_2O, pH 1.2)
404 and/or 2x synthetic intestinal fluid (SI -13.6 g of monobasic potassium phosphate in 250 ml
405 ddH_2O, 77 ml 0.4N NaOH, 500 ml ddH_2O, then 20.0g pancreatin (Sigma), adjusted to pH 6.8,
406 then diluted with water to 1 liter). Virus suspension was incubated for 1 hour at 37°C. For
407 sequential treatments, virus was treated with synthetic gastric fluid for 1 hour at 37°C, then
408 mixed 1:1 with synthetic intestinal fluid and incubated for an additional hour. To neutralize pH
409 and proteases, complete MEM containing 10% FBS was added to final volume of 0.3 ml.
410 HEK293 cells were infected with treated or mock treated virus for 1 hour at 37°C. Inoculum
411 was removed, cells were harvested 19-21 hours post infection, and fixed with 2% PFA for flow

412 cytometry. GFP expression was analysed with an LSR Fortessa (BD Biosciences) and data was
413 analysed using FACSDiva (BD Biosciences) gating out doublets. 10 000 cells were counted per
414 well. Each experiment was done in duplicate.

415

416 *Extrinsic fluorescence thermostability assays*

417 The data for HAdV-C5 have been reported previously (Hernando-Perez *et al.*, 2020). The data
418 for HAdV-F41 were collected in the same time period, as part of a large study comparing the
419 thermostability of different AdV specimens. HAdV-C5 and F41 samples ($5x10^9$ virus particles
420 in a final volume of 800 µl of buffer) were incubated overnight at 4 °C in 8 mM Na_2HPO_4, 2
421 mM KH_2PO_4, 150 mM NaCl, and 0.1 mM EDTA, pH 7.4. Fluorescence emission spectra from
422 virus samples mixed with YOYO-1 were obtained every 2 ºC along a temperature range from
423 20 to 70 ºC in a Horiba FluoroLog spectrophotometer. The dye was excited at 490 nm and
424 maximal emission intensity was achieved at 509 nm. Raw spectra were corrected by subtraction
425 of the YOYO-1 spectrum in buffer at each tested temperature. A total of six (HAdV-C5) and
426 four (HAdV-F41) independent experiments were considered for each averaged curve. The half
427 transition temperatures ($T_{0.5}$) were calculated from the fitting of the averaged fluorescence
428 intensity curve as a function of temperature to a Boltzmann sigmoid equation as described
429 (Hernando-Perez *et al.*, 2020).

430

431 *Cryo-EM sample preparation and data collection*

432 Pooled and concentrated HAdV-F41 samples (as described above) were dialyzed against 50
433 mM Tris-HCl pH 7.8, 150 mM NaCl, 10mM $MgCl_2$ for 1 hour at 4ºC and further concentrated
434 by spinning in a Microcon YM-100 device for 15 min at 4ºC, for a final estimated concentration
435 of $3.1x10^{12}$ vp/ml. Samples were deposited in glow discharged, Quantifoil R2/4 300 mesh
436 Cu/Rh grids and vitrified in liquid ethane after manual blotting in a Leica CPC device. Cryo-

EM images (**Table S1**) were recorded at the CEITEC facility (Brno, Czech Republic) using a 300 kV Titan Krios microscope equipped with a Falcon II detector, with a total dose of 42 e-/Å^2 distributed over 25 frames, at nominal pixel size 1.38 Å and defocus range between -1 and -2.5 µm.

Cryo-EM data analysis

All image processing and 3D reconstruction tasks were performed within the Scipion framework (de la Rosa-Trevin *et al.*, 2016). Frames 2-24 of each movie were aligned using whole-image motion correction implemented in Xmipp, followed by correction of local movements using Optical Flow (Abrishami *et al.*, 2015). The contrast transfer function (CTF) was estimated using CTFFIND4 (Rohou and Grigorieff, 2015). Particles were semi-automatically picked from micrographs corrected for the phase oscillations of the CTF (phase-flipped), extracted into 780×780 pixel boxes, normalized and downsampled by a factor of 2, using Xmipp (de la Rosa-Trevin *et al.*, 2013). All 2D and 3D classifications and refinements were performed using RELION (Scheres, 2012). 2D classification was used to discard low quality particles, and run for 25 iterations, with 50 classes, angular sampling 5 and regularization parameter T = 2. Classification in 3D was run for 40 iterations, with 3 classes, starting with an angular sampling of 3.7 degrees and sequentially decreasing to 0.5, and regularization parameter T = 4. Icosahedral symmetry was imposed throughout the refinement process. The initial reference for 3D classification was a non-human adenovirus cryo-EM map (unpublished), low-pass filtered to 60 Å resolution. The class yielding the best resolution was individually refined using the original 780 px boxed particles and the map obtained during the 3D classification as a reference, producing a final map at 4.0 Å resolution, as estimated according to the gold-standard FSC = 0.143 criterion implemented in RELION auto-refine and postprocess routines (Chen *et al.*, 2013). A global B-factor was estimated after dividing the map Fourier transform by the modulation transfer function (MTF) of the Falcon II detector. The

actual sampling for the map was estimated by comparison with a HAdV-C5 model (PDB ID 6B1T) (Dai *et al.*, 2017) in UCSF Chimera (Pettersen *et al.*, 2004), yielding a value of 1.36 Å/px.

Focused classification was carried out following the general procedures described in (Ilca *et al.*, 2015), using RELION as implemented in Scipion (Abrishami *et al.*, 2020). First, a spherical mask excluding the region of interest was applied to the 3D map obtained from the refinement with icosahedral symmetry enforced. The masked map was projected in the directions determined from the icosahedral refinement, and the projections subtracted from the corresponding experimental images. Then, the regions containing the feature of interest were located by projecting their position in the 3D map, extracted from the subtracted images in small boxes (100x100 pixels) and subjected to 3D classification without orientation change or symmetry enforcement.

Model building and analysis

Interpretation of the HAdV-F41 3D map was performed using the molecular modelling workflow based on sequence homology in Scipion (Martinez *et al.*, 2020). To facilitate chain tracing, we used MonoRes (Vilas *et al.*, 2018) and Localdeblur (Ramirez-Aportela *et al.*, 2020) to sharpen the map according to the local resolution. The initial model for each polypeptide chain was predicted with Modeller (Webb and Sali, 2016), using as input template the structure of the respective HAdV-C5 homolog chain (PDB ID 6B1T). UCSF Chimera (Pettersen *et al.*, 2004) was used to perform a rigid fitting of each chain initial model into the sharpened map. Next, the fitted model of each chain was refined using Coot (Emsley *et al.*, 2010) and Phenix *real space refine* (Afonine *et al.*, 2018). Validation metrics to assess the quality of the atomic structure were computed with the Phenix *comprehensive validation (cryo-EM)* algorithm. Once we generated the whole structure of the AU, Chimera *findclash* (integrated in the Scipion protocol *chimera-contacts*) was executed to identify possible contacts between pairs of chains

contained in the AU or between a chain in the AU and a chain from a neighbouring AU (**Supplementary file S1**). The default parameters, cut-off (-0.4) and allowance (0.0), were used to report as possible bonds all pair of atoms whose distance is no higher than the sum of the corresponding van der Waals radii plus 0.4 Å. To identify additional unmodeled densities, we generated remnant maps with Chimera in two alternative manners: (1) subtracting from the cryo-EM map a map generated with *molmap* around the modelled atoms (resolution 2.0 gridSpacing 1.36); or (2) masking off from the initial map a region within a 3, 4 or 8 Å radius from the modelled atoms (*vol zone invert true*).

Protein sequence alignments were performed with Clustal W2 (http://www.clustal.org/clustal2/) integrated in Geneious version 2019.0 created by Biomatters (https://www.geneious.com) (Larkin *et al.*, 2007). For secondary structure prediction we used Jpred4 Server Jnet version 2.3.1 (http://www.compbio.dundee.ac.uk/jpred) (Drozdetskiy *et al.*, 2015). The coiled-coil prediction programs Multicoil and Marcoil implemented in Waggawagga server (https://waggawagga.motorprotein.de/) were used to predict heptad repeats and the probability to form coiled coils (Delorenzi and Speed, 2002; Simm *et al.*, 2015; Wolf *et al.*, 1997). Molecular graphics and analyses including molecule superposition, RMSD calculation, map rendering, map normalization, coloring and measuring surface blobs and surface electrostatic potential rendering were performed with UCSF Chimera and ChimeraX (Goddard *et al.*, 2018; Pettersen *et al.*, 2004). UCSF Chimera *Hide dust* was used for clarity when composing figures representing density maps.

Database deposition

The HAdV-F41 cryo-EM map and model are deposited at the Electron Microscopy Data Bank (EMDB, http://www.ebi.ac.uk/pdbe/emdb) and the Protein Data Bank (PDB, http://www.ebi.uk/pdbe) with accession numbers EMD-10768 and 6YBA, respectively. The validation report is included as **Supplementary file S2**.

References

Abrishami, V., Ilca, S.L., Gomez-Blanco, J., Rissanen, I., Miguel de la Rosa-Trevín, J., Reddy, V.S., Carazo, J.-M. and Huiskonen, J.T. (2020) Localized reconstruction in scipion expedites the analysis of symmetry mismatches in Cryo-EM data. *Progress in Biophysics and Molecular Biology*.

Abrishami, V., Vargas, J., Li, X., Cheng, Y., Marabini, R., Sorzano, C.O. and Carazo, J.M. (2015) Alignment of direct detection device micrographs using a robust Optical Flow approach. *J Struct Biol*, **189**, 163-176.

Afonine, P.V., Poon, B.K., Read, R.J., Sobolev, O.V., Terwilliger, T.C., Urzhumtsev, A. and Adams, P.D. (2018) Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallographica Section D*, **74**, 531-544.

Alba, R., Hearing, P., Bosch, A. and Chillon, M. (2007) Differential amplification of adenovirus vectors by flanking the packaging signal with attB/attP-PhiC31 sequences: implications for helper-dependent adenovirus production. *Virology*, **367**, 51-58.

Albinsson, B. and Kidd, A.H. (1999) Adenovirus type 41 lacks an RGD alpha(v)-integrin binding motif on the penton base and undergoes delayed uptake in A549 cells. *Virus Res*, **64**, 125-136.

Appaiahgari, M.B. and Vrati, S. (2015) Adenoviruses as gene/vaccine delivery vectors: promises and pitfalls. *Expert Opin Biol Ther*, **15**, 337-351.

Baker, A.T., Aguirre-Hernandez, C., Hallden, G. and Parker, A.L. (2018) Designer Oncolytic Adenovirus: Coming of Age. *Cancers (Basel)*, **10**.

Brandt, C.D., Kim, H.W., Rodriguez, W.J., Arrobio, J.O., Jeffries, B.C., Stallings, E.P., Lewis, C., Miles, A.J., Gardner, M.K. and Parrott, R.H. (1985) Adenoviruses and pediatric gastroenteritis. *J Infect Dis*, **151**, 437-443.

Brown, M. (1985) Selection of nonfastidious adenovirus species in 293 cells inoculated with stool specimens containing adenovirus 40. *Journal of Clinical Microbiology*, **22**, 205-209.

Brown, M. (1990) Laboratory identification of adenoviruses associated with gastroenteritis in Canada from 1983 to 1986. *J Clin Microbiol*, **28**, 1525-1529.

Chen, S., McMullan, G., Faruqi, A.R., Murshudov, G.N., Short, J.M., Scheres, S.H. and Henderson, R. (2013) High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy*, **135**, 24-35.

Cheng, L., Huang, X., Li, X., Xiong, W., Sun, W., Yang, C., Zhang, K., Wang, Y., Liu, H., Huang, X., Ji, G., Sun, F., Zheng, C. and Zhu, P. (2014) Cryo-EM structures of two bovine adenovirus type 3 intermediates. *Virology*, **450-451**, 174-181.

Chhabra, P., Payne, D.C., Szilagyi, P.G., Edwards, K.M., Staat, M.A., Shirley, S.H., Wikswo, M., Nix, W.A., Lu, X., Parashar, U.D. and Vinje, J. (2013) Etiology of viral gastroenteritis in children <5 years of age in the United States, 2008-2009. *J Infect Dis*, **208**, 790-800.

Condezo, G.N., Marabini, R., Ayora, S., Carazo, J.M., Alba, R., Chillón, M. and San Martín, C. (2015) Structures of Adenovirus Incomplete Particles Clarify Capsid Architecture and Show Maturation Changes of Packaging Protein L1 52/55k. *J Virol*, **89**, 9653-9664.

Corcoran, M.S., van Well, G.T. and van Loo, I.H. (2014) Diagnosis of viral gastroenteritis in children: interpretation of real-time PCR results and relation to clinical symptoms. *Eur J Clin Microbiol Infect Dis*, **33**, 1663-1673.

Crawford-Miksza, L. and Schnurr, D.P. (1996) Analysis of 15 adenovirus hexon proteins reveals the location and structure of seven hypervariable regions containing serotype-specific residues. *J Virol*, **70**, 1836-1844.

Dai, X., Wu, L., Sun, R. and Zhou, Z.H. (2017) Atomic Structures of Minor Proteins VI and VII in Human Adenovirus. *J Virol*, **91**.

567 de la Rosa-Trevin, J.M., Oton, J., Marabini, R., Zaldivar, A., Vargas, J., Carazo, J.M. and
568 Sorzano, C.O. (2013) Xmipp 3.0: an improved software suite for image processing in
569 electron microscopy. *J Struct Biol*, **184**, 321-328.
570 de la Rosa-Trevin, J.M., Quintana, A., Del Cano, L., Zaldivar, A., Foche, I., Gutierrez, J.,
571 Gomez-Blanco, J., Burguet-Castell, J., Cuenca-Alba, J., Abrishami, V., Vargas, J.,
572 Oton, J., Sharov, G., Vilas, J.L., Navas, J., Conesa, P., Kazemi, M., Marabini, R.,
573 Sorzano, C.O. and Carazo, J.M. (2016) Scipion: A software framework toward
574 integration, reproducibility and validation in 3D electron microscopy. *J Struct Biol*, **195**,
575 93-99.
576 Delorenzi, M. and Speed, T. (2002) An HMM model for coiled-coil domains and a comparison
577 with PSSM-based predictions. *Bioinformatics*, **18**, 617-625.
578 Drozdetskiy, A., Cole, C., Procter, J. and Barton, G.J. (2015) JPred4: a protein secondary
579 structure prediction server. *Nucleic Acids Res*, **43**, W389-394.
580 Ebner, K., Pinsker, W. and Lion, T. (2005) Comparative sequence analysis of the hexon gene in
581 the entire spectrum of human adenovirus serotypes: phylogenetic, taxonomic, and
582 clinical implications. *J Virol*, **79**, 12635-12642.
583 Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of
584 Coot. *Acta Crystallographica Section D*, **66**, 486-501.
585 Fabry, C.M., Rosa-Calatrava, M., Moriscot, C., Ruigrok, R.W., Boulanger, P. and Schoehn, G.
586 (2009) The C-terminal domains of adenovirus serotype 5 protein IX assemble into an
587 antiparallel structure on the facets of the capsid. *J Virol*, **83**, 1135-1139.
588 Favier, A.-L., Schoehn, G., Jaquinod, M., Harsi, C. and Chroboczek, J. (2002) Structural studies
589 of human enteric adenovirus type 41. *Virology*, **293**, 75-85.
590 Favier, A.L., Burmeister, W.P. and Chroboczek, J. (2004) Unique physicochemical properties
591 of human enteric Ad41 responsible for its survival and replication in the gastrointestinal
592 tract. *Virology*, **322**, 93-104.

593 Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H. and Ferrin,
594 T.E. (2018) UCSF ChimeraX: Meeting modern challenges in visualization and analysis.
595 *Protein Sci*, **27**, 14-25.
596 Grydsuk, J.D., Fortsas, E., Petric, M. and Brown, M. (1996) Common epitope on protein VI of
597 enteric adenoviruses from subgenera A and F. *J Gen Virol*, **77 (Pt 8)**, 1811-1819.
598 Hackenbrack, N., Rogers, M.B., Ashley, R.E., Keel, M.K., Kubiski, S.V., Bryan, J.A., Ghedin,
599 E., Holmes, E.C., Hafenstein, S.L. and Allison, A.B. (2016) Evolution and cryo-EM
600 capsid structure of a North American bat adenovirus and its relationship to other
601 mastadenoviruses. *Journal of Virology*.
602 Hernando-Perez, M., Martin-Gonzalez, N., Perez-Illana, M., Suomalainen, M., Condezo, G.N.,
603 Ostapchuk, P., Gallardo, J., Menendez, M., Greber, U.F., Hearing, P., de Pablo, P.J. and
604 San Martin, C. (2020) Dynamic competition for hexon binding between core protein
605 VII and lytic protein VI promotes adenovirus maturation and entry. *Proc Natl Acad Sci*
606 *U S A*, 201920896.
607 Ilca, S.L., Kotecha, A., Sun, X., Poranen, M.M., Stuart, D.I. and Huiskonen, J.T. (2015)
608 Localized reconstruction of subunits from electron cryomicroscopy images of
609 macromolecular complexes. *Nat Commun*, **6**, 8843.
610 Kidd, A.H., Chroboczek, J., Cusack, S. and Ruigrok, R.W.H. (1993) Adenovirus type 40 virions
611 contain two distinct fibers. *Virology*, **192**, 73-84.
612 Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H.,
613 Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and
614 Higgins, D.G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947-
615 2948.
616 Lion, T. (2014) Adenovirus Infections in Immunocompetent and Immunocompromised Patients.
617 *Clin Microbiol Rev*, **27**, 441-462.
618 Liu, H., Jin, L., Koh, S.B., Atanasov, I., Schein, S., Wu, L. and Zhou, Z.H. (2010) Atomic
619 structure of human adenovirus by cryo-EM reveals interactions among protein
620 networks. *Science*, **329**, 1038-1043.

621 Liu, H., Wu, L. and Zhou, Z.H. (2011) Model of the trimeric fiber and its interactions with the
622 pentameric penton base of human adenovirus by cryo-electron microscopy. *J Mol Biol*,
623 **406**, 764-774.
624 Ma, H.C. and Hearing, P. (2011) Adenovirus structural protein IIIa is involved in the serotype
625 specificity of viral DNA packaging. *J Virol*, **85**, 7849-7855.
626 Maizel, J.V., Jr., White, D.O. and Scharff, M.D. (1968) The polypeptides of adenovirus. I.
627 Evidence for multiple protein components in the virion and a comparison of types 2,
628 7A, and 12. *Virology*, **36**, 115-125.
629 Mangel, W.F. and San Martín, C. (2014) Structure, function and dynamics in adenovirus
630 maturation. *Viruses*, **6**, 4536-4570.
631 Marsh, M.P., Campos, S.K., Baker, M.L., Chen, C.Y., Chiu, W. and Barry, M.A. (2006)
632 CryoEM of Protein IX-Modified Adenoviruses Suggests a New Position for the C-
633 terminus of Protein IX. *J Virol*, **80**, 11881-11886.
634 Martinez, M., Jimenez-Moreno, A., Maluenda, D., Ramirez-Aportela, E., Melero, R., Cuervo,
635 A., Conesa, P., Del Cano, L., Fonseca, Y.C., Sanchez-Garcia, R., Strelak, D., Conesa,
636 J.J., Fernandez-Gimenez, E., de Isidro, F., Sorzano, C.O.S., Carazo, J.M. and Marabini,
637 R. (2020) Integration of Cryo-EM Model Building Software in Scipion. *J Chem Inf*
638 *Model*.
639 Mateu, M.G. (2013) Assembly, stability and dynamics of virus capsids. *Archives of*
640 *Biochemistry and Biophysics*, **531**, 65-79.
641 Matteson, N.L., Barry, M.A. and Reddy, V.S. (2018) Structure-based assessment of protein-
642 protein interactions and accessibility of protein IX in adenoviruses with implications for
643 antigen display. *Virology*, **516**, 102-107.
644 Nicklin, S.A., Wu, E., Nemerow, G.R. and Baker, A.H. (2005) The influence of adenovirus
645 fiber structure and function on vector development for gene therapy. *Mol Ther*, **12**, 384-
646 393.

647 Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and
648 Ferrin, T.E. (2004) UCSF Chimera--a visualization system for exploratory research and
649 analysis. *J Comput Chem*, **25**, 1605-1612.
650 Rajan, A., Persson, B.D., Frangsmyr, L., Olofsson, A., Sandblad, L., Heino, J., Takada, Y.,
651 Mould, A.P., Schnapp, L.M., Gall, J. and Arnberg, N. (2018) Enteric Species F Human
652 Adenoviruses use Laminin-Binding Integrins as Co-Receptors for Infection of Ht-29
653 Cells. *Sci Rep*, **8**, 10019.
654 Ramirez-Aportela, E., Vilas, J.L., Glukhova, A., Melero, R., Conesa, P., Martinez, M.,
655 Maluenda, D., Mota, J., Jimenez, A., Vargas, J., Marabini, R., Sexton, P.M., Carazo,
656 J.M. and Sorzano, C.O.S. (2020) Automatic local resolution-based sharpening of cryo-
657 EM maps. *Bioinformatics*, **36**, 765-772.
658 Reddy, V.S. (2017) The Role of Hexon Protein as a Molecular Mold in Patterning the Protein
659 IX Organization in Human Adenoviruses. *J Mol Biol*, **429**, 2747-2751.
660 Reed, L.J. and Muench, H. (1938) A Simple Method of Estimating Fifty Per Cent Endpoints12.
661 *American Journal of Epidemiology*, **27**, 493-497.
662 Robinson, C.M., Zhou, X., Rajaiya, J., Yousuf, M.A., Singh, G., DeSerres, J.J., Walsh, M.P.,
663 Wong, S., Seto, D., Dyer, D.W., Chodosh, J. and Jones, M.S. (2013) Predicting the next
664 eye pathogen: analysis of a novel adenovirus. *mBio*, **4**, e00595-00512.
665 Rodriguez, E., Romero, C., Rio, A., Miralles, M., Raventos, A., Planells, L., Burgueno, J.F.,
666 Hamada, H., Perales, J.C., Bosch, A., Gassull, M.A., Fernandez, E. and Chillon, M.
667 (2013) Short-fiber protein of ad40 confers enteric tropism and protection against acidic
668 gastrointestinal conditions. *Hum Gene Ther Methods*, **24**, 195-204.
669 Rohou, A. and Grigorieff, N. (2015) CTFFIND4: Fast and accurate defocus estimation from
670 electron micrographs. *J Struct Biol*, **192**, 216-221.
671 San Martín, C. (2012) Latest Insights on Adenovirus Structure and Assembly. *Viruses*, **4**, 847-
672 877.
673 Scheres, S.H. (2012) RELION: implementation of a Bayesian approach to cryo-EM structure
674 determination. *J Struct Biol*, **180**, 519-530.

675 Schoehn, G., El Bakkouri, M., Fabry, C.M., Billet, O., Estrozi, L.F., Le, L., Curiel, D.T.,
676 Kajava, A.V., Ruigrok, R.W. and Kremer, E.J. (2008) Three-dimensional structure of
677 canine adenovirus serotype 2 capsid. *J Virol*, **82**, 3192-3203.
678 Simm, D., Hatje, K. and Kollmar, M. (2015) Waggawagga: comparative visualization of coiled-
679 coil predictions and detection of stable single alpha-helices (SAH domains).
680 *Bioinformatics*, **31**, 767-769.
681 Uhnoo, I., Wadell, G., Svensson, L. and Johansson, M.E. (1984) Importance of enteric
682 adenoviruses 40 and 41 in acute gastroenteritis in infants and young children. *J Clin*
683 *Microbiol*, **20**, 365-372.
684 Vellinga, J., van den Wollenberg, D.J., van der Heijdt, S., Rabelink, M.J. and Hoeben, R.C.
685 (2005) The coiled-coil domain of the adenovirus type 5 protein IX is dispensable for
686 capsid incorporation and thermostability. *J Virol*, **79**, 3206-3210.
687 Vilas, J.L., Gomez-Blanco, J., Conesa, P., Melero, R., Miguel de la Rosa-Trevin, J., Oton, J.,
688 Cuenca, J., Marabini, R., Carazo, J.M., Vargas, J. and Sorzano, C.O.S. (2018)
689 MonoRes: Automatic and Accurate Estimation of Local Resolution for Electron
690 Microscopy Maps. *Structure*, **26**, 337-344 e334.
691 Webb, B. and Sali, A. (2016) Comparative Protein Structure Modeling Using MODELLER.
692 *Curr Protoc Protein Sci*, **86**, 2 9 1-2 9 37.
693 Wolf, E., Kim, P.S. and Berger, B. (1997) MultiCoil: a program for predicting two- and three-
694 stranded coiled coils. *Protein Sci*, **6**, 1179-1189.
695 Yu, X., Veesler, D., Campbell, M.G., Barry, M.E., Asturias, F.J., Barry, M.A. and Reddy, V.S.
696 (2017) Cryo-EM structure of human adenovirus D26 reveals the conservation of
697 structural organization among human adenoviruses. *Sci Adv*, **3**, e1602670.
698 Zubieta, C., Schoehn, G., Chroboczek, J. and Cusack, S. (2005) The structure of the human
699 adenovirus 2 penton. *Mol Cell*, **17**, 121-135.

700

701

Acknowledgements

This work was supported by grants PID2019-104098GB-I00 and BFU2016-74868-P, co-funded by the Spanish State Research Agency and the European Regional Development Fund, as well as BFU2013-41249-P and BIO2015-68990-REDT (the Spanish Adenovirus Network, AdenoNet) from the Spanish Ministry of Economy, Industry and Competitiveness to C.S.M. Work in M.B's. lab was supported by grant 194562-08 from the Natural Sciences and Engineering Research Council of Canada. M.H.-P. is a recipient of a Juan de la Cierva postdoctoral contract funded by the Spanish State Research Agency. M.P.-I. holds a predoctoral contract from La Caixa Foundation (ID 100010434), under agreement LCF/BQ/SO16/52270032).

We acknowledge Jiri Novacek and Miroslav Peterek at the CEITEC facility (Brno, Czech Republic) for cryo-EM data collection. Access to CEITEC was supported by iNEXT, project number 653706, funded by the Horizon 2020 programme of the European Union. CEITEC Cryo-electron Microscopy and Tomography core facility is supported by MEYS CR (LM2018127). The initial seed for HAdV-F41-EGFP was a gift from Douglas Brough (GenVec/Precigen). We thank Mark J. van Raaij (CNB-CSIC) for careful reading and insightful comments on the manuscript, Margarita Menéndez (IQFR-CSIC), Milagros Castellanos and Luis A. Campos (CNB-CSIC) for advice with fluorescence measurements and analyses, Mara I. Laguna (CNB-CSIC) for expert technical help and José M. Carazo (CNB-CSIC) for constant support.

Author contributions

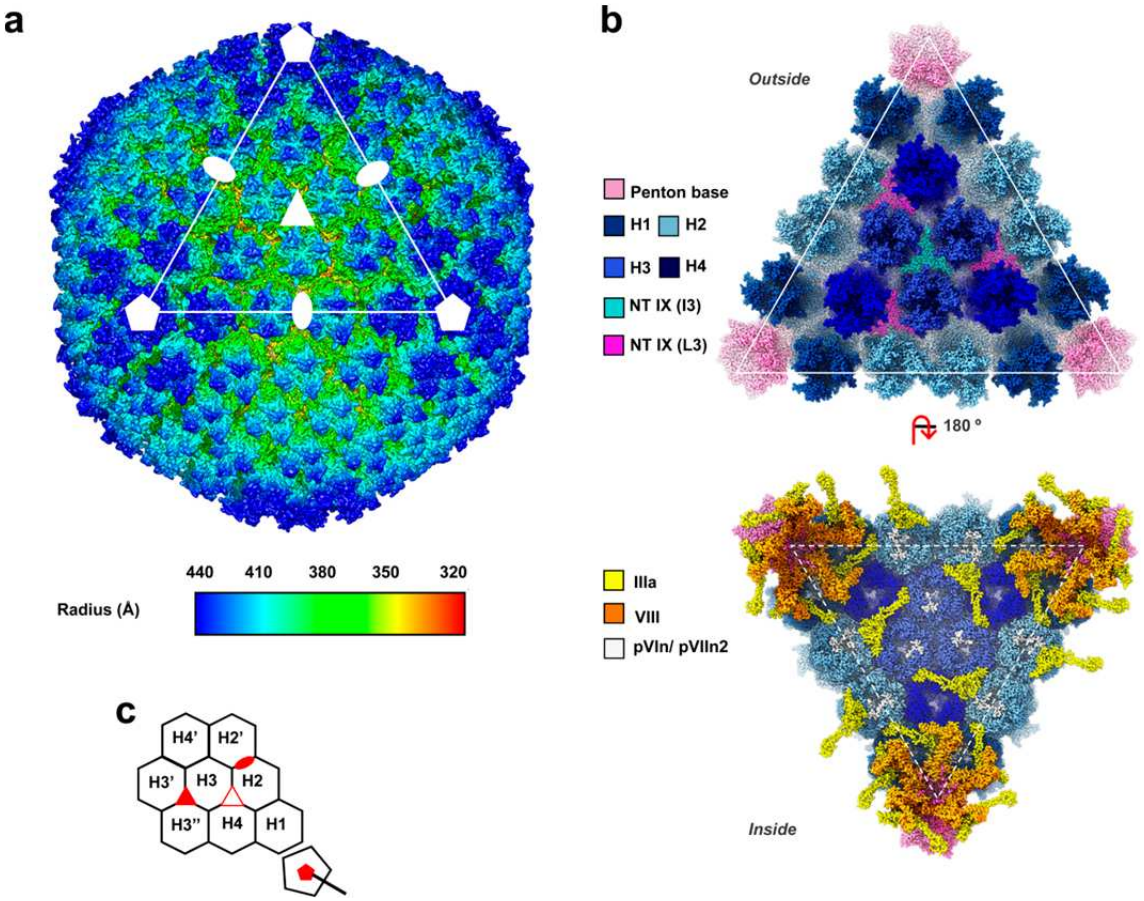
M.B., R.M. and C.S.M. designed research. M. P. -I., M. M., G.N.C, M.H.-P., C.M., M.B., R.M. and C.S.M. performed research. M. P. -I., M. M., M.H.-P., C.M., M.B., R.M. and C.S.M. analysed data. M.M. and R.M. contributed analysis tools. M.P.-I. and C.S.M. wrote the paper, with input from all other authors.

730 **Competing interests statement**

731 The authors declare no competing interests.

732

733 **Figures**

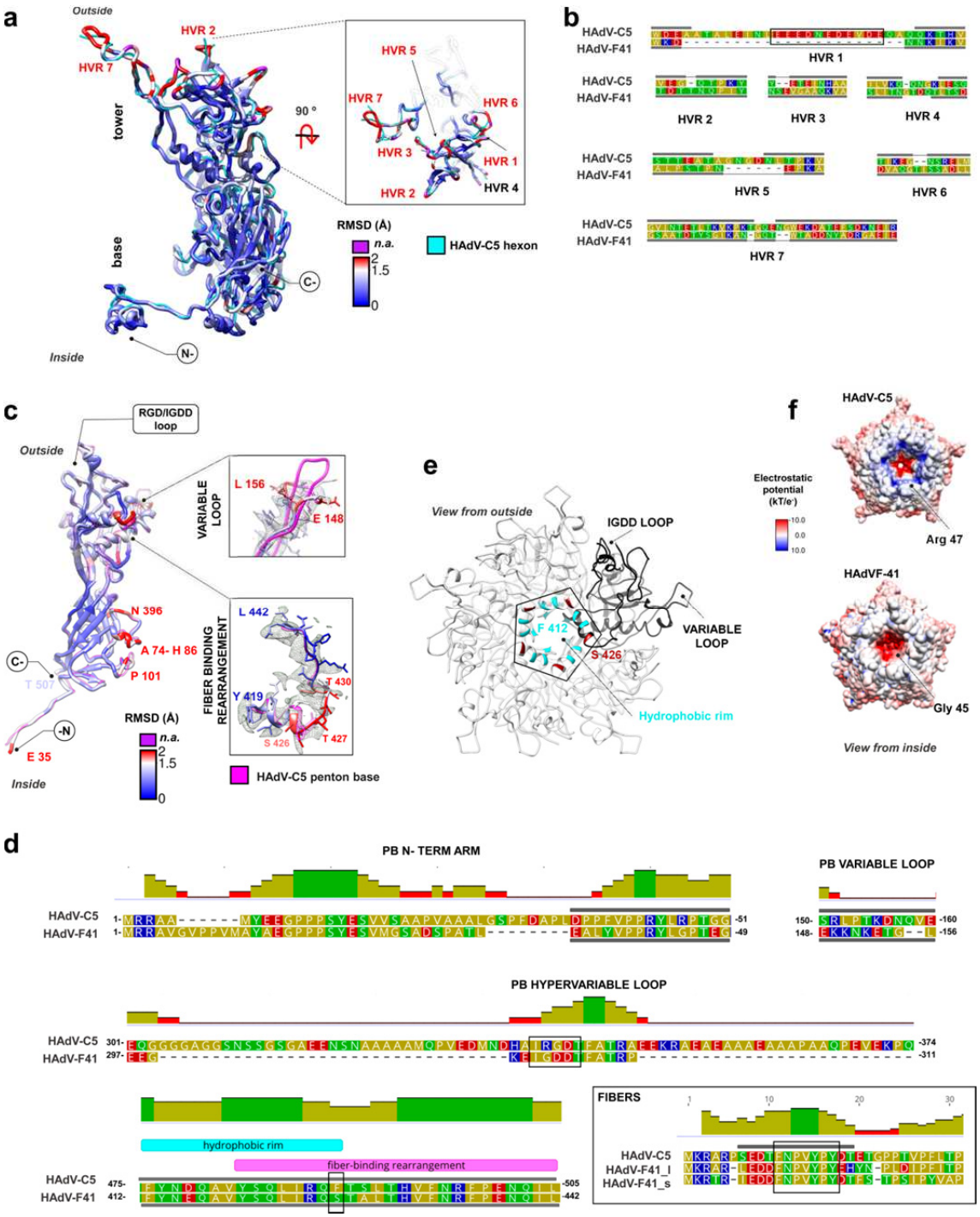


734

735 **Figure 1. The cryo-EM structure of HAdV-F41 at 4.0 Å.**

736 **(a)** Radially coloured surface rendering of the HAdV-F41 map viewed down the 2-fold
737 icosahedral axis. One facet is identified by a white triangle. Icosahedral symmetry axes are
738 indicated with white symbols: 5-fold (pentagon), 3-fold (triangle) and 2-fold (oval). A Gaussian
739 filter has been applied to the map for a clearer appreciation of the main features.

740 **(b)** The molecular models of the major and minor coat proteins traced in the icosahedral HAdV-
741 F41 facet. Top: view from outside the capsid. Bottom: view inside the capsid. Protein colours as
742 indicated by the legend at the left hand side. The four hexon trimers in the AU are numbered
743 H1-H4. **(c)** Cartoon depicting the icosahedral AU and surrounding hexons, as seen from outside
744 the capsid. The icosahedral symmetry axes are indicated with red filled symbols. A hollow red
745 triangle indicates the L3 axis between hexons 2, 3 and 4.



746

747 **Figure 2. Hexon and penton base structures.**

748 **(a)** Superposition of the HAdV-C5 hexon monomer (PDB ID: 6B1T Chain A, cyan) with that of
749 HAdV-F41, coloured by RMSD from blue to red as indicated. Residues exceeding the Chimera
750 RMSD cut-off, or where RMSD calculation is not possible because they are not traced in
751 HAdV-C5, are depicted in purple. The six (out of seven) HVRs traced in the enteric virus are
752 labelled in red, and HVR4 is labelled in black. The **base** and **tower** regions of the capsomer, as

753 well as the orientation with respect to the capsid (**outside**/**inside**) are indicated. The tower
754 region is shown in the right hand side inset as seen from outside the capsid.

755 **(b)** Sequence alignment of hexon HVRs from HAdV-C5 and HAdV-F41. Grey bars indicate
756 traced regions. The HAdV-C5 acidic region is highlighted by a black rectangle. Amino acids are
757 coloured by polarity (yellow: non-polar; green: polar uncharged; red: polar acidic and blue:
758 polar basic.

759 **(c)** Superposition of the HAdV-C5 penton base monomer (PDB ID: 6B1T Chain M, pink) and
760 HAdV-F41 coloured by RMSD from blue to red and purple as in (a). At the right hand side,
761 panels show the variable loop (top) and fibre binding rearrangement (bottom) regions with the
762 HAdV-F41 map in grey mesh.

763 **(d)** Sequence alignments focusing on the penton base (PB) N-terminal arm, variable and
764 hypervariable loops and hydrophobic rim, and on the fibre N-terminal peptide, as indicated.
765 Black rectangles highlight the RGD/IGDD sequence motifs, the conserved fibre peptide, and the
766 non-conserved Ser426 in the hydrophobic rim. Grey horizontal bars indicate regions traced in
767 the structure. For fibre, the bar indicates traced residues in the HAdV-C5 fibre (PDB ID: 3IZO);
768 suffixes **_l** and **_s** refer to the long and short fibers in HAdV-F41. The histogram indicates the
769 mean pairwise identity over all pairs in the column (green: 100% identity, olive: at least 30%
770 and under 100% identity, red: below 30% identity).

771 **(e)** The penton base pentamer is shown as viewed from outside the capsid, with one monomer in
772 black and the rest in white ribbon. Hydrophobic residues conserved with HAdV-C5 in the fibre
773 binding ring are coloured cyan (Phe412, Tyr413, Tyr419 and Leu422) or maroon (non-
774 conserved, Ser426).

775 **(f)** Penton base pentamer in HAdV-C5 and HAdV-F41 coloured by surface electrostatic
776 potential and viewed from inside the virion, showing the differences in charge in the inner
777 penton cavity.

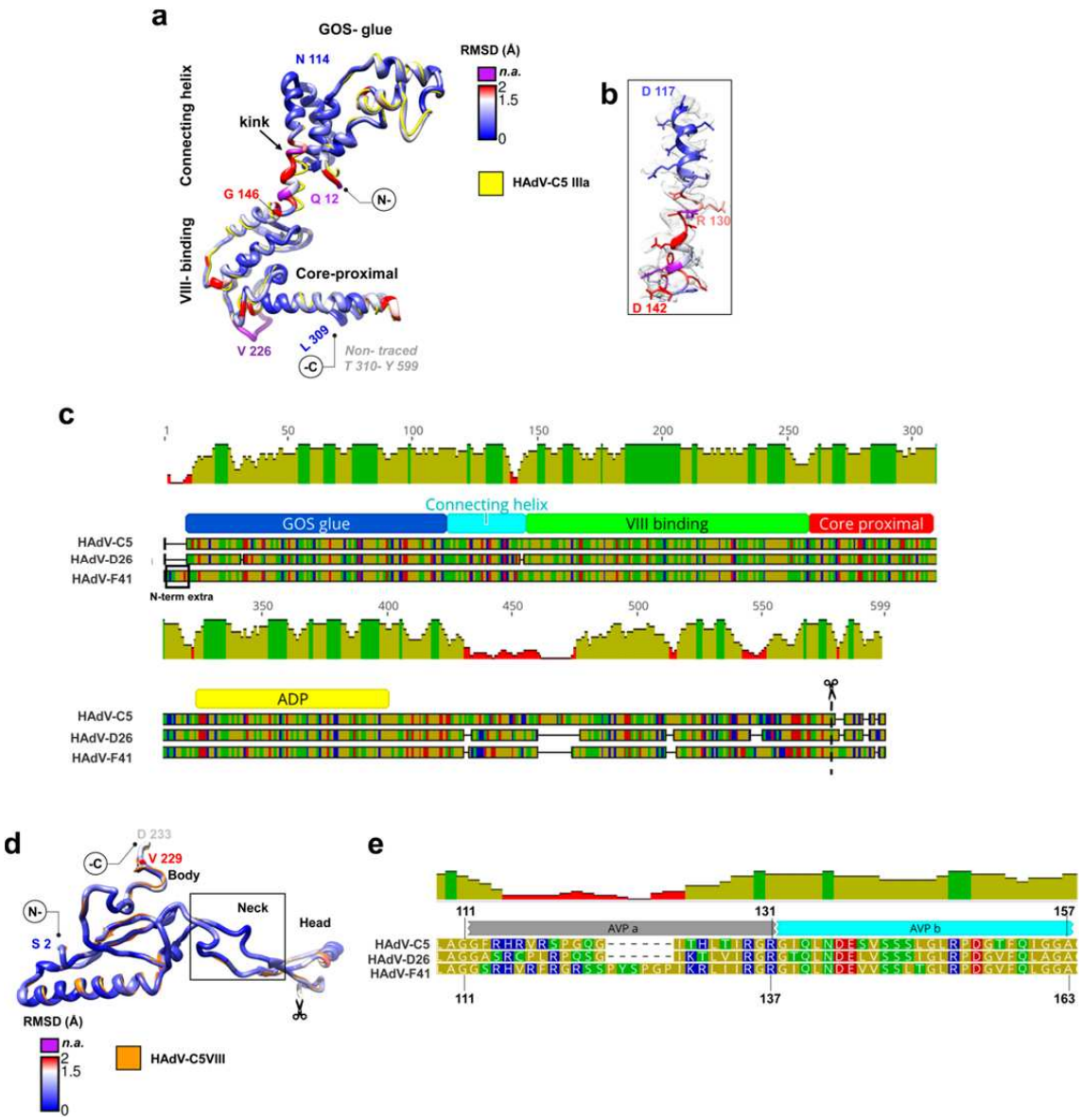


Figure 3. Proteins IIIa and VIII.

(a) Superposition of HAdV-C5 IIIa (PDB ID: 6B1T, Chain N) in yellow and HAdV-F41 IIIa coloured by RMSD as in Fig. 2. The **GOS-glue**, **connecting helix**, **VIII-binding** and **core proximal** domains are indicated.

(b) Detail of the connecting helix to show its fit into the cryo-EM map.

(c) Schematics showing the alignment of protein IIIa sequences in HAdV-C5, D26 and F41. Traced domains are indicated below, in different colours. **APD** is the appendage domain traced

786 in HAdV-D26, but not in the other two viruses. The N-terminal extension in HAdV-F41 (**N-**
787 **term extra**) and maturation cleavage site (**scissors**) are also indicated.

788 **(d)** Superposition of HAdV-C5 protein VIII (PDB ID: 6B1T Chain O) in orange and HAdV-
789 F41 VIII coloured by RMSD as above. The **body**, **neck** and **head** domains are indicated, as well
790 as the gap corresponding to the peptide cleaved during maturation (**scissors**).

791 **(e)** Sequence alignment showing the two central peptides of protein VIII cleaved by AVP
792 (**AVPa, AVPb**) in HAdV-C5, D26 and F41. Amino acid and mean pairwise identity histogram
793 colour schemes are the same as in Fig. 2.

794

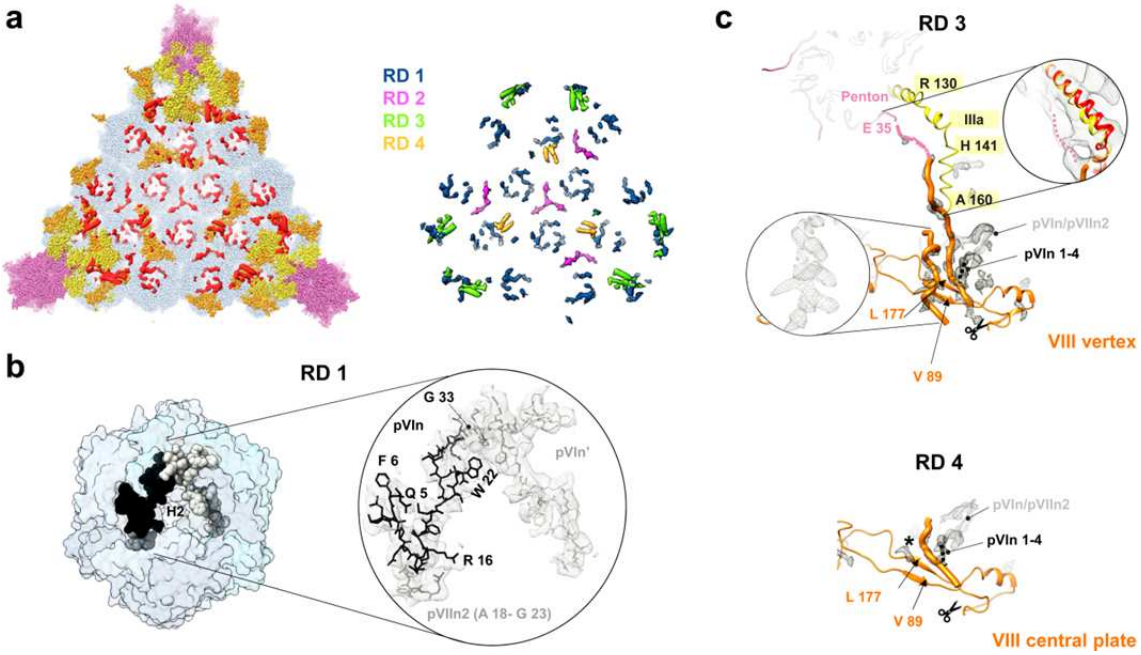


Figure 4. Additional internal components of the HAdV-F41 capsid.

(a) The left hand side panel shows the remnant densities in one facet (red), not occupied by traced proteins hexon (light blue), penton (light pink), IIIa (yellow) or VIII (orange). In the right hand side panel, components of the remnant map have been classified as RD1-4 (see text) and coloured as indicated. Densities correspond to the unsharpened map contoured at 1.5σ.

(b) N-terminal peptides of proteins VI (**pVIn**, two copies) and VII (**pVIIn$_2$**) tentatively traced in RD1, inside the hexon cavity. The surface of one hexon trimer is depicted as seen from inside the capsid.

(c) Proposed interpretation of remnant densities RD3 and RD4. Except for the inset at the right hand side, remnant density in the sharpened map is shown, contoured at 2.5σ. Top, a view of the vertex region seen from inside the capsid showing the location of RD3. Bottom, view from an equivalent point of view of the protein VIII copy located beneath the central plate of hexons. The traced part of the two copies of protein VIII (VIII vertex and VIII central plate) is shown as a thin ribbon, and the poly-Ala peptides modelled in the unassigned densities proposed to correspond to the central peptide of VIII are in orange thick worms. The inset at the left shows the possibly α-helical shape of one of the remnant densities in RD3. This density is much

812 weaker in RD4, where we have placed a black star to indicate its position. Scissors indicate the
813 AVP cleavage site in protein VIII. Peptide pVIn is represented as a grey ribbon, with the
814 possible location of its four untraced N-terminal residues as a black dotted line. In the RD3
815 panel, the IIIa connecting helix is in yellow, the first residue traced in the penton base protein
816 (E35) is colored pink, and a dotted pink line indicates the possible trajectory of the untraced 34
817 residues. In the framed zoom at the right hand side, the IIIa connecting helix is shown
818 overlapped with the HAdV-C5 helix in red, and the unsharpened map contoured at 1σ is shown
819 as a grey mesh, to emphasize low resolution density proposed to correspond to the penton base
820 N-terminal arm.

821

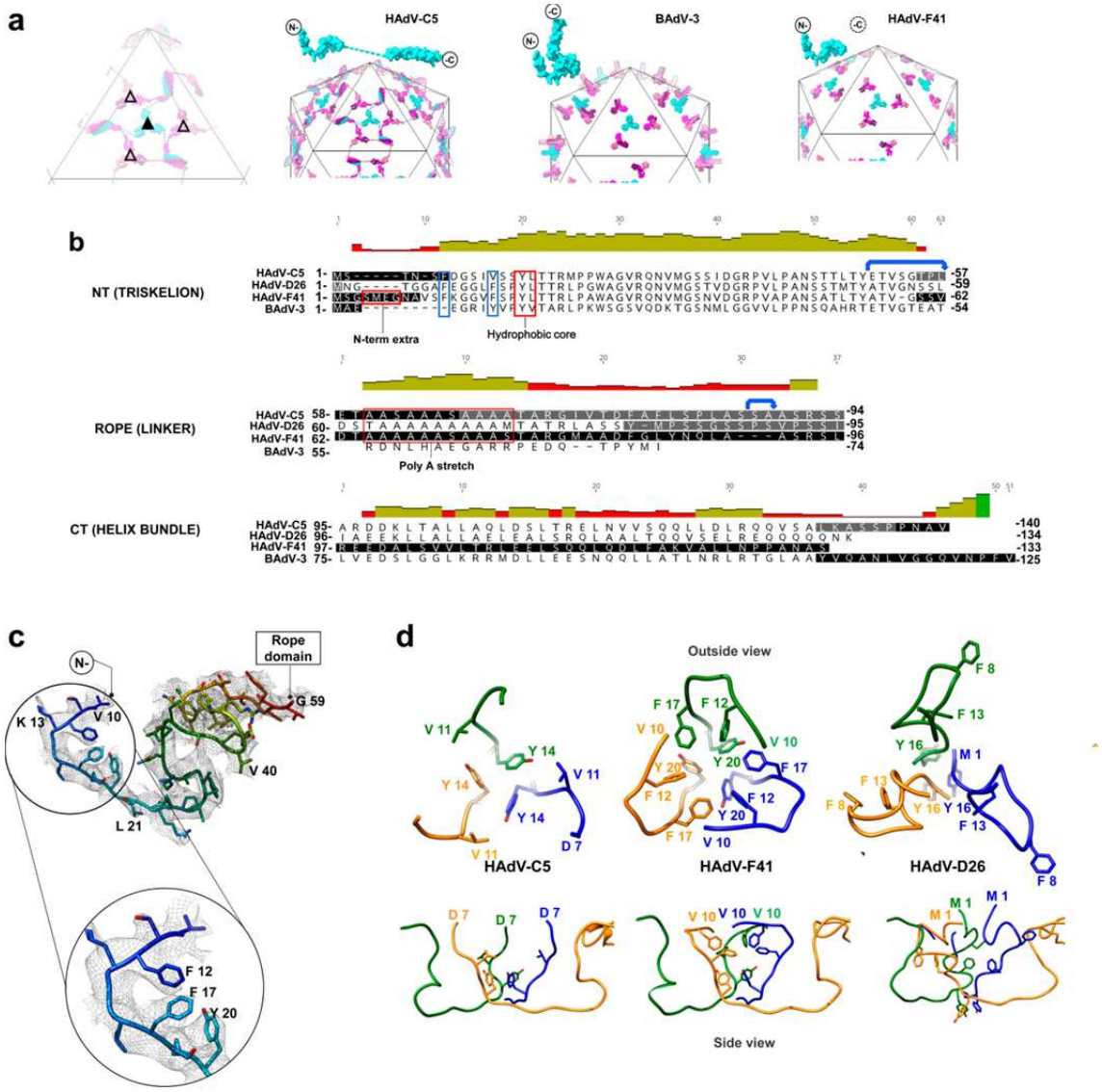


Figure 5. Protein IX triskelion structure.

(a) Schematics showing the organization of protein IX in HAdV-C5, BAdV-3 and HAdV-F41, where only the N-terminal domain is ordered. Triskelions located at the I3 symmetry axis (filled triangle) are depicted in cyan, and those located at the L3 axis (empty triangle) in several shades of pink. The rope domains of HAdV-C5 (dashed lines indicating non-modelled residues) crawl around the hexons on the surface until they reach the edge, where they form 4-helix bundles. In contrast, BAdV-3 rope domains fold back and form 3-helix bundles directly on top of their triskelions. One IX monomer is depicted in cyan on top of each schematic, with the N and C termini indicated.

832 **(b)** Structure-based sequence alignment of protein IX in HAdV-C5, HAdV-D26, HAdV-F41
833 and BAdV-3. Regions spanning the amino terminal (**NT**), **rope** and carboxy terminal (**CT**)
834 domains are indicated. Black text corresponds to residues modelled in the different structures.
835 White text on black background corresponds to regions not modelled. On grey background,
836 regions modelled, but not in all independent copies of IX. The N-terminal extra residues in
837 HAdV-F41, the triskelion hydrophobic core and the poly-Ala stretch (absent in BAdV-3) are
838 indicated with red boxes. Blue boxes highlight the two Phe residues discussed in the text. The
839 two flexible bends are indicated with blue arrows.

840 **(c)** N-terminal domain of HAdV-F41 protein IX rainbow coloured from blue (N-term) to red
841 (last traced residue), with the density map in grey mesh, and a zoom into the hydrophobic
842 residues discussed in the text. **(d)** Comparison of the triskelions in HAdV-C5 (PDB ID: 6B1T),
843 HAdV-F41 and HAdV-D26 (PDB ID: 5TX1). Top row: triskelions as seen from outside the
844 capsid. Bottom: a view rotated by 90º, so that the hexon shell surface would be located at the
845 bottom. Only the first residue traced for each virus is labelled in this view.

846

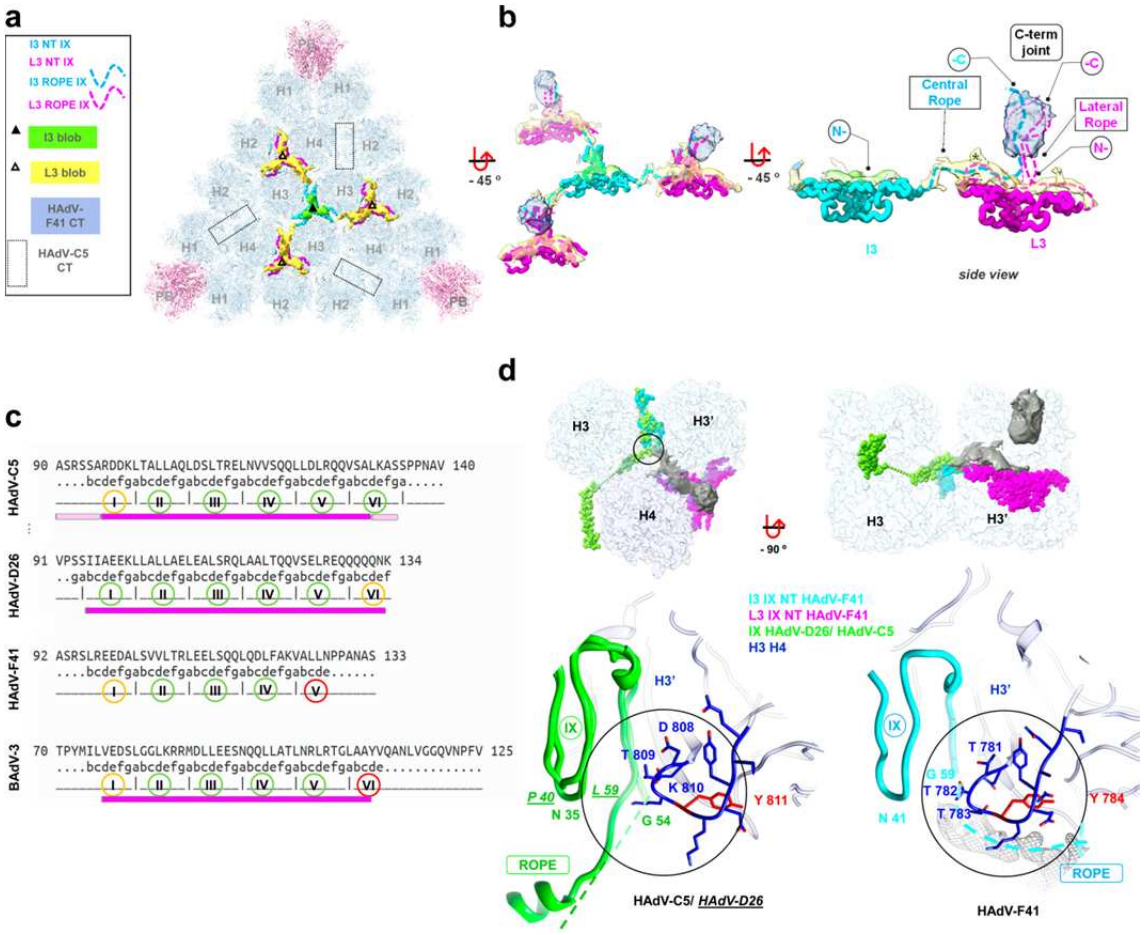


Figure 6. Organization of protein IX in the HAdV-F41 capsid.

(a) Remnant density map as seen from outside the capsid. Penton (**PB**) and hexon (**H1-H4**) models in one facet are shown in pale colours. The modelled triskelions are shown in cyan (triskelion at the I3 symmetry axis, **I3 NT**) and magenta (triskelions at the L3 symmetry axes, **L3 NT**). Remnant unassigned density corresponding to the unsharpened map contoured at 0.2σ is coloured green and yellow. The location of helix bundles in HAdV-C5 is indicated with dotted rectangles.

(b) Interpretation of the remnant map. Apart from the elements shown in (a), the density protruding at the L3 axes is shown in blue, and the proposed path for the rope domains corresponding to each triskelion is in cyan and magenta dashed lines. Note that the green density would only harbour the I3 extra N-terminal residues, whereas the yellow density would

accommodate both the extra N-term of the L3 triskelions and all rope domains. The star indicates where the I3 molecule would join the L3 one.

(c) Heptad repeat prediction for the C-terminal domain of protein IX in HAdV-C5, HAdV-D26, HAdV-F41, and BAdV-3. For each virus, the first line shows the protein sequence, and the second shows the predicted coiled-coil regions whose amino acids are classified as a heptad pattern (abcdefg). Absence of coiled-coil is denoted as a dot. In the third line, heptads are numbered with roman numbers within a green (complete repeat), yellow (almost complete) or red (incomplete) circle. The fourth row represents α-helices traced in the structures, in magenta if the region is modelled for all copies of IX in the AU, and in light pink if it is not. Note that HAdV-F41 predicted coiled-coil region is the shortest, with only three complete repeats.

(d) The rope domain of the IX molecules forming the I3 triskelion follows different paths in HAdV-C5/D26 and HAdV-F41. Top: one of the IX monomers forming the HAdV-D26 triskelion at the I3 axis is shown in green; the equivalent monomer in HAdV-F41 is in cyan; and the HAdV-F41 L3 triskelion is shown in magenta. Surrounding hexons (H3, H3' and H4) are depicted in transparent blue, and the HAdV-F41 remnant map is in grey surface. The view in the left hand side panel is as seen from outside the capsid. For clarity, H4 is not depicted in the right hand side view. Bottom: zoom in on the region where the triskelion ends and the rope domain turns. The structures of IX in HAdV-C5 and HAdV-D26 are overlapped and depicted in green. HAdV-D26 residue labels are underlined. The rope domains in HAdV-C5 and HAdV-F41 are not traced, but are represented as dashed lines. Remnant density of HAdV-F41 corresponding to its rope domain is depicted as a grey mesh. The hexon (H3) amino acids with RMSD above 2 Å are depicted in red. Note the threonine triplet in the case of HAdV-F41 (781-783), absent in HAdV-C5 and D26.

883 **List of supplementary material**

884 - Supplementary figures S1 to S8

885 - Supplementary tables S1 to S5

886 - Supplementary files S1 and S2

887 - Supplementary references

888